Universidade de Aveiro Departamento de Línguas e Culturas
2009

Catarina Alexandra Monteiro de Oliveira

Do Grafema ao Gesto

Contributos Linguísticos para um Sistema de Síntese de Base Articulatória

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Linguística, realizada sob a orientação científica do Doutor António Joaquim da Silva Teixeira, Professor Auxiliar do Departamento de Electrónica Telecomunicações e Informática da Universidade de Aveiro e do Doutor João Manuel Nunes Torrão, Professor Catedrático do Departamento de Línguas e Culturas da Universidade de Aveiro.

Apoio financeiro da FCT, no âmbito do
Programa POCTI, integrado no III
Quadro Comunitário de Apoio.

À minha irmã

Frater a fratre adiutus quasi firma civitas.

(Tomás a Celano, Vita Secunda 33.9)

Ao Ricardo

Pelas tuas mãos medi o mundo

E na balança pura dos teus ombros

Pesei o ouro do Sol e a palidez da Lua.

*(*Sophia de Melo Breyner*, Vida Dividida)*

o júri

presidente

Doutora Ana Maria Vieira da Silva Viana Cavaleiro
Professora Catedrática da Universidade de Aveiro

vogais

Doutor João Manuel Nunes Torrão
Professor Catedrático da Universidade de Aveiro (co-orientador)

Doutor Plínio de Almeida Barbosa
Professor Associado do Departamento de Linguística, Instituto de Estudos da Linguagem, Universidade de Campinas, Brasil

Doutor João Manuel Pires da Silva e Almeida Veloso
Professor Auxiliar da Faculdade de Letras da Universidade do Porto

Doutora Maria Aldina de Bessa Ferreira Rodrigues Marques
Professora Auxiliar do Instituto de Letras e Ciências Humanas da Universidade do Minho

Doutor António Joaquim da Silva Teixeira
Professor Auxiliar da Universidade de Aveiro (orientador)

agradecimentos

Durante a realização deste trabalho, contei com o apoio e colaboração de diversas pessoas e instituições, a quem não posso deixar de manifestar o meu grato reconhecimento.

Reservo o meu primeiro agradecimento ao meu orientador, Doutor António Joaquim da Silva Teixeira, a quem muito devo para além da orientação desta dissertação. O seu apoio incondicional, o seu incentivo e a sua amizade foram determinantes em todas as fases de execução desta tese, mas, sobretudo, nos momentos em que deixei de acreditar num fim possível. Estou-lhe grata pelos múltiplos ensinamentos, conselhos e sugestões; pela grande confiança que sempre em mim depositou; por me ter ajudado a explorar novos campos de investigação e conhecimento; por me ter ensinado a perseverar e a nunca desistir perante os obstáculos (e foram muitos!) ou as complexidades de um problema; por me ter propiciado o contacto com a comunidade científica a trabalhar nesta área, através de idas a congressos, cursos de Verão e deslocações a universidades estrangeiras; pela forma dedicada e paciente com que sempre atendeu às minhas dúvidas e preocupações; pelo entusiasmo e rigor científico com que acompanhou todas as etapas do meu trabalho; e, sobretudo, por nunca ter desistido de mim e da orientação da minha dissertação.

Ao meu co-orientador, Doutor João Manuel Nunes Torrão, eu agradeço o incentivo inicial para a prossecução dos estudos ao nível do doutoramento, a disponibilidade para atender aos meus pedidos e a forma pronta e decidida com que aceitou a orientação da minha tese, quando esta estava já numa fase de execução avançada.

No plano institucional, os meus agradecimentos vão para a FCT, para o Departamento de Línguas e Culturas e, sobretudo, para o IEETA.

O apoio financeiro da FCT, através de uma bolsa de doutoramento, foi absolutamente fundamental para o desenvolvimento da investigação conducente a este trabalho, viabilizando, além do mais, a minha participação em congressos e reuniões científicas, que em muito contribuiram para o meu enriquecimento científico. Através do financiamento do projecto HERON, a mesma instituição tornou possível a recolha de dados articulatórios, o acesso a meios materiais e a recursos bibliográficos, impossíveis de obter de qualquer outra forma.

Ao Departamento de Línguas e Culturas, onde iniciei a minha jornada de estudos pós-graduados, agradeço as condições materiais, que apesar das dificuldades e limitações, foram colocadas ao meu dispor.

Ao IEETA e a todos quantos lá trabalham, tenho a agradecer a forma generosa com que me acolheram, o ambiente de trabalho tranquilo que me proporcionaram e os recursos materiais que, ao longo destes anos, me disponibilizaram. O contacto com outras metodologias de trabalho e outras áreas do saber resultou numa experiência altamente enriquecedora e gratificante, com consequências importantes na minha própria forma de fazer ciência. Reservo uma palavra de agradecimento especial para o Nuno e a Anabela, profissionais dedicados, sempre prontos a auxiliar-me nas tarefas logísticas e outras necessidades; para a Rita, o Samuel e a Sara, pelos momentos de convívio e descontracção.

Obrigado, também, a todos os informantes e ouvintes que pacientemente colaboraram na recolha dos corpora e nos testes perceptuais, pois, sem eles, este trabalho não teria sido possível. Entre estes, cumpre-me destacar a Paula e a Liliana, pelo seu empenho nas referidas tarefas, mas sobretudo pela amizade e companheirismo.

Estendo este agradecimento aos meus amigos mais próximos, por estarem sempre presentes, mesmo quando fisicamente ausentes, e aos meus alunos de FAA II por manterem viva em mim a paixão pelo ensino.

Durante a realização deste trabalho contei ainda com a inestimável ajuda técnica e precioso aconselhamento de várias pessoas. À Doutora Solange Rossato, do GIPSA-LAB, em Grenoble, eu agradeço o caloroso acolhimento no referido laboratório, a ajuda na concepção experimental do trabalho de aquisição das vogais nasais, todas as sugestões e empréstimos bibliográficos. Ao Doutor Christophe Savariaux, eu fico grata pela disponibilidade e apoio técnico na recolha dos dados. Ao Doutor João Veloso, agradeço a simpatia com que me recebeu nas suas aulas, todos os ensinamentos sobre a sílaba e o interesse manifestado pelo meu trabalho. Ao Doutor Francisco Vaz, agradeço o inestimável apoio e intervenção em momentos cruciais do meu percurso de doutoramento. Finalmente, gostaria ainda de agradecer ao Doutor Plínio Barbosa, pelas trocas de impressões sobre sistemas dinâmicos e pelas várias sugestões e críticas a este trabalho.

Por último, um agradecimento muito especial à minha família, o meu suporte emocional ao longo deste atribulado percurso. À minha irmã, quero aqui agradecer a infinita paciência, a generosidade e o carinho constante, que fazem dela uma pessoa tão especial. Aos meus pais, eu devo ter chegado até aqui. Agradeço-lhes pelos valores morais que me outorgaram, pelos múltiplos sacrifícios e dificuldades que em prol da minha educação tiveram de ultrapassar e, fundamentalmente, pelo amor incondicional. Ao Ricardo, eu agradeço a paciência, a compreensão, o estímulo constante e tudo o que as palavras não conseguem expressar.

palavras-chave

Síntese articulatória, Fonologia Articulatória, Português, Vogais Nasais, Silabificação Automática, Conversão Grafema-Fone.

resumo

Motivados pelo propósito central de contribuir para a construção, a longo prazo, de um sistema completo de conversão de texto para fala, baseado em síntese articulatória, desenvolvemos um modelo linguístico para o português europeu (PE), com base no sistema TADA (*TAsk Dynamic Application*), que visou a obtenção automática da trajectória dos articuladores a partir do texto de entrada.

A concretização deste objectivo ditou o desenvolvimento de um conjunto de tarefas, nomeadamente 1) a implementação e avaliação de dois sistemas de silabificação automática e de transcrição fonética, tendo em vista a transformação do texto de entrada num formato adequado ao TADA; 2) a criação de um dicionário gestual para os sons do PE, de modo a que cada fone obtido à saída do conversor grafema-fone pudesse ter correspondência com um conjunto de gestos articulatórios adaptados para o PE; 3) a análise do fenómeno da nasalidade à luz dos princípios dinâmicos da Fonologia Articulatória (FA), com base num estudo articulatório e perceptivo.

Os dois algoritmos de silabificação automática implementados e testados fizeram apelo a conhecimentos de natureza fonológica sobre a estrutura da sílaba, sendo o primeiro baseado em transdutores de estados finitos e o segundo uma implementação fiel das propostas de Mateus & d'Andrade (2000). O desempenho destes algoritmos – sobretudo do segundo – mostrou-se similar ao de outros sistemas com as mesmas potencialidades.

Quanto à conversão grafema-fone, seguimos uma metodologia baseada em regras de reescrita combinada com uma técnica de aprendizagem automática. Os resultados da avaliação deste sistema motivaram a exploração posterior de outros métodos automáticos, procurando também avaliar o impacto da integração de informação silábica nos sistemas.

A descrição dinâmica dos sons do PE, ancorada nos princípios teóricos e metodológicos da FA, baseou-se essencialmente na análise de dados de ressonância magnética, a partir dos quais foram realizadas todas as medições, com vista à obtenção de parâmetros articulatórios quantitativos. Foi tentada uma primeira validação das várias configurações gestuais propostas, através de um pequeno teste perceptual, que permitiu identificar os principais problemas subjacentes à proposta gestual. Este trabalho propiciou, pela primeira vez para o PE, o desenvolvimento de um primeiro sistema de conversão de texto para fala, de base articulatória.

A descrição dinâmica das vogais nasais contou, quer com os dados de ressonância magnética, para caracterização dos gestos orais, quer com os dados obtidos através de articulografia electromagnética (EMA), para estudo da dinâmica do velo e da sua relação com os restantes articuladores. Para além disso, foi efectuado um teste perceptivo, usando o TADA e o SAPWindows, para avaliar a sensibilidade dos ouvintes portugueses às variações na altura do velo e alterações na coordenação intergestual. Este estudo serviu de base a uma interpretação abstracta (em termos gestuais) das vogais nasais do PE e permitiu também esclarecer aspectos cruciais relacionados com a sua produção e percepção.

keywords

Articulatory Synthesis, Articulatory Phonology, Portuguese, Nasal Vowels, Automatic Syllabification, Grapheme-to-Phoneme Conversion

abstract

Motivated by the central purpose of contributing for the construction, in the long term, of a complete text-to-speech system based in articulatory synthesis, we develop a linguistic model for European Portuguese (EP), based on TADA system (*TAsk Dynamic Application*), that aimed at the automatic attainment of the articulators trajectory from the input text.

The specification of this purpose determined the development of a set of tasks, namely the 1) implementation and evaluation of two automatic syllabification systems and two grapheme-to-phoneme (G2P) conversion systems, in view of the transformation of the input in an appropriate format to the TADA; 2) the creation of a gestural database for the EP sounds, in so that each phone obtained at the output of the g2p system could have correspondence with a set of articulatory gestures adapted for EP; 3) the dynamic analysis of nasality, on the basis of an articulatory and perceptive study.

The two automatic syllabification algorithms implemented and tested make appeal to phonological knowledge on the structure of the syllable, being the first one based in finite state transducers and the second one a faithful implementation of Mateus & d'Andrade (2000) proposals. The performance of these algorithms – especially the second - was similar to the one of other systems with the same potentialities.

Regarding grapheme-to-phone conversion, we follow a methodology based on manual rules combined with an automatic learning technique. The evaluation results of this system motivated the exploitation of others automatic approaches, finding also to evaluate the impact of the syllabic information integration in the systems.

The gestural description of the European Portuguese sounds, anchored on the theoretical and methodological tenets of the Articulatory Phonology, was based essentially on the analysis of magnetic resonance data (MRI), from which all the measurements were carried out, aiming to obtain the quantitative articulatory parameters. The several gestural configurations proposed have been validated, through a small perceptual test, which allowed identifying the main underlying problems of the gestural proposal.

This work provided, for the first time to PE, the development of a first articulatory based text-to-speech system.

The dynamic description of nasal vowels relied either on the magnetic resonance data, for characterization of the oral gestures, either on the data obtained through electromagnetic articulography (EMA), for the study of the velum dynamic and of its relation with the remaining articulators. Besides that, a perceptive test was performed, using TADA and SAPWindows, to evaluate the sensibility of the Portuguese listeners to the variations in the height of velum and alterations in the intergestural coordination. This study supported an abstract interpretation (in gestural terms) of the EP nasal vowels and allowed also to clarify crucial aspects related with its production and perception.

mots-clés

Synthèse Articulatoire, Phonologie Articulatoire, Voyelles Nasales, Division Syllabique Automatique, Transcription Orthographique-Phonétique

resumé

Guidés par l'objectif principal d'aider à bâtir, à long terme, un système complet de conversion du texte en parole, fondé sur la synthèse articulatoire, nous avons développé un modèle linguistique pour le portugais européen (PE), sur la base du système TADA (Task Dynamic Application), qui visait l'obtention automatique du mouvement des articulateurs, en partant du texte d'entrée.

La réalisation de cet objectif a dicté le développement d'un ensemble de tâches, y compris: 1) la mise en oeuvre et l'évaluation de deux systèmes de division syllabique automatique et de la transcription phonétique, en vue de modifier, selon un format adapté au système TADA, le texte d'entrée; 2) la création d'un dictionnaire gestuel pour les sons du PE, de sorte que chaque phonème obtenu à la sortie du module de transcription orthographique phonémique puisse correspondre à un ensemble de gestes articulatoires appropriés pour le PE; 3) l'analyse du phénomène de la nasalité, aux principes de la phonologie articulatoire dynamique (FA), sur la base d'une étude articulatoire et perceptive.

Les deux algorithmes de la division syllabique automatique, mis en oeuvre et testés, ont fait appel aux connaissances phonologiques sur la structure de la syllabe, le premier étant basé sur les transducteurs à états finis, dans le deuxième une application fidèle des propositions de Mateus & d'Andrade (2000). La performance de ces algorithmes – surtout le second – s'est révélée semblable à celles des autres systèmes avec le même potentiel.

En ce qui concerne la conversion du graphème-phone, nous avons suivi une méthodologie basée sur des règles de réécriture associée à une technique d'apprentissage automatique. Les résultats de l'évaluation de ce système ont conduit à la poursuite de l'exploration postérieure d'autres méthodes automatiques, cherchant également à évaluer l'impact de l'intégration de l'information syllabique dans les systèmes.

La description dynamique des sons du PE, ancrée dans les principes théoriques et méthodologiques de la FA, a été fondée principalement sur l'analyse des données de résonance magnétique, à partir de laquelle toutes les mesures ont été efectuées, afin d'obtenir des paramètres quantitatifs articulatoires. Une première tentative de validation des différentes configurations des gestes proposés a été essayée, par un petit test de perception qui a permis l'identification des principaux problèmes qui sous-tendent la proposition gestuelle. Ce travail a fourni, pour la première fois, pour le PE, le développement d'un premier système de conversion du texte en parole de base articulatoire.

La description dynamique des voyelles nasales a compté, que ce soit avec les données de imagerie par résonance magnétique pour caractériser les gestes oraux, soit avec les données obtenues grâce à l'articulographie électromagnétique (EMA), pour étudier la dynamique du voile du palais et sa relation avec d'autres articulateurs. De plus, un test de perception a été réalisé, en utilisant le système TADA et le SAPWindows pour évaluer la sensibilité des auditeurs portugais aux variations en hauteur du voile du palais et aux changements de la coordination intergestuelle. Cette étude a servi de base à l'interprétation abstraite (en termes gestuels) des voyelles nasales du PE et a également permis de clarifier des aspects cruciaux liés à leur production et à leur perception.

Conteúdo

Lista de Figuras

Lista de Tabelas

Lista das abreviaturas utilizadas

AFI Alfabeto Fonético International

ANOVA *Analysis of Variance*

ANR *Agence Nationale de la Recherche*

CASY *Configurable Articulatory Synthesizer*

CD *Constriction Degree* Grau de Constrição

CL *Constriction Location* Local de Constrição

CLUL Centro de Linguística da Universidade de Lisboa

DAVO *Dynamic Analog of the VOcal Tract*

EMA *ElectroMagnetic Articulography*

EMMA *ElectroMagnetic Midsagittal Articulography*

ESSUA Escola Superior de Saúde da Universidade de Aveiro

F0 Frequência Fundamental

FA Fonologia Articulatória

FSTs *Finite State Transducers*

G2P *Grapheme-to- Phone(me)*

HLsyn *High Level Parameter Speech Synthesis System*

HMMs *Hidden Markov Models*

HNR *Harmonics-to-Noise Ratio*

IEETA Instituto de Engenharia Electrónica e Telemática de Aveiro

LPC *Linear Predictive Coding*

LTS *Letter-to-Sound Rules*

MBL *Memory-based Learning*

MBROLA *Multi-Band Re-synthesis Overlap-Add*

MIT *Massachussets Institute of Technology*

OVE *Orator Verbis Eletricis*

PAT *Parametric Artificial Talker*

PB português do Brasil

PE português europeu

PF Português Fundamental

PLE Português Língua Estrangeira

PSOLA *Pitch Synchronous Overlap-Add*

RM Ressonância Magnética

SAMPA *Speech Assessment Methods Phonetic Alphabet*

SAPWindows Sintetizador Articulatório do Português para Windows

SPSS *Statistical Package for the Social Sciences*

SRL *Start-to-Release Lag*

TADA *TAsk Dynamics Application*

TBL *Transformation-Based Learning*

TTL *Target-to-Target Lag*

T2TL *Target2-to-Target Lag*

TTS *Text to Speech*

UA Universidade de Aveiro

UCLA *University of California, Los Angeles*

UPSID *UCLA Phonological Segment Inventory Database*

ULSID *UCLA Lexical and Syllabic Inventory Database*

VOCODER *Voice Coder*

VODER *Voice Operator Demonstrator*

Nota: Nas transcrições fonéticas e fonológicas presentes no texto (e nas tabelas), foram utilizados os símbolos e convenções do *Alfabeto Fonético Internacional* (AFI). Nas imagens e gráficos, o alfabeto fonético normalmente utilizado foi o SAMPA.

Capítulo 1

Introdução

It would be a considerable invention indeed, that of a machine able to mimic speech, with its sounds and articulations. I think it is not impossible.

Leonhard Euler (1761)

l'art peut aller jusqu'à former une machine qui articulerait des paroles semblables à celles que je prononce.

Géraud de Cordemoy (1777[1668])

O interesse pela construção de máquinas dotadas da capacidade de fala remonta à Antiguidade. Estátuas falantes de deuses e heróis míticos eram então usadas pelos sacerdotes para impressionar os seus fiéis (Cook *et alii*, 2006). O colosso de Memnon [1] atraiu milhares de peregrinos a Tebas nos primeiros séculos da Era Cristã (Liénard, 1991; Pettorino, 1999), seduzidos pela "voz", que emanava de uma das duas enormes estátuas de arenito [2].

O fascínio pela produção artificial de fala faz-se sentir sobre figuras ilustres - como Gerbert d'Aurillac (ca. 946- 1003), coroado papa (Silvestre II) no ano de 999, o mestre-regente e chanceler da Universidade de Oxford Robert Greathead ou Robert Grosseteste (1168-1253) [3], o filósofo Albertus

[1] Colosso de Memnon é a designação atribuída a duas estátuas do faraó Amen-hotep III, situadas na antiga cidade de Tebas, a oeste da cidade de Luxor, no Egipto. Com cerca de 18 metros e um peso de 1300 toneladas, as estátuas funcionariam como guardiãs do templo funerário do faraó.

[2] O fenómeno - atestado pelas 108 epígrafes esculpidas nas pernas do monumento, bem como pelas descrições de Estrabão, Pausânias e Tácito - foi, durante vários séculos, atribuído a causas naturais, mas Pettorino (1999) acredita que o efeito pode ter sido criado pelo matemático grego Heron de Alexandria, que elabora um sistema para dar voz à estátua, baseado num recipiente com água colocado sob o joelho esquerdo do monumento. O vento, canalizado por um orifício ainda hoje visível, ajudaria a recriar o efeito sonoro. O dispositivo assemelhar-se-ia a um dos muitos mecanismos sonoros descritos pelo grego no tratado de mecânica *Pneumatica*.

[3] Segundo os versos do *De Confessione Amantis* (1390), da autoria de John Gower, o filósofo e teólogo Robert Grosseteste teria dedicado sete anos da sua vida à construção de uma cabeça falante (Pettorino & Giannini, 1999).

Magnus (1193-1280) [4] ou o reputado frade franciscano Roger Bacon (1214-1294) [5] - a quem se atribui a construção de míticas *talking heads* (Mattingly, 1974; Barbosa, 2001; Pettorino & Giannini, 1999), interpretadas, pelos seus contemporâneos, como autênticos actos de *hybris* e associadas a práticas mágicas.

No século XVIII, o tema dos autómatos e das máquinas falantes sustenta acaloradas discussões filosóficas e o desafio de criar um dispositivo capaz de simular *vox humana* domina a atenção da comunidade científica. Entre as diversas personagens que acalentaram o sonho de construir uma "máquina falante", contam-se o multifacetado cientista Erasmus Darwin, cujas pesquisas incluem uma discussão sobre o mecanismo de produção dos sons e a descrição de uma "máquina falante" por ele construída (Jackson, 2005); o professor russo Christian Kratzenstein, responsável pela construção de cinco ressoadores acústicos para simulação dos sons vocálicos (Lemmetty, 1999; Schröeder, 1993; Flanagan, 1972) [6]; o clérigo Mical, que, em 1783, por ocasião da competição anual aventada pela Academia de Ciências de Paris, apresenta duas "cabeças falantes", esculpidas em bronze, dotadas da capacidade de articular um número limitado de frases (Liénard, 1991) [7], e o nobre húngaro Wolfgang von Kempelen (Lemmetty, 1999; Schröeder, 1993; Flanagan, 1972; Liénard, 1991). A "máquina falante" desenvolvida por este último, ao longo de 20 anos de investigação, impressiona pela perfeição técnica (mais do que pela qualidade do som produzido) e inspira vários cientistas, como o físico Charles Wheatstone, o inventor do telefone Alexander Graham Bell ou o matemático Joseph Faber. Considerada um marco incontornável na história da Fonética Experimental e da Síntese de Fala (Barbosa, 2001), a notável invenção de Kempelen inaugura o início de uma nova era, em que o termo "Speaking Machine", com tudo o que a designação encerra de mítico e lendário, cede lugar à "Síntese de Fala", enquanto disciplina científica (Liénard, 1991).

Outros eventos viriam ainda a marcar a pré-história da Síntese de Fala - como a apresentação do sintetizador *Voice Operator Demonstrator* (VODER), em 1939, pelo engenheiro Homer Dudley, ou a invenção do *Pattern Playback* (por Franklin S. Cooper e colegas), um dispositivo capaz de efectuar a leitura óptica de espectrogramas de banda larga desenhados numa correia transparente, transformando-os em sinal sonoro - mas o verdadeiro ponto de viragem ocorreria apenas na segunda metade do século XX, fruto do aparecimento do computador e da tecnologia digital (e, mais tarde, da Internet), bem como do desenvolvimento da Teoria Acústica da Produção de Fala, formalizada por

[4] Do dispositivo construído pelo filósofo Albertus Magnus pouco se sabe para além da história da destruição do mesmo por um dos seus alunos, S. Tomás de Aquino, que terá interpretado a invenção como uma heresia (Rubin & Vatikiotis-Bateson, 2006; Liénard, 1991).

[5] O invento de Roger Bacon, construído com a ajuda de outro franciscano, Thomas Bungey, parece não passar de uma lenda, cuja versão mais conhecida é atribuída a Robert Green (*Friar Bacon and Friar Bungay*, 1589). Na peça, o *Doctor Mirabilis* (nome por que é conhecido Roger Bacon) é o autor de uma *brazen talking head*, com o poder de falar e responder a qualquer questão sobre o futuro. Deixada ao cuidado do assistente de Roger Bacon, enquanto o monge dormia, o engenho acaba por se partir, depois de ter pronunciado a enigmática frase "Time is. Time was. Time past". (http://www.bbc.co.uk/dna/h2g2/A577523).

[6] Uma versão moderna dos ressoadores de Kratzenstein pode ser encontrada no site do *Exploratorium Vocal Vowels* (http://www.exploratorium.edu/exhibits/vocal_vowels/vocal_vowels.html).

[7] O mecanismo seria capaz de articular o seguinte diálogo, em louvor de Luís XVI: "Le Roi a donné la Paix à l'Europe"; "La Paix couronne le Roi de Gloire" (Liénard, 1991, p.19).

Fant, em 1960. O velho sonho de produzir fala de forma artificial inspira o famoso computador HAL 9000, que, no conhecido filme de ficção científica "2001: Odisseia no Espaço", realizado em 1968, pelo cineasta norte-americano Stanley Kubrik, é protagonista de várias conversas com os dois astronautas, companheiros da viagem a bordo da nave espacial *Discovery*, rumo ao planeta Júpiter (Olive, 1996)[8].

Os primeiros sistemas de síntese digitais são americanos, desenvolvidos em instituições como o *Massachussets Institute of Technology* (MIT) - e.g. *Dynamic Analog of the VOcal Tract* (DAVO) e *MITTalk* - ou os *Bell Labs* - berço do VODER - mas, a partir da década de 80, multiplicam-se as empresas e laboratórios interessados no desenvolvimento de tecnologias de síntese de fala, em diversas línguas, estimulados pela crescente procura de meios alternativos de acesso à informação. O crescimento do Processamento de Fala é, portanto, uma exigência natural do mundo global em que actualmente vivemos e onde o computador tem vindo a adquirir uma importância crescente. A fala, enquanto meio de comunicação por excelência, torna o processo de interacção entre o utilizador e a máquina, não só mais ágil e natural, como mais humano [9]. Neste sentido, a síntese (e o reconhecimento) de fala terão obrigatoriamente, mais tarde ou mais cedo, de fazer parte do protocolo de diálogo entre o homem e a máquina.

O recurso à fala é particularmente vantajoso em contexto de realização de tarefas simultâneas, quando os olhos e/ou as mãos estão ocupados e o acesso a teclados, ratos ou ecrãs se encontra comprometido. No interior do *cockpit* dos aviões, por exemplo, os avisos sonoros podem servir de complemento à apresentação visual da informação e "speech synthesis can be used to transmit important information to pilots that can be missed in the middle of other displayed information" (Teixeira, 2005, p.20). Numa linha de produção, o trabalhador pode receber instruções emitidas por um computador, enquanto executa a tarefa que lhe foi atribuída. No carro, é possível conduzir, contando com as indicações de voz de um sistema GPS.

Mas as aplicações da síntese de fala não se ficam por aqui e incluem, entre outras potencialidades, o acesso a informações através do telefone (e.g. transacções bancárias, leitura do e-mail), o ensino assistido por computador (Keller & Keller, 2000a,b; Handley & Hamel, 2005) [10], a tradução automática, o auxílio a indivíduos com deficiência - nomeadamente cegos ou amblíopes [11] e pessoas

[8] Segundo Olive (1996, p.102), "2001 presented the possibility that future computers would speak and function like human beings.".

[9] Segundo Zue (1994, apud Lajoie *et alii*, 2000, p.2), "there are a number of advantages in applying language technology to produce an interface between human and computers. Speech is natural; humans speak before they can read and write. Speech is efficient; generally 3 words or 10 phonemes are uttered per second, whereas typing speed rarely exceeds 60 words per minute. Finally, speech is flexible; it provides hands-free interaction.".

[10] Para uma revisão das aplicações da síntese de fala no âmbito do ensino assistido por computador (dicionários electrónicos, *talking texts*, ditado automático, interlocutor de um diálogo e treino de pronunciação) e do apoio a alunos com dificuldades de aprendizagem (problemas de leitura e compreensão, distúrbios de atenção, dificuldades na expressão escrita), consultar Teixeira *et alii* (2006).

[11] Segundo Sproat *et alii* (1999, p.8), "screen readers and voice browsers for the visually impaired constitute one of the best established applications of speech synthesis technology". A primeira aplicação comercial deste género terá sido a máquina de leitura para cegos, desenvolvida por Raymond Kurzweil, no final dos anos 70.

portadoras de patologias vocais [12] - sistemas de telecomunicações (Levinson *et alii*, 1993), jogos e brinquedos [13] e até síntese de línguas extintas - como o latim clássico (Bianchi, 2005; Bohlenius, 2005) - para fins relacionados com a reconstrução de cenários históricos, "for documentary television, museums, as well as university and secondary education" (Keller & Bianchi, 2002, p.2).

Como teremos oportunidade de ver em maior profundidade no segundo capítulo deste trabalho, a questão da geração automática do sinal de fala tem conhecido abordagens distintas. De entre os métodos de síntese mais utilizados, destaca-se a *síntese por regras* (ou *síntese por formantes*), cujo princípio de funcionamento se baseia na teoria fonte-filtro da produção de fala. Não obstante a boa qualidade e flexibilidade oferecida por esta técnica, a síntese de formantes tem vindo a perder terreno para a *síntese concatenativa*. A relativa simplicidade deste último método de síntese - que se fundamenta na possibilidade de concatenar segmentos de fala natural, pré-armazenados e organizados num dicionário - aliada à qualidade e naturalidade do sinal de fala gerado, em muitos casos similar ou superior à dos melhores sistemas baseados em síntese por regras, justificam o seu estatuto dominante. O actual sucesso da síntese concatenativa não significa que esta esteja isenta de limitações [14], mas tão somente que ela representa, actualmente, a melhor solução disponível. É neste sentido que vão as palavras de Santen & Sproat (1998, p.3) "we use concatenative synthesis because that is currently the best available method to produce synthetic speech of consistently high quality (...) we also believe that in the long run concatenative synthesis is not the answer". A maioria dos investigadores acredita mesmo que, a longo prazo, a resposta aos problemas da síntese de fala - mais do que pela síntese por concatenação - deverá passar pela *síntese articulatória* [15].

Ao contrários de outros métodos, a síntese articulatória preocupa-se em modelar, de uma

[12] O desenvolvimento de sistemas de síntese de fala contribui poderosamente para aumentar a capacidade de comunicação dos indivíduos portadores de patologia vocal (provocada por afasia, laringectomia, paralisia cerebral, acidente vascular cerebral ou outros), com efeitos muito positivos, do ponto de vista psicossocial, embora alguns problemas relacionados, quer com as especificidades da doença, quer com a qualidade do sintetizador, possam impedir a sua utilização generalizada. O relativo insucesso de muitos sistemas *Text to Speech* (TTS) pode ser explicado, pelo menos em parte, pela falta de qualidade e naturalidade da voz sintética produzida. Mas, para além desta exigência capital, para o uso generalizado deste tipo de tecnologia, contribui a possibilidade de ajuste das características da voz ao usuário (Klatt, 1987; Lemmetty, 1999), no sentido de disponiblizar vozes ao gosto do consumidor, em termos de género, idade e mesmo de dialecto. Os utilizadores destes sistemas "can be frustrated by an inability to convey emotions such as urgency or friendliness by voice" (Klatt, 1987, p.779). Segundo Pardo (2004, p.8), "estas personas requieren que la voz que utilizan para comunicarse refleje los estados emocionales de ellos mismos: tristeza, enfado, alegría, sorpresa, etc. Y no que la voz siempre tenga un estado de ánimo neutral".

[13] Na origem dos *talking toys*, que vão desde as bonecas falantes até às modernas consolas de jogos, está o revolucionário *Speak & Spell*, o primeiro de um conjunto de brinquedos electrónicos criados pela *Texas Instruments*, no final dos anos 70 (Frantz & Wiggins, 1982).

[14] As principais desvantagens da síntese por concatenação estão relacionadas com as distorções introduzidas no sinal em consequência do processo de concatenção - cujo efeito pode ser atenuado, mas jamais eliminado na totalidade - e, sobretudo, com a falta de flexibilidade, que se traduz numa dependência excessiva (e, portanto, indesejável) do material pré-gravado e na impossibilidade de proceder a alterações prosódicas profundas, como aquelas que permitiriam, por exemplo, veicular emoções como a alegria ou a raiva.

[15] Num questionário acerca do futuro da Síntese de Fala, realizado por Benoît & Pols (1995), metade dos 38 investigadores inquiridos aposta no crescimento da área do *articulatory modelling*.

forma directa, não as características acústicas do sinal de voz, mas o sistema de produção de fala em si mesmo, ou seja, tem como meta a simulação fiel de todas as manobras físicas realizadas pelo aparelho fonador, no sentido da obtenção da mensagem sonora. Trata-se de uma área de estudos bastante promissora (Shadle & Damper, 2001), que tem conhecido notáveis avanços nas últimas duas décadas, mas ainda com um logo caminho a percorrer até que se possa afirmar como uma alternativa às demais técnicas de síntese. Sabe-se hoje que os articuladores implicados na fonação estão em permanente movimento, mas o conhecimento acerca do processo de produção de fala, enquanto sistema dinâmico, é, ainda, muito limitado, não obstante o advento de novas técnicas para medição da geometria do tracto vocal (e.g. *ElectroMagnetic Articulography* (EMA), ultrasonografia, Ressonância Magnética (RM) tridimensional). Na prática, a aquisição de dados úteis à determinação de regras e modelos articulatórios é ainda uma tarefa muito complicada de executar (Taylor, no prelo).

Para além do estudo pormenorizado da dinâmica dos articuladores, outros desafios se impõem actualmente aos investigadores interessados no desenvolvimento de sistemas de síntese de base articulatória, como, por exemplo, a simulação e controle da fonte sonora (sobretudo ruído de fricção e efeitos de turbulência) ou o estudo do papel da percepção na selecção dos gestos articulatórios que podem ser produzidos. Paul Taylor resume assim as dificuldades que o tema continua a suscitar (Taylor, no prelo, p.422): “Mimicking the human system closely can be very complex and computational intractable.”.

Tendo em vista estas dificuldades - e apesar das eventuais potencialidades da síntese articulatória - continua a ser escasso o número de sintetizadores desenvolvidos com base nesta técnica. Um deles “fala” português, chama-se Sintetizador Articulatório do Português para Windows (SAPWindows), nasceu na Universidade de Aveiro (UA) e representa um rumo novo na história da síntese de fala em Portugal, até então exclusivamente vocacionada para a síntese de formantes (Oliveira, 1996) e, mais recentemente, para a síntese por concatenação (e.g Carvalho *et alii*, 1998). Originalmente desenvolvido para a síntese dos sons nasais (Teixeira, 2000), o sistema tem vindo a sofrer várias actualizações, nomeadamente a adaptação do sintetizador a um ambiente computacional de utilização mais generalizada, o *Windows*, e criação de uma interface amigável, que permita uma fácil interacção com o utilizador (Silva, 2001), ou a evolução de alguns modelos, no sentido de incluir a possibilidade de síntese de outras classes de sons, como as fricativas (Teixeira *et alii*, 2005). Longe de estar resolvido - como haveremos de constatar, em momento oportuno deste trabalho, pela fraca qualidade do som gerado - o problema da síntese das fricativas depende da obtenção de novos dados articulatórios (estáticos e dinâmicos), que permitam, de algum modo, melhorar o conhecimento acerca do processo de produção dos referidos sons.

O trabalho de recolha de dados articulatórios para validação e aperfeiçoamento dos modelos utilizados - iniciado no ano 2000 com a recolha de um *corpus* através de *ElectroMagnetic Midsagittal Articulography* (EMMA) para o estudo das vogais nasais (Teixeira, 2001) - foi retomado recente-

mente, no âmbito do projecto HERON [16], e culminou na aquisição e análise de uma base de dados de RM. Uma outra questão, também equacionada neste projecto - cujo nome homenageia o conhecido arquitecto da Grécia antiga, alegadamente responsável pela construção do mecanismo que dá "voz" a uma das estátuas de Memnon (Pettorino & Giannini, 1999; Pettorino, 1999) - relaciona-se com a necessidade de estimar automaticamente os parâmetros do sintetizador articulatório a partir do texto. O problema da obtenção da trajectória dos articuladores a partir de níveis mais abstractos de representação e o interesse em modelar as complexas relações entre a representação fonológica e a implementação articulatória estão também na génese desta dissertação.

A possibilidade de contribuir para a evolução do sintetizador articulatório da UA, através do desenvolvimento de um método automático de controle dos articuladores - que, a longo termo, viabilizará a construção de um sistema completo de conversão de texto para fala (em inglês, *text-to-speech system* ou simplesmente *TTS*), para o português europeu, baseado em síntese articulatória - constitui a principal motivação para a realização desta pesquisa, cujos propósitos vão de encontro às questões levantadas no seio do projecto HERON. Mas mais do que um importante estímulo à concretização deste trabalho, a participação no referido projecto garantiu o acesso a dados articulatórios - imprescindíveis ao desenvolvimento do modelo de produção de base gestual e à investigação acerca a organização temporal das vogais nasais do português - e, acima de tudo, permitiu o contacto com profissionais oriundos de áreas científicas distintas, habituados a metodologias de trabalho e instrumentos de validação diferentes, mas dispostos a trabalhar em conjunto para um fim comum. A aventura de integração de um linguista numa equipa multidisciplinar tem tanto de estimulante, como de exigente e implica uma atitude de abertura em relação a métodos e técnicas de disciplinas mais afeitas às Ciências e Tecnologias, para além de uma forte disposição para adquirir novos conhecimentos, nem sempre fáceis de assimilar, em boa parte, devido à actual política académica, que determina a separação entre os estudos humanísticos e tecnológicos. A possibilidade de trabalhar numa área pluri-inter-disciplinar como a Síntese de Fala, uma ciência no cruzamento de outras ciências - entre as quais se contam a Engenharia, a Linguística ou a Matemática - foi outro dos factores que motivou a elaboração deste trabalho e resultou numa experiência altamente enriquecedora, embora marcada por uma série de obstáculos, essencialmente relacionados com a já referida falta de formação interdisciplinar, que dificulta a efectiva integração dos linguistas em equipas mistas. Cabe notar que os planos de estudos vigentes nas principais faculdades de letras portuguesas, tanto ao nível da formação inicial, como do ensino pós-graduado, não favorecem o desenvolvimento de uma série de competências básicas, consideradas essenciais à investigação interdisciplinar [17], uma situação que gostaríamos de ver alterada, em prol do

[16] O projecto HERON - *A Framework for Portuguese Articulatory Synthesis Research* - financiado pela Fundação para a Ciência e a Tecnologia (POSC/ PLP/57680/2004) decorreu no período compreendido entre 2005 e 2007, sob a coordenação do Prof. Dr. António Teixeira, e teve como principais objectivos: 1) a aquisição de dados de produção para validação e desenvolvimento dos modelos articulatórios; 2) a evolução do sintetizador articulatório da Universidade de Aveiro, tendo em vista a síntese de outras classes de sons, nomeadamente as fricativas; 3) o desenvolvimento de módulos linguísticos, mediante a exploração de novas teorias linguísticas (como a Fonologia Articulatória), visando a construção, a longo prazo, de um sistema completo de síntese de fala a partir do texto, baseado em síntese articulatória. Para mais informações sobre este projecto, consultar www.ieeta.pt/heron.

[17] O actual responsável pela investigação em tecnologias de fala da Microsoft define assim o perfil do linguista com aspira-

acesso dos linguistas a áreas, que, entre outras vantagens, lhes poderiam oferecer saídas profissionais alternativas.

Como resultará claro do exposto anteriormente, o propósito central de desenvolver um conjunto de módulos linguísticos, passíveis de integração com o sintetizador articulatório SAPWindows, que subsidiem fazer a passagem do texto ao som, inscreve a nossa pesquisa no domínio disciplinar da Linguística Aplicada à Síntese de Fala, em que os pressupostos teóricos da Linguística se fundem com conhecimentos e metodologias advenientes de outras áreas do conhecimento. Isto significa que à Linguística fomos pragmaticamente colher todas as contribuições que se nos afiguraram úteis para a resolução de cada um dos problemas enfrentados, independentemente da corrente teórica, praticando uma estratégia que se poderia denominar de "hibridismo metodológico". O módulo de silabificação automática, por exemplo, toma como ponto de partida um algoritmo de inspiração generativa (Mateus, 1994; Mateus & d'Andrade, 1998, 2000). Já o modelo de controle do sintetizador articulatório é totalmente baseado na Fonologia Articulatória (FA) de Browman & Goldstein (e.g. 1986, 1989, 1990b, 1992). No estudo experimental acerca da organização temporal das vogais nasais, apresentado no capítulo 5, foram adoptados procedimentos oriundos do campo da Fonética Experimental, a que se reconhece o mérito de nos fazer compreender melhor a necessidade de sustentar todas as considerações acerca das vogais nasais (e sons em geral) num estudo rigoroso dos dados empíricos, um objectivo que almejámos atingir também na descrição gestual apresentada no capítulo 4. De outras disciplinas, nomeadamente da Engenharia, recebemos ensinamentos importantes, como, por exemplo, a importância de avaliar, sempre que possível quantitativamente, todos os procedimentos adoptados. Daí que todos os módulos implementados tenham sido sujeitos a validação, mediante testes de desempenho ou testes de percepção [18].

O problema eminentemente teórico da relação entre as unidades fonológicas (discretas e invariantes) e as suas manifestações articulatórias e acústicas (contínuas e dependentes do contexto) representa uma das preocupações centrais da Linguística e está na origem de um pequeno conjunto de modelos dinâmicos de produção (Saltzman & Munhall, 1989; Browman & Goldstein, 1992; Nam & Saltzman, 2003; Saltzman *et alii*, 2008; Birkholz *et alii*, 2006; Birkholz, 2007a; Kröger & Birkholz, 2007; Birkholz *et alii*, 2007), que procuram unificar os dois níveis de representação - o fonológico e o fonético - elegendo o *gesto articulatório* como unidade básica de produção.

O conceito de gesto, não sendo completamente novo na literatura fonética, foi redefinido no âmbito da FA (e.g. Browman & Goldstein, 1986, 1989, 1990b, 1992), e identifica-se, em termos gerais, com a formação de uma determinada constrição num dos subsistemas do tracto vocal (e.g.

ções nesta área: "... a successful phonetician working on a spoken language system will need some knowledge of computers, algorithms, statistics and signal processing (...) Also desired is proficiency with common computing environments such as Windows, UNIX and Macintosh, text editors, and speech analysis packages." (Acero, 1995, p.175).

[18] Uma metodologia semelhante, assente na avaliação dos vários módulos do sistema, foi adoptada num trabalho de doutoramento recente, desenvolvido na área da Linguística aplicada à Síntese de Fala (Braga, 2008).

oclusão labial para a realização da consoante /b/), através da acção coordenada de um conjunto de articuladores (e.g. mandíbula, lábio inferior e lábio superior, no caso do gesto de oclusão labial).

Enquanto "átomos" de um sistema combinatório, os gestos combinam-se entre si, no sentido de formar "estruturas moleculares" de nível superior - sejam elas segmentos, sílabas ou palavras - podendo sobrepôr-se uns aos outros total ou parcialmente. As relações de coordenação entre os gestos são formalizadas em "pautas gestuais" (*gestural scores*), diagramas a duas dimensões, onde cada gesto é representado por uma pequena caixa, cujas dimensões vertical e horizontal traduzem, respectivamente, a magnitude e o tempo de activação (duração).

De entre os modelos dinâmicos que elegem o gesto articulatório como primitivo de análise, destaca-se o auto-intitulado *task-dynamic model of speech production* (Saltzman & Munhall, 1989; Browman & Goldstein, 1992; Nam & Saltzman, 2003; Saltzman *et alii*, 2008), em desenvolvimento nos Laboratórios Haskins.

A implementação computacional deste modelo subdivide-se em três níveis funcionalmente distintos, mas interligados entre si. Em primeiro lugar, o *modelo gestual* transforma o texto de entrada numa pauta gestual, que especifica os intervalos de activação e as relações de coordenação entre os gestos que compõem um determinado enunciado. Na versão actual do modelo, quer as variáveis de activação, quer a coordenação intergestual são determinadas por um conjunto de *gestural planning oscillators*. Seguidamente, o modelo *task-dynamic* gera as trajectórias individuais de cada um dos articuladores associados a uma determinada variável do tracto, com base na informação contida na pauta gestual, nomeadamente os parâmetros dinâmicos invariantes (e.g. *target*, *damping ratio* e *stiffness*), que definem os gestos. No final, o modelo integra um *sintetizador articulatório strictu sensu*, que, a partir da posição dos articuladores, calcula as funções de área do tracto vocal, as frequências de ressonância, as larguras de banda e, finalmente, produz o sinal acústico.

A recente disponibilização do sistema *TAsk Dynamics Application* (TADA) - um *software* que implementa o referido modelo dinâmico para os sons do inglês (Nam *et alii*, 2004) - viabilizou o desenvolvimento de um modelo articulatório dinâmico para o português europeu (PE), através da criação de um novo dicionário gestual, específico para esta língua, entre outras adaptações, relacionadas, por exemplo, com a transcrição fonética e silabificação automática do texto de entrada ou a integração das vogais nasais no sistema. O mesmo é dizer que fizemos incidir a nossa atenção sobre o primeiro componente do modelo computacional - o modelo gestual - empreendendo um conjunto de actividades, que visaram a obtenção de uma pauta gestual para as palavras do PE. Como afirmámos já, a pauta gestual não é mais do que um conjunto de instruções idealizadas para os articuladores, necessitando, portanto, de ser interpretada e transformada em trajectórias reais. Cabe ao segundo módulo do sistema executar as referidas instruções da pauta gestual e calcular o movimento dos articuladores, de acordo com os princípios da *task-dynamics*, que, por serem universais, não necessitam de qualquer tipo de adaptação para que possam funcionar para outras línguas.

Inscrevendo-se a nossa pesquisa no âmbito disciplinar da Linguística, não é (e nem poderia ser) nosso propósito implementar um novo modelo de produção de fala ou sequer introduzir alterações significativas nos modelos existentes - uma tarefa que vai muito além das nossas competências e capacidades e que caberá aos Engenheiros e especialistas em Sistemas Dinâmicos operacionalizar -, mas descrever e analisar o léxico do PE, à luz de um modelo dinâmico, de base gestual, tendo em vista a obtenção de pautas gestuais a partir do texto escrito, e contribuindo, desta forma, para o desenvolvimento, a longo prazo, de um sistema TTS, de base articulatória.

Tendo em mente os objectivos e limitações expostos até aqui, socorremo-nos de um programa computacional que, ao mesmo tempo que implementa os princípios e conceitos da FA, oferece a possibilidade de modelar dinamicamente o léxico de outras línguas, para além do inglês, através da criação de novos dicionários gestuais - e, em última instância, da manipulação, *a posteriori*, dos parâmetros dinâmicos, que regem a construção das pautas gestuais - sem que sejam necessárias modificações de cariz mais técnico, mais concretamente a implementação de alterações ao modelo *task-dynamics*, que não estaríamos habilitados a concretizar. Esta abordagem permitiu-nos: 1) averiguar até que ponto o modelo gestual, proposto pela FA, é adequado para representar os segmentos do PE, em geral, e para dar conta do processo de nasalização, em particular e 2) identificar alguns problemas inerentes a este mesmo modelo. Reconhecer a existência de fragilidades, não significa negar a eficácia desta ferramenta. Ao contrário, julgamos até ter reunido algumas pistas e argumentos em favor de uma tratamento dinâmico do PE, sobretudo no tocante ao fenómeno da nasalidade, em relação ao qual empreendemos um estudo de produção mais aprofundado. Outros dados há, que parecem, contudo, interrogar e até desafiar as capacidades actuais do modelo. O caso da interpretação gestual dos róticos é - como teremos oportunidade de analisar, embora sem o grau de profundidade que gostaríamos - sintomático da necessidade de revisão de alguns parâmetros do modelo e sua substituição por outros mais adequados.

A concretização dos objectivos gizados anteriormente implicou a realização de um conjunto de tarefas, que passamos a apresentar.

A primeira actividade prende-se com a silabificação automática e transcrição fonética do texto de entrada e responde a objectivos eminentemente práticos. Na versão original do TADA, a transcrição fonética do *input* é assegurada pela associação da palavra a transcrever (em formato ortográfico normalizado) com a entrada de um dicionário. Uma vez que este não inclui informação sobre as fronteiras da sílaba, é aplicado um algoritmo, com vista à silabificação automática das entradas do léxico. Nos casos em que a palavra (ou logátomo) não consta do dicionário, esta pode ser directamente introduzida no sistema, já em formato fonético e com referência explícita às fronteiras silábicas.

No caso do PE, a ausência de recursos similares - nomeadamente um dicionário anotado foneticamente e um sistema de silabificação automática - que possam servir eficazmente os nossos propósitos (e os interesses da restante comunidade científica que se dedica ao processamento computacional do português), determinou o desenvolvimento de um conjunto de ferramentas, que visaram

a transcrição fonética e divisão silábica das palavras portuguesas a processar pelo TADA. Para além disso, a informação silábica é essencial ao correcto funcionamento do modelo gestual, baseado em osciladores acoplados.

Apesar de acessórios, na medida em que não fazem parte do núcleo central do sistema computacional, e, em última análise, dispensáveis, pois, como já referimos, é sempre possível introduzir a informação requerida directamente no sistema, estes dois procedimentos são muito importantes para o cumprimento do nosso propósito de construir, num futuro que se quer próximo, um sistema completo de conversão de texto em fala, baseado em síntese articulatória. Ainda que cumpram um objectivo muito concreto - relacionado, como já vimos, com a transformação do texto de entrada num formato adequado ao TADA - os módulos desenvolvidos funcionam como produtos independentes do sistema que nos serve de base, podendo facilmente ser integrados num qualquer sistema TTS.

Como veremos mais adiante, a conversão grafema-fone - aqui entendida em sentido amplo, no sentido em que inclui não só a transcrição ortográfico-fonética, mas também a divisão silábica e a determinação do acento lexical - é uma das etapas clássicas do processo de síntese de fala a partir do texto e uma das mais férteis no que à incorporação de conhecimentos linguísticos diz respeito, embora, nos últimos anos, a abordagem linguística tenha vindo a perder terreno para as técnicas de aprendizagem automática, baseadas em grandes *corpora* de fala. Longe de estar esgotado, o tema continua a atrair a atenção da comunidade científica, sendo que, no plano nacional, é possível contar já com várias soluções para muitos dos problemas implicados na conversão grafema-fone. Contudo, ao que foi possível apurar, a grande maioria dessas ferramentas não é do domínio público e/ou não são conhecidos os resultados da sua avaliação, o que, por si só, justifica o nosso investimento no desenvolvimento dos referidos módulos.

Como convém a uma tese em Linguística, optámos, em termos gerais, por uma abordagem guiada por critérios fonéticos e fonológicos. Os algoritmos de divisão silábica fazem apelo a conhecimentos de ordem fonológica sobre a estrutura da sílaba em PE, sendo que o segundo método descrito procura mesmo ser uma implementação fiel do algoritmo de silabificação de base proposto por Mateus & d'Andrade (2000). No tocante à transcrição grafema-fone, seguimos, numa primeira fase, uma metodologia baseada em regras de reescrita - tirando partido da elevada regularidade da ortografia portuguesa - combinada com uma uma técnica de aprendizagem automática. Esta abordagem, que se pretende complementar a outras propostas similares (e.g. Oliveira *et alii*, 1991; Viana *et alii*, 1991; Oliveira, 1996; Teixeira, 2004; Braga *et alii*, 2006; Braga, 2008), é, ao mesmo tempo económica, na medida em que reduz substancialmente o número de regras necessárias, e eficaz, já que muitos dos erros decorrentes da aplicação das regras são, numa segunda passagem, corrigidos pelo método de aprendizagem automática. Com efeito, esta solução viabilizou o rápido desenvolvimento de um primeiro sistema de fonetização automática, com um desempenho minimamente aceitável. Motivados pelos resultados desta primeira avaliação, enveredámos, numa segunda fase, pela exploração de outros métodos automáticos, em formato isolado ou combinados entre si, procurando igualmente avaliar o impacto da utilização de informação silábica sobre o desempenho global do sistema.

A segunda tarefa realizada no âmbito deste estudo está relacionada com a criação de um dicionário gestual, específico para o PE, com base no formato pré-definido pelo TADA. Cada um dos segmentos gerados pelo módulo de transcrição automática foi associado a um conjunto de gestos articulatórios, adaptados para o PE. Assim, a oclusiva velar [k], por exemplo, passa a ser representada por um gesto glotal de abertura em total sincronismo com um gesto de corpo da língua fechado na região velar.

A descrição gestual dos segmentos do PE, ancorada nos princípios teóricos da FA, contou, numa primeira aproximação, com os dados de produção disseminados na literatura fonética sobre o português. Face à escassez de estudos articulatórios, socorremo-nos também, sempre que se justificou, quer de dados acústicos - já que a partir deles podem ser feitas inferências sobre a configuração do tracto vocal - quer de descrições relativas a outras línguas.

Todas as informações recolhidas foram, posteriormente, confrontadas com os dados de ressonância magnética, adquiridos no âmbito do projecto HERON, a partir dos quais foram realizadas todas as medições, com vista à obtenção de parâmetros articulatórios quantitativos.

Em casos muito particulares - em que se verifica a mais absoluta ausência de informação articulatória ou o volume de dados é particularmente reduzido - foi a configuração original, relativa ao inglês, que nos serviu de base, legitimada quer pela observação impressionista dos contornos simulados com recurso ao TADA e ao SAPWindows, quer pela apreciação informal e iterativa da qualidade do som sintetizado, a partir de um primeiro conjunto de gestos.

A adaptação do modelo gestual ao PE, com vista à obtenção de pautas gestuais, implica, como foi já referido, não só a definição dos gestos associados a um determinado item lexical, mas também a especificação das relações de coordenação estabelecidas entre eles. A (quase) total ausência de dados acerca da organização temporal do PE comprometeu a tarefa de determinação do tipo de relação temporal estabelecida entre dois gestos consecutivos. Na impossibilidade de recolher dados relativos a esta matéria para todos os segmentos do PE (ou pelo menos para os que estão associados a uma configuração gestual complexa) e avaliar simultaneamente a influência da sua filiação silábica - uma tarefa que não seria viável no âmbito de uma única dissertação de doutoramento, naturalmente sujeita a constrangimentos temporais, e que terá que resultar forçosamente do empenho de uma equipa de investigadores, munida dos meios materiais e humanos adequados - optámos por manter, salvo pequenas alterações apontadas oportunamente, os princípios de composição gestual previstos para o inglês americano. Esta solução, não sendo a ideal, afigurou-se como a única opção metodológica ao nosso alcance, tendo em mente o objectivo de desenvolver um primeiro modelo articulatório para o PE, baseado na fonologia gestual de Browman e Goldstein.

Face às limitações expostas a propósito da coordenação gestual, decidimos circunscrever a nossa análise às vogais nasais. A escolha do fenómeno linguístico a estudar foi determinada por três premissas fundamentais: características do SAPWindows; nasalidade enquanto fenómeno característico do português; possibilidade de aquisição de novos dados EMMA.

Como foi já referido, o sintetizador articulatório da UA está especialmente vocacionado para a síntese de sons nasais, sendo que as simulações e os testes perceptuais já efectuados revelam resultados muitos promissores quanto à qualidade do som gerado (Teixeira, 2000; Teixeira *et alii*, 2005), o que nos garante a possibilidade de testar as nossa proposta, sem quaisquer constrangimentos advenientes de problemas com o sintetizador.

Ao contrário do inglês (em que a nasalidade não é usada com função distintiva e, por conseguinte, não faz parte do *cahier de charges* da Fonologia Articulatória), a nasalidade é um dos traços mais marcantes do sistema vocálico do português (Strevens, 1954) e nunca a nossa proposta de caracterização gestual do português poderia ser considerada completa - o que não significa dizer acabada - sem que este fenómeno fosse tido em linha de conta.

Para além do *corpus* de RM, que não foi especificamente desenhado para os objectivos desta dissertação, a participação no projecto HERON - e subsequentes contactos com investigadores franceses (do *GIPSA- LAB, Université Stendhal*, Grenoble), interessados em estudar o fenómeno da nasalidade nas suas várias vertentes - proporcionou-nos a oportunidade de desenhar e recolher um *corpus*, usando EMMA, que visou, simultaneamente, responder às nossas necessidades de obtenção de informação dinâmica sobre o comportamento do velo e a sua relação com os restantes articuladores e possibilitar a realização de estudos comparativos sobre a nasalidade do francês e do português. Numa fase preliminar da pesquisa, chegámos ainda a recorrer a um outro acervo de dados EMMA, adquirido no âmbito do projecto "Síntese Articulatória do Português" (Teixeira, 2001), mas cedo chegámos à conclusão que este não cumpria alguns dos requisitos essenciais à prossecução do estudo. Esta experiência revelou-se, contudo, fundamental, no que toca, quer à definição dos contextos a incluir no novo *corpus*, quer à aferição da metodologia a utilizar na análise do mesmo.

1.1 Estrutura da dissertação

A estrutura desta dissertação espelha as diferentes tarefas realizadas para alcançar os objectivos traçados. Depois desta Introdução, o trabalho está organizado em cinco capítulos.

O **capítulo II** - intitulado "Fundamentos Teóricos e Arquitectura do Sistema" - inclui uma revisão de todas as questões teóricas (algumas já introduzidas nesta Introdução), consideradas fundamentais, quer para a contextualização do tema geral, quer para a caracterização do modelo fonológico de base gestual, e uma descrição da arquitectura do sistema de produção do PE, desenvolvido a partir do TADA.

Após uma breve introdução ao conceito de síntese de fala, segue-se uma secção dedicada à descrição da arquitectura geral dos sistemas TTS, que, de um modo geral, inclui um bloco de Processamento de Linguagem Natural e um bloco de Processamento de Sinal (Dutoit, 1997). No âmbito da apresentação do primeiro componente, foram contemplados vários aspectos relacionados com a representação fonológica do texto escrito, com especial destaque para a conversão grafema-fone, um

tema que centralizará a nossa atenção na primeira fase do estudo subsequente. Quanto à etapa de processamento de sinal, daremos a conhecer as diferentes estratégias actualmente utilizadas para a geração de fala sintética, nomeadamente a síntese de formantes, a síntese baseada na concatenação de unidades e a síntese articulatória. Depois da caracterização de cada um destes métodos em particular, as secções seguintes centram-se na descrição do sintetizador articulatório SAPWindows e das adaptações introduzidas posteriormente em alguns dos seus modelos, nomeadamente as que visaram a integração do sintetizador com o sistema TADA. Este último modelo computacional - ponto de partida para o desenvolvimento de um modelo articulatório do PE - procura implementar, em termos genéricos, os princípios da Fonologia Articulatória, pelo que uma parte importante deste capítulo foi inevitavelmente dedicada à revisão dos pressupostos teóricos subjacentes a este modelo linguístico, até agora praticamente desconhecido dos investigadores portugueses. Quanto a este aspecto em particular, começámos por nos ocupar do conceito de gesto articulatório, do seu modelamento através de uma equação dinâmica e ligações à *task dynamics*, do seu estatuto enquanto unidade simultaneamente discreta e dinâmica, terminando com uma análise dos princípios que regem a coordenação entre os gestos constitutivos de um determinado item lexical e a sua relação com a sílaba. O capítulo culmina com a descrição dos vários módulos que compõem o TADA e, finalmente, com a apresentação da arquitectura geral do modelo articulatório desenvolvido no âmbito deste trabalho.

O **capítulo III** centra-se no desenvolvimento e teste de dois módulos distintos, tendo em vista a silabificação automática e a transcrição fonética do texto escrito a processar pelo modelo gestual do TADA.

Depois de uma breve introdução - onde são abordadas, entre outros tópicos, questões relacionadas o papel do módulo de divisão silábica no âmbito dos sistemas TTS e os quadros teóricos disponíveis para resolver o problema da silabificação automática - a primeira parte do capítulo tem início com uma secção dedicada à descrição da estrutura interna da sílaba do PE, segundo o chamado modelo de "Ataque-Rima". Na medida em que um dos divisores silábicos desenvolvidos pretende ser uma implementação mais ou menos fiel do algoritmo de silabificação do português proposto por Mateus & d'Andrade (2000), a secção seguinte integra uma descrição pormenorizada do conjunto de "convenções" que o compõem. No sentido de contextualizar o nosso trabalho no panorama da investigação nacional, são também referenciados alguns estudos que se debruçam sobre a temática da divisão silábica automática. Seguidamente, descreve-se o essencial da implementação dos dois algoritmos de silabificação desenvolvidos: o primeiro com base em transdutores de estados finitos e o segundo a partir da aplicação automática de um conjunto de regras de silabificação originalmente propostas por Mateus & d'Andrade (2000). As secções subsequentes centram-se na explicitação dos critérios metodológicos adoptados na avaliação dos dois sistemas e na análise e discussão dos resultados obtidos.

A segunda parte do capítulo - dedicada ao desenvolvimento de um módulo de transcrição ortográfico-fonética - encerra uma pequena reflexão em torno da relação entre a ortografia e a fonologia e um resumo dos estudos realizados em Portugal com o intuito de determinar automaticamente a pronúncia das palavras. Segue-se uma descrição dos prodedimentos adoptados no desenvolvimento

e avaliação de um primeiro algoritmo de conversão grafema-fone, baseado em regras manuais, corrigidas mediante a aplicação de uma técnica de aprendizagem automática. Após uma análise do desempenho do referido algoritmo, ilustrada com exemplos do processamento efectuado, apresenta-se o segundo algoritmo proposto, desta feita assente, essencialmente, em métodos de aprendizagem automática. Também este último sistema foi avaliado, segundo a metodologia então explicitada, sendo os resultados obtidos - com destaque especial para o impacto da informação silábica sobre o desempenho dos vários sistemas - analisados no final do capítulo.

O **capítulo IV** trata da caracterização gestual dos sons do PE, de modo a que cada fone obtido à saída do módulo de transcrição automática possa ter uma correspondência com um conjunto de gestos articulatórios adaptados para o português. Em primeiro lugar, procede-se a uma descrição detalhada do modo de funcionamento do modelo gestual do TADA, mais especificamente dos dois componentes adaptados por nós, no sentido de lidar com os sons do PE: 1) *o dicionário gestual*, que especifica os gestos associados aos segmentos de entrada (representados sob a forma de variáveis do tracto com os respectivos parâmetros dinâmicos e peso dos articuladores); e 2) o *dicionário de coordenação inter-gestual*, que inclui informações acerca dos osciladores associados aos gestos e modo de acoplamento entre gestos consecutivos, em função da posição silábica. Depois de esclarecidos os procedimentos metodológicos adoptados na descrição gestual dos segmentos do PE, é apresentada a nossa proposta de caracterização gestual, sendo reservada uma secção isolada para cada uma das classes de sons do PE. Cada secção toma como ponto de partida os dados articulatórios disponíveis na literatura fonética, respeitantes à classe de som em análise, e culmina com a apresentação da nossa proposta - sob a forma de uma tabela, contendo os vários parâmetros consignados no TADA - com base na análise informal de imagens RM e medições realizadas sobre os perfis articulatórios de um dos informantes.

No restante, o capítulo é dedicado à construção e aplicação de um teste perceptual para avaliar a inteligibilidade dos segmentos sintetizados a partir das configurações gestuais apresentadas. A análise dos resultados obtidos permitiu identificar os principais problemas, subjacentes à caracterização gestual do inventário fonémico do PE, e ter uma ideia geral do funcionamento da infraestrutura.

O **capítulo V** ocupa-se da descrição e análise da nasalidade vocálica à luz dos princípios da FA. Assim, após um breve enquadramento teórico do tema abordado - que permitiu a formulação das questões de partida do estudo EMA e o próprio planeamento do trabalho experimental - a secção 5.3 centra-se na análise do comportamento assumido pelos articuladores orais, nomeadamente o dorso da língua, durante a produção das vogais nasais. A caracterização da dimensão oral baseia-se, tal como no capítulo anterior, na informação obtida através de RM.

De seguida, procede-se à apresentação do estudo experimental, baseado em EMA, a partir do qual procurámos caracterizar o gesto nasal, quer em termos de parâmetros dinâmicos, quer na sua relação temporal com os demais articuladores. Depois de formuladas as questões de partida e hipóteses a testar, damos conta dos procedimentos metodológicos adoptados no desenvolvimento do trabalho experimental, desde a recolha do *corpus* até ao processamento dos dados recolhidos. Os

resultados relativos a cada uma das variáveis em análise - amplitude, duração, *stiffness*, coordenação temporal - são apresentados e brevemente comentados em secções separadas.

A última parte do capítulo diz respeito ao teste perceptivo efectuado, tendo em vista o esclarecimento de algumas questões suscitadas pelo estudo exploratório anterior. Depois de descrito o teste e apresentados os principais resultados, acompanhados de um primeiro comentário, o capítulo encerra com uma discussão global dos dados recolhidos, à luz das questões teóricas enunciadas e da bibliografia disponível.

No **capítulo final, o sexto**, faz-se uma síntese do trabalho realizado, apresentam-se as principais conclusões decorrentes deste estudo e tecem-se algumas considerações acerca das limitações e problemas enfrentados no decurso desta dissertação. Para além disso, apontam-se algumas sugestões para continuação do trabalho.

1.2 Publicações

O trabalho efectuado está na génese de um conjunto de publicações, que passamos a apresentar.

Uma primeira versão do sistema de conversão de texto para fala, para o PE, baseado numa integração do TADA com o SAPWindows, incluindo os vários módulos linguísticos desenvolvidos, tendo em vista a conversão das unidades linguísticas discretas nas trajectórias contínuas dos articuladores, foi apresentado na *International Conference on Computational Processing of Portuguese* (PROPOR 2008):

- Teixeira, A., Oliveira, C., Barbosa, P., "European Portuguese Articulatory Based Text-to-Speech: First Results" in António Teixeira, Vera Lúcia Strube de Lima, Luís Caldas de Oliveira & Paulo Quaresma (Eds), *Computacional Processing of the Portuguese Language (Proceedings of the 8th International Conference on Computational Processing of the Portuguese Language)*, Springer-Verlag, 2008, 101-111.

A questão da silabificação automática do PE foi abordada nas seguintes publicações:

- Oliveira, C., Moutinho, L., Teixeira, A., "On European Portuguese Automatic Syllabification", *Proceedings of Interspeech'2005- 9th European Conference on Speech Communication and Technology*, Lisboa, 2005, 2933-2936.

- Oliveira, C., Moutinho, L., Teixeira, A., "On Automatic European Portuguese Syllabification" in Manuel González González, Elisa Fernández Rei, Begoña González Rei (Coords), *Actas do III Congreso Internacional de Fonética Experimental*, Santiago de Compostela, Xunta de Galicia, 2007, 461-473.

O essencial da implementação e avaliação dos dois módulos de transcrição ortográfico-fonética desenvolvidos encontra-se publicado em:

- Teixeira, A., Oliveira, C., Moutinho, L., "On the Use of Machine Learning and Syllable Information in European Portuguese Grapheme- Phone Conversion" in Renata Vieira, Paulo Quaresma, Maria das Graças Volpe Nunes, Nuno Mamede, Cláudia Oliveira & Maria Carmelita Dias (Eds), *Computacional Processing of the Portuguese Language (Proceedings of the 7th International Workshop - PROPOR 2006)*, Springer-Verlag, 2006, 212-215.

- Teixeira, A., Oliveira, C., Moutinho, L., "Machine Learning of European Portuguese Grapheme-To-Phone Conversion using a Richer Feature Set", *Revista do DETUA*, Vol. 4, n^o 6, Aveiro, 2006, 746-751.

- Oliveira, C., Moutinho, L., Teixeira, A., "Um novo sistema de conversão grafema-fone para o PE baseado em transdutores", *Anais do II Congresso Internacional/VIII Congresso Nacional de Fonética e Fonologia* (2004), São Luís/Maranhão, Brasil, 2007.

O décimo sexto *International Congress of Phonetic Sciences* foi palco da apresentação oral dos primeiros resultados relativos à sincronização dos gestos que compõem as vogais nasais, obtidos com base da análise de um *corpus* EMMA, adquirido no ano 2000:

- Oliveira, C., Teixeira, A., "On Gestures Timing in European Portuguese Nasals", in Jürgen Trouvain, William Barry (Coord.), *Proceedings of 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, 2007, 405-408.

No âmbito do projecto HERON, foram realizados, em conjunto com o seu coordenador, dois relatórios técnicos, que descrevem: 1) o protocolo de aquisição de dois *corpora*, com base em EMMA, relativos a sons nasais e consoantes laterais; 2) os procedimentos adoptados na anotação automática do movimento dos articuladores, numa base de dados EMMA, recolhida no seio de um projecto anterior:

- Oliveira, C., Teixeira, A., "Nova Base de Dados EMMA relativa às Nasais e Laterais do Português Europeu", *Projecto POSC/PLP/57680/2004 - HERON: A Framework for Portuguese Articulatory Synthesis Research*, IEETA, Universidade de Aveiro, 2007.

- Oliveira, C., Teixeira, A., "Base de Dados EMMA com Anotação Automática de Gestos", *Projecto POSC/PLP/57680/2004 - HERON: A Framework for Portuguese Articulatory Synthesis Research*, IEETA, Universidade de Aveiro, 2006.

Finalmente, destacamos três artigos, cujo conteúdo está, de algum modo relacionado, com a temática central desta dissertação. O primeiro descreve um estudo de coarticulação, baseado em

ressonância magnética. O segundo pretende ser uma revisão das aplicações da síntese de fala no campo do ensino assistido por computador e como ferramenta ao serviço da investigação nas áreas da Linguística e Psicolinguística. O último encerra um reflexão sobre o modo de oralização das siglas, tendo como preocupação o ensino do Português Língua Estrangeira (PLE).

- Teixeira, A., Martins, P., Carbone, I., Oliveira, C., Silva, A., "An MRI Study of European Portuguese Lingual Coarticulation", capítulo de livro (no prelo).
- Teixeira, A., Oliveira, C., Moutinho, L., "A síntese de voz aplicada ao ensino das línguas", in Susan Howcroft (Coord.), *Actas do Encontro Internacional de Linguística Aplicada*, Universidade de Aveiro, Aveiro, 2006, 199-213.
- Mendes, H. M., Oliveira, C., Teixeira, A., "PLE: uma sigla para ler ou soletrar?", in Lurdes Moutinho (Coord.), *Cadernos de PLE 3*, Universidade de Aveiro, Aveiro, 2003, 121-139.

Capítulo 2

Fundamentos Teóricos e Arquitectura do Sistema

Omnium rerum principia parua sunt.

Cícero, *De Finibus Bonorum et Malorum*, 7.21

Como o objectivo que nos move é o desenvolvimento de um conjunto de módulos linguísticos para suportar a conversão automática do texto em fala, far-se-á necessária uma pequena introdução ao tema da síntese de fala. Depois de uma breve definição do conceito, ocupar-nos-emos da descrição da arquitectura geral de um sistema TTS, nomeadamente o *bloco de Processamento de Linguagem Natural* e um *bloco de Processamento de Sinal*. O sintetizador articulatório SAPWindows será apresentado em seguida, juntamente com os procedimentos efectuados no âmbito do projecto HERON, no sentido da sua integração com o TADA.

Uma vez que o modelo linguístico que nos serve de referência é a Fonologia Articulatória, os pontos seguintes estão reservados à apresentação dos pressupostos teóricos que enformam esta teoria fonológica. Tal apresentação não pretende ser exaustiva, mas tem antes como finalidade dar a conhecer alguns dos seus aspectos básicos, necessários à compreensão do trabalho efectuado. De entre eles, destaca-se o conceito de gesto e suas propriedades constitutivas essenciais: dimensão espacio-temporal; possibilidade de coordenação com outros gestos através de relações de fase bem definidas; dupla função como unidade de informação (contraste fonológico) e acção (produção de constrições no tracto oral).

A secção seguinte será dedicada à apresentação dos vários componentes do sistema TADA.

Finalmente, os diferentes módulos que compõem o modelo de produção para o PE, desenvolvido com base na integração do TADA com o SAPWindows, serão descritos num ponto autónomo na parte final do capítulo.

2.1 Síntese de fala

Embora o termo *síntese de fala* englobe um conjunto muito diversificado de processos, de um modo muito geral, este pode ser definido como a capacidade de produção de fala através de mecanismos artificiais. Este ponto de vista é partilhado, por exemplo, por Dutoit & Stylianou (2003) e Olive (1996):

> *Text-to-speech (TTS) synthesis is the art of designing talking machines.* (Dutoit & Stylianou, 2003, p.324)

> *Speech synthesis, or sound synthesis, refers to the process of creating a sound by machine or computer, rather than by such natural means as the human voice or a musical instrument.* (Olive, 1996, p.102)

A estratégia usada na implementação de um sistema de síntese de fala depende de vários factores, entre eles a inteligibilidade e naturalidade do sinal de fala a ser gerado, o tamanho do vocabulário com o qual o sistema trabalha, a velocidade de execução e o custo computacional (Simões, 1999). A complexidade de um sistema de síntese de fala pode também variar significativamente em função do domínio de aplicação. A figura 2.1 ilustra as diferentes técnicas geralmente utilizadas na produção artificial do sinal de fala.

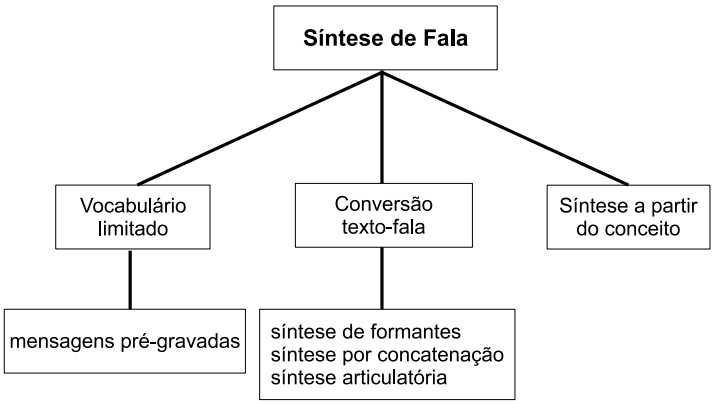


Figura 2.1: Estratégias utilizadas na produção artificial do sinal de fala (adaptado de Simões, 1999).

O modo mais elementar de produzir um sinal de fala consiste na **reprodução de mensagens pré-gravadas**. Este tipo de interfaces proporciona uma qualidade alta, com uma complexidade mínima, mas está naturalmente limitado a aplicações muito específicas.

Uma variante deste tipo de sistemas consiste na concatenação de um número reduzido de palavras ou frases gravadas anteriormente para produzir novas mensagens. Também aqui a qualidade do sinal é elevada e o tempo de resposta bastante curto, ainda que à custa de uma versatilidade baixa.

O número de frases passível de ser gerado através desta técnica é limitado, consistindo basicamente na combinação das mensagens pré-gravadas entre si. Por outro lado, não é possível proceder a qualquer tipo de manipulação prosódica, pelo que a selecção, etiquetagem e gravação do *corpus* se revestem da maior importância, na medida em que podem afectar seriamente a qualidade do sinal gerado. Também os custos de armazenamento inerentes à implementação deste tipo de sistemas são bastante elevados. Esta estratégia mostra-se perfeitamente adequada a algumas aplicações mais simples, onde o vocabulário requerido é circunscrito e as orações a pronunciar de estrutura simples e pré-definida, como é o caso dos sistemas de acesso ao saldo bancário via telefone, mas é absolutamente inviável para aplicações mais gerais, onde o inventário de mensagens é ilimitado. Segundo alguns (e.g. Trancoso *et alii*, 2000; Lemmetty, 1999), um tal sistema não deveria receber a designação de *sintetizador*, estando esta tipicamente reservada para um sistema capaz de sintetizar fala a partir de qualquer tipo de texto.

Os **sistemas de conversão de texto em fala** (TTS) têm como finalidade a transformação de qualquer mensagem escrita na sua correspondente realização sonora (Dutoit, 1997). Este tipo de sistemas procura, de algum modo, mimetizar o processo realizado pelo leitor humano no momento de oralização de um texto. A maior vantagem desta técnica será, porventura, a sua flexibilidade, que se traduz na capacidade de gerar um número ilimitado de frases, aliada a um custo de armazenamento relativamente baixo, pelo menos quando comparado com a técnica acima descrita. As diversas particularidades dos sistemas de conversão texto-fala serão discutidas com mais detalhe nas secções que se seguem.

Uma alternativa aos sistemas de conversão texto-fala são os **sistemas de síntese a partir do conceito** (CTS, do inglês *concept-to-speech*) (Fallside & Young, 1979). Neste caso, a informação a ser manipulada não está representada na forma de texto, sendo mapeada sob a forma de um conceito de entrada, i.e., uma estrutura de formato pré-definido, cujo objectivo principal é padronizar o formato de representação da informação, para que esta possa ser manipulada adequadamente pelos subsequentes estágios do processo de síntese. Nas palavras de Fallside & Young (1979): "The speech synthesis from concept system converts an input concept into speech by using a transformational grammar to generate a well-formed English sentence and a word concatenation synthesizer to generate the actual speech output.". Este tipo de sistemas tem vindo a ser usado com sucesso em Sistemas Automáticos de Diálogo e também na área da tradução automática.

2.1.1 Arquitectura geral de um sistema TTS

Apesar da diversidade, a maioria dos TTS apresenta uma estrutura comum. Podemos distinguir dois módulos principais: a) **bloco de Processamento de Linguagem Natural**, capaz de gerar a representação fonológica da mensagem a partir da sua forma escrita e calcular parâmetros prosódicos (*high-level synthesis*); b) **bloco de Processamento de Sinal**, responsável pela transformação da informação sim-

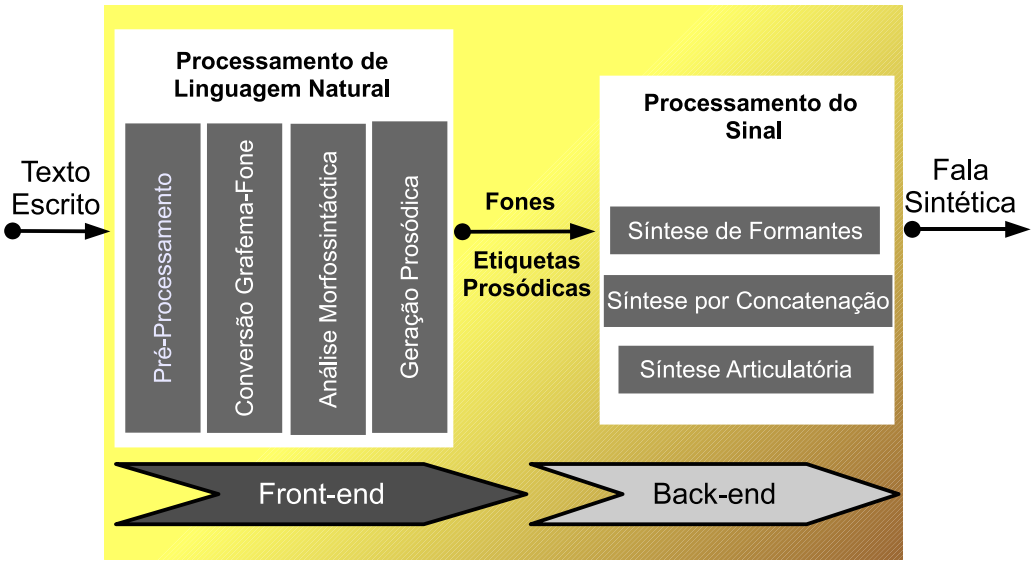


Figura 2.2: Arquitectura geral dos sistemas TTS.

bólica em sinal acústico (*low-level synthesis*) (Dutoit, 1997) [1]. Uma versão simplificada do processo é apresentada no diagrama de blocos da figura 2.2.

Numa primeira fase, é necessário dividir o texto de entrada em unidades menores e assegurar que este se encontra num formato legível. O primeiro módulo encarrega-se, portanto, de separar o texto em frases (*sentence tokenization*) e lidar com as idiossincrasias das abreviaturas, números, etc., num processo designado por *normalização do texto* ou *pré-processamento*. Segue-se a análise morfossintática e a conversão grafema-fone (*Grapheme-to- Phone(me)* (G2P)); e a geração de informação prosódica, que especifica o ritmo e entoação desejadas.

A última etapa corresponde à geração do sinal de fala propriamente dito, através de técnicas de processamento de sinal, tais como a síntese por regras (também designada de síntese de formantes), a síntese por concatenação e a síntese articulatória.

A fim de melhor compreender todo este processo, cada uma das etapas supramencionadas será descrita em pormenor nas secções seguintes, com especial ênfase para os módulos de processamento linguístico, nomeadamente as questões da conversão grafema-fone e divisão silábica, em relação às quais apresentaremos novas soluções no capítulo 3.

2.1.2 Pré-processamento do texto

A primeira tarefa a ser realizada no processo de conversão texto-fala consiste na normalização do texto de entrada, que inclui a divisão do texto em unidades menores e a expansão de elementos não lexicais (números, símbolos, siglas, abreviaturas, fórmulas matemáticas, etc.) em segmentos ortográficos.

[1] Os blocos de **Processamento de Linguagem Natural** e de **Processamento do Sinal** são também chamados de *front-end* e *back-end*, respectivamente (vd. figura 2.2).

Na grande maioria das línguas, o final das frases é marcado por um conjunto de sinais de pontuação (./ ?/ !), seguidos de um espaço em branco e de uma palavra em letras maiúsculas, pelo que a separação do texto em frases é um processo relativamente trivial. Grande parte das ambiguidades está relacionada com o duplo papel desempenhado pelo ".", que nem sempre assinala o final de uma frase, estando, muitas vezes, associado a números, a endereços de internet e, sobretudo, a abreviaturas. A solução poderá passar por adiar para mais tarde - e.g. para depois da expansão das abreviaturas ou quando novas infomações sintáctico-semânticas estiverem disponíveis - a decisão de assinalar (ou não) o final de uma frase ou por aplicar um algoritmo de desambiguação (Jurafsky & Martin, 2008).

A principal característica de um sistema TTS prende-se com o carácter irrestrito do texto a ser sintetizado, sendo que deste podem naturalmente fazer parte mais do que simples sequências de palavras. Para que o texto de entrada seja adequadamente interpretado pelo módulo de transcrição grafema-fone, é pois necessário minimizar a sua variabilidade, convertendo todos os números, símbolos e abreviaturas no seu equivalente textual.

Este processo é também levado a cabo no módulo de pré-processamento e compreende, pelo menos, três fases: 1) identificação dos potenciais elementos especiais; 2) classificação dos mesmos em diferentes categorias (numerais, abreviaturas, siglas, endereços electrónicos), usando uma técnica de etiquetagem prévia ou através de reconhecimento automático; 3) expansão numa sequência de caracteres ortográficos (Jurafsky & Martin, 2008).

A substituição de símbolos (e.g. '#'), abreviaturas (e.g. 'Dr.') e convenções (e.g. '15/11/89'), por palavras está longe de ser uma tarefa simples e envolve algumas complexidades que procuraremos ilustrar.

O processamento dos **números**, por exemplo, é particularmente complexo, variando a sua leitura em função do contexto: numerais cardinais vs ordinais (1- um; 1^o- primeiro), horas (19:25- dezanove horas e vinte e cinco minutos), valores monetários (5 £- cinco libras), reais em notação de vírgula fixa (11,2 - onze vírgula dois).

A tradução de algumas sequências pode mesmo revelar-se uma tarefa bastante difícil, determinada apenas pelo contexto (12/11- doze sobre onze; doze a dividir por onze; doze de Novembro; ou doze do onze).

No tratamento das **abreviaturas**, é comum o recurso a um dicionário, que relaciona a abreviatura com a representação por extenso. Neste sentido, depois de identificada, uma sequência como "sr." será convertida em "senhor". Apesar de, no geral, a expansão de abreviaturas se revelar uma tarefa mais simples do que o processamento de números, há que contar com algumas dificuldades: por um lado, algumas abreviaturas carregam consigo informação de número (1 cm = um centímetro; 2 cm = dois centímetros); e a sua expansão pode também envolver algumas ambiguidades (Cap. =

capítulo "cap. I" ou capitão "cap. Gancho").

Já a principal dificuldade encontrada no tratamento das **siglas** [2] é decidir se elas devem ser lidas ou soletradas[3]. Muitas vezes, as siglas não são oralizadas de acordo com as regras gerais de conversão grafema-som (Yvon, 1994). Mesmo entre falantes de uma mesma língua, a pronunciação de acrónimos está sujeita a uma grande variabilidade.

O problema tem motivado diversas reflexões: enquanto uns estão mais empenhados na identificação dos mecanismos linguísticos que presidem à oralização das siglas (Plénat, 1993, 1998; Diamond, 1994); outros parecem mover-se por interesses mais práticos, nomeadamente a definição de regras de transcrição grafema-fone para a pronunciação automática das siglas, tendo em vista a melhoria da qualidade dos sistemas de síntese (Mareüil, 1994, 1995; Mareüil & Floricic, 2001; Trancoso & Viana, 1997; Yvon, 1994; Barbosa *et alii*, 2003c), ou o ensino de uma língua estrangeira (Mendes *et alii*, 2003).

No que respeita ao português, a grande maioria das pesquisas acerca da oralização de acrónimos foram realizadas no âmbito dos projectos ONOMASTICA (Trancoso *et alii*, 1994; Viana *et alii*, 1996; Trancoso *et alii*, 1995) e VODIS (Trancoso & Viana, 1997). No que toca a este tópico em particular, o objectivo consistiu em delinear um conjunto de regras para predizer o modo de oralização das siglas, para posterior integração no sistema de síntese de fala DIXI. Na sua versão inicial, este sistema soletrava todas as siglas constituídas apenas por sequências de consoantes, deixando os procedimentos de transcrição fonética actuar sobre as restantes. A presença de uma vogal é, efectivamente, uma condição necessária para que uma sigla possa ser lida, mas não é o único critério a considerar na altura de seleccionar o processo adequado de oralização. A extensão e a estrutura do acrónimo são factores a ter em conta (Trancoso *et alii*, 1994; Viana *et alii*, 1996; Trancoso *et alii*, 1995). São geralmente soletradas todas as siglas com menos de três letras e preferencialmente lidas as que têm mais de cinco. As duas formas de oralização são possíveis com siglas de três ou quatro letras, parecendo a escolha depender da estrutura da sigla. Para além disso, "para que uma determinada sequência segmental possa ser lida, tem de se prestar a uma análise silábica concordante com o conjunto de princípios gerais e com as restrições específicas da língua, mas tem também de corresponder a um padrão de palavra possível em extensão e peso". (Viana *et alii*, 1996). Com base em princípios deste tipo, foi criado um conjunto de regras que predizem o modo de oralização das siglas, com resultados bastante

[2] O acrónimo distingue-se da sigla no que diz respeito ao processo de formação e ao modo de realização oral. Enquanto aquele é formado por uma ou mais letras, sílabas ou até morfemas iniciais (**I**nstituto **S**uperior da **Mai**a), a sigla inclui apenas a primeira letra do pequeno conjunto de signos que se pretende abreviar (**T**ransportes **A**éreos **P**ortugueses). As siglas podem ser lidas ou soletradas, ao passo que os acrónimos são criados para serem lidos. Uma vez que não existe consenso relativamente a estas definições e a tendência actual é tratar as siglas como acrónimos, não faremos qualquer distinção entre ambos.

[3] Embora pouco frequentes, existem também siglas, cuja a oralização é mista, i.e., uma parte da sequência é soletrada e a outra parte é lida (e.g. sigla CMOS - *Complementary Metal-Oxide-Semiconductor* - que se pronuncia "Cê-Mós").

aceitáveis (Trancoso *et alii*, 1994; Viana *et alii*, 1996; Trancoso *et alii*, 1995).

Ainda em relação ao tópico do pré-processamento, destaca-se também o sistema desenvolvido por Braga (2008), que, para além de um conversor de símbolos e caracteres especiais e um leitor de numerais, inclui ainda um algoritmo de leitura de siglas e acrónimos. O mesmo módulo é usado também para a segmentação do texto em unidades menores, essencialmente a partir dos sinais de pontuação.

2.1.3 Conversão grafema-fone

O módulo de conversão grafema-fone tem como objectivo transformar o texto de entrada, que se apresenta sob o formato ortográfico convencional, numa sucessão de fones. Para realizar esta tarefa, o sistema tem de dispor da mesma informação que permite a um falante ler um texto: interpretação fónica dos grafemas, conhecimento da estrutura silábica da língua e posição do acento, processos fonológicos (Mestre, 1998; Dutoit, 1997). Neste sentido, o desenvolvimento de um sistema de transcrição fonética automática pode beneficiar em muito dos métodos e conhecimentos da Linguística.

Geralmente, o modelo de pronunciação adoptado deve ajustar-se a uma variedade considerada padrão, que oferece a vantagem de ser uma norma comum a todos os falantes, ainda que se possa optar, assim o exija a aplicação, por uma transcrição marcada por rasgos dialectais (cf., por exemplo, Savino *et alii*, 2005)

Longe de ser uma tarefa simples, a conversão grafema-fone comporta algumas dificuldades, uma vez que não existe uma correspondência bi-unívoca entre os grafemas e a sua realização oral. O processo de conversão grafema-fonema é particularmente árduo em línguas em que o mapeamento das letras em fonemas é altamente irregular (e.g. inglês e francês), mas mesmo em línguas com ortografias de base fonológica, como é o caso do português (d'Andrade & Viana, 1985; Viana *et alii*, 1991; Mateus, 2006), a correspondência entre o sistema ortográfico e o sistema fonético não é perfeita. Consequentemente, o problema da conversão grafema-fone continua a ser alvo de atenção por parte dos especialistas, quer sejam linguistas ou engenheiros.

Uma das estratégias para determinar a pronúncia de uma palavra passa por armazenar a maior quantidade possível de informação fonológica num **dicionário** (cf., por exemplo, Laporte, 1988; Coker *et alii*, 1973) [4]. Apesar de absolutamente infalível, no caso da palavra a transcrever fazer parte do dicionário, os custos computacionais desta abordagem são muitíssimo elevados. Por outro lado, seria necessário actualizar constantemente o conteúdo do dicionário, no sentido de dar conta dos estrangeirismos, neologismos e novas siglas frequentemente incorporados à língua.

[4] Um dos dicionários actualmente mais utilizados nos sistemas TTS (e.g. versão americana do sistema de síntese multilíngua *Festival*) é o *CMU Pronouncing Dictionary* (também conhecido por *cmudict*), que contém cerca de 127,069 palavras e respectiva transcrição fonética (Carnegie Mellon University, 1993).

A fim de reduzir o número de entradas, é comum optar por um dicionário de morfemas, em vez de um dicionário de pronúncias completo. Neste caso, a transcrição é obtida a partir de regras concatenação, que especificam as modificações sofridas pelos constituintes morfémicos quando combinados em palavras. Esta foi a abordagem seguida no sistema *MITTALK* (Allen *et alii*, 1987), onde um dicionário com cerca de 12 000 morfemas permite dar conta da maioria das palavras da língua [5].

Uma outra possibilidade para resolver o problema da conversão grafema-fone é a aplicação de **regras de transcrição** (*letter-to-sound rules*). Esta abordagem, embora mais económica em termos de memória, não permite dar conta de todas as formas de transcrição, pelo que é frequente o recurso a um **dicionário de excepções**.

Este é composto por palavras de pronúncia irregular, mas pode incluir outros items, tais como siglas - cuja leitura normal seja diferente da produzida pelas regras de leitura de siglas (se as houver) - ou estrangeirismos. O dicionário de excepções pode ainda armazenar outras informações (como, por exemplo, a classe gramatical das palavras), passíveis de serem usadas em implementações de análise sintáctica.

Um pequeno dicionário de excepções, contendo as palavras mais frequentes da língua, pode chegar a cobrir uma grande parte das palavras de um texto: em inglês, por exemplo, verifica-se que um dicionário de 200 entradas abrange mais de 50% das palavras de um texto, enquanto um de 10000 entradas pode chegar a cobrir 90% desse mesmo texto (Hunnicut, 1980; Klatt, 1987). Contudo, para aumentar essa abrangência de 90% para 93% seriam necessárias cerca de 60000 palavras adicionais. No caso do português, segundo os dados de Oliveira (1996), as 568 palavras do dicionário cobrem, em termos de frequência, 55% do *corpus* de teste.

Klatt (1987) conclui que o tamanho do dicionário está intimamente relacionado com o desempenho das regras de conversão, tomando como exemplos 1) o sistema *Prose-2000*, com um dicionário de excepções de 3000 palavras e um conjunto de regras de transcrição desenvolvidas por Bernstein com uma taxa de acerto de 85%, resultando num sistema com um desempenho global superior a 97%; e 2) a primeira versão do *DECtalk*, que integra as regras de Hunnicut (1980), com um desempenho de 65%, e um dicionário com 6000 entradas, e que apresentou resultados de 95%.

As regras seguem, por vezes, um formalismo importado da teoria fonológica. São normalmente regras de reescrita de aplicação contextual - dado que a realização de um segmento fonológico está condicionada pelas propriedades dos sons adjacentes - com a seguinte forma:

Grafema → Fone / contexto esquerdo _ contexto direito

O problema da eleição de um destes dois procedimentos - dicionários ou regras - para a transcrição do francês e do inglês, é abordado por Divay & Vitale (1997), no excerto que passamos a

[5] Segundo Klatt (1987), um conjunto de 12000 morfemas pode representar mais de 100000 palavras inglesas.

transcrever:

> *Memory is increasingly less expensive and we now have the capability to store in memory a large number of words (along with their phonetic equivalent, grammatical class, and meaning). Why not then store all words (or certainly all of the words that would be commonly encountered in text) in memory? First, if we include derived forms and technical jargon, there are well over three-quarters of a million words in the English or French language. It would be an extremely difficult task to create such a list. More importantly, new words come into the language every day and from these are generated many derived forms. Lastly, when we factor in items that may not even be found in a dictionary, such as proper nouns (first names, surnames, place names, names of corporations, etc.), the necessity of a rule-governed approach quickly becomes apparent.* (Divay & Vitale, 1997, p.497)

Com o advento da inteligência artificial, a par destas duas abordagens tradicionais, surgem novas soluções que se propõem também resolver o problema da transcrição fonética. São técnicas baseadas em aprendizagem automática, capazes de determinar uma transcrição fonética a partir de grandes conjunto de dados, a que alguns chamam **métodos *data-driven*** (Damper *et alii*, 1998; Taylor, 2005) [6]. Entre eles contam-se, por exemplo, a pronunciação por analogia (Marchand & Damper, 2000), as abordagens baseadas em redes neuronais (Trancoso *et alii*, 1994) ou as técnicas baseadas em *Hidden Markov Models* (HMMs) (Taylor, 2005). A diferença principal entre a abordagem baseada em regras e a grande maioria destes métodos reside simplesmente no modo como as regras são desenvolvidas: no primeiro caso, estas são explicitamente formuladas à mão por especialistas na matéria, enquanto os métodos automáticos implicam uma aprendizagem das mesmas a partir dos dados.

Em todo o caso, o grau de dificuldade enfrentado na tarefa de transcrição, a eleição da estratégia a adoptar e até à eficácia da operação são condicionados pela natureza (relação grafema-fonema) da ortografia da língua envolvida.

A transcrição mediante regras parece ser uma estratégia bastante eficaz em línguas que se caracterizam por uma elevada regularidade entre a representação ortográfica e a transcrição fonética.

Assim se justifica a utilização preferencial de regras em sistemas de síntese para o português (Oliveira, 1996; Viana *et alii*, 1991; Oliveira *et alii*, 1992; Teixeira *et alii*, 1998; Barbosa *et alii*, 2003b; Simões, 1999; Albano & Moreira, 1996; Albano & Aquino, 1997) ou o espanhol (Martí & Niñerola, 1987; Pérez & Vidal, 1991; Conejo & van Coile, 1991; Rodríguez *et alii*, 1993; Lopez-Gonzalo *et alii*, 1993), por exemplo.

No pólo oposto, estão línguas como o inglês e o francês, cuja complexidade das correspondências grafema-fone dificulta, mas não invalida, a tarefa de contrução das *Letter-to-Sound*

[6] Taylor (2005) distingue ainda uma terceira metodologia - a par dos métodos *data-driven* e das regras formuladas manualmente - para a conversão grafema-fone, que designa de abordagem estatística.

Rules (LTS).

Em muitas línguas, a informação sobre a **localização do acento lexical** e a **posição silábica** é crucial para uma correcta determinação da sequência de fones, mas tem ainda um grande impacto ao nível da análise e geração da prosódia. Assim sendo, é comum o conversor grafema-fone incluir rotinas para identificação da posição do acento e procedimentos de divisão automática da palavra em sílabas (cf., por exemplo, Braga, 2008). Não obstante, em alguns sistemas, as regras de silabificação e acentuação são aplicadas no módulo de pré-processamento (e.g. Castejón *et alii*, 1994).

Esta questão será retomada mais adiante, no capítulo 3, onde nos propomos descrever os algoritmos desenvolvidos para a transcrição fonética e silabificação automáticas do PE.

O conjunto de procedimentos descritos anteriormente não permite resolver todos os casos de ambiguidade fonológica, nomeadamente a que se verifica para os homógrafos heterófonos (palavras com ortografia equivalente, mas pronúncias distintas). A fonetização correcta de palavras como "gosto", "seco", "namoro" implica o recurso a um ***parser* morfossintáctico** [7], que consiga determinar a sua classe gramatical [8]. Liberman & Church (1992) estimam que o conhecimento da categoria gramatical da palavra seja suficiente para resolver a maioria dos casos de homografia.

Uma outra função deste componente é justamente a de determinar a estrutura sintáctica do texto de entrada. A informação gerada será usada pelo módulo prosódico, de modo a permitir a decomposição do enunciado em constituintes prosódicos (*prosodic phrasing*).

Em casos em que os homógrafos pertencem à mesma classe gramatical (e.g. "sede": "reunião na sede" e "tenho muita sede"), só através de uma **análise semântica** e/ ou discursiva é possível aceder à pronunciação correcta.

Sobre o impacto da análise morfossintáctica no desempenho dos sistemas TTS, vale a pena referir o trabalho de Ribeiro *et alii* (2002, 2003), que desenvolve um *POS tagger* para o PE, composto por dois módulos: 1) um analisador morfológico (*Palavroso*) (Medeiros, 1995); 2) um desambiguador morfossintáctico (MARv), de estrutura híbrida (regras linguísticas combinadas com uma abordagem probabilística).

Especialmente dedicado à desambiguação de homógrafos heterófonos, destaca-se ainda o trabalho de Braga *et alii* (2007) e Braga (2007), que propõe dois tipos de soluções para lidar com o problema: a primeira baseada na análise morfossintáctica, para desambiguar pares de homógrafos que pertencem a classes gramaticais distintas; a segunda, que resulta do desenvolvimento de algoritmos de base semântica para tratar os casos de homógrafos pertencentes à mesma classe gramatical.

A proposta de Barbosa *et alii* (2003a) para o português do Brasil (PB), desenvolvida no âm-

[7] Do inglês, *part-of-speech (POS) tagger/parser.*

[8] Para mais informações sobre tipologias de homógrafos e técnicas tradicionalmente usadas para resolver este problema, consultar Yarowsky (1997).

bito da chamada Gramática Cognitiva, embora interessante, foi testada com base num único exemplo (['sɛdi] vs ['sedɨ]) e carece naturalmente de generalização.

A última etapa do processo de conversão grafema-fonema consiste, normalmente, numa **análise pós-lexical**, que procura dar conta das modificações produzidas entre palavras consecutivas no interior de uma frase (e.g. fusão vocálica/ crase [kaz**a**mɐrɛlɐ]) [9].

2.1.4 Processamento prosódico

Uma mensagem não pode ser analisada apenas em função dos segmentos fonéticos que a constituem e das relações que estes estabelecem com os segmentos vizinhos (efeitos de coarticulação). Existem outras características da fala que estão associadas a unidades mais amplas, como a sílaba, a palavra ou a frase. O processamento prosódico é, portanto, um processamento de natureza predominantemente suprassegmental e refere-se especialmente a parâmetros acústicos como a duração, a Frequência Fundamental (F0) e a intensidade.

A prosódia possui inúmeras funções no processo de codificação da informação da mensagem falada: indicação do tipo de frase; estruturação do enunciado, i.e. divisão em blocos menores, facilitando a sua compreensão por parte do ouvinte; expressão de atitudes e emoções do locutor; distinção perceptiva entre informação nova e importante e informação conhecida e menos relevante num texto.

O processamento prosódico é, portanto, essencial para garantir a inteligibilidade e, sobretudo, a naturalidade do sinal de fala.

Geralmente, a análise prosódica decorre em duas fases: 1) geração de uma representação abstracta/ simbólica de aspectos prosódicos importantes para a síntese de fala (e.g. fraseamento prosódico, proeminência e melodia); 2) predição dos valores de duração, F0 e intensidade a partir dessa representação abstrata (Jurafsky & Martin, 2008).

O processamento prosódico não faz parte dos objectivos desta dissertação, pelo que esta questão não será aqui abordada em profundidade. Limitamo-nos a introduzir o essencial sobre os três parâmetros acústicos normalmente considerados no âmbito dos sistemas TTS: F0, duração e intensidade. Mais informações sobre este assunto podem ser encontradas, por exemplo, em Jurafsky & Martin (2008, cap.8) e Huang *et alii* (2001, cap.15).

2.1.4.1 Frequência fundamental

Uma das funções do módulo de processamento prosódico é atribuir, de uma forma automática, um contorno de frequência fundamental apropriado a cada uma das frases a ser sintetizada.

[9] Para uma revisão dos fenómenos de *sandhi* em PE, consultar Frota (2000).

Um controlo adequado deste parâmetro (e da prosódia em geral) é fundamental, já o dissemos, para garantir a qualidade e naturalidade do sinal de fala.

Desde que Mattingly (1966, apud Klatt, 1987) desenvolveu aquele que é considerado o primeiro algoritmo para determinar o contorno melódico, diversas estratégias têm sido propostas para determinar os movimentos da frequência fundamental, sendo que a tendência actual vai no sentido do desenvolvimento de modelos mais sofisticados, que procuram especificar o contorno de F0 a partir de uma representação abstrata da estrutura prosódica.

Descrevemos, em seguida, alguns dos sistemas de representação mais conhecidos (cf. Llisterri *et alii*, 2003, 2004):

- O sistema TOBI (**T**one and **B**reak **I**ndices) (Silverman *et alii*, 1992) é um dos mais conhecidos modelos de representação simbólica da entoação, baseando-se no trabalho pioneiro de Pierrehumbert (1981). Segundo este modelo, as frases estão organizadas em constituintes entoacionais, que, por sua vez, se dividem em um ou mais constituintes intermédios. Cada um dos constituintes está associado a uma configuração de dois tons de fronteira (L-L%, L-H%, H-H% e H-L%) e cada uma das palavras constantes do enunciado pode, opcionalmente, receber um dos cinco graus de acento disponíveis. Para além disso, distinguem-se ainda cinco níveis de segmentação prosódica, que expressam o grau de coesão da juntura prosódica. Originalmente concebido para a anotação do inglês americano, este sistema de etiquetagem prosódica tem vindo a ser adaptado a outras línguas, inclusive ao português (Frota, 2000). O sistema tem servido de base a vários modelos que procuram determinar a curva melódica, quer através de regras (Jilka *et alii*, 1999), quer através de métodos de aprendizagem automática (Black & Hunt, 1996);

- O modelo INTSINT (**IN**ternational **T**ranscription **S**ystem for **INT**onation) (Hirst, 1994), desenvolvido no *Centre National de la Recherche Scientifique de la Université de Provence* compreende quatro níveis de representação: acústico, fonético, fonológico superficial e fonológico profundo. Cruz-Ferreira (1999b) descreve o sistema prosódico do PE à luz deste modelo. Em termos de aplicação à síntese de fala, a título de exemplo, destacamos duas versões deste sistema, ambas para o francês: uma baseada em regras (Cristo *et alii*, 2000) e outra baseada num método probabilístico (Véronis *et alii*, 1998). Tanto uma como outra são compostas por duas partes distintas: um módulo linguístico, responsável pela análise do texto e atribuição de etiquetas prosódicas; e um módulo fonético, que associa as etiquetas prosódicas abstractas a valores acústicos;

- O modelo fonético de Fujisaki (Fujisaki & Nagashima, 1969) para a geração de F0 sobrepõe aditivamente, numa escala logarítmica, diversos componentes: um valor básico para a frequência fundamental, outro relacionado com a segmentação prosódica e um terceiro para o acento

tonal [10]. Originalmente desenhado para dar conta dos contornos de F0 em japonês, o sistema foi adaptado e estendido a várias línguas como, por exemplo, o grego (Fujisaki *et alii*, 1997), o alemão (Mixdorff & Fujisaki, 1994) e o português europeu (Teixeira, 2004).

2.1.4.2 Duração

Uma das funções do módulo de processamento prosódico é determinar, de forma automática, a duração de cada um dos segmentos e pausas que compõem o enunciado.

Um modelo de duração adequado deve levar em conta os vários factores, intrínsecos e extrínsecos, que afectam a duração dos segmentos (Lehiste, 1996). É sabido que cada um dos segmentos fonéticos tem uma duração inerente (normalmente, na ordem das dezenas a centenas de milissegundos), condicionada pelas suas características articulatórias. Para além disso, a duração de um mesmo som varia de acordo com a taxa de elocução, posição do acento, ponto e modo de articulação dos sons adjacentes, posição na frase, estrutura silábica, etc.

Seguindo a classificação de Teixeira (2004, cap.3) [11], existem várias estratégias que podem ser usadas para determinar os valores de duração dos constituintes de um enunciado: 1) modelos baseados em regras (Klatt, 1979; Zellner, 1998); 2) modelos matemáticos (Santen, 1994); 3) e modelos estatísticos. Entre as principais técnicas estatísticas usadas para modelar a duração contam-se as árvores de classificação e regressão (Riley, 1992; Chung, 2002) e as redes neuronais (Riedi, 1998; Teixeira, 2004; Córdoba *et alii*, 1999).

De entre os modelos fundados em regras, o mais conhecido é, porventura, o desenvolvido por Klatt (1979). Este baseia-se no pressuposto que cada segmento fonético possui uma duração intrínseca, que pode ser modificada através da aplicação consecutiva de um conjunto de regras. Nenhum segmento pode ser encurtado aquém do seu valor mínimo de duração. Modelos deste tipo foram desenvolvidos para um conjunto variado de línguas, incluindo o francês (Bartkova & Sorin, 1987) e o português (Simões, 1990), apenas para citar alguns exemplos. O resultado obtido com a aplicação do modelo depende muito do conjunto de regras que o constitui. Para além disso, as regras são obviamente dependentes da língua com que se está a trabalhar.

Os modelos estatísticos, por outro lado, necessitam de um extenso *corpus* de fala adequadamente etiquetado. A partir dessa informação, o modelo é capaz de calcular automaticamente o padrão de duração das frases a sintetizar. Um dos problemas desta abordagem advém da dificuldade em elaborar um *corpus* representativo de todos os fenómenos prosódicos de uma língua e lidar com contextos fonético-prosódicos mais raros, ausentes do *corpus* de treino.

Vale a pena citar ainda outros modelos mais complexos como o de Campbell (1992a,b) ou o de Barbosa (1997). O primeiro funciona em duas etapas: 1) obtenção da duração silábica através

[10] Uma descrição pormenorizada das bases fisiológicas e dos princípios matemáticos implicados no modelo de Fujisaki podem ser encontrados em Teixeira (2004, cap.4).

[11] Para uma revisão dos modelos de duração, consultar também Barbosa (1994b, cap.2).

de redes neuronais; 2) distribuição da duração entre os diversos constituintes da sílaba mediante a aplicação de um modelo estatístico. Já Barbosa propõe o recurso a duas unidades rítmicas alternativas, a sílaba e o IPCG (*inter-perceptual-center-group*). A geração automática da duração divide-se também em duas fases: primeiro procede-se ao cálculo da duração de cada uma das unidades rítmicas do enunciado, usando uma rede neuronal; em seguida, faz-se a distribuição da duração das unidades entre os segmentos que as constituem.

Um outro problema a ter conta na modelação da duração está relacionado com o tamanho da unidade rítmica utilizada. Unidades distintas têm servido de base para a predição da duração (Barbosa, 1994b): fonema (Klatt, 1979; O'Shaughnessy *et alii*, 1988; Bartkova & Sorin, 1987; Riley, 1992; Santen, 1994); sílaba (Campbell & Isard, 1991; Campbell, 1992a,b); IPCG (Barbosa, 1997).

2.1.4.3 Intensidade

A intensidade está associada à amplitude da forma de onda. Este parâmetro permite distinguir os sons fortes dos sons fracos. As vogais das sílabas tónicas, para além de possuirem, uma maior duração e valores mais elevados de F0 do que as sílabas átonas, caracterizam-se por um padrão de energia mais elevado. O mesmo acontece com elementos portadores de informação nova, que apresentam geralmente um incremento da intensidade. Sabe-se ainda que a intensidade varia segundo a posição do segmento em relação às pausas, posição no enunciado, acentuação e tamanho da sequência (Blecua & Acín, 1995).

Poder-se-ia, então concluir que a intensidade é um parâmetro prosódico tão importante como os demais. Contudo, o que efectivamente se observa é que a intensidade do sinal tem uma função de contraste muito menos significativa do que outros parâmetros prosódicos: o fenómeno da acentuação está mais relacionado com contrastes nos valores de duração e frequência fundamental, entre outros factores, do que com a intensidade em si.

Assim, a maioria dos sistemas de síntese de fala prescinde do modelamento da intensidade, atendo-se, em contrapartida, ao tratamento dos padrões de duração e frequência fundamental. Na melhor das hipóteses, é efectuada a normalização da intensidade das unidades concatenadas, de modo a atenuar as distorções do sinal. Não obstante, é possível encontrar alguns modelos que procuram predizer a intensidade (Bartkova *et alii*, 1993; Bagshaw, 1998; Dohalská *et alii*, 2002).

Actualmente, o modelamento da intensidade ganhou um novo fôlego, graças à preocupação em reproduzir voz com emoção (Montero *et alii*, 1999) [12]. Também o fenómeno de diminuição da pressão pulmonar ao longo do enunciado, que resulta numa lenta diminuição da intensidade ao longo da frase, é simulado em alguns sintetizadores (Oliveira, 1996).

[12] O modelamento das emoções é actuamente um tópico de investigação de grande interesse (Bailly *et alii*, 2003). Para além da duração, da F0 e da intensidade, a emoção está associada a outros parâmetros acústicos como, por exemplo, o *jitter*, o *shimmer* (vd. nota 29), o *Harmonics-to-Noise Ratio* (HNR) e o quociente de abertura.

2.1.5 Geração do sinal

A última etapa a ser executada durante o processo de conversão texto-fala é a síntese do sinal, que consiste, basicamente, na geração de um sinal acústico a partir da sequência de fones determinada pelo módulo de transcrição fonética e das variáveis prosódicas calculadas durante a fase de processamento prosódico.

Existem, actualmente, diferentes estratégias que podem ser usadas para a geração de fala sintética, nomeadamente a **síntese de formantes**, a **síntese baseada na concatenação de unidades** e a **síntese articulatória** [13].

Ao longo da presente secção, serão analisadas as particularidades, bem como as vantagens e desvantagens, associadas a cada um dos métodos acima mencionados. Adicionalmente, apresentaremos alguns dos principais eventos históricos que marcaram o desenvolvimento de cada uma das técnicas, com especial ênfase para a síntese articulatória.

2.1.5.1 Sintetizadores de formantes

A síntese de formantes é baseada no modelo fonte-filtro da teoria acústica da produção de fala (Fant, 1960). De acordo com este modelo, o sinal de fala é o resultado da modificação de uma fonte de excitação por um sistema de filtros, cuja função de transferência é determinada pela configuração do tracto vocal.

Os sintetizadores de formantes procuram modelar e controlar os vários parâmetros que actuam durante a produção do sinal acústico, tanto os relacionados com a fonte de excitação (período de *pitch*, amplitude, eventual presença de aspiração, ...), como também os ligados à configuração do tracto vocal (frequência, amplitude, largura de banda, presença de pólos e zeros nasais, ...).

Para a simulação da função de transferência, estes sintetizadores dispõem de uma sequência de filtros que modelam as ressonâncias e anti-ressonâncias das cavidades vocal e nasal. Estes filtros podem associar-se em paralelo ou em cascata, embora em alguns sistemas se tenha optado por uma abordagem mista (Klatt, 1980). A estrutura em cascata é mais apropriada para simular a produção de segmentos sonoros (não-nasais), enquanto a associação em paralelo funciona melhor para as oclusivas, fricativas e nasais.

A passagem que passamos a transcrever, da autoria de Styger & Keller (1994), resume as principais particularidades deste método:

> *In formant synthesis, the basic assumption is that the vocal tract transfer function can be satisfactorily modelled by simulating formant frequencies and formant amplitudes.*

[13] Os sistemas de síntese podem também ser classificados de acordo com o o grau de intervenção manual (Huang *et alii*, 2001, cap.16): no caso da **síntese por regra**, é usado um conjunto de regras manuais para controlar o sintetizador, enquanto na **síntese *data-driven*** os parâmetros são obtidos automaticamente a partir de um conjunto de dados reais.

> *The synthesis thus consists of the artificial reconstruction of the formant characteristics to be produced. This is done by exciting a set of resonators by a voicing source or noise generator to achieve the desired speech spectrum, and by controlling the excitation source to simulate either voicing and voicelessness. The addition of a set of anti-resonators furthermore allows the simulation of nasal tract effects, fricatives and plosives.* (Styger & Keller, 1994, p.111)

Os primeiros sintetizadores de formantes surgem nos anos 50. São eles o *Parametric Artificial Talker* (PAT), desenvolvido por Walter Lawrence, e o *Orator Verbis Eletricis* (OVE) I, da autoria de Gunnar Fant [14] (Lemmetty, 1999; Klatt, 1987).

O emblemático sintetizador de formantes de Klatt seria apresentado duas décadas depois (Klatt, 1980), sendo posteriormente sujeito a diversas actualizações, no sentido de aperfeiçoar a qualidade da fala sintetizada, principalmente no tocante a vozes femininas (Klatt, 1990, apud Styger & Keller, 1994). O sistema é composto por 39 parâmetros de controlo, actualizados a cada 5 ms, que permitem regular as principais características acústicas do sinal de fala. A estrutura básica do modelo é formada por um conjunto de ressoadores, associados em paralelo ou em cascata, para simular a função de transferência do tracto vocal. Adicionalmente, ao modelo do tracto vocal em cascata foi acrescentado um ressoador e um anti-ressoador, no sentido de simular os sons nasais. Quanto à fonte de excitação, dependendo das suas características - fonte sonora ou ruído - pode ser modelada à custa de um gerador de impulsos, separados por um intervalo de tempo igual ao período de *pitch*, ou através de um gerador de números aleatórios. Ambos os sinais podem combinar-se, de forma a simular a produção de sons com características mistas, como as fricativas vozeadas.

A elevada qualidade do sistema justificou que este tenha sido largamente usado pela comunidade científica - para efeitos de síntese e experiências de percepção - e servido de base a várias tecnologias recentes, como o *MITalk*, o *DECtalk* ou o *Prose-2000* (Lemmetty, 1999).

O potencial deste método de síntese foi demonstrado por John Holmes (1973, apud Simões, 1999), que terá conseguido gerar, através de um sintetizador de formantes com circuito paralelo (versão aperfeiçoada do PAT), um sinal de fala sintética indistinguível do sinal natural gravado por um informante masculino. Já no que diz respeito à voz feminina, o investigador ter-se-á deparado com maiores dificuldades (Klatt, 1987).

Uma outra vantagem - para além desta possibilidade em gerar sinais de fala de elevada qualidade, mediante um controlo eficaz dos parâmetros do sintetizador - prende-se com a grande flexibilidade desta técnica no que respeita, por exemplo, à simulação de diferentes qualidades de voz ou distintos estilos de fala, através do ajuste dos parâmetros relevantes.

Não obstante estas vantagens, a dificuldade em estimar os referidos parâmetros de controlo

[14] Dez anos mais tarde, mais precisamente em 1962, Fant apresenta uma nova versão do OVE, um sintetizador denominado de OVE II. O actual sistema comercial Infovox é originalmente descendente destes dois sistemas pioneiros (Lemmetty, 1999).

do sintetizador a partir de amostras de fala natural e o grande número de regras e parâmetros que é necessário manipular fazem da síntese de formantes uma técnica bastante complexa (Dutoit, 1997). Até que se atinjam os valores adequados, é, muitas vezes, necessário ajustar o modelo, mediante um processo de tentativa e erro, que poderá ser bastante moroso.

2.1.5.2 Sintetizadores baseados em concatenação de unidades

Os sintetizadores concatenativos produzem um sinal de fala através da concatenação de segmentos de fala natural, previamente gravados e armazenados numa base de dados.

Um dos principais aspectos a ter em conta neste método diz respeito ao tamanho das unidades de fala a concatenar. Em se tratando de síntese de fala a partir de texto irrestrito, o recurso a *palavras* é totalmente inviável: a gravação de milhões de formas lexicais e a sua rápida recuperação para a síntese implicaria custos de armazenamento imensos, já para não falar da necessidade em actualizar constantemente o inventário de unidades, de modo a contemplar as novas siglas e neologismos que todos os dias são incorporados à língua. O mesmo acontece em relação às *sílabas*, cujo inventário, embora menor, continua a ser demasiado elevado para poder ser usado num sistema TTS. Quanto aos *fones*, embora em número restrito, têm o problema de gerar descontinuidades espectrais muito significativas, quando usados num contexto fonético muito distinto do original, pondo em causa a inteligibilidade do sinal (Harris, 1953, apud Barbosa, 2001).

O maior desafio enfrentado no processo de elaboração de um inventário de unidades é, então, "to capture key coarticulation phenomena while, at the same time keeping the number of units small" (Sproat, 1998, p.200-201). A solução encontrada para lidar com o fenómeno de coarticulação e ao mesmo tempo controlar o tamanho do inventário, passa, actualmente, pela utilização de *demissílabas*, *difones*, ou outro tipo de unidades com características mistas, que designaremos genericamente por *polifones*.

Quando comparadas com os fones e difones, as *demissílabas* (Fujimura & Lovins, 1978) - unidades que representam metade de uma sílaba, dividida no centro da vogal - implicam substancialmente menos pontos de concatenação, sendo simultaneamente capazes de capturar grande parte do fenómeno coarticulatório. As desvantagens da utilização das demissílabas como blocos constituintes básicos do sinal de fala sintetizada continuam a estar relacionadas com os custos computacionais (o número de demissílabas é superior aos difones) e, sobretudo, com a impossibilidade de sintetizar todas as palavras possíveis, apenas com base num sistema baseado em demissílabas (Lemmetty, 1999). Contudo, estas podem ser usadas com sucesso em sintetizadores que recorrem simultaneamente a unidades de tamanho variável (Portele *et alii*, 1992).

Entre as unidades mais usadas nos sistemas de síntese por concatenação estão os *difones* ou *díades* (Peterson *et alii*, 1958), unidades acústicas que se estendem da região estável de um fone até à região estável do fone seguinte. A principal vantagem dos difones reside na minimização das referidas descontinuidades, já que a transição entre os fones é inteiramente preservada. Isto significa

que o processo de junção das unidades tem lugar precisamente nas regiões mais estáveis do sinal, o que reduz drasticamente as distorções decorrentes do processo de concatenação. Estima-se que, em português, um dicionário de cerca de 1000 elementos seja suficiente para sintetizar todas as palavras da língua (Simões, 1999), enquanto para o francês e o inglês, o número aumenta para 1200 (Dutoit, 1997, p.187) e 1300 difones (Huang *et alii*, 2001, p.790), respectivamente.

Se a utilização de difones permitiu melhorar substancialmente a qualidade da síntese, os problemas de inteligibilidade persistem em relação a alguns sons, nomeadamente aqueles com uma duração muito curta ou que, em virtude das suas características dinâmicas, não possuem uma região estável. Uma das alternativas passa por considerar sequências de tamanho superior ao difone (e.g. as vogais átonas podem ter de ser inseridas num trifone). Numa altura em que a memória computacional não é mais um problema, unidades de tamanho variável - designadas de *poliphones* ou *N-phone units* (Holmes & Holmes, 2001, p.72) - são uma das soluções mais populares.

Uma vez seleccionado o conjunto de unidades a concatenar, o processo de criação do dicionário de unidades básicas consiste em 1) efectuar a gravação de amostras de fala natural, contendo as unidades alvo; 2) segmentar as amostras, manual ou automaticamente, de forma a isolar as unidades escolhidas; 3) e armazenar as unidades numa base de dados, juntamente com todas as informações úteis, para posterior utilização pelo sistema de síntese.

Segue-se a fase de síntese propriamente dita, durante a qual os trechos de som pré-gravados - sejam eles demissílabas, difones ou outros - necessários para realizar o enunciado a sintetizar são seleccionados, a partir do dicionário, e concatenados, procurando suavizar as descontinuidades espectrais nas junções. Parte deste problema pode ser minimizado mediante uma escolha adequada das unidades base - como foi já referido anteriomente - e do controle rigoroso das condições de gravação e segmentação (seleccção de contextos foneticamente neutros, mediante o recurso a logátomos e frases de suporte; leitura das frases, usando uma taxa de elocução e uma F0 constantes ao longo de toda a gravação; segmentação das unidades em regiões espectralmente estáveis do sinal; escolha adequada do informante). Para além disso, a eficiência do processo de concatenação depende ainda da técnica de processamento do sinal utilizada.

A par da concatenação das unidades pré-gravadas, durante esta fase, é ainda necessário promover uma actualização dos parâmetros prosódicos, de modo a que os segmentos que fazem parte do enunciado a sintetizar venham a ter o contorno prosódico determinado durante a etapa de processamento prosódico.

Cabe notar que, não obstante a sua simplicidade [15] e qualidade do sinal gerado, o método de síntese em causa não admite alterações prosódicas muito pronunciadas, sob pena de introduzir

[15] A elaboração do inventário de unidades e respectiva segmentação constitui, eventualmente, a tarefa mais complexa e demorada de todo o processo. A introdução de algoritmos automáticos para a selecção e segmentação das unidades veio facilitar este trabalho, ainda que os resultados atingidos não sejam totalmente fiáveis. Apesar disso, é actualmente possível construir sistemas de síntese baseados em concatenação de unidades para a maioria das línguas, num curto período de tempo (Sproat, 1998).

graves distorções no sinal. Neste sentido, a síntese por concatenação de unidades mostra-se muito menos flexível do que a síntese de formantes, na medida em que esta última dispõe, como já vimos, de parâmetros que permitem controlar livremente as características da fonte glotal e do tracto oral/ nasal. A esta desvantagem vêm somar-se as já relatadas descontinuidades espectrais decorrentes do processo de concatenação, responsáveis pela sensação de "voz metálica". Este tipo de efeitos pode ser atenuado - mas jamais compensado totalmente - mediante técnicas de síntese como o *Linear Predictive Coding* (LPC), os algoritmos *Pitch Synchronous Overlap-Add* (PSOLA) ou o mais recente método *Multi-Band Re-synthesis Overlap-Add* (MBROLA) (Simões, 1999; Lemmetty, 1999).

Muito embora a teoria seja bastante anterior (Peterson *et alii*, 1958), os primeiros sistemas de síntese de fala por concatenação com base em difones surgem nos finais dos anos 60 (e.g. Dixon & Maxey, 1968, apud Klatt, 1987).

O advento de novos métodos (e.g. predição linear multipulso e método PSOLA), ou a gravação de unidades mais longas (e.g. demissílabas ou trifones) implicará novos ganhos para este tipo de dispositivos, de tal forma que "concatenative synthesis is now the leading approach in speech synthesis, based on numbers of researchers pursuing that approach and numbers of commercial speech synthesizers using it." (Shadle & Damper, 2001).

A seguinte citação resume, de forma que consideramos esclarecedora, o actual estado de desenvolvimento do método de síntese que nos tem vindo a ocupar:

> *Currently, the most successful approach for speech generation in the commercial sector is concatenative synthesis. Concatenative synthesizers store segments of natural speech, which are pieced together to form the desired speech output. The best speech quality is currently achieved by so called unit-selection synthesizers. However, all concatenative synthesizers depend on the prerecorded speech material, which can only be modified moderately without a loss of quality. This makes it difficult to simulate arbitrary voices speaking arbitrary languages and to express emotions like happiness or anger.* (Birkholz, 2007b)

2.1.5.3 Sintetizadores articulatórios

Os modelos baseados em síntese articulatória procuram simular, de forma realista, os mecanismos fisiológicos de produção de fala. Neste sentido, a síntese de base articulatória é considerada o modo mais "natural" de produzir fala (Taylor, no prelo, p.422).

Este método é definido por Teixeira *et alii* (2005) nos seguintes termos:

> *Articulatory synthesis generates the speech signal through modeling of physical, anatomical, and physiological characteristics of the organs involved in human voice production. (...) In the articulatory approach, the system is modeled instead of the signal or*

> *its acoustics characteristics. Approaches based on the signal try to reproduce the signal of a natural voice as faithfully as possible with few or no concern about how it is produced. In contrast, a model based on the production system uses physical laws to describe the sound propagation in the vocal tract and models mechanical and aeroacoustic phenomena to describe the oscillation of the vocal folds.* (Teixeira *et alii*, 2005, p.1436)

Este tipo de sintetizadores incluem, regra geral, dois componentes (Teixeira, 2000): 1) um modelo anatómico-fisiológico das estruturas implicadas na produção de fala, que transforma a posição dos vários articuladores (maxilar, língua, velo, etc.) em áreas transversais do tracto vocal; 2) e um modelo de propagação dos sons nessas mesmas estruturas, que descreve as propriedades acústicas do sistema vocal através de um conjunto de equações. Este segundo modelo engloba, por sua vez, diversas subtarefas que vão desde a criação de uma fonte excitação glotal e fontes de ruído até à simulação da radiação da energia acústica nos lábios e/ ou narinas, passando pela propagação do som nas cavidades sub e supra-glotais [16].

A posição dos articuladores (e respectivas áreas) pode ser estimada a partir de métodos directos - como a radiografia simples, actualmente substituída pela ressonância magnética, ou outras técnicas (e.g. tomografia computorizada, X-Ray Microbeam [17] ou articulografia electromagnética) (Teixeira, 2000; Shadle & Damper, 2001) - ou pode ser obtida com base no sinal acústico.

Não obstante o enorme potencial desta abordagem, enfatizado no conjunto de citações adiante transcritas, e os recentes avanços na área, há ainda um longo caminho a percorrer até que a síntese articulatória se constitua como uma verdadeira alternativa aos métodos actualmente utilizados nos sistemas de conversão de texto para fala. Mais do que uma tecnologia comercialmente viável, a síntese articulatória é considerada, antes de mais, uma das mais importantes e poderosas ferramentas ao serviço da investigação em áreas como a produção de fala ou a síntese audio-visual (ou síntese multi-modal) (Taylor, no prelo, p.417) [18].

> *By giving us a better understanding of the speech production mechanisms, articulatory synthesis has the long term potential to solve problems affecting the current approaches*

[16] Para uma descrição pormenorizada dos vários modelos (articulatórios e acústicos) disponíveis, consultar Teixeira (2000).

[17] O *X-ray Microbeam* é uma técnica imagiológica, inventada e testada por Osamu Fujimura (Universidade de Tóquio), entre 1973 e 1975, que permite produzir representações parametrizadas de vários pontos anatómicos estáticos e dinâmicos. Foi desenvolvida com o intuito de reduzir as doses de radiação, emitidas pelos sistemas de radiografia simples e cine-radiografia, e simplificar a análise da informação. Um sistema de segunda geração foi desenvolvido na Universidade de Wisconsin (Westbury *et alii*, 1994).

[18] O objectivo geral da síntese audio-visual é a construção de *talking-heads* i.e. sistemas em que a síntese de fala se combina com modelos paramétricos da face humana. A componente visual do discurso (*visible speech*) pode aumentar, em muito, a inteligibilidade da mensagem, nomeadamente em ambientes comunicativos ruidosos - como foi, aliás, demonstrado pelos estudos perceptuais conduzidos por Siciliano *et alii* (2003) ou Massaro (2002) - mas, acima de tudo, o *visible speech* é um excelente canal de comunicação para os indivíduos com perdas auditivas. O *Baldi* é um dos mais conhecidos sistemas de síntese audio-visual, com reconhecido potencial, não só no apoio a crianças com necessidades educativas especiais (Barker, 2003; Bosseler & Massaro, 2003; Massaro & Light, 2004), mas também como suporte ao ensino de uma língua estrangeira (Massaro & Light, 2003).

in speech synthesis. (Gabioud, 1994, p.215)

Ultimately, concatenative synthesis is not the answer. In the long term, articulatory synthesis has more potential, not only for extending our knowledge of speech science, but for high-quality speech synthesis. (Shadle & Damper, 2001)

It has long been conjectured that synthesis based on articulatory models is the most versatile synthesis method and will ultimately produce the most natural-sounding speech. There are several reasons for the belief in this conjecture: such models control the same slowly varying parameters that are controlled in human speech production; the interaction between the vocal cords and the vocal tract is natural and should lead to more natural excitation; the parameters of the model are well suited for interpolation and also well suited for modification in order to produce various voices. (Sondhi & Sinder, 2005, p.75-76)

Entre os principais factores que dificultam o desenvolvimento de tais modelos estão ainda a falta de dados articulatórios sobre o processo de produção de fala - a maior parte dos dados disponíveis dizem respeito a configurações estáticas, enquanto a informação sobre a dinâmica dos articuladores é ainda muito escassa - e de estratégias de controlo apropriadas (Carlson, 1994). Outras dificuldades estão relacionadas com a ausência de um processo de inversão completo, para obtenção dos parâmetros articulatórios a partir de fala natural, e a complexidade e morosidade dos cálculos necessários à simulação (Teixeira, 2000).

As origens da síntese articulatória remontam às "máquinas falantes" do século XVIII. De entre estas, a mais conhecida é, porventura, a "Acoustical Mechanical Speech Machine", arquitectada pelo multifacetado Wolfgang Ritter von Kempelen (Lemmetty, 1999; Schröeder, 1993; Flanagan, 1972; Liénard, 1991) e capaz de produzir sons isolados e até "several hundreds of words, clearly and distinctly. For instance Papa, Mama, Marianna, Roma, Maladie, Santé, Astronomie ... as well as long and difficult words such as Constantinopolis, Monomotapa, Mississipi, Astrakan, Anastasius, etc..." (Kempelen, 1791, apud Liénard, 1991, p.21), para além de um número limitado de frases.

Motivado por questões relacionadas com a educação dos surdos-mudos, Kempelen iniciou a construção da sua máquina em 1769, mas só a terminou 20 anos depois, em 1791[19].

De um modo geral, a máquina era constituída por um fole, que funcionava como fonte de ar para uma caixa de ressonância; uma palheta vibratória de metal para simular as cordas vocais; e um tubo de couro flexível para o tracto vocal [20]. Através da manipulação da forma do tubo, era

[19]Kempelen terá apresentado versões parciais da sua "máquina falante" numa *tournée* efectuada pela Europa entre 1783 e 1785 (Barbosa, 2005; Pompino-Marschall, 2005).

[20]Kempelen descreve pormenorizadamente a sua "máquina falante" no livro "Mechanismus der menschlichen Sprache nebst Beschreibung einer sprechenden Maschine", publicado, numa edição paralela alemão-francês, em 1791. Este inclui ainda várias reflexões sobre o mecanismo de produção de fala, de modo que é considerado "a milestone in the history of phonetics, incorporating many insightful observations on articulatory mechanisms, whereas the speaking machine clearly a milestone in audio engineering." (Pompino-Marschall, 2005, p.155).

possível simular o som das várias vogais, que, contudo, apresentavam problemas de inteligibilidade. Para produzir as diferentes consoantes, incluindo as nasais, existiam quatro constrições ao longo do tubo, controladas manualmente, através das mãos do operador [21].

Se até então, a laringe era considerada o elemento central na produção de voz, as experiências de Kempelen vieram salientar o papel fulcral do tracto vocal no processo de articulação dos sons.

O fascínio do engenho de Kempelen faz-se sentir ao longo de quase dois séculos e várias foram as suas reproduções, destacando-se a do físico britânico Charles Wheatstone (Lemmetty, 1999; Flanagan, 1972) [22], a do inventor do telefone Alexander Graham Bell (Schröeder, 1993; Flanagan, 1972) [23] e do imigrante alemão Joseph Faber [24].

Já na era dos sintetizadores eléctricos, mais precisamente em 1922, Stewart apresenta o primeiro análogo eléctrico do tracto vocal capaz de gerar sons sintéticos (Lemmetty, 1999; Klatt, 1987).

Contudo, o grande marco na história da síntese articulatória (e da síntese de fala em geral) aconteceu, sem dúvida, em 1939 (Schröeder, 1993; Lemmetty, 1999; Klatt, 1987; Liénard, 1991), quando o engenheiro dos Laboratórios Bell, Homer Dudley, deu a conhecer à comunidade científica o sistema de síntese por ele desenvolvido, denominado VODER. O dispositivo, exibido na Exposição Universal (1939) de Nova Iorque, apresentava-se como o primeiro capaz de gerar uma frase completa e dispunha de um interruptor para seleccionar o sinal de entrada, um pedal que permitia controlar a frequência fundamental, um teclado, a partir do qual o operador controlava a amplitude dos dez filtros passa-banda, e um amplificador. O correcto manuseio do sintetizador exigia bastante treino e habilidade, de modo que as operadoras responsáveis pela demonstração do equipamento na referida exposição precisaram de um ano de preparação. Embora a inteligibilidade do sinal gerado fosse bastante reduzida, ficaram demonstradas as potencialidades do sistema para produzir fala artificial. O sistema VODER tinha como inspiração um mecanismo de análise do sinal de voz, também desenhado por Dudley poucos anos antes, o *Voice Coder* (VOCODER).

[21] Uma das muitas versões da "máquina falante" construída por Kempelen está patente no departamento dedicado aos instrumentos musicais, no *Deutsches Museum*, em Munique (http://www.ling.su.se/staff/hartmut/kemplne.htm). Destacam-se, ainda, as reconstituições de Liénard (1967) e Broecke (1983), bem como as réplicas recentes de Nikléczy & Olaszy (2003) e Brackhane & Trouvain (2008).

[22] Em 1835, Charles Wheatstone apresenta, em Dublin, uma nova versão da "máquina falante" de Von Kempelen (Lemmetty, 1999; Flanagan, 1972). De arquitectura complexa, o engenho mecânico era, de um modo geral, dotado de todos os componentes presentes no modelo original (fole, palheta e tubo de couro).

[23] O sintetizador mecânico de Bell incluia uma réplica de todos os orgãos envolvidos no mecanismo de produção de voz: lábios de arame cobertos de borracha, língua de madeira, palato, dentes, faringe e velo (Schröeder, 1993; Flanagan, 1972). Segundo Bell, o dispositivo era capaz de produzir vogais, consoantes nasais e pequenos enunciados simples. São também conhecidas as suas inusitadas experiências com o seu cão Skye, na tentativa de induzir o animal a produzir voz humana.

[24] O sofisticado aparelho de Faber, conhecido como "*Amazing Talking Machine*", incluia uma cabeça e um busto de homem vestido à maneira turca e, no interior, foles, uma glote, uma língua de marfim, uma câmara de ressonância e uma cavidade vocal com palato de borracha, maxilar inferior e bochechas (Riskin, 2003). O artefacto de grandes dimensões, capaz de produzir voz normal e murmurada e até de cantar, era controlado através de pedais e de um teclado de 17 teclas.

A partir do modelo eléctrico para simulação do tracto vocal [25], desenvolvido por Dunn (1950) e aperfeiçoado por Stevens *et alii* (1953), Rosen (1958) contrói, no MIT, o primeiro circuito para a realização de síntese articulatória de forma automática (Klatt, 1987). Contrariamente aos dispositivos eléctricos iniciais, o DAVO era capaz de produzir sons contínuos e incluia um modelo do tracto nasal.

Já na década de 60, são apresentados os primeiros modelos que representam a cavidade oral no plano sagital. O modelo desenvolvido por Coker (1967), bem como os propostos por Mermelstein (1973) e Flanagan *et alii* (1975), estão entre os mais usados, ainda hoje, pelos investigadores da área da síntese articulatória (Carlson, 1994).

O modelo articulatório de Mermelstein (1973) esteve na base do primeiro sistema TTS completo para a língua inglesa, desenvolvido por Teranishi & Umeda (1968), no *Electrotechnical Laboratory*, no Japão. Apresentado no *6th Internacional Congress on Acoustics*, em Tóquio, o dispositivo incluia um módulo de análise sintáctica bastante sofisticado, mas a qualidade do som não era a melhor (Lemmetty, 1999; Klatt, 1987).

É também neste modelo computacional do tracto vocal que se baseia o sintetizador articulatório de Rubin *et alii* (1981), desenhado nos Laboratórios *Haskins*, com vista à realização de estudos de produção e percepção.

A síntese articulatória continuará a desenvolver-se ininterruptamente, tendo-se assistido, nos últimos anos, ao aperfeiçoamento de modelos já existentes, como o *Configurable Articulatory Synthesizer* (CASY) (Rubin *et alii*, 1996; Iskarous *et alii*, 2003) ou o *High Level Parameter Speech Synthesis System* (HLsyn) (Stevens & Hanson, 2003); ao aparecimento de modelos tridimensionais do tracto (e.g. Engwall, 1999; Birkholz *et alii*, 2006; Bailly *et alii*, 2002), com versões adaptadas para a síntese de canto (Birkholz, 2007a); à criação de modelos flexíveis para a simulação de estruturas complexas como a língua (Engwall, 2004); ao desenvolvimento de sintetizadores com capacidade de simular o crescimento do tracto vocal, desde a infância até à idade adulta (Birkholz & Kröger, 2007), e ao aparecimento de novos modelos acústicos.

A grande maioria destes desenvolvimentos só foi possível graças ao advento de novas técnicas para medição da geometria do tracto vocal - e.g. a articulografia electromagnética 3D, a ultrasonografia e a ressonância magnética tridimensional - que possibilitam não só um conhecimento mais detalhado da relação acústico-articulatória, como também uma medição mais precisa dos articuladores.

[25] O "*Electrical Vocal Tract*" criado por Dunn (1950) era alimentado por uma fonte sonora e composto por um conjunto de circuitos, para modelar as ressonâncias do tracto vocal, e uma "tongue component" (Cook *et alii*, 2006), capaz de se mover ao longo dos filtros ressoadores. Apesar da incapacidade de produzir consoantes, a qualidade das vogais sintetizadas era reconhecidamente elevada (Cook *et alii*, 2006; Rubin & Vatikiotis-Bateson, 2006).

2.1.6 Sintetizador articulatório da Universidade de Aveiro

Na Universidade de Aveiro, a investigação na área da síntese articulatória teve início em 1995 e culminou na construção de um sintetizador articulatório 2D, intitulado de SAPWindows, que visa, em primeira instância, servir de suporte a estudos de produção e percepção.

Originalmente vocacionado para a síntese de sons nasais, o sistema tem sofrido algumas actualizações: os modelos actualmente implementados permitem também a síntese de vogais orais e fricativas (Teixeira *et alii*, 2005), embora com algumas limitações.

Um sintetizador deste tipo necessita de modelar, pelo menos, três componentes do processo de produção de fala: 1) geometria das cavidades acima da glote; 2) propagação das ondas sonoras nas cavidades; 3) fonte de excitação. Cada um destes módulos será brevemente descrito nas secções que se seguem [26].

2.1.6.1 Modelos anatómicos

O modelo anatómico do **tracto oral**, apresentado na figura 2.3, é uma versão aperfeiçoada do modelo articulatório desenvolvido na Universidade da Flórida, que, por sua vez, é baseada no modelo de Mermelstein (1973). Os parâmetros articulatórios, relacionados com a posição dos articuladores, são o centro do corpo da língua, o ápice da língua, o maxilar, os lábios, o osso hióide e o véu palatino.

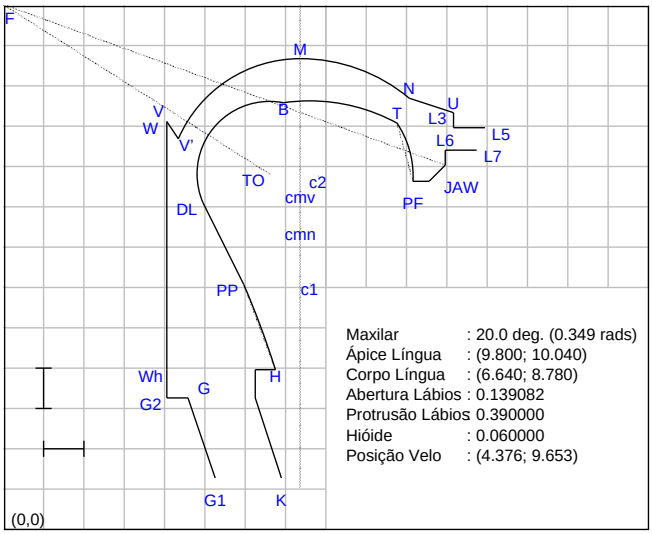


Figura 2.3: Parâmetros Articulatórios do SAPWindows (fonte: Teixeira, 2000). Encontram-se representados os pontos e parâmetros articulatórios necessários à construção do modelo, incluindo o ápice da língua (T), o centro do corpo da língua (T0), a abertura e protrusão labiais (L3 a L7), o hióide (PP) e o velo (V'). Os quadrados têm 1cm de lado.

A informação tridimensional relativa ao comprimento e área das várias secções do tracto

[26] Para uma descrição pormenorizada dos vários modelos implementados, consultar Teixeira (2000) e Teixeira *et alii* (2005).

é obtida a partir do modelo sagital bidimensional, aplicando uma grelha variável. A grelha utilizada divide o tracto oral em 60 secções, repartidas por 6 zonas (Teixeira, 2000).

O modelo adoptado para a simulação do **tracto nasal** é similar ao proposto por Chen (1997). A cavidade nasal é modelada de forma muito semelhante ao tracto oral, sendo que a maior diferença se regista ao nível da função de área, que é fixa para a maior parte do tracto nasal, com excepção da zona do velo. Este último parâmetro varia em função do grau de acoplamento nasal. Cabe notar também a inclusão no modelo de um seio peri-nasal (o seio maxilar) e de uma área de radiação reduzida ($0.5\ cm^2$).

As configurações do tracto relativas às vogais (orais e nasais), definidas pelos parâmetros articulatórios, foram obtidas mediante um processo de inversão, mais concretamente um método baseado em optimização (Teixeira, 2000). Para as consoantes nasais, os parâmetros foram definidos manualmente com base em descrições articulatórias.

2.1.6.2 Modelo acústico

Em virtude do pendor demasiado técnico do modelo acústico, limitamo-nos a referir algumas informações básicas. Mais detalhes podem ser encontrados em Teixeira (2000) e Teixeira *et alii* (2005).

Neste modelo, foi usada uma análise no domínio da frequência e um método de síntese no domínio do tempo, tradicionalmente designado de método híbrido (Sondhi & Schroeter, 1987).

O problema geral da existência de várias fontes foi decomposto em vários problemas simples, usando o princípio da sobreposição. Assim, no sentido de calcular a pressão irradiada nos lábios devido a cada uma das fontes de ruído, o tracto vocal foi dividido em três secções: secção faríngea, região entre o ponto de acoplamento do velo e a fonte de ruído e zona posterior à fonte. Para calcular a resposta impulsional do tracto vocal foram utilizadas matrizes ABCD, calculadas com base na função de área de cada secção (Sondhi & Schroeter, 1987). A fala sintética resulta da aplicação do processo para as várias fontes (glotal e de ruído) intervenientes, da soma de todas as contribuições e, finalmente, do cálculo da derivada (por aproximação numérica), simulando a radiação.

2.1.6.3 Modelo da fonte glotal

O modelo de excitação glotal desenvolvido corresponde a uma versão aperfeiçoada do trabalho de Allen & Strong (1985) e implica o modelamento dos vários subsistemas envolvidos: pulmões, cavidades subglotais, pregas vocais e tracto supraglotal.

Os pulmões são representados no modelo por uma fonte de pressão pulmonar, sendo que os valores de pressão constantes da proposta original (Allen & Strong, 1985) foram ajustados.

Para simular a região subglotal, incluindo a traqueia, foram usados três circuitos RLC para-

lelo em cascata [27].

Quanto às pregas vocais, optou-se por recorrer a um modelo paramétrico de duas massas, seguindo de perto as propostas de Prado (1991).

Todo o sistema acima da glote foi simulado mediante a utilização de uma impedância de entrada [28], que permite modelar de forma mais eficaz as perdas dependentes da frequência.

O modelo da fonte glotal inclui ainda a possibilidade de modelar outros parâmetros como o *jitter* e o *shimmer* [29] ou o ruído de aspiração.

O modelo implementado é controlado por dois tipos de parâmetros: 1) parâmetros variáveis ao longo do tempo (e.g. pressão pulmonar, aspiração), de comportamento similar aos parâmetros articulatórios do tracto e que podem ser usados para controlar, por exemplo, a qualidade de voz; 2) parâmetros invariáveis no tempo (e.g. resistência pulmonar, dimensões da glote), cujos valores podem ser directamente alterados num ficheiro de configuração.

Os vários módulos que constituem o sintetizador articulatório foram integrados numa interface gráfica (Silva, 2001), que tem como objectivo ilustrar todo o processo de síntese, mostrando alguns dos sinais daí derivados. A representação simultânea da informação articulatória e acústica tem várias outras vantagens, para além da monitorização do processo para efeitos de aperfeiçoamento do sistema, e poderá ser de grande interesse para o ensino da Fonética e para a área da Terapia da Fala.

A referida interface é exibida na figura 2.4. Esta divide-se em duas partes: no painel da esquerda, encontram-se representados o contorno sagital do tracto vocal e a trajectória dos lábios, do velo e do hióide; no painel da direita, é possível visualizar o sinal das pregas vocais (duas primeiras linhas), o sinal acústico, o espectrograma e uma representação tridimensional da função de área, designada por areagrama.

2.1.6.4 Desenvolvimentos no âmbito do projecto HERON

O cumprimento dos objectivos delineados na Introdução a este trabalho - nomeadamente o desenho de uma estratégia para a obtenção e controlo dos parâmetros articulatórios, a partir do texto escrito - implicou um processo de integração do sintetizador SAPWindows com o sistema TADA, descrito em pormenor na secção 2.2.5.

Esta tarefa foi executada no âmbito do projecto Heron e consistiu, em termos muitos gerais, no seguinte:

[27] Os circuitos RLC paralelo são circuitos eléctricos constituídos por uma resistência (R), uma bobine (L) e um condensador (C), montados em paralelo, que possuem uma frequência de ressonância, ajustada através dos valores de L e C, para simular uma das ressonâncias subglotais.

[28] De uma forma muito sucinta, a impedância de entrada é a "resistência" que o tracto, na sua globalidade, apresenta à passagem do fluxo de ar.

[29] O *jitter* é definido como a variação da frequência fundamental (ou período) de um ciclo para o seguinte, enquanto o *shimmer* se refere à variação da amplitude do pulso glotal entre dois períodos consecutivos (Baken & Orlikoff, 1999).

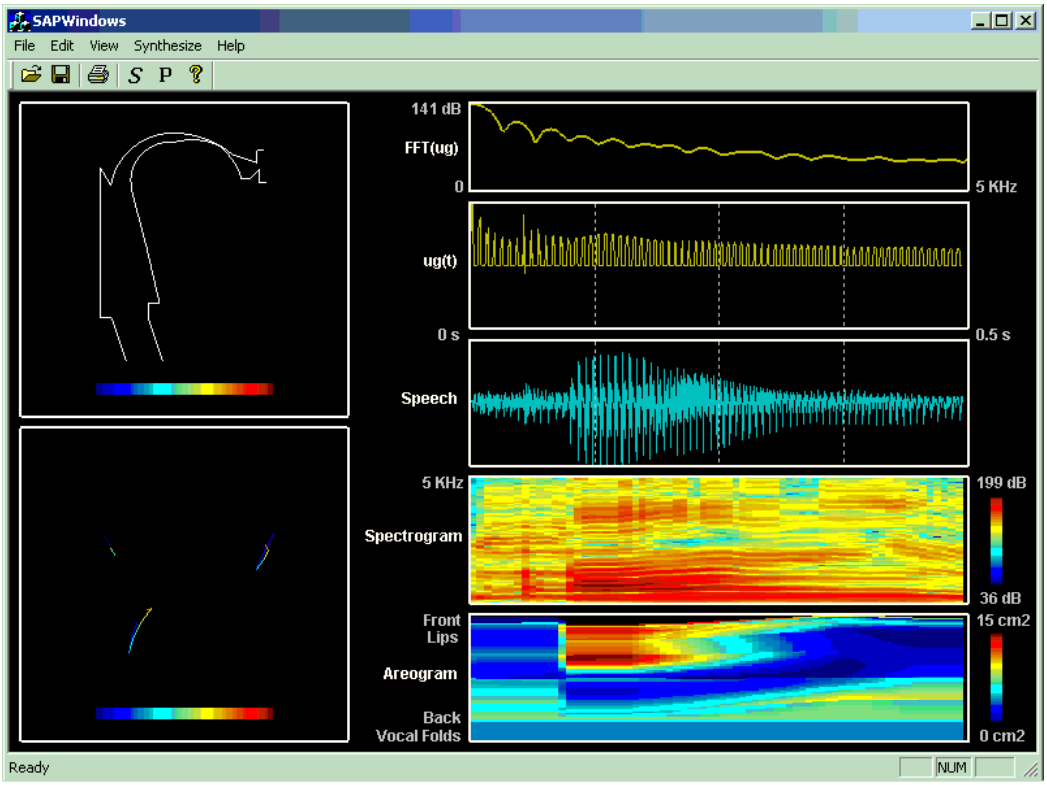


Figura 2.4: Interface gráfica do sintetizador SAPWindows (fonte: Silva, 2001). À esquerda, encontra-se representado o contorno sagital do tracto vocal (em cima) e a evolução de três parâmetros articulatórios (lábios, velo e hióide) ao longo do tempo (em baixo). À direita, podem ver-se (de cima para baixo) duas representações da excitação glotal, o sinal de voz, o espectrograma e o areagrama.

1. adaptação do formato de saída do TADA, de modo a que este pudesse ser usado como *input* para o SAPWindows - no âmbito do sistema TADA, o modelamento da configuração do tracto vocal ao longo do tempo e geração do sinal acústico está a cargo do sintetizador CASY(Iskarous *et alii*, 2003), embora o sistema inclua também a possibilidade de interface com o sintetizador HLsyn (Stevens & Hanson, 2003). Neste sentido, foi necessário proceder aos necessários mapeamentos entre os parâmetros do CASY e os usados pelo SAPWindows. Esta tarefa foi facilitada pelo facto de ambos os sintetizadores assentarem em modelos articulatórios derivados do modelo proposto por Mermelstein (1973);

2. ajuste dos modelos do SAPWindows - foram necessários alguns ajustes às dimensões das estruturas fixas do modelo articulatório e alterações no funcionamento dos parâmetros relacionados com os lábios (abertura e protrusão), para além de adaptações ao modelo da fonte glotal para lidar com sons não vozeados.

2.2 Introdução à fonologia articulatória

O modelo linguístico proposto por Browman & Goldstein (1986, 1989, 1990b, 1992, 2000) sugere uma revisão drástica na heurística do tratamento dos fenómenos fonológicos, pugnando por uma visão unificadora dos aspectos fonéticos e fonológicos, encarados como níveis diferentes - o macroscópico e o microscópico - do mesmo sistema complexo. A integração da dimensão física e da dimensão cognitiva da linguagem - consideradas inconciliáveis pela teoria generativa - é conseguida à custa de um novo primitivo de análise, o *gesto articulatório*, uma unidade fonológica dinâmica, que vem substituir o tradicional traço distintivo.

Inspirados nos estudos de motricidade geral, conduzidos nos anos 70 e 80, que advogam o carácter abstracto das acções motoras, organizadas sob a forma de estruturas coordenadas (Fowler, 1977; Turvey, 1977; Kugler *et alii*, 1982), Browman e Goldstein modelam o gesto articulatório através de uma equação dinâmica e introduzem a noção de *pauta gestual* como forma de representação das relações de coordenação entre os vários gestos que constituem um determinado item lexical.

Uma grande variedade de fenómenos linguísticos, como variações alofónicas e vários tipos de coarticulação, entre outro tipo de processos ditos de fala rápida (e.g. assimilações, apagamentos ou inserções de segmentos) pode ser explicada a partir de alterações na magnitude e grau de sobreposição entre os gestos (Browman & Goldstein, 1990b; Gick, 1999b) [30]. Em resultado da sua dimensão espacio-temporal, os gestos não podem ser apagados (ou inseridos) das representações linguísticas, estando tão somente sujeitos a alterações na sua magnitude e/ou grau de sobreposição, o que se traduz em mudanças acústicas e perceptuais.

A secção que se segue será inteiramente dedicada à apresentação deste modelo gestual, até agora praticamente desconhecido dos investigadores portugueses. Esta estará inevitavelmente sujeita a algumas omissões e muitas simplificações, nomeadamente no que respeita ao sistema dinâmico que serve de suporte à implementação dos gestos. Para uma descrição completa da fonologia gestual de Browman e Goldstein e modelos associados sugere-se a leitura das referências originais, citadas ao longo do texto.

2.2.1 Do gesto abstracto às trajectórias dos articuladores

De acordo com os princípios da FA, delineados por Browman & Goldstein (1986, 1989, 1990b, 1992, 2000), a unidade fonológica por excelência é o *gesto articulatório*. Os gestos da FA correspondem, essencialmente, a uma intenção de movimento, no sentido de realizar uma determinada tarefa no tracto vocal: por exemplo, elevação do dorso da língua ou oclusão labial.

[30] No que respeita ao português, destaca-se o recente trabalho de Meireles & Barbosa (2008), que desenvolvem um estudo acústico-articulatório, no sentido de analisar a influência da taxa de elocução na reorganização lexical (reestruturação silábica) do PB (e.g. “abóbora”> “abóbra”). De acordo com os dados, o fenómeno, considerado tradicionalmente como uma queda da vogal pós-acentuada, resulta afinal da redução da magnitude dos gestos consonânticos pós-tónicos e do aumento da sobreposição entre eles, em virtude do aumento da taxa de elocução.

Variável do tracto	Articuladores Envolvidos
LP Protrusão Labial	lábios superior e inferior, mandíbula
LA Abertura Labial	lábios superior e inferior, mandíbula
TTCL Local de constrição ponta da língua	ponta e corpo da língua, mandíbula
TTCD Grau de constrição ponta da língua	ponta e corpo da língua, mandíbula
TBCL Ponto de constrição corpo da língua	corpo da língua, mandíbula
TBCD Grau de constrição corpo da língua	corpo da língua, mandíbula
VEL Abertura do velo	véu palatino
GLO Abertura glotal	glote

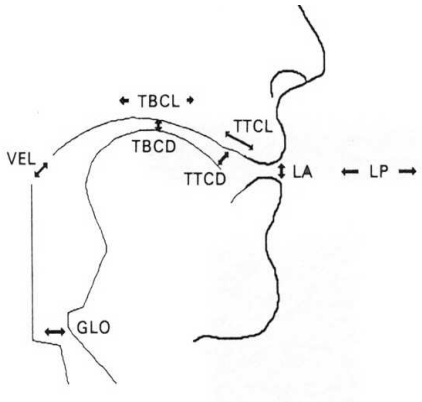


Figura 2.5: Variáveis do tracto e correspondência com os articuladores (traduzido de Browman & Goldstein (1990b, p.344). As aberviaturas referem-se à terminologia inglesa.

Esta intenção abstracta é especificada por um conjunto de cinco variáveis do tracto - ponta da língua (TT), corpo da língua (TB), velo (VEL), glote (GLO) e lábios (L) [31] - totalmente independentes umas das outras e que se referem, simultaneamente, a uma ou duas dimensões. Os gestos orais traduzem-se num par de variáveis do tracto, o local de constrição (do inglês *constriction location*) e o grau de constrição (do inglês *constriction degree*) [32], que se referem a duas dimensões da mesma constrição, estando, por isso, relacionadas entre si [33].

As variáveis do tracto consideradas pelo modelo e os articuladores envolvidos na realização da tarefa são apresentados na figura 2.5.

A variável do tracto *Constriction Location* (CL) pode assumir os valores: `[protruso]`, `[labial]`, `[dental]`, `[alveolar]`, `[pós-alveolar]`, `[palatal]`, `[velar]`, `[uvular]` e `[faríngeo]`.

Quanto ao *Constriction Degree* (CD), a FA prevê cinco valores distintos: `[fechado]`, `[crítico]`, `[estreito]`, `[médio]` e `[largo]`. Os primeiros dois caracterizam as oclusivas e fricativas, respectivamente. As restantes três etiquetas são usadas para descrever os contrastes de

[31] As abreviaturas referem-se à terminologia original, em inglês: *lips* (L), *tongue tip* (TT), *tongue body* (TB), *velum* (VEL) e *glottis* (GLO).

[32] Haveria ainda um terceira dimensão, a *forma de constrição*, que nunca chegou a ser incorporada ao modelo.

[33] Por exemplo, o local e grau de constrição do corpo da língua são duas dimensões da mesma constrição - corpo da língua.

altura entre vogais, embora o valor `[médio]` possa também aplicar-se às consoantes aproximantes.

As trajectórias das variáveis do tracto são explicitamente geradas por um modelo matemático conhecido como *task dynamics* (Saltzman, 1986; Saltzman & Munhall, 1989), originalmente usado para modelar diferentes tipos de movimento humano - quer seja andar, mastigar, estender um braço, etc. - e, mais recentemente, aplicado à produção de fala [34]. Segundo este modelo, o movimento é definido não em termos das estruturas anatómicas envolvidas, mas da "tarefa" abstracta a ser cumprida. Existem, contudo, muitos *graus de liberdade*, i.e., os articuladores envolvidos podem combinar-se entre si de maneiras diferentes, de modo a efectivar uma determinada tarefa. Se o movimento de um destes articuladores é bloqueado ou perturbado, os restantes articuladores responsáveis pela tarefa de imediato se ajustam, no sentido de cumprir o objectivo final e atingir o alvo (ou posição de equilíbrio).

No caso da fala em particular, há um conjunto de articuladores anatomicamente relacionados (*coordenative structure*) (Fowler, 1977) que actuam em conjunto e têm como "tarefa" a formação (e distensão) de uma determinada constrição em diferentes regiões do tracto vocal.

Por exemplo, a tarefa de oclusão labial, para a produção do /b/, resulta da acção coordenada de três articuladores - lábio inferior, lábio superior e mandíbula (vd. tabela 2.5) - que constituem o *effector system*. Os lábios destacam-se, neste caso, como *terminal devices* ou *end-effectors*, já que a sua posição define directamente a abertura labial. Independentemente do contexto segmental (e.g. [aba] ou [ibi]), a tarefa para o [b] - bem como o *effector system* e os *terminal devices* - é sempre a mesma (oclusão bilabial). O modo como os articuladores se coordenam entre si para a realizar é, contudo, bastante flexível: numa sequência como [ibi], a mandíbula tenderá a exibir uma posição elevada para a realização da vogal [i], pelo que a contribuição dos lábios no sentido de efectivar a oclusão bilabial será menor do que em contexto de vogal baixa (e.g. [aba]).

Assim, é o movimento das variáveis do tracto em direcção a um determinado alvo (ou ponto de equilíbrio), e não o movimento de cada um dos articuladores, que é caracterizado dinamicamente.

Na actual formulação do modelo, a formação (e distensão) de uma constrição gestual é modelada através de uma equação dinâmica simples, do tipo massa mola com amortecimento crítico (2.1):

$$m\ddot{x} + b\dot{x} + k(x - x_0) = 0 \tag{2.1}$$

em que,

m representa massa do objecto, considerada constante e adquirindo, portanto, o valor arbitrário de 1,

b representa o amortecimento do sistema (*damping ratio*), que está relacionado com a trajectória

[34] Uma introdução aos conceitos e pressupostos implicados na *task dynamics*, tendo como alvo não especialistas na matéria, pode ser encontrada em Hawkins (1992).

efectuada pela variável do tracto na aproximação ao *target*: na versão actual do modelo, todos os gestos orais são assumidamente criticamente amortecidos, o que significa que a variável do tracto se aproxima assimptoticamente do *target*, sem nunca oscilar em torno dele. Por outras palavras, a massa movimenta-se em direcção ao *target*, mas não chega jamais a atingi-lo.

k representa a rigidez da mola (*stiffness*), que, em conjunto com o amortecimento (b), é o parâmetro mais directamente responsável pela frequência de oscilação e pela duração da trajectória da variável do tracto: quanto maior a rigidez, mais alta a frequência de oscilação e menor a duração do movimento. Apesar de formalmente incorporado no modelo *task dynamics*, que serve de base à FA, permanecem ainda muitas dúvidas sobre o papel do *stiffness* na produção, a sua relação com x_0 ou a necessidade de considerar valores distintos para duas variáveis do tracto relacionadas (CL e CD) (Browman & Goldstein, 1990a). Browman & Goldstein (1990a, p. 306) especulam ainda sobre a possibilidade do *stiffness* "could form the basis for natural classes. For example, gestures for glides might differ from those for vowels primarly in their stiffnesses (glides being stiffer); similarly, gestures for stops (and affricates) might be stiffer than those for fricatives.". Investigações recentes, com base em articulografia electromagnética 3D (Roon *et alii*, 2007), indicam, contudo, que o referido parâmetro não depende do modo de articulação, mas varia, sobretudo em função do articulador envolvido, estando os movimentos do corpo da língua associados a um *stiffness* significativamente inferior ao de outros articuladores, como a ponta da língua ou os lábios, independentemente do falante e do contexto.

$\ddot{x}$ representa a aceleração instantânea da variável do tracto,

$\dot{x}$ representa a velocidade instântanea da variável do tracto,

x representa o deslocamento instantâneo da variável do tracto,

x_0 representa a posição de equilíbrio (ou *target*) da variável do tracto, i.e., a posição em direcção à qual todo o sistema se movimenta. Este parâmetro está intimamente relacionado com a tradicional caracterização dos segmentos fonéticos em termos de modo e ponto de articulação, análogos do CD e CL, respectivamente.

A especificação de um gesto implica, portanto, a referência aos valores dos parâmetros dinâmicos associados à variável do tracto em causa [35], para além da especificação da contribuição relativa (os chamados *weights*) de cada um dos articuladores para o movimento da variável do tracto.

2.2.2 Do gesto à palavra

Os gestos podem combinar-se entre si, de modo a formar estruturas gestuais mais vastas (segmentos, sílabas, palavras), sendo que as palavras são consideradas pela FA como "organized "molecules"

[35] Uma vez que os gestos orais se traduzem em CL e CD, é necessário definir o valor dos parâmetros referidos para ambas as variáveis do tracto, o que quer dizer que cada uma delas é modelada por uma equação dinâmica distinta.

composed of multiple articulatory gestures (the "atomic" units)" (Goldstein *et alii*, 2006, p.224).

Cabe notar que, embora os segmentos possam ser interpretados como um conjunto organizado de gestos, as palavras, traduzidas em "moléculas" ou "constelações gestuais", não resultam da simples concatenação dos gestos que formam cada um dos segmentos da sequência. Uma das evidências neste sentido provém da análise do grupo consonântico [sp] - presente em palavras inglesas como *spot* ou *spill* - no qual dois gestos orais, necessários à produção das consoantes, se combinam, não com dois (conforme seria de esperar numa análise tradicional), mas como um único gesto de abertura glotal (Browman & Goldstein, 1986). Só assim se justifica a ausência de aspiração (que habitualmente caracteriza as oclusivas do inglês) nas sequências [sp], [st] e [sk].

2.2.3 O gesto articulatório: unidade de acção e informação

Do que ficou dito até então, ressalta o duplo papel desempenhado pelos gestos articulatórios, enquanto "units of action and units of information (contrast and encoding)" (Goldstein *et alii*, 2006, p.217). Sendo o gesto articulatório uma unidade ao mesmo tempo dinâmica - já que definido por uma equação dinâmica que modela o movimento dos articuladores ao longo do tempo, no sentido da formação de uma determinada constrição no tracto vocal - e discreta - na medida em que pode ser usado para distinguir e contrastar enunciados - é possível, através dele, fazer a ponte entre o nível fonológico e o nível da implementação fonética, concebidos como duas dimensões (macroscópica e microscópica, respectivamente) de um mesmo sistema complexo. Por outras palavras, em consequência deste isomorfismo entre as propriedades fonológicas e fonéticas, estas correspondem à dimensão macroscópica e microscópica de um mesmo sistema, pelo que não há necessidade de qualquer tipo de "transdução" entre o nível fonológico e fonético.

Os gestos controlam articuladores (*organs*) independentes, tais como os lábios, a ponta da língua, o corpo da língua, a raiz da língua, o velo e a glote e, neste sentido, "articulatory gestures of distinct organs have the capacity to function as discretely different" (Goldstein *et alii*, 2006, p.222). Para além disso, o contraste lexical pode também ser assegurado por diferenças na própria parametrização dos gestos (Goldstein *et alii*, 2006). Nas palavras "pato" e "cato", o contraste resulta da substituição, em Ataque de sílaba, de um gesto labial por um gesto de corpo da língua. Já "saco" e "taco" partilham o mesmo articulador - a ponta da língua - e a diferença reside somente no valor de um dos parâmetros dinâmicos, o *target* da variável grau de constrição: um valor `[crítico]`, gerador de fricção, no primeiro caso, vs. um valor `[fechado]`, que se traduz numa oclusão completa, que caracteriza a oclusiva. Existem ainda outras possibilidades de diferenciação lexical que envolvem, por um lado, a presença/ausência de um gesto e, por outro, divergências no modo de coordenação de um mesmo conjunto de gestos (Browman & Goldstein, 1992, 1995a). Comparando, por exemplo, "pato" com "bato", há um gesto de abertura glotal, responsável pela ausência de vozeamento no [p], que não é activado durante a produção do [b] [36]. As palavras inglesas *mad* e *ban* são exemplos de "moléculas"

[36] A FA assume que "in speech mode, the larynx is positioned appropriately for voicing unless otherwise instructed"

compostas pelo mesmo conjunto de gestos, embora organizados de forma distinta (Goldstein *et alii*, 2006).

Do mesmo modo, a especificação do gesto em termos de tarefa abstracta a ser cumprida visa conferir ao gesto um carácter discreto. A formação da constrição é modelada através de uma equação dinâmica, cujos parâmetros (e.g. *target, stiffness, damping*) são independentes do contexto, i.e., permanecem inalterados ao longo de todo o intervalo de activação do gesto. Os valores dos parâmetros dinâmicos são, portanto, variáveis categorizáveis e definem as propriedades macroscópicas do gesto, permitindo-lhe estabelecer relações de contraste e oposição com outros gestos.

As trajectórias dos articuladores emergem a partir dos gestos, especificados como unidades dinâmicas invariantes, e dependem do contexto segmental, na medida em que os gestos se sobrepõem uns aos outros. Nas palavras de Goldstein (2005), "invariant dynamics at the task level shapes the time-varying, context-dependent dynamics at lower levels of the system (articulators and muscles)".

Quando dois (ou mais) gestos se encontram simultaneamente activos, estes podem envolver variáveis do tracto distintas ou partilhar a mesma variável (Saltzman & Munhall, 1989). No primeiro caso, os gestos vão atingir invariavelmente o *target* especificado, embora a contribuição de cada um dos articuladores usados para alcançar esse objectivo possa variar em função do contexto. Em situação de partilha da mesma variável do tracto (*blending*) - em que dois gestos competem entre si, tentando realizar tarefas distintas com estruturas articulatórias idênticas - os parâmetros associados a cada gesto, combinam-se entre si "either by simple averaging, weighted averaging or addition" (Hawkins, 1992).

Browman & Goldstein (1992) sumarizam, assim, a capacidade do gesto articulatório em unificar os dois aspectos da linguagem - o mental (considerado propriamente linguístico) e o físico - sem necessidade de regras *ad-hoc* ou quaisquer outras implementações adicionais:

> *Gestures can give rise to context-dependent articulatory and acoustic trajectories, without having to posit any 'implementation rules' for converting specific invariant (phonological) units into variable (physical) parameters. The variation follows directly from the definition of the units as parameterized task-dynamical systems, their phonological organization (pattern of overlap), and the general principles of how overlapping units blend. The same gesture structures simultaneously characterize phonological properties of the utterance (contrastive units and syntagmatic organization) and physical properties.* (Browman & Goldstein, 1992, p.165)

2.2.4 Coordenação intergestual

Ao contrário de outras correntes fonológicas, na FA, os gestos não se combinam entre si de uma forma sequencial, mas encontram-se organizados, como já referimos, em estruturas "moleculares" complexas, podendo sobrepôr-se uns aos outros, parcial ou totalmente. Assim, não basta determinar

(Browman & Goldstein, 1992, p.157).

o conjunto de gestos constitutivos de um segmento (ou de uma sequência deles), sendo igualmente importante caracterizar as relações de coordenação temporal estabelecidas entre eles. Browman e Goldstein adoptam o termo *gestural constellation* para referir o conjunto de gestos associados a uma determinada unidade lexical, cujas relações de coordenação intergestual se encontram explicitamente especificadas.

2.2.4.1 Relações de fase

Na versão original do modelo (e.g. Browman & Goldstein, 1990b), a coordenação intergestual é determinada por um conjunto de regras que especificam as *relações de fase* entre os dois sistemas dinâmicos que controlam a produção do gestos. Isto significa que um ponto particular da estrutura temporal (ou do movimento sinusoidal) de um gesto está sincronizado com um ponto específico de outro gesto. Os pontos de coordenação gestual são definidos por Browman e Goldstein com base num ciclo "virtual" de 360^o [37], cuja duração é determinada apenas pelo *stiffness*.

Para efeitos de sincronização, nem todos os pontos assumem a mesma relevância: o *target* (240^o) ou o início do movimento em direcção a este (0^o) são, normalmente, suficientes para assegurar a coordenação entre a maioria dos gestos (Browman & Goldstein, 1990a).

No âmbito do modelo, este mesmo ciclo desempenha (a par com o *stiffness*) uma outra função relacionada com o cálculo dos intervalos de activação dos gestos: cada um dos gestos permanece activo durante uma determinada porção do seu ciclo virtual, sendo que esta porção é assumidamente diferente para vogais e consoantes.

Gafos (2002) propõe uma notação mais intuitiva para indicar o tipo de relação temporal [38] estabelecida entre gestos consecutivos. Os vários pontos - denominados pelo investigador de *landmarks* - identificados por Gafos (2002) durante a trajectória de um gesto são ilustrados na figura 2.6, em comparação com o ciclo "virtual", previsto por Browman & Goldstein (1990b). O *onset* corresponde ao ponto em que o gesto inicia o seu movimento em direcção a um determinnado alvo; o *target* diz respeito ao momento em o gesto atinge o alvo; a *release* acontece quando a constrição se desfaz e o gesto começa a afastar-se do alvo; e a *release-offset* é definida como o ponto em que o gesto deixa simplesmente de estar activo. A região compreendida entre o *target* e a *release* denomina-se *gestural plateau* e corresponde a um período da trajectória em que o movimento se mantém mais ou menos constante. O ponto médio do *plateau* é chamado de *c-center*.

Na esteira das simulações conduzidas por Browman (1994) [39], Gafos (2002) assume que a *release* é parte integrante da especificação do gesto, na medida em que esta é activamente controlada, tal como o movimento em direcção em alvo: "a gesture's active control regime is the interval between the onset of movement and the release offset landmarks. During this interval, the movement of the

[37] Este seria o ciclo gerado se o sistema fosse não-amortecido (b=0).

[38] Gafos (2002) define a coordenação intergestual como "a relation between two gestures stating that a specified landmark (within the temporal structure) of one gesture is synchronous with a specified landmark of another gesture" .

[39] Recentemente, também Nam (2007) analisou e modelou o comportamento da *release*, no contexto da *task dynamics*.

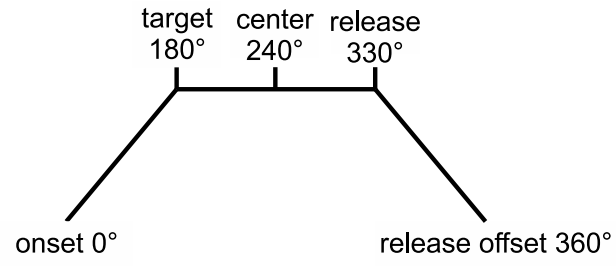


Figura 2.6: Marcas gestuais propostas por Gafos (2002) *versus* ciclo oscilatório de Browman & Goldstein (1990b) (fonte: Davidson, 2003).

gesture is actively controlled. Movement may continue after the release offset, due to the articulators' biomechanical inertia or due to movement associated with a different gesture. But such movement is not part of the linguistically-significant goal of the associated gesture." (Gafos, 2002).

Browman & Goldstein (1990b) determinam que, nas sequências CV, o centro do gesto consonantal (240^o) está sincronizado com o início da vogal (0^o), enquanto nas sequências VC, o centro da consoante (240^o) coincide com a *release* da vogal (330^o).

Em relação às sequências de consoantes (CC) em Ataque de sílaba, várias investigações (Browman & Goldstein, 1988; Byrd, 1994; Honorof & Browman, 1995) demonstraram que os gestos orais pré-vocálicos se comportam como um todo em relação à vogal seguinte, exibindo o chamado efeito *c-center*. Este é calculado da seguinte forma: "for every consonantal gesture, the (temporal) midpoint between the left and right edges of the plateau was computed. The c-center of a sequence is the mean of all the midpoints of the gestures in that sequence." (Browman & Goldstein, 1988, p.144). Segundo Browman & Goldstein (2000), a estabilidade do *c-center* resulta da acção conjunta de dois tipos de coordenação, que exercem pressão sobre o Ataque silábico: por um lado, cada um dos gestos que compõem o Ataque está sincronizado individualmente com o gesto vocálico (coordenação CV) e, ao mesmo tempo, os gestos consonânticos estabelecem entre si uma relação temporal (coordenação CC). Browman & Goldstein (2000) concluem que esta última coordenação se sobrepõe à primeira, na medida em que o rigoroso cumprimento da sincronização CV implicaria que os dois gestos consonânticos tivessem lugar ao mesmo tempo, o que poria em risco a recuperação perceptual das consoantes.

De acordo com os mesmos dados, o efeito *c-center*, geralmente, não se manifesta em contexto de Coda complexa. Neste caso, apenas a consoante mais à esquerda se coordena com a vogal.

As várias tipologias de coordenação gestual encontram-se representadas, de forma esquemática, nas figuras 2.7 e 2.8, usando a notação de Gafos (2002).

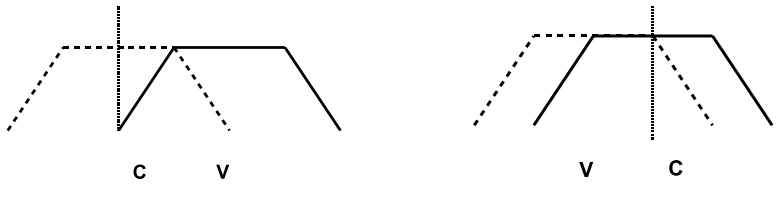


Figura 2.7: Relações de coordenação entre consoantes (C) e vogais (V) em sequências CV e VC (linha= consoantes; tracejado= vogais; a linha vertical assinala o ponto de alinhamento) (fonte: Davidson, 2003).

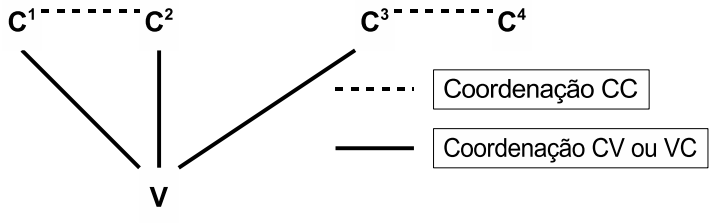


Figura 2.8: Relações de coordenação entre consoantes (C) e vogais (V) em Ataque e Coda complexas (linha= coordenação CV ou VC; tracejado= coordenação CC) (fonte: Davidson, 2003).

2.2.4.2 Osciladores acoplados

O modelo de coordenação intergestual, baseado no alinhamento manual de pontos específicos dos ciclos oscilatórios de cada um dos gestos, foi revisto recentemente. Actualmente, quer os intervalos de activação dos gestos, quer a coordenação intergestual - que subjazem à formação da pauta gestual (vd. mais informações nos parágrafos abaixo) - são controladas (ou planeadas) por um conjunto de osciladores acoplados. Inicialmente desenvolvido para coordenar pares de gestos (i.e. dois osciladores), *o modelo dinâmico dos osciladores acoplados*, proposto por Saltzman & Byrd (2000), foi alargado por Nam & Saltzman (2003), no sentido de lidar com acoplamentos múltiplos (potencialmente em competição).

Face à complexidade deste modelo, limitamo-nos a descrever, sucintamente, a ideia central que subjaz ao seu funcionamento. Para mais informações, recomenda-se a leitura dos artigos referenciados ao longo desta exposição. Pormenores adicionais poderão ainda ser encontrados no capítulo 4.

Em termos muito gerais, cada gesto está associado a um oscilador (ou *clock*) com uma determinada frequência. No início do processo, cada um dos osciladores entra, isoladamente, em movimento numa fase aleatória do seu ciclo. A coordenação entre dois gestos é determinada pelo acoplamento entre os osciladores correspondentes. O acoplamento entre um determinado par de gestos é controlado por uma função potencial (função coseno), definida a partir da diferença de fase (fase relativa) entre os dois osciladores e um mínimo, que corresponde ao *target* pretendido. Sempre que a fase relativa é diferente do *target*, a função potencial actua sobre cada um dos osciladores, no sentido de forçar a aproximação ao *target* definido. Assim que este é atingido, e o processo de coordenação entre os dois osciladores estabiliza, os gestos são activados pelo respectivo oscilador, o que acontece

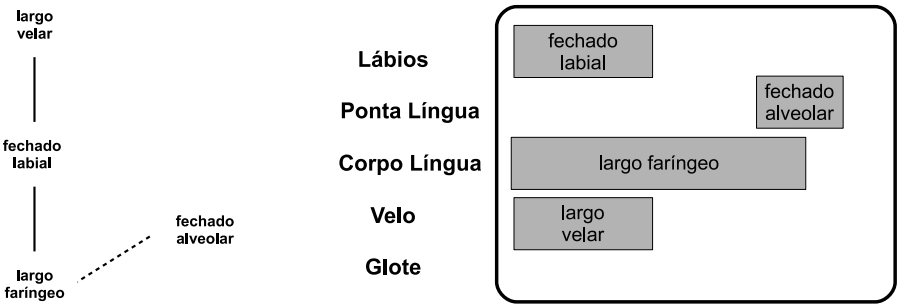


Figura 2.9: *Coupling graph* (à esquerda) e pauta gestual (à direita) relativos à palavra "mar". No *coupling graph*, as linhas a negrito indicam um acoplamento em fase, enquanto a linha a tracejado se refere a um acoplamento em anti-fase.

normalmente aos 0^o.

As relações de acoplamento entre gestos são especificadas num *coupling graph*, considerado parte integrante da representação fonológica de um enunciado.

No *coupling graph* da figura 2.9, relativo à palavra "mar", as linhas indicam o tipo de acoplamento estabelecido entre os dois osciladores que controlam os gestos. O oscilador `[fechado labial]` encontra-se acoplado ao oscilador vocálico `[faríngeo largo]` e também ao oscilador do velo `[largo]`. A linha a negrito refere-se a um acoplamento em-fase (0^o) entre estes três osciladores, enquanto a linha a tracejado, que associa o oscilador `[fechado alveolar]` à vogal, representa um modo de acoplamento anti-fase (180^o).

Os intervalos de activação dos gestos são calculados a partir destes *coupling graphs* e representados em "pautas gestuais" (do inglês *gestural scores*), diagramas bidimensionais, em que os gestos de diferentes estruturas articulatórias estão dispostos em "camadas" (*tiers*) distintas. Cada rectângulo corresponde ao intervalo de activação da variável do tracto associada ao gesto.

Retomando o exemplo da palavra "mar", na pauta gestual da figura 2.9, estão representados quatro gestos: um gesto `[fechado labial]`, praticamente concomitante com um gesto `[largo]` do velo, o que perfigura uma consoante nasal ([m]), um gesto `[faríngeo largo]` do corpo da língua (vogal [a]) e um gesto `[fechado alveolar]` de ponta da língua ([ɾ]).

No interior dos rectângulos, as etiquetas remetem para os parâmetros dinâmicos, nomeadamente o valor numérico da posição de equilíbrio que caracteriza uma determinada variável do tracto (vd. capítulo 4). No caso dos gestos orais, há duas etiquetas, uma para o CD e outra para o CL. Por exemplo, no tocante ao gesto de ponta da língua, especificado como `[fechado alveolar]`, `[fechado]` remete para um grau de constrição de -3.5mm - o que significa que a língua comprime o palato - enquanto `[alveolar]` indica o local de constrição, neste caso 56^o [40].

De acordo com as experiências sobre a coordenação motora em geral - nomeadamente dos movimentos oscilatórios das pernas, dedos ou braços (Turvey, 1990) - são apenas dois os modos de

[40] 90^o corresponde a uma constrição palatal.

coordenação espontaneamente disponíveis: coordenação em fase (0^o) ou em anti-fase (180^o). Outras relações de fase podem, em princípio, ser adquiridas, mas somente à custa de algum tipo de aprendizagem.

Sempre que a frequência de oscilação aumenta, observam-se transições espontâneas do modo anti-fase para o modo em fase, mas o contrário não se verifica (Haken *et alii*, 1985), o que quer dizer que a frequência não exerce qualquer tipo de influência sobre o modo de coordenação em fase. Com base nestes dados, este último modo de coordenação (0^o) é considerado mais estável do que a coordenação anti-fase (180^o).

Segundo Goldstein *et alii* (2006), sempre que possível, os sistemas fonológicos exploram estes dois modos de coordenação básicos, que são a base da estrutura interna da sílaba. O Ataque e a Coda representam as duas possibilidades de coordenar um gesto consonântico (C) com um gesto vocálico (V): em fase e em anti-fase, respectivamente.

Do mesmo modo, também o Ataque complexo é definido por uma relação em fase entre cada um dos gestos C e a vogal (Núcleo de sílaba), o que implica um total sincronismo entre todos os gestos, inclusivé entre os gestos consonânticos (vd. secção 2.2.4.1). Se a combinação de gestos definidos como `[crítico]` (fricativas) ou `[fechado]` (oclusivas) com outros gestos, como o velo ou a glote, resulta em estruturas perfeitamente recuperáveis do ponto de vista perceptual - tradicionalmente analisadas como segmentos únicos (e.g. nasais, laterais) - noutros casos a produção simultânea de vários gestos C compromete a diferenciação perceptiva entre consoantes. Daí a necessidade de prever, a par da coordenação CV, uma relação anti-fase entre as consoantes (CC), que garanta, pelo menos parcialmente, a realização sequencial dos gestos. A relação de competição entre dois modos de coordenação distintos traduz-se no já referido (vd. secção 2.2.4.1) efeito *c-center* (Browman & Goldstein, 2000), atestado já para várias línguas (Browman & Goldstein, 1988; Honorof & Browman, 1995; Marin & Pouplier, 2008; Hermes *et alii*, 2008; Pouplier, 2008). Este determina que "adding additional consonants to an onset changes the resultant phase of all consonant gestures with respect to the vowel in a way that preserves the overall timing of the *center* of the consonant sequence with respect to the vowel" (Nam & Saltzman, 2003, p.2253).

Também já aqui foi dito (vd. secção 2.2.4.1) que - ao contrário do Ataque - em Coda, o efeito *c-center* não se manifesta de uma forma consistente [41]. Isto significa que não há qualquer competição e apenas a primeira consoante está coordenada (em anti-fase) com a vogal, enquanto as consoantes estão coordenadas entre si, também em anti-fase.

As simulações de Nam & Saltzman (2003), a partir dos *coupling graphs* apresentados na figura 2.10, mostram que o modelo de osciladores acoplados é eficaz no modelamento das assimetrias fonéticas que caracterizam a estrutura silábica. Adicionalmente, este é capaz de prever as diferenças na variabilidade dos gestos em Ataque e Coda: a coordenação CC é mais estável em Ataque do que

[41] Os resultados de Byrd (1995), para o inglês, e de Pouplier (2008), para o alemão, sugerem que também as Codas complexas podem, em alguns casos, exibir efeito de *c-center*.

em Coda.

Figura 2.10: Coordenações gestuais entre gestos vocálicos (V) e consonânticos(C) em Ataque (à esquerda) e Coda (à direita) complexas (fonte: Nam & Saltzman, 2003).

Dados experimentais recentes (Goldstein *et alii*, 2007; Shaw & Gafos, 2008; Hermes *et alii*, 2008) sugerem ainda que o efeito *c-center* pode ser um bom indicador da afiliação silábica. Goldstein *et alii* (2007), por exemplo, verificam a ocorrência do *c-center* nos grupos consonânticos em início de sílaba do georgiano, mas não do tashlhiyt berber [42]. Este resultado está de acordo com as descrições fonológicas anteriores, que sugerem que - ao contrário do que acontece no georgiano (e no inglês) - no berber, os Ataques ramificados não são autorizados.

Vários outros trabalhos (e.g. Chitoran *et alii*, 2002; Zeroual *et alii*, 2008) mostram uma influência de vários factores - como a posição na palavra, o ponto de articulação, o modo de articulação e o vozeamento - na organização temporal das sequências CC.

De acordo com Goldstein *et alii* (2006), uma grande parte das propriedades fonológicas da sílaba (e.g. formato não-marcado da sílaba CV, possibilidades combinatórias, ressilabificação) poderá ser explicada à luz desta teoria.

Na grande maioria das línguas, as consoantes em Ataque de sílaba podem combinar-se livremente com as vogais do Núcleo, enquanto as sequências VC ou CC (Coda ramificada ou Ataque ramificado) estão, normalmente, sujeitas a grande restrições. No sentido de justificar essas assimetrias - que, a par com outros argumentos, sustentam a tradicional divisão binária da sílaba em Ataque e Rima - Goldstein *et alii* (2006) defendem uma relação entre acoplamento e restrições combinatórias: os gestos podem combinar-se entre si livremente, desde que estejam coordenados em fase, já que este corresponde ao modo de coordenação mais estável. Consequentemente, quase todas as combinações CV são possíveis, devido, quer à estabilidade e disponibilidade do acoplamento em fase, quer à própria natureza dos gestos vocálicos e consonânticos: porque possuem propriedades distintas, C e V podem ser realizados em fase (portanto, em total sincronismo), sem perdas de informação, ao nível perceptual, o que não acontece quando em causa estão dois gestos consonânticos, como referimos em parágrafos anteriores. Os gestos vocálicos distinguem-se dos gestos consonânticos, quanto ao grau de constrição, mas são também mais lentos a atingir o *target* (com naturais implicações ao nível do valor do *stiffness*), permanecendo activos durante mais tempo.

A estabilidade da coordenação em fase justifica também a universalidade da sílaba CV, em comparação com o formato VC.

A dependência entre gestos V e C, em sequências VC, é muito maior, em virtude da instabilidade que caracteriza o modo de coordenação anti-fase. Já os Ataques e Codas ramificadas são

[42] O tashlhiyt berber é uma língua falada no ocidente de Marrocos, por cerca de 5 milhões de pessoas.

definidas por modos de coordenação não-espontâneos e, consequentemente, adquiridos através da aprendizagem, em fases posteriores da aquisição.

Esta abordagem permite ainda explicar a tendência de ressilabificação de consoantes intervocálicas em Coda para Ataque, em consequência do aumento da taxa de elocução, como um resultado da preferência por um modo de coordenação mais estável.

2.2.5 Sistema TADA (*Task Dynamics Application*)

Os princípios da FA, incluindo o modelo de osciladores acoplados - que rege a coordenação intergestual - foram integrados num modelo dinâmico da produção de fala, em desenvolvimento nos Laboratórios Haskins (Saltzman & Munhall, 1989; Browman & Goldstein, 1992; Saltzman & Byrd, 2000; Nam & Saltzman, 2003).

A aplicação TADA, desenvolvida em ambiente MATLAB (Nam *et alii*, 2004), é a mais recente implementação do modelo computacional e incorpora os seguintes componentes (cf. Browman *et alii*, 2001-2006):

1. *Syllable structure-based* ***gestural coupling model*** - este modelo recebe como entrada texto (na versão original, texto em inglês), em formato ortográfico ou fonético (ARPABET) [43], e gera, à saída um *coupling graph*, especificando não só os gestos (representados em termos de parâmetros dinâmicos que caracterizam a variável do tracto e peso relativo dos articuladores) associados aos segmentos de entrada e a sua posição na sílaba, como também as relações de coordenação intergestual, determinadas pelo acoplamento entre os osciladores;

2. ***Coupled oscillator model*** *of intergestural coordination* - a partir do *coupling graph*, o modelo calcula os intervalos de activação para cada um dos gestos e gera a respectiva pauta gestual;

3. ***Task dynamical model*** *of inter-articulator coordination* - com base na pauta gestual, o modelo calcula as trajectórias dos articuladores que constituem a estrutura coordenada, de acordo com os princípios gerais (e universais) da *task dynamics*.

Numa fase terminal, o sistema TADA integra um módulo, o sintetizador *strictu sensu*, que manipula parâmetros físicos concretos correspondentes aos movimentos dos articuladores da fala para a produção de sinal acústico.

Originalmente, o modelamento da configuração do tracto ao longo do tempo é da responsabilidade do sistema CASY (Iskarous *et alii*, 2003). As frequências de ressonância e as larguras de banda, calculadas a partir das funções de área, são, posteriormente usadas para gerar o sinal acústico.

[43] O alfabeto fonético ARPABET foi desenvolvido pela ARPA (*Advanced Research Projects Agency*, actual *Defense Advanced Research Projects Agency* - DARPA), no âmbito do projecto *Speech Understanding Project* (1971-1976). Cada fone é representado por uma ou duas letras maiúsculas e os números, colocados logo após a vogal, são usados como indicadores de acento.

O modelo acústico utilizado apresenta, no entanto, algumas limitações - relacionadas com o controlo da fonte (fricção, aspiração, F0) e ausência de nasalidade - naturalmente supridas pela possibilidade de interface com o pseudo-sintetizador articulatório HLsyn (Stevens & Hanson, 2003).

Recentemente (Saltzman *et alii*, 2008), este modelo foi adaptado no sentido de acomodar vários aspectos da estrutura prosódica (pé, palavra, sintagma, frase). Um novo conjunto de *modulation gestures* (μ gestures) foi adicionado ao modelo, de modo a controlar as propriedades espaciais e temporais de todos os gestos activos. Uma vez que a interacção prosódia-segmentos não será tida em conta neste trabalho, a questão do gesto prosódico e do funcionamento deste novo componente não será aqui descrito em detalhe. Para mais pormenores sobre esta matéria recomenda-se a leitura de Saltzman *et alii* (2008) e Nam *et alii* (2006).

2.3 Modelo de produção para o português europeu

O diagrama do modelo de produção do PE, desenvolvido com base no sistema TADA, é apresentado na figura 2.11. Este é constituído por três grandes blocos: 1) módulos de processamento linguístico; 2) sistema TADA, adaptado para o PE; 3) sintetizador articulatório SAPWindows [44].

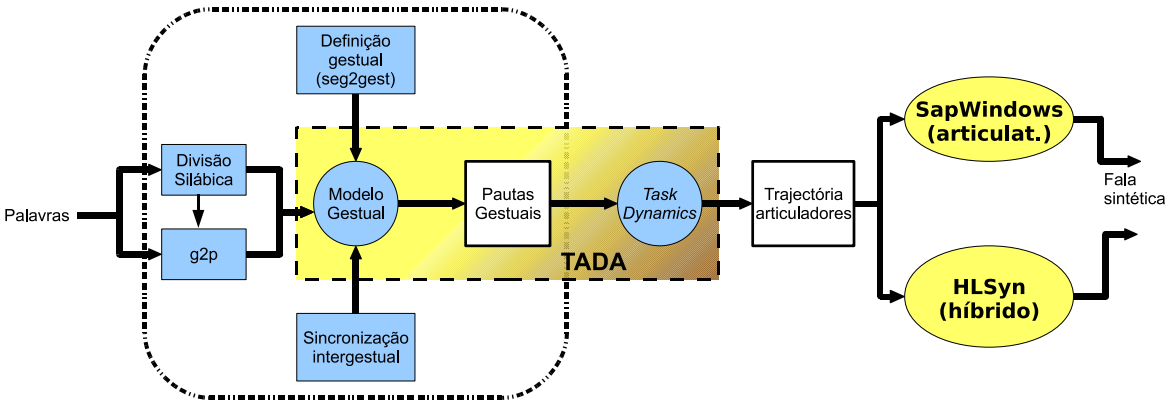


Figura 2.11: Diagrama do sistema de conversão texto-fala de base articulatória, para o português europeu, desenvolvido com base na integração do TADA com o SAPWindows.

O *processamento linguístico* inclui um módulo de transcrição ortográfico-fonética (G2P) e um módulo de divisão silábica. Esta primeira etapa - como já referimos na Introdução a este trabalho - tem como objectivo a transformação do texto de entrada num formato adequado ao TADA e é essencial ao correcto funcionamento do modelo de coordenação intergestual, baseado na informação silábica. Informações detalhadas sobre o desenvolvimento e avaliação de um conjunto de procedimentos para a silabificação e transcrição automáticas do léxico do PE serão apresentadas no capítulo 3.

O segundo módulo desenvolvido consiste na adaptação do sistema TADA ao PE, que se

[44] Como já referimos, o SAPWindows apresenta algumas limitações, no que se refere ao tipo de sons que permite sintetizar, pelo que, ocasionalmente, recorremos ao sintetizador pseudo-articulatório HLsyn.

traduziu na introdução de algumas alterações ao modelo gestual - que, como já vimos, está dividido em duas componentes, *syllable structure-based gestural coupling model* e *coupled oscillator model of intergestural coordination* (vd. 2.2.5) - nomeadamente a criação de um novo dicionário gestual (seg2gest), específico para o português, e a modificação de alguns padrões de coordenação, tendo em vista a obtenção de uma pauta gestual a partir do texto escrito.

Cabe depois ao terceiro módulo do sistema TADA calcular o movimento dos articuladores, a partir da pauta gestual, de acordo com os princípios da *task-dynamics*. Sendo estes universais, o que significa que funcionam para todas as línguas de igual forma, não foi necessário levar a cabo qualquer tipo de alteração ao modelo matemático, responsável pelo cálculo das trajectórias dos articuladores.

Capítulo 3

Processamento Linguístico

Como deixámos expresso na Introdução, o principal objectivo deste trabalho reside no desenvolvimento de um modelo articulatório dinâmico para o PE, com base num sistema pré-existente, auto-intitulado TADA, que oferece a possibilidade de modelar dinamicamente o léxico de outras línguas, mediante a criação de novos dicionários gestuais. Antes de empreender tal tarefa, foi, no entanto, necessário, criar um conjunto de recursos que visaram a normalização do texto de entrada do TADA, i.e, a transformação do texto escrito num formato adequado ao sistema computacional que nos serve de referência.

Assim, o presente capítulo é dedicado à apresentação dos módulos que asseguram a silabificação automática e a transcrição ortográfico-fonética das palavras portuguesas a processar pelo TADA. Ainda que necessários, tendo em conta as dificuldades de acesso a ferramentas deste tipo para o PE, os referidos módulos não fazem parte do núcleo central do sistema computacional, funcionando de forma totalmente independente.

A primeira parte do capítulo é reservada à questão da divisão silábica e nela se descreve o essencial da implementação e teste de dois algoritmos de silabificação automática: o primeiro, baseado em transdutores de estados finitos, e o segundo desenvolvido a partir das propostas de Mateus & d'Andrade (2000). Adicionalmente, pareceu-nos também pertinente incluir, nesta primeira parte, algumas informações teóricas adicionais acerca da estrutura interna da sílaba do PE. Dar-se-á notícia não só das certezas consagradas sobre a estrutura da sílaba do PE, mas também dos casos problemáticos, que ainda não foi possível integrar no quadro dos princípios fonológicos universais que regem o seu funcionamento. Este enquadramento teórico terminará com uma síntese dos principais estudos relativos à silabificação automática do português.

Os módulos de conversão grafema-fone, baseados quer em regras linguísticas, quer em métodos de aprendizagem automática, são apresentados na segunda parte do capítulo, após um breve enquadramento teórico do trabalho efectuado, que constou essencialmente de uma revisão dos estudos anteriores que se dedicaram à questão da transcrição ortográfico-fonética do PE.

3.1 Silabificação automática

> *Just as vowels with consonants are the matter of syllables, so also syllables are the matter for the construction of nouns and verbs, and of the elements which are made of them.*
>
> António, Retórico de Tagrit, *Knowledge of Rhetoric*, Book V, Canon I

3.1.1 Introdução

Apesar das abundantes evidências, provenientes das mais variadas áreas da linguística, sobre a importância fundamental da sílaba na estrutura das línguas, esta continua a carecer de uma definição abrangente e amplamente aceite por todos.

A fonologia, a fonética e a psicolinguística desenvolveram abordagens diversificadas, ainda que complementares, no sentido de evidenciar o papel da sílaba enquanto unidade linguística e esclarecer questões relativas à sua natureza e papel no funcionamento das línguas.

Esta natureza problemática da sílaba terá levado Chomsky & Halle (1968) a dispensar-lhe muito pouca atenção [1].

Também em Portugal, os anos 60 e 70 se caracterizam pela aplicação de modelos teóricos que não reconhecem a importância e funcionalidade desta unidade fonológica (Barbosa, 1965; Mateus, 1975; d'Andrade, 1977).

Evidências provenientes dos mais variados quadrantes avolumaram a necessidade da fonologia dispensar uma maior atenção a tal unidade. A partir dos trabalhos de Fudge (1999), Hooper (1972) e, sobretudo, de Kahn (1976), a sílaba foi gradativamente sendo reconhecida como unidade fonológica. Actualmente, a sua pertinência é amplamente reconhecida pela maioria das correntes fonológicas, nomeadamente pelos chamados modelos "pós-*SPE*" ou correntes não-lineares (multilineares).

É na segunda metade da década de 80 que começam também a surgir os primeiros trabalhos sobre a estrututura silábica do PE (e.g. Barbeiro, 1986; Mateus, 1993, 1994; Mateus & d'Andrade, 1998; Vigário & Falé, 1993). O modelo mais adoptado para a descrição desta língua é a chamada representação em "Ataque-Rima".

Paradoxalmente à complexidade implicada na sua organização e à falta de uma definição

[1] Chomsky e Halle são os responsáveis pela formulação da teoria generativa na sua versão clássica (designada por "modelo *SPE*", abreviatura de *The Sound Pattern of English*). Em virtude de uma visão estritamente segmental e unilinear dos fenómenos fonológicos, unidades de extensão superior à do segmento, como a sílaba, são totalmente ignoradas. Blevins (1995, p.234, nota 2) considera, no entanto, que a referência à classe das vogais como o único grupo de segmentos [+silábico] é, por si só, um indício do importante papel desempenhado pela sílaba, no âmbito da fonologia generativa tradicional: "The use of the symbol V, a [+vocalic] segment, in countless phonological rules in *SPE*, with subsequent recognition that this natural class might be referred to [+syllabic] (...) can be viewed as acknowledgment of the syllable's important role in phonological theory".

clara e definitiva, a sílaba é frequentemente apontada como uma unidade facilmente intuída pelos falantes. Tal intuição revela-se através da capacidade que qualquer falante de uma língua parece possuir, independentemente do seu grau de instrução, para identificar as sílabas na cadeia fónica [2], "quer através da sua segmentação explícita (...), quer através de outras manipulações explícitas do material verbal baseadas em critérios fonéticos e fonológicos, como a contagem, o apagamento ou a inversão" (Veloso, 2003, p.82-83). O carácter fortemente intuitivo da sílaba é, porventura, um dos principais argumentos que fundamenta a defesa da sílaba enquanto unidade fonológica (Veloso, 2003; Freitas & Santos, 2001; Barroso, 1999; Mateus & d'Andrade, 2000; Blevins, 1995) [3].

Esta natureza intuitiva da sílaba contrasta, contudo, com a reconhecida falta de acordo dos falantes no que toca à delimitação de fronteiras silábicas. O problema agudiza-se quando está em causa a silabificação de uma ou mais consoantes intervocálicas (e.g. VCV, VCCV) (Blevins, 1995) ou em línguas onde o fenómeno de ambissilabicidade é mais recorrente.

O português, apesar de pertencer ao grupo de línguas que permite consoantes em Coda, não apresenta grandes dificuldades na silabificação de sequências VCV. De acordo com Vigário & Falé (1993), a consoante intervocálica é indubitavelmente Ataque de sílaba (V.CV). O padrão CV corresponde ao formato silábico preferencial em todas as línguas do mundo [4], inclusive do português [5], razão pela qual ele é tido como "formato não-marcado" (Blevins, 1995). Isto significa que todas as línguas têm Ataque [6], mas nem todas têm Coda. Com base nesta observação, numa sequência do tipo VCV, a consoante deve preencher a posição silábica de Ataque e não de Coda, tendência esta

[2]Recorde-se, por exemplo, como prova da consciência linguística da sílaba, por parte dos falantes: 1) os chamados *lapsus linguae*, i.e., trocas de informação linguística no interior de ou entre palavras (Veloso, 2003; Freitas & Santos, 2001; Mateus & d'Andrade, 2000; MacKay, 1978; Schiller, 1998; Beaulieu, 2001), onde os erros de produção não são aleatórios, mas são condicionados pela posição silábica; 2) as experiências sobre o chamado fenómeno *tip-of-the-tongue* (TOT), que mostram que os falantes, são, muitas vezes, capazes de identificar o número de sílabas de uma determinada palavra-alvo, sem que a consigam, no entanto, pronunciar e reconhecer os segmentos que a compõem (Burke *et alii*, 1991; Brown, 1991; Schiller, 1998; Krakow, 1989); 3) alguns lapsos de escrita, que parecem estar directamente relacionados com a dificuldade em reconhecer padrões silábicos complexos (Freitas & Santos, 2001); 4) as escritas de tipo silábico (Veloso, 2003; Barroso, 1999; Blevins, 1995); 5) a manipulação silábica em jogos linguísticos (Hombert, 1986; Demolin, 1991; Rousset, 2004; Veloso, 2003; Blevins, 1995; Mateus & d'Andrade, 2000; Freitas & Santos, 2001; Schiller, 1998; Bagemihl, 1995), como a língua dos "pês" (Veloso, 2003; Mateus & d'Andrade, 2000; Bagemihl, 1995; Rousset, 2004), a *Bi-Sprache* (Rousset, 2004) ou a "língua secreta de Alfama" (Veloso, 2003; Mateus & d'Andrade, 2000) ; 6) o papel da estrutura silábica CV, tanto nas primeiras produções das crianças, como na fala dos afásicos (Jakobson, 1969; MacNeilage, 1998; Freitas, 1997).

[3]As afirmações de Blevins (1995, p.209-210) ilustram a importância da intuição dos falantes enquanto evidência da presença da sílaba ao nível do seu conhecimento fonológico, o que justifica, por sua vez, a sua aceitação como objecto de estudo da fonologia: "In a number of languages, native speakers have clear intuitions regarding the number of syllables in a word or utterance, and in some of these, generally clear intuitions as to where syllable breaks occur. (...) If phonology is in part the study of the mental representations of sound structure, then such intuitions support the view of the syllable as a plausible phonological constituent.".

[4]Segundo o levantamento estatístico obtido a partir da base de dados *UCLA Lexical and Syllabic Inventory Database* (ULSID), especificamente desenvolvida com o intento de estudar as tendências de organização silábica em várias línguas do mundo, o tipo silábico CV é, sem dúvida, o mais favorecido, contabilizando, no total, 5299 ocorrências (cerca de 50% do *corpus*). Apenas 5 das 16 línguas analisadas possuem maioritariamente sílabas do tipo CVC (Rousset, 2004). Estes dois padrões silábicos - CV e CVC - perfazem cerca de 85 % das sílabas do *corpus*.

[5]Os trabalhos sobre a frequência dos tipos silábicos em PE (d'Andrade & Viana, 1993b; Vigário & Falé, 1993; Viana *et alii*, 1996; Vigário *et alii*, 2006) mostram que o formato CV é o mais representado nos *corpora* analisados.

[6]Pesquisas recentes sobre a língua Arrernte (ou Aranda) (Breen & Pensalfini, 1999) - falada na região de Alice Springs, na Australia - contrariam, de alguma forma, a pretensa universalidade do formato CV.

contemplada no *Princípio do Ataque Máximo* [7].

As ambiguidades e discrepâncias surgem em relação ao grupos/encontros consonânticos (VCCV), passíveis de serem considerados tautossilábicos (V.CCV) ou heterossilábicos (VC.CV) (Vigário & Falé, 1993; Barroso, 1999; Veloso, 2003).

Existem, no entanto, outros argumentos, de carácter intrinsecamente fonológico, que justificam o interesse da sílaba: por um lado, a sílaba precisa necessariamente de ser evocada na explicitação e formalização de certos processos fonológicas (Veloso, 2003; Blevins, 1995; Selkirk, 1999; Clements & Keyser, 1999; Goldsmith, 1990; Krakow, 1989; Kahn, 1976), por outro, é em relação aos constituintes silábicos que se verifica a maior parte das restrições fonotácticas (Veloso, 2003; Fudge, 1999; Blevins, 1995; Selkirk, 1999; Krakow, 1989).

As regras fonológicas são, de algum modo, condicionadas pela estrutura da sílaba e são muitos os processos descritos na literatura passíveis de afectar apenas os segmentos posicionados num determinado contexto silábico (Goldsmith, 1990; Blevins, 1995) [8]. Será este o caso da aspiração, muitas vezes associada às fronteiras silábicas (Blevins, 1995) ou da velarização da lateral /l/ em posição de Coda, um fenómeno recorrente em muitas línguas (cf., por exemplo, Sproat & Fujimura, 1993; Recasens *et alii*, 1995; Recasens & Espinosa, 2005), inclusive no português (Cunha & Cintra, 1997; Andrade, 1999, 1998).

Também a concepção de sílaba como combinação de segmentos que se organizam em torno de um elemento nuclear, detentor de uma sonoridade mais proeminente, parece reunir o consenso dos fonólogos (Veloso, 2003; Freitas & Santos, 2001; Selkirk, 1984; Blevins, 1995; Barroso, 1999; Goldsmith, 1990; Krakow, 1989). A partir deste "pico de sonoridade", os segmentos sucedem-se, em relação à margem direita e esquerda da sílaba, em sonoridade decrescente. Esta organização dos sons, no interior da sílaba, em função do seu grau de sonoridade, é conhecido por *Princípio de Sonoridade* e pode ser definido da seguinte forma [9]:

> *Numa sílaba, a sonoridade dos segmentos tem de decrescer a partir do núcleo até às suas extremidades. A sonoridade dos segmentos é definida pela seguinte escala, apresentada por ordem decrescente de sonoridade: Vogais - Líquidas - Nasais - Fricativas - Oclusivas.* (Vigário & Falé, 1993, p.473) [10]

Esta relação entre sílaba e sonoridade, reconhecida, pelo menos, desde finais do século XIX, está na base de algumas definições de sílaba. Transcrevemos duas:

[7] O *Princípio do Ataque Máximo* postula que é preferível o preenchimento dos Ataques ao preenchimento das Codas (cf. Veloso, 2003, para uma versão portuguesa deste princípio silábico).

[8] Segundo Blevins (1995, p.209) "phonological properties with the syllable as their domain include pharyngealization, stress, tone, and ballisticity".

[9] Adoptámos a formulação de Vigário & Falé (1993), baseada essencialmente em Selkirk (1984). Outras versões podem ser encontradas, por exemplo, em Goldsmith (1990) ou Blevins (1995).

[10] A escala de sonoridade silábica proposta pelas autoras e adoptada de Selkirk (1984), não inclui as semivogais, pelo que Mateus & d'Andrade (2000), Freitas (1997) e Freitas & Santos (2001) propõem uma outra escala que considera esta classe de segmentos: Oclusivas< Fricativas< Nasais< Líquidas< Semivogais< Vogais.

> *[A sílaba é] uma unidade rítmica, constituída por uma sequência de segmentos que se agrupam em torno de um segmento a que está associado maior grau de proeminência.* (Mateus *et alii*, 1990, p.211)

> *The syllable then is the phonological unit which organizes segmental melodies in terms of sonority; syllabic segments are equivalent to sonority peaks within these organizational units.* (Blevins, 1995, p.207)

O *Princípio de Sonoridade* está ligado a um outro instrumento de descrição silábica, a *Condição de Dissemelhança* (também chamada de *Princípio de Dissimilaridade*), que determina, para cada língua, a diferença de sonoridade permitida entre dois segmentos adjacentes dentro da mesma sílaba (Veloso, 2003; Mateus & d'Andrade, 2000; Freitas, 1997; Freitas & Santos, 2001; Selkirk, 1984; Goldsmith, 1990).

Vigário & Falé (1993) propõem uma Escala de Sonoridade indexada para o PE, com base na proposta de Selkirk (1984) para o inglês, e formulam a seguinte definição da *Condição de Dissemelhança*:

> *Os segmentos adjacentes numa mesma sílaba têm de ter entre si uma diferença de sonoridade igual ou superior a 4 (...), sendo sempre preferível um valor superior e sendo sempre marcada (ou impossível) uma sequência com um valor inferior.* (Vigário & Falé, 1993, p.474)

Com base neste pressuposto, uma sílaba ideal, tendo em conta a universalidade do formato CV, deverá ser constituída por uma oclusiva seguida de uma vogal (Freitas & Santos, 2001; Mateus *et alii*, 2005), uma vez que os segmentos se encontram nos extremos opostos da escala de sonoridade.

Estes dois mecanismos (*Princípio de Sonoridade* e *Condição de Dissemelhança*), em conjunto com outros princípios gerais formulados e aceites pela teoria fonológica [11], estão na base da atribuição de posições silábicas a cada um dos segmentos das palavras. Ao longo do presente capítulo, estes serão regularmente invocados, na medida em que explicam a naturalidade *versus* irregularidade de alguns agrupamentos segmentais, mas também porque regulam a definição e implementação do algoritmo de silabificação do português.

Do ponto de vista acústico e articulatório, a descrição da estrutura silábica continua a apresentar-se como um dos maiores desafios da Fonética. A ausência de critérios fonéticos claros, que permitam delimitar rigorosamente as sílabas no *continuum* sonoro, é evidenciada por vários autores (Barbosa, 1965, 1994a; Vigário & Falé, 1993; Goldsmith, 1990; Barroso, 1999) e contrasta com a já referida capacidade intuitiva, partilhada por todos os falantes, de identificação silábica.

Se a grande maioria dos estudos falhou na tentativa de provar que a sílaba é a unidade básica da produção (e.g. Stetson, 1951; Draper *et alii*, 1959), mais recentemente, uma nova abordagem sugere

[11] Para uma definição portuguesa destes e doutros princípios silábicos, consultar Veloso (2003).

que a sílaba pode estar associada a padrões articulatórios específicos (cf. 2.2.4.2). Neste sentido, a unidade sílaba é articulatoriamente definida e determinada pelas relações temporais entre classes específicas de gestos, ou, nas palavras de Browman & Goldstein (1995b, p.20), "syllable structure is a characteristic pattern of coordination among gestures".

No âmbito da aplicação TADA, que nos serve de referência, a entrada do sistema é um conjunto arbitrário de palavras em língua inglesa. Cada uma das formas é associada a uma entrada do *CMU English Pronunciation Dictionary* (Carnegie Mellon University, 1993). Como este não inclui qualquer tipo de informação silábica, é aplicado um algoritmo de silabificação automática.

No caso específico do português, até onde nos é dado conhecer, não é do domínio público qualquer dicionário electrónico com a representação fonética e/ou silábica das palavras da língua. Assim, este capítulo serve um objectivo eminentemente prático: o da silabificação automática de todas as palavras portuguesas que constituem o *input* do nosso sistema. Mais do que avaliar a eficácia dos instrumentos teóricos adoptados ou contribuir, de algum modo, para qualquer teoria linguística da sílaba, é nosso propósito desenvolver um algoritmo de silabificação automática eficaz, tendo em vista uma aplicação prática.

De um modo geral, o módulo de silabificação desempenha um papel fundamental em qualquer sistema TTS (Sproat, 1998) [12]: se, por um lado, a pronunciação de um fonema pode depender, em grande parte, da sua localização na estrutura da sílaba [13], muitas características prosódicas, nomeadamente a duração [14], são mais facilmente modeladas com base na sílaba e na sua organização interna.

A ideia de que a informação silábica pode contribuir, de algum modo, para o aumento do desempenho do módulo de conversão grafema-fone, constitui, no nosso caso, uma motivação adicional para investir num algoritmo de silabificação poderoso e eficaz.

O problema da silabificação automática tem conhecido várias abordagens (Marchand & Damper, 2007; Marchand *et alii*, 2007), que podem dividir-se, essencialmente, em: 1) soluções baseadas em regras (*rule-based*), que procuram implementar noções como o *Princípio de Sonoridade* ou a *Condição de Dissemelhança*, partindo das restrições fonotácticas da língua (Fisher, 1996; Weerasinghe *et alii*, 2005); e 2) paradigmas baseados em *corpora* (*data-driven*), em que as novas silabificações são inferidas à custa de grandes léxicos já com informação silábica (Marchand & Damper, 2007;

[12] Observem-se, a este propósito, as afirmações de Marchand *et alii* (2007): "However it is defined, and whatever the rights or wrongs of theorising about its linguistic status, syllable knowledge aids word modeling in automatic speech recognition and/or the unit selection and composition process of concatenative synthesis.".

[13] Um resumo esclarecedor desta perspectiva é-nos dado por Müller *et alii* (2000, p.225): "syllable structure represents valuable information for pronunciation systems". Os resultados de Marchand & Damper (2007) vão exactamente neste sentido, pois mostram que "including good quality information on syllabification of words can enhance the performance of a pronunciation system for use in text-to-speech and similar applications." (Marchand & Damper, 2007, p.22).

[14] Santen *et alii* (1997), por exemplo, mostram que a duração de um fone é afectada pelo seu posicionamento na sílaba.

Daelemans & Bosch, 1992; Tian, 2004; Kiraz & Möbius, 1998; Müller *et alii*, 2000; Müller, 2001; Bouma, 2002).

Em se tratando de uma investigação desenvolvida na área da Linguística e não sendo possível contar, para o PE, com um "*gold standard*" *corpus*, essencial para a aplicação de métodos de aprendizagem automática (bem como para a avaliação fidedigna de um algoritmo de silabificação), a escolha recaiu necessariamente sobre algoritmos de silabificação de base linguística, cujo objectivo principal consiste em atribuir, de forma categórica, não ambígua, silabificações a todos os segmentos das palavras, respeitando princípios silábicos consagrados (Blevins, 1995).

3.1.2 A sílaba no português europeu

O cumprimento dos objectivos supra-mencionados determina, antes de mais, uma breve apresentação do chamado modelo de "Ataque-Rima", com vista à descrição da estrutura interna da sílaba do PE. Em seguida, usando os instrumentos teóricos fornecidos por este modelo, será efectuada uma caracterização dos vários constituintes silábicos do PE: Ataque, Rima, Núcleo, Coda. Serão consideradas, não só as estruturas silábicas mais frequentes, mas também outras sequências alvo de dúvidas e ambiguidades. Finalmente, após a pequena revisão teórica com que se inicia a presente secção, procederemos à apresentação da última versão do algoritmo de silabificação do português (Mateus & d'Andrade, 2000), nos moldes em que ele foi delineado pelos seus proponentes.

3.1.2.1 Estrutura interna da sílaba no português europeu

Das várias propostas de estruturação silábica disponíveis [15], a concepção de sílaba como estrutura prosódica hierarquicamente organizada em constituintes silábicos corresponde ao modelo mais comumente aceite pela teoria fonológica [16]. Esta mesma concepção, representada no esquema arbóreo da figura 3.1, e designada por modelo "Ataque-Rima", tem-se revelado também bastante produtiva e funcional na descrição das estruturas silábicas do PE (Mateus, 1994; Mateus & d'Andrade, 2000; Freitas, 1997; Freitas & Santos, 2001; Veloso, 2003) [17].

De acordo com este modelo, a sílaba (σ) ramifica em Ataque (A) e Rima (R) e esta última ramifica em Núcleo (Nu) e Coda (Cd). Cada constituinte silábico está associado a uma, ou no máximo duas, posições do esqueleto, representadas por um X na árvore silábica [18]. Ao nível do esqueleto, associam-se os elementos fonológicos da língua. Este nível terminal de representação pode estar

[15] Para uma revisão de outros modelos de organização interna da sílaba, cf., por exemplo, Blevins (1995).

[16] Para uma análise dos argumentos que fundamentam a defesa desta concepção de sílaba, cf. Freitas (1997), Veloso (2003), Mateus (1994), Blevins (1995) e Mateus *et alii* (2005).

[17] A definição de sílaba, filiada no modelo "Ataque-Rima", proposta por Mateus & d'Andrade (2000, p.38), é a seguinte: "syllable is a multidimensional object with an internal structure that has a hierarchical organization where the onset and the rhyme constitute a branching structure".

[18] No quadro da teoria autossegmental, o nível do esqueleto é interpretado como uma sequência de unidades de tempo abstractas.

preenchido ou vazio (Ø). O Núcleo é, na maioria das vezes, preenchido por segmentos vocálicos ou, mais raramente, por consoantes, enquanto as posições de Ataque e Coda estão sempre associadas a segmentos consonânticos.

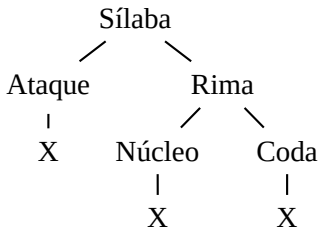


Figura 3.1: Representação esquemática da organização interna da sílaba.

Ataque

O constituinte silábico Ataque pode ser:

1. **não ramificado**, se estiver associado a uma posição do esqueleto. Neste caso, o Ataque pode ainda ser:
   - **simples**, por estar associado a material segmental;
   - **vazio**, se não tiver qualquer segmento associado.
2. **ramificado ou complexo**, no caso de estar associado a mais do que uma posição do esqueleto.

Segundo Vigário & Falé (1993), a tendência é para que o Ataque seja preenchido por uma única consoante, tanto nas palavras monossilábicas, como polissilábicas (e.g. "pé" [ˈpɛ], "pato" [ˈpa.tu]). Qualquer consoante do PE, com qualquer ponto de articulação, pode ocupar um Ataque não ramificado simples.

Os Ataques não ramificados segmentalmente vazios, à semelhança do que acontece com os simples, podem ocorrer no início ou no interior da palavra (e.g. "uva" [ˈu.vɐ], "pia" [ˈpi.ɐ]).

O português apresenta ainda Ataques ramificados, constituídos por duas consoantes. Neste caso, as únicas sequências consonânticas admitidas são as combinações de oclusiva+líquida (e.g. "prato" [ˈpratu]) e fricativa+líquida (e.g. "flor" [ˈflor]), por respeitarem o *Princípio de Sonoridade* e a *Condição de Dissemelhança*, ainda que o primeiro grupo seja muito mais frequente do que o segundo (Vigário & Falé, 1993). As diferenças de frequência entre os grupos consonânticos que podem preencher o Ataque [19] e a preferência do PE por Ataques ramificados do tipo oclusiva+líquida encontram justificação na diferença de sonoridade entre elementos adjacentes: na escala de sonoridade, uma oclusiva está mais distante de uma líquida do que de uma fricativa (vd. 3.1.1).

[19] Segundo Vigário & Falé (1993), 94% dos ataques ramificados correspondem a sequências de oclusiva+líquida e apenas 7% são constituídos por grupos de fricativa+líquida.

Já outros grupos consonânticos, que combinam sequências de obstruinte+obstruinte (e.g. "raptor" [rapˈtoɾ], "afta" [ˈaftɐ]) ou obstruinte+nasal (e.g. "pneu" [ˈpnew], "admirar" [ɐdmiˈraɾ]), violam claramente os referidos princípios de boa formação silábica e são tidos, na literatura sobre o tema, como problemáticos para efeitos da determinação das fronteiras silábicas.

De acordo com a interpretação fonológica de alguns autores (Mateus, 1994; Mateus & d'Andrade, 1998, 2000; Câmara, 1973, 1971) [20], estes grupos consonânticos não se constituem como sequências tautossilábicas, co-ocorrentes em Ataque Ramificado, mas devem antes aceitar-se como Ataques simples de duas sílabas distintas e adjacentes (sequências heterossilábicas). Entre as duas consoantes postula-se a existência de um *Núcleo Vazio*, representado por ∅ [21].

A proposta de divisão heterossilábica das sequências em análise encontra-se consignada na *Convenção de Criação de Núcleos vazios* (vd. 3.1.2.2) do algoritmo de silabificação do português que nos serve de base (Mateus & d'Andrade, 2000), o que justifica um olhar atento sobre os principais argumentos normalmente aventados em favor desta análise. Destacamos os seguintes:

- Inserção de uma vogal epentética entre as duas consoantes do grupo (Mateus & d'Andrade, 1998, 2000; Câmara, 1973, 1971; Barbosa, 1965): no caso do português do Brasil, a vogal em causa é normalmente um [i], mas, também em registos coloquiais do português europeu, é possível detectar um [ɨ] [22];

[20] Em trabalhos anteriores, Mateus (1993) admite a possibilidade de tais sequências poderem realizar, em português, Ataques ramificados, a par das combinações de obstruinte+líquida, enfatizando o facto de violarem o *Princípio de Sonoridade*. Na *Introdução à Linguística Geral e Portuguesa* (Faria *et alii*, 1996), num capítulo dedicado à Fonologia, esta investigadora, a propósito de sequências de consoantes pouco habituais em início de palavra (e.g. "gnomo", "psicologia", etc.), refere que "são, indubitavelmente, ataque de sílaba" (Faria *et alii*, 1996, p.178), e estende a explicação às consoantes no interior das palavras. A explicação de Cunha & Cintra (1997), na *Nova Gramática do Português Contemporâneo*, vai no mesmo sentido, ao considerar os encontros consonantais em questão "naturalmente inseparáveis" (Cunha & Cintra, 1997, p.51), quando iniciais. Em posição medial, "em pronúncia tensa, podem ser articulados numa só sílaba, ou em sílabas distintas" (Cunha & Cintra, 1997, p.52).

[21] Veloso (2003, 2006) dá conta de algumas semelhanças importantes entre as sequências interpretadas, nas descrições fonológicas do português, como heterossilábicas e as sequências de obstruinte+lateral, normalmente analisadas como tautossilábicas: 1) evolução histórica: ao contrário das combinações obstruinte+vibrante, a sequência obstruinte+lateral em caso algum foi mantida, na evolução espontânea do latim para o português; 2) a introdução de uma vogal epentética entre os dois elementos da sequência, bastante frequente no registo coloquial dos adultos, nas produções das crianças (cf. Freitas, 1997) e nas criações poéticas versificadas dos chamados "poetas populares"; 3) divisões silábicas explícitas: de acordo com o estudo de Veloso (2003), a divisão silábica da estrutura obstruinte+lateral como tautossilábica só emerge após a aprendizagem das regras de translineação gráfica, sendo que antes essas sequências são preferencialmente analisadas pelas crianças como heterossilábicas. Com base nestes argumentos, o autor afirma (Veloso, 2006, p.151-152) "parece-nos razoavelmente justificado que se encontre alguma forma de se admitir que no conhecimento fonológico de um número significativo de falantes nativos da língua exista uma diferença entre as sequências Obstruinte+Vibrante e as sequências Obstruinte+Lateral no que diz respeito à sua divisão silábica, sendo as primeiras preferencialmente consideradas como *genuinamente tautossilábicas* e as segundas como *genuinamente heterossilábicas*.".

[22] Como facilmente se pode depreender das palavras do fonólogo estruturalista Câmara (1973, p.46-47) - "Restam dois problemas muito importantes para a fixação das estruturas silábicas portuguesas. O primeiro se refere aos vocábulos, diacronicamente de origem "erudita" (isto é, introduzidos através da língua escrita, a partir do séc. XV, como empréstimos ao latim clássico). São os de tipo - *compacto, apto, ritmo, afta*, e assim por diante. (...) Na realidade há entre uma e outra consoante a intercalação de uma vogal, que não parece poder ser fonemicamente desprezada, apesar da tendência a reduzir a sua emissão no registro formal da língua culta. Ela é /i/ na área do Rio de Janeiro e /e/ ([a] neutro em Portugal).") - para este autor a vogal interconsonântica tem estatuto fonémico. Segundo Mateus & d'Andrade (1998, 2000), trata-se da realização fonética de um Núcleo fonologicamente vazio.

- Dificuldades/ hesitações/ dúvidas dos falantes na determinação das fronteiras silábicas, quando estão em causa grupos consonânticos problemáticos (Mateus & d'Andrade, 1998, 2000), com repercussões directas na tarefa de translineação gráfica (d'Andrade & Viana, 1993a) [23];

- Dados relativos à aquisição das estruturas silábicas do português europeu (Freitas, 1997) revelam que os grupos consonânticos formados por obstruinte+líquida têm natureza distinta da das sequências não legitimadas, apesar da pouca representatividade destas últimas no *corpus* que serviu de base à investigação: as crianças produzem palavras com Ataques ramificados desde cedo, ainda que recorrendo a algumas *estratégias de reconstrução* silábica, sobretudo o apagamento da segunda consoante. O mesmo não se verifica com os outros grupos consonânticos problemáticos, cuja utilização é despoletada tardiamente no processo de aquisição linguística da criança, sem que sejam atestadas produções com supressão do segundo elemento do grupo. Tal comportamento tem sido interpretado como uma prova da coesão das unidades tautossilábicas (Mateus & d'Andrade, 1998, 2000) [24];

- Palavras com sequências CC não legítimas são pouco frequentes em português (Barbosa, 1965, 1994a; Câmara, 1973, 1971; Vigário & Falé, 1993; Freitas, 1997; Veloso, 2003), quando comparadas com os grupos consonânticos constituídos por obstruinte+líquida, e foram introduzidas tardiamente na língua, por via erudita, a partir de outras línguas, sobretudo do latim (Veloso, 2003) [25]. Os falantes tendem a dividir estas combinações em duas sílabas sucessivas, lançando mão de estratégias como a epêntese interconsonântica. Dados diacrónicos relativos à evolução da língua (e.g. semivocalização do primeiro elemento consonântico: lat. *nocte*> port. *noite*) ilustram esta tendência da língua em libertar-se dos encontros consonantais que incorrem em violações dos princípios expostos (Vigário & Falé, 1993).

Rima

A Rima é um constituinte silábico não terminal, que integra obrigatoriamente um Núcleo, e, opcionalmente, à direita deste último, uma Coda. A postulação de um constituinte hierarquicamente superior na estrutura da sílaba visa captar a forte relação entre o Núcleo e a Coda, que não se verifica entre o Ataque e o resto da sílaba. Esta interacção entre os segmentos ocorrentes no Núcleo e os segmentos presentes na Coda pode ser ilustrada através de dois processos fonológicos (Mateus *et alii*, 2005): i) o bloqueio imposto pela lateral em Coda à elevação das vogais átonas em português; ii) o processo

[23] O processo de translineação é de base silábica (a divisão gráfica da palavra corresponde, normalmente, às fronteiras interssilábicas (cf. d'Andrade & Viana, 1993a).

[24] Outro aspecto invocado por Mateus & d'Andrade (1998, 2000) é a inserção, na linguagem infantil (e verificada também em erros de ortografia), de uma vogal entre as duas consoantes que compõem a sequência não admitida (e.g. "pneu" *[pɨnew]). Refira-se, contudo, que a epêntese da vogal é uma estratégia recorrente também no caso de estruturas consonânticas legítimas, num estádio de aquisição intermédio, confirmando o predomínio da estrutura CV (Freitas, 1997, 2003, 2002).

[25] Cf. as afirmações de Câmara (1973), nota 22.

diacrónico que envolve as vogais nasais do português. Estas são o resultado da transferência da nasalidade da Coda latina para o Núcleo. Relativamente ao comportamento independente do Ataque e da Rima, Mateus (1994, p.289-90) afirma "os segmentos incluídos no ataque são irrelevantes relativamente ao número máximo de segmentos permitidos pela Rima". No mesmo sentido vão as afirmações de Veloso (2003), que chama a atenção para as restrições entre o Núcleo e a Coda, como a que impede a ocorrência de um ditongo em Núcleo associado a uma lateral ou vibrante em Coda, limitação esta inexistente entre o Ataque e o Núcleo, onde todas as combinações são permitidas. Os estudos desenvolvidos na área da Psicolinguística, citados pelo mesmo autor (Veloso, 2003), são, contudo, contraditórios no que toca a fornecer evidências para a divisão da sílaba em Rima e Ataque.

Assim, a Rima pode assumir as seguintes tipologias:

1. **não ramificada**, se estiver associada apenas a um Núcleo;
2. **ramificada**, quando composta por Núcleo + Coda.

A estrutura destes dois constituintes silábicos (Núcleo e Coda) é apresentada e discutida nas secções seguintes.

Núcleo

O Núcleo goza de um estatuto especial, visto ser de preenchimento obrigatório e definir, por si só, a identidade da sílaba. Por outras palavras, o Núcleo implica a existência da Rima e, por conseguinte, da própria sílaba.

Este constituinte silábico pode estar associado a um mínimo de uma e a um máximo de duas posições no nível do esqueleto, i.e., pode apresentar um formato:

1. **não ramificado**, quando preenchido por uma vogal;
2. **ramificado**, quando associado a uma vogal seguida de uma glide, sequência tradicionalmente designada por ditongo decrescente (VG). Neste caso, a única consoante que pode ocorrer em Coda é o /S/ (e.g. "mais", "fausto").

Tal como na grande maioria das línguas, também no português, o Núcleo é preenchido exclusivamente por segmentos [-consonânticos]. Isto significa que não há, nesta língua, consoantes silábicas [26].

Qualquer vogal (oral ou nasal) pode assumir este papel, configurando, desta forma, um Núcleo não ramificado.

[26]Delgado-Martins (1994) sugere a possibilidade de conferir o estatuto silábico a algumas consoantes como o [ʃ], em palavras como [ʃ'plisitu], à semelhança do que acontece já em línguas como o francês e o inglês.

No que diz respeito aos Núcleos ramificados, os dois elementos do ditongo decrescente pertencem ao Núcleo [27]. De acordo com a interpretação de Mateus & d'Andrade (2000), a semivogal fonética é uma vogal [+alta] no nível fonológico [28]: "if a high vowel is marked and if it is preceded by another vowel, it becomes a glide at the phonetic level and it is integrated in the syllable nucleus with the preceding vowel" (Mateus & d'Andrade, 2000, p.48).

Um dos argumentos a favor desta análise são os ditongos nasais, já que ambos os elementos, vogal e glide, sofrem os efeitos da nasalidade e devem, por isso, ser integrados no mesmo Núcleo.

Conforme salientado por Mateus & d'Andrade (2000), os ditongos em geral, e os ditongos nasais em particular (sobretudo em final de palavra), são estruturas muito frequentes em português [29].

Contrariamente aos ditongos decrescentes, cujos elementos estão sempre representados no Núcleo, as sequências de glide+vogal (GV), tradicionalmente designadas de ditongos crescentes, afiguram-se mais problemáticas no que concerne à estipulação de fronteiras silábicas.

A explicação de Mateus (1993) e Mateus & d'Andrade (1998, 2000) vai no sentido de considerar que, neste último caso, a semivogal tem na base uma vogal fonológica e existe uma fronteira silábica entre os dois elementos do ditongo (V.V). Em estrutura de superfície, a primeira vogal poderá ocupar sozinha um Núcleo de sílaba (V.V) ou agregar-se à vogal seguinte, constituindo assim um ditongo crescente (GV).

Esta análise é suportada essencialmente por dois argumentos: 1) por um lado, o acento pode recair sobre a vogal alta em palavras morfologicamente aparentadas (e.g. "suo" [ˈsuw]/ "suar" [suˈar]) [30]; 2) por outro, enquanto nos ditongos decrescentes nasais, os dois elementos da sequência são nasalizados, nos ditongos crescentes, a semivogal não é afectada pela nasalidade [31].

Freitas & Santos (2001), no sentido de demonstrar a natureza distinta dos ditongos decrescentes e crescentes, salientam ainda as diferenças na translineação gráfica destas duas estruturas. Na escrita, não é permitido separar os grafemas que correspondem a um ditongo decrescente, já que ambos pertencem à mesma sílaba (e.g. pau-ta/ *pa-uta). No caso dos ditongos crescentes, esta partição é possível (e.g. pi-ano), o que reflecte, de alguma forma, uma representação de base em que existe,

[27] De acordo com a perspectiva estruturalista, a glide do ditongo pertence à Coda e não ao Núcleo (Barbosa, 1994a; Barroso, 1999; Câmara, 1971).

[28] A glide fonética é considerada um vogal no nível fonológico quando: a) é antecedida de outra vogal e b) é marcada lexicalmente como não-acentuável. Por outro lado, a vogal e a glide não contrastam fonologicamente, uma vez que não existem, em português, pares mínimos que suportem a oposição ditongo decrescente e sequência de duas vogais ([ˈpaj] vs.*[paˈi]). Foneticamente, as glides têm características idênticas às suas correspondentes vocálicas, mas uma duração inferior. Do ponto de vista silábico, não podem constituir isoladas Núcleo de sílaba (cf. Mateus *et alii*, 2003).

[29] As questões relativas ao número de posições métricas ocupadas pelos ditongos decrescentes ultrapassam o âmbito deste trabalho. Para mais informações sobre este assunto, cf., por exemplo, Bisol (1989, 1994) e Mateus & d'Andrade (2000).

[30] O exemplo foi retirado de Mateus & d'Andrade (2000, p.49)

[31] Também a classificação de palavras do tipo "farmácia" como proparoxítonas indicia que, para efeitos de atribuição do acento, o segmento que precede a última vogal é entendido como silábico e, consequentemente, são contabilizadas duas sílabas depois do acento. Os resultados acústicos de Drenska (1986) e Silva (1987) mostram, no entanto, que, no contexto pós-tónico, a alternância GV/VV não se verifica e a semivocalização parece ter carácter obrigatório. O comportamento das crianças, apesar de pouco esclarecedor quanto a esta estrutura, sugere também que as crianças "a processam preferencialmente como estando associada a uma semivogal" (Freitas, 1997, p.330).

entre as duas vogais, uma fronteira interssilábica.

Em consequência de factores como o ritmo e a velocidade de elocução, entre outros, este tipo de encontros vocálicos pode ser pronunciado em ditongo ou em hiato ("piada" [piˈadɐ]/ [ˈpjadɐ]) [32], o que não se verifica em relação aos ditongos decrescentes.

As dúvidas residem precisamente no lugar ocupado pela semivogal - que resulta da perda do traço silábico e semivocalização do primeiro elemento do ditongo com consequente apagamento da fronteira inter-silábica (V.V -> GV) - na estrutura da sílaba.

Alguns autores (Mateus, 1993; Mateus & d'Andrade, 2000; d'Andrade & Viana, 1993b) sugerem que o G seja dominado por um Ataque ramificado e não pela Rima, uma vez que:

- em palavras como "reauscultar" [rjawʃkulˈtaɾ], a integração da glide no Núcleo seguinte aumentaria para quatro o número de segmentos permitidos na Rima, quando está estabelecido que, no português, a Rima deve conter, no máximo, três elementos. A solução, para estes autores, passa então por integrar a glide no Ataque, ao invés de aumentar o número de segmentos previstos na Rima, uma vez que "the number of the elements in the onset is irrelevant for the maximum number of elements in the rhyme" (Mateus & d'Andrade, 2000, p.51);
- contrariamente ao que acontece nos ditongos decrescentes, quando a glide ocorre antes de uma vogal ou ditongo nasal, esta não sofre o efeito da nasalidade (e.g. "leão" [ˈljɐ̃w̃]). Assume-se, portanto, que o autossegmento nasal projecta a nasalidade no nó Núcleo, não atingindo a glide da esquerda por esta não se encontrar neste constituinte.

Segundo (Freitas, 1997), esta análise tem a desvantagem de gerar ataques triposicionais em palavras como "criado" ([ˈkɾjadu]) e padrões silábicos excepcionais do tipo CG, que não existem em qualquer outro contexto.

Por outro lado, recordando os dados de Freitas (1997), verifica-se que esta interpretação não encontra eco no comportamento das crianças, na medida em que, durante o processo de aquisição das estruturas silábicas, estas associam a glide do ditongo ao Núcleo da vogal adjacente à direita, ao invés de a processarem como parte de um Ataque ramificado.

Em contextos restritos, que envolvem a semivogal arredondada [w], após oclusiva velar [k] e [g] (e.g. "quatro" [ˈkwatru] e "água" [ˈagwɐ]), a referida alternância fonética GV/ VV não se verifica e a produção da semivogal é obrigatória. Várias hipóteses têm sido formuladas sobre a natureza fonológica e papel silábico desta semivogal. Freitas (1997, 2000) sumariza as diferentes propostas, disseminadas na literatura do português:

1. A estrutura GV conforma aquilo que tradicionalmente se designa como ditongo crescente e,

[32] Freitas (1997, p.307) considera mesmo que "a produção preferencial da estrutura-alvo V.V, no Português Europeu coloquial, é GV". Segundo Freitas & Santos (2001, p.44, nota 19), "estas sequências são pouco naturais nas línguas e tendem a alterar-se, quer por redução (...), quer por alteração das propriedades segmentais (...), quer ainda por ditongação".

como tal, a semivogal integra o Núcleo ramificado [33];

2. A semivogal [w] é um segmento autónomo, de natureza consonântica, sendo representada no domínio de um Ataque ramificado. Esta análise decorre da interpretação de Barbosa (1965) quanto ao estatuto das semivogais, tidas pelo autor como consoantes;

3. A estrutura CG é considerada uma unidade monofonemática, i.e., uma oclusiva velar labializada ([k^w], [g^w]) (Bisol, 2001) e está associada a um Ataque não ramificado.

Freitas (2000) sublinha que as três hipóteses implicam análises excepcionais no contexto do sistema fonológico do PE: a optar pela primeira hipótese, este seria o único caso de ditongo fonológico crescente no português, limitando-se a um contexto muito particular (quando o Ataque é preenchido por uma oclusiva velar); por outro lado, a atribuição do estatuto de consoante à semivogal traduzir-se-ia num caso ímpar de Ataque ramificado com a estrutura CG no PE; por fim, a consideração da labialização da consoante velar acarretaria o aumento, pouco económico, do inventário segmental fonológico. A nosso ver esta última hipótese explicativa carece igualmente de um estudo experimental sobre o modo como se processa a labialização da oclusiva.

Freitas (1997, 2000), no estudo de aquisição já referido, apoia-se nos comportamentos verbais das crianças para concluir que as estruturas em causa são processadas como consoantes velares labializadas, ocupando, como tal, a posição de Ataque não ramificado.

Coda

Como noutras línguas naturais, também em português o constituinte silábico Coda está sujeito a restrições segmentais fortes [34]: apenas três consoantes [35] - o /l/, o /S/ [36] e o /ɾ/ - podem figurar nesta posição (Freitas, 1997; Freitas & Santos, 2001; Mateus, 1994; Mateus & d'Andrade, 1998, 2000; Bar-

[33] Cf. as afirmações Cunha & Cintra (1997, p. 48) acerca dos ditongos crescentes: "apenas apresentam estabilidade os que têm a semivogal [w] precedida de [k] (grafado q) ou de [g], como em quase e igual.". Também em Mateus (1989) se refere que as sequências [kw] e [gw] são "casos particulares que praticamente não põem dúvidas quanto à natureza semivocálica do segundo elemento" (Mateus, 1989, p.352).

[34] Na maioria das línguas, o número de consoantes em Coda é inferior ao número de consoantes admitidas em Ataque (Blevins, 1995; Prince & Smolensky, 2004). Quando comparadas com as línguas germânicas, as línguas românicas parecem ser ainda mais restritivas no tocante à admissão de consoantes em Coda (Mateescu, 2003). O português destaca-se como uma das línguas românicas mais limitativas a este respeito (Veloso, 2008). Uma série de processos diacrónicos, como o apagamento ou semivocalização das Codas latinas [-sonoras] e alterações fonológicas (e.g. transferência da Coda etimológica para o Ataque da sílaba seguinte em empréstimos do inglês) ilustram esta tendência do PE (e ainda em maior grau do PB) para evitar o preenchimento do constituinte silábico Coda. A este fenómeno Veloso (2008) chama "*coda-avoiding*".

[35] Em consequência do apagamento, a nível fonético, das vogais átonas, nomeadamente o [ɨ] final, qualquer consoante pode ocorrer em Coda (e.g. "bate" [ˈbat], "pode" [ˈpɔd]), mas, na realidade, ao nível subjacente, a vogal está presente, mesmo que raramente se realize do ponto de vista fonético. Sequências do tipo "mexe bem" ([ˈmɛʃ ˈbẽj̃]) são a prova disso mesmo: se a consoante [ʃ] estivesse mesmo em final absoluto de palavra, estaria sujeita à regra de assimilação do traço de vozeamento da consoante seguinte, o que na realidade não acontece, supostamente devido à presença de uma vogal sujacente (Mateus *et alii*, 2003).

[36] O /S/ realiza-se foneticamente como [ʃ] ou como [ʒ], em função do vozeamento da consoante seguinte, no caso de existir alguma.

roso, 1999) [37]. De entre estas, a consoante /S/ é a única capaz de suceder a um Núcleo ramificado (e.g. "claustro"). O PE admite, portanto, Rimas triposicionais (Mateus & d'Andrade, 2000).

Assume-se, assim, que, em português, o constituinte silábico Coda não ramifica [38], dominando um elenco de consoantes reduzido.

Palavras como "perspectiva" e "solstício", que apresentam duas consoantes à direita da vogal da primeira sílaba, são interpretados como "casos raros no sistema, sendo considerados excepções ao comportamento regular, i.e., não há Codas ramificadas no Português" (Freitas & Santos, 2001, p.47).

Esta análise é partilhada por Mateus (1993) [39] , que, no entanto, em trabalhos posteriores (Mateus & d'Andrade, 2000), subscreve uma outra interpretação fonológica: a única maneira de resolver o problema da violação do *Princípio de Sonoridade* e *Condição de Dissemelhança*, decorrente da análise destas sequências como Codas ramificadas, consiste em aceitar uma fronteira silábica entre as duas consoantes, sendo que a primeira consoante do grupo é o Ataque de uma sílaba de *Núcleo vazio*.

A tradicional simetria entre as três consoantes em final de sílaba advogada por Mateus & d'Andrade (2000) é questionada por Freitas (1997, 1998) e Correia (2003, 2004a,b), com base nas produções silábicas de crianças portuguesas. Restrições fonotácticas e distribucionais entre fricativas e líquidas (e.g. depois de ditongo ou Núcleo nasal apenas a consoante obstruinte é admitida; no processo de formação do plural, a lateral é substituída por uma semivogal) indiciam já a possibilidade destas desempenharem papéis silábicos distintos (Freitas, 1998; Correia, 2003, 2004a). Esta hipótese é avaliada à luz dos dados da aquisição: apesar de as crianças possuirem já, no seu inventário segmental, as consoantes líquidas e se encontrar disponível, num dado momento da aquisição, o constituinte silábico Coda, estas começam a produzir líquidas muito depois do início da produção das fricativas. Este comportamento revela que as consoantes fricativas devem ser representadas em Coda, mas as líquidas não estão a ser processadas como Codas e devem antes ser integradas num Núcleo ramificado (Freitas, 1997, 1998; Correia, 2003, 2004a,b).

3.1.2.2 O algoritmo de silabificação do português europeu segundo Mateus & d'Andrade (2000)

O processo de silabificação consiste na aplicação de uma série de mecanismos ("convenções") que procuram associar todos os segmentos das representações lexicais a um determinado constituinte si-

[37] Os argumentos que fundamentam esta posição podem ser encontrados em Mateus (1994), Mateus & d'Andrade (2000) e Freitas (1997).

[38] Do ponto de vista estatístico, os grupos consonânticos são, por si, só desfavorecidos, mas, a existirem, são mais frequentes em Ataque do que em Coda (Rousset, 2004; Zerling, 2000). Foneticamente, este fenómeno pode ser explicado pelo próprio ciclo mandibular (Redford, 1999): a fase de abaixamento é mais longa do que a de subida, deixando mais espaço para a articulação consonântica em início de sílaba.

[39] Mateus (1993) admite a ocorrência de Codas compostas em palavras como "abstrair" e "perspectiva". Segundo a fonóloga, esta estrutura silábica resultaria, caracteristicamente, de combinações de C+/S/, ocorrendo, sobretudo em sílabas pré-acentuadas, com Núcleos não-ramificados.

lábico. O algoritmo de silabificação de base proposto por Mateus & d'Andrade (2000) [40] para o PE foi desenvolvido de acordo com os princípios teóricos já referidos, tendo em atenção todas as restrições enunciadas na secção anterior. De acordo com o enquadramento teórico adoptado (modelo "Ataque-Rima"), a atribuição de papéis silábicos (nível da Rima) aos segmentos (nível dos segmentos) é mediado pelo nível do esqueleto (X). Na abordagem escolhida, tradicionalmente chamada de "all nuclei first approach" (cf. Goldsmith, 1990), começa-se pela construção das Rimas, de acordo com as restrições da língua. Seguem-se as restantes regras que compõem o algoritmo, apresentado em pormenor já a seguir:

1. **Convenção de Associação de Núcleos**

   - Associar todos os X [-cons], simultaneamente não marcados lexicalmente e não precedidos por um [-cons], a um Núcleo.
   - Juntar os restantes X [-cons] ao Núcleo que se encontra à sua esquerda. Com a criação do Núcleo, constrói-se automaticamente uma Rima.

2. **Convenção de Associação de Ataques**

   - Associar a um Ataque todos os X [+cons] que precedam imediatamente o Núcleo;
   - Juntar no mesmo Ataque o X [+cons] anterior, se o grupo consonântico respeitar o Princípio de Sonoridade e a Condição de Dissemelhança. Isto significa que uma sequência de duas consoantes só se insere no mesmo Ataque se estiver de acordo com as restrições da língua.

3. **Convenção de Criação de Núcleos Vazios**

   - Criar um Núcleo à esquerda do Ataque, com a correspondente posição esqueletal, se este for precedido por X [+cons] sem posição silábica atribuída especificada para o vozeamento. Nos restantes casos, criar um Núcleo à esquerda dessa mesma posição não associada.

4. **Convenção de Criação de Ataques Vazios**

   - Criar um Ataque à esquerda da Rima, com a correspondente posição na fiada do esqueleto, no caso desta não ser precedida por um Ataque. Assume-se assim que a sílaba no PE é obrigatoriamente constituída por um Ataque e por uma Rima, ainda que um destes constituintes (mas não ambos) possa estar vazio (cf. Mateus & d'Andrade, 2000, p.58).

[40] A formalização do processo de silabificação de base pode ser encontrada em Mateus (1994) e Mateus & d'Andrade (1998, 2000). A nossa apresentação do algoritmo segue de perto a proposta publicada em Mateus & d'Andrade (2000) e traduzida para o português por Veloso (2003), que encerra algumas diferenças relativamente a versões anteriores.

5. **Convenção de Criação de Associação de Codas**

   - Associar os X [-cons] ainda sem posição silábica à Coda da Rima precedente.

A aplicação sequencial de regras de silabificação e a criação de sílabas sucessivas com a estrutura CV, prevalentes no português, permitem concluir que "base syllables in Portuguese are CV syllables, despite the apparent violations of European Portuguese at the phonetic level. Consequently, the underlying syllables differ crucially from those on the phonetic level, namely for EP, as the number of CV syllables is obviously higher underlyingly." (Mateus & d'Andrade, 2000, p.64).

3.1.3 Silabificação automática em português europeu

Ao contrário de outras línguas, a investigação em torno da silabificação automática do PE não tem despoletado grande interesse. Deixamos aqui alguns exemplos de trabalhos desenvolvidos nesta área, para a língua portuguesa. Ao que foi possível apurar, a grande maioria das ferramentas não é do domínio público e não são conhecidos os resultados da sua avaliação, o que, por um lado, invalida qualquer tipo de estudo comparativo e, por outro, justifica o investimento no desenvolvimento de um módulo de silabificação automática.

No âmbito da sua tese de Doutoramento, dedicada ao desenvolvimento de um sistema completo de síntese de fala a partir de texto (DIXI), Oliveira (1996) faz uso de um conjunto de onze regras para a silabificação automática das palavras. Este procedimento, necessário ao módulo de análise prosódica, actua directamente sobre o nível ortográfico marcado com o acento lexical. Contrariando a tendência actual - que reconhece a importância fundamental da sílaba, quer na área da síntese, quer no campo dos sistemas automáticos de reconhecimento de fala (vd. secção 3.1.1) - o investigador acredita que "a colocação exacta da fronteira silábica não é particularmente importante" (Oliveira, 1996, p.86), considerando o método de síntese que utiliza, o que talvez justifique a escassez de informação sobre o processo de implementação e/ou avaliação do algoritmo de divisão silábica.

Este mesmo sistema de conversão grafema-fone (DIXI), que inclui o módulo de silabificação, foi, posteriormente utilizado num estudo, enquadrado no projecto europeu ONOMASTICA, que visava uma análise comparativa do desempenho de um sistema automático de transcrição fonética aplicado a nomes próprios e léxico comum (Viana *et alii*, 1996). A análise da distribuição dos padrões silábicos nos dois *corpora* ditou o esclarecimento de alguns dos critérios de silabificação, na sua generalidade concordantes com as observações de Vigário & Falé (1993) e d'Andrade & Viana (1993a). Ficamos, assim, a saber, por exemplo, que não são admitidas obstruintes em Coda e os ditongos crescentes são contemplados. Quanto ao desempenho em si, foram testadas duas metodologias: uma baseada em regras e outra em redes neuronais. O desempenho das regras na silabificação de nomes próprios e léxico comum é similar (0.3%); o desempenho das redes foi inferior ao das regras em 0.7% (nomes próprios) e 0.5% (léxico comum).

Já Teixeira (2004) considera que a sílaba é uma unidade relevante na determinação de parâmetros prosódicos, pelo que apresenta um algoritmo de divisão automática do texto em sílabas (Gouveia *et alii*, 2000), enquanto etapa preliminar ao desenvolvimento de um sistema de análise prosódica para o PE (Teixeira, 2004). As regras de separação silábica podem ser aplicadas directamente sobre o texto escrito ou actuar sobre uma sequência de fones, resultado da leitura desse mesmo texto por um locutor. Em ambos os casos, os critérios de silabificação baseiam-se no pressuposto de que, em PE, apenas alguns tipos silábicos são admitidos. Na divisão silábica do texto escrito, as ambiguidades decorrentes da silabificação de consoantes intervocálicas são resolvidas graças a um conjunto de regras adicionais. A separação silábica do texto falado acarreta novas dificuldades relacionadas com a elisão de vogais, que determinam não só a assunção de novos formatos silábicos, como a introdução de novas regras a aplicar em situações específicas. De acordo com os resultados da avaliação, o desempenho do algoritmo de silabificação, na versão aplicada aos fones, é ligeiramente inferior ao observado para o algoritmo aplicado ao texto escrito (0.06% e 0.089%, respectivamente), diferença esta que parece encontrar justificação na dificuldade do algoritmo (na versão aplicada ao texto falado) em lidar com a referida queda de vogais.

Também o Projecto Natura (Almeida *et alii*, 2003) [41] tem à disposição da comunidade científica um algoritmo de silabificação automática, que não se faz, infelizmente, acompanhar de informação detalhada sobre a sua implementação e/ou avaliação.

Uma outra ferramenta electrónica - desenvolvida no âmbito do projecto "Padrões de Frequência na Fonologia do Português. Investigação e Aplicações" - que dispõe da funcionalidade de dividir o texto escrito em sílabas é o *Frep* [42]. Esta aplicação visa, essencialmente, a obtenção de informação sobre a frequência de ocorrência de unidades fonológicas, permitindo, entre muitas outras potencialidades, identificar e contar os tipos silábicos, em função 1) da posição na palavra, 2) do acento, 3) da posição na palavra e do acento (Vigário *et alii*, 2005, 2006). O processo de delimitação das fronteiras silábicas segue de perto as propostas e análises de Mateus & d'Andrade (2000) e Vigário & Falé (1993). Assim, em linha com estes trabalhos: a) as glides entre duas vogais (e.g. "areia") são consideradas ambissilábicas; b) entre duas consoantes adjacentes que incorrem em violações dos princípios de silabificação foi introduzida uma posição *V-slot*, i.e., uma posição vocálica vazia susceptível de ser preenchida por vogais epentéticas (e.g. "obter" > o.bV.ter); c) as sequências de grafemas <gu> e <qu> são analisadas como oclusivas (labializadas, como em "quando", ou não-labializadas, como em "líquido"); d) os ditongos decrescentes são obrigatórios a par dos ditongos crescentes pós-tónicos (e.g. "família"), já que, em ambos os casos, a semivocalização é obrigatória. Sabe-se ainda que a fiabilidade do sistema no tocante à divisão silábica é de 99.709%, mas não são conhecidos quaisquer pormenores quanto aos critérios de avaliação.

[41] O projecto Natura (cf. http://natura.di.uminho.pt) desenvolve trabalho na área do Processamento de Linguagem Natural (PLN), com ênfase no português.

[42] Uma demonstração desta ferramenta encontra-se disponível online em www.fl.ul.pt/LaboratorioFonetica/FreP. Na mesma página, é ainda possível encontrar informações várias sobre o sistema, inclusive um manual do programa. Infelizmente, e apesar de concebida como ferramenta de domínio público para fins de investigação, todos os contactos no sentido da sua obtenção se revelaram infrutíferos.

Resta-nos referir o trabalho de Meinedo (Meinedo *et alii*, 1999; Meinedo, 2000; Meinedo & Neto, 2000), que desenvolve diferentes métodos automáticos de segmentação silábica, tendo em vista um aumento do desempenho global dos sistemas de reconhecimento de fala para o PE.

3.1.4 Sistemas de silabificação automática desenvolvidos

Nas secções seguintes, descreve-se o essencial da implementação dos dois algoritmos de silabificação desenvolvidos: o primeiro com base em transdutores de estados finitos e o segundo a partir da aplicação automática de um conjunto de regras de silabificação originalmente propostas por Mateus & d'Andrade (2000).

Apresentam-se ainda os critérios que presidiram à avaliação dos vários sistemas desenvolvidos, bem como os *corpora* utilizados no desenvolvimento e teste dos algoritmos.

Finalmente, sintetizam-se os resultados relativos aos níveis de desempenho dos diferentes métodos.

3.1.4.1 Silabificação automática baseada em transdutores de estados finitos

O módulo de silabificação automática que agora se apresenta faz parte de um conjunto de procedimentos que visam servir de suporte ao desenvolvimento de um sistema de conversão grafema-fone, descrito em pormenor na secção 3.2.3.1.

A divisão silábica constitui-se assim, a par do módulo de atribuição do acento lexical, como um primeiro passo para a conversão automática do texto em fones, realizada através da aplicação de regras definidas manualmente, implementadas através de *Finite State Transducers* (FSTs) e corrigidas via *Transformation-Based Learning* (TBL) (Brill, 1995).

Neste módulo é já operada uma primeira "normalização". Assim, os dígrafos que correspondem a um único fone são desde logo transformados no símbolo SAMPA [43] correspondente: por exemplo, o <ss> passa a [S] e o <rr> é convertido em [R].

Os grafemas <m> e <n>, em final de sílaba (e.g. "caNtar"), são convertidos num elemento abstracto, representado por /N/, que obrigatoriamente nasaliza a vogal precedente, desaparecendo posteriormente.

Assume-se também, seguindo a abordagem de d'Andrade & Viana (1993b) e Bisol (2001), que as sequências de grafemas <qu> e <gu> representam uma única consoante - labializada ("quatro") ou não-labializada (e.g. "quinto") (cf. Vigário *et alii*, 2005, 2006) - e não correspondem a sequências de oclusiva e semivogal (vd. secção 3.1.2.1).

[43] O *Speech Assessment Methods Phonetic Alphabet* (SAMPA) é um alfabeto fonético, especialmente concebido para o tratamento das línguas europeias por computador. Desenvolvido sob o signo do projecto ESPRIT, no final da década de 80, foi aplicado, em primeira instância, ao dinamarquês, holandês, inglês, francês, alemão e italiano. Só mais tarde foi alargado a outras línguas, incluindo o português (1993). Para mais informações, cf. http://www.phon.ucl.ac.uk/home/sampa.

O processo por nós implementado, baseado na proposta de Bouma (2002), consiste na composição de três transdutores, que actuam directamente sobre o nível ortográfico. Estes são definidos a partir de expressões regulares e têm como objectivo final a delimitação das palavras de entrada em sílabas, mediante a inserção de uma marca. Assim, a) o primeiro assinala o Núcleo; b) o seguinte insere marcas de fronteira de sílaba; c) e o último remove as marcas de Núcleo.

Toda a implementação assenta numa descrição simplificada da estrutura interna da sílaba do PE, segundo o modelo "Ataque-Rima", apresentado na secção 3.1.2.1.

Numa primeira fase, enunciaram-se as consoantes e grupos consonânticos que, de acordo a descrição fonológica do PE, podem figurar em posição de Ataque.

Como é sabido, em posição de Ataque simples, qualquer consoante do PE é admitida. Há ainda que contar com os grupos consonânticos que formam Ataques ramificados. Neste caso, as únicas combinações consideradas foram as sequências de oclusiva+líquida ou fricativa+líquida, por respeitarem os princípios de boa formação silábica (vd. secção 3.1.2.1).

No sentido de dar conta de combinações consonânticas (obstruinte+obstruinte ou obstruinte+nasal) que violam o *Príncípio de Sonoridade* e *Condição de Dissemelhança*, não sendo, por conseguinte, aceites como Ataques ramificados, postulámos a existência de um Núcleo vazio entre as duas consoantes adjacentes (assinalado, no sistema, com um "V"), subscrevendo, assim, a proposta defendida por Mateus & d'Andrade (2000) e explicitada na secção 3.1.2.1. A lista de sequências consonânticas infractoras dos referidos princípios não é exaustiva, pelo que será possível a qualquer momento acrescentar novos grupos ao inventário.

Deste modo, a lista de possíveis Núcleos inclui não só as vogais presentes a nível fonológico, mas também as sequências de vogal+segmento nasal /N/, bem como o próprio o símbolo "V", que representa uma posição vocálica vazia, passível de ser preenchida por uma vogal epentética. Para além disso, considerámos ainda os chamados ditongos decrescentes, orais e nasais, cujos elementos, segundo a interpretação fonológica de Mateus & d'Andrade (2000) são sempre representados no Núcleo. Os ditongos crescentes, que resultam de um processo opcional de semivocalização, foram ignorados (vd. secção 3.1.2.1).

Quanto aos segmentos associados à Coda, que, em português, como vimos anteriormente (3.1.2.1), tem uma estrutura simplificada, são eles o <l>, o <r> e o <s>/<z>/<x>.

Definidas as listas dos grafemas passíveis de preencher os diferentes constituintes silábicos, os três transdutores actuam em sequência, com base no comando *replace*:

- Marcação do Núcleo, inserindo um "@" antes e depois dos grafemas constantes da lista de possíveis Núcleos:

```
replace([ []: @, id(nucleo), []:@, [], [])
```

- Marcação das fronteiras silábicas (através do símbolo "/"), com base na estrutura da sílaba

tradicionalmente aceite: Rima com Núcleo obrigatório e Ataque e Coda opcionais (o símbolo "^" assinala a não obrigatoriedade).

```
replace([]:'|', [@, coda^], [onset^ , @])
```

- Finalmente, remoção das marcas delimitadoras do Núcleo:

```
replace(#:[], [], []).
```

Se tomarmos como exemplo a palavra "raptar", teremos sucessivamente:

```
r@a@p@V@t@a@r
r@a@|p@V@|t@a@r
ra|pV|tar
```

3.1.4.2 Implementação do algoritmo de silabificação de Mateus & d'Andrade (2000)

O algoritmo de silabificação proposto por Mateus & d'Andrade (2000) foi implementado em duas versões: uma que actua directamente sobre o texto escrito, beneficiando da natureza fonológica da ortografia portuguesa (Viana *et alii*, 1991; Mateus, 2006); uma outra capaz de silabificar uma sequência de fones, que somente poderá ter lugar após a conversão grafema-fone.

Os vários procedimentos que compõem o algoritmo - implementados, por agora, sob a forma de programas independentes, de modo a permitir uma cuidadosa monitorização e inspecção dos resultados da aplicação de cada uma das etapas de processamento - procuram, na sua generalidade, ser um espelho das convenções propostas por Mateus & d'Andrade (2000) para a silabificação do PE. O processo de silabificação do texto de entrada - esteja ele no formato ortográfico ou fonético - é assim totalmente transparente para o utilizador.

O resultado da execução de cada um dos referidos procedimentos é um documento XML (*e**X**tensible **M**arkup **L**anguage*) [44], uma metalinguagem de anotação que garante grande flexibilidade ao sistema, já que permite o acesso às etapas intermédias do processamento. Este é realizado em Perl, através do XML::DOM. Abordagens similares foram adoptadas no desenvolvimento de sistemas TTS como o *Mary* (Schröder & Trouvain, 2003) ou o *Prosynth* (Ogden *et alii*, 2000).

Passamos, em seguida, a descrever o essencial das propriedades e funcionamento de cada uma das etapas do processamento. As informações dizem respeito ao algoritmo de silabificação de base ortográfica, mas aplicam-se também, de modo similar, à segunda versão do algoritmo, que corre directamente sobre os fones.

[44] O XML é uma linguagem de anotação extensível, desenvolvida sob a égide do W3C (*World Wide Web Consortium*, com o objectivo essencial de servir de suporte a um leque muito variado de aplicações (incluindo a criação de conteúdos para a Internet). Um documento em XML é visto como uma estrutura lógica, em que os elementos estão hierarquicamente organizados numa espécie de árvore (Ramalho & Henriques, 2002).

- **Etapa 0 - Pré-processamento**

  O primeiro passo, que podemos designar por pré-processamento da informação, consiste na criação de um nodo (ou nó) na hierarquia XML para cada uma das palavras constantes da lista de entrada. Em seguida, o sistema divide cada destas palavras nos respectivos grafemas (*letter*), fazendo-lhes corresponder um sub-nodo (ou sub-nó) *syllable*, o que significa que cada um dos grafemas é considerado uma potencial sílaba.

  Esta última é caracterizada por um conjunto de atributos relacionados com 1) a localização do acento de palavra (*stress*), marcado com um "1"; 2) os traços silábico (*Syllabic*) e consonântico (*Consonantic*), que indiciam quais os segmentos que poderão preencher o Núcleo; 3) e o constituinte silábico (*PartSyllable*).

  O processo de determinação do acento não foi alvo de um trabalho aturado e resulta da adaptação do código disponibilizado no âmbito do Projecto Natura (Almeida *et alii*, 2003). Por razões de eficiência, o conjunto de regras de atribuição do acento lexical foi implementado em Perl, usando expressões regulares.

  As grafias com duas letras, i.e. os dígrafos (e.g. <ss> ou <rr>), são agrupadas e associadas a uma única sílaba potencial. Em relação ao <qu> e <gu> foi seguida uma abordagem similar, sendo estas sequências de grafemas analisadas, para efeito de processamento, como uma única consoante.

  Tomando como exemplo a palavra "claustro", em particular a primeira letra (<c>), sabe-se já que o acento de palavra não recai sobre este grafema, assinalado também como segmento [+consonântico], pelo que não poderá jamais constituir-se como Núcleo, embora o lugar ocupado por esta consoante na estrutura da sílaba permaneça, por enquanto, desconhecido ("?").

```
  <?xml version="1.0" encoding="iso-8859-1" ?>
- <xml>
- <PHRASE>
- <WORD SPELLING="claustro" PRONUNCIATION="?" STRESS="cla1us|tro">
- <syllable>
  <letter Stress="-"  Consonantic="+" Syllabic="-" PartSyllable="?">c</letter>
  </syllable>
- <syllable>
  <letter Stress="-" Consonantic="+" Syllabic="-" PartSyllable="?">l</letter>
  </syllable>
- <syllable>
  <letter Stress="1" Consonantic="-" Syllabic="+" PartSyllable="?">a</letter>
  </syllable>
- <syllable>
  <letter Stress="-"  Consonantic="-" Syllabic="+" PartSyllable="?">u</letter>
  </syllable>
- <syllable>
  <letter Stress="-" Consonantic="+" Syllabic="-" PartSyllable="?">s</letter>
  </syllable>
- <syllable>
  <letter Stress="-" Consonantic="+" Syllabic="-" PartSyllable="?">t</letter>
```

```
  </syllable>
- <syllable>
  <letter Stress="-" Consonantic="+" Syllabic="-" PartSyllable="?">r</letter>
  </syllable>
- <syllable>
  <letter Stress="-" Consonantic="-" Syllabic="+" PartSyllable="?">o</letter>
  </syllable>
  </WORD>
  </PHRASE>
```

- **Etapa 1 - Núcleo**

  Depois da etapa preliminar, todos os grafemas assinalados como [+silábicos] foram referenciados como Núcleo, conforme consignado na *Convenção de Associação de Núcleos* do algoritmo que nos serve de referência (Mateus & d'Andrade, 2000). Adicionalmente, os restantes grafemas [+ silábicos], simultaneamente [-acentuados] e precedidos de um [+silábico], são marcados como Núcleo2 e associados ao Núcleo da esquerda. Esta abordagem é coerente com o princípio segundo o qual a semivogal fonética do ditongo decrescente tem na base uma vogal fonológica, sendo ambos os elementos do ditongo dominados pelo mesmo Núcleo.

  Retomando o exemplo anterior, o <a> de "claustro" é processado como *Nucleus* (N), enquanto o <u> seguinte é assinalado como *Nucleus_2* (N2). Esta última vogal está sujeita a um processo de semivocalização obrigatória no plano fonético e vai associar-se ao Núcleo da sílaba anterior.

```
  <?xml version="1.0" encoding="iso-8859-1" ?>
- <xml>
- <PHRASE>
- <WORD SPELLING="claustro" PRONUNCIATION="?" STRESS="cla1us|tro">
- <syllable>
  <letter Stress="-"  Consonantic="+" Syllabic="-" PartSyllable="?">c</letter>
  </syllable>
- <syllable>
  <letter Stress="-"  Consonantic="+" Syllabic="-" PartSyllable="?">l</letter>
  </syllable>
+ <syllable>
  <letter Stress="1"  Consonantic="-" Syllabic="+" PartSyllable="N">a</letter>
  <letter Stress="-"  Consonantic="-" Syllabic="+" PartSyllable="N2">u</letter>
  </syllable>
- <syllable>
  <letter Stress="-"  Consonantic="+" Syllabic="-" PartSyllable="?">s</letter>
  </syllable>
- <syllable>
  <letter Stress="-"  Consonantic="+" Syllabic="-" PartSyllable="?">t</letter>
  </syllable>
- <syllable>
  <letter Stress="-"  Consonantic="+" Syllabic="-" PartSyllable="?">r</letter>
  </syllable>
- <syllable>
  <letter Stress="-"  Consonantic="-" Syllabic="+" PartSyllable="N">o</letter>
  </syllable>
  </WORD>
  </PHRASE>
```

- **Etapa 2 - Ataque**

  Tal como previsto na *Convenção de Associação de Ataques* do algoritmo original (Mateus & d'Andrade, 2000), na segunda etapa do processamento, todos os grafemas assinalados como [+ consonânticos] que precedem o Núcleo são associados ao Ataque e, por conseguinte, marcados como "Ataque" no atributo "PartSyllable".

  Seguidamente, dando cumprimento à segunda parte do algoritmo de Mateus & d'Andrade (2000), os segmentos [+ consonânticos] que precedem o Ataque recentemente criado, são avaliados quanto à possibilidade de constituirem, juntamente com a consoante seguinte, um Ataque ramificado. No âmbito do nosso sistema - seguindo a interpretação fonológica geralmente aceite para o PE - a emergência de Ataques ramificados assenta no respeito não só pelo *Princípio de Sonoridade*, mas também pela *Condição de Dissemelhança*. Assim, aos segmentos foram atribuídos valores distintos de sonoridade, de acordo com a escala proposta por Vigário & Falé (1993). Sempre que o diferencial mínimo de sonoridade silábica entre as duas consoantes se manteve, o segmento foi classificado como "Ataque" e associado ao Ataque da sílaba seguinte.

  Com efeito, no exemplo de "claustro", os segmentos <l> e <r> são classificados como *Ataque* pelo sistema. Depois de avaliada a distância mínima de sonoridade em relação aos segmentos precedentes, os grafemas [+consonânticos] (respectivamente <c> e <t>) vêm juntar-se à consoante seguinte para formar um Ataque ramificado.

```
  <?xml version="1.0" encoding="iso-8859-1" ?>
- <xml>
- <PHRASE>
- <WORD SPELLING="claustro" PRONUNCIATION="?" STRESS="cla1us|tro">
- <syllable>
  <letter Stress="-"  Consonantic="+" Syllabic="-" PartSyllable="?">c</letter>
  </syllable>
- <syllable>
  <letter Stress="-" Consonantic="+" Syllabic="-" PartSyllable="ATAQUE">l</letter>
  <letter Stress="1" Consonantic="-" Syllabic="+" PartSyllable="N">a</letter>
  <letter Stress="-"  Consonantic="-" Syllabic="+" PartSyllable="N2">u</letter>
  </syllable>
- <syllable>
  <letter Stress="-"  Consonantic="+" Syllabic="-" PartSyllable="?">s</letter>
  </syllable>
- <syllable>
  <letter Stress="-"  Consonantic="+" Syllabic="-" PartSyllable="?">t</letter>
  </syllable>
- <syllable>
  <letter Stress="-"  Consonantic="+" Syllabic="-" PartSyllable="ATAQUE">r</letter>
  <letter Stress="-" Consonantic="-" Syllabic="+" PartSyllable="N">o</letter>
  </syllable>
  </WORD>
  </PHRASE>


  <?xml version="1.0" encoding="iso-8859-1" ?>
- <xml>
```

```
- <PHRASE>
- <WORD SPELLING="claustro" PRONUNCIATION="?" STRESS="cla1us|tro">
- <syllable>
  <letter Stress="-" Consonantic="+" Syllabic="-" PartSyllable="ATAQUE">c</letter>
  <letter Stress="-" Consonantic="+" Syllabic="-" PartSyllable="ATAQUE">l</letter>
  <letter Stress="1" Consonantic="-" Syllabic="+" PartSyllable="N">a</letter>
  <letter Stress="-" Consonantic="-" Syllabic="+" PartSyllable="N2">u</letter>
  </syllable>
- <syllable>
  <letter Stress="-" Consonantic="+" Syllabic="-" PartSyllable="?">s</letter>
  </syllable>
- <syllable>
  <letter Stress="-" Consonantic="+" Syllabic="-" PartSyllable="ATAQUE">t</letter>
  <letter Stress="-" Consonantic="+" Syllabic="-" PartSyllable="ATAQUE">r</letter>
  <letter Stress="-" Consonantic="-" Syllabic="+" PartSyllable="N">o</letter>
  </syllable>
  </WORD>
  </PHRASE>
```

- **Etapa 3 - Coda**

  Os restantes segmentos [+consonânticos] - ainda sem posição silábica atribuída e pertencentes ao restrito conjunto de consoantes admitidas em Coda - são especificados como Coda no respectivo nó *PartSyllable* e associados à sílaba anterior.

  Com o objectivo de definir o inventário de Codas possíveis e, sobretudo, de aferir a pertinência da escala de sonoridade silábica adoptada - nomeadamente dos valores de sonoridade atribuídos a cada segmento e da distância mínima de sonoridade permitida, com vista ao respeito da *Condição de Dissemelhança* - foram processadas cerca de 100.000 palavras, monitorizando todas as decisões tomadas pelo programa à saída destas duas últimas etapas do processamento.

  A título ilustrativo dos efeitos desta etapa do processamento, em “claustro”, a única consoante marcada como Coda é o <s>, já que pertence ao inventário de Codas pré-definido.

```
  <?xml version="1.0" encoding="iso-8859-1" ?>
- <xml>
- <PHRASE>
- <WORD SPELLING="claustro" PRONUNCIATION="?" STRESS="cla1us|tro">
- <syllable>
  <letter Stress="-"  Consonantic="+" Syllabic="-" PartSyllable="ATAQUE">c</letter>
  <letter Stress="-"  Consonantic="+" Syllabic="-" PartSyllable="ATAQUE">l</letter>
  <letter Stress="1"  Consonantic="-" Syllabic="+" PartSyllable="N">a</letter>
  <letter Stress="-"  Consonantic="-" Syllabic="+" PartSyllable="N2">u</letter>
  <letter Stress="-"  Consonantic="+" Syllabic="-" PartSyllable="CODA">s</letter>
  </syllable>
- <syllable>
  <letter Stress="-"  Consonantic="+" Syllabic="-" PartSyllable="ATAQUE">t</letter>
  <letter Stress="-"  Consonantic="+" Syllabic="-" PartSyllable="ATAQUE">r</letter>
  <letter Stress="-"  Consonantic="-" Syllabic="+" PartSyllable="N">o</letter>
  </syllable>
  </WORD>
  </PHRASE>
```

- **Etapa 4 - Núcleo vazio**

  À luz das descrições fonológicas do português, na impossibilidade de aceitar certas combinações consonânticas como Ataques ramificados por violarem os princípios de boa formação silábica, a primeira consoante da sequência deve ser entendida como o Ataque de uma primeira sílaba com Núcleo vazio. Esta solução encontra-se consignada na *Convenção de Criação de Núcleos Vazios* do algoritmo de Mateus & d'Andrade (2000) e é replicada na última etapa de processamento do nosso sistema [45].

  Com efeito, à direita dos grafemas [+ consonânticos], cuja *PartSyllable* permanece ainda por apurar, foi introduzida uma nova posição silábica ("V"), que pretende representar um Núcleo vazio, superficialmente preenchido por uma vogal epentética.

  Escolhemos o exemplo de "afta" para ilustrar o processamento efectuado, no sentido de implementar computacionalmente a solução prescrita na literatura fonológica do PE para resolver as referidas violações dos princípios silábicos.

  Já depois da avaliação formal do algoritmo, este foi alterado de modo a não mais considerar Núcleos vazios, deixando que a coordenação gestual entre gestos consonantais habilite ou não a emergência de uma vogal epentética.

O *output* do sistema são árvores silábicas geradas automaticamente, de que são exemplo os esquemas arbóreos seguintes, referentes à silabificação de "claustro" e "afta":

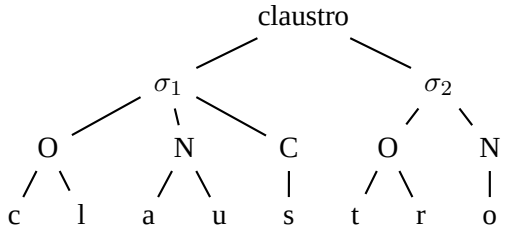


Figura 3.2: Árvore silábica da palavra "claustro".

3.1.4.3 Avaliação

Parâmetros de avaliação

A avaliação dos dois algoritmos desenvolvidos teve em conta, não só a concordância na estipulação dos limites silábicos, mas também a introdução de marcas espúrias (i.e. marcas de fronteira de sílaba inexistentes), ambas expressas em percentagem.

[45] A ordem de aplicação das regras de silabificação na última versão do algoritmo de Mateus & d'Andrade (2000) e na implementação automática do mesmo algoritmo não é exactamente coincidente, nomeadamente no tocante à associação de Codas e criação de Núcleos vazios. Por razões de eficiência, optámos, em primeiro lugar, por integrar na estrutura da sílaba as Codas, antes mesmo de tratar dos Núcleos vazios, conforme descrito em versões mais antigas do mesmo algoritmo (Mateus, 1994).

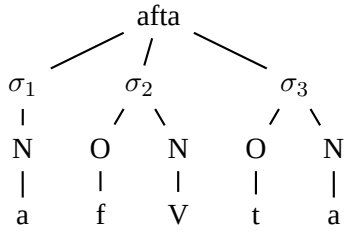


Figura 3.3: Árvore silábica da palavra “afta”.

Todos os parâmetros de avaliação foram obtidos automaticamente, usando um programa em Perl, especialmente desenvolvido para o efeito.

A percentagem de divisões silábicas acertadas foi definida como a razão entre o número de fronteiras silábicas correctamente determinadas e o número total de sílabas. O segundo parâmetro (percentagem de inserções) consiste na diferença entre o número total de marcas de fronteira de sílaba e o número de marcas introduzidas. O desempenho global (ou precisão) do sistema foi obtido à custa da subtracção destes dois parâmetros. Finalmente, foi também calculada a percentagem de palavras correctamente processadas.

No seguimento destas observações, conclui-se que a avaliação se ficou pela análise dos limites silábicos, reservando-se para uma fase posterior o desenvolvimento de procedimentos conducentes à avaliação detalhada dos vários constituintes silábicos.

Corpora

O sistema foi desenvolvido com base num pequeno *corpus* com cerca de 1000 palavras.

Os resultados resultam da aplicação do programa a dois *corpora* de teste: o primeiro (*Teste 1*) consiste numa pequena amostra - mais exactamente 2076 entradas - do Português Fundamental (PF) (Nascimento *et alii*, 1987); e o segundo (*Teste 2*) contém 1300 palavras aleatoriamente seleccionadas do *corpus* CETEMPúblico (excluindo estrangeirismos, erros de grafia, siglas e acrónimos), construído pela Linguateca [46], a partir das edições *online* do jornal com o mesmo nome. Em termos gerais, desta última lista fazem parte palavras mais longas e de uso menos comum.

A cada entrada em formato ortográfico foi adicionada a respectiva transcrição fonética, usando o alfabeto fonético SAMPA. As transcrições foram geradas automaticamente com o sistema de regras descrito no capítulo seguinte e, depois, processadas manualmente para correcção dos valores fonéticos atribuídos aos segmentos.

À transcrição fonética, juntou-se, também manualmente, informação sobre a localização das fronteiras de sílaba [47]. De modo a tornar o processo mais rápido, as marcas silábicas foram

[46] Cf. http://www.linguateca.pt.

[47] As fronteiras silábicas foram assinaladas com um ponto (.), seguindo as convenções do Alfabeto Fonético International

automaticamente transpostas para o nível ortográfico e, posteriormente, sujeitas a correcção manual. Este processo foi precedido de uma etapa de alinhamento automático entre as formas ortográficas e as suas respectivas transcrições fonéticas, uma vez que a cada símbolo de entrada (grafema) nem sempre corresponde apenas um símbolo de saída (fone) e vice-versa. Torna-se, assim, necessário indicar, por exemplo, que certos grafemas não têm realização fonética (e.g. <h> inicial) ou que a um só grafema pode corresponder uma sequência de fones (e.g. <x> [ks]).

A necessidade de constituição de um *corpus*, com transcrição fonética e divisão silábica, decorre da impossibilidade de aceder a um *corpus* que contemple esta informação e permita aferir os resultados do nosso sistema. Esta estratégia de construção de um *corpus* específico para treino e teste dos métodos automáticos de silabificação implementados compromete, de algum modo, uma medida objectiva do desempenho desses mesmos sistemas, mas afigurou-se como a única solução ao nosso alcance, em face da indisponibilidade de *corpora* desta natureza

Como foi já referido, e existência de *corpora* é crucial para testar a eficácia dos sistemas, mas também nos casos em que o paradigma *data-driven* é aplicado, já que constitui a base (ou o modelo) de aprendizagem para novas silabificações. A ausência de um "*gold standard*" *corpus* impõe-se, assim, como uma grande limitação para quem quer desenvolver e/ou testar sistemas automáticos deste género, já para não falar da inibição da comparação entre sistemas, com base no mesmo *corpus* de teste.

3.1.4.4 Resultados

Tabela 3.1: Resultados (em percentagem) da avaliação dos vários métodos de silabificação automática, nos dois corpora de teste (teste 1 e teste 2).

Corpus		M&A Grafemas	FSTs (Grafemas)	PT::PLN (Grafemas)	M&A Fones
Teste 1	Divisões Correctas (%)	99.77	99.27	96.62	98.46
	Inserções (%)	0.08	1.17	0.19	0.26
	Palavras Correctas (%)	99.57	98.80	93.93	97.88
Teste 2	Divisões Correctas (%)	99.59	98.92	96.06	98.80
	Inserções (%)	0.15	0.75	0.15	0.03
	Palavras Correctas (%)	98.85	97.40	88.95	96.47

Os resultados da avaliação dos dois algoritmos - o primeiro dos quais em duas versões (grafemas e fones) - segundo os parâmetros previamente enunciados - percentagem de divisões silábicas correctas (*Divisões Correctas (%)*), percentagem de inserções (*Inserções (%)*) e percentagem de palavras correctamente processadas (*Palavras Correctas (%)*) - são apresentados na tabela 3.1, para os dois *corpora* de teste. Exclusivamente para efeitos comparativos, foram ainda incluídos os resultados do desempenho do sistema de silabificação disponibilizado pelo Projecto Natura (vd. secção 3.1.3).

(AFI).

Da observação destes valores é possível concluir que: 1) o desempenho do algoritmo de Mateus & d'Andrade (2000) aplicado à ortografia é claramente superior, quando comparado com os restantes três métodos, não só ao nível da determinação de fronteiras silábicas e número de palavras correctamente silabificadas, mas também no que toca à baixa percentagem de inserções; 2) no caso da versão original do algoritmo de Mateus & d'Andrade (2000), com um desempenho ligeiramente inferior, o parâmetro de avaliação mais afectado é o número de palavras correctamente processadas, enquanto que, no algoritmo implementado com base em FSTs, a maior taxa de erro se regista ao nível das inserções; 3) de um modo geral, o desempenho dos vários algoritmos de silabificação é inferior para o segundo *corpus* de teste, o que poderá estar relacionado com as diferenças no tamanho e complexidade das palavras.

Tendo em conta estes dados, convém salientar que a fiabilidade do nosso melhor algoritmo de silabificação (M&A Grafemas) é muito semelhante à de outros sistemas com as mesmas potencialidades e que correm também sobre o texto escrito, nomeadamente o proposto no âmbito do Frep (Vigário *et alii*, 2005), cujo desempenho global se situa à volta dos 99.709%.

Grande parte dos erros detectados ficam a dever-se não a falhas directas na estipulação das fronteiras silábicas, mas a problemas na atribuição do acento ou alinhamento entre grafemas e fones, que inevitavelmente fomentam erros de silabificação.

Sublinhamos ainda um outro tipo de erro, recorrente no algoritmo M&A aplicado aos grafemas, que resulta da manutenção, na ortografia do PE, de consoantes etimológicas mudas (e.g. <p>, como em "óptimo"). Com efeito, estas são tratadas pelo sistema automático de silabificação como o primeiro elemento de uma sequência de duas obstruintes, entre as quais é introduzida uma posição vocálica vazia.

Já o algoritmo de silabificação baseado em FSTs não é absolutamente eficaz na identificação da vogal nasal como Núcleo de sílaba e no tratamento das sequências <qu> e <gu>. Por outro lado, como foi já referido, a lista de sequências consonânticas infractoras não é exaustiva.

3.1.5 Comentários finais

Motivações práticas que se prendem com a silabificação automática do *input* do TADA e o aumento do desempenho do módulo de conversão grafema-fone, a par com as dificuldades de acesso a um sistema aberto para fins de investigação, que servisse, ao mesmo tempo, as nossas necessidades, ditaram o desenvolvimento e teste de dois algoritmos de divisão silábica.

Ambos fazem apelo a conhecimentos de ordem fonológica sobre a estrutura da sílaba em PE, sendo que o segundo método aqui descrito (3.1.4.2) procura mesmo ser uma implementação fiel do algoritmo de silabificação de base proposto por Mateus & d'Andrade (2000). Dado o seu carácter eminentemente categórico, este tipo de algoritmo presta-se naturalmente ao tratamento automático, permitindo atingir desempenhos superiores aos de outros sistemas testados, nomeadamente o algo-

ritmo baseado em FSTs ou o sistema de silabificação automática disponibilizado pela Linguateca, cujo princípio de funcionamento assenta noutros critérios, nem sempre muito claros, mas não exclusivamente de base fonológica.

Este é, portanto, um exemplo claro de como o carácter utilitário deste tipo de sistemas não implica necessariamente que estes se desvinculem dos estudos linguísticos teóricos e de como os sistemas automáticos podem e devem servir para implementar e, sobretudo, testar as propostas fonológicas. O procedimento é semelhante ao seguido em outros trabalhos - como o de Howard & Goldman (1994), cujas regras de silabificação automática seguem o algoritmo de Hualde (1991), inscrito na fonologia generativa - em que se recorre à metodologia científica e ferramentas da engenharia para comprovar a adequação das regras fonológicas propostas.

Muitos erros poderão ainda ser evitados se os procedimentos de atribuição do acento lexical e alinhamento automático entre grafemas e fones vierem a ser melhorados. Uma análise exaustiva dos mecanismos que permitiriam minimizar este tipo de erros não se justifica de momento, pois, se, por um lado, o algoritmo de identificação do acento utilizado, ainda que longe de atingir resultados perfeitos, serve as necessidades básicas relacionadas com a silabificação e a transcrição fonética, por outro, a informação relativa ao acento não é usada na actual versão do TADA. Quando ao alinhamento automático, é um problema eminentemente técnico, a ser resolvido fora do âmbito desta tese.

Ainda que o sistema de silabificação automática sirva objectivos muito concretos, em última análise, o módulo agora proposto pode facilmente ser integrado num qualquer sistema TTS ou mesmo acoplado a dicionários electrónicos para gerar informação silábica, já que funciona de forma totalmente independente. Para além disso, esta ferramenta pode ser facilmente adaptada para extrair informação sobre a frequência de tipos silábicos do PE.

3.2 Conversão grafema-fone

> *Two main justifications are commonly given for the continuing need for a GTP component. Firstly, there will always be genuinely new words ("blairism", "email", "yuppie") created in the course of time. In addition there are many words which may not be new, but were ignored when the system was originally built and have now become common enough to require proper pronunciation (e.g. "bin laden"). In such cases GTP conversion will always be required.*
>
> Taylor, 2005

3.2.1 Introdução

Conforme mencionado em 2.1.3 (página 25), o processo de mapeamento de uma sequência de grafemas numa sequência de fones é uma tarefa complexa, sujeita a algumas ambiguidades, que decorrem

essencialmente do facto de não existir uma correspondência directa e biunívoca entre a grafia e os sons.

Mesmo no caso do português, cuja relação grafia-fonia é caracterizada por uma elevada regularidade (d'Andrade & Viana, 1985; Viana *et alii*, 1991; Mateus, 2006) [48], é possível encontrar alguns exemplos em que o processo de passagem do símbolo gráfico ao som não é, de modo algum, linear. O mesmo grafema <c> pode representar vários sons distintos, como nas palavras "cereja" e "cama". Inversamente, um mesmo som pode ter mais do que uma grafia. Ilustram esta possibilidade a equivalência sonora de <s>, <ss>, <c>, <ç> e <x>, em "saco", "massa", "cebola", "laço" e "próximo".

A escolha do método de conversão grafema-fone a adoptar depende, pois, da natureza do sistema ortográfico da língua a transcrever, entre outros factores, de entre os quais se destaca o próprio objectivo e características do sistema de síntese. Tal como tivemos ocasião de referir em 2.1.3, entre as várias soluções disponíveis, contam-se 1) a consulta sistemática de um dicionário; 2) a aplicação de regras de transcrição (eventualmente combinadas com um dicionário de excepções); e 3) o recurso a técnicas de aprendizagem automática (*machine learning*).

Em línguas como o inglês, por exemplo, a falta de correspondência entre o sistema ortográfico e o sistema fonológico impõe o recurso a grandes léxicos, já que a aplicação de regras de transcrição, se revela, na maior parte dos casos, ineficaz.

Já o português, em virtude da já citada regularidade ortográfica, presta-se mais a um tratamento baseado em regras de reescrita (Oliveira, 1996; Viana *et alii*, 1991; Oliveira *et alii*, 1992; Teixeira *et alii*, 1998; Barbosa *et alii*, 2003a; Simões, 1999; Albano & Moreira, 1996; Albano & Aquino, 1997), muito embora tenham sido propostas outras abordagens para resolver o problema da conversão grafema-fone (vd. próxima secção).

O desenvolvimento de um novo sistema de conversão grafema-fone(ma), no âmbito deste trabalho, significa, por um lado, que o assunto está longe de estar resolvido - continuando a motivar várias publicações, como facilmente se depreenderá do ponto 3.2.2 - e, por outro lado, que, apesar dos esforços empreendidos nesse sentido, não foi possível aceder a nenhum dos sistemas já existentes, para posterior modificação, extensão e, principalmente, integração no TADA e no sintetizador articulatório da UA.

Numa primeira fase do nosso trabalho, optámos por uma das abordagens mais populares na conversão grafema-fone, em geral, e na transcrição ortográfico-fonética do PE, em particular. Referimo-nos, naturalmente, à aplicação de regras de reescrita, que no nosso caso específico, foram combinadas com uma técnica de aprendizagem automática (vd. ponto 3.2.3.1). Segundo Braga

[48] Embora a ortografia portugesa seja considerada de "natureza fonológica" (Mateus, 2006), isto não significa que esta se baseie exclusivamente em critérios fonológicos. Tal como qualquer outro sistema ortográfico, também a ortografia portuguesa resulta de um compromisso convencionado entre a fonética e a etimologia, regendo-se por um conjunto de regras específicas, muitas delas alheias à relação com a oralidade. A este propósito, cf. Rui Vieira de Castro (1987, apud Mateus, 2006, nota 13): "...os sistemas ortográficos movem-se na tensão entre dois pólos: a tendência para uma representação fonética da língua e a tendência para uma representação histórica, marcada pela manutenção ou recuperação de matrizes etimológicas".

(2008), o recurso a regras linguísticas é, sem dúvida, a melhor estratégia para resolver o problema da conversão grafema-fone em PE: se, por um lado, é mais económica em termos de memória computacional do que uma abordagem baseada num dicionário - o que, no caso concreto dos ambientes móveis, é uma vantagem bastante importante - permite, concomitantemente, atingir desempenhos superiores aos obtidos por métodos automáticos. Esta mesma técnica é também bastante eficaz a lidar com as palavras novas, que vão sendo permanentemente associadas ao léxico (Taylor, 2005).

O desenvolvimento de algoritmos de conversão grafema-fone, com base em regras, é ainda, acrescentaríamos nós, uma matéria de investigação de grande interesse para a Linguística. Um sistema deste género pode e deve ser encarado como uma ferramenta de ensaios e avaliação de diferentes modelos e teorias linguísticas, em que o investigador tem a oportunidade de implementar computacionalmente uma gramática ou um modelo fonológico e aceder a ele, sempre que necessário, para monitorizar e modificar as regras em tempo real.

No tocante ao nosso sistema em particular, a abordagem mista, baseada em regras manuais, corrigidas por uma técnica automática, viabilizou, acima de tudo, o rápido desenvolvimento de um sistema de conversão grafema-fone, adequado às nossas necessidades práticas.

Impulsionados por essa mesma motivação, enveredámos, numa segunda fase, pela exploração de outros métodos de aprendizagem automática, avaliados em comparação com a abordagem baseada em regras. Esta experiência é relatada e comentada no ponto 3.2.3.2 deste capítulo.

3.2.2 Conversão grafema-fone para o português europeu

No respeitante ao PE, várias têm sido as abordagens adoptadas ao longo da última década para resolver o problema da conversão grafema-fone, sobretudo - mas não exclusivamente - no âmbito do sistema DIXI, o primeiro sintetizador de fala a partir do texto, desenhado de raiz para a língua portuguesa, resultado de uma investigação multidisciplinar e de um trabalho de colaboração entre a equipa de Processamento de Fala do INESC e o grupo de Fonética e Fonologia do Centro de Linguística da Universidade de Lisboa (CLUL).

A primeira versão do sistema DIXI (Oliveira *et alii*, 1991, 1992, 1993; Oliveira, 1996), desenhada a partir do sintetizador de formantes de Klatt, compreende um módulo de transcrição fonética, baseado num conjunto de regras de reescrita, mais concretamente 18 regras de atribuição do acento lexical, escritas directamente em linguagem C, e cerca de 200 regras de transcrição fonética automática [49], codificadas através do compilador de regras SCYLA (Viana *et alii*, 1991).

Mais tarde, esta abordagem baseada em regras foi comparada com outros dois métodos de aprendizagem automática, nomeadamente redes neuronais e busca em tabelas (Trancoso *et alii*, 1994, 1995; Trancoso & Viana, 1995). Em causa esteve ainda o desempenho dos vários sistemas na

[49] As regras de transcrição fonética que integram o DIXI são muito similares às propostas por d'Andrade & Viana (1985), no âmbito do desenvolvimento do CORSO I (Conversor de texto ortográfico em código fonético para o português), um programa que tem por objectivo gerar a representação fonética de superfície partindo da representação ortográfica.

transcrição de nomes próprios, por oposição ao léxico comum. Não obstante a diminuta taxa de erro conseguida pelas redes neuronais, os melhores resultados foram obtidos com a abordagem clássica, baseada em regras.

A experiência com redes neuronais e busca em tabelas não chegou a integrar o sintetizador (Trancoso *et alii*, 1994), ao contrário da abordagem testada numa fase posterior, baseada em *Classification and Regression Trees* ou simplesmente CARTs (Oliveira *et alii*, 2001). O módulo de conversão grafema-fone da versão actual do DIXI, desenvolvida no âmbito do projecto Tecnovoz, recupera esta técnica, com resultados de 93.28% ao nível da palavra, e de 98.12%, ao nível dos fones (Paulo *et alii*, 2008).

Recentemente, um conjunto de técnicas (regras manuais, *data-driven* e um método híbrido) foi implementado através de FSTs (Caseiro *et alii*, 2002), sendo que esta mesma abordagem foi, posteriormente, aplicada ao tratamento do mirandês (Caseiro *et alii*, 2003).

A respeito da conversão grafema-fone do PE, importa ainda destacar o trabalho de Teixeira *et alii* (1998), cujo sistema de síntese (MULTIVOX) inclui um módulo de transcrição fonética com cerca de 600 regras. Uma grande parte destas regras foi, posteriormente, incorporada no sintetizador da Faculdade de Engenharia da Universidade do Porto e os algoritmos subjacentes ao módulo de conversão parcialmente publicados em Teixeira (2004).

Intimamente relacionado com este trabalho estão as propostas de Braga *et alii* (2006) e Braga (2006, 2008), igualmente baseadas em regras linguísticas. Para além do PE, os algoritmos de conversão grafema-fone propostos foram adaptados a outras línguas/variedades, nomeadamente o português do Brasil e o galego (Silva *et alii*, 2006; Braga & Coelho, 2006; Braga, 2008).

Finalmente, a questão da conversão grafema-fone foi ainda abordada por Almeida & Simões (2001) - que sugere um conjunto de regras para converter as palavras em fones, contemplando ainda alguns fenómenos de Shandi - e por Barros & Weiss (2006), cujos algoritmos de transcrição fonética, de determinação do acento e das fronteiras da sílaba se fundam em modelos de máxima entropia.

3.2.3 Sistemas de conversão grafema-fone implementados

3.2.3.1 Sistema de conversão grafema-fone baseado em transdutores de estados finitos

O primeiro sistema G2P desenvolvido é baseado na aplicação de um pequeno conjunto de regras definidas manualmente e implementadas através de FSTs, seguindo as propostas de Bouma (2000), para a língua holandesa. Os erros de transcrição produzidos pelas regras manuais são, posteriormente, corrigidos, mediante uma técnica de aprendizagem automática, o TBL (Brill, 1995). O processo de conversão dos grafemas em fones é suportado por um conjunto de subtarefas, mais concretamente a atribuição do acento lexical e a divisão silábica. Todos os procedimentos que asseguram a silabificação automática das palavras de entrada foram descritos em pormenor na secção 3.1.4.1. Quanto ao acento, muito embora a sua marcação seja absolutamente essencial ao eficaz funcionamento do processo de

transcrição fonética, o módulo de acentuação não foi ainda alvo de um trabalho aturado e resulta da adaptação de um código disponibilizado pelo Projecto Natura (Almeida *et alii*, 2003). O acento primário é assinalado com um "1" a seguir à vogal.

Os vários processos foram integrados num programa desenvolvido em linguagem Perl. Este programa utiliza uma base de dados (em linguagem SQL) para guardar informação e viabilizar procuras posteriores. O sistema corre actualmente em ambiente Linux.

A transcrição fonética é efectuada, com base na informação disponibilizada pelos módulos de acentuação e silabificação automática.

Antes de qualquer outra operação, o sistema verifica se a palavra a transcrever faz, ou não, parte do dicionário de excepções. Se tal se verificar é-lhe, de imediato, associada a entrada lexical correspondente. O dicionário contém um conjunto de formas que implicam um tratamento particular e que, de algum modo, escapam às regras gerais de transcrição ou de acentuação, mesmo após sucessivos refinamentos. Actualmente, a lista de excepções não é muito extensa, mas admitimos a possibilidade de vir a completá-la com novas palavras.

De seguida, as regras de transcrição, desenvolvidas manualmente e implementadas através de um transdutor de estados finitos, actuam sobre aquelas palavras que não constam do dicionário, de modo a transformar a sequência de entrada na sequência fonética equivalente. Neste momento, as regras operam apenas ao nível lexical, ou seja, cada palavra é transcrita isoladamente de acordo com o contexto próximo. A versão actual do sistema não faz nenhum tipo de tratamento pós-lexical, ou seja, não são considerados os fenómenos decorrentes de encontros consonantais e vocálicos em fronteira de palavra.

Por fim, as transcrições erradas são corrigidas através da aplicação de novas regras obtidas por aprendizagem automática.

Regras manuais compiladas através de FSTs

Na linha do proposto por Bouma (2000), o transdutor resultante da composição das várias regras foi definido com base no comando *replace*, que permite transformar uma sequência de entrada numa determinada sequência de saída, sempre que esta se encontra num contexto específico. No exemplo que se segue, o grafema <a>, em posição tónica, quando seguido de <m,n,nh> , passa a [6] [50](e.g. "cama", "dano", "banho").

```
g2p( a x '6', [],  [1,'|',{m,n,'J'}])
```

As regras são aplicadas de forma sequencial, pelo que, enquanto as primeiras operam apenas sobre os grafemas, as últimas actuam, essencialmente, sobre os fones, entretanto gerados. Este pro-

[50]Com o intuito de evitar problemas de processamento, foi usado alfabeto fonético SAMPA, em vez do Alfabeto Fonético Internacional.

cesso obriga, também, a que se ordenem as regras do específico para o geral, evitando que a aplicação das mesmas crie ou destrua o contexto necessário à aplicação de outras regras subsequentes.

Sendo que as regras actuam ao nível das palavras, a informação acerca dos limites das mesmas é essencial, pelo que o processo tem início com a marcação das fronteiras de palavra (símbolo "#"). Para que o fone obtido não volte a ser alvo da acção de outras regras, na linha do proposto por Bouma (2000), é inserida uma marca, mais concretamente um "-", após cada grafema. Sempre que este sofre a acção de uma regra e é convertido num fone, a marca que lhe sucede transforma-se em "+". O exemplo (tabela 3.2), ilustra o processamento da palavra "rasga":

Tabela 3.2: Exemplo de processamento da palavra "rasga".

Silabificação	ras|ga
Marcação dos limites da palavra	#ras|ga#
Inserção de "-" (segmentação)	#r-a-s-|g-a-#
Aplicação de uma regra: o <s> antes de segmento [+vozeado] realiza-se como [Z] `g2p( s x 'Z', [], ['|',consv])`	 #r-a-Z+|g-a-#

No total, foram definidas 162 regras para proceder à transcrição dos grafemas do PE, que resultam num transdutor de 468 estados.

Apesar de, nesta fase do trabalho, continuar a existir espaço para mais desenvolvimentos, o grau de complexidade da tarefa tende a aumentar consideravelmente, com a definição de regras de aplicação cada vez mais restrita, através da análise minuciosa dos resultados das várias etapas do desenvolvimento. Foi então que a mudança de uma abordagem linguística para um método de aprendizagem automática nos pareceu adequada.

Correcção automática das regras manuais TBL

Com o objectivo de corrigir alguns dos erros de transcrição resultantes da aplicação das regras manuais, recorremos, conforme foi já referido, a uma técnica de aprendizagem automática, o TBL. Trata-se de um método originalmente proposto por Brill (1995) para a anotação automática da categoria morfossintáctica das palavras, entretanto aplicado a outras tarefas de processamento da linguagem natural, incluindo a conversão grafema-fone (Bouma, 2000). O processamento tem lugar em várias etapas: primeiro, o texto é anotado por um sistema inicial (*base-line system*), mais ou menos complexo; em seguida, os resultados deste sistema são alinhados com a "verdade", i.e. são comparados com um *corpus* de referência manualmente anotado; finalmente, o TBL "aprende" um conjunto de regras de transformação, que podem ser aplicadas ao sistema inicial, de modo a que a saída deste se aproxime cada vez mais do *corpus* manual.

No caso específico da conversão grafema-fone, o sistema inicial é o transdutor de estados

finitos descrito anteriormente, usado para transcrever foneticamente uma lista de palavras. Os resultados gerados pela aplicação de regras de reescrita são, depois, alinhados e comparados com a transcrição fonética correcta, i.e. introduzida manualmente, num *corpus* de treino. Deste confronto resulta a aprendizagem automática de regras que corrigem os erros de pronunciação realizados pelo primeiro sistema. De acordo com estas, um símbolo fonético será, então, substituído por outro, sempre que um determinado número de condições esteja reunida. A definição do contexto é da responsabilidade do utilizador. Num bloco de declarações são determinados os contextos a considerar na aplicação da regra, que podem abranger desde os fones e grafemas contíguos até uma janela de quatro fones/grafemas à direita e à esquerda do fone a processar. Depois de considerar os exemplos a favor e contra a implementação de cada regra possível, o sistema selecciona a regra mais pontuada e aplica-a.

A título de exemplo (tabela 3.3), consideremos as palavras "caixeiro" e "encaixe", transcritas respectivamente como [kajx6jru] e [e~kajx@], após a aplicação das regras manuais. Durante o processo automático, o TBL "aprendeu" uma regra de substituição do [x]* pelo fone [S], quando este é precedido de [j], baseada em quatro exemplos a favor e nenhum contra. A aplicação da nova "regra" transforma [kajx6jru] e [e~kajx@] em [kajS6jru] e [e~kajS@], respectivamente.

Tabela 3.3: Exemplo de aplicação de TBL às palavras "caixeiro" e "encaixe".

Resultado das regras manuais	Após regras TBL
kajx6jru e~kajx@	kajS6jru e~kajS@
`GOOD:4 BAD:0 SCORE:4` `RULE: fone_-1=j fone_0=x => fone=S`	

Corpora

***Corpus* para desenvolvimento das regras manuais** - As formas incluídas no *corpus* usado para desenvolvimento das regras correspondem, *grosso modo*, a palavras seleccionadas, automaticamente, de textos da edição *online* do jornal Público, disponibilizados pela Linguateca [51] e previamente processados. A cada entrada do léxico foi adicionada, manualmente, a transcrição fonética (para além de uma marca de identificação da sílaba tónica e informação acerca dos limites da sílaba), usando os símbolos do alfabeto fonético SAMPA. A comparação entre a transcrição fonética realizada pelo sistema, com base nas regras, e a anotação manual permitiu a monitorização, passo a passo, do funcionamento das regras definidas com base na nossa própria competência de falantes e nas informações publicadas na literatura sobre a Fonética e a Fonologia do PE (e.g Faria *et alii*, 1996; Mateus & d'Andrade, 2000; Mateus *et alii*, 2003, 2005). Já que o sistema possibilita também a visualização das regras e dos resultados da sua aplicação, foi possível, sempre que necessário, modificá-las ou mesmo criar novas regras

[51] Cf. http://www.linguateca.pt.

ainda não consideradas.

***Corpus* para treino de TBL** - A necessidade de um *corpus* mais extenso para treino de TBL ditou a criação de um léxico de pronunciação de cerca de 8000 palavras, com transcrição gerada automaticamente por um sistema baseado em árvores e modelos de previsão (CARTs) e verificada manualmente. Este *corpus* corresponde a parte do *Ispell Dictionary*, versão portuguesa (disponibilizada pelo Projecto Natura) [52].

***Corpus* de teste** - Durante a fase de teste foi usado um *corpus*, constituído também a partir da edição *online* do jornal Público, com cerca de 1000 entradas seleccionadas aleatoriamente, de modo automático, a que foram retirados os estrangeirismos, as siglas e os acrónimos.

Um dos requisitos indispensáveis à aplicação de técnicas de aprendizagem automática, neste caso o TBL, consiste no alinhamento dos dados para treino e teste. Tendo em conta que as sequências de entrada e saída nem sempre são iguais - na medida em que há grafemas que não têm realização fonética (e.g. <h>), casos em que uma sequência de grafemas corresponde a um só fone (e.g. dígrafos) e vice-versa - a correspondência entre as cadeias de grafemas e respectivas transcrições fonéticas é, normalmente, garantida à custa da inserção de símbolos nulos.

No nosso caso específico, a tarefa de alinhamento do TBL com o sistema de base (regras manuais implementados através de um transdutor de estados finitos) foi, em grande parte, simplificada pela introdução do símbolo "-" depois de cada grafema. A questão revelou-se bem mais problemática no dicionário anotado manualmente, que obrigou a um alinhamento forçado, com ajustes manuais.

Resultados

Na presente versão, a aplicação das regras manuais ao *corpus* de teste produz 73.2% de resultados fonéticos correctos, ao nível da palavra. No que diz respeito aos fones, estes são correctamente transcritos em 96.3% dos casos.

A maior parte dos erros de transcrição está relacionada com uma transcrição errada da consoante <x> e das vogais <o> e <e>.

A consoante <x> está sujeita a uma enorme variabilidade [53] na transcrição e, de acordo com os resultados do teste, o conjunto de regras actualmente definidas não é suficiente para dar conta de todas as possibilidades de realização. Para cada uma das regras consideradas é quase sempre possível identificar um conjunto de formas de excepção, o que não só dificulta a definição das próprias regras, como faz aumentar a taxa de erro.

O desempenho do sistema relativamente às vogais nasais é claramente positivo, atingindo

[52] Cf. http://natura.di.uminho.pt

[53] O grafema <x> pode ser realizado como [s] ("auxílio"), [S] ("xarope"), [z] ("exacto") ou [ks] ("táxi").

quase os 100% de acerto. O mesmo não se verifica com as vogais orais, nomeadamente com o <e> e <o> quando tónicos. No que diz respeito ao <e>, a dificuldade reside em decidir quando se realiza como [e] ou como [E]. Do mesmo modo, um grande número de erros prende-se com a transcrição do <o> como [O] ou como [o]. Também as regras gerais de transcrição das vogais átonas são, em alguns casos, incorrectamente aplicadas (e.g. uma pequena percentagem de <a> são convertidos em [6], quando, na verdade, correspondem a [a]).

Alguns destes problemas decorrem directamente do facto de o sistema não incorporar, até ao momento, um analisador morfológico. Uma análise deste tipo permitiria resolver grande parte dos casos de ambiguidade fonológica.

Devido a constrangimentos relacionados com o formato diferenciado do *corpus* de treino e de teste, apenas nos foi possível avaliar o desempenho das regras automáticas com base no *corpus* de treino. Esta situação impede, obviamente, a generalização dos resultados apresentados, que devem, portanto, ser analisados com alguma cautela.

Segundo a tabela 3.4, a taxa de acerto aumenta consideravelmente, após a aplicação das regras automáticas, que corrigem os erros de transcrição fonética gerados pelo processamento manual anterior. Ainda assim, o desempenho do nosso sistema fica um pouco abaixo dos resultados atingidos por Bouma (2000), uma diferença que poderá estar relacionada com o tamanho do *corpus* de treino (8000 entradas para o PE vs 20000 entradas para o holandês). Com efeito, quanto mais extenso o *corpus* de treino, mais completo o espaço amostral dos fenómenos fonéticos de uma língua e maior a possibilidade de aprendizagem de novas regras.

Tabela 3.4: Comparação entre o desempenho das regras manuais e o TBL, no PE e no holandês (resultados de Bouma, 2000).

	Resultados para PE		Resultados para o Holandês (Bouma, 2000)	
	Regras	TBL	Regras	TBL
Palavras	73.2	89.3	60.6	86.1
Fones	96.3	98.7	93.6	98.0

3.2.3.2 Sistema de conversão grafema-fone baseado em métodos automáticos

Numa segunda fase do trabalho, a abordagem baseada em regras linguísticas cedeu lugar ao desenvolvimento de um conjunto de métodos automáticos para a conversão grafema-fone do PE. Foram implementados e testados dois métodos de auto-aprendizagem, o *Memory-based Learning* (MBL) e o TBL, considerados individualmente ou combinados entre si (em paralelo e em cascata). Simultaneamente, foi também investigado o interesse em incorporar informação silábica neste tipo de abordagem automática.

Todas as experiências de transcrição fonética automática do PE foram efectuadas em Perl, usando XML::DOM.

Num primeiro momento, as palavras foram silabificadas automaticamente e a informação estruturada num documento XML, de modo a possibilitar a extracção das propriedades e a estruturação da saída num formato adequado. Depois de treinados e testados com as ferramentas e algoritmos seleccionados, os diferentes sistemas foram analisados quanto ao seu desempenho, usando um conjunto de *scripts* Perl.

Transformation-Based Learning - O funcionamento deste método foi já descrito no ponto 3.2.3.1. Para a realização das várias experiências, seleccionámos o pacote *fnTBL*, uma ferramenta disponível em código fonte, originalmente orientada para tarefas de processamento de linguagem natural (Florian & Ngai, 2001).

Memory-Based Learning - De acordo com a metodologia MBL, a transcrição fonética de uma palavra de entrada pode ser determinada com base em exemplos de palavras com pronúncia similar.

O algoritmo MBL "take a set of examples (fixed-length patterns of feature-values and their associated class) as input, and produce a classifier which can classify new, previously unseen, input patterns" (Daelemans *et alii*, 2004, p.10).

O nosso sistema de conversão grafema-fone para o PE foi desenvolvido com base no TiMBL (Daelemans *et alii*, 2004), uma ferramenta usada com relativo sucesso na conversão grafema-fone de várias línguas (Daelemans & Bosch, 1997). Esta ferramenta implementa diferentes algoritmos de auto-aprendizagem (IB1, IB2, IGTREE, TRIBL e TRIBL2), cuja característica comum reside na capacidade de armazenamento em memória de algum tipo de representação do conjunto de treino. Durante a fase de teste, os novos casos são classificados de acordo com o exemplo mais próximo, armazenado na memória. As principais diferenças entre os vários algoritmos implementados pelo TiMBL residem: 1) na definição de similaridade; 2) no modo como as representações são guardadas em memória; e 3) na maneira como a pesquisa através da memória é conduzida.

3.2.3.3 Propriedades

As características (em inglês *features*) usadas para o treino dos modelos, inspiradas em Reichel & Schiel (2005), foram:

- GRAFEMA - identifica o grafema actual;
- POS_NA_SILABA - determina a posição ocupada pelo grafema na estrutura da sílaba. As várias possibilidades são: Ataque, Núcleo e Coda;
- SIL_FRONTEIRA - especifica se há ou não uma fronteira de sílaba a seguir ao grafema;

- SIL_POS_PALAVRA - determina qual a posição ocupada pela sílaba que contém o grafema (medida em percentagem do número de sílabas da palavra);
- LEX_ACENTO - assinala a posição do acento lexical. No MBL, a sílaba tónica é marcada com um 0, a sílaba pré-tónica com um -1, etc. No TBL, esta característica é especificada apenas como "acentuado" ou "não-acentuado".

Para o MBL, as características GRAFEMA e POS_NA_SILABA foram extraídas usando uma janela de comprimento 11, centrada no grafema actual.

Para o TBL, por razões de complexidade das regras admissíveis, não foram consideradas as propriedades SIL_FRONTEIRA and SIL_POS_PALAVRA.

3.2.3.4 Parâmetros de avaliação

A avaliação dos vários métodos automáticos na conversão grafema-fone do PE teve em conta a percentagem de erro ao nível das palavras (WER) e a percentagem de erro ao nível dos fones (PER). Nos sistemas desenvolvidos, a saída do sistema automático está rigorosamente alinhada com a transcrição fonética manual, o que de algum modo previne que os erros decorrentes de omissões ou inserções se propaguem a toda a cadeia de fones subsequentes, o que seria altamente penalizante para o desempenho dos sistemas.

Ainda a pensar neste último problema, foi ainda usado um terceiro parâmetro de avaliação, originalmente proposto por Reichel & Schiel (2005) e designado de *Mean Normalized Levenshtein Distance* (MNLD). Este é definido como "the minimum number of edit operations (the Levenshtein distance) to convert one string into the other divided by the lenght of the reference string" (Reichel & Schiel, 2005, p.1939). A referência para comparação com a saída do sistema é a transcrição original.

3.2.3.5 Corpora

A fase de teste foi precedida de um processo semi-automático de alinhamento entre os grafemas e respectivas transcrições fonéticas, de modo a viabilizar comparações automáticas entre os dois níveis.

***Corpora* para treino dos métodos automáticos** - Os *corpora* usados para treino dos diferentes métodos automáticos foram constituídos com base na versão portuguesa do dicionário Ispell, um recurso nascido da colaboração entre a Linguateca e o Projecto Natura, da Universidade do Minho. A par da forma ortográfica, cada entrada contém indicação da categoria gramatical.

O primeiro *corpus* de treino, correspondente a parte do dicionário Ispell, inclui 6500 entradas diferentes, às quais foram manualmente adicionadas as respectivas transcrições fonéticas.

Já no decorrer da experiência, foi iniciada a criação de um segundo *corpus* de treino, também com base no dicionário Ispell, com o intuito de fazer face ao problema da exiguidade de dados

disponíveis para treino dos sistemas. Neste caso, a transcrição fonética foi gerada automaticamente por um dos sistemas de conversão grafema-fone entretanto desenvolvidos, com base em MBL, e, depois, verificada manualmente. Durante este processo, algumas entradas, correspondentes a estrangeirismos ou simples erros ortográficos, foram eliminadas. Este *corpus* é constituído, actualmente, por cerca de 4000 entradas, que combinadas com o primeiro *corpus* de treino, deram origem a uma base de dados de 10.5 k palavras.

***Corpora* para teste dos sistemas G2P** - A avaliação dos sistemas baseou-se nos dois *corpora* - usados para teste dos sistemas de silabificação automática - descritos em 3.1.4.3. Recordamos que o primeiro corresponde a um subconjunto do PF (Nascimento *et alii*, 1987) e é composto por 2076 palavras e o segundo inclui cerca de 1300 palavras retiradas do *corpus* CETEMPúblico, elaborado e disponibilizado pela Linguateca.

3.2.3.6 Resultados

Para além de avaliar o desempenho dos dois métodos de auto-aprendizagem (MBL e TBL) na conversão grafema-fone do PE, através das experiências descritas abaixo, procurámos também investigar o interesse e o impacto de incorporar informação silábica aos sistemas automáticos. A par dos sistemas básicos, foram ainda testadas várias combinações entre os dois métodos automáticos e o sistema de conversão grafema-fone baseado em regras linguísticas.

Sistema MBL

O primeiro sistema proposto assenta na exploração do método MBL e na análise do impacto da informação silábica sobre o desempenho final do sistema.

Depois de efectuados alguns testes preliminares com os vários algoritmos implementados pelo TiMBL, usando as configurações por defeito, foi seleccionado, com base nos resultados obtidos, o algoritmo TRIBL2. O sistema foi treinado com dois *corpora*, de dimensão distinta, o primeiro composto por 6.500 palavras e o segundo contendo cerca de 10.500.

Os resultados com os dois *corpora* de teste (teste 1 e teste 2) são apresentados na tabela 3.5.

A partir da análise da tabela, conclui-se que tanto a inclusão informação silábica como o aumento do tamanho do *corpus* de teste (de 6.5 k para 10.5 k) têm um impacto positivo sobre os resultados, que se traduz numa diminuição da percentagem de erro ao nível das palavras (WER), dos fones (PER) e da MNLD.

É também importante notar que a taxa de erro nas palavras e nos fones (WER e PER) é ligeiramente mais baixa para o teste 1 do que para o teste 2. Acreditamos que esta diferença possa estar relacionada com as características intrínsecas dos dois *corpora*: enquanto o teste 1 é constituído

Tabela 3.5: Resultados do MBL (algoritmo TRIBL2) nos dois *corpora* de teste.

Sistema				Teste 1			Teste 2		
No.	Treino	Sílaba	Algoritmo	PER%	WER%	MNLD	PER%	WER%	MNLD
s1	6.5k	Não	TRIBL2	5.01	27.26	0.056	6.68	44.43	0.063
s2		Sim	TRIBL2	3.88	22.06	0.045	5.67	37.51	0.051
s3	10.5k	Não	TRIBL2	4.33	24.95	0.050	5.36	37.74	0.049
s4		Sim	TRIBL2	**3.76**	**21.63**	0.043	4.79	32.36	**0.042**

por palavras de uso comum, do teste 2 fazem parte vocábulos mais longos, alguns deles de origem dita "erudita" ou pertencentes a domínios mais técnicos.

Sistema TBL

No desenvolvimento dos sistemas baseados em TBL, para além da questão da informação silábica, foram equacionados outros aspectos, nomeadamente a importância do ponto de partida da aprendizagem. Como vimos, este último pode variar bastante no que respeita ao nível de complexidade. Para o caso específico da conversão G2P do PE, como base para o treino do TBL, foram consideradas duas alternativas distintas: 1) uma tabela de conversão grafema-fone básica, com as correspondências mais comuns entre os grafemas e os fones; 2) o sistema de transcrição ortográfico-fonética, baseado em regras manuais, descrito anteriormente (secção 3.2.3.1).

Depois de treinado, o sistema foi testado com os mesmos *corpora* já mencionados (teste 1 e teste 2), tendo em conta as várias variáveis em estudo: informação silábica, dimensão do *corpus* de treino, ponto de partida para a aprendizagem. Os resultados constam da tabela 3.6.

Tabela 3.6: Resultados do TBL com as duas listas de teste (teste 1 e teste 2), em função da dimensão do *corpus* de treino, do ponto de partida da aprendizagem (tabela ou regras) e da informação silábica.

Sistema				Teste 1			Teste 2		
No.	Treino	P. Partida	Sílaba	PER%	WER%	MNLD	PER%	WER%	MNLD
s5	6.5k	Tabela	não	7.90	43.00	0.088	8.66	56.07	0.091
s6			sim	5.09	27.73	0.057	5.15	36.48	0.055
s7		Regras	não	5.42	29.23	0.061	5.94	38.33	0.063
s8			sim	4.85	26.96	0.055	4.65	33.18	0.051
s9	10.5k	Tabela	não	6.71	37.22	0.077	7.04	48.23	0.076
s10			sim	4.26	**23.74**	**0.049**	4.03	29.34	**0.043**
s11		Regras	não	4.58	25.13	0.053	5.01	33.56	0.055
s12			sim	**4.19**	**23.74**	**0.049**	**3.89**	**28.42**	**0.043**

Tal como no sistema baseado em MBL, a primeira conclusão a reter da tabela prende-se com a diminuição da taxa de erro (WER, PER, MNLD), em consequência da introdução no sistema

Tabela 3.7: Resultados da abordagem WTA, em que os dois métodos *data-driven* (TBL e MBL) se combinam com o sistema baseado em regras manuais.

Sistema			Teste 1			Teste 2		
No.	Treino	Sílaba	PER%	WER%	MNLD	PER%	WER%	MNLD
s13	6.5k	não	3.36	19.16	0.038	6.53	44.44	0.061
s14		sim	2.77	16.23	0.032	3.13	22.91	0.027
s15	10.5k	não	2.79	16.42	0.032	5.30	38.20	0.049
s16		sim	2.66	15.75	0.031	2.91	21.37	0.025

de informação acerca dos limites silábicos e da utilização de um *corpus* de treino de maior dimensão.

Quanto à questão do ponto de partida para aprendizagem do TBL, verifica-se que o desempenho do sistema é, em geral, superior, quando a base para o treino são as regras manuais. O ponto de partida do TBL parece ainda ter um influência determinante no grau de impacto da informação silábica sobre o desempenho do sistema G2P. Com efeito, embora a introdução de informação silábica resulte sempre numa diminuição da percentagem de erro, este efeito é muito mais evidente, quando a base para a aprendizagem é uma tabela de correspondências.

Sistemas híbridos

O primeiro dos sistemas híbridos desenvolvido resulta da combinação dos dois sistemas de auto-aprendizagem acima descritos (MBL e TBL) com a abordagem baseada em regras, apresentada na secção 3.2.3.1.

Neste sistema, auto-intitulado "*Winner Take All*" (WTA), cada palavra é processada, em paralelo, pelos dois métodos automáticos e pelo sistema baseado em regras linguísticas, sendo que, no final, é retida a decisão da maioria.

Os resultados da avaliação deste sistema encontram-se na tabela 3.7.

Olhando para os resultados tabela 3.7, mais uma vez se assiste a uma diminuição da taxa de erro, em virtude da utilização de informação silábica e do aumento do tamanho do *corpus* de treino. Destacamos os resultados obtidos para o teste 2 (*corpus* de treino de 6.5k), em que a inclusão de informação silábica se traduz numa diferença de 21.53 %, para a WER, e de 0.034, para o MNLD, entre o sistema 13 e o 14.

Com o intuito de aprofundar estes dados e de apurar as contribuições individuais de cada sistema para a decisão do WTA, foi calculado o número de vezes em que 1) a concordância entre os três sistemas foi total; 2) a decisão final resulta do acordo entre dois sistemas, sejam eles quais forem; 3) não houve qualquer tipo de acordo.

Em termos globais, os resultados indicam que a percentagem de concordância entre os três sistemas é superior (cerca de 93%), quando a informação silábica está disponível. Caso contrário,

Tabela 3.8: Resultados da combinação em cascata dos dois métodos automáticos, usando o MBL como ponto de partida para a aplicação de TBL.

Sistema				Teste 1			Teste 2		
No.	Treino MBL	Treino TBL	Sílaba	PER%	WER%	MNLD	PER%	WER%	MNLD
s17	6.5k	4k	não	17.16	60.38	0.178	17.93	74.50	0.185
s18			sim	3.83	21.86	0.044	4.71	33.56	0.049
s19	4k	6.5k	não	17.33	59.51	0.182	18.67	74.89	0.194
s20			sim	4.49	25.81	0.051	4.49	33.41	0.048

i.e., havendo apenas informações relativas aos grafemas e respectivos contextos, a decisão do WTA decorre, na maioria das vezes, do acordo entre o sistema de regras e o MBL.

Um outro aspecto que merece ser salientado diz respeito à percentagem de des(acerto) da decisões do WTA. Verificou-se que, quando a informação silábica está acessível e existe, ao mesmo tempo, uma correspondência absoluta entre as saídas dos três sistemas, a percentagem de erro ronda os 0.6 %. Contudo, em alguns casos, nomeadamente quando não há qualquer tipo de acordo entre os sistemas e é seleccionado o *output* do MBL, esta mesma percentagem ascende aos 78 % .

O segundo sistema hibrído - e último sistema desenvolvido - consiste numa combinação em cascata dos dois métodos de auto-aprendizagem (MBL e TBL), explorando as funcionalidades básicas do TBL, enquanto método que “aprende” regras de correcção. Os resultados das experiências com este sistema constam da tabela 3.8.

Os resultados seguem, mais uma vez, a tendência relativa à informação silábica e dimensão do *corpus* de treino e, em geral, são muito inferiores aos obtidos a partir da abordagem WTA.

Claramente, o TBL não só não é capaz de corrigir os erros gerados pelo MBL, como é responsável pela introdução de novos erros. Com efeito, o desempenho da combinação do MBL com o TBL é inferior à obtida pelo MBL isoladamente.

Ainda assim, parece preferível usar um maior volume de dados para treinar o MBL do que o TBL.

3.2.4 Comentários finais

O primeiro sistema de conversão grafema-fone apresentado baseia-se na combinação de um conjunto de regras - definidas manualmente e implementadas através de FSTs - com um método de aprendizagem automática (TBL). Esta estratégia viabilizou o desenvolvimento rápido de um primeiro sistema G2P para o PE, dando assim seguimento a um dos objectivos enunciados na Introdução a este trabalho.

Embora o trabalho de aperfeiçoamento e definição de novas regras linguísticas apresentasse margem para mais desenvolvimentos - nomeadamente no que respeita à atribuição do acento lexical - a tarefa veio a revelar-se bastante complexa e demorada, em grande parte devido ao problema de aplicação sequencial das regras. A solução encontrada passou, então, por corrigir as regras manuais através de um método de auto-aprendizagem (TBL), seguindo as propostas de Bouma (2000). Uma das principais vantagens do TBL reside na capacidade de atingir desempenhos elevados, usando um *corpus* de treino relativamente pequeno (Bouma, 2000), o que efectivamente se veio a comprovar nos nossos resultados: como vimos, o recurso ao TBL conduziu a um aumento considerável da taxa de acerto, tanto a nível das palavras como dos fones. Também aqui haveria espaço para melhorias, sendo, contudo, necessário recorrer a um *corpus* de dimensão mais alargada, de que, actualmente, não dispomos.

Embora as regras manuais sejam consideradas a melhor estratégia para resolver o problema da conversão grafema-fone em línguas com ortografias de base fonológica, como é o caso do português, "no set of rules, no matter how extensive and complicated, can ever describe or account for the totality of any individual's or group's language" (Hall, 1972, p.42). É com base neste pressuposto, que, nos últimos anos, a abordagem linguística tem vindo a perder terreno para as técnicas *data-driven*, cujos resultados suplantam, pelo menos em algumas línguas (e.g. inglês) (Damper *et alii*, 1999), o desempenho dos métodos fundados em regras.

Tendo presentes estes argumentos, enveredámos, numa segunda fase, pela exploração de outros métodos automáticos - para além do TBL - isolados ou combinados entre si, procurando igualmente avaliar o impacto da utilização de informação silábica sobre o desempenho global do sistema.

Comparando o desempenho dos vários sistemas implementados e testados, conclui-se que o método WTA - resultado da combinação de dois sistemas de aprendizagem automática com o sistema de regras linguísticas - é o que apresenta resultados mais interessantes (WER de 15.75 % no teste 1 e MNLD igual a 0.025 no teste 2). Considerando apenas os sistemas individuais, os melhores resultados pertencem ao MBL (PER de 3.76 %, WER de 21.63 % e MNLD igual a 0.043 %, no teste 1).

Em todos os sistemas G2P com desempenho superior, os 10 erros mais frequentes dizem respeito à transcrição das vogais <e>, <o> e, algumas vezes, <a> (teste 1). No teste 2, evidenciam-se as mesmas dificuldades em lidar com as vogais, para além de problemas com a conversão do <x> e do <s>.

À semelhança do que acontece para outras línguas - e.g. inglês (Marchand & Damper, 2007; Bartlett *et alii*, 2008), francês (Beringer, 2004) e alemão (Libossek & Schiel, 2000) - também no caso do PE, o desempenho dos sistemas G2P aumenta significativamente com a integração de informação silábica. Uma outra via a explorar, tendo em vista o aumento do desempenho deste tipo de sistemas poderia passar por incorporar também informação morfológica (Reichel & Schiel, 2005).

Para além da sílaba, também a dimensão do *corpus* de treino parece ter uma influência determinante nos resultados. De acordo com estes, o alargamento do *corpus* de treino tem sempre como

consequência uma diminuição da percentagem de erro, tanto ao nível dos fones como das palavras.

É no tamanho do *corpus* de treino (e de teste) que reside também uma das maiores fragilidades deste trabalho. Como é sabido, o desenvolvimento de métodos *data-driven* assenta em *corpora* de treino de grandes dimensões - que funcionam como modelo para a aprendizagem - mas que no caso do PE, não estão publicamente disponíveis. O desenvolvimento de recursos deste tipo é uma tarefa muito morosa e dispendiosa do ponto de vista material e humano, apenas ao alcance de uma grande equipa. Face a este problema, a solução passou pela criação de um pequeno *corpus* de treino próprio, lançando mão de uma estratégia que começou por ser totalmente manual, passando depois a automática - através da aplicação do MBL - com posterior verificação manual. Apesar de todos os esforços desenvolvidos, a dimensão deste *corpus* está ainda longe de ser a ideal, o que significa que os nossos sistemas automáticos podem ter potencialmente desempenhos superiores, desde que se utilize para treino um *corpus* mais alargado.

Uma forma eficaz e rápida de ampliar este *corpus* - ou até de produzir um *"gold standard" corpus* para a língua portuguesa, para utilização em tarefas de processamento de linguagem natural - poderá mesmo passar por aplicar algoritmos de conversão grafema-fone automáticos, com correcção manual *a posteriori*, em vez da tradicional anotação manual.

Também no que respeita ao *corpus* de teste se reconhecem algumas limitações, já mencionadas anteriormente a propósito da avaliação dos sistemas de silabificação automática. As dificuldades de acesso a um *corpus standard* para teste dos sistemas, motivou o recurso a dois *corpora* de teste, também desenvolvido por nós, o que, de algum modo, compromete a objectividade da avaliação.

Não obstante as dificuldades supra-mencionadas, julgamos ter cumprido minimamente os objectivos a que nos propusémos ao desenvolver um conjunto de sistemas G2P para o PE, com um desempenho razoável, pelo menos para as nossas actuais necessidades.

Capítulo 4

Modelo Gestual para o Português Europeu

> The tongue, the lips articulate; the throat
> With soft vibration modulates the note.
>
> Darwin, *The Temple of Nature*, Canto III, l.367

O presente capítulo ocupar-se-á da caracterização gestual dos vários sons do PE, de modo a que cada fone obtido à saída do módulo de transcrição automática possa ter uma correspondência com um conjunto de gestos articulatórios adequados para o português.

Depois de uma breve descrição do funcionamento do modelo gestual do TADA - que, como vimos no capítulo 2 (secção 2.2.5), se divide em dois componentes distintos, o *Syllable structure-based gestural coupling model*, que gera um *coupling graph*, especificando os gestos associados aos segmentos de entrada e as relações de coordenação intergestual, e o *Coupled oscillator model of inter-gestural coordination*, que calcula os intervalos de activação de cada um dos gestos e gera a respectiva pauta gestual - procederemos à apresentação da metodologia geral que presidiu à definição dos gestos. Esta definição será, sobretudo, suportada por dados de produção, nomeadamente informação obtida através de ressonância magnética. Sempre que se revelar necessário, e em virtude da escassez de dados articulatórios relativos ao PE, serão tidos em conta dados acústicos - já que a partir deles podem ser feitas inferências sobre a configuração do tracto vocal - ou mesmo estudos realizados para outras línguas.

As configurações gestuais para cada um dos segmentos do PE serão apresentadas e fundamentadas, em secções distintas, começando pelas vogais e passando pelas várias classes de consoantes. Sempre que se revelar necessário, serão ainda fornecidas informações adicionais sobre alterações aos padrões de coordenação pré-definidos no TADA.

A avaliação das propostas gestuais - e de todas as modificações introduzidas nos parâmetros de entrada do TADA - decorreu em duas fases distintas: 1) apreciação informal, pelo investigador, da qualidade do som gerado, a partir da configuração gestual proposta; 2) teste de inteligibilidade com

vários sujeitos. A última parte do capítulo centrar-se-á na descrição da construção e aplicação deste último teste perceptivo, bem como na análise e discussão dos resultados obtidos.

4.1 Modelo gestual no TADA

No âmbito do sistema TADA, a geração automática do *coupling graph* associado ao texto de entrada tem duas componentes distintas: 1) criação de uma lista de gestos associados ao *input*, representados em termos de parâmetros dinâmicos que caracterizam a variável do tracto, peso relativo dos articuladores e *blending*; 2) especificação da coordenação intergestual entre os osciladores associados aos gestos, com base na estrurura silábica.

O funcionamento destes dois componentes será descrito, sumariamente, em seguida.

4.1.1 Composição gestual

Depois de convertido em transcrição fonética (ARPABET) e automaticamente silabificado, mediante a aplicação de um algoritmo, o texto de entrada é associado a um conjunto de gestos, segundo as especificações de um dicionário, que contém a configuração gestual de cada um dos segmentos do inglês.

Neste dicionário, os gestos são representados simbolicamente por quatro parâmetros fundamentais: articulador envolvido na produção (*organ*), tipo de oscilador (*osc*), variável do tracto (*TV*) e constrição (*constr*), conforme exemplificado no quadro 4.1, para a fricativa [Z] [1]. O exemplo foi directamente extraído do dicionário gestual (*TADA/gest/english/gestures_english.xl*s) do TADA (Browman *et alii*, 2001-2006), referente ao inglês.

Tabela 4.1: Definição gestual da consoante inglesa [Z], segundo os parâmetros do dicionário gestual do TADA (fonte: Browman *et alii*, 2001-2006). A consoante encontra-se foneticamente representada em ARPABET (ARPA) e os gestos a ela associados são caracterizados pelo articulador (*Organ*), oscilador (*Osc*), variável do tracto (*TV*) e constrição (*Const*). Os pontos indicam que estão a ser utilizados os valores de *target* e *stiff* (rigidez da mola) pré-definidos em ficheiro independente.

ARPA	Organ	Osc	TV	Const	Target	Stiff
Z	TT	crt	TTCL	ALV	.	.
	TT	ctr	TTCD	CRIT	0,16	.
	TT	rel	TTCL	REL	.	.
	TT	rel	TTCD	REL	.	.
	TB	crt	TBCL	VEL	.	.
	TB	crt	TBCD	WIDE	.	.
	Velum	crt	VEL	CLO	.	.

Seguindo os pressupostos da FA (vd. capítulo 2, ponto 2.2.1), os gestos traduzem-se em

[1] A consoante [Z] (ARPABET) corresponde ao [ʒ] (AFI).

cinco possíveis *variáveis do tracto* (TV) - Lábios (L), Ponta da Língua (TT), Corpo da Língua (TB), Velo (VEL) e Glote (GLO) - que se referem simultaneamente a duas dimensões, local (CL) e grau de constrição (CD), ambas especificadas sob a designação geral de *constrição* (*Const*) [2].

Quanto à variável CL, estão previstos os seguintes descritores gestuais: os lábios podem receber a especificação de dental `[DENT]`, protruído `[PRO]`, distendido `[REL]`; à ponta da língua são atribuídas as etiquetas dental `[DENT]`, alveolar `[ALV]`, alveolo-palatal `[ALVPAL]`, palatal `[PAL]` e distendido `[REL]`; o corpo da língua pode ser palatal `[PAL]`, velar `[VEL]`, uvular `[UVU]`, uvofaríngeo `[UVUPHAR]` e faríngeo `[PHAR]`.

Já o CD, divide-se em fechado `[CLO]`, crítico `[CRIT]`, estreito `[NAR]`, largo `[WIDE]` e vocálico `[V]`.

O parâmetro *constrição* é quantitativamente especificado em valores dinâmicos, num ficheiro independente (vd. anexo A), de acordo com os campos enumerados a seguir:

- *target*, que corresponde à posição de equilíbrio da variável do tracto e é definido em milímetros (CD) e em graus (CL);
- *alpha*, um índice numérico que traduz o peso do gesto, quando este se sobrepõe a um outro gesto pertencente à mesma variável do tracto (*blending*). Quanto mais alto o valor, maior a contribuição do gesto em causa.
- *art_wts*, um indicador, fundamentado em dados reais, do envolvimento (ou peso) de cada articulador na produção de uma determinada constrição. Os valores das variáveis são distribuídos percentualmente entre os verdadeiros articuladores (e.g. LX- protrusão labial; JA- ângulo da mandíbula, UH - movimento do lábio superior, LH - movimento do lábio inferior, etc.), no sentido de gerar a trajectória dos corpos físicos concretos. A execução de um gesto labial, por exemplo, implica a mobilização não só dos lábios, mas também da mandíbula. Considera-se, no entanto, estes participam na tarefa em diferentes proporções: quanto maior o valor associado a um determinado articulador, menor o seu movimento no sentido da execução da constrição em causa (vd. anexo A). Os pontos indicam tão somente que um determinado articulador não foi activado.

Adicionalmente a estes parâmetros, são definidos outros valores por defeito, mais directamente relacionados com a equação dinâmica simples, do tipo massa-mola, que modela os gestos. São eles a *frequência natural da variável do tracto* (*TV frequency*), estipulada em 4 HZ para todos os gestos vocálicos (associados ao oscilador v) e em 8 Hz para todos os outros, e o *amortecimento do sistema* (*damping ratio*), cujo valor é 1.

Excepções a estes valores podem ser directamente assinaladas no dicionário de segmentos, que inclui, para além dos tópicos já referidos - *organ*, *osc*, *TV* e *constr* - campos para o *target* e o

[2] Lembramos, uma vez mais, que as abreviaturas usadas se referem à terminologia inglesa.

stiffness [3]. Esta possibilidade é exemplificada na tabela 4.1, a propósito do [Z], onde o valor do *target* do TTCD - definido como `[CRIT]`, o que corresponde a um *target* de 1 mm (vd. anexo A) - foi alterado para 0.16 mm, de modo a gerar mais fricção.

O *tipo de oscilador* (*osc*) identifica o oscilador associado ao gesto e é representativo do comportamento deste em termos de coordenação temporal. As várias classes de osciladores são especificadas logo após a identificação dos articuladores envolvidos na produção, agrupados sob a designação geral de *organ*. Às oclusivas, fricativas e glides correspondem, respectivamente, os osciladores *clo*, *crt*, *nar*, associados, quase sempre (mas não obrigatoriamente), a um oscilador de *release* (*rel*). A segunda articulação das líquidas é representada por um oscilador de tipo *voc*, enquanto os gestos da glote e do véu palatino implicam os osciladores *h* e *n*, respectivamente. Já as vogais estão associadas a uma constrição do corpo da língua (oscilador *v*) e outra dos lábios (oscilador *v_round*).

A composição gestual de Ataques e Codas complexas é definida como a combinatória dos gestos que compõem cada um dos segmentos da sequência. A possibilidade de um destes gestos desaparecer, em virtude da posição ocupada na estrutura da sílaba, é contemplada num conjunto de regras de excepção, que determinam o gesto a eliminar, bem como a TV subjacente [4]. Paralelamente, este tipo de regras permite estipular novos valores para os parâmetros constantes do dicionários de segmentos.

4.1.2 Coordenação intergestual

Durante a geração dos *coupling graph*, a informação sobre o tipo de oscilador correspondente a uma determinada constrição é usada na determinação automática do tipo de sincronização entre os gestos. Após o cálculo dos *parâmetros de cada oscilador*, que determinam os intervalos de activação dos respectivos gestos, são especificados, numa segunda fase, os *padrões de coordenação entre gestos* consecutivos.

A explicitação dos parâmetros do ciclo oscilatório tem como referência a equação dinâmica enunciada a seguir:

$$\ddot{x}_{I,i} = -\alpha_i \dot{x} + \beta x_i \dot{x}_i + \gamma \dot{x}_i^3 + \omega_{0i}^2 x_i \tag{4.1}$$

Os campos considerados incluem:

[3] Tal como referido no capítulo 2 (ponto 2.2.1), o *stiffness* determina as características temporais dos gestos. Segundo Roon *et alii* (2007, p.409), "Stiffness is a measurement of articulator movement that characterizes speed independent of its displacement (...). In the motor control literature, it is an abstract control parameter with a complex of consequences in the time-space behavior of the system. For an intuitive idea of what stiffness is, imagine two springs alike in all aspects other than the material they are made of. If each spring is extended the same distance, the one that returns to its resting position faster has higher stiffness".

[4] Um dos exemplos destas regras é aquela que actua sobre o já referido (capítulo 2, ponto 2.2.2) grupo consonântico /sp/, em que o gesto de abertura glotal do /p/ é apagado.

- *NatFreq*, definida como a frequência natural do oscilador ω_0 , expressa em Hz. O algoritmo automático fixa em 6 Hz a frequência natural dos osciladores associados a gestos consonânticos, enquanto os osciladores vocálicos assumem o valor por defeito de 3 Hz;

- *m:n*, que corresponde à frequência do oscilador expressa em valores inteiros, usada no cálculo da razão entre *m* e *n* de qualquer par de osciladores, que, por sua vez, está na base da relação de fase entre ambos. Tendo em conta as referidas diferenças de *NatFreq* entre osciladores vocálicos e consonânticos, os valores correspondentes para *m:n* são 1 e 2, respectivamente;

- *escap*, parâmetro utilizado no cálculo dos coeficientes α, β e γ do ciclo do oscilador;

- *amp_init*, que corresponde à amplitude em t_0 e é sempre igual a 1;

- *phase-init*, que indica a fase do oscilador em t_0, cujo valor é aleatório.

Os restantes campos (*riseramp, plateau, fallramp*) especificam, em graus, as fases de activação e desactivação dos gestos. De um modo geral, os valores indicados por defeito mostram que os gestos V permanecem activos durante uma porção maior do seu ciclo oscilatório do que os gestos C, a que correspondem intervalos de activação mais curtos. Por sua vez, os gestos *clo* estão activos durante mais tempo do que as respectivas *rel*. Finalmente, a posição na sílaba - Ataque ou Coda - determina intervalos de activação distintos para os osciladores, sendo que em Coda estes são menores do que em Ataque.

Quanto ao tipo de coordenação estabelecida entre dois gestos consecutivos, são considerados três tipos de relações intergestuais, seguindo a proposta inicial de Browman & Goldstein (1986, 1989, 1990b), desenvolvida *a posteriori* por Gafos (2002): 1) coordenação entre gestos pertencentes à mesma sílaba; 2) interligações entre osciladores de sílabas distintas; 3) relações de gestos em fronteira de palavra. As várias possibilidades de coordenação entre os osciladores associados aos gestos, contempladas no TADA, encontram-se no anexo B.

Tomando como entrada um determinado *coupling graph*, o sistema TADA é capaz de calcular automaticamente os intervalos de activação (entrada para o modelo *task-dynamics*) de cada um dos gestos presentes num dado enunciado, conforme ilustrado na figura 4.1.

As linhas coloridas simbolizam as relações de sincronismo estabelecidas entre gestos consecutivos: o verde representa uma coordenação em fase (0 graus); o vermelho está associado a coordenações desfasadas 180 graus; e o amarelo refere-se a todas as outras coordenações possíveis.

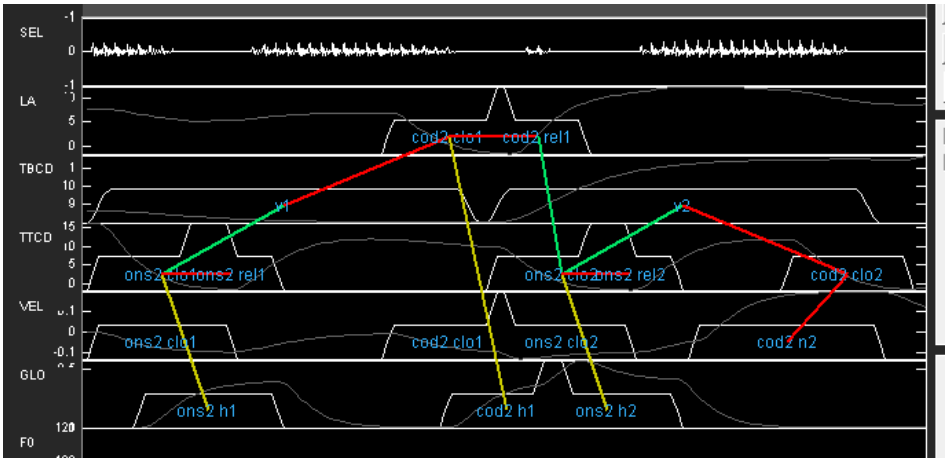


Figura 4.1: Pauta gestual e *coupling graph* da sequência *tip ten*, geradas automaticamente pelo TADA (fonte: Browman *et alii*, 2001-2006). Encontram-se representados o sinal acústico (linha superior) e pauta gestual (linhas 2 a 5). Cada rectângulo corresponde ao intervalo de activação de uma determinada variável do tracto (LA, TBCD, TTCD, VEL, GLO). As linhas coloridas simbolizam as coordenações entre os gestos.

4.2 Caracterização gestual dos sons do português europeu

4.2.1 Metodologia

A presente secção responde ao objectivo - formulado na Introdução a este capítulo - de apresentar uma primeira proposta de caracterização gestual dos sons do PE, com vista à sua integração na infra-estrutura do TADA. O mesmo é dizer que construímos um novo dicionário gestual para o PE, que especifica a composição gestual de cada fonema da língua.

Cada um dos fones do PE - começando pelas vogais e passando pelas várias classes de consoantes - será associado a um conjunto de gestos articulatórios, especificados para as variáveis do tracto nas suas duas dimensões (CL e CD).

Cada uma das configurações gestuais aqui apresentada, toma sempre, como ponto de partida, as descrições articulatórias disponíveis na literatura.

Estas serão, depois, confrontadas com os dados articulatórios (RM) de que dispomos, que servirão ao mesmo tempo de suporte à nossa proposta, apresentada no final da secção e resumida sob a forma de tabela.

Todas as medições, tendo em vista a obtenção de parâmetros articulatórios dinâmicos, correspondentes ao local (CL) e grau de constrição (CD), foram realizadas sobre os perfis articulatórios do informante AND. Esta decisão funda-se na proximidade entre as dimensões do tracto vocal deste falante e as fixadas para o modelo articulatório. As restantes imagens apresentadas servem, sobretudo, como ponto de comparação para aferir e validar os locais de máxima constrição.

Os contornos apresentados foram obtidos semi-automaticamente, através do método *Live-*

Wire (Martins *et alii*, 2008a).

No sentido de estimar o CD e o CL, foi automaticamente determinado o ponto de maior aproximação entre o contorno superior e inferior. O CD corresponde à distância em milímetros (mm) e o CL é o ângulo formado pelo segmento de recta que une os dois pontos (em graus).

Cumpre, no entanto, notar que os valores apresentados são meramente indicativos e "such analyses can only provide approximations to the gestural specification since gestures are comparatively abstract - they are not the articulatory movements themselves, but rather the functions underlying the observed movements." (Browman & Goldstein, 1990a, p.306).

No que respeita à coordenação temporal entre os gestos, na mais absoluta ausência de dados, optámos por manter - salvo pequenas alterações apontadas oportunamente - os princípios de composição gestual (vd. anexo B) previstos para o inglês americano. Até ao momento, os resultados desta opção têm-se revelado satisfatórios, pelo menos no tratamento de estruturas silábicas simples (CV e VC), o que não invalida a possibilidade de revisões futuras, caso tal se nos afigure necessário e novos dados sobre esta matéria se tornem disponíveis.

4.2.2 Dados de ressonância magnética

A descrição articulatória do PE assenta, essencialmente, na observação de imagens RM (2D estáticas) do tracto vocal, no plano sagital, durante a produção dos vários sons do PE.

Sempre que necessário será feito apelo a outro tipo de dados que possam, de algum modo, contribuir para o refinamento da proposta em causa. Face a situações que envolvem articulações complexas, em que se verifica a mais absoluta ausência de dados recentes e seguros, foi, muitas vezes, a configuração do inglês que nos serviu de base, legitimada quer pela observação impressionista dos contornos simulados com recurso ao SAPWindows, quer pela apreciação informal e iterativa da qualidade do som gerado pelos sintetizadores (SAPWindows e HLsyn), a partir de um primeiro conjunto de gestos.

As imagens de RM que nos servem de referência foram adquiridas em duas sessões distintas, ao longo de dois anos (2006-2007), no Sector de Ressonância Magnética do Serviço de Imagiologia dos Hospitais da Universidade de Coimbra [5].

No total, estão disponíveis dados articulatórios relativos a três informantes, dois do sexo masculino (AND e JHM) e um do sexo feminino (RAQ), sendo que nem todos adquiriram a totalidade do *corpus*. Os resultados obtidos para o sujeito AND (o primeiro a adquirir o *corpus*) foram já alvo de análise e discussão num estudo sobre a produção dos vários sons do PE, desenvolvido com o mesmo

[5] Pormenores sobre o protocolo de aquisição das imagens de RM podem ser encontradas em Martins (2007).

propósito geral de contribuir para o aperfeiçoamento e evolução do sintetizador de voz SAPWindows (Martins, 2007).

O *corpus* adquirido por meio de RM foi seleccionado de modo a integrar quase todos os sons do PE, à excepção das vibrantes, obedecendo a limitações temporais e critérios de ordem técnica. Convém, no entanto, ressalvar que este foi desenhado de modo a cumprir objectivos mais gerais, que não os especificamente delineados para este estudo, o que justifica, por exemplo, a inclusão de contextos para avaliar efeitos coarticulatórios, cuja análise sai claramente fora do âmbito deste trabalho.

Para além disso, da primeira (informante AND) para a segunda sessão (informantes RAQ e JHM), o *corpus* foi sujeito a alguns ajustes de pormenor, conforme se pode depreender da análise das tabelas 4.2 e 4.3.

Conforme especificado nessas mesmas tabelas, o material linguístico seleccionado subdivide-se em dois sub-*corpora*: *corpus* 2D sagital e *corpus* 3D. A informação proveniente dos dois *corpora* é estática, i.e, resulta da produção de sons de forma sustentada ou da manutenção artificial de uma articulação durante o período de aquisição.

Para efeitos de análise, no seio desta dissertação, interessa-nos sobretudo o *corpus* 2D, adquirido com todos os informantes.

Tabela 4.2: Corpus MRI (Parte I): vogais orais e nasais. Apresenta-se informação sobre as vogais adquiridas, a palavra de referência (em formato ortográfico e fonético) a partir da qual estas foram elicitadas, e o tipo de sub-*corpora* (2D ou 3D) adquirido por cada um dos informantes (AND, JHM e RAQ).

Fone	Palavra	Transcrição	AND	JHM	RAQ
Vogais Orais:					
[i]	pipo	[pipu]	2D,3D	2D	2D,3D
[e]	pêca	[pekɐ]	2D,3D	2D	2D,3D
[ɛ]	leva	[lɛvɐ]	2D	2D	2D,3D
[ɨ]	devi	[dɨvi]	2D	2D	2D,3D
[ɐ]	cada	[kɐdɐ]	2D,3D	2D	2D,3D
[a]	pato	[patu]	2D,3D	2D	2D,3D
[u]	buda	[budɐ]	2D,3D	2D	2D,3D
[o]	tôpo	[topu]	2D,3D	2D	2D,3D
[ɔ]	pote	[pɔtɨ]	2D,3D	2D	2D,3D
Vogais Nasais:					
[ĩ]	pinta	[pĩtɐ]	2D,3D	2D	2D,3D
[ẽ]	pente	[pẽtɨ]	2D,3D	2D	2D,3D
[ɐ̃]	canto	[kɐ̃tu]	2D,3D	2D	2D,3D
[ũ]	punto	[pũtu]	2D,3D	2D	2D,3D
[õ]	ponte	[põtɨ]	2D,3D	2D	2D,3D

Tabela 4.3: Corpus MRI (Parte II): consoantes. Apresenta-se informação sobre as consoantes adquiridas, a palavra ou logátomo (em formato ortográfico e fonético) de referência, e o tipo de sub-*corpora* (2D ou 3D) adquirido por cada um dos informantes (AND, JHM e RAQ).

Fones	Palavras/Logátomos	Transcrição	AND	JHM	RAQ
		Oclusivas:			
[p]	[apa], [ipi], [upu]		2D	2D	2D,3D
[t]	[ata], [iti], [utu]		2D	2D	2D,3D
[k]	[aka], [iki], [uku]		2D	2D	2D,3D
[b]	[aba], [ibi], [ubu]		2D	2D	2D,3D
[d]	[ada], [idi], [udu]		2D	2D	2D,3D
[g]	[aga], [igi], [ugu]		2D	2D	2D,3D
		Nasais:			
[m]	cama	[kɐmɐ]	2D,3D	2D	2D,3D
[n]	cana	[kɐnɐ]	2D,3D	2D	2D,3D
[ɲ]	canha	[kɐɲɐ]	2D,3D	2D	2D,3D
[m]	[ama], [imi], [umu]			2D	2D,3D
[n]	[ana], [ini], [unu]			2D	2D,3D
[ɲ]	[aɲa], [iɲi], [uɲu]			2D	2D,3D
		Fricativas:			
[f]	fala	[falɐ]	2D	2D	2D,3D
[s]	sala	[salɐ]	2D	2D	2D,3D
[ʃ]	chá	[ʃa]	2D	2D	2D,3D
[v]	vaca	[vakɐ]	2D	2D	2D,3D
[z]	zarpa	[zarpɐ]	2D	2D	2D,3D
[ʒ]	jacto	[ʒatu]	2D	2D	2D,3D
[f]	[afa], [ifi], [ufu]		2D,3D	2D	2D,3D
[s]	[asa], [isi], [usu]		2D,3D	2D	2D,3D
[ʃ]	[aʃa], [iʃi], [uʃu]		2D,3D	2D	2D,3D
[v]	[ava], [ivi], [uvu]		2D	2D	2D,3D
[z]	[aza], [izi], [uzu]		2D	2D	2D,3D
[ʒ]	[aʒa], [iʒi], [uʒu]		2D	2D	2D,3D
		Laterais:			
[l]	laço	[lasu]	2D,3D		
[l]	pála	[palɐ]	3D	2D	2D,3D
[ɫ]	mal	[maɫ]	2D,3D	2D	2D,3D
[ʎ]	falha	[faʎɐ]	2D	2D	2D,3D
[ʎ]	palha	[paʎɐ]		2D 2D	2D,3D
[l]	[ala], [ili], [ulu]			2D	2D,3D
[ʎ]	[aʎa], [iʎi], [uʎu]			2D	2D,3D

4.2.3 Inventário fonémico

Do inventário de sons considerado neste estudo fazem parte todas as formas fonéticas características do dialecto padrão do português europeu, estando, portanto, excluídas todas as restantes variações

Tabela 4.4: Quadro geral da classificação tradicional das vogais orais do PE.

	Anteriores	Centrais	Posteriores
Fechadas	[i]	[ɨ]	[u]
Semi-Fechadas	[e]	[ɐ]	[o]
Semi-Abertas	[ɛ]		[ɔ]
Abertas		[a]	

dialectais.

Ficam também afastados desta descrição os segmentos habitualmente designados como *glides*, *semivogais* ou *semiconsoantes*. Na bibliografia portuguesa, estes são normalmente considerados realizações contextualmente determinadas de "vogais plenas", não fazendo, portanto, parte do inventário fonémico do português (Mateus, 1975). Independentemente do seu estatuto fonológico, as semivogais nunca ocorrem sozinhas, mas formam, juntamente com as vogais que as precedem, os chamados *ditongos decrescentes*. O modelamento linguístico dos ditongos, no quadro teórico da FA, levanta problemas de coordenação específicos (cf. Marin, 2005, 2007), que está fora do alcance desta dissertação abordar, o que justifica, por sua vez, as omissões relativamente às semivogais.

4.2.4 Gestos vocálicos

As condições de escoamento do ar através do tracto vocal permitem distinguir três grandes classes de sons: as consoantes, as vogais e as glides. No tocante às vogais, estas são produzidas sem constrições significativas à passagem do fluxo do ar e com vibração das pregas vocais, sendo, por isso, considerados sons vozeados ou sonoros. De entre todos os articuladores implicados na sua produção, destaca-se o papel do véu palatino - responsável pela distinção entre vogais orais e nasais - dos lábios e do dorso da língua. É sobretudo com base na posição deste último articulador que se define a classificação fonética tradicional das vogais. Assim, em função do movimento da língua no eixo antero-posterior, as vogais são divididas em anteriores (ou palatais), centrais e posteriores (também designadas de recuadas ou velares). A altura do dorso da língua em conjunto com a abertura do maxilar inferior determina o seu grau de abertura e permite classificá-las nas seguintes categorias: abertas, semiabertas, semifechadas e fechadas.

Do sistema vocálico do PE (norma-padrão) fazem parte nove vogais orais: [i], [e], [ɛ], [a], [ɐ], [ɨ], [ɔ], [o], [u]. No quadro 4.4, é apresentada uma síntese da sua classificação articulatória, de acordo com os critérios supra-mencionados.

A categorização das vogais orais apresentada segue, no essencial, a classificação vulgarmente encontrada nas publicações de fonética portuguesa (Mateus *et alii*, 1990, 2005; Cruz-Ferreira, 1999a; Moutinho, 2000; Veloso, 1999; Cunha & Cintra, 1997), que, pesem embora algumas variações, se funda nos dois critérios enunciados. Estes aproximam-se dos parâmetros contemplados no

AFI, onde as vogais são representadas num quadrilátero que visa reflectir o espaço articulatório possível em termos da posição longitudinal da língua, em abcissa, e do grau de abertura/ fechamento, em ordenada. O parâmetro arredondamento labial - que dá conta do papel dos lábios na produção da vogal, e que, no português, se encontra especificado exclusivamente para as vogais posteriores - é igualmente considerado na disposição gráfica adoptada no AFI: os símbolos fonéticos ocorrentes à esquerda das linhas do quadrilátero referem-se a vogais não-arredondadas, enquanto os símbolos à direita transcrevem vogais arredondadas.

No mais rigoroso respeito pela doutrina geral do AFI, Veloso (1999) insere a vogal [a] no eixo das vogais anteriores. Optámos, contudo, neste trabalho, por adoptar a classificação tradicional, que considera esta vogal como central (Mateus *et alii*, 1990, 2005; Cruz-Ferreira, 1995, 1999a; Moutinho, 2000; Emiliano, 2006; Barbosa, 1994a; Barroso, 1999).

Também a vogal [ɨ] levanta alguns problemas de classificação articulatória - motivados, em larga medida, pela ausência de estudos articulatórios sobre a matéria - o que se reflecte, naturalmente, no tipo de notação usada para transcrever este segmento. A maioria dos autores descreve esta vogal como central, fechada, alta e não-arredondada (Barbosa, 1994a; Barroso, 1999; Cunha & Cintra, 1997). O símbolo que lhe corresponderia no AFI seria então o [ɨ], adoptado por autores como Veloso (1999), Barroso (1999), Andrade & Viana (1996), Mateus *et alii* (2005), Mateus & d'Andrade (2000). Outros há (Viana, 1973a; Lacerda & Hammarström, 1952; Cunha & Cintra, 1997; Mateus *et alii*, 1990; Barbosa, 1994a; Mateus, 1975; d'Andrade, 1977; Barbosa, 1965), que apesar de aceitarem esta descrição como a mais adequada, optam, implícita ou explicitamente, pelo [ə], símbolo usado, em diversas línguas, para representar o *schwa*. A referida classificação é questionada por Cruz-Ferreira (1995, 1999a), que, com base em dados acústicos, considera a vogal "a fronted and lowered high back unrounded vowel" (Cruz-Ferreira, 1999a, p.127) e usa o símbolo [ɯ]. A subscrição de uma ou outra proposta fica, por agora, adiada, pelo menos até que um olhar mais atento sobre os dados articulatórios nos permita uma tomada de posição mais fundamentada.

No quadro da FA, o tratamento das vogais implica a referência a quatro variáveis do tracto: velo (VEL), glote (GLO), lábios (L) e corpo da língua (TB). Segundo este modelo teórico, por defeito, o véu palatino encontra-se levantado contra a parede posterior da faringe e a glote apresenta-se fechada, pelo que, na caracterização gestual das vogais, enquanto sons orais e vozeados, não há necessidade de especificar estes dois parâmetros. Quanto ao corpo da língua, os graus de constrição reconhecidos pela FA para os gestos vocálicos são: `[estreito]` (`[narrow]`), `[médio]` (`[mid]`), `[largo]` (`[wide]`). No âmbito do TADA, estes são substituídos por um descritor geral, designado simplesmente de "vocálico" (`[V]`), exclusivo para as vogais, e associado a um oscilador "v". As tradicionais quatro alturas (ou aberturas) vocálicas são depois obtidas à custa da especificação do *target* em valores articulatórios reais, definidos em milímetros. Para que se compreendam cabalmente as descrições gestuais dos segmentos vocálicos, apresentadas a seguir, lembramos aqui (vd. anexo A) os pontos de constrição, e respectivos ângulos, referentes ao corpo da língua: `[palatal]` (95 graus), `[velar]` (100 graus), `[uvular]` (125 graus), `[uvo-faríngeo]` (150 graus) e `[faríngeo]`

(180 graus). Para além disso, como já vimos, as vogais podem manifestar projecção labial. Nestes casos, a variável do tracto dorso da língua combina-se com os lábios, representados, no caso das vogais, pelo oscilador "v_rnd". Ambos os gestos que compõem a vogal - gesto de protrusão labial e o gesto do dorso da língua - têm início no mesmo ponto do tempo (vd. anexo B).

4.2.4.1 Gestos para as vogais posteriores

No sentido de determinar o local de máxima constrição das vogais tradicionalmente designadas por posteriores, apresentam-se em seguida (figura 4.2) os perfis articulatórios dos informantes AND, JHM e RAQ, referentes à produção das vogais [u], [o] e [ɔ], em versão sustentada, induzida a partir de uma palavra de referência (neste caso "buda", "topo" e "pote").

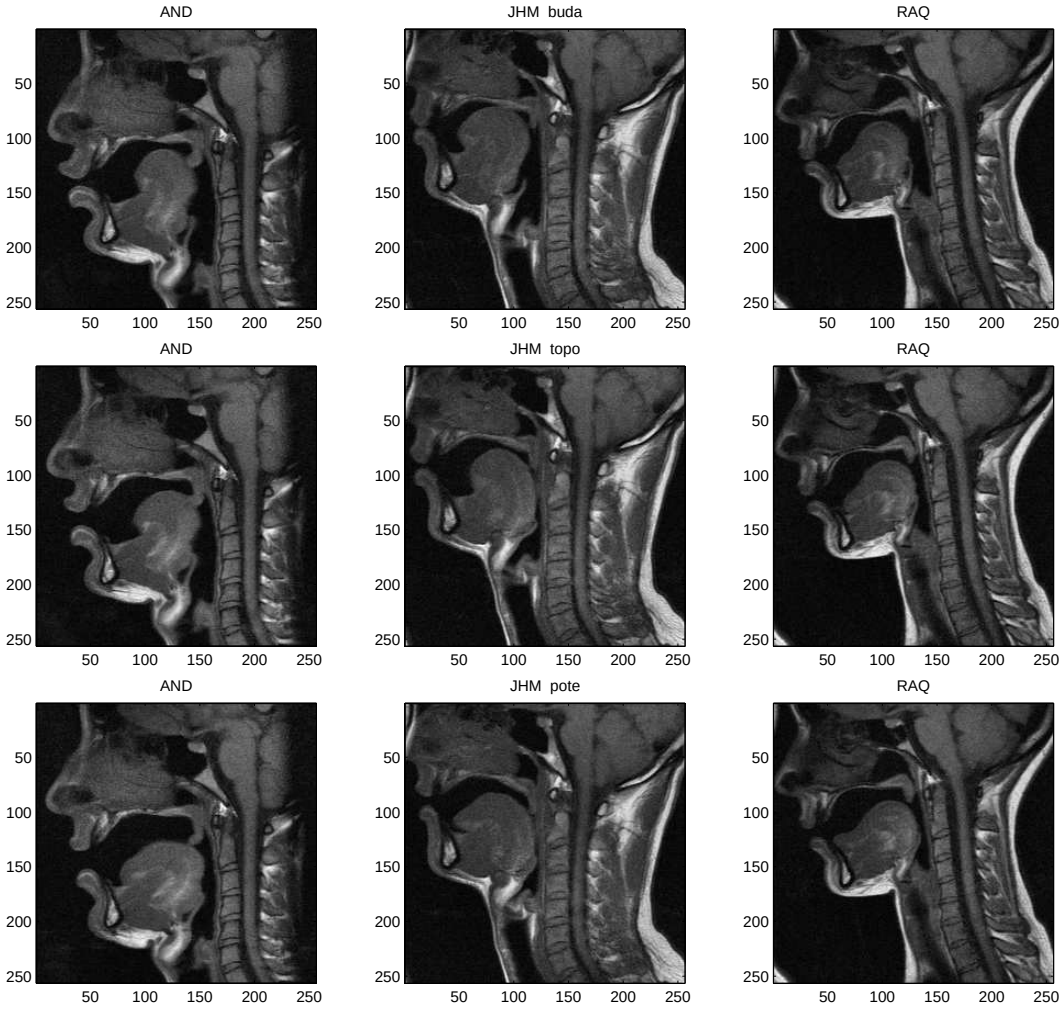


Figura 4.2: Perfis articulatórios dos sujeitos AND, JHM e RAQ, durante a produção das vogais [u], [o] e [ɔ], a partir das palavras de referência "buda" (em cima), "topo" (ao centro) e "pote" (em baixo).

De uma forma geral, a observação informal das imagens permite determinar, sem surpresas,

uma zona de articulação posterior para as vogais em estudo, conforme previsto na descrição articulatória tradicional e ilustrado por meio de diagramas articulatórios (Barroso, 1999) e radiografias avulsas (Barbosa, 1994a; Martins, 1977).

Tal como foi já assinalado por Barbosa (1994a), no caso do [u], a parte posterior do dorso da língua encontra-se numa posição elevada e aproxima-se do palato mole, enquanto que para as vogais [o] e [ɔ] o ponto de maior constrição se situa entre a língua e a região uvo-faríngea. Consequentemente, o volume da cavidade faríngea vai diminuindo de acordo com a seguinte proporção [u]>[o]>[ɔ], ao passo que a abertura da cavidade oral aumenta, segundo a mesma ordem.

No restante, são ainda de salientar as diferenças ao nível da abertura e protrusão labiais: a uma maior projecção labial, visível para as vogais [u] e [o], corresponde, por sua vez, uma menor abertura, enquanto que em relação ao [ɔ] se verifica exactamente o inverso.

O ponto de máxima constrição, intuído a partir das imagens anteriores, só foi determinado com exactidão a partir das medidas efectuadas sobre o perfil articulatório do sujeito AND (figura 4.3).

Em se tratando de vogais arredondadas, houve ainda necessidade de aferir valores de protrusão e abertura labiais associados a cada vogal.

Do ponto de vista gestual, todas as vogais posteriores se caracterizam pela acção combinada de dois articuladores: o dorso da língua e os lábios, representados, respectivamente, pelos osciladores "v" e "v_rnd".

Quanto à variável de corpo da língua, considerada nas suas duas dimensões (ponto e grau de constrição), a vogal [u] caracteriza-se por um ponto de constrição velar ([VEL]) - especificado nos 110 graus, na medida em que a zona de articulação é um pouco mais posterior (vd. figura 4.3, imagem à esquerda) - sendo que a distância entre a língua e o palato mole se fixou nos 5 mm. A vogal [o] é definida por uma constrição uvo-faríngea ([UVOPHAR]), localizada nos 140 graus, e um valor de TBCD vocálico de 9 mm (vd. figura 4.3, imagem ao centro). A vogal [ɔ] apresenta uma constrição claramente faríngea ([PHAR]), cujo *target* foi também alterado directamente no dicionário de segmentos para os 170 graus, e uma abertura faríngea de 7 mm (vd. figura 4.3, imagem à direita).

As vogais em análise distinguem-se entre si também em relação ao valores de abertura (LA) e protrusão (LP) labiais. Relativamente ao primeiro parâmetro, a maior abertura regista-se para a vogal [ɔ], logo seguida do [o] e, finalmente, do [u], pelo que se propõem os valores de 5 mm, 4 mm e 2 mm, para cada uma das vogais, respectivamente. Inversamente, a protrusão labial é superior no [u] e no [o] (12 mm), quando comparada com o [ɔ] (8 mm).

A informação gestual relativa às vogais posteriores foi compilada na tabela 4.5, respondendo assim, às exigência formais do modelo computacional.

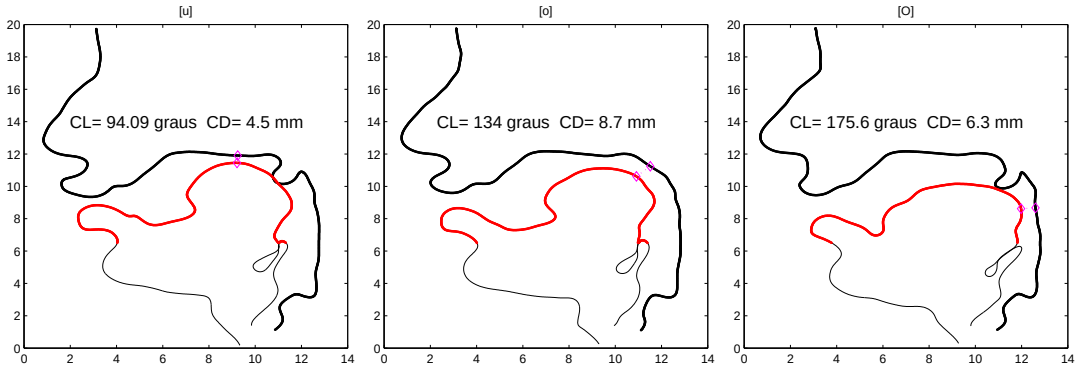


Figura 4.3: Medidas de *constriction location* (CL), em graus, e *constriction degree* (CD), em milímetros (mm), para as vogais [u], [o] e [ɔ], efectuadas a partir dos contornos sagitais do informante AND.

Tabela 4.5: Gestos associados às vogais posteriores do PE. As vogais encontram-se foneticamente representadas em AFI e os gestos a elas associados são caracterizados pelo articulador (*Organ*), oscilador (*Osc*), variável do tracto (*TV*), constrição (*Const*). O *target* da constrição e o *stiffness* encontram-se pré-definidos num ficheiro independente (“.”), mas podem ser directamente especificados no dicionário.

Vogal	Organ	Osc	TV	Const	Target	Stiff
u	TB	v	TBCL	VEL	110	.
	TB	v	TBCD	V	5	.
	Lips	v_rnd	LP	PRO	12	.
	Lips	v_rnd	LA	NAR	2	.
o	TB	v	TBCL	UVUPHAR	140	.
	TB	v	TBCD	V	9	.
	Lips	v_rnd	LP	PRO	12	.
	Lips	v_rnd	LA	NAR	4	.
ɔ	TB	v	TBCL	PHAR	170	.
	TB	v	TBCD	V	7	.
	Lips	v_rnd	LP	PRO	8	.
	Lips	v_rnd	LA	NAR	5	.

4.2.4.2 Gestos para as vogais centrais

Segundo os perfis sagitais dos informantes AND, JHM e RAQ (figura 4.4), nas chamadas vogais centrais, produzidas a partir das palavras de referência “devi”, “cada” e “pato”, a parte central do dorso da língua eleva-se em direcção ao palato, sendo este movimento mais acentuado no [ɨ], do que nas outras duas vogais. Ao mesmo tempo, no [ɐ] e, sobretudo, no [a], a raiz da língua recua - pelo que o ponto de máxima constrição ocorre exactamente na zona da faringe - o que não se verifica em relação ao [ɨ], em relação ao qual é difícil identificar um ponto onde a distância médio-sagital é claramente menor (pelo menos nos informantes AND e RAQ).

Por outro lado, a mandíbula encontra-se visivelmente mais aberta para o [ɐ] e, principalmente, para o [a], tradicionalmente consideradas vogais mais abertas do que o [ɨ].

Havíamos referido, na secção 4.2.4, que a classificação do [a], enquanto vogal central, não é totalmente consensual. Em face dos dados articulatórios RM, podemos observar, a este respeito, que, considerando unicamente o movimento do dorso da língua, esta vogal se encontra efectivamente numa posição central, a par de outras vogais indubitavelmente centrais como o [ɐ] e o [ɨ]. Para sermos mais precisos, a vogal em causa denota mesmo, mais do que qualquer outra vogal central, uma certa tendência para a posteriorização.

Este movimento de recuo é captado pelo quadrilátero vocálico proposto por Cruz-Ferreira (1999a), mas não pelo registo gráfico de Veloso (1999), que no escrupuloso cumprimento das convenções do AFI, mas sem atender aos factos reais da língua, insere a vogal [a] no eixo das vogais anteriores. Lembramos ainda aqui, a este propósito, que não é de todo invulgar encontrar, entre as ilustrações do AFI constantes do *Handbook of the International Phonetic Association* (International Phonetic Association, 1999), situações de línguas (e.g. catalão, galego, hebreu) em que esta vogal é descrita e representada como central, ainda que esta classificação subverta, de algum modo, as categorias previstas no AFI.

Uma análise das medidas articulatórias realizadas a partir do perfil médio-sagital do sujeito AND (vd. figura 4.5) permitirá determinar os parâmetros necessários ao modelo gestual.

Assim, às três vogais em estudo está associado um gesto dorsal, cujo local de constrição (TBCL) foi identificado como `[faríngeo]` para o [a] e para o [ɐ], sendo que o *target* desta última vogal foi rectificado para os 170 graus, já que ocorre mais próximo da úvula.

Estas duas vogais distinguem-se ainda em relação ao grau de constrição (TBCD): em consequência do movimento de posteriorização da raiz da língua, a distância entre esta e a parede faríngea foi quantificado nos 11 mm para o [a], enquanto para o [ɐ], o valor foi fixado nos 15 mm.

Os gestos associados ao [ɨ] foram muito mais complicados de determinar, em virtude da dificuldade em definir um ponto de constrição, capaz de induzir a percepção acústica de um [ɨ].

A realização dos informantes AND e RAQ parece corresponder à descrição fonética tradicional do [ɨ] como vogal fechada central (cf. Andrade, 1996). As imagens do informante JHM mostram uma vogal ainda mais fechada e ligeiramente mais recuada, mas ainda assim mais central do que o [ɐ] e o [a] [6].

De acordo com os resultados da síntese articulatória, o recurso a um gesto palatal (vd. figura 4.5) para representar esta vogal não é, contudo, capaz de captar as diferenças entre o [ɨ] e outras vogais anteriores, nomeadamente o [i] e o [e] e a utilização de um gesto velar - mais recuado - aproxima esta vogal do [u]. Assim, depois de um conjunto de experiências, que implicaram a apreciação informal e iterativa da qualidade do som gerado pelos sintetizadores (SAPWindows e HLsyn), a partir de uma

[6] Os dados RM dos informantes AND e RAQ justificam a utilização do símbolo fonético [ə] para descrever o comportamento fonético da vogal central portuguesa, traduzindo simultaneamente o seu carácter instável, que aproxima esta vogal do português do *schwa* de outras línguas. Se tivermos em conta as imagens do informante JHM, então, a transcrição adoptada pela maioria dos linguistas portugueses ([ɨ]) seria talvez a mais adequada. O [ɯ], usado por Cruz-Ferreira (1995, 1999a), não traduz o carácter central da referida vogal, quando comparada com as outras duas vogais da mesma série ([ɐ] e [a]).

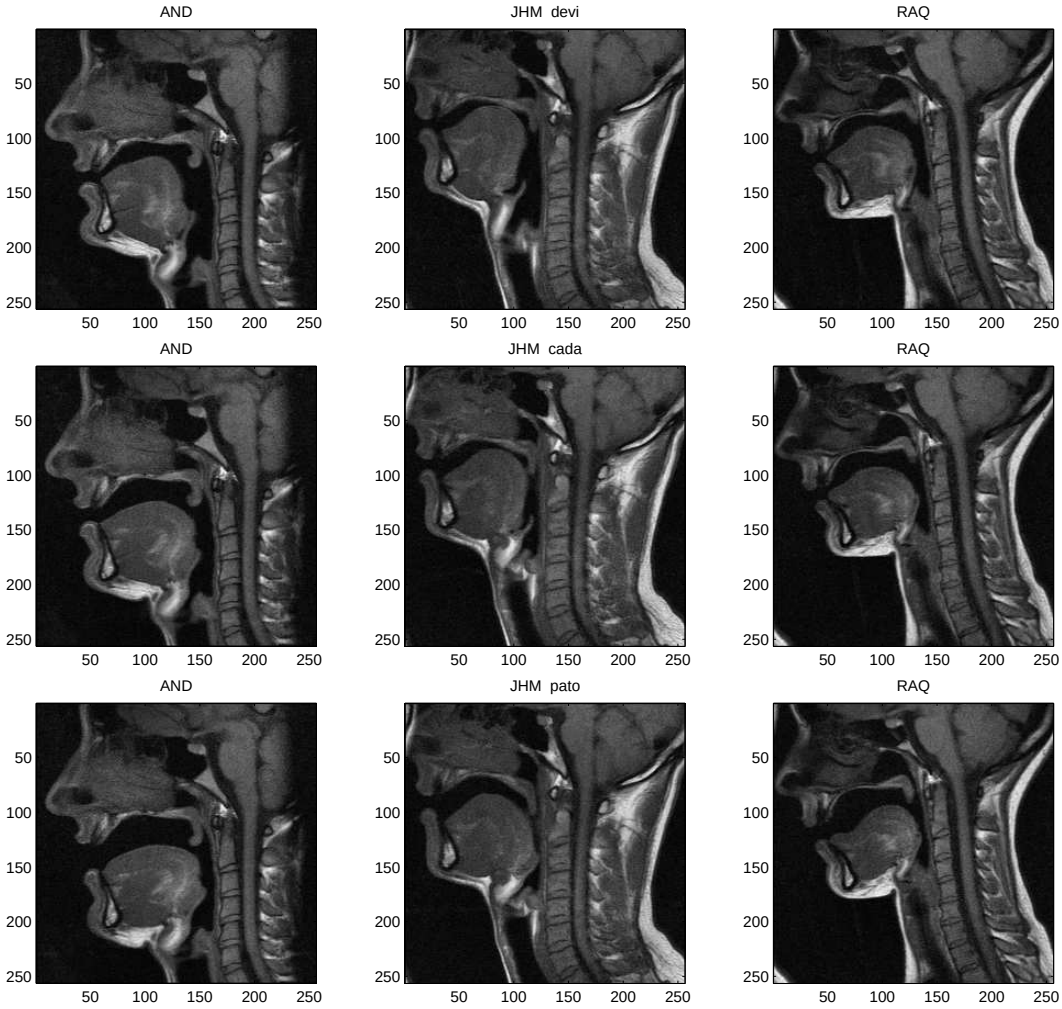


Figura 4.4: Perfis articulatórios dos sujeitos AND, JHM e RAQ, durante a produção das vogais [ɨ], [ɐ] e [a], a partir das palavras de referência "devi" (em cima), "cada" (ao centro) e "pato" (em baixo).

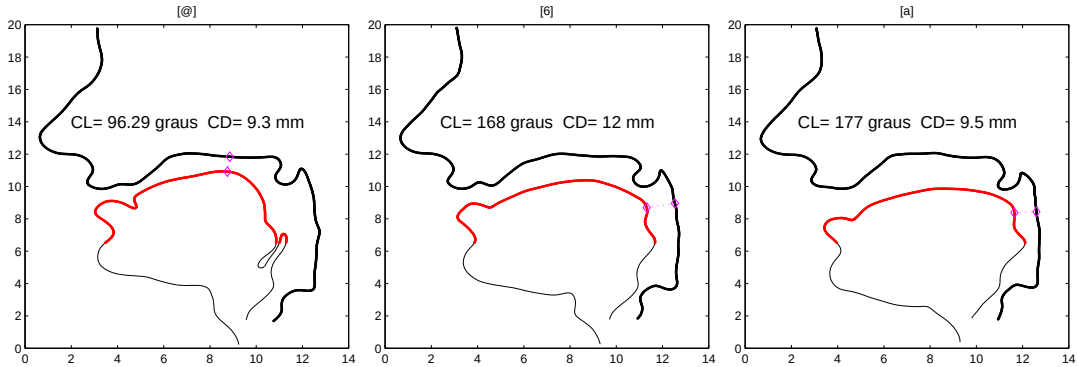


Figura 4.5: Medidas de *constriction location* (CL), em graus, e *constriction degree* (CD), em milímetros (mm), para as vogais [ɨ], [ɐ] e [a], efectuadas a partir dos contornos sagitais do informante AND.

determinada configuração gestual, optámos - tomando como referência a definição original do TADA para o *schwa* do inglês - por um TBCL uvular (UVU), fechado (V=3.5 mm). Perceptualmente, esta configuração é relativamente eficaz e, em termos de produção, traduz o carácter mais central e fechado do [ɨ], em relação às restantes vogais centrais ([ɐ] e [a]).

Dado o carácter iminentemente breve do [ɨ], com forte tendência para a queda, foi ainda necessário aumentar os valores de *stiffness*, o que, como vimos, se traduz numa diminuição da duração do gesto de corpo da língua (vd. tabela 4.6).

Seguindo o formalismo imposto pelo TADA, chegámos à tabela 4.6.

Tabela 4.6: Gestos associados às vogais centrais do PE. As vogais encontram-se foneticamente representadas em AFI e os gestos a elas associados são caracterizados pelo articulador (*Organ*), oscilador (*Osc*), variável do tracto (*TV*), constrição (*Const*). O *target* da constrição e o *stiffness* encontram-se pré-definidos num ficheiro independente ("."), mas podem ser directamente especificados no dicionário.

Vogal	Organ	Osc	TV	Const	Target	Stiff
ɨ	TB	v	TBCL	UVU	.	3
	TB	v	TBCD	V	3.5	3
ɐ	TB	v	TBCL	PHAR	170	.
	TB	v	TBCD	V	15	.
a	TB	v	TBCL	PHAR	.	.
	TB	v	TBCD	V	11	.

4.2.4.3 Gestos para as vogais anteriores

A partir da observação das imagens sagitais relativas às vogais [i], [e] e [ɛ], produzidas pelos três falantes (figura 4.6), conclui-se que, para todas as vogais, o dorso da língua se eleva em direcção ao palato duro. Ao mesmo tempo, a língua movimenta-se ligeiramente para a frente em diferentes proporções, o que vai de encontro ao vinculado na literatura (cf. Barbosa, 1994a): do [ɛ] para o [i], passando pelo [e], a língua sobe e anterioriza-se, diminuindo o volume da cavidade anterior e aumentando o da posterior. O referido movimento encontra-se totalmente reflectido na disposição gráfica adoptada pelo AFI para representar as vogais e adaptada para o PE por Veloso (1999). A arrumação destas três vogais na categoria geral das anteriores (palatais) parece assim plenamente justificada.

As medidas articulatórias relativas ao perfil sagital do informante AND (figura 4.7) permitem confirmar, sem margem para dúvidas, um ponto de máxima constrição, comum às três vogais, localizado na zona palatal.

A informação articulatória veiculada anteriormente, relativamente às vogais [i], [e] e [ɛ], traduz-se, assim, em termos de descrição gestual - vd. tabela 4.7 - num gesto de dorso da língua, especificado como palatal ([PAL]), quanto à variável do tracto local de constrição. No sentido de dar conta dos diferentes graus de anteriorização da língua ([i]>[e]>[ɛ]) referidos anteriormente, o *target*

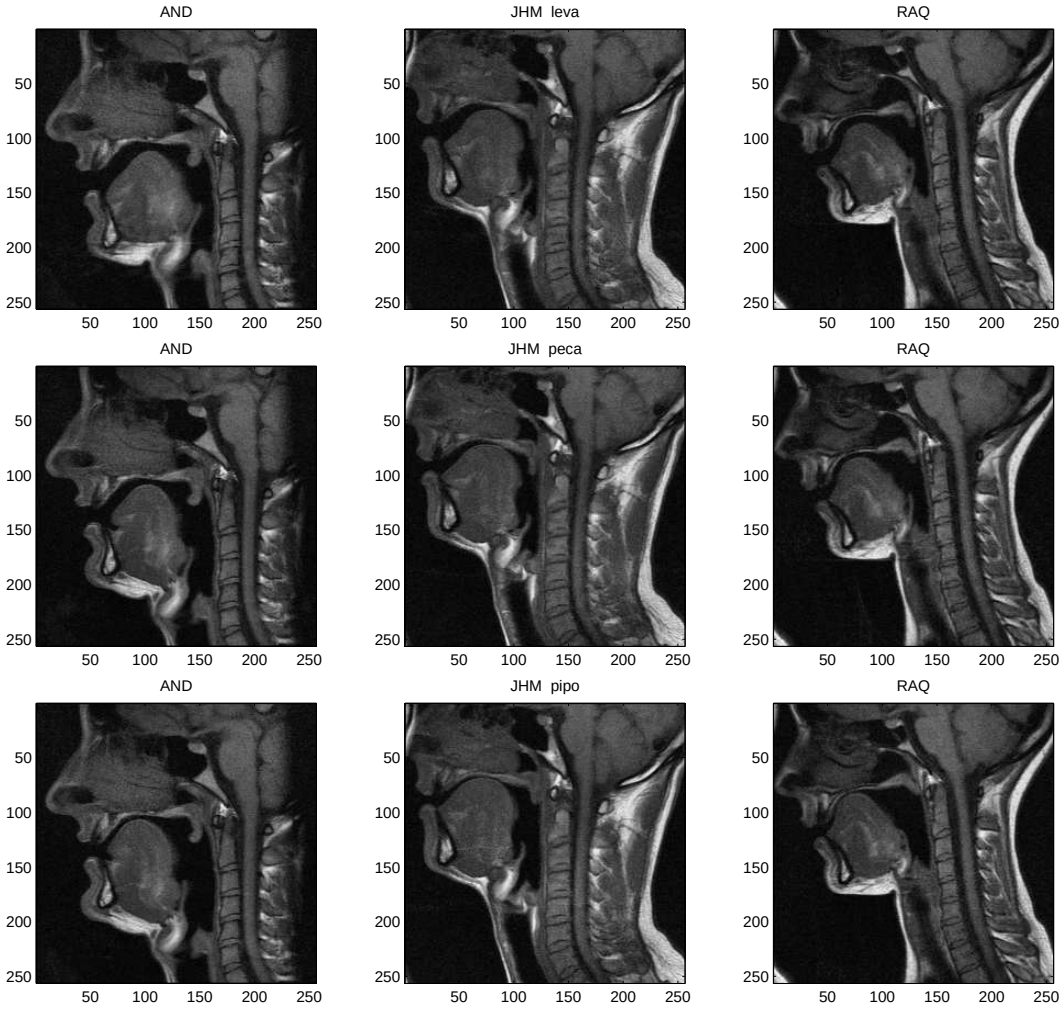


Figura 4.6: Perfis articulatórios dos sujeitos AND, JHM e RAQ, durante a produção das vogais [ɛ], [e] e [i], a partir das palavras de referência “leva” (em cima), “pêca” (ao centro) e “pipo” (em baixo).

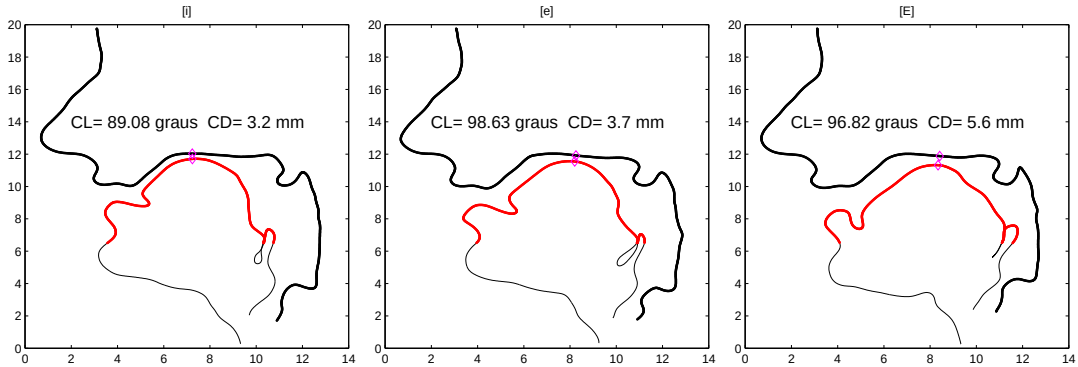


Figura 4.7: Medidas de *constriction location* (CL), em graus, e *constriction degree* (CD), em milímetros (mm), para as vogais [ɛ], [e] e [i], efectuadas a partir dos contornos sagitais do informante AND.

foi ajustado - para os 80 graus, no caso do [i], e para os 90 graus, no caso do [e]. No respeitante ao grau de constrição, definiram-se, a partir das medidas efectuadas sobre as imagens, as seguintes distâncias entre a língua e o palato duro - que confirmam o carácter mais aberto do [ɛ]: 9 mm para o [ɛ], 6 mm para o [e] e 3 mm para o [i].

Tabela 4.7: Gestos associados às vogais anteriores do PE. As vogais encontram-se foneticamente representadas em AFI e os gestos a elas associados são caracterizados pelo articulador (*Organ*), oscilador (*Osc*), variável do tracto (*TV*), constrição (*Const*). O *target* da constrição e o *stiffness* encontram-se pré-definidos num ficheiro independente ("."), mas podem ser directamente especificados no dicionário.

Vogal	Organ	Osc	TV	Const	Target	Stiff
i	TB	v	TBCL	PAL	80	.
	TB	v	TBCD	V	3	.
e	TB	v	TBCL	PAL	90	.
	TB	v	TBCD	V	6	.
ɛ	TB	v	TBCL	PAL	.	.
	TB	v	TBCD	V	9	.

4.2.5 Gestos consonânticos

4.2.5.1 Consoantes oclusivas

As consoantes oclusivas são produzidas mediante a oclusão completa e momentânea do canal bucal, sendo que, no caso das oclusivas orais, também a passagem nasal se encontra fechada. A distensão abrupta da constrição dá origem a um ruído breve, semelhante a uma curta explosão, facto este que está na origem do epíteto *explosivas*, tantas vezes usado para designar esta classe de consoantes.

O sistema consonântico do PE integra três oclusivas surdas - [p], [t], [k] - e três oclusivas sonoras - [b], [d] e [g].

Quanto ao ponto de articulação, o [p] e o [b] são classificadas como bilabiais, na medida em que a interrupção da corrente expiratória tem lugar nos lábios; o [t] e o [d] são descritas como dentais (Lacerda & Hammarström, 1952; Sá Nogueira, 1938; Mateus *et alii*, 1990, 2005; Moutinho, 2000; Cruz-Ferreira, 1999a) ou alveolares (Viana, 1973b; Veloso, 1999), sendo que o ápice/ lâmina da língua toca alternativamente na parte interna dos dentes superiores ou nos alvéolos[7]; o [k] e o [g] são, geralmente, consideradas velares, já que o dorso da língua se encosta à região do palato mole.

Apresentam-se, em seguida, as imagens articulatórias referentes à produção das consoantes oclusivas (figura 4.8 a 4.10), pelos três informantes, no contexto das três vogais cardinais [i], [u] e [a].

[7] Veloso (1994) interpreta as flutuações na atribuição de um ponto de articulação a /t/ e /d/ como indício de uma variação alofónica (relacionada quer com causas individuais, quer com factores contextuais), não pertinente do ponto de vista fonológico e perceptivo. Barbosa (1994a) ilustra esta variação, a maioria das vezes imperceptível ao ouvido, com imagens radiográficas, atribuindo-a expressamente aos hábitos articulatórios individuais de cada falante. Sá Nogueira (1938), por sua vez, parte da análise de palatogramas para concluir que "a superfície tocada da região alveolar pode ser maior ou menor, conforme a energia com que se proferem êsses fonemas [oclusivas /t/ e /d/], e conforme o ponto e o modo de articulação do fonema que se lhes segue no corpo da palavra" (Sá Nogueira, 1938, p.33).

Confrontando os dados RM com a informação anterior, estamos em condições de identificar as marcas articulatórias que individualizam cada um destes segmentos consonânticos.

A averiguar pelas configurações sagitais que se seguem (figura 4.8), a descrição canónica do [p] (e do [b]) como consoante bilabial parece perfeitamente pertinente e adequada. Assim, na tabela gestual 4.8, o [p] e o [b] podem ser representados por um gesto de oclusão labial. No caso do [p], enquanto som não-vozeado, há ainda a acrescentar um gesto glotal de abertura em total sincronismo com o gesto dos lábios.

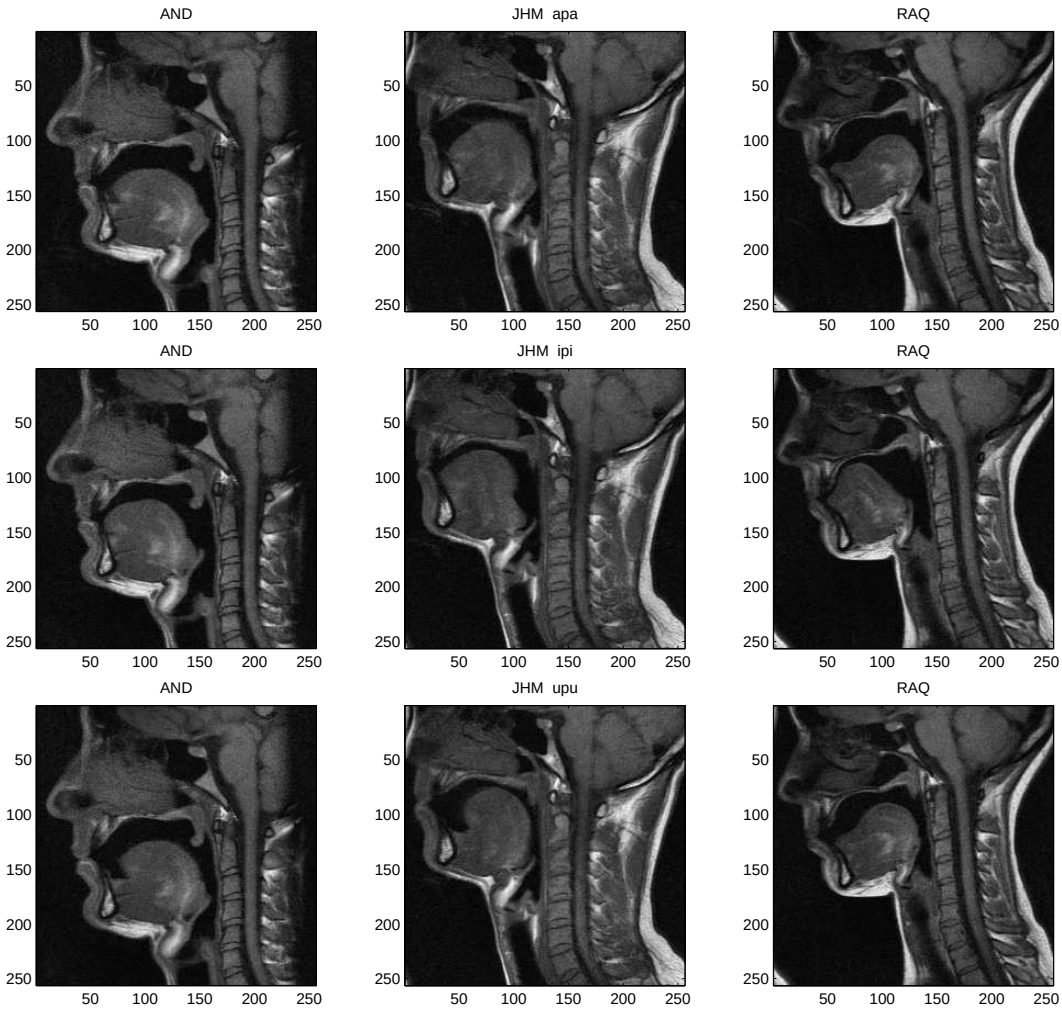


Figura 4.8: Perfis articulatórios dos sujeitos AND, JHM e RAQ, durante a produção da consoante [p], no contexto de [a] (em cima), [i] (ao centro) e [u] (em baixo).

Na produção do [t] e do [d], a ponta/ lâmina da língua aproxima-se da parte interna da arcada dentária superior (figura 4.9). Contudo, e uma vez que o contorno dos dentes não é visível na imagem, a definição exacta do ponto de máxima constrição só se tornará possível através do recurso a outro tipo de métodos instrumentais, nomeadamente a electropalatografia.

Em resultado de uma pequena avaliação auditiva, verificámos, no entanto, que a variação

entre um *target* alveolar ou dental não parece acarretar consigo diferenças acústicas perceptivamente pertinentes. Assim, na representação gestual destas consoantes (tabela 4.8), optámos por considerar um gesto de ponta da língua fechado na região dental - em linha com a maioria das obras de fonética portuguesa (Lacerda & Hammarström, 1952; Sá Nogueira, 1938; Mateus *et alii*, 1990, 2005; Moutinho, 2000; Cruz-Ferreira, 1999a) e conforme sugestionado por, pelo menos, parte das imagens - o que não significa rejeitar a possibilidade de um *target* alveolar, eventualmente mais adequado para alguns falantes ou contextos.

À semelhança do [p], também no caso do [t], foi necessário considerar, a par do gesto de ponta da língua, um gesto de abertura glotal, que tem como consequência o não vozeamento.

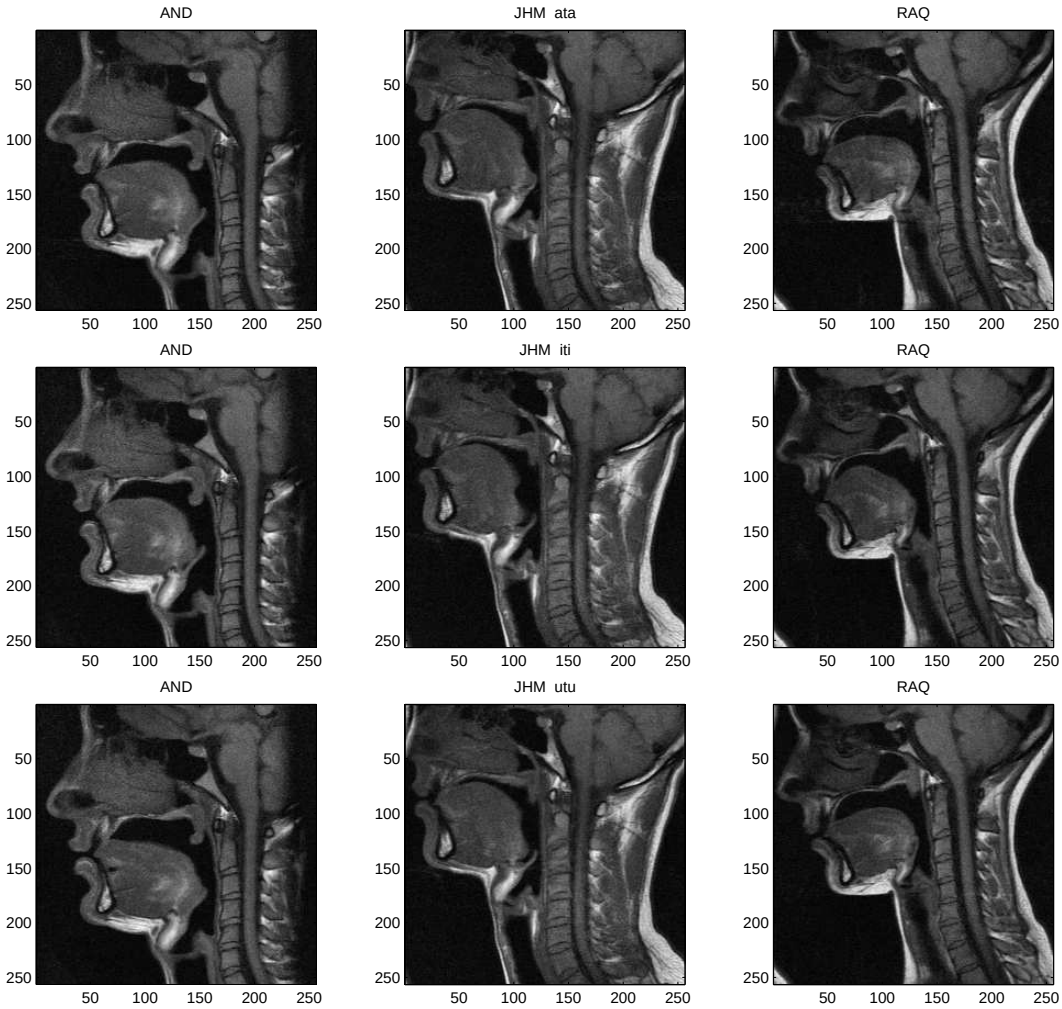


Figura 4.9: Perfis articulatórios dos sujeitos AND, JHM e RAQ, durante a produção da consoante [t], no contexto de [a] (em cima), [i] (ao centro) e [u] (em baixo).

A oclusão, nas designadas consoantes velares, acontece quando a parte posterior do dorso da língua toca a região velar. O local exacto da constrição parece, no entanto, variar em função do falante: a simples observação dos dados (figura 4.10) sugere que as consoantes produzidas pelo falante

JHM têm o seu ponto de oclusão numa região mais anterior do que as articuladas pelos outros dois sujeitos. Nestes, a língua entra em contacto com o palato na zona de transição entre o palato duro e o véu palatino.

Paralelamente, tal como as restantes oclusivas, também as dorso-velares parecem bastante susceptíveis a efeitos coarticulatórios (Barbosa, 1994a; Sá Nogueira, 1938). Segundo as imagens apresentadas, este facto parece repercutir-se quer na zona da ponta da língua, para o falante AND (cf. Martins, 2007), quer na sua parte posterior, onde ocorre a oclusão em si mesma, no caso dos restantes informantes (cf. Barbosa, 1994a; Sá Nogueira, 1938).

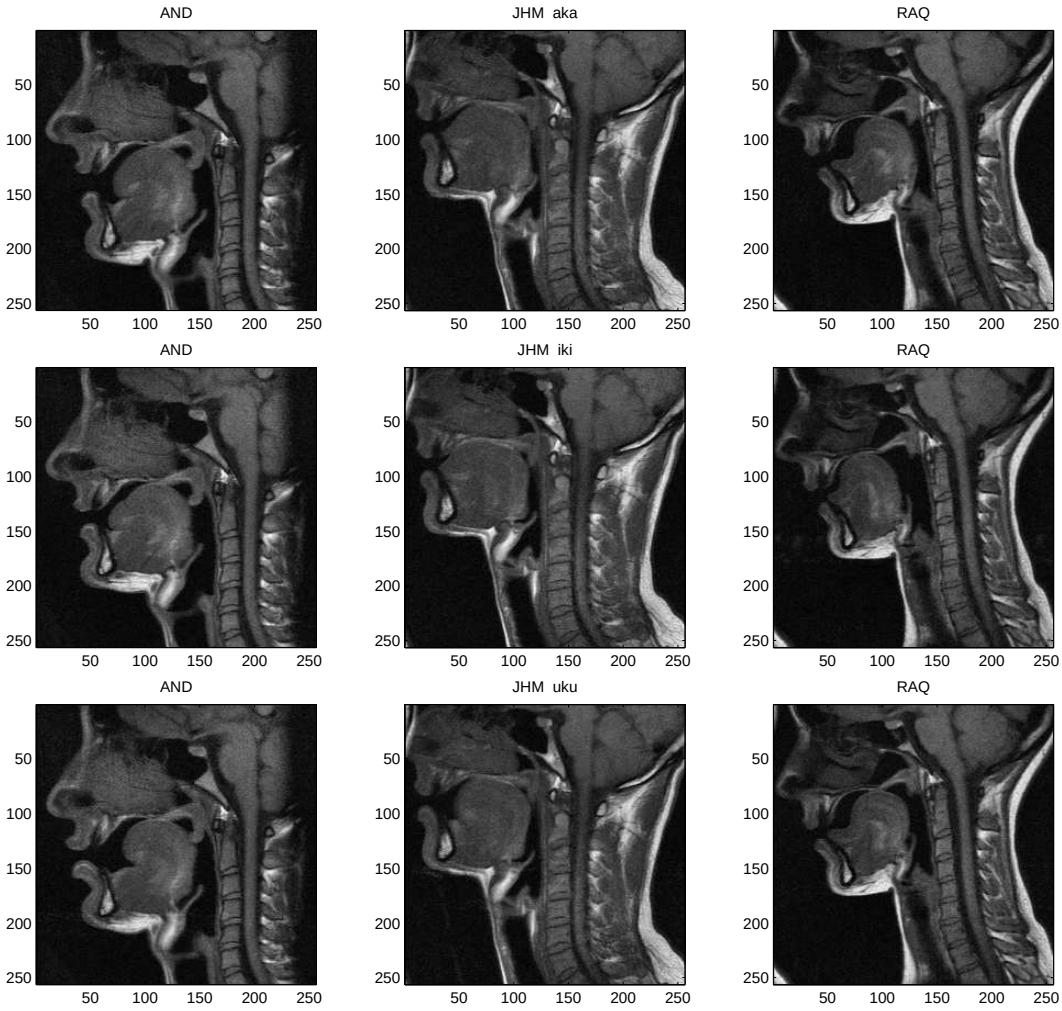


Figura 4.10: Perfis articulatórios dos sujeitos AND, JHM e RAQ, durante a produção da consoante [k], no contexto de [a] (em cima), [i] (ao centro) e [u] (em baixo).

Assim, às consoantes [k] e [g], foi associado um gesto de corpo da língua, genericamente descrito como `[velar]`, quanto ao local de constrição e evidentemente `[fechado]`, no respeita ao grau de constrição (uma característica comum a todas as oclusivas). Tal como as outras duas oclusivas surdas, a produção do [k] implica ainda um gesto glotal (tabela 4.8).

Não se verificando qualquer tipo de aspiração, o gesto de afastamento das pregas vocais será praticamente concomitante com o gesto oral e, consequentemente, "the maximum is reached at about the mid-point of the oral closure duration and the vocal folds return to a voicing position again at about the moment of release" (Ladefoged & Maddieson, 1995, p.66).

Pelo contrário, no inglês, em consequência da aspiração que caracteriza as oclusivas surdas, prevê-se um atraso na coordenação entre o gesto glotal e a articulação oral, o que significa dizer que o gesto de afastamento das pregas vocais continua activo depois do *offset* do gesto oral (Browman & Goldstein, 1986).

Esta diferença entre as duas línguas implicou inevitavelmente uma modificação nos princípios de composição gestual subjacentes a este tipo de segmentos (vd. anexo B).

A completa caracterização do grupo das oclusivas - apresentada na tabela 4.8 - implica ainda a referência à variável do tracto `[VEL]`, que dá conta da acção do véu palatino. Em se tratando de segmentos orais, este encontra-se naturalmente encostado à parede da faringe - e, portanto, especificado como `[CLO]` - garantindo, a par com outros factores, as condições de pressão adequadas à ocorrência da explosão, que caracteriza estes segmentos (vd. capítulo 5, secção 5.2.4).

4.2.5.2 Consoantes nasais

As consoantes nasais são produzidas, a exemplo do que ocorre com as oclusivas, mediante uma oclusão total do tracto vocal. Geralmente, as pregas vocais mantêm-se em vibração e, neste sentido, são bastantes as semelhanças entre consoantes nasais e as consoantes oclusivas sonoras. Contudo, durante a produção das primeiras, o véu palatino encontra-se descido e afastado da parede posterior da faringe, permitindo a passagem do ar pelas cavidades nasais [8].

As semelhanças articulatórias - nomeadamente a oclusão oral - entre consoantes oclusivas e consoantes nasais justificam que estas últimas sejam tratadas em alguma literatura fonética como oclusivas nasais.

Contudo, as consoantes nasais, tal como as vogais, podem ser sustidas durante um período de tempo relativamente longo, o que não acontece normalmente com as oclusivas.

Do sistema consonantal do PE fazem parte três consoantes nasais: [m], [n], [ɲ], as mesmas consideradas no nosso dicionário gestual.

No que respeita à sua distribuição, todas ocorrem em posição medial, mas em início de palavra, a nasal /ɲ/ está confinada a um número reduzido de vocábulos (e.g. "nhu") (Mateus & d'Andrade, 2000) [9].

Quanto ao ponto de articulação, a consoante [m] resulta do contacto dos lábios inferior e

[8] Cf., a este propósito, a definição de consoante nasal de Stevens (1998, p.305): "A nasal consonant is produced with a velopharyngeal opening but with a complete closure of the main vocal tract at some point within the oral cavity.".

[9] Segundo Mateus & d'Andrade (2000), as palavras começadas por /ɲ/ são mais frequentes no PB do que no PE. Contudo, no PB, antes da consoante é, muitas vezes, produzido um [i].

Tabela 4.8: Gestos associados às consoantes oclusivas do PE. As consoantes encontram-se foneticamente representadas em AFI e os gestos a elas associados são caracterizados pelo articulador (*Organ*), oscilador (*Osc*), variável do tracto (*TV*), constrição (*Const*). O *target* da constrição e o *stiffness* encontram-se pré-definidos num ficheiro independente (“.”), mas podem ser directamente especificados no dicionário.

Vogal	Organ	Osc	TV	Const	Target	Stiff
p	Lips	clo	LA	CLO	.	.
	Lips	rel	LA	REL	.	.
	Glottis	h	GLO	WIDE	.	.
	Velum	clo	VEL	CLO	.	.
t	TT	clo	TTCL	DENT	.	.
	TT	clo	TTCD	CLO	.	.
	TT	rel	TTCL	REL	.	.
	TT	rel	TTCD	REL	.	.
	Glottis	h	GLO	WIDE	.	.
	Velum	clo	VEL	CLO	.	.
k	TB	clo	TBCL	VEL	.	.
	TB	clo	TBCD	CLO	.	.
	TB	rel	TBCD	REL	.	.
	Glottis	h	GLO	WIDE	.	.
	Velum	clo	VEL	CLO	.	.
b	Lips	clo	LA	CLO	.	.
	Lips	rel	LA	REL	.	.
	Velum	clo	VEL	CLO	.	.
d	TT	clo	TTCL	DENT	.	.
	TT	clo	TTCD	CLO	.	.
	TT	rel	TTCL	REL	.	.
	TT	rel	TTCD	REL	.	.
	Velum	clo	VEL	CLO	.	.
g	TB	clo	TBCL	VEL	.	.
	TB	clo	TBCD	CLO	.	.
	TB	rel	TBCD	REL	.	.
	Velum	clo	VEL	CLO	.	.

superior (bilabial); o [n] é tradicionalmente classificado como ápico-dental (Cruz-Ferreira, 1999a; Sá Nogueira, 1938; Barroso, 1999) ou ápico-alveolar (Moutinho, 2000; Emiliano, 2006; Cunha & Cintra, 1997; Mateus *et alii*, 2005; Veloso, 1999); e o /ɲ/ é tido como consoante palatal.

Os dados articulatórios disponíveis, obtidos a partir de ressonância magnética, permitem confirmar os pontos de articulação tradicionalmente referenciados. As configurações nasais foram adquiridas, para os três informantes, em versão sustentada, a partir de uma palavra de referência. No caso dos informantes JMH e RAQ - que adquiriram o *corpus* na segunda sessão, já depois de introduzidos novos contextos - a consoante alvo foi produzida, à semelhança das oclusivas, também

em contexto VCV (figura 4.11 a 4.13) [10].

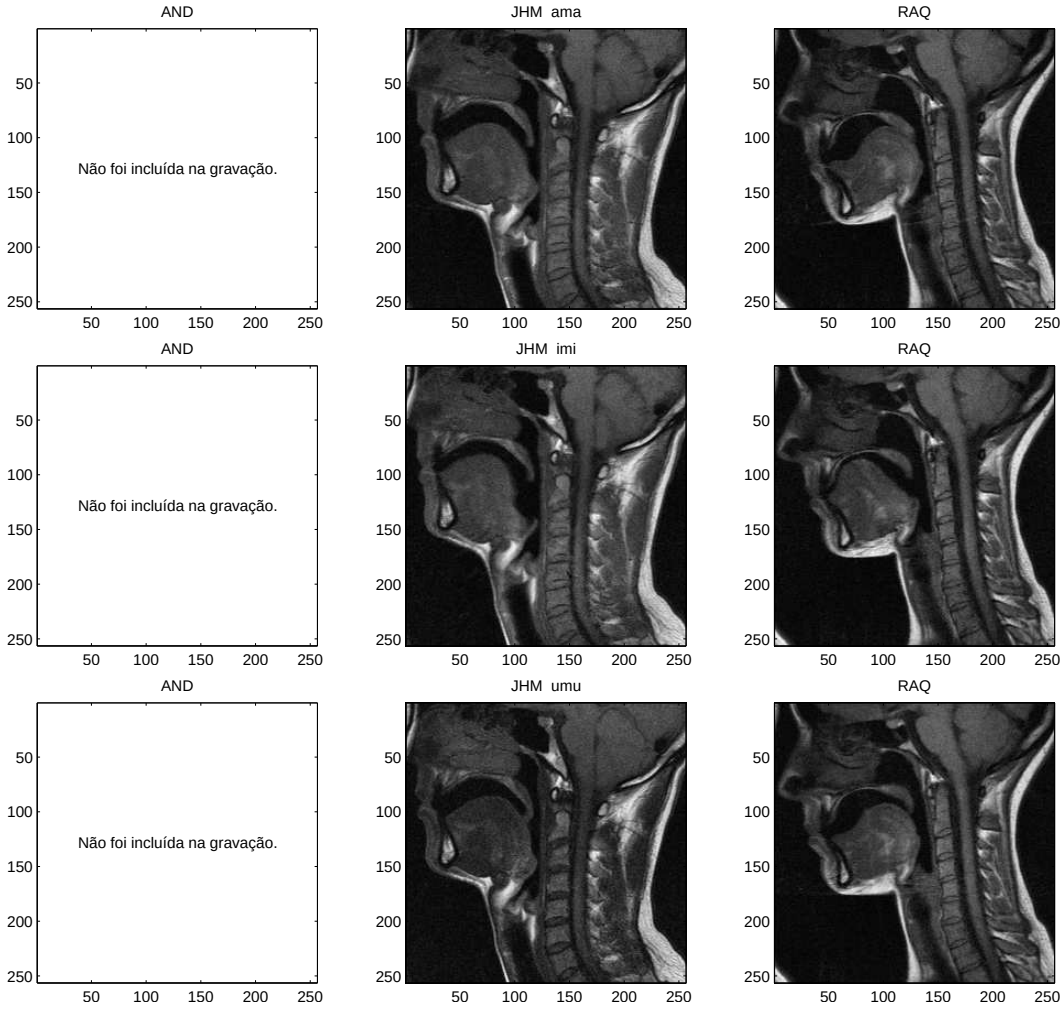


Figura 4.11: Perfis articulatórios dos sujeitos JHM e RAQ, durante a produção da consoante [m], no contexto de [a] (em cima), [i] (ao centro) e [u] (em baixo). O *corpus* gravado pelo informante AND, numa sessão independente, não incluiu este contexto.

No que respeita ao ponto de articulação, o [m] é nitidamente bilabial (figura 4.11), o que justifica que, na sua caracterização gestual (tabela 4.9), seja considerado um gesto de fechamento (e *release*) dos lábios.

Quanto ao [n] - produzido de modo similar às oclusivas orais com o mesmo ponto de articulação - a língua toca a região dental/ alveolar superior (figura 4.12). O problema da atribuição de um ponto de articulação (dental ou alveolar), referido a propósito das oclusivas orais, parece estender-se também a esta consoante. Face à dificuldade em identificar o local exacto de oclusão, a partir das imagens de RM, a solução adoptada passou - a exemplo do sucedido anteriormente e depois de al-

[10]Uma vez que, para o informante AND, não foram adquiridas oclusivas em contexto CVC, nas imagens não constam dados relativos a este informante.

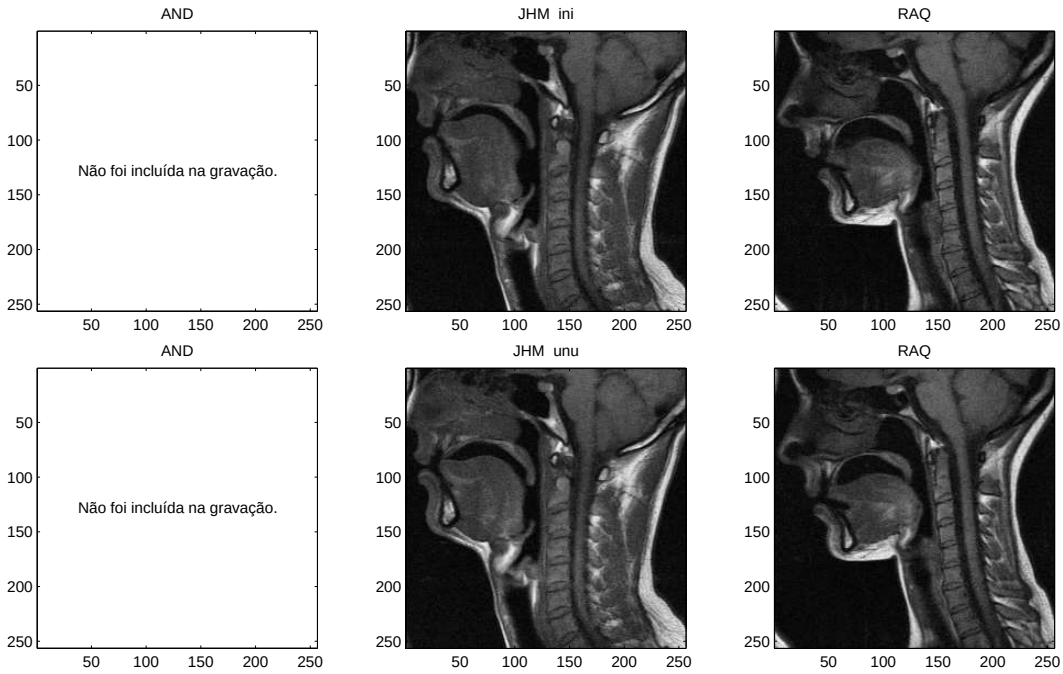


Figura 4.12: Perfis articulatórios dos sujeitos JHM e RAQ, durante a produção da consoante [n], no contexto de [i] (em cima) e [u] (em baixo). O *corpus* gravado pelo informante AND, numa sessão independente, não incluiu este contexto. As imagens relativas à consoante nasal, produzida pela informante RAQ, em contexto de [a], apresentam problemas de qualidade, pelo que optámos por não incluir este contexto.

guns testes auditivos informais - por associar um gesto de ponta da língua, fechado na região dental, à consoante em causa (tabela 4.9).

Na produção do [ɲ], o dorso da língua adapta-se efectivamente ao palato duro para impedir a saída do ar pela boca (figura 4.13).

Em virtude desta configuração, esta consoante é tradicionalmente classificada como dorso-palatal, a par do [ʎ]. Apesar disso, as duas consoantes apresentam diferenças significativas: no [ʎ], o dorso da língua encontra-se numa posição mais anterior do que no [ɲ] (cf. Martins *et alii*, 2008a). Os dados RM corroboram, assim, as observações prévias de Sá Nogueira (1938, p.42), feitas com base na análise de palatogramas: "a articulação do $\dot{n}$ e do $\dot{l}$ são muito semelhantes, (...), mas não são perfeitamente idênticos: o ponto de articulação do $\dot{l}$ é mais anterior que o do $\dot{n}$ (...). No $\dot{n}$ o contacto faz-se do médio-dorso com o médio-palato; no $\dot{l}$ faz-se do pre-dorso com o pre-palato".

A mesma progressão foi também atestada por Recasens & Espinosa (2006, p.297), para outras línguas românicas: "the main prediction regarding place of articulation for palatal consonants is that closure fronting should decrease in progression [ʎ]>[ɲ]>[c]".

Neste sentido, esta consoante encontra-se especificada, do ponto de vista gestual (tabela 4.9), através de um gesto de corpo da língua, fechado na região palatal.

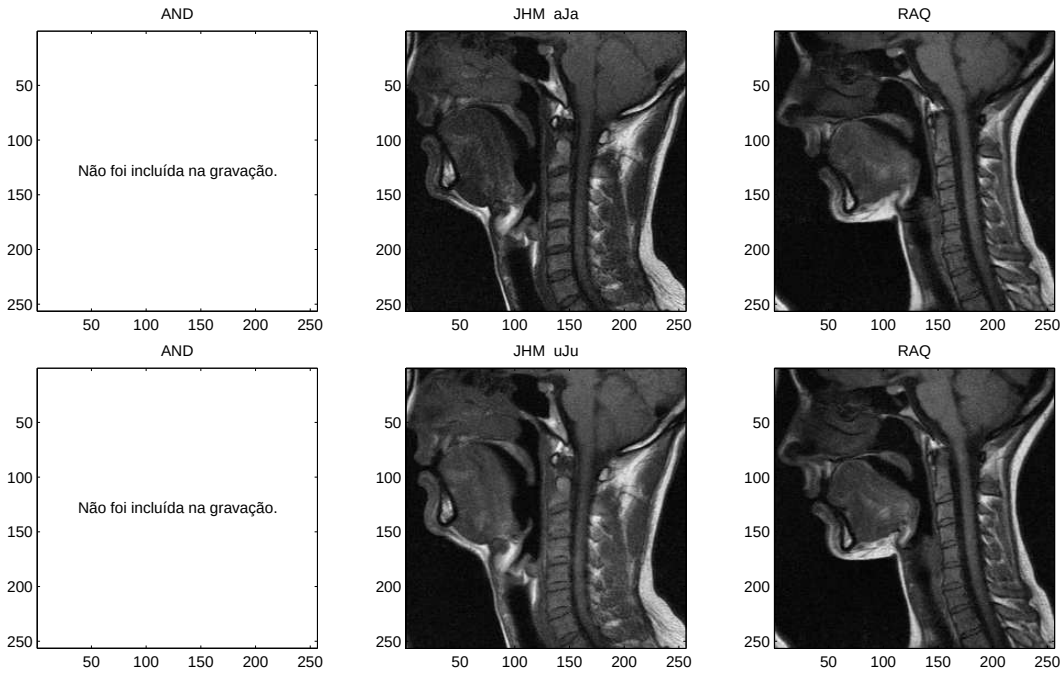


Figura 4.13: Perfis articulatórios dos sujeitos JHM e RAQ, durante a produção da consoante [ɲ], no contexto de [a] (em cima) e [u] (em baixo). O *corpus* gravado pelo informante AND, numa sessão independente, não incluiu este contexto. As imagens relativas à consoante nasal, produzida pela informante RAQ, em contexto de [i], apresentam problemas de qualidade, pelo que optámos por não incluir este contexto.

Para além do gesto de oclusão oral - dos lábios, da ponta da língua ou do corpo da língua - as consoantes nasais do português são caracterizadas por um gesto do velo, que, conforme se pode facilmente concluir a partir da observação das imagens RM acima apresentadas, se encontra claramente descido e afastado da parede da faringe.

Na tabela 4.9, este último foi genericamente especificado como `[wide]`. Numa fase posterior do trabalho (vd. capítulo 5), a questão das diferenças na altura do velo durante a produção das consoantes e vogais nasais do PE será devidamente equacionada, o que poderá justificar uma revisão dos parâmetros dinâmicos associados ao gesto do velo.

Em termos de coordenação entre os dois gestos, o TADA prevê (vd. anexo B) - em linha com os estudos articulatórios realizados sobre esta matéria, para o inglês americano (Krakow, 1989, 1993; Byrd *et alii*, no prelo) - que, em Ataque de sílaba, o velo e o articulador oral atinjam o *target* sensivelmente ao mesmo tempo (coordenação síncrona), ao passo que, em posição de Coda silábica, o gesto velar deve preceder o oral (coordenação sequencial). Os dados articulatórios relativos ao PB indiciam um comportamento similar (Oliveira & Marin, 2005). Assim, até que novos estudos sejam realizados, optámos por manter inalterado este padrão de coordenação.

Concluindo, no âmbito da nossa base de dados, as consoantes nasais do português foram analisadas da seguinte forma (tabela 4.9):

Tabela 4.9: Gestos associados às consoantes nasais do PE. As consoantes encontram-se foneticamente representadas em AFI e os gestos a elas associados são caracterizados pelo articulador (*Organ*), oscilador (*Osc*), variável do tracto (*TV*), constrição (*Const*). O *target* da constrição e o *stiffness* encontram-se pré-definidos num ficheiro independente (“.”), mas podem ser directamente especificados no dicionário.

Vogal	Organ	Osc	TV	Const	Target	Stiff
m	Lips	clo	LA	CLO	.	.
	Lips	rel	LA	REL	.	.
	Velum	n	VEL	WIDE	.	.
n	TT	clo	TTCL	DENT	.	.
	TT	clo	TTCD	CLO	.	.
	TT	rel	TTCL	REL	.	.
	TT	rel	TTCD	REL	.	.
	Velum	n	VEL	WIDE	.	.
ɲ	TB	clo	TBCL	PAL	.	.
	TB	clo	TBCD	CLO	.	.
	TB	rel	TBCD	REL	.	.
	Velum	n	VEL	WIDE	.	.

4.2.5.3 Consoantes fricativas

As consoantes que envolvem uma constrição supralaríngea suficientemente estreita para gerar ruído de fricção, aquando da passagem do fluxo do ar, são tradicionalmente designadas por fricativas.

Do sistema consonântico do PE fazem parte três fricativas surdas - [f], [s] e [ʃ] - e três fricativas vozeadas - [v], [z] e [ʒ].

Quanto ao ponto de articulação, o [f] e o [v] são tradicionalmente classificados como lábio-dentais, já que o lábio inferior se aproxima dos incisivos superiores, enquanto o [s] e o [z] são tidas como alveolares (Barbosa, 1994a; Cruz-Ferreira, 1999a; Veloso, 1999) ou dentais (Lacerda & Hammarström, 1952; Moutinho, 2000; Mateus *et alii*, 2005), pois a coroa da língua aproxima-se da região dento-alveolar. Já o [ʃ] e o [ʒ] são alvo de descrições algo divergentes, sendo possível encontrar designações como “palato-alveolares” (Cruz-Ferreira, 1999a), “palatais” (Lacerda & Hammarström, 1952; Mateus *et alii*, 2005), “pré-palatais” (Moutinho, 2000) e “predorso-prepalatais” (Veloso, 1999). Apesar de toda esta diversidade, de um modo geral, as várias nomenclaturas parecem indiciar um ponto de articulação mais recuado para estas consoantes, quando comparadas com o [s] e o [z].

Em algumas regiões do norte de Portugal, o sistema descrito assume contornos um pouco mais complexos. Com efeito, em alguns dialectos mais conservadores (Barbosa, 1994a; Ferreira *et alii*, 1996; Mateus & d’Andrade, 2000) é possível distinguir não um, mas dois pares de sibilantes na região dento-alveolar, ou seja, duas consoantes ápico-alveolares e duas consoantes com articulação predorso-dental, que correspondem a representações gráficas distintas: “o [ʃ] corresponde ao s- inicial de palavra e a -ss- gráficos; o [ʒ] corresponde a -s- intervocálico gráfico. As predorsodentais têm as correspondências seguintes: o [s] corresponde a $c^{e,i}$ e a -ç- ; o [z] corresponde a z gráfico em posição

inicial ou medial" (Ferreira *et alii*, 1996, p.494).

Paralelamente, em alguns desses dialectos a africada [ʧ] (graficamente representada por <ch>) opõe-se à fricativa palatal [ʃ] (que corresponde, em termos de grafia ao <x>), assegurando o contraste entre "buxo" [ˈbuʃu] e "bucho" [ˈbuʧu] [11].

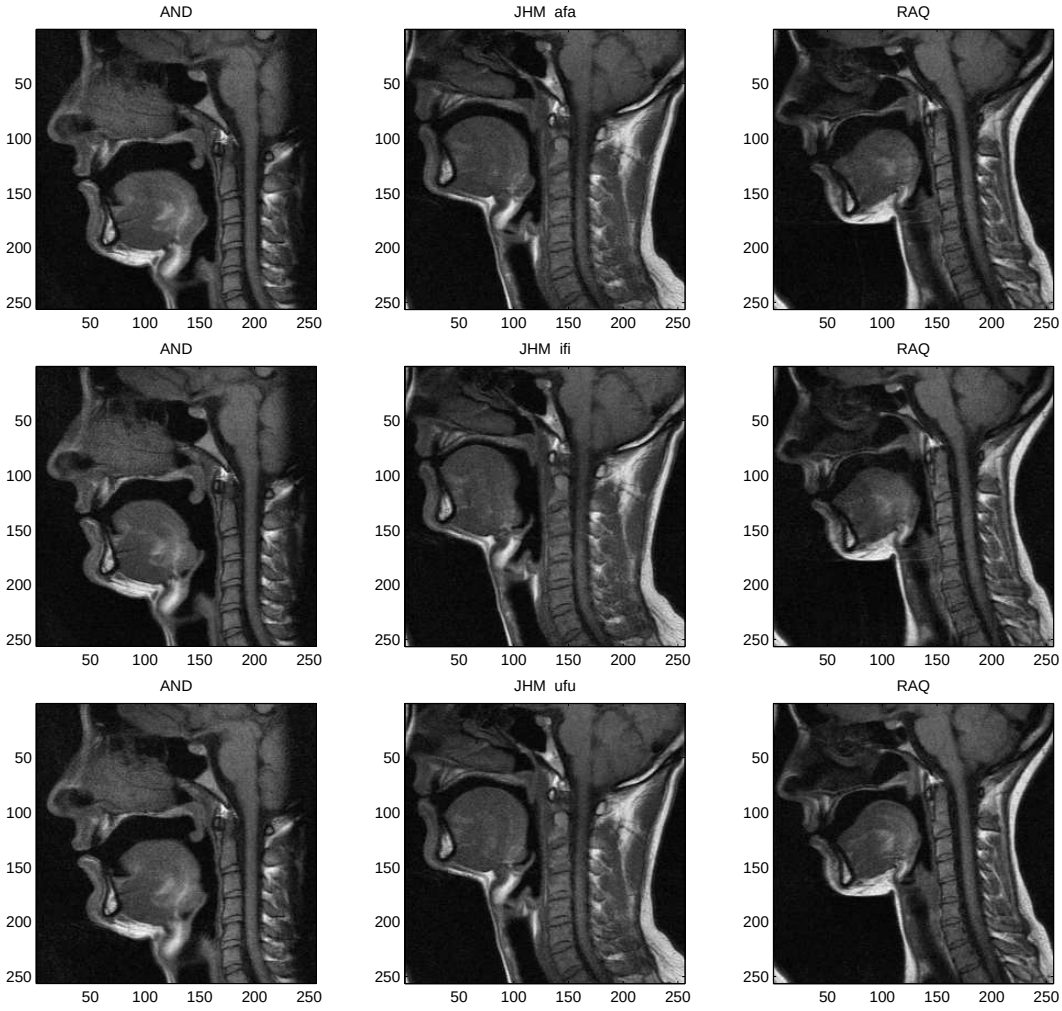


Figura 4.14: Perfis articulatórios dos sujeitos AND, JHM e RAQ, durante a produção da consoante [f], no contexto de [a] (em cima), [i] (ao centro) e [u] (em baixo).

Partimos, como habitualmente, destas classificações tradicionais para a caracterização gestual das consoantes fricativas, confrontando as descrições articulatórias da literatura (relativas ao português e outras línguas) com os dados RM.

Devido à ausência dos dentes nas imagens, a análise das imagens de ressonância magnética relativas às fricativas lábio-dentais (figura 4.14) não nos permite muito mais do que inferências gerais acerca do mecanismo de produção desta classe de sons. A posição dos lábios parece indiciar uma aproximação entre o lábio inferior e o sector anterior da arcada dentária superior. Este movimento,

[11] O exemplo é da autoria de Mateus & d'Andrade (2000, p.13).

de acordo com Ladefoged & Maddieson (1995, p.140-141), é da exclusiva responsabilidade do lábio inferior, já que o lábio superior permanece numa posição elevada, enquanto o lábio inferior sobe e recua em direcção aos dentes. Uma outra alternativa para que se produzam articulações lábio-dentais passa pela aproximação da parte interna do lábio inferior com a superfície frontal dos incisivos (Ladefoged & Maddieson, 1995, p.141).

De acordo com estas informações, o [f] e o [v] podem ser representados por um gesto labial, especificado como `[crítico]` (`[CRIT]`) em relação ao grau de constrição (`[LA]`) e como `[dental]` (`[DENT]`), no que toca ao local de constrição (`[LP]`). Em se tratando de sons não-nasais, o véu palatino encontra-se, naturalmente `[fechado]` (`[CLO]`) [12]. As pregas vocais vibram para o [v], mas não para o [f], pelo que, para esta última consoante, foi também considerado um novo gesto glotal `[aberto]`.

No tocante às fricativas [s] e [z], conforme verificámos em parágrafos anteriores, as diversas obras de fonética do português oscilam tanto na sua classificação como *dentais* ou *alveolares*, como na referência ao articulador passivo (ápice ou lâmina/pré-dorso da língua).

Mais uma vez, os dados articulatórios adquiridos por meio de RM não são absolutamente esclarecedores relativamente a esta matéria - em boa parte devido à ausência do contorno dos dentes - embora as imagens, apresentadas em seguida (figura 4.15), pareçam apontar para uma realização lâmino-alveolar. Já os palatogramas disponibilizados por Sá Nogueira (1938, p.46) evidenciam alguma variabilidade entre sujeitos, motivada, segundo a explicação do autor, pelas diferenças na "forma da concavidade da abóbada palatina". Com base nestes dados, poderíamo supor que, tal como no inglês e no chinês (Ladefoged & Maddieson, 1995), também no PE, estas fricativas estão sujeitas a alguma variabilidade no ponto de articulação.

Em face dos dados, na elaboração da tabela gestual referente ao [s] e [z], optámos por uma solução intermédia que passou por definir um gesto de ponta da língua, `[crítico]` (`[CRIT]`) e `[alveolar]` (`[ALV]`), mas ajustado para os 50 graus, de modo a reflectir a tendência para uma articulação mais anterior. No caso da fricativa alveolar vozeada, também os valores relativos ao TTCD, originalmente fixados em 1 mm, foram diminuídos para os 0.16 mm, de modo a simular, eficazmente, a ocorrência de turbulência à passagem do ar pela constrição.

É sabido que, durante a produção das fricativas, toda a configuração do tracto vocal está sujeita a um elevado grau de precisão [13] que se traduz numa elevada resistência aos efeitos de co-

[12] Como veremos adiante (capítulo 5, secção 5.2.4), as consoantes obstruintes são os segmentos menos compatíveis com a nasalidade (Ohala, 1975; Ohala & Ohala, 1993). No caso específico das fricativas, o abaixamento do velo e consequente saída de ar pelas fossas nasais reduz a pressão intraoral, o que pode impedir a geração de ruído (Ohala *et alii*, 1998b; Solé, 1999). Nas palavras de Solé (1999), "to the extend that a fricative is a good fricative perceptually, it cannot be nasalized (without added biomechanical cost, *e.g.*, increased subglottal pressure). Thus, the features frication and nasalization bleed each other aerodynamically and do not combine into a sufficiently discriminable percept.".

[13] "The gesture forming the constriction in many fricatives has a greater degree of articulatory precision than that required in stops and nasals (...) in a fricative a variation of one milimeter in the position of the target for the crucial part of the vocal tract makes a great deal of difference. There has to be a very precisely shaped channel for a turbulent airstream to be produced." (Ladefoged & Maddieson, 1995, p.137).

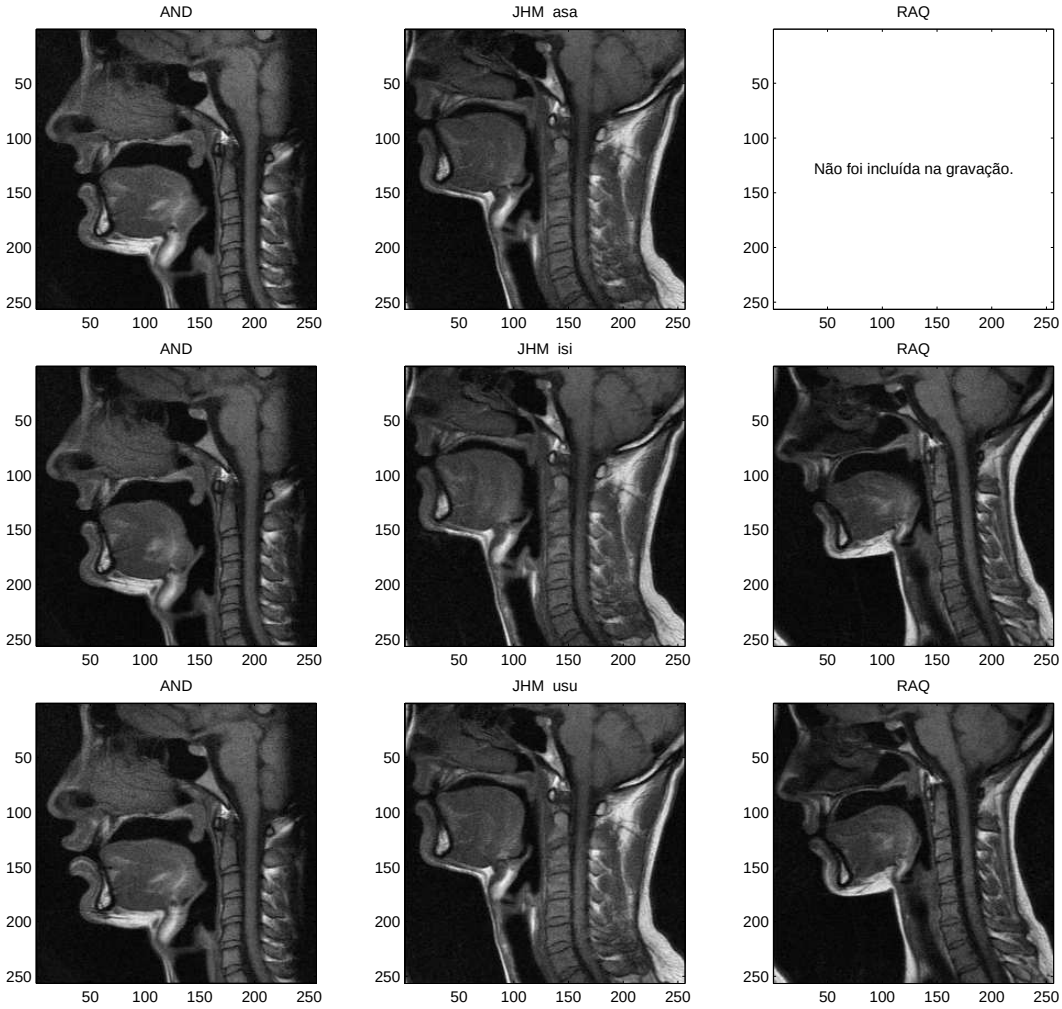


Figura 4.15: Perfis articulatórios dos sujeitos AND, JHM e RAQ, durante a produção da consoante [s], no contexto de [a] (em cima), [i] (ao centro) e [u] (em baixo). A imagem relativa à consoante [s], produzida pela informante RAQ, perdeu-se no processo de aquisição e tratamento dos dados.

articulação impostos pelas vogais adjacentes (Martins, 2007; Martins *et alii*, 2008b). Este facto é particularmente evidente no caso das fricativas alveolares e (pré-)palatais (Martins, 2007).

Assim, a criação das condições articulatórias necessárias à produção das fricativas dento-alveolares exige ainda a especificação de um gesto dorsal, a par do gesto de ponta da língua. Para tanto, baseámo-nos na proposta original referente às fricativas do inglês americano, que assume um local de constrição `[velar]` (`[VEL]`) e um grau de constrição `[wide]`, correspondente a uma abertura de 10 mm, para a variável do tracto TB.

À semelhança das fricativas lábio-dentais, também a passagem do ar para as fossas nasais se encontra, naturalmente, vedada, daí o gesto velar, `[fechado]` (`[CLO]`) associado às representações gestuais do [z] e [s]. No caso desta última consoante, as pregas vocais não vibram, pelo que se torna ainda necessária referência à abertura glotal.

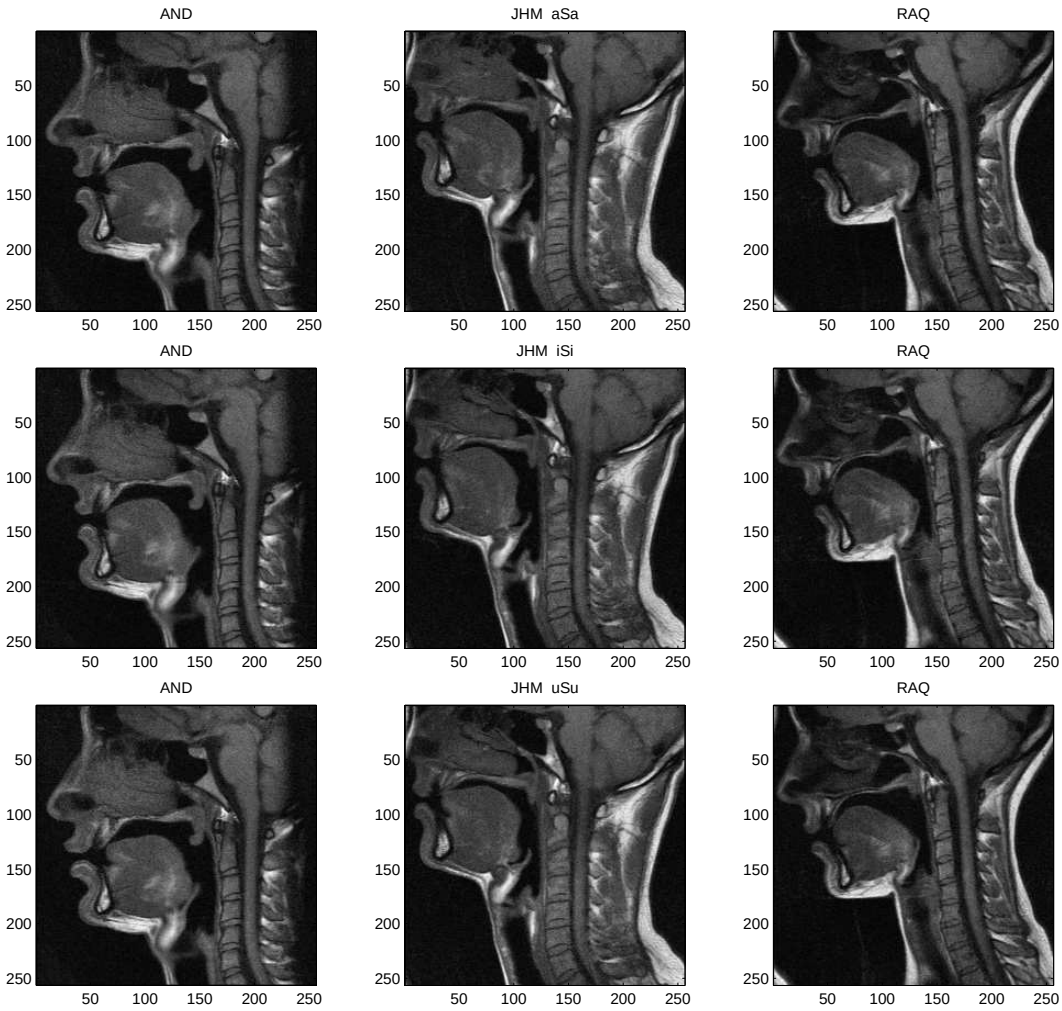


Figura 4.16: Perfis articulatórios dos sujeitos AND, JHM e RAQ, durante a produção da consoante [ʃ], no contexto de [a] (em cima), [i] (ao centro) e [u] (em baixo).

Finalizaremos a descrição dos gestos associados às fricativas com a caraterização das consoantes [ʃ] e [ʒ]. Apesar da diversidade de termos usados na literatura para descrever estes sons, é ponto assente que o [ʃ] e [ʒ] são as fricativas mais posteriores do português [14].

Ainda que o ponto de articulação não possa ser determinado com exactidão, apenas com base na observação das imagens RM, é mais ou menos evidente que a língua se aproxima do palato numa zona ligeiramente mais posterior do que o verificado em relação às denominadas fricativas alveolares (figura 4.16). O ponto de máxima constrição não chega, contudo, a ser palatal - pelo que esta classificação é de evitar, ficando reservada a consoantes, cuja articulação ocorre numa zona mais recuada - parecendo antes ter lugar na região que Ladefoged & Maddieson (1995, p.148) chamam de "post-alveolar region, that is, on the center of the alveolar protuberance". Para além disso, a parte

[14] O [ʃ] e [ʒ] são as fricativas mais posteriores do português, se não tivermos em conta a possibilidade de realização do /r/ como fricativa uvular.

frontal e central da língua encontra-se numa posição mais elevada para as fricativas pós-alveolares do que para as fricativas alveolares.

Com base nestes dados, às consoantes em análise foram associados dois gestos distintos: um gesto de ponta da língua (TT) `[crítico]` (`[CRIT]`), localizado na região alveolo-palatal (`[ALVPAL]`) [15]; e um gesto de corpo da língua (TB), especificado como `[estreito]` (`[NAR]`) e `[palatal]` (`[PAL]`), necessário para modelar com exactidão a configuração do tracto imediatamente atrás da constrição e assegurar as condições aerodinâmicas essenciais à geração de ruído. Os valores do *target* do TBCD foram aumentados de 2 mm para 8 mm, seguindo as orientações para o inglês.

As observações acerca do gesto do véu palatino feitas anteriormente a propósito das consoantes lábio-dentais e alveolares aplicam-se de forma similar a esta última classe de fricativas.

Quanto ao gesto da glote, está presente apenas no caso da fricativa surda [ʃ].

Toda a descrição gestual é resumida na tabela 4.10.

4.2.5.4 Consoantes laterais

As laterais são tradicionalmente caracterizadas pela presença de uma constrição ao longo da linha médio-sagital, com passagem da corrente expiratória através de um ou dos dois lados do dorso da língua (Ladefoged & Maddieson, 1995; Narayanan *et alii*, 1997). Ainda que sejam atestados, pelo menos, nove pontos de articulação no que respeita às laterais aproximantes, na maioria das línguas, a oclusão tende a verificar-se na região dento-alveolar (Ladefoged & Maddieson, 1995).

No português, distinguem-se dois tipos de consoantes laterais, em função do ponto de articulação: a lateral dorso-palatal /ʎ/ ; e a lateral /l/, com uma constrição na parte anterior do tracto vocal (Mateus *et alii*, 1990, 2005; Barbosa, 1965; Cruz-Ferreira, 1999a).

Em relação a esta última, tal como referenciado a propósito de outras consoantes, as propostas de atribuição de um ponto de articulação não são inteiramente coincidentes [16]. Em Mateus *et alii* (2005), Moutinho (2000), Emiliano (2006), Veloso (1999), Barroso (1999), a lateral /l/ é considerada alveolar, ao passo que Cruz-Ferreira (1999a) a classifica como dental.

Também em relação à manifestação fonética da consoante /l/, as opiniões dos estudiosos

[15] A etiqueta `[ALVPAL]` corresponde a um ângulo de 60 graus, sendo este valor apenas ligeiramente superior ao considerado para o local de constrição `[ALV]`, fixado nos 56 graus.

[16] A oscilação na atribuição de um ponto de articulação fica patente nas afirmações de autores portugueses como Barbosa (1965, p.170), que caracteriza o /l/ "comme une latérale **apicoalvéolaire ou apicodentale**, généralement sonore plus vélarisée en position finále de syllable qu'ailleurs"; ou de Sá Nogueira (1938, p.53), segundo o qual, na produção da referida consoante, "**o ápice da língua adapta-se aos incisivos superiores ou à região alveolar dêstes**, para impedir a saída do ar segundo a linha média do canal bucal..." (sublinhado nosso).

Tabela 4.10: Gestos associados às consoantes fricativas do PE. As consoantes encontram-se foneticamente representadas em AFI e os gestos a elas associados são caracterizados pelo articulador (*Organ*), oscilador (*Osc*), variável do tracto (*TV*), constrição (*Const*). O *target* da constrição e o *stiffness* encontram-se pré-definidos num ficheiro independente (“.”), mas podem ser directamente especificados no dicionário.

Vogal	Organ	Osc	TV	Const	Target	Stiff
f	Lips	crt	LA	CRIT	.	.
	Lips	rel	LA	REL	.	.
	Lips	crt	LP	DENT	.	.
	Lips	rel	LP	REL	.	.
	Glottis	h	GLO	WIDE	.	.
	Velum	crt	VEL	CLO	.	.
v	Lips	crt	LA	CRIT	.	.
	Lips	rel	LA	REL	.	.
	Lips	crt	LP	DENT	.	.
	Lips	rel	LP	REL	.	.
	Velum	crt	VEL	CLO	.	.
s	TT	crt	TTCL	ALV	.	.
	TT	crt	TTCD	CRIT	.	.
	TT	rel	TTCL	REL	.	.
	TT	rel	TTCD	REL	.	.
	TB	crt	TBCL	VEL	.	.
	TB	crt	TBCD	WIDE	.	.
	Glottis	h	GLO	WIDE	.	.
	Velum	crt	VEL	CLO	.	.
z	TT	crt	TTCL	ALV	.	.
	TT	crt	TTCD	CRIT	0,16	.
	TT	rel	TTCL	REL	.	.
	TT	rel	TTCD	REL	.	.
	TB	crt	TBCL	VEL	.	.
	TB	crt	TBCD	WIDE	.	.
	Velum	crt	VEL	CLO	.	.
ʃ	TT	crt	TTCL	ALVPAL	.	.
	TT	crt	TTCD	CRIT	.	.
	TT	rel	TTCL	REL	.	.
	TT	rel	TTCD	REL	.	.
	TB	crt	TBCL	PAL	.	.
	TB	crt	TBCD	NAR	8	.
	Glottis	h	GLO	WIDE	.	.
	Velum	crt	VEL	CLO	.	.
ʒ	TT	crt	TTCL	ALVPAL	.	.
	TT	crt	TTCD	CRIT	0,5	.
	TT	rel	TTCL	REL	.	.
	TT	rel	TTCD	REL	.	.
	TB	crt	TBCL	PAL	.	.
	TB	crt	TBCD	NAR	8	.
	Velum	crt	VEL	CLO	.	.

portugueses parecem divergir. Alguns estudos reclamam para a consoante apical aquilo que Andrade (1998) chama de “comportamento binário a nível fonético”, i.e., tal como acontece noutras línguas,

nomeadamente no inglês, também no PE, o /l/ estaria categoricamente associado a dois alofones, um não-velarizado ("clear", "light", "non-velarized", "non-pharyngealized") em Ataque ([l]) e outro velarizado ("dark", "velarized", "pharyngealized") em posição de Coda ([ɬ]) (Lacerda & Hammarström, 1952; Cunha & Cintra, 1997; Faria *et alii*, 1996; Mateus & d'Andrade, 2000).

Contudo, de acordo com outras descrições, o fenómeno de velarização pode ser atestado em posição intervocálica, dependendo da qualidade da vogal anterior e do próprio falante (Viana, 1973a).

Barbosa (1965, 1994a), por seu turno, defende que a variação em causa deve ser interpretada como um fenómeno gradual (vd. nota 16).

Outros ainda não hesitam em reconhecer "the extremely "dark" quality of the commoner variety of l-sound" (Strevens, 1954, p.6), como uma das características distintivas do português. Descrições mais recentes, baseadas na pronunciação lisboeta formal, parecem também apontar neste sentido (Cruz-Ferreira, 1999a; Emiliano, 2006).

No sentido de esclarecer esta questão, na caracterização dos gestos envolvidos na produção das consoantes laterais, haveremos de contar não só com dados acústico-articulatórios disponíveis para o português, mas também com informações veiculadas para outras línguas, nomeadamente aquelas em que o /l/ parece manifestar comportamentos fonéticos semelhantes aos do PE.

Vários estudos articulatórios (Sá Nogueira, 1938; Giles & Moll, 1975; Gartenberg, 1984; Browman & Goldstein, 1995b; Sproat & Fujimura, 1993; Narayanan *et alii*, 1997) reportam que a principal diferença entre as duas variantes do /l/ reside na configuração do dorso da língua atrás da constrição principal: maior recuo e elevação em Coda do que em Ataque [17].

Por outro lado, o alofone pós-vocálico é sistematicamente produzido com redução da amplitude do gesto de ponta da língua, com possibilidade de perda de contacto entre este articulador e a região anterior do palato duro (Giles & Moll, 1975; Browman & Goldstein, 1995b; Sproat & Fujimura, 1993; Gick, 2003; Narayanan *et alii*, 1997) [18].

Relativamente ao problema da velarização do /l/, as configurações do tracto vocal dos falantes AND e RAQ, induzidas a partir das palavras "mal" e "laço" (figuras 4.17 e 4.18), indiciam que a diminuição da área na região velo-faríngea, em consequência da elevação do corpo da língua em direcção à região velar, não é exclusiva do /l/ final. Um elevado grau de velarização sobressai também das imagens adquiridas em contexto [VCV] (figura 4.19), sobretudo no caso da informante feminina. Contudo, tendo em conta a natureza do *corpus* adquirido, nomeadamente o modo de elicitação dos

[17] Sá Nogueira (1938, p.54) descreve assim o fenómeno de velarização: "[As laterais velarizadas] pronunciam-se na generalidade do mesmo modo que o *l* ápico-dental. A diferença está em que neste o ápice da língua recua um pouco mais, e o pos-dorso eleva-se um tanto, o que determinou que se chamasse *velar* a êste *l*".

[18] A reportada redução articulatória em final de sílaba é consistente com o chamado fenómeno de "vocalização" da lateral, caracterizada pela completa ausência da articulação alveolar. Esta parece verificar-se em algumas variedades do inglês britânico (Wrench & Scobbie, 2003a,b; Hardcastle & Barry, 1989), em certos dialectos do norte de Portugal (Mateus & d'Andrade, 2000; Boléo & Silva, 1962) e no PB (Feldman, 1972). Nesta última variedade, a lateral em Coda é sistematicamente realizada como uma aproximante labio-velar, dando origem a homófonos como [maw] ("mal") *vs* [maw] ("mau") (Barbosa & Albano, 2004). O aparente antagonismo entre o gesto de velarização e a elevação da ponta da língua e recuo do dorso podem estar na origem do fenómeno.

sons, não é possível, somente a partir dos dados RM para o PE, averiguar com toda a clareza qual a verdadeira natureza do /l/ e respectiva relação com a posição silábica.

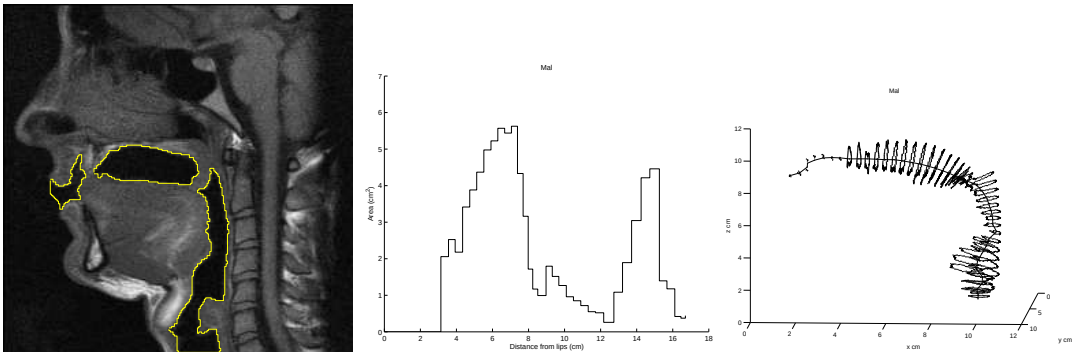


Figura 4.17: Contornos sagitais e funções de área do sujeito AND, durante a produção do /l/, na palavra “mal” (fonte: Martins *et alii*, 2008a).

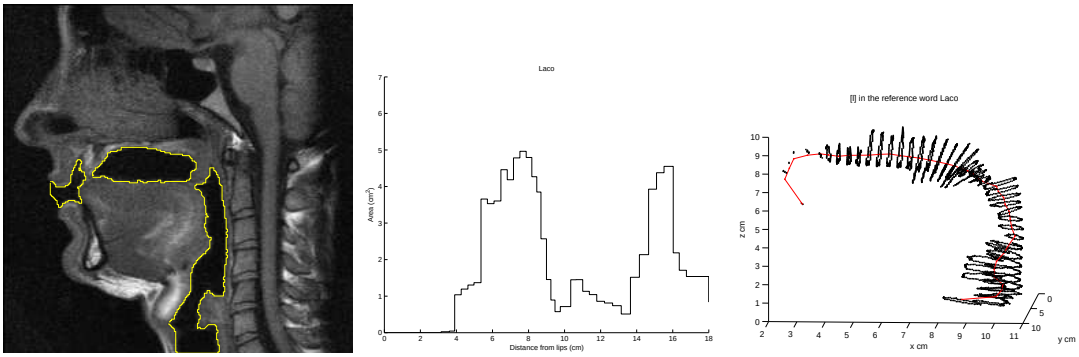


Figura 4.18: Contornos sagitais e funções de área do sujeito AND, durante a produção do /l/, na palavra “laço” (fonte: Martins *et alii*, 2008a).

As referidas imagens permitem, no entanto, afirmar, com toda a certeza, que, tal qual no inglês (cf. Narayanan *et alii*, 1997), também a produção da lateral em PE implica uma oclusão dento-alveolar, embora a ausência dos dentes nas imagens não permita grandes conclusões sobre a localização exacta dessa constrição.

Contrariamente, os perfis articulatórios [19] obtidos por Sá Nogueira (1938), a partir de imagens radiológicas, revelam que o dorso da língua apresenta maior recuo e elevação em direcção ao velo em [al] do que em [la], o que parece corroborar, de algum modo, a ideia de que, também em PE, o /l/ está categoricamente associado a dois alofones, que se distribuem de forma complementar: um velarizado, em Coda, e outro não-velarizado, em Ataque. Contudo, os dados relativos à lateral em grupo consonântico ([pla]), que mostram uma lateral ainda mais velarizada do que noutros contextos, parecem contrariar, de alguma forma, a hipótese binária. O facto leva Andrade (1998) a antecipar uma outra explicação, a de que a velarização se manifestaria essencialmente nas “margens silábicas”,

[19] Para além dos diagramas articulatórios (cortes sagitais), o autor apresenta (Sá Nogueira, 1938) os palatogramas correspondentes.

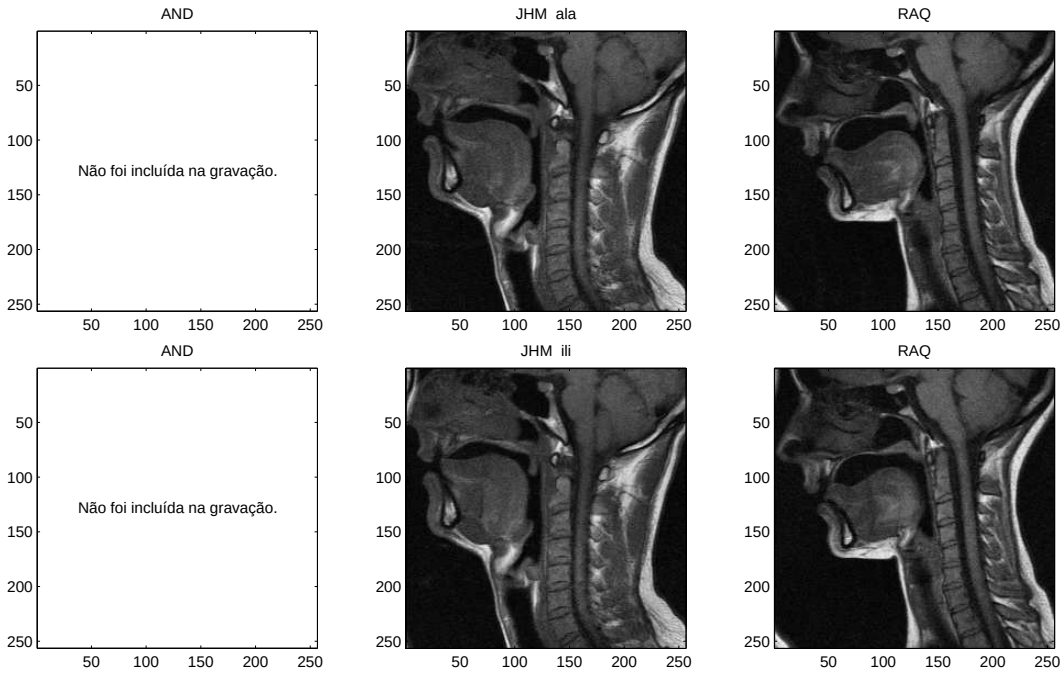


Figura 4.19: Perfis articulatórios dos sujeitos JHM e RAQ, durante a produção da consoante [l], no contexto de [a] (em cima) e [i] (em baixo). O *corpus* gravado pelo informante AND, numa sessão independente, não incluiu este contexto. As imagens relativas à consoante lateral alveolar, produzida pela informante RAQ, em contexto de [u], apresentam problemas de qualidade, pelo que optámos por não incluir este contexto.

hipótese esta infirmada pelos resultados acústicos da própria investigadora.

Neste caso particular, em que é evidente a necessidade de proceder à recolha e análise de novos dados articulatórios, outros parâmetros acústicos devem ser considerados, no sentido de esclarecer a questão da realização da lateral em PE e apoiar alguma das hipóteses até agora avançadas.

Na medida em que as frequências dos formantes são sensíveis às modificações na configuração da língua, o grau de velarização pode, efectivamente ser inferido a partir de dados acústicos. Entre F2 e as características do sistema de ressonância atrás da constrição existe uma relação íntima (Fant, 1960): o /l/ velarizado apresenta sempre um segundo formante mais baixo do que a variante não velarizada, em consequência da formação de uma constrição dorsal velo-faríngea (e.g Lehiste, 1964; Bladon & Al-Bamerni, 1976; Epsy-Wilson, 1992). O F1, por seu turno, tende a subir, em virtude do abaixamento do pré-dorso da língua.

Neste sentido, e tendo em consideração os resultados publicados para línguas que contrastam laterais velarizadas e não velarizadas (e.g. russo, búlgaro e albanês) (Andrade, 1998; Recasens & Espinosa, 2005; Ladefoged & Maddieson, 1995) e línguas com um *clear* (e.g. castelhano, francês, italiano, alemão) (Recasens *et alii*, 1995, 1996; Recasens & Espinosa, 2005) ou um *dark* (e.g cata-

lão e algumas variedades do inglês) (Recasens & Espinosa, 2005) [20] /l/, os valores de F2 [21] obtidos para o português (Andrade, 1998, 1999) são próprios de uma lateral velarizada [22], embora a análise dos dados individuais revele a co-ocorrência de comportamentos distintos, i.e., de vários graus de velarização.

Por outro lado, é sabido que a realização da lateral é em grande parte condicionada pelo contexto adjacente, sobretudo pelas características das vogais à sua direita (cf. Bladon & Al-Bamerni, 1976) [23].

O grau de influência da vogal sobre a lateral depende do grau de velarização desta última (cf. Bladon & Al-Bamerni, 1976). A variante velarizada mostra-se mais resistente à coarticulação vocálica do que a congénere não velarizada. O efeito é particularmente notório em contexto de [i], na medida os gestos dorsais envolvidos na produção do /l/ velarizado e da vogal são inerentemente antagónicos: o abaixamento e recuo do corpo da língua implicados na velarização da lateral bloqueiam, em larga medida, a elevação e anteriorização, requeridas para a produção da vogal (Recasens *et alii*, 1995, 1996; Recasens & Espinosa, 2005). Assim, o /l/ velarizado tende a exibir uma configuração da língua muito similar, independentemente da qualidade da vogal adjacente [24].

Relativamente a esta questão de coarticulação, os resultados acústicos publicados para o PE (Andrade, 1998, 1999) parecem sugerir que a lateral é, a despeito de alguma variabilidade entre sujeitos, tendencialmente velarizada nas diferentes posições silábicas consideradas e, consequentemente, pouco sensível ao contexto adjacente. Os dados carecem, no entanto, de um tratamento estatístico adequado, sendo que para a avaliação do grau de resistência articulatória seriam também de considerar outros parâmetros, como o já referido MDC (vd. nota 24).

Cabe ainda notar que os resultados dizem apenas respeito ao português de Lisboa, que apesar de metonimicamente referido pelos linguistas como *Português Europeu*, diatopicamente é apenas uma das muitas variedades do PE, elevada à categoria de norma-padrão, em virtude de circunstâncias históricas e sociais diversas.

Ficam, assim, por apurar as características acústicas e articulatórias da lateral noutras variedades, que não a variedade lisboeta, embora os recentes dados de ressonância magnética [25] pareçam apontar para um padrão de produção comum a todo o território português.

[20] Em Recasens & Espinosa (2005) podem ser encontrados valores de F2 publicados na literatura para várias línguas, juntamente com referências bibliográficas de interesse.

[21] Segundo Recasens *et alii* (1995), as duas variantes da lateral podem ser distinguidas em função do valor do segundo formante, tendo o limiar de 1450/1500Hz como referência.

[22] Recasens & Espinosa (2005) incluem o português no grupo das línguas com uma "strongly dark variety of /l/ in all positions", tendo em conta os dados publicados por Andrade (1999).

[23] Para além dos conhecidos efeitos de coarticulação entre laterais e vogais adjacentes, outros estudos (West, 1999, 2000) demonstram que as consoantes líquidas exercem efeitos coarticulatórios de longa-distância, quer nas vogais, quer nas consoantes.

[24] Para além da frequência de F2 e percentagem de contacto dorso-palatal, o índice quantitativo MDC (*mean articulatory distance*) tem sido utilizado para avaliar o grau de resistência coarticulatória da lateral no contexto de [i] e [a] (Bladon & Al-Bamerni, 1976; Recasens, 2004; Recasens & Espinosa, 2005).

[25] O informante AND é originário da zona norte do País, enquanto os informantes JHM e RAQ são da zona centro.

Independente do seu carácter velarizado, no quadro da FA, a lateral consiste sempre em dois gestos (Browman & Goldstein, 1995b): um gesto de ponta da língua e outro envolvendo o corpo da língua (cf. Sproat & Fujimura, 1993). De acordo com esta perspectiva, a emergência dos dois canais seria uma consequência meramente secundária desta configuração, em que a elevação da ponta da língua é acompanhada da retracção do dorso da língua (Browman & Goldstein, 1995b). As variantes lexicais possíveis do /l/ decorreriam de alterações temporais e/ou espaciais a esta configuração base (Albano, 2001).

Em linha com estas observações, o desenho gestual da lateral do PE implicará um gesto de ponta da língua (TT) e outro de corpo da língua (TB).

Permaneçam embora muitas incertezas, partindo do pressuposto que o /l/ em PE é sempre velarizado - como parecem sugerir os dados acústicos e, pelo menos, parte dos resultados RM - a constrição secundária, que tem lugar na região uvo-faríngea (`[UVU]`), encontra-se especificada como estreita (`[NAR]`) (cf. Albano, 2001) [26].

O gesto de ponta da língua, por sua vez, foi considerado `[alveolar]` (`[ALV]`), quanto ao ponto de constrição, e estreito (`[NAR]`) no que toca ao grau, a etiqueta reservada no TADA para as aproximantes (vd. tabela 4.11).

Um teste perceptual informal, usando o sintetizador HLsyn, demonstra que a especificação destes dois gestos, sem qualquer referência aos canais laterais, parece, efectivamente, induzir a percepção acústica de um [l], conforme previsto e testado, ainda que informalmente, por Browman & Goldstein (1995b).

Segundo vários estudos (Sproat & Fujimura, 1993; Browman & Goldstein, 1995b; Gick, 1999a, 2003; Gick *et alii*, 2006), a principal diferença entre as duas variedades do /l/ reside, contudo, no modo como os gestos que compõem o segmento se organizam entre si, em função da filiação silábica.

No tocante à produção das laterais, as investigações (Sproat & Fujimura, 1993; Browman & Goldstein, 1995b; Gick, 1999a, 2003; Gick *et alii*, 2006) sugerem que o gesto de ponta da língua tende a ocorrer antes (ou simultaneamente com) do gesto do dorso da língua em posição inicial, enquanto que em Coda se verifica precisamente o inverso. Sproat & Fujimura (1993) atribuem a este efeito de Coda o nome de "*tip delay*", na medida em que a constrição mais anterior sofre um "atraso" temporal em relação à constrição dorsal.

No sentido de justificar a diferença de paradigmas temporais entre consoantes em início e fim de sílaba, Sproat & Fujimura (1993) distinguem os gestos quanto à sua natureza vocálica ou consonântica, com base no grau de constrição: os gestos consonânticos implicam uma oclusão no plano médio-sagital; os vocálicos são produzidos sem constrições significativas à passagem do fluxo

[26] A caracterização de uma lateral alveolar não velarizada ("l claro") implicaria também, segundo Albano (2001), a activação de um gesto na região dorso-faríngea, especificado como `[faríngeo]` e `[médio]`, para além do gesto de ponta da língua. As propostas da linguista brasileira baseiam-se em dados acústicos.

do ar no tracto vocal. O gesto apical do /l/ é um gesto consonântico, ao passo que o gesto dorsal é considerado vocálico [27].

Os gestos consonânticos tendem a ser mais "fortes" em Ataque do que em Coda, enquanto que, no caso dos gestos vocálicos, acontece precisamente o inverso.

Segundo o princípio da "*gestural affinity*", os gestos vocálicos ocorrem mais perto do Núcleo da sílaba, enquanto os consonânticos tendem a aproximar-se da margens silábicas. Assim, em posição inicial de sílaba, o gesto consonantal (a ponta da língua), mais proeminente, precede o gesto vocálico (o dorso da língua), menos proeminente, enquanto que em sílaba final, as laterais são produzidas com um gesto dorsal, mais proeminente, que ocorre antes do gesto apical, mais fraco, com possibilidade de "*undershoot*" (vd. figura 4.20).

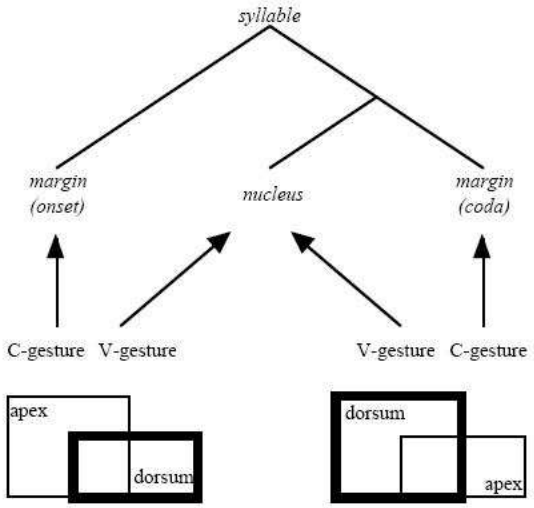


Figura 4.20: Coordenação gestual do /l/, segundo Sproat & Fujimura (1993). Os gestos consonânticos são atraídos para as margens da sílaba; os gestos vocálicos são atraídos para o núcleo da sílaba. As caixas mais altas representam os gestos mais proeminentes (fonte: Carter, 2002).

Este esquema de coordenação temporal encontra-se implementado no TADA (vd. anexo B) com toda a legitimidade, já que a grande maioria dos estudos referidos se reporta ao inglês americano.

A questão que agora se coloca é a de saber qual a viabilidade de estender a referida matriz temporal também ao português. Ao momento, face à ausência de dados sobre este assunto, não será possível outra opção, que não a de fazer algumas inferências gerais, com base nos indícios recolhidos para outras línguas. Obviamente que estas devem ser interpretadas como conjecturas, meras hipóteses teóricas, que os dados haverão ou não de confirmar.

[27] Giles & Moll (1975) haviam já considerado o /l/ final vocálico por natureza: "The post-vocalic type allophones may be regarded as vowel-like for they show relatively slow movement characteristics and undershoot of articulatory positions with increases in the utterance rate [...]. The same undershoot characteristics have not been noted for pre-vocalic allophones, and they seem to be similar to consonants because of their relatively high rate of articulatory movement." (Giles & Moll, 1975, p.223).

O problema é tanto mais difícil de solucionar, quanto as pesquisas de Gick *et alii* (2006), centradas nos efeitos da filiação silábica sobre a coordenação intergestual das líquidas, põem em causa a presunção universalista de Krakow (1999), pelo menos, no que concerne a esta classe de sons. Efectivamente, não existe uma matriz temporal única que possa aplicar-se às seis línguas analisadas, mas é, contudo, possível identificar algumas tendências comuns: 1) em todas as líquidas estudadas foi possível reconhecer e medir dois gestos dorsais em posição pós-vocálica; 2) os padrões de coordenação gestual são, de um modo geral, assimétricos em posição pré e pós-vocálica; 3) a anterioridade, e não o ponto e grau de constrição (cf. Sproat & Fujimura, 1993), é tida como o factor determinante na organização dos gestos, sendo que a constrição mais anterior (e que, normalmente, envolve oclusão total) ocorre mais perifericamente na sílaba, enquanto a menos anterior tende a aproximar-se do Núcleo da sílaba [28]; 4) em posição intervocálica, sempre que estão presentes dois gestos, estes ocorrem simultaneamente.

Com base nestas observações, é possível supor que, em PE, os dois gestos em posição prévocálica têm lugar simultaneamente ou com um pequeno atraso negativo (i.e. o gesto anterior precede o gesto posterior), como acontece para a maioria das línguas (e.g inglês americano ou o sérvio-croata) analisadas por Gick *et alii* (2006).

Por sua vez, em posição de Coda, e seguindo a tendência geral, prevê-se que os dois gestos sejam produzidos com um atraso positivo pronunciado (i.e. o gesto posterior precede o gesto anterior).

No restante, esta matriz corresponde, como já referimos, ao padrão de coordenação previsto no TADA, pelo que não foi necessário proceder a qualquer tipo de alteração significativa. Convém, contudo, ter em mente, o carácter eminentemente teórico e especulativo da nossa argumentação e a necessidade de empreender uma pesquisa experimental para verificação da hipótese formulada.

A consoante [ʎ] ocorre em sílaba medial, estando limitada a um pequeno número de formas em posição inicial de palavra, na sua maioria empréstimos, principalmente do espanhol (e.g. “lhano”, “lhama”).

Quanto ao ponto de articulação, a consoante é tradicionalmente descrita como lateral dorso-palatal (e.g. Mateus *et alii*, 1990, 2005; Barbosa, 1965; Cruz-Ferreira, 1999a).

Contudo, os dados de outras línguas românicas (e.g. catalão, francês, italiano) sugerem uma constrição mais anterior, na zona alveolo-palatal ou mesmo na região alveolar (e.g. espanhol) ou dento-alveolar (catalão de Maiorca) (Recasens & Espinosa, 2006), o que poderá estar relacionado com constrangimentos aerodinâmicos: um ponto de articulação mais anterior favoreceria a formação

[28] A hipótese da anterioridade dá conta da coordenação dos gestos envolvidos na produção da glide [w], sem que haja necessidade de considerar o gesto dos lábios [consonântico] (Gick, 2003). Segundo a proposta de Sproat & Fujimura (1993), os dois gestos deveriam ocorrer simultaneamente, sem “*lag effect*”, na medida em que são ambos [vocálicos]. Os resultados de Gick (2003) indicam, no entanto, que os gestos que compõem o /w/ têm um comportamento semelhante aos do /l/, com o gesto dos lábios a aproximar-se das margens da sílaba.

dos canais laterais atrás dos dentes molares.

Isso mesmo parece indicar o palatograma apresentado por Sá Nogueira (1938, p.54), onde o completo selamento na zona lateral do palato aponta para um escoamento do fluxo oral através de canais formados na região pós-palatal/ velar. Mais do que palatal, o ponto de articulação é alveolo-palatal.

Os dados de ressonância magnética apontam também neste sentido, com a área de contacto a estender-se desde o pré-palato até à região dental (figura 4.21).

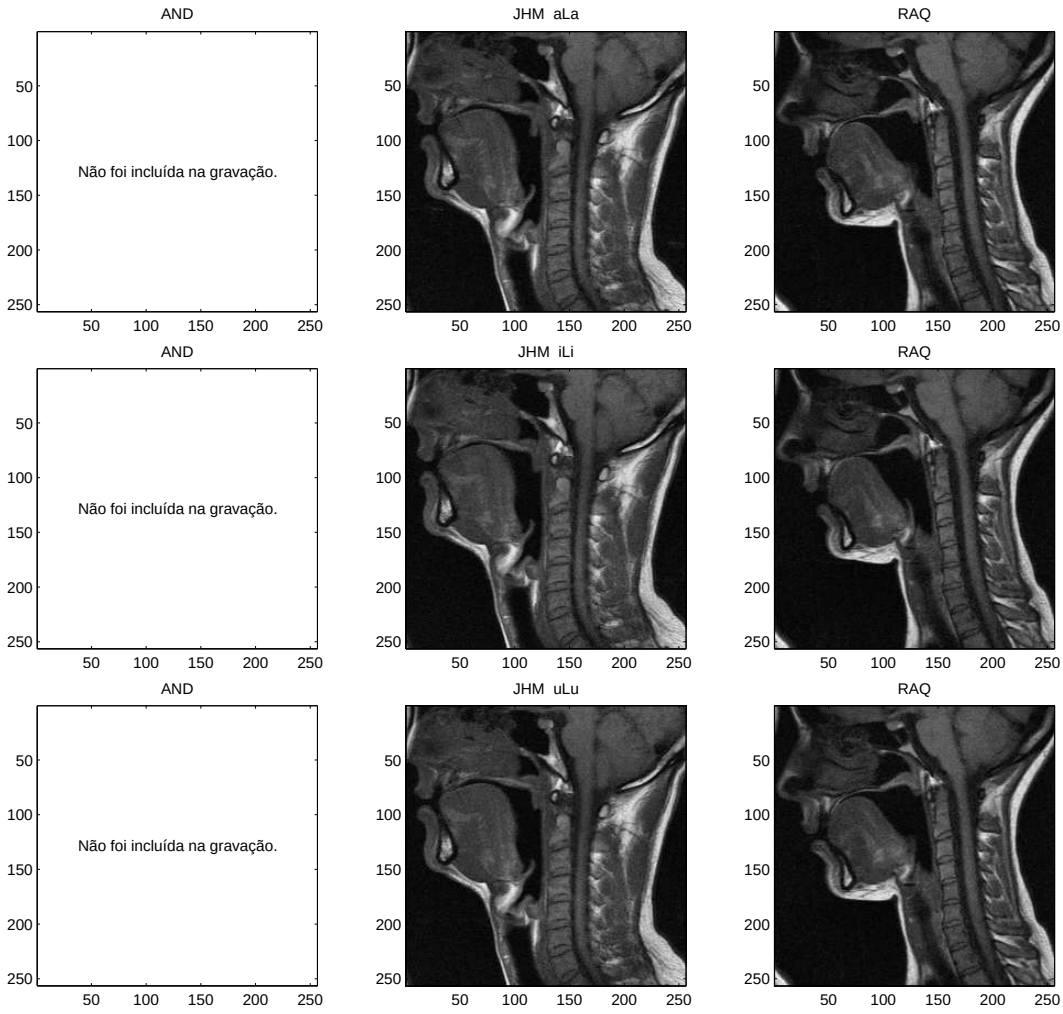


Figura 4.21: Perfis articulatórios dos sujeitos JHM e RAQ, durante a produção da consoante [ʎ], no contexto de [a] (em cima), [i] (ao centro) e [u] (em baixo). O *corpus* gravado pelo informante AND, numa sessão independente, não incluiu este contexto.

Na definição gestual, a especificação do *target* `[palatal]` (`[PAL]`) da variável TBCL para os 85 mm traduz a referida anterioridade do gesto de corpo da língua, salientando, ao mesmo tempo, as diferenças em relação ao [ɲ] (vd. secção 4.2.5.2).

O grau de constrição foi fixado como `[estreito]` (`[NAR]`), como convém a uma lateral

aproximante.

Em todo o caso, ao contrário da lateral alveolar, em que a especificação da dupla constrição, sem qualquer referência às passagens laterais, ajuda a recriar um efeito acústico próximo do /l/, é impossível simular o [ʎ] - em relação ao qual é forçado prever uma segunda articulação (cf. Albano, 2001) - apenas com base na descrição supramencionada, sem qualquer alusão à *forma de constrição*. Esta última variável, embora prevista no modelo gestual, nunca chegou a ser incorporada à FA, nem tão pouco ao TADA.

Em face desta limitação, a formulação da tabela 4.11 - onde foi incorporada, para uma descrição satisfatória do [ʎ], a variável *tongue body shape* (TBSH), especificada como `[lateral]` (`[LAT]`) e associada a um oscilador *sh* - não passa de uma proposta teórica e, portanto, impossível de testar no estado actual do sistema.

Se no caso particular do [ʎ], a incorporação de uma variável dessa natureza nos parece absolutamente indispensável no sentido de criar o efeito acústico pretendido, ela possibilitaria também um modelamento mais satisfatório das laterais, de uma forma geral.

Tabela 4.11: Gestos associados às consoantes laterais do PE. As consoantes encontram-se foneticamente representadas em AFI e os gestos a elas associados são caracterizados pelo articulador (*Organ*), oscilador (*Osc*), variável do tracto (*TV*), constrição (*Const*). O *target* da constrição e o *stiffness* encontram-se pré-definidos num ficheiro independente (“.”), mas podem ser directamente especificados no dicionário.

Vogal	Organ	Osc	TV	Const	Target	Stiff
l	TT	nar	TTCL	ALV	.	.
	TT	nar	TTCD	NAR	.	.
	TT	rel	TTCL	REL	.	.
	TT	rel	TTCD	REL	.	.
	TB	voc	TBCL	VEL	.	.
	TB	voc	TBCD	NAR	.	.
ʎ	TB	nar	TBCL	PAL	85	.
	TB	nar	TTCD	NAR	.	.
	TB	rel	TBCD	REL	.	.
	TB	sh	TBSH	LAT	.	.

4.2.5.5 Consoantes vibrantes

Do inventário do PE fazem também parte outras duas consoantes, classificadas, segundo a tradição portuguesa, como vibrantes: uma vibrante simples alveolar /ɾ/ (*tap* ou *flap* nos termos do inglês) e uma vibrante múltipa /r/ (designada, no inglês, por *trill*), que unicamente se opõem em posição intervocálica (e.g. “carro” vs “caro”) (Mateus & d’Andrade, 2000).

Como não são conhecidos estudos articulatórios relativos à produção das vibrantes em PE - nem sequer dados obtidos a partir de RM - para a caracterização gestual destes sons, valemo-nos

dos relatos impressionistas dos autores portugueses, dos escassos dados acústicos publicados (Jesus & Shadle, 2005) e, sobretudo, da comparação com línguas aparentadas (nomeadamente o espanhol).

Os dados coligidos são unânimes em considerar que a vibrante simples resulta de um único movimento balístico do ápice da língua contra a região dento-alveolar, enquanto a congénere múltipla é produzida em resultado da vibração de um articulador - seja a ponta da língua ou a úvula - que tem lugar sempre que as condições aerodinâmicas estejam reunidas.

A descrição articulatória da vibrante simples coincide com aquilo a que outras línguas chamam de *tap* ou *flap*. Estes dois tipos de sons distinguem-se entre si pelo tipo de movimento efectuado pelo ápice da língua, ainda que muitos linguistas não cheguem a fazer distinções entre as duas designações, tendo em conta a subtileza das diferenças (Ladefoged & Maddieson, 1995, p. 230-231) [29].

A maioria dos autores portugueses classifica a vibrante simples do PE como vozeada, ainda que sejam conhecidas algumas referências a variantes surdas (Sá Nogueira, 1938; Andrade, 1996) [30]. Para além disso, está descrita uma variante fricativizada desta consoante, que ocorre, sobretudo, em posição de final de palavra, cujas características principais são 1) a perda do contacto alveolar e 2) a geração de ruído de baixa amplitude (Jesus & Shadle, 2005).

Em português, a vibrante simples pode ainda, formar, em conjunto com uma oclusiva ou uma fricativa - [v] ou [f] - um Ataque silábico complexo. Nestes casos, é possível observar, entre a obstruinte e a vibrante, um segmento com características similares a uma vogal breve.

A vibrante múltipla é composta por duas ou mais vibrações (ou ciclos), estando este número relacionado com factores como a proeminência prosódica ou o estilo de fala (Mota, 1991b,a; Blecua, 1999). Para além disso, ao contrário do *tap*, a vibrante múltipla está associada um abaixamento pronunciado do pré-dorso e a um movimento de retracção da raiz da língua (Recasens, 1991; Recasens & Pallarès, 1999). Deste ponto de vista, o *trill* apresenta muitas semelhanças com a lateral velarizada.

Adicionalmente, durante a produção do /r/, o posicionamento da língua é rigorosamente controlado, o que se traduz numa grande resistência à coarticulação com as vogais vizinhas.

Estes factos levam muitos autores (Recasens, 1991; Catford, 1977; Bradley, 2001) a considerar que o *trill* não pode ser interpretado como uma simples sequência de *taps*, visto que as duas consoantes são realizadas mediante mecanismos de produção distintos [31].

> *A flap ... is a single ballistic flick or hit-and-run gesture. A trill ... is a maintained and prolongable posture: the vibrations that occur in a trill are aerodynamically imposed on*

[29] Segundo Ladefoged & Maddieson (1995, p. 230-231), o *flap* caracteriza-se por um movimento tangencial do articulador activo em direcção ao local de contacto, enquanto o *tap* é realizado, movendo o articulador directamente contra o palato. A consoante dental/alveolar da palavra espanhola ['kaɾu] seria um *tap*, enquanto a pronúncia das oclusivas dentais em posição intervocálica no inglês americano (e.g. *city*) corresponderia a um *flap*.

[30] Andrade (1996) refere que os quatro informantes do seu estudo desvozeiam o segundo rótico da palavra [ɐpiɾtɐˈɾa].

[31] Segundo Recasens & Pallarès (1999), as diferenças entre vibrantes simples e múltiplas, quer ao nível dos gestos implicados na sua produção, quer no que respeita à coarticulação, não permitem suportar a tradicional análise fonológica, segundo a qual o *trill* é tratado como uma versão geminada do /ɾ/, resultando da aplicação de uma série de regras (Mateus & d'Andrade, 2000, p.15-16).

> *the posture. Any idea that a trill is a "rapid series of flaps", or that a flap is just an "ultra-short trill" is quite wrong.* (Catford, 1977, apud Bradley, 2001, p.107)

O mecanismo de vibração associado ao *trill* é similar ao envolvido no processo de vibração das pregas vocais (Ladefoged & Maddieson, 1995) e, ao contrário do *tap*, que envolve um movimento muscular activo da língua, é causado por forças aerodinâmicas.

Para que este mecanismo de vibração (*trilling*) se inicie (e mantenha) é necessário um controlo muito preciso, não só do posicionamento dos articuladores, mas também das condições de *stiffness* e de pressão (Catford, 1977; Ladefoged & Maddieson, 1995; Barry, 1997; Solé, 2002) [32].

O processo de produção das vibrantes múltiplas é descrito por Ladefoged & Maddieson (1995) na citação que passamos a transcrever:

> *The primary characteristic of a trill is that it is the vibration of one speech organ against another, driven by the aerodynamic conditions. One of the soft moveable parts of the vocal tract is placed close enough o another surface, so that when a current of air of the right strength passes through the aperture created by this configuration, a repeating pattern of closing and opening of the flow channel occurs. (...) In its essentials this is very similar to the vibration of the vocal folds during voicing; in both cases there is no muscular action that controls each single vibration, but a sufficiently narrow aperture must be created and an adequate airflow through the aperture must occur. The aperture size and airflow must fall within critical limits for trilling to occur, and quite small deviations mean that it will fail. As a result, trills tend to vary with non-trilled pronunciations.* (Ladefoged & Maddieson, 1995, p.217).

Assim sendo, é perfeitamente natural que o *trill* esteja entre os sons que mais problemas de aprendizagem suscitam, quer se trate de falantes nativos, quer de estrangeiros a aprender uma segunda língua. A bibliografia sobre a aquisição da fonologia revela que as líquidas emergem tardiamente na produção e, de entre estas, o /r ʀ/ é mesmo a última classe a estabilizar (Freitas, 1997; Vihman, 1996).

Neste sentido, as vibrantes são particularmente sensíveis a pequenas variações das condições articulatórias e aerodinâmicas: o mínimo desajuste dos referidos parâmetros impede, por norma, a ocorrência da dita vibração (cf. Ohala *et alii*, 1998a; Solé, 2002). Em resultado, as vibrantes tendem a alternar, sincronica e diacronicamente, com outro tipo de realizações, em que não se chega a produzir *trilling*, nomeadamente *taps*, aproximantes e fricativas.

À semelhança de outras línguas, também no PE se observam realizações da vibrante múltipla, em que a típica vibração simplesmente não acontece. A vibrante múltipla alveolar [r] coexiste

[32] Ladefoged & Maddieson (1995, p. 218) destacam também o papel da massa do articulador, enquanto elemento facilitador do mecanismo vibratório: "Trills are much more easily produced if the vibrating articulator has relatively small mass, hence the most common trills involve the tongue tip vibrating against a contact point in the dental/alveolar region, or the uvula vibrating against the back of the tongue.

com a vibrante múltipla uvular /ʀ/, que se encontra actualmente muito difundida em todo o território nacional (Emiliano, 2006), a par com a fricativa uvular surda ([χ]) ou sonora ([ʁ]).

Segundo Mateus & d'Andrade (2000, p.11), este último alofone estará mesmo em vias de se tornar dominante na região de Lisboa. Essa parece ser também a interpretação de Cruz-Ferreira (1995), que, na sua descrição do PE, com base do português de Lisboa, apenas considera a fricativa uvular. A mesma proposta, suportada em dados acústicos, é defendida por Jesus & Shadle (2005), mas, desta feita, para o PE em geral.

No caso em que a realização mais comum do rótico parece ser a fricativa uvular, podemos supor que não há qualquer contacto, mas apenas uma aproximação entre os articuladores (úvula e região posterior da língua), acompanhada de geração de fricção (Ladefoged & Maddieson, 1995). As razões que motivam a ausência de vibração poderão estar relacionadas com a redução da pressão subglotal, posicionamento inadequado dos articuladores ou aumento da tensão da úvula (cf. Solé, 2002). Não estando reunidas todas as condições articulatórias e aerodinâmicas necessárias à ocorrência de vibração, o resultado é uma fricativa.

A caracterização e simulação das vibrantes no âmbito do TADA acarreta sérias dificuldades relacionadas, não só com a falta de dados que permitam sustentar a nossa proposta gestual, mas sobretudo com aquelas que parecem ser as reais limitações do próprio modelo computacional.

Com efeito, as trajectórias dos articuladores são modeladas a partir de uma equação dinâmica simples, do tipo massa-mola, com amortecimento crítico, o que significa que a massa não oscila en torno do alvo, mas apenas se aproxima dele.

Uma da soluções para lidar com as vibrantes do português passaria então, conforme sugerido por Albano (2001, p.13-131), por introduzir algumas alterações na equação dinâmica, de modo a considerar um amortecimento menor do que o denominado *crítico*, que actualmente caracteriza os gestos do TADA. Esta mesma abordagem é sugerida por Browman & Goldstein (1990a, 1992) para o tratamento dos *flaps*.

Segundo a mesma autora, para além de dois graus de amortecimento não crítico, necessários para distinguir a vibrante simples da múltipla, uma correcta descrição desta classe de sons implicaria ainda, à semelhança das laterais, a referência a duas constrições: "uma constrição faríngea é obrigatória para todos os róticos, podendo ser acompanhada de uma constrição coronal, como nas vibrantes alveolares, ou de uma constrição dorsal, como nas vibrantes uvulares" (Albano, 2001, p.130-131).

Conforme vimos em parágrafos anteriores, há dados articulatórios (Recasens, 1991; Recasens & Pallarès, 1999) - embora não dedicados ao português em particular - que permitem sustentar a obrigatoriedade de um gesto faríngeo na representação gestual das vibrantes múltiplas alveolares. Este teria como objectivo específico facilitar o *trilling*.

Apesar da falta de estudos de produção, é também perfeitamente plausível aceitar, com base no referenciado para o *trill* alveolar, uma especificação posterior obrigatória em relação à variante

uvular [33].

Assumir uma dupla especificação articulatória para os demais róticos - nomeadamente a vibrante simples alveolar - parece-nos, contudo, um pouco mais problemático, na medida em que esta interpretação não encontra suporte empírico nos estudos articulatórios (Recasens, 1991; Recasens & Pallarès, 1999) e, sobretudo, não reflecte as diferenças de produção - já aqui devidamente apontadas - subjacentes ao *tap* e ao *trill*.

Ainda que, durante a produção do *tap* tenha sido detectada uma certa depressão do pré-dorso da língua, que eventualmente não se verifica para as demais alveolares, o movimento não chega a atingir a dimensão observada para o *trill*, cujo processo de produção parece ser muito similar ao da lateral velarizada (Recasens & Pallarès, 1999), em relação à qual considerámos, efectivamente, uma dupla articulação.

Face às informações disponíveis ao momento, para efeitos de caracterização gestual da vibrante simples, optámos por fazer referência exclusivamente ao gesto de ponta da língua (vd. tabela 4.12), pelo menos até que se encontrem motivos suficientes, de preferência ancorados em novos dados articulatórios especificamente dedicados ao PE, para sustentar a dupla especificação gestual do *tap*.

Adicionalmente, tendo em conta o carácter inerentemente curto desta consoante, a duração do movimento da variável do tracto associada ao gesto foi diminuída (vd. tabela 4.12), mediante um aumento do valor do *stiffness*. Teoricamente, para atingir este mesmo fim, poderia igualmente optar-se por fazer acompanhar esta alteração do *stiffness* por uma diminuição dos valores da *damping ratio*, de modo a permitir, por um lado, a oscilação da massa em torno do alvo e, controlando, por outro, a frequência de oscilação (Browman & Goldstein, 1990a).

A validade desta última via permanece, contudo, por testar, já que, até onde nos foi possível averiguar, o TADA não permite a modificação dos valores da *damping ratio*. Independentemente disso, a representação da vibrante simples como um gesto oscilatório não está muito de acordo com as características articulatórias da consoante: já aqui mencionámos que o *tap* resulta de um movimento muscular activo e não da acção de forças aerodinâmicas como no caso do *trill* [34].

Ainda que não possa ser testada em virtude da referida limitação do sistema, a nosso ver, a interpretação gestual das vibrantes - enquanto constelação formada obrigatoriamente por dois gestos, estando um deles associado a um amortecimento menor que crítico (Albano, 2001) - será, portanto, mais adequada para descrever o *trill* do que o *tap*, nomeadamente o *trill* alveolar, já que o caso da variante uvular nos parece bem mais difícil de acomodar à referida representação gestual.

Com efeito, Albano (2001) preconiza para a vibrante uvular dois gestos, um gesto faríngeo, comum aos demais róticos, e um gesto dorsal, sendo que, em relação a este último, a autora defende a

[33] Ladefoged & Maddieson (1995, p.225) refere, com base nos estudos radiológicos de Delattre (1971) para o francês e o alemão, que as vibrantes uvulares são produzidas "by an initial movement of the tongue root backwards followed by an upward movement toward the uvula, which is also moved forward to a position where trilling can occur.".

[34] A sermos rigorosos, as características articulatórias do *tap* não justificam sequer o termo "vibrante", tradicionalmente usado, em português, para denominar esta classe de consoantes (cf. Mateus & Rodrigues, 2003).

necessidade de se proceder a uma reformulação dos valores da *damping ratio*, de modo a promover o movimento oscilatório do articulador. Independentemente da legitimidade em considerar uma dupla especificação gestual, quando não existem muitos indícios que a fundamentem, cabe notar que, no caso particular das vibrantes uvulares, é a úvula que entra em vibração e não o dorso da língua, ainda que, para que tal aconteça, seja, efectivamente, imprescindível um movimento deste último articulador (cf., p.225 Ladefoged & Maddieson, 1995).

Em nossa opinião, uma representação gestual mais adequada da consoante em causa (e das vibrantes em geral), implicaria, tão somente, a criação das condições articulatórias adequadas à vibração da úvula, mais concretamente a especificação de um gesto dorsal, fechado, na zona da úvula, eventualmente acompanhado de um gesto faríngeo. Este movimento activo do corpo da língua estará na origem do movimento vibratório - a par das já referidas condições aerodinâmicas -, enquanto a constrição faríngea visaria apenas subsidiar um posicionamento mais preciso dos articuladores.

O rigoroso cumprimento dos restantes requisitos necessários à vibração da úvula - nomeadamente o rigoroso controlo da pressão subglotal, que irá desencadear o efeito de Bernoulli - deverá ser assegurado pelo modelo articulatório em si mesmo, que, para simular o efeito de *trilling*, terá obrigatoriamente de possuir, como principal característica, uma grande elasticidade dos articuladores, sob pena da vibração não ser possível. Uma abordagem deste género foi já adoptada no sintetizador articulatório desenhado por Boersma (1995), capaz de simular vibrantes múltiplas com qualquer ponto de articulação à custa de um processo semelhante ao usado para replicar o mecanismo de vibração das pregas vocais.

Nenhum dos sintetizadores ao nosso dispor é dotado dessa capacidade, pelo que a validação desta proposta não está ao nosso alcance.

Nos casos em que o /r/ é produzido como uma fricativa - ao que parece a realização mais usual do /r/, pelo menos na região de Lisboa - é cabível supor que a configuração gestual não se altera substancialmente, a não ser no que respeita ao grau de constrição do gesto dorsal, que ao passar de fechado a crítico, fará diminuir os níveis de pressão subglotal, tendo como consequência provável a geração de ruído, em vez do *trilling*.

Todas as observações feitas nos parágrafos anteriores acerca da vibrante múltipla e respectivas manifestações carecem de suporte experimental, pelo que nos parece absolutamente necessário prosseguir com a investigação, antes de adiantar alguma resposta definitiva sobre a sua representação gestual. Como tal, tendo em conta o carácter eminentemente teórico e especulativo da nossa argumentação e todas incertezas que envolvem a caracterização gestual da vibrante múltipla, optámos por não a incluir na tabela gestual 4.12 referente às vibrantes do PE.

Tabela 4.12: Gestos associados ao *tap*. A consoante encontra-se foneticamente representada em AFI e os gestos a ela associados são caracterizados pelo articulador (*Organ*), oscilador (*Osc*), variável do tracto (*TV*), constrição (*Const*). O *target* da constrição e o *stiffness* encontram-se pré-definidos num ficheiro independente ("."), mas podem ser directamente especificados no dicionário.

Vogal	Organ	Osc	TV	Const	Target	Stiff
ɾ	TT	clo	TTCL	ALV	.	5
	TT	clo	TTCD	CLO	.	5
	TT	rel	TTCL	REL	.	5
	TT	rel	TTCD	REL	.	5

4.3 Primeira avaliação perceptiva da proposta

Definidos os gestos associados a cada um dos segmentos do PE, impõe-se uma validação da proposta apresentada. Esta passou pela aplicação de um teste perceptual para avaliar a inteligibilidade dos segmentos sintetizados - quer pelo SAPWindows quer pelo HLsyn - a partir das configurações gestuais indicadas. O recurso a um segundo sintetizador, para além do SAPWindows, durante a avaliação, prende-se essencialmente com as limitações deste último no tocante ao tratamento de parte das classes de sons do PE.

Tendo em conta a natureza dos sintetizadores em causa - lembramos que a síntese articulatória é, porventura, a técnica mais promissora, mas ao mesmo tempo a menos desenvolvida - a única opção possível passou um teste perceptivo de inteligibilidade, estando os testes de qualidade - mais adequados para sistemas de síntese de outro tipo, completamente inteligíveis do ponto de vista segmental, mas a necessitar de uma implementação mais satisfatória ao nível da naturalidade - reservados para uma fase posterior do desenvolvimento.

A concretização do objectivo enunciado - averiguar a pertinência da nossa proposta gestual - esteve, contudo, sujeita a uma série de constrangimentos, que passaremos a explicitar em seguida, e que se reflectiram, tanto no desenho do próprio teste como nos resultados do mesmo:

- Limitações do sintetizador SAPWindows - tal como referido anteriormente, o SAPWindows não está preparado para produzir grande parte dos sons do PE. Originalmente vocacionado para a síntese dos sons nasais, o sistema é actualmente capaz de gerar vogais orais e consoantes oclusivas. No caso destas últimas, persistem alguns problemas no modelamento da excitação glotal. Apesar das evoluções recentes (Teixeira *et alii*, 2005), também no tocante às fricativas, o sintetizador apresenta margem para evoluções. Como consequência, apenas uma pequena parte do *corpus* foi processada com recurso a este sintetizador.

- Limitações do TADA

  – Ausência de informação prosódica - ainda que, na transcrição fonética gerada automaticamente, a localização do acento esteja explicitamente assinalada através de um número

colocado a seguir à vogal, esta informação não é usada na versão actual do TADA. Para além disso, o sistema é totalmente desprovido de informação prosódica, embora seja possível editar, à *posteriori*, as pautas gestuais, de modo a controlar os gestos de F0 e os chamados "pi-gestures" (Byrd & Saltzman, 2003). A falta de entoação e ausência de acento reflecte-se inevitavelmente na qualidade e naturalidade do *output*, que adquire um tom metálico e monocórdico, dificultando em muito a identificação das palavras.

- Impossibilidade de testar algumas configurações gestuais - como foi já adiantado, face às limitações actuais do TADA, não foi possível testar duas das configurações gestuais propostas: no caso do [ʎ], o modelo não incorpora a variável do tracto "forma de constrição", pelo que as passagens laterais não podem ser simuladas; uma das possibilidades para caracterizar o [r] alveolar implicaria assumir um amortecimento menor do que o actualmente previsto no sistema (Albano, 2001), mas, até onde nos foi dado perceber, alterações a este parâmetro não são actualmente permitidas. Neste sentido, foram excluídos do *corpus* de teste todos os estímulos com vibrantes múltiplas e laterais palatais.
- Vogais nasais - as vogais nasais não fazem parte do inventário fonológico do inglês e, como tal, estão ausentes do TADA. Projecta-se para breve a integração das vogais nasais no sistema, ficando reservada para o capítulo seguinte a descrição e avaliação de todos os esforços empreendidos nesse sentido. Na fase actual do desenvolvimento, as constelações gestuais subjacentes às vogais nasais estão ainda por definir, pelo que estas não fazem parte dos estímulos seleccionados.

- Sincronização temporal entre gestos - a adopção dos padrões de coordenação propostos para o inglês, em virtude da escassez de estudos articulatórios imporá, necessariamente, limitações à avaliação da nossa proposta gestual, na medida em que será difícil, senão impossível, aferir qual o peso relativo de uma coordenação deficiente nos resultados finais de uma simulação.

As referidas limitações fazem do processo de validação da proposta gestual uma tarefa arriscada, mas ainda assim necessária, porquanto permitirá identificar os principais problemas subjacentes à caracterização gestual do inventário fonémico do PE e ter uma ideia geral do funcionamento da infraestrutura.

4.3.1 Estímulos

Os cinquenta estímulos incluídos no teste perceptual de identificação foram seleccionados aleatoriamente a partir de uma pequena amostra do *corpus* PF (Nascimento *et alii*, 1987) [35], previamente processada de modo a excluir todas as palavras com ditongos, vogais nasais, vibrantes múltiplas, laterais palatais e grupos consonânticos.

[35] Referimo-nos à mesma amostra usada no capítulo 3, para avaliação dos algoritmos de silabificação automática.

No tocante ao número de sílabas, 50% das palavras do *corpus* de teste são dissilábicas, enquanto 30% correspondem a palavras com três sílabas e apenas 20% têm um número igual ou superior a quatro sílabas. Considerando o reduzido número de formas monossilábicas constantes da lista do PF [36], optámos também por eliminar do nosso *corpus* as palavras com apenas uma sílaba.

Quanto à estrutura silábica dos estímulos, 80% do material linguístico seleccionado é constituído por sílabas abertas, com ou sem Ataque preenchido - isto é estímulos CV ou simplesmente V - ao passo que os casos de palavras com sílabas fechadas perfazem 20% da totalidade do *corpus* de teste.

A tabela 4.13 inclui a lista de estímulos escolhidos para o teste de identificação, de acordo com o tamanho e a estrutura silábica das palavras.

Tabela 4.13: Estímulos seleccionados para o teste de inteligibilidade, de acordo com o número de sílabas e o tipo de sílaba (aberta ou fechada). Apresenta-se ainda informação relativa ao(s) sintetizador(es) usados para produzir os estímulos.

Sílabas	Tipo Sílaba	Num	Palavras	HLSyn	SAP
2	Aberta	9	olá, bica, lago, louro, bota, linha, duro, pato, mapa	X	X
		10	doze, sofá, nove, louça, sete, seco, você, vinho, vosso, tosse	X	
	Fechada	1	arte	X	X
		5	desde, polvo, virar, secar, uvas	X	
3	Aberta	8	cebola, músico, bagaço, assado, tesoura, desejo, vizinho, desenho	X	
		5	mínimo, agora, bocado, macaco, medida	X	X
	Fechada	2	vomitar, vistoso	X	
≥ 4	Aberta	5	chocolate, dezassete, camisola, cerâmica, velocidade	X	
		3	naturalidade, política, unidade	X	X
	Fechada	1	analisar	X	
		1	gabardine	X	X

4.3.2 Construção do teste

Seleccionados os estímulos, optámos por introduzir as palavras directamente no dicionário - associando-lhes, de forma manual, a respectiva representação fonética e silábica, de acordo com os requisitos do TADA - em vez de recorrer aos programas que asseguram a conversão grafema-fone e divisão silábica (vd. capítulo 3), evitando assim possíveis erros decorrentes dos procedimentos automáticos.

[36] De acordo com Vigário (2003, p.159), das sete mil formas flexionadas do PF, apenas 138 palavras (lexicais) são monossilábicas, e destas apenas 28 constituídas por sílaba aberta.

Tendo em mente a possibilidade de realização de outros testes perceptuais, com novos estímulos, todos os procedimentos conducentes à geração das pautas gestuais no TADA foram automatizados. Estes consistem, genericamente, em 1) criação da lista de gestos associados a cada um dos elementos de entrada, especificados quanto aos parâmetros dinâmicos, tipo de oscilador e posição silábica ; 2) especificação dos parâmetros do oscilador, a par com as fases de activação dos gestos e tipos de coordenação estabelecidas entre eles. De seguida, procedeu-se à criação das pautas gestuais para cada uma das entradas do *corpus*, geração das trajectórias dos articuladores e respectivos *outputs* para os sintetizadores.

Todos os estímulos foram sintetizados com recurso ao HLsyn, mas o SAPWindows, em virtude das suas limitações, foi somente usado para produzir versões sintetizadas de dezanove das cinquenta palavras do *corpus* de teste.

Para a construção do teste de identificação, e uma vez preparados os estímulos, utilizou-se um programa, em Tcl/Tk (Teixeira & Vaz, 2000), especialmente adaptado para este fim. Uma das principais vantagens desta plataforma reside na geração automática dos resultados do teste, logo após a sua conclusão, de modo a que estes possam ser imediatamente exportados para o programa de tratamento estatístico, neste caso o *Statistical Package for the Social Sciences* (SPSS).

A interface concebida e ilustrada na figura 4.22 permitiu, para além da identificação do interveniente no teste, uma monitorização da progressão do mesmo, mais concretamente do número de estímulos já avaliados pelo ouvinte. Em se tratanto de um teste perceptual de resposta aberta (*open response test*) - concebido para que os ouvintes identifiquem o estímulo sem qualquer tipo de condicionamento ou orientação - a referida interface dispunha ainda de um local para registo da resposta dos ouvintes à pergunta “Qual a palavra que ouviu?”.

Também a repetição do estímulo foi possível, bastando, para isso, pressionar o botão “Ouvir novamente”.

A apresentação das palavras obedeceu a uma ordem aleatória ditada pelo programa informático que sustenta o teste. Os sete primeiros serviram apenas de treino, tendo como objectivo a habituação dos participantes ao teste em geral, e à voz sintética em particular. Depois de identificado um estímulo e registada a resposta em local próprio, o computador passou automaticamente à reprodução da palavra seguinte.

4.3.3 Aplicação do teste

Participaram no teste de percepção nove indivíduos, cinco do sexo feminino e quatro do sexo masculino, quase todos naturais e residentes na zona norte do País, na sua grande maioria com habilitações literárias iguais ou superiores ao Mestrado, sem história conhecida de perturbações auditivas. Todos os elementos recrutados para a realização da experiência foram informados dos objectivos do estudo e aceitaram participar no teste.

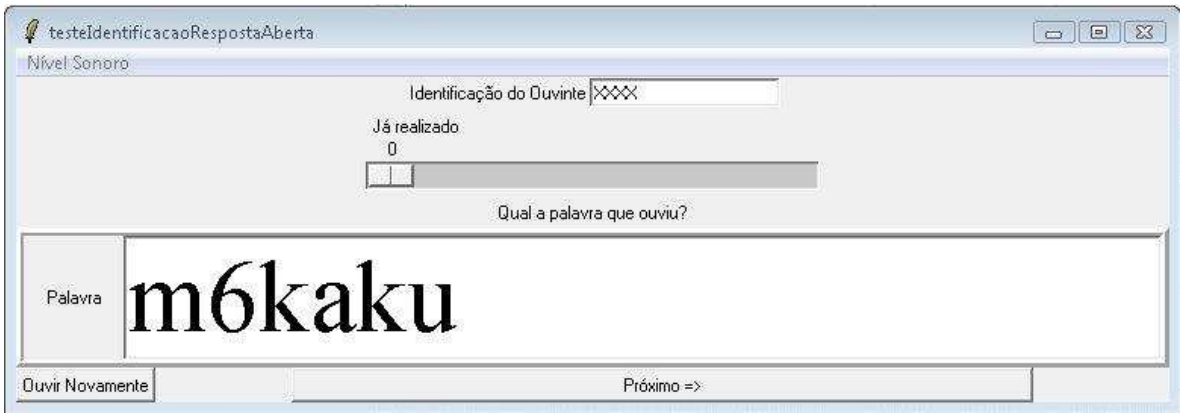


Figura 4.22: Interface gráfica do teste perceptivo de identificação. Para além da identificação do ouvinte (em cima), a interface dispõe de uma barra móvel ("Já realizado") para monitorização dos estímulos avaliados (ao centro), um botão para ouvir novamente os estímulos (em baixo à esquerda) e uma zona (ao centro) para registar a resposta do ouvinte à pergunta "Qual a palavra que ouviu?".

Este foi aplicado individualmente, no Instituto de Engenharia Electrónica e Telemática de Aveiro (IEETA), num gabinete com um nível de ruído baixo a moderado, sob a supervisão da experimentadora.

Pediu-se aos ouvintes que identificassem um estímulo, apresentado aos sujeitos através de auscultadores. Na explicação da tarefa, o experimentador chamou a atenção para algumas características dos estímulos, nomeadamente os problemas ao nível do acento e da prosódia, em grande parte responsáveis pela falta de naturalidade das palavras.

A repetição dos estímulos (mediante recurso ao referido comando "Ouvir novamente"), sempre que o participante do teste sentiu essa necessidade, ficou a cargo da experimentadora. Esta foi também responsável pelo registo escrito das respostas dos sujeitos, em transcrição fonética (usando SAMPA) e em local próprio da interface.

Não foi imposto nenhum limite de tempo para a realização do teste perceptual, que, em média, durou cerca de vinte minutos para cada ouvinte.

4.3.4 Resultados

Nesta secção, proceder-se-á à apresentação dos resultados do teste perceptual, tendo em conta aspectos como: taxa de acerto global; percentagem de identificação ao nível do segmento; eventual interferência de variáveis como número de sílabas e estrutura silábica do estímulo; possíveis diferenças de desempenho entre os dois sintetizadores.

Tal como seria de esperar, as várias limitações do modelo linguístico e dos sintetizadores articulatórios em si mesmos traduzem-se numa percentagem de erros elevada: apenas 25.3% dos estímulos são correctamente identificados, na sua totalidade. Ainda no tocante à taxa de acerto ao nível da palavra, verifica-se uma diferença entre os dois sintetizadores, sendo os resultados do SAPWindows,

em média, superiores aos do HLsyn (33.3% versus 22.2%, respectivamente). De acordo com o teste *t*, esta diferença é estatisticamente significativa (p= 0.002).

Quanto ao desempenho individual dos sujeitos durante a realização da tarefa, o gráfico 4.23 indica que os valores variam entre os 14.6% e os 44.6% de estímulos correctamente reconhecidos.

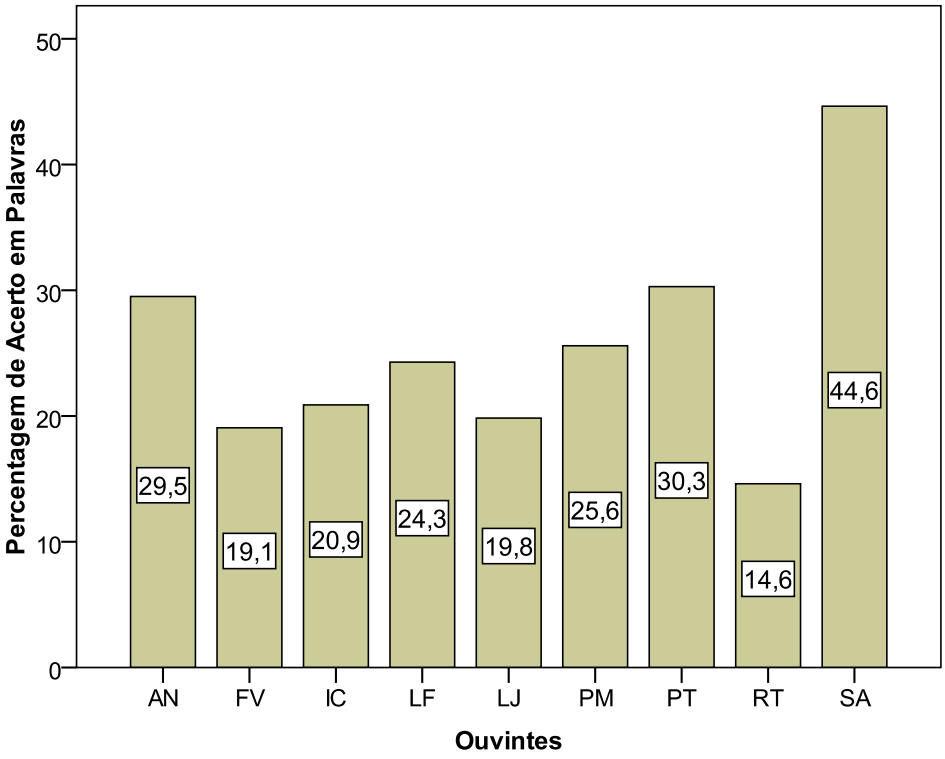


Figura 4.23: Percentagem de palavras correctamente identificadas por cada um dos ouvintes participantes no teste de identificação.

A percentagem de acerto foi mais elevada sempre que estiveram em causa palavras como "mínimo" [ˈminimu] (94.4%), "camisola" [kɐmiˈzɔlɐ] (77.8%), "olá" [ɔˈla] (77.8%), "sofá" [suˈfa] (77.8%), "chocolate" [ʃukuˈlatɨ] (77.8%), "pato" [ˈpatu] (77.8%) e "mapa" [ˈmapɐ] (66.7%). Entre os estímulos mais difíceis de identificar pelos ouvintes que realizaram o teste estão palavras como "virar" [viˈraɾ], "vosso" [ˈvɔsu] ou "seco" [ˈsɐku].

Considerou-se também a taxa de acerto ao nível segmental. Na figura 4.24 , é apresentada, sob a forma de um gráfico de barras, a percentagem de respostas certas em função da classe de som envolvida e do sintetizador em causa.

Como facilmente se pode concluir a partir da observação do gráfico, para ambos os sintetizadores, a percentagem de respostas correctas é mais elevada para as vogais (orais) e as consoantes nasais, atingindo valores próximos dos 70%. Estes resultados estão em perfeita consonância com a opinião geral dos próprios participantes, que, no decorrer do teste, de forma sistemática, assinalaram a facilidade em identificar as vogais, por oposição às consoantes.

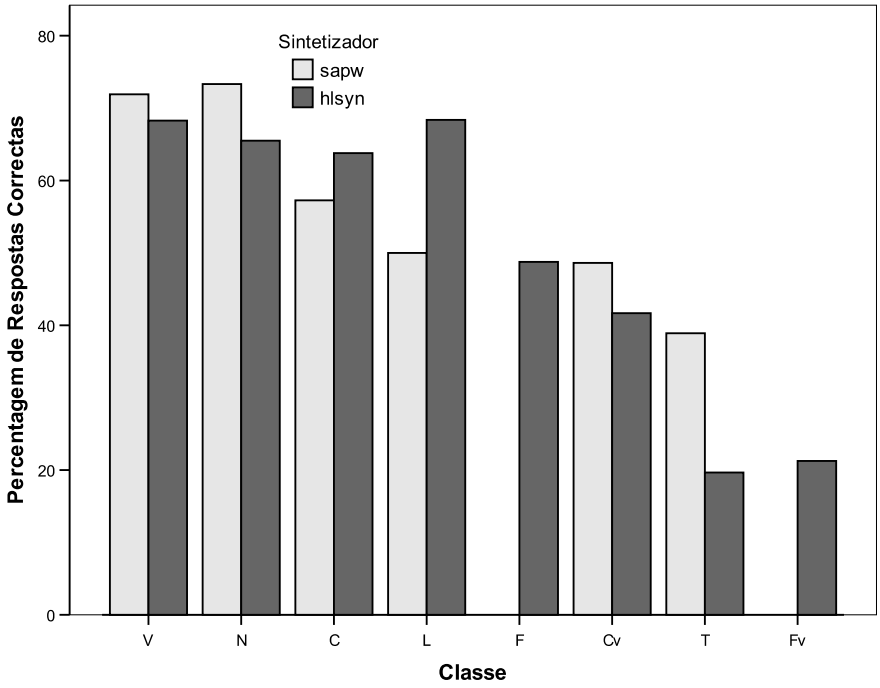


Figura 4.24: Percentagem de respostas correctas em função da classe de sons e do sintetizador utilizado (HLsyn, a cinzento escuro, ou SAPWindows, a cinzento claro). As classes de sons são: vogais (V), consoantes nasais (N), oclusivas surdas (C), laterais (L), fricativas surdas (F), oclusivas sonoras (Cv), vibrantes simples (T) e fricativas sonoras (Fv).

Quanto às laterais (neste caso, apenas o /l/), os resultados dependem muito do sintetizador usado para gerar os estímulos: as taxas de acerto para o HLsyn não só são superiores às do SAPWindows, como chegam a ombrear com os valores obtidos para as consoantes nasais e as vogais.

É em relação às fricativas vozeadas e ao *tap* que se registam as percentagens de identificação mais baixas. No que se refere ao [ɾ], fica também patente uma diferença significativa entre os resultados obtidos para o HLsyn e o SAPWindows, sendo que, no caso deste último, os valores são claramente superiores.

Repare-se, igualmente, que as oclusivas e fricativas vozeadas, quando comparadas com as congéneres não-vozeadas, estão sempre associadas a percentagens de acerto inferiores, não importa qual o sistema de síntese considerado.

Conforme ilustrado no gráfico 4.25, entre os fones mais facilmente percepcionados estão o [f], o [u] e o [m], todos com percentagens acima dos 80%. No fundo da tabela dos sons que apresentam maiores dificuldades aos ouvintes, com taxas de acerto inferiores a 25%, aparecem o [ɲ], o [b], o [v], [ʒ] e o [e].

Apurou-se, igualmente, a relevância do tamanho da palavra em termos de sílabas para a correcta discriminação dos estímulos e dos segmentos que os constituem. Os resultados são apresen-

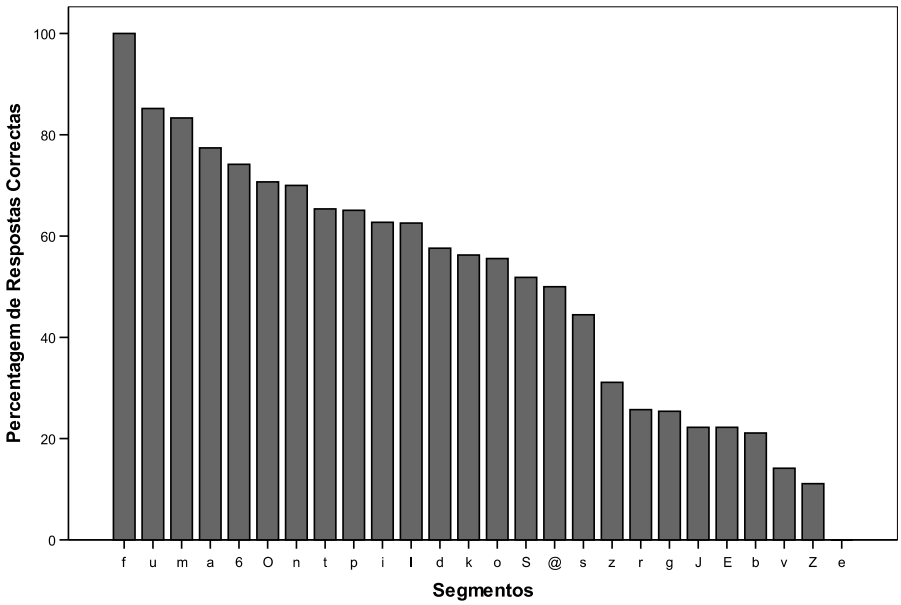


Figura 4.25: Percentagem de respostas correctas em função do fone.

tados na figura 4.26. A percentagem de respostas correctas, tanto ao nível da palavra como ao nível do segmento, foi maior sempre que a palavra apresentada era constituída por seis sílabas. Pelo contrário, os ouvintes revelaram muitos problemas em identificar estímulos com cinco sílabas. Este facto poderá estar relacionado não tanto com o número de sílabas, mas com a composição segmental dos próprios estímulos.

Apresentados os resultados globais, fizemos incidir a nossa atenção sobre as respostas dos ouvintes para cada fone em particular. Se é verdade que, na grande maioria das vezes, os participantes no estudo se mostraram incapazes de identificar o segmento em causa - como foi possível verificar a partir dos resultados globais inferiores a 30% - casos há em que este foi sistematicamente confundido e substituído por outros sons. A análise das circunstâncias em que estas trocas acontecem permitirá formular algumas hipóteses acerca dos factores subjacentes às dificuldades na identificação de alguns segmentos.

Assim, é possível constatar que, não raro, as consoantes oclusivas e fricativas surdas são confundidas com os respectivos pares sonoros e vice-versa, o que deixa antever problemas relacionados com a simulação do vozeamento.

Para além disso, os ouvintes parecem manifestar algumas dificuldades em distinguir o [s] do [ʃ] e, em menor grau, o [z] do [ʒ]. Este facto corrobora os resultados da avaliação perceptiva preliminar, realizada informalmente pelo próprio experimentador ao longo do processo de determinação das configurações gestuais associadas a cada segmento do PE.

Genericamente, verifica-se também uma tendência dos indivíduos participantes no teste em

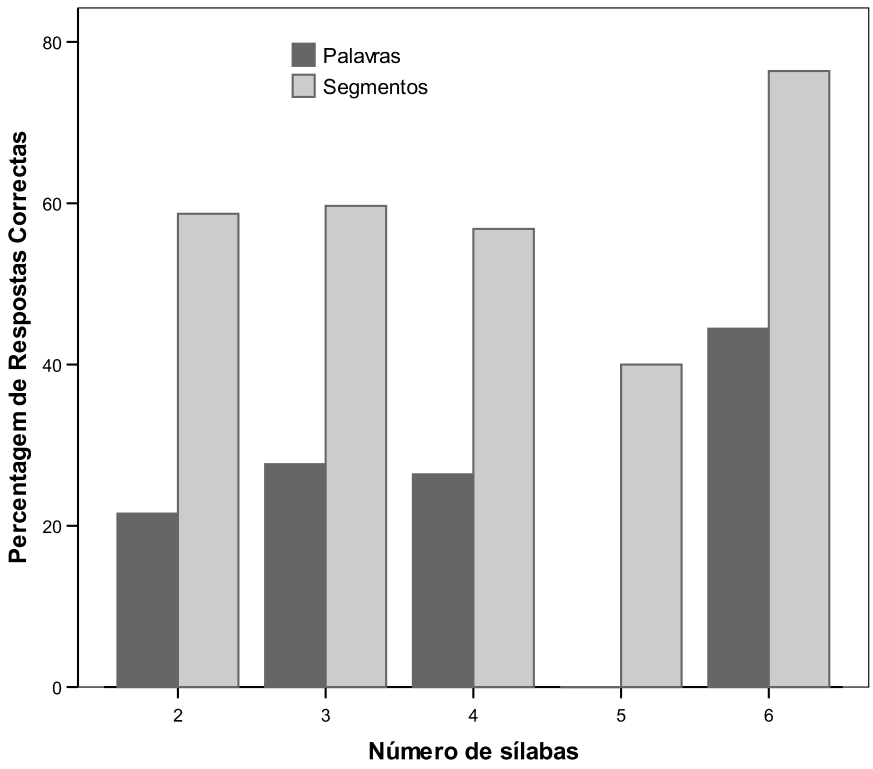


Figura 4.26: Percentagem de palavras (a cinzento escuro) e segmentos (a cinzento claro) correctamente identificados em função do número de sílabas.

associar uma determinada consoante a outras com o mesmo ponto de articulação: é o caso, por exemplo, do [l] frequentemente confundido com o [ɾ] ou ao contrário; ou do [m], algumas vezes associado ao [b].

No que respeita às vogais, ainda que facilmente reconhecidas na maioria dos casos, remanescem algumas confusões que reflectem problemas relacionados com distinções de altura. A questão é mais premente em relação às vogais centrais ([a] vs [ɐ]) e posteriores ([o] vs [ɔ]). Por outro lado, são de assinalar as dificuldades dos ouvintes em dissociar o [u] do [ɨ].

4.3.5 Discussão

Durante a realização dos testes, os ouvintes evidenciaram alguma dificuldade em identificar os estímulos apresentados. Tal dificuldade traduziu-se, por um lado, na audição repetida dos estímulos, antes da resposta à questão formulada, e, por outro, numa percentagem de acerto relativamente baixa, a rondar os 25%.

A dificultar a tarefa dos indivíduos recrutados para participarem no teste estiveram vários factores, com especial relevo para a falta de entoação e a ausência de acento. Apesar disso, se ti-

vermos em consideração as diferenças entre o número de respostas correctas do sujeito SA e dos restantes indivíduos, ilustradas na figura 4.23, é perfeitamente cabível supor que o próprio perfil do ouvinte ou o grau de empenhamento na realização da tarefa possam também influenciar, de algum modo, os resultados. Para comprovar a influência deste tipo de variáveis, seria, contudo, necessário recorrer a uma população mais alargada e diversificada, controlando rigorosamente factores como idade, proveniência geográfica, nível de escolaridade, entre outros.

Uma outra questão que teria sido pertinente investigar está relacionada com o apuramento da consistência das respostas dos sujeitos. Este procedimento, embora mais correcto do ponto de vista metodológico, revelou-se totalmente inviável devido, quer à reconhecida falta de qualidade dos estímulos apresentados, quer à necessidade de recorrer a dois sistemas distintos para síntese das palavras. Fazemos notar, no entanto, que, em consequência deste último facto, parte dos estímulos foi apresentada aos ouvintes em duplicado, compensando, de algum modo, a referida limitação. A repetição das palavras para cada um dos sintetizadores faria aumentar em muito o tempo requerido para a realização dos testes, sob pena de aborrecer os ouvintes e pôr em causa a acuidade das respostas.

Estas mesmas razões, que nos fizeram desistir de avaliar a consistência das respostas dos intervenientes no teste, determinaram também que, ao nível do material linguístico, nos tivéssemos circunscrito a um número de palavras relativamente reduzido. Os dados apresentados anteriormente foram obtidos com base num conjunto de apenas cinquenta estímulos. Ainda que o recurso a um número mais vasto de palavras (e de sujeitos) seja sempre preferível neste tipo de estudos, imposições externas relacionadas com a necessidade de reduzir ao mínimo o tempo de realização dos testes, para não dispersar a atenção dos ouvintes e evitar assim respostas aleatórias, motivaram um rigoroso controlo do número de palavras do teste.

Pensamos, no entanto, que este aspecto do alargamento do material linguístico pode e deve ser equacionado em investigações futuras, já com os problemas de falta de qualidade dos estímulos e limitações funcionais do SAPWindows devidamente resolvidas. Reunidas estas duas condições, não só um maior número de palavras poderá vir a ser admitido, facilitando a análise estatística posterior, como também outro tipo de material verbal (incluindo, por exemplo, palavras com laterais palatais e vibrantes múltiplas) deverá necessariamente ser considerado, logo que o TADA seja igualmente sujeito a alguns ajustes e correcções.

Embora a origem dos problemas nem sempre seja fácil de averiguar, tendo em conta as várias limitações que cada um dos componentes do sistema apresenta, alguns dos resultados acima expostos parecem estar directamente relacionados com as características dos próprios sintetizadores.

Tendo o SAPWindows sido originalmente concebido para síntese e estudo dos sons nasais, é sem surpresa que constatamos que, no tocante a este sintetizador, os melhores resultados obtidos dizem precisamente respeito a esta classe de sons. Estima-se, portanto, que também a agregação das vogais nasais ao sistema, prevista para o capítulo seguinte, decorra sem entraves de maior ou que, pelo menos, a proposta possa ser devidamente testada sem qualquer tipo de constrangimentos advenientes

de problemas com o sintetizador.

Paralelamente, os resultados revelam que também as vogais tendem a ser facilmente identificadas pelos ouvintes.

Com deixámos expresso anteriormente (secção 4.2.1), as variáveis do tracto - e respectivos valores de *target* - associadas às vogais foram estimadas com base em medidas articulatórias concretas e rigorosas (cf. Browman & Goldstein, 1990a), efectuadas a partir do perfil articulatório do informante AND. Adicionalmente, os perfis sagitais deste sujeito foram comparados com outras imagens RM, relativas ao mesmo contexto, mas adquiridas para outros informantes, de modo a aferir e validar os locais de máxima constrição considerados. Todo o processo esteve sujeito a uma cuidadosa apreciação auditiva, procurando monitorizar os efeitos de eventuais reajustes nos valores de ponto e, sobretudo, de grau de constrição. O rigoroso cumprimento dos referidos procedimentos metodológicos não invalida, contudo, a possibilidade de rectificar alguns dos valores previstos, nomeadamente em relação aos pares [a]/[ɐ] e [o]/[ɔ]. Cabe ainda salientar, a este respeito, que a tarefa de determinar um *target* vocálico numérico é, em grande parte, dificultada pela variação a que a pronúncia das vogais parece estar sujeita e que resulta não só do contexto, mas também de factores regionais, entre outros.

A percentagem de respostas correctas em relação às consoantes oclusivas surdas foi também satisfatória, mas os resuldados das oclusivas sonoras - baixa taxa de acerto e confusão sistemática entre consoantes surdas e sonoras - indiciam problemas ao nível da simulação do vozeamento. Esta limitação é comum aos dois sintetizadores e verifica-se, de igual modo, a respeito das consoantes fricativas (neste caso, apenas para o HLsyn). Para as dificuldades na identificação destas últimas muito terão contribuído ainda os problemas na geração de ruído. Conforme deixámos expresso em momentos anteriores deste capítulo, parece-nos clara a necessidade de se aperfeiçoar ainda mais o conhecimento dos mecanismos de produção destes segmentos, de modo a incrementar o desempenho dos sintetizadores articulatórios relativamente a estes sons.

No grupo dos segmentos mais difíceis de identificar, a par das fricativas surdas, inserem-se ainda a lateral alveolar e a vibrante simples. Se na base dos resultados obtidos para as fricativas surdas estão, a nosso ver, as características dos sistemas de síntese envolvidos na geração dos estímulos - de que avultam as deficiências na produção de vozeamento e fricção - as dificuldades na identificação de laterais e vibrantes parecem antes emanar de problemas associados ao TADA e às configurações gestuais em si mesmas. Já aqui fizemos notar que, na simulação do /l/ e do /ɾ/, lançámos mão de configurações gestuais alternativas, por forma a colmatar as limitações subjacentes ao TADA. Esta estratégia, não sendo a ideal, acabou por condicionar, conforme seria aliás de esperar, os resultados do teste perceptual.

Quando se observa o número de respostas correctas obtidas para a lateral alveolar, facilmente se verifica que o sucesso da solução implementada para fazer face à impossibilidade de especificar a *forma de constrição* no TADA depende do sintetizador usado. Se é verdade que a discrepância entre os resultados do HLsyn e do SAPWindows pode ser justificada pela falta de preparação deste

último para lidar com esta classe de sons, de antemão sabemos também que esta solução de recriar o efeito acústico do /l/ à custa da referência a uma dupla articulação não passa de um artifício. Embora esta abordagem funcione de forma satisfatória para o /l/ - pelo menos para o HLsyn - ela não é extensível a outros sons laterais, como oportunamente tivemos ocasião de referir.

Os mesmos argumentos são válidos em relação à vibrante simples, simulada, voltamos a lembrar, mediante uma alteração do valor do *stiffness*. Esta solução afigurou-se como a única ao nosso alcance para tentar ultrapassar a impossibilidade de testar outras alternativas, como o controle da designada *damping ratio*. Da análise dos resultados parece, no entanto, lícito arguir que o recurso a tal metodologia acabou por se revelar pouco eficaz. Concomitantemente, e à semelhança das laterais, esta abordagem não é generalizável às vibrantes múltiplas.

Destas observações decorre, portanto, a necessidade de implementar e avaliar outras vias complementares de representação gestual das laterais e vibrantes, até agora apenas equacionadas sob um ponto de vista teórico (vd. secção 4.2.5.4 e 4.2.5.5).

Não sendo este o objectivo primordial do estudo perceptual, também não ficou clara a interferência de variáveis como o número de sílabas da palavra ou estrutura dos estímulos na tarefa de identificação das palavras e dos segmentos. Para que se pudesse proceder a análises estatísticas mais poderosas e, assim, formular conclusões sólidas acerca desta matéria, seria necessário considerar não só um maior número de palavras, como também recorrer a populações mais alargadas.

A tarefa de validação da nossa proposta gestual, mediante a aplicação de um teste perceptual, apesar de arriscada por estar sujeita a numerosos constrangimentos reconhecidos anteriormente - que desde sempre fizeram prever uma percentagem de acerto global reduzida - revelou-se útil na identificação dos principais problemas e futuras linhas de intervenção. Confirma-se a validade e interesse em adaptar o TADA, tornando-o apto a lidar com os sons do PE.

Capítulo 5

Para uma Abordagem Gestual das Vogais Nasais do Português Europeu

Durante a emissão das vogais fechadas..., o véu palatino pode abaixar-se permitindo que a corrente expiratória se escape também pelas fossas nasais. A vibração laríngea é então ressoada por duas câmaras, a bucal e a nasal. Chamam-se vogais nasais aos fonemas assim pronunciados.

Oliveira Guimarães (1927)

5.1 Introdução

Terminada a longa e complexa tarefa de caracterização gestual dos vários segmentos do PE - condicionada, em larga medida, pela escassez de dados articulatórios e de informação acerca da organização temporal do PE - propomo-nos agora avançar para o cumprimento daquele que é um dos objectivos primordiais desta dissertação: a descrição e análise do fenómeno da nasalidade à luz dos princípios dinâmicos da FA, com base num estudo experimental, especialmente desenhado para o efeito.

Contrariamente ao capítulo anterior, mais do que especificar os gestos associados às vogais nasais, foi nosso propósito obter dados acerca da organização temporal dos próprios gestos. Adicionalmente, ao coligir pistas sobre o mecanismo de produção das vogais nasais, esperamos também poder contribuir para o aprofundamento do conhecimento das propriedades fonéticas do material estudado e, em última análise, aduzir alguns argumentos úteis à reflexão teórica em torno das vogais nasais do português.

A decisão de circunscrever este tipo de análise às vogais nasais está relacionada, antes de mais, com a importância da própria nasalidade, frequentemente invocada como uma das idiossincra-

sias mais características do vocalismo do português (Strevens, 1954). Existem, contudo, outras razões - já enunciadas parcialmente na Introdução a este trabalho - que justificam a opção tomada, nomeadamente as características do sintetizador articulatório SAPWindows e a dificuldade de acesso a dados articulatórios desta natureza (EMA), facilitado, no caso particular das vogais nasais, pelos contactos estabelecidos com os investigadores do GIPSA-LAB, no âmbito de dois projectos de investigação (projecto HERON e projecto *"Dynamique de la nasalité. Émergence et phonologisation des voyelles nasales"* [1]). Além destes aspectos, a nasalidade do português continua a suscitar algumas dúvidas e controvérsias, resultantes, em boa medida, da escassez de estudos recentes acerca desta matéria.

No quadro teórico da FA, o tratamento das vogais nasais implica uma referência a, pelo menos, duas variáveis do tracto: o corpo da língua (TB) e o velo (VEL). Para a caracterização da dimensão oral contámos, como habitualmente, com a informação obtida através de RM. Já a descrição do movimento do velo e da sua relação com os demais articuladores assenta num estudo EMA, desenhado, entre outros objectivos, com o intuito específico de dar resposta às questões formuladas neste capítulo. O processo que presidiu à concepção e recolha deste *corpus*, bem como a análise dos resultados obtidos, constitui a essência deste capítulo.

Assim, depois de uma breve introdução teórica à temática da nasalidade, através da revisão de estudos anteriores - dedicados sobretudo ao português, mas com referência, sempre que tal se justificou, a outras línguas onde o fenómeno assume relevância - ocupámo-nos, numa primeira fase, da caracterização das variáveis do tracto grau e local de constrição do corpo da língua, que especificam o gesto vocálico, seguindo uma metodologia em tudo similar à adoptada no capítulo anterior.

Tendo em mente uma série de resultados e questões prévias suscitadas por um conjunto de estudos, que se debruçaram sobre a questão da nasalidade, nas mais variadas perspectivas, delineámos as hipóteses e questões que nos propomos investigar no estudo exploratório do comportamento do velo, baseado em EMA. Posteriormente, cedemos lugar a uma apresentação geral da metodologia seguida ao longo do estudo experimental, desde a escolha do *corpus* até ao pós-processamento e anotação dos dados, passando pela explicitação dos procedimentos adoptados durante a recolha do *corpus* junto dos informantes.

A primeira parte deste capítulo encerra com a apresentação, em secções separadas, dos resultados obtidos para cada umas das variáveis em análise - amplitude, duração, *stiffness*, coordenação temporal - seguida de uma breve discussão dos mesmos.

A segunda parte (e última) do capítulo diz respeito ao estudo perceptivo realizado, tendo em vista o esclarecimento de algumas questões suscitadas no decurso do estudo articulatório.

[1] Projecto financiado pela *Agence Nationale de la Recherche* (ANR) e coordenado por Solange Rossato.

5.2 Conceitos teóricos

A nasalidade encontra-se frequentemente associada às vogais, sendo possível encontrar vogais nasais em cerca de 22% das línguas consideradas no inventário *UCLA Phonological Segment Inventory Database* (UPSID) (Maddieson, 1984). As estatísticas mais recentes (Maddieson, 2007) apontam para 138 línguas, num total de 670, com uma ou mais vogais nasais [2], o que corresponde a um pouco mais de um quinto do total das línguas consideradas. Segundo os dados de Hajek (2008), o contraste entre vogais orais e vogais nasais está presente em 26% das línguas estudadas (64 línguas em 244).

Em praticamente todos os casos, o número de vogais nasais não ultrapassa o número de vogais orais [3]. Para além disso, não existem línguas apenas com vogais nasais, sem vogais orais.

O português faz parte do grupo de línguas que incluem, no seu inventário fonológico, vogais e consoantes nasais [4].

5.2.1 Sistema vocálico nasal do português europeu

Do sistema vocálico do português fazem parte cinco vogais nasais [5] - [ɐ̃], [ẽ], [ĩ], [ũ], [õ] - tradicionalmente classificadas como fechadas ou semi-fechadas [6]. Em Viana *et alii* (1993, 1996) e Teixeira *et alii* (2001) podem ser encontradas algumas estatísticas relativas à frequência de ocorrência dos fones do PE, em dois corpora distintos. Na tabela 5.1, damos conta dos dados relativos às vogais nasais.

Corpus	[ɐ̃]	[ẽ]	[ĩ]	[ũ]	[õ]
Teixeira et al, 2001	2.9%	1.2%	0.7%	0.6%	0.9%
Viana et al., 1993	3.3%	1.2%	0.7%	0.8%	0.7%
Viana et al., 1994	2.5%	1.7%	0.7%	0.2%	0.9%

Tabela 5.1: Frequência de ocorrência das vogais nasais do PE, segundo os dados publicados em Viana *et alii* (1993, 1996) e Teixeira *et alii* (2001).

Em relação ao subsistema oral, verifica-se uma redução no número de vogais. Este compor-

[2] O número de vogais nasais nas várias línguas vai de uma vogal nasal (Gwari, Cherokee e Copainalá Zoque) até dez vogais nasais (Sindhi e Koromfe) (Maddieson, 2007).

[3] A única excepção parece ser a língua Koyra Chiini (Songhay, Mali), que apresenta quatro tipos de contraste, envolvendo nasalidade e duração, e que tem mais vogais nasais breves do que vogais orais breves (Hajek, 2008).

[4] Clements & Osu (2003) agrupam as línguas em quatro grupos, em função do seu sistema nasal: tipo 1 - nem consoantes nasais, nem vogais nasais (1%); tipo 2 - unicamente vogais nasais, sem consoantes nasais (1%); tipo 3 - somente consoantes nasais, sem vogais nasais (76%); tipo 4 - consoantes nasais e vogais nasais (22%). Apesar da polémica em torno da interpretação fonológica das vogais nasais (vd. secção 5.2.3), pensamos que o português poderá ser integrado neste último grupo.

[5] Este número corresponde, segundo Maddieson (2007), ao padrão mais frequente, atestado para a maioria das línguas com vogais nasais.

[6] Em variedades regionais encontram-se vogais abertas ou semi-abertas (Sampson, 1999; Barbosa, 1994a; Cunha & Cintra, 1997). Para uma revisão das alterações fonéticas das vogais nasais, condicionadas por factores dialectais, consultar Sampson (1999). Segundo este mesmo autor, as diferenças dialectais na articulação das vogais nasais decorrem das assimetrias na evolução histórica destes sons, ao longo do território nacional.

tamento é comum à maioria das línguas analisadas por Maddieson (2007) e por Hajek (2008) [7].

Contudo, todas as vogais do português podem ser nasalizadas [8] por coarticulação, i.e., as vogais orais em contacto com as consoantes nasais, quer em posição anterior (nasalização regressiva), quer em posição posterior (nasalização progressiva), sofrem os efeitos da nasalidade. A nasalização contextual - sobretudo a regressiva - é um processo muito actuante no PB, mas, no PE, a sua accção é menos marcada (Moraes, 2003).

Embora alguns autores (Nobiling, 1974; Lacerda & Strevens, 1956; Câmara, 1977; Mateus, 1975) atribuam às vogais nasalizadas um grau de nasalidade mais fraco e um papel secundário, não são conhecidos muitos estudos sobre o fenómeno de nasalização por coarticulação em PE. Destacam-se as conclusões de Lacerda & Head (1966, p.67):

1. *Uma vogal oral antes de consoante nasal nem sempre manifesta nasalização: frequentemente o seu decurso é inteiramente oral. Quando há nasalização, esta aparece, quase sempre, apenas durante um breve subtrecho final.*
2. *Uma vogal precedida de consoante nasal manifesta sempre um decurso nasal ou nasal-oral.*

Para além da nasalidade de tipo "fonêmico" e da nasalidade "alofônica", que ocorre quando a vogal precede uma consoante nasal na sílaba seguinte, Moraes (1997, 2003) distingue ainda um terceiro tipo de nasalidade potencial, a que chama "nasalidade regressiva por coarticulação": enquanto a "nasalidade alofónica" acontece nas sílabas acentuadas ("cama" [ˈkɐ̃.mɐ]) [9], a "nasalidade por coarticulação" ocorre nas sílabas átonas ("camada" [kɐ.ˈmɐ.dɐ]). O estudo articulatório Moraes (1997), relativo ao PB, revela que a amplitude do gesto vélico é similar na nasalidade fonémica e alofónica, o que sugere que ambos os processos são o resultado da aplicação de regras fonológicas de nasalização em PB. Pelo contrário, em sílaba átona (nasalidade regressiva por coarticulação), o movimento do velo é claramente inferior, o que aponta para um fenómeno fonético transitório, universal e não-perceptível (Solé & Ohala, 1991).

Os resultados do estudo perceptivo (Moraes, 2003) corroboram os do estudo articulatório, mostrando que, no PB, a nasalidade "fonêmica" e a nasalidade "alofônica" são processos similares,

[7] Razões de natureza perceptual podem ajudar a explicar a diminuição do número de vogais nasais em relação às vogais orais (Beddor, 1993; Sampson, 1999; Maddieson, 2007): "the addition of nasal coupling causes the first peak of spectral proeminence, F1, which is critical for vowel quality identification, to lose the intensity and spread in bandwith. As a result, the differences between neighbouring but distinct vowel types can be masked to some extent, such that there is a greater possibility amongst hearers for associating and equating originally different nasal vowel types." (Sampson, 1999, p.12).

[8] A fonética tradicional (Ladefoged, 1975) reserva o termo nasal para os sons produzidos com a passagem nasal aberta e uma oclusão completa no tracto oral, i.e. sons em que o fluxo o ar é libertado, na sua totalidade, através das cavidades nasais. Os outros sons, articulados com os canais oral e nasal abertos são referenciados como nasalizados. No entanto, neste trabalho, na esteira da tradição fonológica portuguesa (e.g. Lacerda & Head, 1966; Almeida, 1976; Wetzels, 1997), um som é classificado como nasal sempre que ocorra um contraste fonológico entre esse som e um qualquer outro da mesma língua, determinado pela posição do véu palatino. Um som é considerado nasalizado quando realizado com o véu palatino abaixado, mas devido a condicionamentos fonéticos, i.e., por efeitos de assimilação com uma nasal adjacente. Assim, teremos vogais nasais fonológicas e vogais nasalizadas por acção do contexto.

[9] Os exemplos e respectivas transcrições fonéticas foram retirados de Moraes (1997).

enquanto a nasalidade "co-articulatória" é um fenómeno puramente fonético. Contudo, no PE "a nasalização dita alofônica e a co-articulatória se comportam como processos fonéticos, distintos da nasalização fonêmica." (Moraes, 2003).

O fenómeno da nasalidade por coarticulação não será tido em conta neste estudo. Fica o registo da sua ocorrência em português, a par da nasalidade fonológica.

5.2.2 Evolução histórica das vogais nasais

Também do ponto de vista diacrónico, o processo de formação das vogais nasais em português envolve algumas complexidades. O problema não será aqui tratado em detalhe. Contentar-nos-emos em destacar alguns factos da evolução histórica destes sons, que permitam esclarecer e melhor compreender o fenómeno da nasalidade em português. Tomaremos como referência básica a obra de Sampson (1999), a par com outras como Nobiling (1974), Teyssier (1980), Parkinson (1996), Martins (1995), entre outros.

Nesta língua, o aparecimento de vogais nasais acontece, sobretudo, em resultado de um processo de assimilação regressiva: a vogal é nasalizada por acção de uma consoante nasal tautossilábica, que posteriormente desaparece [10]. Os casos de nasalização progressiva são muito mais raros e estão sujeitos a várias restrições, o que aponta, de algum modo, para o estatuto secundário deste segundo processo (Sampson, 1999).

Embora o processo de desenvolvimento das vogais nasais seja bastante mais complexo e detalhado, Sampson (1999) identifica três contextos principais, que estariam na origem destes sons:

1. /VNC/, onde a vogal precede uma consoante nasal que é seguida por outra consoante;
2. /VN#/, onde a vogal precede uma consoante nasal final [11];
3. /VNV/, onde a consoante nasal ocupa uma posição intervocálica.

1. **/VNC/**

   Em contexto VNC, a consoante nasal resistiu durante mais tempo ao desaparecimento do que no contexto /VN#/, acabando, contudo, também por enfraquecer gradualmente, especialmente antes de consoantes [+contínuas], dando origem a um som de transição ou acabando mesmo por cair. No caso das consoantes [-contínuas], o enfraquecimento da consoante resulta numa pequena oclusão, cujo ponto de articulação depende da natureza da consoante seguinte, ou num breve som de transição. Fica por determinar se estas ressonâncias consonânticas são resquícios

[10] De acordo com alguns linguistas (Vaissière, 1995), a tautossilabicidade favorece a nasalização fonológica.

[11] Quando a consoante nasal termina a palavra, a grafia mais comum foi, durante muito tempo o *-n*, embora o til (˜) sirva também frequentemente para indicar a nasalidade das vogais. Contudo, desde o período do galego-português medieval, começa a surgir, nesta posição, a grafia em *-m* (e.g. *razõ, razon, razom*). É esta grafia em *-m* que se vai generalizar em português (Teyssier, 1980).

do *m* e *n* latinos ou se desenvolvem posteriormente a partir das próprias vogais nasais (Sampson, 1999; Nobiling, 1974). Há, no entanto, indícios do apagamento da consoante já no período pré-literário ou início do período literário, sobretudo antes de consoante fricativa. A grafia de formas medievais como *iffante* (< infantem), *cofonder, cofortar, cofujom* (<confundere, confortare, confusionem) aponta para o apagamento da nasal antes de fricativa, embora a interpretação do estatuto destas vogais - enquanto vogais orais ou vogais nasais - seja controversa.

2. **/VN#/**

   Neste contexto, a consoante nasal final foi progressivamente enfraquecendo até desaparecer por completo, transmitindo a nasalidade à vogal precedente. Terá sido este um processo relativamente rápido e estaria praticamente concluído no início da era literária (Sampson, 1999). A apócope tem lugar em palavras em que o *-e* final é precedido por consoantes coronais, incluindo a consoante nasal -N- e a geminada -NN- (e.g. panem > pan(e) > pãn > pãw̃). As evoluções subsequentes dependem da qualidade da vogal anterior: enquanto as vogais altas ([ĩ] e [ũ]) desenvolveram provavelmente sons de transição, que foram posteriormente absorvidos (e.g. finem > fĩn > fĩ), as restantes vogais deram origem a ditongos (e.g. bene > bẽn > bẽ > bẽj̃ (Sampson, 1999).

3. **/VNV/**

   Neste contexto, os *-n* intervocálicos desaparecem, depois de terem nasalizado a vogal precedente (e.g. manu> mãnu> mão) [12]. Esta queda do *-n* intervocálico é um fenómeno exclusivo do galego-português (Teyssier, 1980; Sampson, 1999). Não se documenta em leonês, nem no castelhano, nem nos falares moçárabes [13]. Nos dialectos do sul, tanto na toponímia (e.g. *Madroneira* (Beja) em vez de *Madroeira*), como em algumas palavras de uso comum (e.g. *maçanera* por *macieira* (Algarve e Baixo Alentejo)) o *-n* intervocálico não raro permaneceu. Os *-n* intervocálicos mantêm-se também em várias palavras de origem árabe (e.g. *azeitona, alfenim*).

   A datação do fenómeno não reúne o consenso dos especialistas. Segundo Teyssier (1980), este teve lugar no início no século XI, mas no século XII estaria ainda em curso. A maioria dos estudiosos tende, no entanto, a optar pelo século X como data mais provável para a ocorrência do desaparecimento da consoante nasal (Sampson, 1999). Apesar das dúvidas, os dados disponíveis sugerem que o apagamento do *-n* poderá ter acontecido logo a partir do século IX, com algumas restrições numa primeira fase, generalizando-se posteriormente a todo o território galaico-português.

   Da queda do *-n* intervocálico resulta um grande número de hiatos ou sequências de vogal nasal seguida de vogal oral (ṼV). Numa primeira fase, as duas vogais em contacto pertenceriam a

[12] O apagamento do -n intervocálico encontra paralelo no comportamento do -l nesta mesma posição. Ambos os fenómenos têm lugar no período pré-literário. Por altura do aparecimento dos primeiros textos (século XIII), não há qualquer vestígio da queda do -l intervocálico, já a consoante nasal deixa a sua marca na vogal precedente.

[13] A conservação do -n e do -l intervocálicos na zona moçárabe é um facto bem conhecido (cf. Teyssier, 1980, nota 5).

sílabas distintas, o que pode ser facilmente atestado através da escansão dos versos dos textos dos *Cancioneiros* (Teyssier, 1980) [14]. Contudo, estes grupos de vogais em hiato são por natureza muito instáveis, pelo que cedo se iniciam as evoluções, que teriam como fim a supressão da maioria dos encontros vocálicos.

As soluções postas em prática para atingir este objectivo são muito variadas e dependem da acentuação, da altura da vogal nasal e das diferenças na posição da língua entre as vogais contíguas (cf. Sampson, 1999, para uma descrição mais detalhada).

Tal como foi já referido, um conjunto limitado de palavras com vogais nasais resulta de um processo de assimilação progressiva. Neste caso, o /m/ e, mais raramente, o /n/ nasalizam a vogal seguinte, sobretudo se esta for alta e acentuada (mihi > mĩ). O processo de integração destas formas na língua padrão foi, no entanto, bastante lento, estendendo-se ao longo de vários séculos. Em alguns casos, a variante nasal co-existe com a variante oral (e.g. [ˈmẽzɐ] ou [ˈmezɐ]), mas somente esta última acaba por resistir e integrar a língua-padrão.

5.2.3 Estatuto fonológico das vogais nasais

O estatuto fonológico das vogais nasais do português está longe de reunir o consenso. Mais uma vez, a questão não será aqui abordada em profundidade, na medida em que está fora do âmbito deste trabalho. Limitar-nos-emos a traçar as linhas centrais desta discussão, a fim de melhor caracterizar o nosso objecto de estudo. A controvérsia decorre, em boa parte, do conhecimento imperfeito das características articulatórias destes segmentos e, neste sentido, o nosso trabalho poderá ajudar a esclarecer os pontos mais obscuros da teoria.

Observando os pares mínimos [ˈkatu]/[ˈkɐ̃tu], [ˈtetɐ]/[ˈtẽtɐ], [ˈpitɐ]/[ˈpĩtɐ], [ˈrõbu]/[ˈrobu], [ˈtubɐ]/[ˈtũbɐ], importa saber em que consiste a oposição entre as formas.

Existem basicamente três interpretações distintas do fenómeno da nasalidade em português:

1. A maioria dos fonólogos portugueses (e.g. Barbosa, 1961, 1965; Mateus, 1975; Barroso, 1999; Mateus & d'Andrade, 2000) opta por uma análise bifonémica das vogais nasais, em que estas são entendidas não como fonemas distintos, mas como uma combinação abstracta de uma vogal e um elemento nasal ([Ṽ]=/V+N/). Este segmento nasal abstracto é interpretado como um "arquifonema nasal" (Câmara, 1971, 1977; Cagliari, 1977; Barbosa, 1961, 1965) ou como um "autossegmento" (Mateus & d'Andrade, 2000; d'Andrade, 1994). Isto significaria que "underlyingly, there are no nasal vowels in Portuguese" (Mateus & d'Andrade, 2000, p.21) e "underlyingly, Portuguese nasal vowels receive their nasality from a nasal segment that is deleted at the phonetic level" (Mateus & d'Andrade, 2000, p.23).

[14] Os Cancioneiros são três - *Cancioneiro da Ajuda* (século XIII/ XIV), *Cancioneiro da Vaticana* (século XVI) e *Cancioneiro da Biblioteca Nacional de Lisboa* (século XVI) - e incluem *cantigas de amigo*, *cantigas de amor* e *cantigas de escárnio e maldizer*.

Esta proposta assenta em vários argumentos. Destacamos os seguintes:

- o comportamento do /r/ intervocálico, que se realiza sempre como [ʀ] após vogal nasal, o que só acontece em sílaba fechada por consoante (e.g. [ˈõʀɐ̃]);
- a resistência da vogal nasal final à crase com a vogal seguinte, contrariando uma tendência frequente entre sequências de vogais;
- a inexistência de vogais nasais em hiato: a nasalidade desaparece ou dá origem a uma consoante de transição. Segundo Barbosa (1961), "Ce n'est donc pas la nasalité en elle-même qui empêche la crase, mais le fait qu'il y a une consonne entre les deux voyelles.";
- a mesma representação de base (V+N) está na origem de palavras derivadas a partir do prefixo <in>, como, por exemplo, "incapaz" ou "inacabado". No primeiro caso, a sequência VN realiza-se foneticamente como uma vogal nasal ([ĩ]) antes de consoante; na segunda palavra, o grupo realiza-se como uma vogal oral seguida de consoante nasal ([in]). A mesma alternância vogal nasal/ vogal+consoante pode ser encontrada nos ditongos (e.g. "irmão" [iɾˈmɐ̃w̃], "irmanar" [iɾmɐˈnaɾ]) [15].

2. Segundo alguns autores (Lüdtke, 1952; Head, 1964; Pontes, 1972, *inter alia*), as vogais nasais têm função distintiva em relação às vogais orais e, constituem, portanto, fonemas independentes do português (interpretação monofonémica). Isto significa que a vogal nasal está presente na representação de base, na matriz fonológica.

3. Para Parkinson (1983), as vogais nasais são, na verdade, ditongos, "made up of two phonological segments, one oral and one nasal, but the second element is a vowel rather than a consonant, and nasal vowels are true diphtongs" (Parkinson, 1983, p.158) [16].

5.2.4 Acerca do gesto do velo

O acesso às cavidades nasais é regulada pela posição do velo. Sempre que este se encontra descido - quer por determinações linguísticas, quer porque esta é a sua posição neutral de repouso - a passagem nasal fica desimpedida e o fluxo do ar escapa-se pelas fossas nasais. Alternativamente, o velo pode elevar-se em direcção às paredes da faringe, em diferentes graus, bloqueando total ou parcialmente a passagem nasal. A distinção fonológica entre um som [+ nasal] e [- nasal] proposta por Chomsky & Halle (1968) assenta precisamente nesta diferença articulatória fundamental, que se aplica tanto às vogais como às consoantes:

[15] Aos argumentos listados, Veloso (2008) acrescenta um outro: "given that in Portuguese no more than one consonant is admitted in coda position (Mateus & d'Andrade 2000: 53), the inhibition of any coda consonant by a nuclear nasal vowel suggests that this prosodic constituent is actually filled ("saturated") by nasality.".

[16] Importa referir, pelas suas semelhanças com a proposta de Parkinson (1983), as interpretações de Louro (1954-1955, p.242): "As outras vogais [nasais] (ã ou ɐ̃, ĩ, ẽ, ũ, õ), quando mediais ou no interior das frases, são geralmente ligadas (ou mesmo substituídas na sua parte final) por um ə̃ (formando com elas uma espécie de discreto ditongo decrescente. São as vibrações deste ə̃ que, nos gráficos, fazem pensar na existência, em português, de verdadeiras consoantes nasais, em fim de sílaba interna.".

> *Nasal sounds are produced with a lowered velum which allows the air to escape through the nose; non nasal sounds are produced with a raised velum so that the air from the lungs can escape only trough the mouth.* (Chomsky & Halle, 1968, p.316)

Na medida em que o velo é responsável por esta distinção fonológica binária, a este articulador foram também incorrectamente atribuídas, durante muito tempo, apenas três possibilidades fonéticas: aberto para as vogais e consoantes nasais; fechado para as consoantes orais; com uma posição inerentemente neutra para as vogais orais, determinada pelo contexto consonântico. Um grande número de estudos experimentais vem, no entanto, contrariar este pressuposto algo simplista, ao apurar que a altura do velo não é uniforme para todas consoantes e vogais orais, nem sequer para todas as consoantes e vogais nasais.

Em primeiro lugar, a produção de segmentos orais não implica que a passagem nasal esteja completamente selada (e.g. Björk, 1961; Huffman, 1989; Cohn, 1990), mas antes que a área seja suficientemente pequena para impedir o acoplamento acústico e aerodinâmico das cavidades oral e nasal (Bell-Berti, 1993). Bell-Berti (1993) estima que a percepção da nasalidade exija aberturas velo-faríngeas acima dos 20 mm2.

Para além disso, a posição do velo varia em função da qualidade dos segmentos (oral vs nasal, consoante vs vogal, ponto e modo de articulação, altura da vogal, vozeamento, etc.), para além de ser influenciada por factores prosódicos - como a posição do segmento na sílaba ou o acento (Krakow, 1989, 1993; Vaissière, 1988) - pela taxa de elocução (Krakow, 1993) e até por estratégias individuais.

Do mais fechado para o mais aberto, e no que respeita à altura intrínseca do velo, a ordem geralmente observada é a seguinte: CONSOANTES ORAIS (oclusivas > fricativas > líquidas) > GLIDES > VOGAIS ORAIS > CONSOANTES NASAIS > VOGAIS NASAIS (Rossato *et alii*, 2006; Amelot & Rossato, 2006; Vaissière, 1995). Esta hierarquia é, *grosso modo*, compatível com as duas escalas - propostas por Walker (2000) e Schourup (1972) - referidas por Solé (2007), que traduzem a harmonia entre os segmentos e a nasalidade. Ambas reflectem constrangimentos aerodinâmicos e acústicos.

De acordo com Ohala (1975, p.300), "Nasalization would be least compatible with oral obstruents... since the noise of fricatives and affricates and burst at the release of stops requires a build up of air pressure in the oral cavity. This would require that no air leak out of the oral cavity into the nasal cavity.". Para que se produza a pressão oral requerida à produção de oclusivas e fricativas, o velo, teria, portanto, de estar subido. Este constrangimento não se aplica às consoantes, cujo ponto de articulação tem lugar atrás da passagem velo-faríngea (oclusivas e fricativas glotais e faríngeas), na medida em que a posição do velo não tem influência na geração desta pressão (Ohala, 1975; Ohala & Ohala, 1993).

Alguns estudos experimentais (e.g. Ohala *et alii*, 1998b; Solé, 1999) demonstram, contudo, que a produção de obstruintes - mais concretamente de fricativas - é compatível com pequenas aber-

turas vélicas. Os resultados de Shosted (2006a) sugerem que as fricativas nasalizadas podem efectivamente ocorrer, mas não sem que as suas características espectrais sejam significativamente alteradas. Em fala espontânea, é possível ter fluxo nasal durante a totalidade das oclusivas e fricativas, que antecedem ou sucedem vogais nasais (Basset *et alii*, 2001). Segundo Warren *et alii* (1994), uma abertura de 10 mm2 durante a produção de oclusivas é perfeitamente admissível e não implica a percepção de nasalidade.

Quanto às líquidas, Solé (2002, 2007) destaca as vibrantes múltiplas como segmentos altamente incompatíveis com a nasalidade. A produção destas consoantes está intimamente dependente de um fluxo oral elevado, que possa propiciar a vibração da língua. Pequenas variações da pressão oral, em consequência do abaixamento do velo, podem impedir essa vibração (Solé, 2002). Este facto justifica que as vibrantes múltiplas nasalizadas não façam parte do repertório fonético das línguas conhecidas. Os constrangimentos aerodinâmicos referenciados não se aplicam à vibrante simples, que envolve um movimento balístico do ápice da língua. Exemplos de vibrantes simples nasalizadas estão amplamente documentados numa grande variedade de línguas (Ladefoged & Maddieson, 1995).

A produção dos demais segmentos (laterais, glides e vogais) congraça-se perfeitamente com uma posição do velo mais baixa, sem que as suas características essenciais sejam comprometidas. O estudo de Moll & Daniloff (1971), por exemplo, sugere que o abaixamento do velo para a produção da consoante nasal pode ter lugar ainda durante o /l/. Amelot & Rossato (2006), por sua vez, constatam que "Some oral segments are produced with an open velopharyngeal port: this concerns some oral vowels following or preceding a nasal consonant, and productions of the oral consonants /l/ or /ʁ/ in nasal vowel contexts.". Contudo, se a abertura da passagem velo-faríngea exceder a área da constrição oral, uma soante nasalizada pode transformar-se numa consoante nasal.

A utilização de posições intermédias do velo não implica, no entanto, a percepção de diferentes níveis de nasalidade. Os exemplos de línguas com um contraste fonológico entre diferentes tipos de nasalidade - como o Acehnese (Indonésia), que distingue nasalidade "forte" e "fraca" (Ladefoged & Maddieson, 1995) - são, aliás, bastante raros.

5.2.4.1 Relação entre amplitude do velo e altura da vogal

No que respeita às vogais, vários estudos experimentais (e.g. Henderson, 1984; Bell-Berti *et alii*, 1979; Moll, 1962; Rossato *et alii*, 2003; Clumeck, 1976; Ohala, 1975) apontam para uma relação directa entre a altura da vogal e a posição do velo: este tenderia a descer mais durante a produção de vogais baixas, do que durante a produção de vogais altas.

Conforme assinalado por Ohala (1975), Clumeck (1976) e Rossato *et alii* (2006), em algumas línguas, as vogais baixas em contextos orais são produzidas com a passagem velo-faríngea aberta. No francês, de acordo com Durand (1953), a vogal baixa em contexto oral (*il l'a*) pode ser produzida com uma distância de 10 mm entre o palato mole e a parede posterior da faringe. O facto é corroborado por dados mais recentes, obtidos a partir da tecnologia EMA (Rossato *et alii*, 2006).

Os dados EMA disponíveis para o português europeu (Teixeira *et alii*, 2001; Rossato *et alii*, 2006) confirmam também esta correlação, na medida em que para a vogal aberta [a] o velo se encontra numa posição mais baixa do que para todas as outras vogais orais, com uma amplitude próxima das consoantes nasais [m] e [n]. A universalidade desta relação entre abertura do velo e altura da vogal em contextos orais é, contudo, posta em causa por Hajek (1997), que refere alguns exemplos de línguas (e.g. francês do norte de África), em que a abertura do velo se estende às vogais orais altas.

Do mesmo modo, também o pressuposto (Henderson, 1984; Bell-Berti, 1993) de que a posição do velo em contextos nasais varia também em função da altura da vogal parece não se aplicar a todas as línguas da mesma forma.

Numa investigação sobre a coarticulação nasal antes de N, em seis línguas (inglês americano, francês, chinês, português do Brasil, sueco e hindi), Clumeck (1976) verifica que apenas cinco dos trinta falantes analisados apresentam diferenças estatisticamente significativas entre as vogais altas e as vogais baixas, no que respeita à altura do velo [17].

Num outro estudo, Al-Bamerni (1983) descobre uma diferença estatisticamente significativa na abertura velar entre vogais altas e vogais baixas, por um lado, e vogais baixas e vogais médias, por outro, apenas em cinco (inglês, francês, curdo, árabe e norueguês) das sete línguas estudadas. Em hindi e gujarati, a altura da vogal não influencia a amplitude do velo, sendo que as maiores aberturas se registam para as vogais posteriores (médias e altas).

Com base nestas evidências, Hajek (1997) sugere que a hipótese de uma correlação universal entre altura da vogal e amplitude do velo deve ser rejeitada, na medida em que não encontra suporte empírico em todas as línguas.

5.2.4.2 A propósito da importância da dinâmica do velo

A necessidade de modelar a trajectória dos articuladores ao longo do tempo decorre da observação de que qualquer articulador mantém uma posição constante apenas durante um breve instante de tempo. Estudos vários revelam contínuas variações na posição do velo, mesmo durante a produção de segmentos orais (cf. Bell-Berti, 1993). Ohala & Ohala (1993) sublinha a importância das pistas dinâmicas para os segmentos nasais. Em hindi, o investigador observa um abaixamento progressivo e mais acentuado do velo durante a produção de vogais nasais do que durante a articulação de vogais nasalizadas por contexto.

Clumeck (1976, p.351), por sua vez, chama a atenção para o papel da dinâmica temporal, mais do que da altura do velo, na percepção da nasalidade: “It might then be the case that the listener’s perception of the presence or absence of the nasalization is more dependent upon the timing of palatal lowering rather than upon actual extent of palatal lowering.”. Teixeira *et alii* (1999) e Teixeira

[17] Clumeck (1976) sugere ainda uma correlação entre o tempo de abaixamento do velo e a altura da vogal: o movimento de descida do velo teria início mais cedo nas vogais baixas do que nas vogais altas. Os próprios resultados do autor indicam, no entanto, que esta relação varia em função da língua estudada.

(2000) demonstram experimentalmente que a qualidade das vogais nasais sintetizadas, bem como a sua percepção, melhora substancialmente, se a variação do velo ao longo do tempo for considerada.

Em português, a dimensão temporal da nasalidade reveste-se de particular interesse, na medida em que às vogais nasais desta língua têm sido comummente associadas três fases : 1) *onset* oral, 2) fase nasal propriamente dita e 3) murmúrio nasal. Esta hipótese tem sido corroborada por vários estudos (e.g. Lacerda & Strevens, 1956; Lacerda & Head, 1966; Drenska, 1988; Sousa, 1994; Silva, 1995; Gregio, 2006; Lovatto *et alii*, 2007), usando metodologias diversas, que vão desde a simples capacidade auditiva do investigador até à investigação articulatória, passando pela análise acústica.

Lacerda foi dos primeiros investigadores a mencionar a manifestação da nasalidade ao longo do tempo, a partir de uma experiência realizada com o auxílio de um *speech-stretcher* ("extensor-sonoro") (Lacerda & Strevens, 1956), que culminaria na investigação de Lacerda & Head (1966), desta feita com recurso a um "pneumocromográfico":

> *Nasal vowels in Portuguese have an initial segment whose degree of nasality varies very greatly. In some cases the nasality may be so slight that in practice we can regard it as being absent; that is to say, we can regard the initial segment as being oral.* (Lacerda & Strevens, 1956, p.15-16)

> *O decurso das vogais nasais e ditongos nasais é sempre oral-nasal com nasalidade médio-final ou final, excepto no caso de ocorrer uma consoante nasal anterior. Neste caso, o decurso é inteiramente nasal, i.e., com vibrações nasais desde início até final.* (Lacerda & Head, 1966, p.67)

A presença/ ausência de uma consoante após a vogal nasal, bem como o seu estatuto, tem vindo a suscitar, desde há muito, uma acesa discussão entre os investigadores portugueses.

A polémica reflecte-se já nas opções ortográficas dos gramáticos do século XVI: Oliveira (1536) assinala a nasalidade com um til (˜), ao passo que Barros (1540) opta em favor de uma consoante nasal após a vogal.

A postulação de um segmento nasal depois da vogal parece depender do contexto considerado: 1) antes de consoante oclusiva; 2) antes de outras consoantes [18]; 3) em posição final, diante de pausa.

Quanto ao primeiro contexto, os especialistas tendem a considerar a presença de uma consoante nasal após a vogal, cujo ponto de articulação depende da oclusiva seguinte (Barbosa, 1961;

[18]Em geral, considera-se que as características intrínsecas da consoante pós-nasal (ponto e modo de articulação e vozeamento) afectam a nasalização da vogal e posterior queda da consoante nasal: as obstruintes surdas, nomeadamente as fricativas, são mais efectivas na promoção da elisão da consoante nasal do que outros segmentos (Busà, 2003; Ohala & Busà, 1995). Assim, a consoante nasal pode ser mais difícil de detectar antes de fricativas surdas do que antes de oclusivas (Ohala & Busà, 1995).

Viana, 1973b; Sá Nogueira, 1938; Lacerda & Hammarström, 1952; Nobiling, 1974) [19]. Louro (1954-1955, p.242) tem, no entanto, uma interpretação diferente: "No entanto, estas consoantes nasais, à semelhança do que se passa em francês devem considerar-se meramente gráficas, quer sejam seguidas de consoante oclusiva, quer de constritiva. É que, embora escrevendo-se na actual ortografia, em nenhum caso há consciência de se pronunciarem e, na realidade, também não lhes correspondem quaisquer movimentos articulatórios próprios".

Já o caso de vogal nasal seguida de outra consoante está longe de reunir o consenso: enquanto Viana (1903) nega em absoluto a existência de uma consoante nasal, outros admitem a sua presença (Barbosa, 1961; Head, 1964).

O mesmo acontece em relação à vogal nasal em posição final: embora alguns pesquisadores tenham assinalado a ocorrência mais ou menos sistemática de um segmento consonantal após a vogal final, nomeadamente no dialecto paulista (Nobiling, 1974) [20], em relação ao português em geral, considera-se que este segmento pode ou não manifestar-se (Sá Nogueira, 1938; Lacerda & Strevens, 1956; Barbosa, 1961) [21].

Pesquisas mais recentes vêm lançar nova luz sobre o comportamento do velo ao longo da produção das vogais nasais e sobre o estatuto destes segmentos consonânticos.

Estudos acústicos sobre as vogais nasais do PE e do PB (Drenska, 1988; Sousa, 1994; Seara, 2000) permitem concluir que a vogal, começa efectivamente, na maioria dos casos, como oral. O expoente nasal abrange cerca de metade a 4/5 do segmento vocálico, podendo a sua realização ser inteiramente nasal, quando este se encontra depois de consoante nasal (Drenska, 1988). Regista-se ainda, na maioria dos casos, a presença de um múrmurio/ apêndice/ consoante nasal (Drenska, 1988; Sousa, 1994; Seara, 2000; Medeiros *et alii*, 2008): de acordo com Drenska (1988), uma consoante nasal muito breve, na maioria dos casos apenas insinuada; segundo Sousa (1994, p.128) "um murmúrio coarticulado à vogal, e sem existência consonantal autónoma em relação a esta vogal", cuja duração varia muito segundo o informante, a vogal e a emissão [22].

Silva (1995) procedeu à análise das vogais nasais em diferentes contextos, recorrendo ao método *Recursive Least Square*. O investigador considera que a produção destes segmentos envolve dois estados estáveis, para além de uma zona de transição: 1) fase de posicionamento dos articuladores para a pronunciação da vogal oral, com uma duração entre os 60 e os 100 ms; 2) fase de transição

[19] Nas palavras de Viana (1973b, p.10-11), "Quando a uma vogal se segue consoante explosiva, além dessa vogal nasal ouve-se atenuada, reduzida, uma consoante nasal, homorgânica com essa explosiva".

[20] Com base em dados aerodinâmicos, Shosted (2003, 2006b) defende que alguns dialectos do Brasil mostram, efectivamente, sinais de emergência de uma coda nasal (usualmente uma consoante velar), após a vogal nasal final. Este processo é, sobretudo, condicionado pela altura e posterioridade da vogal, sendo o segmento consonântico muito mais evidente nas vogais nasais altas e posteriores.

[21] Observe-se a propósito, o exposto por Lacerda & Strevens (1956, p.8, nota 3): "The final segment of a word as *lã* is often followed by a noticeable sound that seems to be the release of an articulatory contact. The release, which is like a weak plosive, sometimes follows at a distance of one or two seconds from the end of the word, when the word is spoken as a statement of fact, in isolation. The contact being released is between the back of the tongue and the soft palate".

[22] Embora a vogal nasal exiba, muitas vezes, três fases, Sousa (1994) e Seara (2000) reportam vários casos em que uma delas está ausente.

correspondente à abertura da passagem velo-faríngea, com uma duração de 30 a 50 ms; 3) fase estável resultante do acoplamento do tracto nasal, com características muito semelhantes para as diferentes vogais nasais.

Do mesmo modo, os dados de ressonância magnética recolhidos para o PB (Gregio, 2006) evidenciam diferentes posições assumidas pelos articuladores no decorrer da produção de vogais nasais. No conjunto, identificam-se três fases distintas: a primeira caracteriza-se pelo posicionamento elevado ou levemente abaixado do palato mole; a segunda fase faz-se acompanhar da abertura velo-faríngea necessária ao acoplamento nasal; a terceira fase evidencia um movimento do dorso da língua em direcção à região palato-alveolar.

Lovatto *et alii* (2007), recorrendo à fibroscopia, distinguem duas fases: a parte vocálica (V) e um "*nasal tail*" (N). Esta última, geralmente mais curta do que V, está presente em 85% dos casos analisados. O velo atinge a sua posição mais baixa antes ou durante o murmúrio nasal.

Também os dados aerodinâmicos (Medeiros *et alii*, 2008) confirmam a presença de um apêndice nasal a seguir à vogal nasal, muito mais curto do que uma consoante nasal em posição inicial de sílaba (tónica ou átona).

5.3 Gestos orais

A presente secção centra-se na descrição e análise do comportamento assumido pelos articuladores orais, nomeadamente o dorso da língua, durante a produção das vogais nasais. Na caracterização articulatória destas últimas, a dimensão oral tem sido, muitas vezes, descurada em detrimento do estudo do movimento do véu palatino. Embora a nasalidade seja, em larga medida, uma consequência do movimento deste articulador, estudos articulatórios, realizados para duas línguas marcadas por fortes índices de nasalidade, como o francês (Delvaux *et alii*, 2002; Demolin *et alii*, 2003) e o português (Matta Machado, 1993; Gregio, 2006; Martins, 2007; Martins *et alii*, 2007, 2008a), evidenciam outros ajustes ao nível da cavidade oral, que permitem diferenciar as vogais nasais das suas correspondentes orais.

À semelhança do capítulo anterior, começaremos por uma descrição articulatória geral, de natureza qualitativa, assente na visualização de imagens de RM. Seguidamente, enveredaremos por uma comparação entre vogais orais e vogais nasais, com base na sobreposição dos contornos sagitais obtidos a partir das imagens. Por fim, avançaremos para aquele que é o nosso objectivo principal: a obtenção dos parâmetros quantitativos referentes ao CD e CL, que viabilizarão uma descrição gestual das vogais nasais.

A análise dos perfis médio-sagitais dos informantes AND, JHM e RAQ, referentes à produção das cinco vogais nasais - [ɐ̃], [ẽ], [ĩ], [õ] e [ũ] - em versão sustida, induzida a partir de uma palavra de referência (neste caso "canto", "pente", "pinto", "ponte" e "punto") (vd. figura 5.1), permite-nos, confirmar, antes de mais, e sem qualquer tipo de surpresa, que o véu palatino se encontra efectiva-

mente descido para todas as vogais nasais, à excepção da vogal [ũ], quando articulada pelo informante JHM. Em relação a este facto particular, cumpre notar que as imagens adquiridas para este informante são, de um modo geral, menos credíveis do que as registadas para os sujeitos AND e RAQ: a falta de treino fonético do informante comprometeu, em alguns casos, a fiabilidade das produções, que não terão sido realizadas em conformidade com as instruções.

Ainda em relação à altura do velo, regista-se alguma variabilidade em função da altura da vogal, um fenómeno já salientado noutros estudos sobre o português (Matta Machado, 1993; Martins *et alii*, 2007, 2008a) [23]. No caso do [õ] e do [ũ], o véu palatino chega mesmo a tocar a região posterior do dorso da língua, não em virtude de um maior abaixamento do velo, mas como consequência da posteriorização e elevação da própria língua. Tal comportamento foi já constatado na literatura por Stevens (1998), Gregio (2006) e Martins (2007).

No que se refere aos lábios, as vogais nasais posteriores, sobretudo o [ũ], são marcadas, à semelhança das suas congéneres orais, por algum arredondamento labial, o que não acontece em relação às restantes vogais nasais.

Quanto à posição da língua, as vogais [ĩ] e [ẽ] são claramente altas e anteriores. Também no caso do [ũ], a língua se encontra numa posição elevada, embora bastante mais posteriorizada, quando comparada com as outras duas vogais ([ĩ] e [ẽ]). Também na produção do [õ] - uma vogal tradicionalmente considerada posterior, a par do [ũ] - se assiste a um recuo do dorso da língua, não tanto em direcção à região do velo, como acontece no [ũ], mas em direcção à parede da faringe. Na passagem de [ũ] para [õ], a língua baixa para uma altura similar à do [ɐ̃]. Contudo, no caso desta última vogal, a tendência para a posteriorização é menos acentuada, o que se reflecte num aumento da área da faringe e numa configuração mais neutra (pelo menos para os informantes AND e RAQ). Considerando o movimento de recuo do dorso da língua em direcção à parede da faringe, confirma-se a seguinte progressão, atestada em estudos anteriores para o português do Brasil (Matta Machado, 1993): [ɐ̃] < [ũ] < [õ].

As comparações entre as vogais nasais e as suas correspondentes orais baseiam-se na sobreposição dos contornos, apresentados nas figuras 5.2 a 5.6.

Quando comparada com [a] e [ɐ], a vogal nasal [ɐ̃] (figura 5.2) é caracterizada por uma anteriorização do dorso da língua, tal como descrito noutros estudos para o PB (Gregio, 2006). Acreditamos que esta mudança na postura do dorso da língua possa estar relacionada com o abaixamento do véu palatino, que, no caso particular da vogal [ɐ̃] é obrigado a descer muito para criar um nível de acoplamento nasal suficiente para a percepção de nasalidade (Maeda, 1993), arrastando consigo a língua. Os ajustes no posicionamento deste último articulador têm efeitos na cavidade faríngea, que é maior na produção da nasal, do que nas duas vogais orais. No restante, a configuração da língua na vogal nasal é muito similar à do [ɐ], pelo menos no caso dos informantes JHM e RAQ. Para o AND,

[23] A questão da altura do velo será alvo de análise em secções posteriores deste estudo.

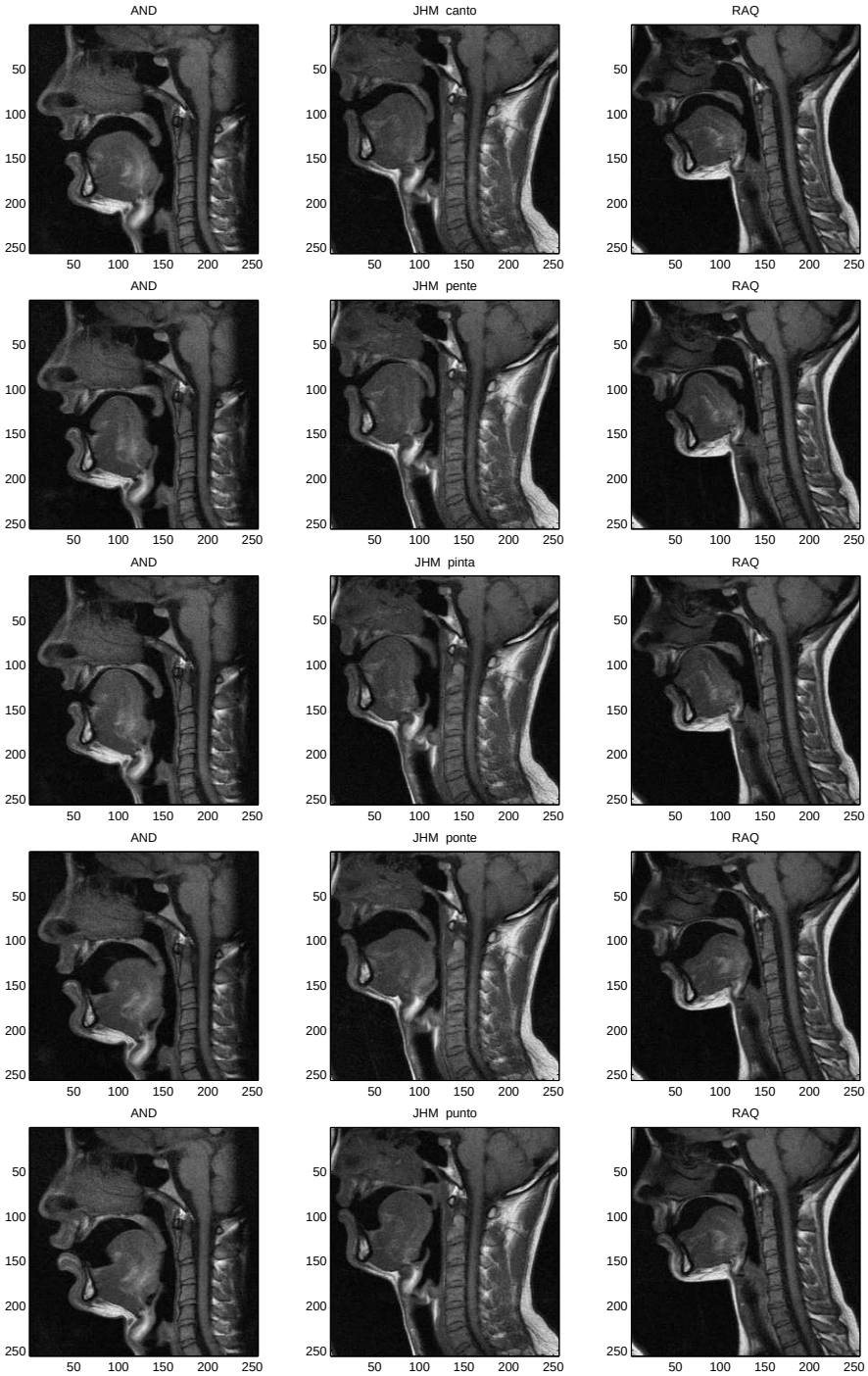
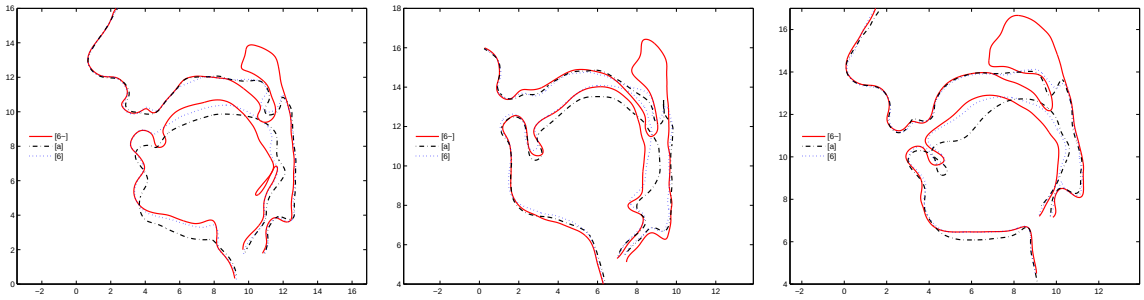


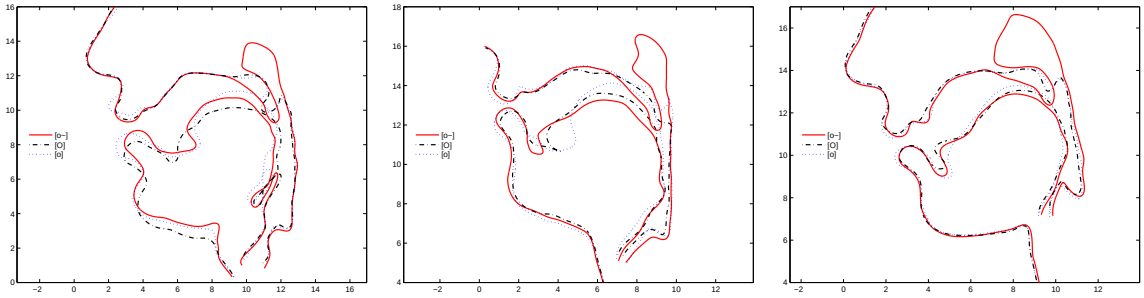
Figura 5.1: Imagens de ressonância magnética das cinco vogais nasais do PE ([ɐ̃], [ẽ], [ĩ], [õ] e [ũ]), produzidas pelos três informantes (AND, JHM e RAQ), a partir das palavras de referência “canto”, “pente”, “pinta”, “ponte” e “punto”.



(a) AND (b) JHM (c) RAQ

Figura 5.2: Sobreposição dos contornos relativos à configuração do tracto vocal para as vogais [ɐ̃] (vermelho), [a] (preto) e [ɐ] (azul), produzidas pelos três informantes: AND (a), JHM (b) e RAQ (c).

a região anterior da língua eleva-se ligeiramente em relação ao [ɐ].



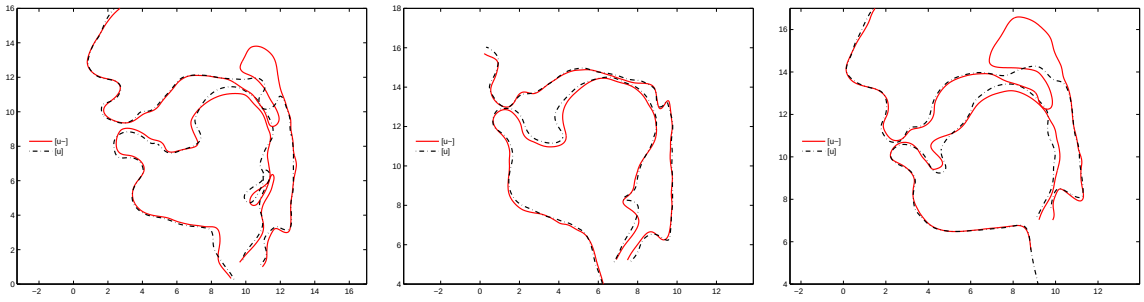
(a) AND (b) JHM (c) RAQ

Figura 5.3: Sobreposição dos contornos relativos à configuração do tracto vocal para as vogais [õ] (vermelho), [o] (azul) e [ɔ] (preto), produzidas pelos três informantes: AND (a), JHM (b) e RAQ (c).

Uma outra vogal a registar diferenças em relação às suas correspondentes orais é a nasal [õ] (figura 5.3). À semelhança do que acontece com o [ɐ̃], o dorso da língua tende a avançar e baixar um pouco em relação ao [o] (e, no caso dos informantes RAQ e JHM, também em relação ao [ɔ]), presumivelmente em consequência do abaixamento do velo, que chega mesmo a tocar a zona posterior do dorso da língua, em dois dos informantes (RAQ e AND). No caso dos sujeitos masculinos ficam ainda patentes alterações significativas na região anterior da língua: no caso do AND, a configuração da vogal nasal aproxima-se da do [o], enquanto que para o JHM tende a assemelhar-se mais com a do [ɔ]. Ao nível da cavidade faríngea, não se observam modificações substanciais entre as três vogais, a não ser um ligeiro estreitamento da vogal nasal em relação ao [o], no caso do informante AND e RAQ, que não se verifica, contudo, para o terceiro informante (JHM).

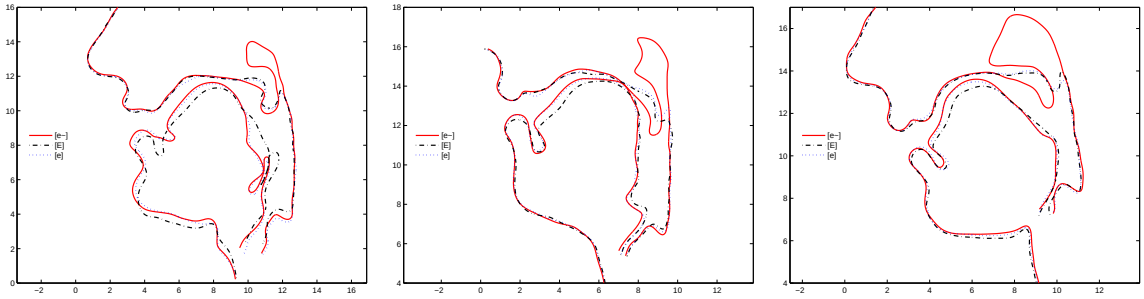
A comparação de [ũ] com [u] (figura 5.4) revela configurações muito similares. Contudo, à semelhança do descrito para o [õ], na vogal nasal, o dorso da língua baixa ligeiramente na zona do velo e a cavidade faríngea sofre um estreitamento [24].

[24]Pelas razões apontadas anteriormente nesta secção, as observações feitas em relação ao [ũ], apenas têm em conta as produções de dois informantes (AND e RAQ).



(a) AND (b) JHM (c) RAQ

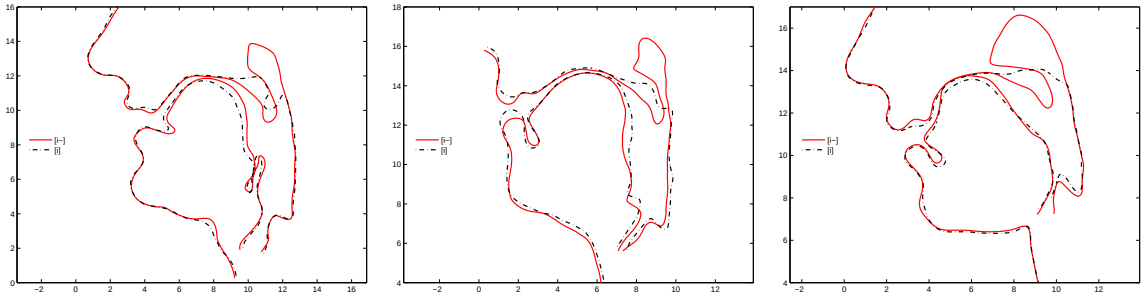
Figura 5.4: Sobreposição dos contornos relativos à configuração do tracto vocal para as vogais [ũ] (vermelho) e [u] (preto), produzidas pelos três informantes: AND (a), JHM (b) e RAQ (c).



(a) AND (b) JHM (c) RAQ

Figura 5.5: Sobreposição dos contornos relativos à configuração do tracto vocal para as vogais [ẽ] (vermelho), [e] (azul) e [ɛ] (preto), produzidas pelos três informantes: AND (a), JHM (b) e RAQ (c).

Para os três informantes analisados, os contornos relativos ao [ẽ], apresentados na figura 5.5, são praticamente indistintos do [e], com excepção do velo, que, no caso da vogal nasal, se encontra numa posição abaixada, favorecendo a passagem do ar pelas fossas nasais. O mesmo é dizer que a configuração oral da vogal nasal se aproxima muito mais do [e] do que da vogal oral [ɛ].



(a) AND (b) JHM (c) RAQ

Figura 5.6: Sobreposição dos contornos relativos à configuração do tracto vocal para as vogais [ĩ] (vermelho) e [i] (preto), produzidas pelos três informantes: AND (a), JHM (b) e RAQ (c).

O mesmo se observa em relação ao par [ĩ] e [i] (figura 5.6), sendo que, também neste caso,

a sobreposição dos respectivos contornos revela configurações muito similares no tocante à postura do dorso da língua. Para além das já esperadas alterações ao nível do velo, há ainda a registar, para o sujeito JHM, um aumento da cavidade faríngea, para a vogal nasal. Inversamente, no caso do informante AND a língua recua um pouco, subindo ligeiramente na região de máxima constrição.

Concluindo, os dados articulatórios acima descritos evidenciam pequenas diferenças entre as vogais nasais e as suas correspondentes orais - para além do posicionamento do velo - sobretudo em relação ao [ɐ̃] e [õ]. Em termos gerais, estes resultados estão em linha com a literatura publicada sobre as vogais nasais do PB (Gregio, 2006; Master *et alii*, 1991), onde se relatam mudanças mais acentuadas nas vogais nasais posteriores do que nas anteriores, de acordo com a seguinte progressão [ɐ̃]> [õ]> [ũ].

Do mesmo modo, e desta feita em relação ao francês, vários estudos (Zerling, 1984; Demolin *et alii*, 1998; Delvaux *et alii*, 2002; Demolin *et alii*, 2003) destacam as diferenças articulatórias entre as vogais nasais e as suas congéneres orais, não só ao nível da posição do velo, mas também no que se refere à postura do dorso da língua (e grau de protrusão e arredondamento dos lábios). Nas vogais nasais, este articulador tende a recuar consideravelmente, com reflexos acentuados ao nível da cavidade faríngea, que se apresenta invariavelmente mais constringida do que nas vogais orais. Segundo Maeda (1993), estas manobras articulatórias desempenham um papel essencial na implementação do contraste de nasalidade e "velum lowering, labialization and tongue backing should therefore be considered as a single articulatory complex used to produce nasal counterparts of these [oral] vowels" (Maeda, 1993, p.165).

Não obstante as mudanças articulatórias atestadas entre as vogais nasais e as vogais orais do PE, essas diferenças estão longe de atingir a dimensão registada para o francês. Enquanto, nesta última língua, a redução da cavidade faríngea parece ser um factor determinante para a distinção articulatória entre vogais nasais e as suas correspondentes orais, no português, o estreitamento faríngeo não segue um padrão consistente para todas as vogais nasais (cf. Matta Machado, 1993) e nem sequer para todos os sujeitos. Paralelamente, não existem ajustes articulatórios dignos de nota ao nível dos lábios.

Os dados mencionados anteriormente - nomeadamente os que indicam que as diferenças articulatórias entre vogais orais e nasais, ao nível do posicionamento da língua, não são muito acentuadas - permitem-nos enveredar por uma caracterização gestual das vogais nasais, baseada na descrição das vogais orais correspondentes.

O resultado das medições, que tiveram como propósito a obtenção dos parâmetros articulatórios quantitativos, correspondentes ao local (CL) e grau de constrição (CD), para as vogais nasais, é apresentado na figura 5.7. Aí se mostram também os valores dos parâmetros obtidos para as vogais orais com a configuração mais próxima, de modo a facilitar as comparações. Lembramos, uma vez mais, que os valores foram estimados com base nos perfis articulatórios do informante AND.

A posição do velo dificulta, em muito, a obtenção do CL e CD a partir dos perfis médio-

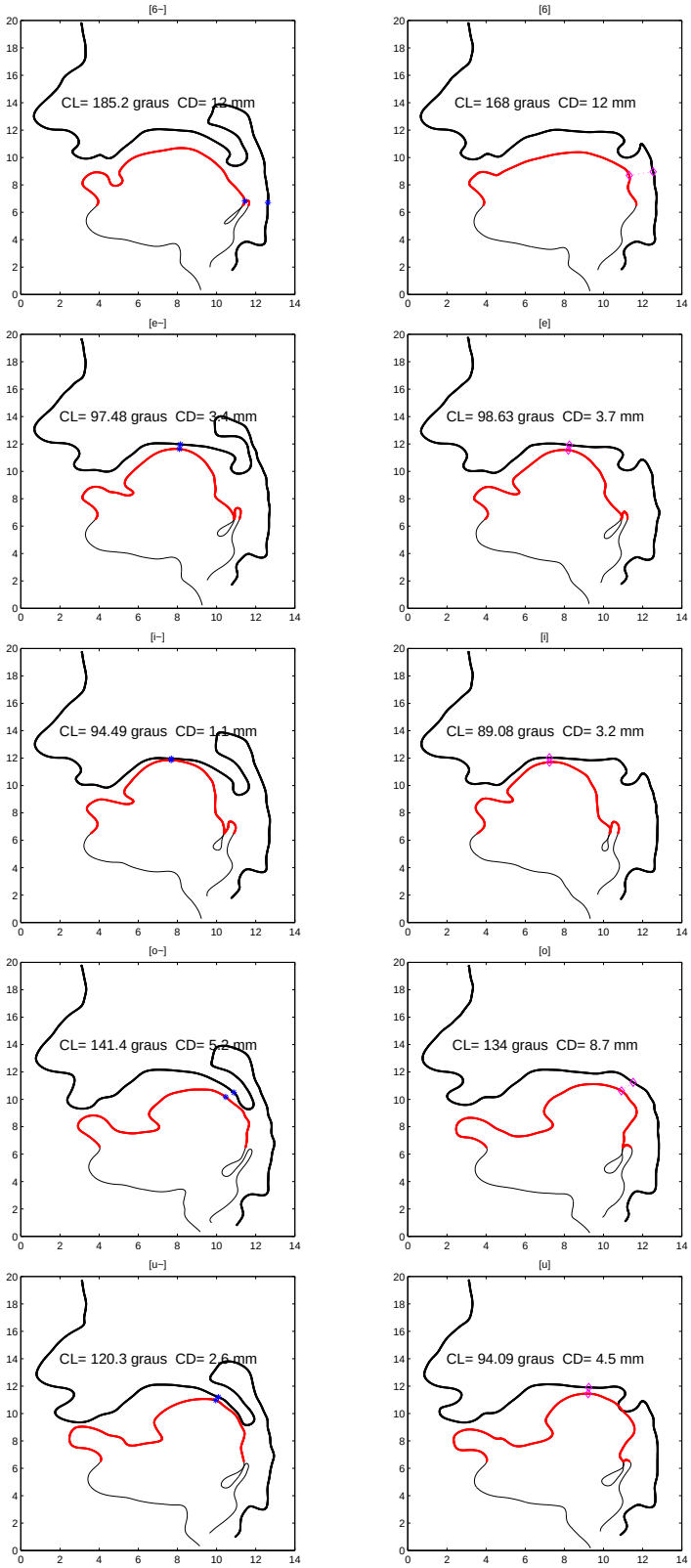


Figura 5.7: Valores de CL (*constriction location*), em graus, e de CD (*constriction degree*), em mm, para as vogais nasais (à esquerda) e para as vogais orais (à direita), obtidos a partir dos contornos médio-sagitais do informante AND.

sagitais, pelo que os valores apresentados devem ser encarados como uma aproximação, um ponto de partida para a definição das configurações gestuais finais. Estas foram obtidas tendo em conta não só estas medições, mas também os testes auditivos preliminares, que ditaram pequenos ajustes aos valores obtidos inicialmente.

Grande parte destas dificuldades poderá ser ultrapassada através do recurso a dados RM 3D, que garantam informações mais precisas sobre a área.

Assim, à vogal nasal [ɐ̃] está associado um gesto de corpo da língua, cujo local de constrição (TBCL) poderá ser identificado como palatal (PAL) ou faríngeo (PHAR). As dificuldades na definição do TBCL são uma consequência directa da já referida anteriorização do corpo da língua, que provoca uma diminuição da área na região palatal. A apreciação prévia da qualidade do som gerado a partir da configuração palatal veio, contudo, demonstrar que a alteração do local de constrição de faríngeo para palatal não permitia distinguir o [ɐ̃] do [ẽ], pelo que optámos por manter o CL oral. A distância entre a língua e o palato fixou-se nos 12 mm.

Quanto à variável de corpo da língua, considerada nas suas duas dimensões (ponto e grau de constrição), a vogal nasal [ẽ], à semelhança da sua congénere oral, caracteriza-se por um TBCL palatal (PAL) e um TBCD, que, tal como na vogal oral, foi ajustado para os 6 mm.

A vogal nasal [ĩ] apresenta uma constrição palatal (PAL) e um valor de TBCD vocálico de 2 mm. Comparativamente à vogal oral [i], cujo TBCD foi fixado nos 3 mm, assiste-se a uma ligeira redução da distância entre o palato duro e o corpo da língua.

A vogal [õ] é definida por uma constrição uvo-faríngea (UVOPHAR) localizada nos 140 graus, e um grau de constrição estimado nos 11 mm. Esta pequena mudança no valor do TBCD - dos 9 mm (vogal oral) para os 11 mm (vogal nasal) - procura reflectir o abaixamento do dorso da língua, em consequência da descida do véu palatino.

Finalmente, a vogal nasal [ũ] foi associada a um TBCL velar (VEL), localizado nos 125 graus, e um TBCD de 6 mm. Tal como no [õ], o dorso da língua, na vogal nasal, tende a baixar um pouco na zona do velo, pelo que o TBCD foi também ligeiramente aumentado.

Quanto ao gesto labial, que caracteriza as vogais nasais posteriores [õ] e [ũ], optámos por manter os valores utilizados para as vogais orais. Muito embora, para o informante AND, se registe uma maior protusão labial (PRO) no [ũ] do que no [õ], esta diferença não é perceptível para os outros dois informantes. Ficam apenas registadas as diferenças ao nível da abertura labial (LA) nas duas vogais nasais, o que se traduz num valor de 4 mm para o [õ] e de 2 mm para o [ũ].

A informação gestual relativa às vogais nasais foi compilada na tabela 5.2, segundo as exigências formais do modelo computacional.

Vogal	Organ	Osc	TV	Const	Target	Stiff
ɐ̃	TB	v	TBCL	PHAR	.	.
	TB	v	TBCD	V	12	.
ẽ	TB	v	TBCL	PAL	.	.
	TB	v	TBCD	V	6	.
ĩ	TB	v	TBCL	PAL	.	.
	TB	v	TBCD	V	2	.
õ	TB	v	TBCL	UVUPHAR	140	.
	TB	v	TBCD	V	11	.
	Lips	v_rnd	LP	PRO	12	.
	Lips	v_rnd	LA	NAR	4	.
ũ	TB	v	TBCL	VEL	125	.
	TB	v	TBCD	V	6	.
	Lips	v_rnd	LP	PRO	12	.
	Lips	v_rnd	LA	NAR	2	.

Tabela 5.2: Definição gestual (gesto do dorso da língua e gesto dos lábios) das vogais nasais do PE.

5.4 Gesto nasal

5.4.1 Caracterização do problema: hipóteses e questões a explorar

Como deixámos expresso na Introdução, o objectivo final deste trabalho reside na descrição da nasalidade vocálica do português à luz da FA. Este modelo linguístico postula - a par de um gesto vocálico, especificado através das variáveis do tracto grau e local de constrição do corpo da língua - um gesto de abertura vélica, que procuraremos caracterizar através de um estudo experimental, quer em termos de parâmetros dinâmicos, quer na sua relação temporal com o gesto vocálico e o gesto consonantal seguinte (no caso deste existir).

Na presente secção, com base na revisão teórica e bibliográfica desenvolvida no início deste capítulo, formulamos e fundamentamos as questões centrais a investigar no estudo experimental acerca do comportamento do velo.

1. *Q1: Existem diferenças na altura do velo durante a produção de vogais nasais e consoantes nasais?*

   Estudos articulatórios anteriores, contemplando línguas como o francês (Rossato *et alii*, 2003; Amelot & Rossato, 2006, 2007) ou o português (Rossato *et alii*, 2006) evidenciam uma maior amplitude vélica para as vogais nasais do que para as consoantes nasais. De acordo com os objectivos a que nos propusemos, a confirmação deste resultado para o português, poderá implicar a atribuição de um terceiro grau de abertura vélica às vogais nasais, para além do fechado (CLO= -0.1 mm) que caracteriza os sons orais (oclusivas e fricativas) e do aberto (WIDE=0.2 mm) que define as consoantes nasais. Para que tal se verifique, a par das diferenças articulatórias, importa ainda avaliar a pertinência de tal distinção, do ponto de vista perceptual.

2. *Q2: Há uma relação entre a altura da vogal e a altura do velo?*

   Como se viu, vários estudos (e.g. Bell-Berti *et alii*, 1979; Moll, 1962; Rossato *et alii*, 2003) estabelecem uma relação entre a altura do velo e a posição da língua. Segundo estes, as vogais baixas tendem a ser produzidas com o velo numa posição mais baixa do que as vogais altas. Porquanto esta relação entre altura da vogal e amplitude do velo não parece assumir padrões consistentes, nem tão pouco universais (cf. Hajek, 1997), interessa-nos analisar o comportamento do velo durante a produção das vogais orais e, sobretudo, das vogais nasais.

   Caso os dados venham a confirmar algum tipo de interacção entre qualidade da vogal nasal e amplitude do velo, níveis individuais de abertura vélica (*targets*) podem vir a ser estabelecidos para cada vogal.

3. *Q3: Qual a duração do gesto de abertura (e fecho) do velo?*

   O estudo da duração do gesto do velo visa, sobretudo, possibilitar as experiências de coordenação desenvolvidas posteriormente. Com base em estudos prévios (Stevens, 1998; Teixeira *et alii*, 2001; Amelot, 2004; Basset *et alii*, 2006; Oliveira & Teixeira, 2007b), estima-se que o ciclo de abertura e fecho do velo dure cerca de 200 a 300 ms. Prevemos, contudo, que essa duração possa ser afectada por factores segmentais, contextuais e, sobretudo, pela taxa de elocução.

4. *Q4: Qual a velocidade do gesto do velo?*

   Embora pouco explorada na literatura fonética, a questão da velocidade dos movimentos articulatórios "is crucial for a better understanding of the temporal aspects of speech production" (Roon *et alii*, 2007, p.409). No modelo *task-dynamics*, usado pela Fonologia Articulatória, o parâmetro relacionado com a velocidade dos gestos articulatórios, com consequências complexas ao nível da sua duração, é o *stiffness*. Apesar do recente interesse dos investigadores por este parâmetro (Roon *et alii*, 2007, 2008) e pelos seus efeitos prosódicos (e.g. Byrd & Saltzman, 1998), não são conhecidos estudos dedicados à relação entre o *stiffness* e o velo. Seguindo a metodologia usada por Roon *et alii* (2007) para outros articuladores, procurámos estimar os valores de *stiffness* para o gesto de abertura e fecho do velo, avaliando possíveis implicações em termos de coarticulação com gestos orais subsequentes.

5. *Q5: Como se organizam os gestos orais, nasais e glotais na produção de vogais nasais, em diversos contextos?*

   Tendo em mente os postulados teóricos da FA e estudos anteriores que evidenciam o carácter dinâmico da nasalidade em português, antecipa-se que o gesto de abertura vélica possa começar depois do início do gesto vocálico, induzindo a chamada fase oral da vogal nasal. Seguir-se-á um gesto de fecho do velo que poderá sobrepor-se ao gesto consonantal seguinte (no caso de o haver), habilitando ou não a emergência de uma consoante nasal, dita intrusiva. A presença e

duração desta última está, pois, dependente do grau de sobreposição entre dois gestos consecutivos (Albano, 1999). Admite-se que este possa ser afectado por vários factores, segmentais e prosódicos, que procuraremos averiguar no decurso do estudo experimental.

5.4.2 Metodologia

Nas secções seguintes, pretendemos dar conta dos métodos e procedimentos adoptados no decurso do estudo exploratório, através do qual procurámos responder a alguns dos problemas e questões levantadas na secção anterior.

Para a definição do protocolo experimental a aplicar inspirámo-nos noutros estudos de base articulatória, na área das vogais nasais (e.g. Amelot & Rossato, 2007), mas contámos também com a experiência prévia, adquirida a partir da análise da base de dados EMA, gravada em 2001 (Teixeira, 2001; Oliveira & Teixeira, 2007b).

Dar-se-á notícia das características do *corpus* gravado em EMA, bem como dos critérios que presidiram à sua elaboração; do perfil dos informantes; dos procedimentos adoptados durante a recolha dos dados; e posterior tratamento da informação que serve de base à análise linguística efectuada.

Das várias técnicas actualmente disponíveis para estudar o comportamento dos diferentes articuladores da fala (incluindo o velo), optámos pela Articulografia Electromagnética 2D. Para além de motivos de ordem prática, relacionados com as facilidades de acesso ao equipamento e garantias de apoio técnico durante a recolha dos dados - proporcionadas pelos contactos estabelecidos com um conjunto de investigadores do GIPSA-lab, no âmbito de um projecto de investigação - outras razões, que se prendem com as características do método em si mesmo, estão na origem desta escolha.

No sistema EMA, um conjunto de bobines transmissoras (ou emissoras) é montado num capacete (*helmet*), especialmente construído para esse efeito, e colocado em torno da cabeça do informante. Cada uma delas gera um campo magnético específico. Tal como num transformador, os campos magnéticos alternados, gerados pelas bobines emissoras, induzem tensão eléctrica nas bobines receptoras, coladas em pontos específicos dos articuladores (língua, lábios, velo,...). A tensão eléctrica induzida em cada uma das bobines receptoras é inversamente proporcional ao cubo da distância entre bobine emissora e receptora.

Entre as principais vantagens da Articulografia Electromagnética contam-se a possibilidade de obtenção de informação dinâmica sobre a trajectória dos vários articuladores (lábios, língua, velo,...) em simultâneo e em tempo real. Esta característica faz do EMA um dos métodos mais eficazes no estudo da coordenação inter-articuladores (Stone, 1999), ainda que a informação obtida se limite ao plano sagital e seja relativa a apenas alguns pontos [25].

[25]Com a articulografia 3D (e.g. sistema AG 500, da Carstens), esta e outras limitações - relacionadas, por exemplo, com a imobilização da cabeça mediante um capacete - são em parte, ultrapassadas, já que é possível recolher informação para lá do plano médio-sagital.

Considerado um método relativamente inócuo, não estão, contudo, totalmente esclarecidos os efeitos biológicos de longo-prazo decorrentes da exposição prolongada a este tipo específico de campos magnéticos. Assim sendo, não se recomenda o recurso a sujeitos portadores de *pace-maker* ou mulheres grávidas (Hoole & Nguyen, 1999).

Uma outra desvantagem associada a esta técnica prende-se com a colagem dos sensores, uma tarefa bastante árdua e morosa, em particular no caso do sensor do velo, cuja fixação exige esforços adicionais e soluções alternativas à cola cirúrgica, como a colagem indirecta (Wrench, 1999) e até a sutura (Engelke, 1991, apud Hoole & Nguyen, 1999).

Embora a interferência dos sensores EMA na articulação possa ser considerada irrelevante e em tudo similar à causada pelos chamados *pellets* do sistema *X-Ray Microbeam* (Hoole & Nguyen, 1999), há que notar que a Articulografia Electromagnética não deixa de ser um método invasivo e, por isso mesmo, pode causar um certo desconforto, pelo menos a alguns sujeitos.

Existem vários sistemas de Articulografia Electromagnética, como o *Electro Magnetic Mid-Sagital Articulography* desenvolvido por Perkell *et alii* (1992), os sistemas Carstens 2-D AG 100, 2-D AG 200 e 3-D AG 500 (http://www.articulograph.de) e o sistema *Movetrack*, que fundamentalmente variam no número de bobines emissoras e receptoras utilizadas.

Devido ao seu custo elevado, não são muitas as universidades a dispor deste tipo de equipamento. No plano nacional, nenhuma universidade possui um sistema deste género.

Nesta experiência, os dados articulatórios foram adquiridos através do sistema AG-100, da Carstens, propriedade do GIPSA-LAB, *Université Stendhal* (Grenoble).

Corpus EMA

Na selecção do material linguístico que integra o *corpus* adquirido foram tidos em consideração vários outros factores, para além do estudo do movimento do velo e da sua relação com os restantes articuladores orais, nomeadamente 1) a necessidade de completar e estender o *corpus* gravado em 2001 (Teixeira, 2001), recorrendo à mesma técnica, e 2) a possibilidade de realizar análises comparativas entre o francês e o português. O mesmo é dizer que a recolha dos dados não se restringiu aos objectivos específicos deste trabalho, estando integrada num projecto mais lato, que visa permitir efectuar vários tipos de estudos sobre a nasalidade, seja ela vocálica ou consonântica. Assim sendo, muito do material linguístico abaixo descrito não será alvo de análise no seio desta dissertação.

Por razões metodológicas - entre outros critérios que se prendem com as razões acima enunciadas e com a gestão e racionalização do tempo disponível para gravação - os dados foram organizados por ordem de prioridade e agrupados em diferentes secções, que passamos a descrever sumariamente:

1. A primeira parte do *corpus* inclui todas as vogais orais e nasais do PE, em contexto isolado.

2. A segunda parte visa cumprir um objectivo mais específico relacionado com a obtenção de medidas de amplitude do velo para todos os sons do PE. Assim, foram consideradas todas as consoantes do português, em contexto [#VCV#], sendo que a V correspondem as cinco vogais nasais ([ɐ̃], [ɛ̃], [ĩ], [õ], [ũ]) e as três vogais cardinais ([i], [u], [a]).

3. O material da terceira parte do *corpus* é composto por sequências com vogais nasais nas três posições lexicais possíveis: inicial, medial e final (e.g. [ˈɐ̃.pɐ], [ˈpɐ̃.pɐ], [ˈpɐ̃]). Quanto ao contexto fonético, a vogal surge sempre em posição acentuada e entre oclusivas (bilabiais ou dentais) ou fricativas, procurando: 1) evitar, dentro do possível, efeitos de coarticulação e garantir uma mais fácil e rigorosa segmentação do sinal acústico; 2) contemplar segmentos, cuja articulação não envolva activamente o dorso da língua, implicado na produção das vogais nasais e alvo privilegiado do nosso estudo, na sua relação com o véu palatino. A opção por contextos absolutamente simétricos implicou, necessariamente, o recurso a logátomos, em vez de palavras. Estes foram inseridos numa frase de suporte ("Diz...três vezes"), com pequenos ajustes que visaram impedir a ocorrência de fenómenos de coarticulação e ressilabificação em posição inicial e final de palavra.

4. A quarta parte do *corpus* almeja possibilitar estudos comparativos sobre o comportamento do velo no francês e no português. Assim, seleccionaram-se contextos em que as vogais (orais e nasais) surgem flanqueadas de consoantes nasais, seguindo a estrutura proposta no *corpus* EMA francês.

5. Os contextos considerados nesta secção correspondem a palavras reais dissilábicas de estrutura [CV.CV], produzidas sem qualquer tipo de frase de suporte. Na primeira palavra da sequência, a primeira vogal é nasal e a consoante seguinte uma oclusiva ou fricativa. Nas sequências seguintes, a vogal é oral e a consoante seguinte mantém o ponto de articulação da palavra anterior, variando apenas quanto à posição do velo (e.g. [ˈkɐ̃.pɐ], [ˈkɐ.mɐ], [ˈkapɐ]). Os objectivos prendem-se com a análise das características do gesto oral implicado na produção da consoante, em contexto de vogal oral e nasal.

6. Seguidamente, consideraram-se vogais nasais precedidas de consoante nasal em confronto com vogais orais antecedidas de consoante nasal, em contextos simétricos, que perfiguram, em alguns casos, algo similar a pares mínimos (e.g. [ˈmɐ̃.tu], [ˈma.tu]). Para além destes, acrescentaram-se ainda algumas palavras comuns, onde a vogal nasal se encontra igualmente precedida de consoante nasal.

7. As duas últimas partes do corpus são muito menos importantes do ponto de vista dos objectivos a alcançar, pelo menos no âmbito do estudo aqui relatado. A primeira, largamente inspirada no *corpus* EMA francês, prende-se com o estudo do espraiamento da nasalidade ao longo da palavra e da resistência dos vários sons à propagação da mesma, pelo que inclui palavras de origem culta, muitas delas com sequências consonânticas pouco frequentes. A última parte do

corpus integra palavras de uso comum e visa a análise das consoantes nasais e, eventualmente, vogais nasalizadas por acção do contexto.

O conteúdo global do *corpus* encontra-se resumido no anexo C.

Informantes

Neste estudo, participaram dois informantes adultos, falantes nativos do português europeu, um do sexo masculino (AT) e outro do sexo feminino (LF), com o seguinte perfil:

(a) AT, indivíduo do sexo masculino com 39 anos de idade, cerca de 99 Kg de peso e 190cm de altura, natural do concelho de Paredes, distrito do Porto, habilitações literárias ao nível do Doutoramento, com treino fonético. Este sujeito tinha já participado, enquanto informante, numa outra experiência similar (Teixeira, 2001). Para além disso, apresenta algumas marcas dialectais que o diferenciam, de alguma forma, da norma-padrão da língua.

(b) LF, informante do sexo feminino com 27 anos de idade, 60Kg de peso e 171cm de altura, também proveniente da zona Norte do país, com formação superior ao nível do mestrado, sem treino fonético.

À data da recolha, nenhum dos informantes apresentava qualquer tipo de perturbação da fala e/ou da linguagem. Ambos foram informados, mediante documento escrito (anexo D), dos propósitos da experiência, bem como das vantagens e contra-indicações da utilização da tecnologia EMA. Todos os sujeitos aceitaram participar gratuitamente na recolha dos dados.

Os dados do informante masculino não foram considerados para o presente estudo, nem, portanto sujeitos a análise, devido a razões que se prendem com o posicionamento pouco favorável do sensor do velo [26], o que levantou sérios problemas ao estudo da sua altura e dinâmica temporal.

Recolha dos dados

A aquisição dos dados decorreu no final de Outubro de 2007, em duas sessões distintas, no GIPSA-LAB, *Université Stendhal*, Grenoble, sob a orientação técnica de Christophe Savariaux. Entre sessões de preparação e recolha propriamente dita, todo o processo levou cerca de 3 dias.

As tarefas preparatórias [27] encontram-se devidamente especificadas em Oliveira & Teixeira

[26] Idealmente, o sensor do velo deve ser colado na zona mais móvel do articulador, aproximadamente entre o início do véu palatino (junção com o palato duro) e a extremidade da úvula. No caso do informante AT, o sensor ficou demasiado próximo do palato duro, numa zona com menos flexibilidade do que seria desejável.

[27] Para mais informações sobre os procedimentos e regras de segurança a ter em conta durante a aquisição de dados EMA, consultar o site do fabricante (www.articulograph.de) e o manual da *University of California, Los Angeles* (UCLA) (www.linguistics.ucla.edu/faciliti/facilities/physiology/Emamual.html).

(2007a) e incluíram:

- **a criação do *corpus* final** - a reunião de preparação para a aquisição dos dados decorreu já em Grenoble, no dia anterior à recolha propriamente dita, e contou com a participação de alguns membros do GIPSA-LAB. Para além de algumas explicações técnicas sobre o EMA e metodologias a adoptar, foram ainda tomadas algumas decisões relativas ao *corpus*. Algumas partes do mesmo foram agrupadas, de maneira a rentabilizar o tempo disponível para a aquisição de cada item [28]. Tendo em conta a duração de cada uma das secções do *corpus*, foi possível calcular a duração aproximada de toda a sessão de recolha. Depois de codificados, os dados foram organizados por ordem de prioridade. Todos os informantes tiveram acesso à versão escrita do material antes da sua aquisição, ainda que o período de familiarização com o *corpus* tenha sido relativamente curto.
- **a limpeza e esterilização dos sensores** - antes de cada sessão, todos os sensores e pinças foram limpos através de imersão numa solução esterilizadora. Posteriormente, os sensores foram mergulhados em latex líquido. Este procedimento protege os sensores, impedindo o seu desgaste, ao mesmo tempo que torna todo o processo mais higiénico.
- **o aquecimento da máquina** - o sistema foi activado cerca de duas horas antes, de modo a permitir o aquecimento das bobinas, de acordo com o especificado no site da Carstens [29].
- **a calibração do sistema**
- **o ajuste da cadeira e do "capacete"** - o capacete que gera os campos magnéticos foi posicionado à volta da cabeça do sujeito. Em virtude do seu peso, o *helmet* foi suspenso, através de uma roldana presa ao tecto. Depois de devidamente ajustado à cabeça de cada informante, o capacete foi de novo removido para a colocação dos sensores.
- **a verificação do plano sagital**
- **a colagem dos sensores** - esta tarefa durou cerca de 15 a 20 minutos e contou com a colaboração de um segundo investigador, para enxugar, com um rolo de algodão, a superfície dos locais onde os sensores foram colados, enquanto o primeiro investigador se dedicou a colocar cola (similar à usada pelos dentistas) nos sensores e a fixá-los com uma pinça, começando pelo mais posterior (velo) e terminando com os sensores de referência. Parte do fio do velo foi colada ao palato, de modo a minimizar o reflexo de vómito e a interferência na produção de fala. Os restantes fios de ligação foram imobilizados exteriormente, através de fita adesiva. Antes de recolocar o capacete, cada um dos sujeitos teve oportunidade de se adaptar aos sensores, produzindo algumas frases de teste. Para cada um dos informantes, foram usados 8 sensores (vd. tabela 5.3).

[28] Convém aqui lembrar que a gravação é efectuada em pequenos segmentos individuais, de duração variável, mas que não ultrapassa, geralmente, os trinta segundos (uma a quatro frases).

[29] www.articulograph.de.

Tabela 5.3: Posicionamento dos 8 sensores EMA nos dois informantes (AT e LF).

Sensor	Informante AT	Informante LF
1	Nariz (referência)	Nariz (referência)
2	Incisivos Superiores (referência)	Incisivos Superiores (referência)
3	Véu Palatino	Véu Palatino
4	Dorso da Língua	Ponta/ Lâmina da Língua *(1.2 cm atrás do ápice)*
5	Ponta/ Lâmina da Língua *(1.2 cm atrás do ápice)*	Dorso da Língua
6	Incisivos inferiores (mandíbula)	Incisivos inferiores (mandíbula)
7	Lábio inferior	Lábio inferior
8	Lábio superior	Lábio superior

- **a ligação dos fios ao pré-amplificador** - cada um dos sensores foi ligado a um canal do pré-amplificador.

A gravação foi efectuada em pequenos segmentos (*sweeps*), de cerca de cinco a vinte segundos, sendo o início e o fim de cada um deles assinalado com um pequeno sinal sonoro. No total, cada sessão rondou, aproximadamente, os quarenta e cinco minutos [30].

A informante feminina foi a primeira a adquirir o *corpus*, tendo recebido previamente algumas instruções sobre a forma de produzir os sons e a taxa de elocução a adoptar nas diferentes secções, visto não possuir treino fonético, embora tenha trabalhado, no âmbito do seu trabalho de mestrado, com vogais nasais. O sujeito AT não necessitou de quaisquer indicações neste sentido, na medida em que participou activamente na definição do *corpus* e planeamento desta e de outras experiências similares (Teixeira, 2001).

Apesar disso, ambos contaram com o apoio do experimentador no interior da câmara insonorizada, para auxílio na leitura dos dados. Estes foram apresentados em versão ortográfica. A comunicação com o exterior processou-se através de um circuito áudio. Todo o processo foi monitorizado pelo pessoal técnico, através desse mesmo circuito.

Após a identificação dos sujeitos e gravação de uma sequência com o tracto oral em repouso, iniciou-se a aquisição dos dados, de acordo com a ordem prevista inicialmente: 1) gravação de todo o material a uma velocidade de elocução considerada normal; 2) produção de uma pequena parte do *corpus*, usando uma segunda taxa de elocução, mais rápida. Sempre que necessário, por iniciativa do próprio falante, ou por indicação do experimentador ou dos técnicos, o item foi repetido. A sequência [#VCV#] relativa às vibrantes foi repetida em item separado para todas as vogais, já que, em muitos casos, o tempo destinado à produção desta sequência para todas as consoantes do PE não foi suficiente.

O sinal de fala foi gravado em simultâneo com a trajectória dos articuladores, a uma frequên-

[30] Para além do *corpus* descrito em 5.4.2, no caso do informante masculino, foi ainda adquirido um pequeno *corpus*, tendo em vista o estudo de questões relacionadas com a coordenação temporal dos gestos implicados na produção de consoantes laterais.

(a) (b)

(c) (d)

Figura 5.8: Processo de recolha dos dados EMA da informante LF no GIPSA- LAB, Université Stendhal, Grenoble: a) ajuste do capacete; b) colagem dos sensores; c) posicionamento dos vários sensores; d) aquisição de pontos de referência para rotação dos dados no plano de oclusão (*bite plane*).

cia de amostragem de 22050 Hz.

Para os dois informantes, todos os sensores mantiveram a posição inicial até ao final da sessão, inclusive o do velo, tradicionalmente mais difícil de manter colado durante períodos longos de tempo.

O sistema de coordenadas EMA é determinado pela posição do capacete durante a aquisição dos dados, pelo que é necessária a aquisição de pontos de referência para rotação dos dados no plano de oclusão (*bite plane*). Assim, no final de cada sessão, foi inserida na boca de cada um dos sujeitos uma placa (*bite plate*) de plástico rígido [31] com dois sensores e gravada uma pequena sequência de poucos segundos.

Por último, foi adquirido o traçado do palato dos informantes.

[31] O *bite plate* assemelha-se a um cartão de crédito de tamanho variável: cerca de 4.5 cm para as mulheres e 5.5 cm para os homens (cf./www.humnet.ucla.edu/humnet/linguistics/faciliti/facilities/physiology/ema.html).

Pós-processamento

Parte do pós-processamento foi efectuada em Grenoble, durante as semanas subsequentes às sessões de recolha, pela equipa responsável pelo equipamento EMA, nomeadamente por Christophe Savariaux. O pós-processamento consistiu essencialmente na divisão, em ficheiros separados, das gravações de voz (usando os sinais sonoros que marcam o início e fim de cada período de aquisição EMA) e mudança de coordenadas dos dados dos sensores EMA para um sistema de coordenadas baseado nos pontos de referência.

A trajectória dos articuladores foi adquirida a 500 Hz e, posteriormente filtrada, de forma a eliminar o ruído [32].

Anotação dos dados

Os ficheiros áudio foram anotados foneticamente, utilizando-se para tal os símbolos do alfabeto fonético SAMPA [33] para o PE [34]. Esta tarefa foi efectuada com recurso ao programa Praat (Boersma & Weenink, 2009) e ficou a cargo de dois anotadores especialmente contratados para o efeito. Muitos dos critérios adoptados na transcrição fonética do material recolhido inspiraram-se nos procedimentos metodológicos seguidos na anotação do *corpus* francês.

Foram considerados dois níveis de anotação, criados automaticamente: "pho" e "seg".

No que respeita ao primeiro nível ("pho"), foram anotados os intervalos correspondentes ao fone a analisar, bem como o contexto fonético de ocorrência. De acordo com os critérios previamente estabelecidos, esta anotação foi obrigatória para todos os ficheiros.

O segundo nível ("seg"), a que corresponde uma anotação por pontos, baseia-se nas transições entre os segmentos (marcadas com um ponto) e parte estável dos mesmos, que, na maioria das vezes, corresponde à zona central. Neste último caso, a etiqueta faz referência ao *target* e sons circundantes (trifone). Este critério foi aplicado, por exemplo, na anotação dos ficheiros [#VCV#] (vd. figura 5.9).

A anotação automática da localização dos gestos dos vários articuladores, incluindo o do velo, foi efectuada, com base nos critérios explicitados em Oliveira & Teixeira (2006).

As velocidades de três dos articuladores (ponta da língua, velo e lábios) foram automaticamente calculadas em Matlab, usando funções especialmente definidas para o efeito. No caso do velo, este processo estava já praticamente implementado (Teixeira, 2001), o mesmo não se verificando para os articuladores orais. Para a ponta da língua, foi tida em conta a velocidade tangencial (Chitoran

[32] Mais concretamente foi efectuado um filtro passa-baixo que elimina todas as frequências abaixo dos 20 Hz.
[33] Cf. http://www.phon.ucl.ac.uk/home/sampa/index.html.
[34] Para uma correspondência entre os símbolos AFI e SAMPA, para o Português Europeu, consultar Jesus *et alii* (2007).

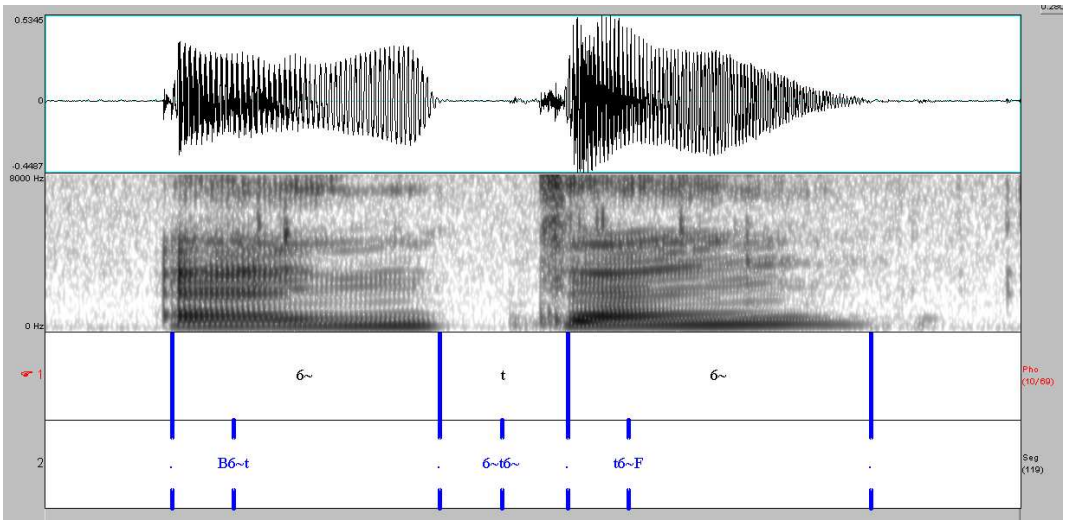


Figura 5.9: Exemplo de anotação fonética dos dados em Praat. Sinal acústico e espectrograma da sequência [ɐ̃tɐ̃], anotada (usando SAMPA) em dois níveis distintos: "pho" (em cima) e "seg" (em baixo).

et alii, 2002) [35] do sensor aí colocado. No caso do velo, foi usado o eixo y do sensor em causa. Finalmente, para os lábios foi calculada a distância vertical (designada de *lip aperture*, em inglês) entre o lábio inferior e superior e respectiva velocidade.

Em Matlab, foram também criados algoritmos para demarcar os movimentos articulatórios. Quatro eventos foram identificados e assinalados: início do movimento (S), *target* (T), *release* (R), *target2* (T2) (cf. Gafos, 2002). O método usado para determinar o início e fim dos gestos baseia-se na passagem da curva da velocidade por um determinado limiar à volta de zero (*noise band*). Este limiar é definido como uma percentagem da velocidade máxima de cada sensor. Para a ponta da língua, foi utilizada a percentagem de 15% da velocidade tangencial (Chitoran *et alii*, 2002), enquanto para o velo e lábios foi aplicado um critério de 20% da velocidade máxima vertical (Hoole, 1996).

Os dados foram posteriormente processados, no sentido de eliminar algumas marcas correspondentes a inflexões ou movimentos complexos. Sequências contíguas de S-T foram excluídas, de modo a manter apenas as etiquetas que correspondem a movimentos significativos dos articuladores. Se esta etapa de pós-processamento funciona de forma aceitável para o velo, o mesmo não se verifica para os articuladores orais, pelo que, no caso dos lábios e da ponta da língua, se optou por usar a versão sem correcções. É precisamente em relação a este articulador, a ponta da língua, que se verificam as maiores imprecisões na anotação automática. Este facto pode ficar a dever-se, pelo menos em parte, à localização pouco precisa do sensor [36].

Apesar das limitações acima indicadas, o recurso a um método de anotação automático acabou por se revelar muito útil para a investigação, na medida em que permitiu reduzir drasticamente

[35] A velocidade tangencial foi calculada de acordo com a fórmula proposta por Chitoran *et alii* (2002): $\sqrt{\dot{x}^2 + \dot{y}^2}$, sendo $\dot{x}$ a velocidade do sensor da ponta da língua no eixo horizontal e $\dot{y}$ a velocidade do mesmo sensor na direcção vertical.

[36] Como fizemos notar anteriormente, o sensor da ponta da língua é, normalmente, colado um pouco mais atrás, já na zona da lâmina, a cerca de 1 cm do alvo.

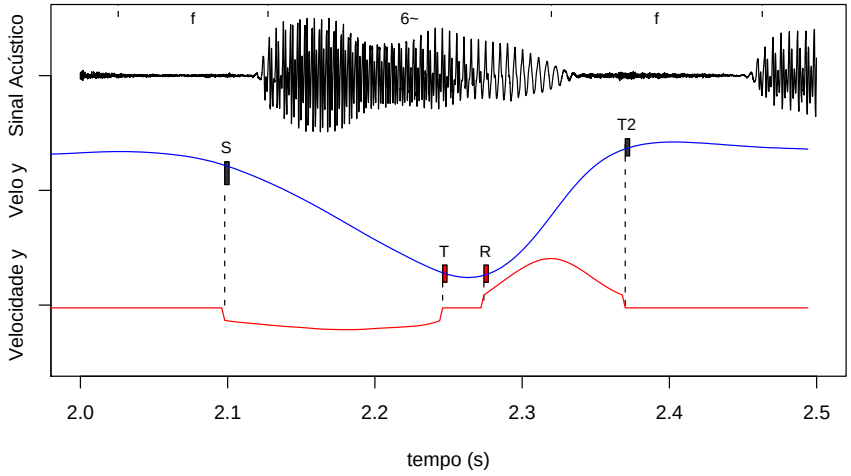


Figura 5.10: Exemplo de anotação automática do movimento do velo: início do movimento (S), *target* (T), *release* (R) e *target2* (T2). Sinal acústico (em cima), movimento do sensor colado no velo (ao meio) e curva de velocidade (em baixo), durante a produção da sequência [fɐ̃f], pela informante LF.

o tempo gasto na tarefa de anotação, garantindo, simultaneamente, um grau de consistência e precisão impossível de alcançar manualmente.

Análises estatísticas

Todas as análises estatísticas apresentadas neste trabalho foram executadas com o programa SPSS (v.16, SPSS Inc, Chicago, IL) e programa R (v. 2.71).

Na escolha, realização e apresentação dos testes estatísticos baseámo-nos, sobretudo, em Maroco (2007).

Com o intuito de avaliar se diversos factores (e.g. taxa de elocução, posição, vogal) afectavam significativamente as diversas variáveis de interesse (e.g. amplitude, duração, TTL), recorreu-se à *Analysis of Variance* (ANOVA), seguida de testes *post-hoc* de Bonferroni [37]. O pressuposto da distribuição normal da variável dependente nos diferentes grupos definidos pelo cruzamento dos factores foi avaliado pelo teste de Shapiro-Wilk, com correcção de Lilliefeors. Para alguns casos, obtiveram-se pequenos desvios à normalidade, mas, uma vez que a ANOVA é robusta a violações suaves deste pressuposto, optámos por não proceder a transformações matemáticas correctivas ou à realização de testes não-paramétricos (cf. Maroco, 2007). Contudo, em casos que ofereceram mais dúvidas, os resultados foram confirmados com testes não-paramétricos.

[37] Segundo Maroco (2007, p.161), em se tratando de amostras pequenas, o teste de Bonferroni é mais adequado do que outros testes *post-hoc* de comparações múltiplas de médias, nomeadamente o teste de Tukey.

O pressuposto de homogeneidade de variância foi validado com o teste de Levene. Em casos em que essa homogeneidade não estava garantida, foram utilizadas as variantes que entram em conta com essas diferenças, nomeadamente testes *post-hoc* de Tamhane, em vez de Bonferroni.

Consideraram-se estatisticamente significativos os efeitos cujo *p-value* foi inferior ou igual a 0.05.

5.4.3 Altura do velo

Nesta seccção, procurámos reunir alguns dados preliminares acerca da altura velo durante a produção dos vários sons do PE, particularmente dos sons nasais. O problema é abordado nas questões 1 e 2 do estudo, que prevêm, respectivamente, diferenças de amplitude entre as vogais nasais e as consoantes nasais e uma possível correlação entre a altura da vogal e o grau de abertura velar.

5.4.3.1 Metodologia

Tendo em conta os objectivos a atingir, o material linguístico usado nesta fase da experiência restringiu-se à secção 2 do *corpus*, que inclui todas as vogais nasais ([ɐ̃ ẽ ĩ õ ũ]) e todas as consoantes do PE ([p t k b d ɡ v f s ʃ z ʒ l ʎ ɾ ɾ m n ɲ]), para além das três vogais cardinais ([i a u], organizadas em sequências [#VCV#]. Estas últimas foram produzidas uma única vez, a uma taxa de elocução considerada normal, pelos dois sujeitos, embora os resultados aqui relatados se refiram, como já explicámos, a um único informante (LF).

Na articulografia electromagnética, as trajectórias dos articuladores são indicadas no plano sagital, através de duas coordenadas (XY). Os valores de amplitude foram extraídos - seguindo a metodologia usada pela equipa de investigação francesa - a partir da posição vertical do sensor do velo (coordenada y), com base nas etiquetas fonéticas do nível "seg", mais exactamente no ponto que assinala a zona estável de cada um dos segmentos. No caso concreto das vogais nasais, foi, contudo, necessário introduzir uma pequena alteração ao método, já que se constatou - através de uma análise preliminar dos dados - que, nas vogais nasais do português, contrariamente às vogais nasais francesas, o ponto de máxima amplitude acontece perto do final da vogal (cf. Lovatto *et alii*, 2007) e não no ponto médio acústico. Assim, para as vogais nasais, a altura do velo foi obtida no ponto mais baixo do sensor do velo (eixo do y). O risco de confundir a posição do velo para a produção da vogal nasal final com a posição do velo durante a respiração [38] foi despistado através de uma ANOVA que indicou que os valores obtidos em posição inicial e final não são estatisticamente diferentes. Para além disso, os valores correspondentes à última vogal nasal da série foram eliminados para efeitos de análise.

[38] É durante a respiração que o velo atinge a abertura máxima, superior à registada para os fonemas nasais (Amelot, 2004).

5.4.3.2 Resultados

O gráfico 5.11 sintetiza os resultados obtidos para a altura do velo (nos gráficos Velo y) nas consoantes orais (C), vogais orais (V), consoantes nasais (N) e vogais nasais (Vn).

Comparando a altura do velo para as várias classes em análise, observa-se a seguinte progressão C (média= -1.96)< V (média= -2.03)< N (média= -2.23)< Vn (média= -2.34). Verifica-se, sem grandes surpresas, que o velo se encontra numa posição mais alta para os sons orais do que para os sons nasais. Para além disso, e considerando apenas os fonemas nasais, é durante a produção das vogais nasais que o velo atinge uma maior amplitude.

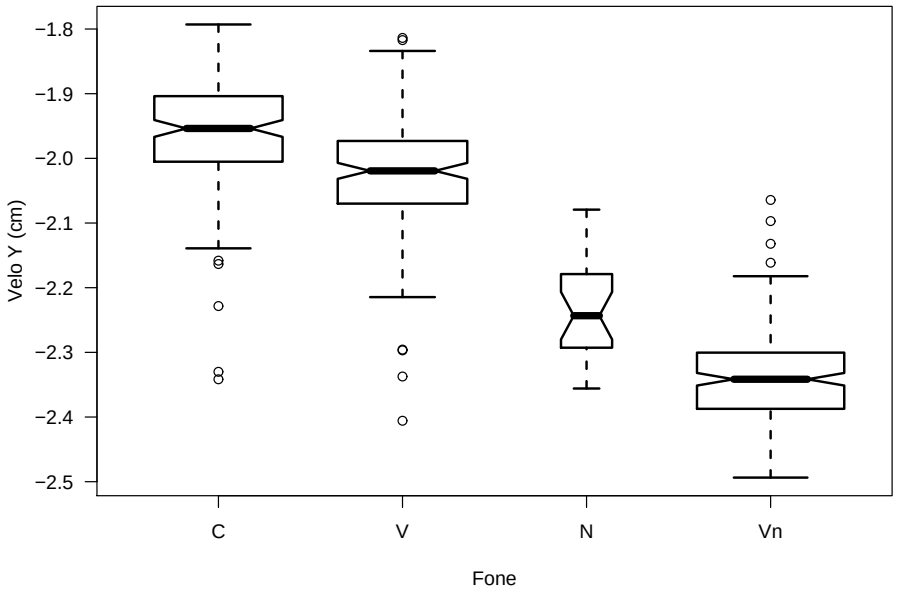


Figura 5.11: Diagrama de extremos e quartis relativo à altura do velo (Velo y) para as diferentes classes de sons do PE: consoantes orais (C), vogais orais (V), consoantes nasais (N) e vogais nasais (Vn).

O resultado dos testes estatísticos ANOVA [F(3,524)=718.5, p<0.001], seguidos de um teste *post-hoc* de Bonferroni ($p < 0.05$), revela que a diferença entre as várias classes de sons, no que respeita à altura do velo, é estatisticamente significativa. Fica assim patente uma menor amplitude no caso das duas classes orais, que, para além disso, são também diferentes entre si. Em segundo lugar, e mais importante do ponto de vista dos objectivos a alcançar com este estudo, as consoantes nasais e as vogais nasais distinguem-se também entre si, quanto à altura do velo.

Depois de uma análise global dos dados, importa agora avaliar a posição assumida pelo velo nas diferentes consoantes, aqui divididas em sete classes distintas: fricativas surdas (Fric sur), fricativas sonoras (Fric son), oclusivas surdas (Ocl sur), oclusivas sonoras (Ocl son), vibrantes (Vibr), laterais (Lat) e nasais (Nas).

Segundo o diagrama de extremos e quartis (do inglês *box plot*) da figura 5.12, o velo atinge a posição mais baixa durante a realização das consoantes nasais. Contudo, fica igualmente patente alguma sobreposição entre consoantes orais - nomeadamente laterais, vibrantes e oclusivas sonoras - e consoantes nasais. Isto significa que, pelo menos em alguns contextos, alguns segmentos classificados como orais poderão ser produzidos com o velo parcialmente aberto, chegando este último a atingir uma altura similar à registada para as consoantes nasais.

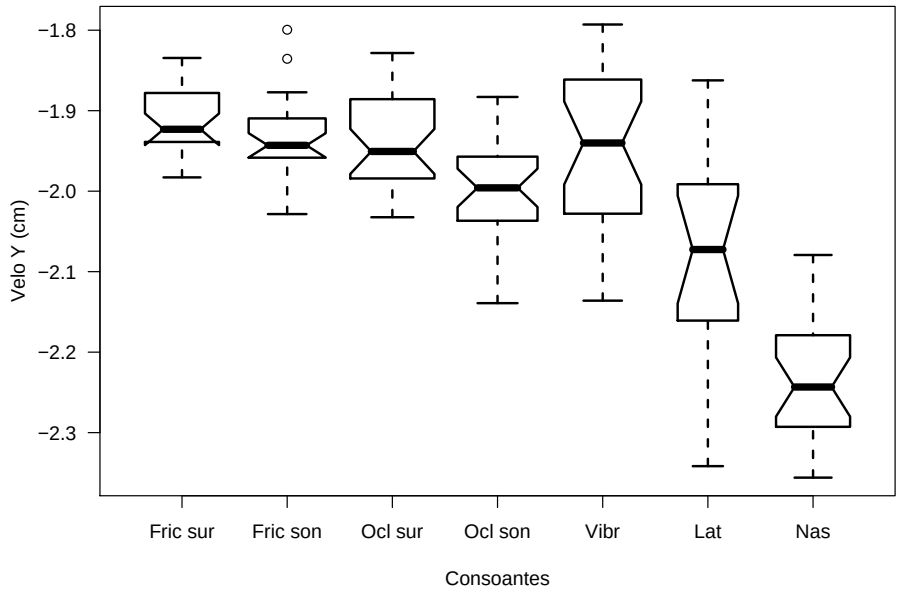


Figura 5.12: Diagrama de extremos e quartis relativo à altura do velo (Velo y) para as diferentes classes de consoantes do PE: fricativas surdas (Fric sur), fricativas sonoras (Fric son), oclusivas surdas (Ocl Sur), oclusivas sonoras (Ocl Son), vibrantes (Vibr), laterais (Lat) e nasais (Nas).

No sentido de comprovar a significância estatística das diferenças entre as várias consoantes, no que respeita à altura do velo, recorreu-se, mais uma vez, à ANOVA, seguida de testes *post-hoc* de Tamhane. A primeira dá conta de uma diferença significativa entre os vários tipos de consoantes [$F(6,168)=55.476$, $p<0.001$]. Já os segundos permitem especificar a natureza dessas diferenças e mostram claramente que: 1) as consoantes nasais são significativamente diferentes das demais; 2) as oclusivas sonoras distinguem-se também de todas as outras consoantes, à excepção das vibrantes e laterais; 3) entre as restantes classes de consoantes não há diferenças significativas.

Seguidamente, fizemos incidir a nossa análise sobre a relação entre a altura do velo e a vogal. A partir da observação do gráfico 5.13, relativo à altura do velo nas três vogais orais consideradas para este estudo, é possível verificar que a vogal [u] é realizada com o velo numa posição consideravelmente mais baixa do que o [i] e o [a].

A referida diferença entre as vogais, veio a revelar-se significativa [$F(2,150)=13.214$, $p<0.001$],

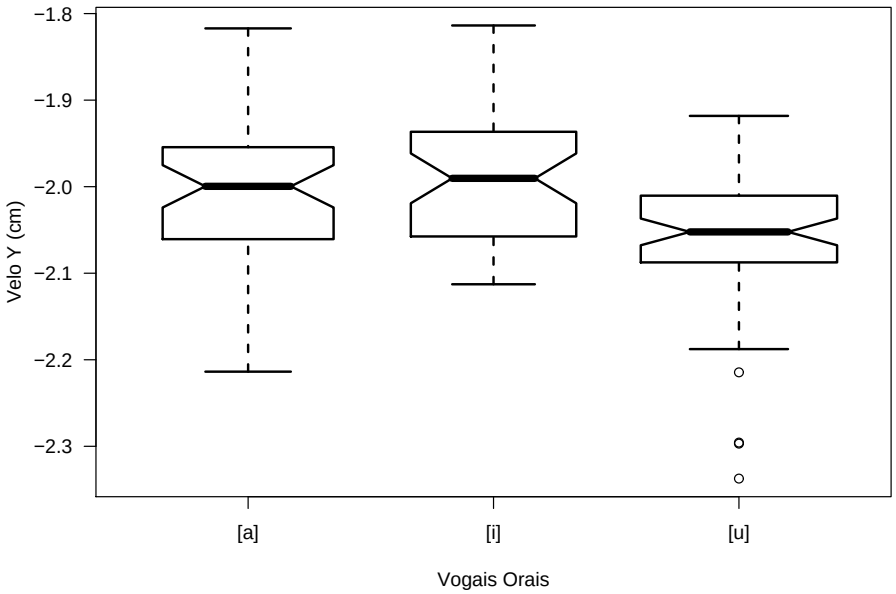


Figura 5.13: Diagrama de extremos e quartis relativo à altura do velo (Velo y) nas vogais orais: [a], [i], [u].

após aplicação dos testes estatísticos ANOVA.

Os testes *post-hoc* de Bonferroni vieram confirmar como estatisticamente significativa a diferença entre a vogal [u] e as outras duas vogais em análise. O mesmo teste não detectou diferenças significativas entre o [a] e o [i].

Sublinhando que, para o esclarecimento da questão 2, importa sobretudo reunir dados relativos ao comportamento do velo durante a produção das vogais nasais, dedicamos alguma atenção, nesta fase do estudo, à relação entre a altura da vogal nasal e a altura do velo.

Os valores obtidos para as cinco vogais nasais [39], encontram-se representados no gráfico 5.14.

A primeira informação a reter do gráfico é a grande sobreposição e dispersão de valores da altura do velo, para praticamente todas as vogais nasais.

Submetendo estes resultados ao teste estatístico, verifica-se que a diferença de altura do velo nas várias vogais nasais não é estatisticamente significativa [F(4,195)=1.12, p=0.35].

[39]Cumpre notar, neste ponto, que a realização da vogal nasal [ẽ], pela informante LF, oscilou entre a vogal plena e o ditongo [ẽj̃]. Esta situação verificou-se, sobretudo, em posição final.

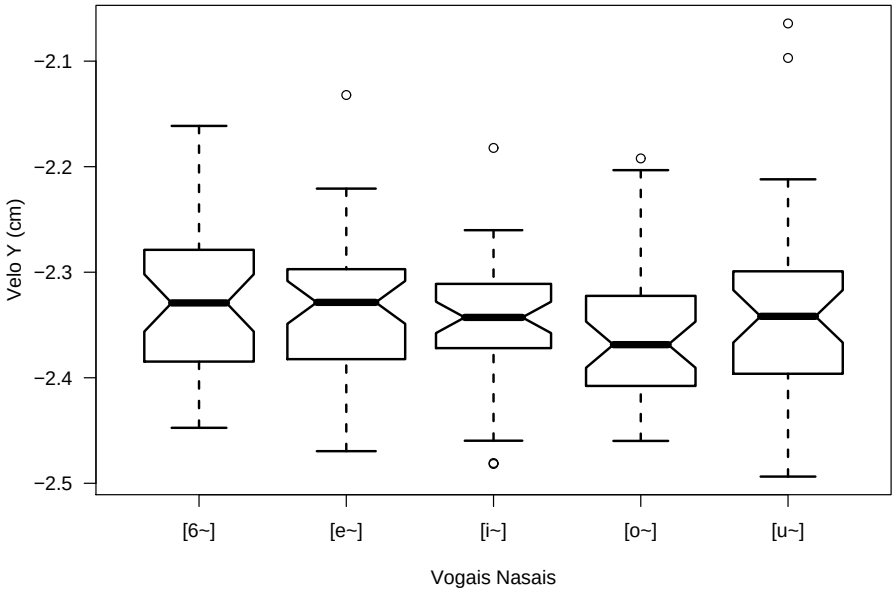


Figura 5.14: Diagrama de extremos e quartis relativo à altura do velo (Velo y) nas vogais nasais do PE: [ɐ̃], [ɛ̃], [ĩ], [õ], [ũ].

5.4.3.3 Comentários aos resultados

Em face dos resultados obtidos, conclui-se que, do ponto de vista articulatório, existe uma diferença entre consoantes nasais e vogais nasais, no que respeita à altura intrínseca do velo. Os dados experimentais recolhidos confirmam, assim, as previsões da questão 1 e apontam para a necessidade de considerar dois graus distintos de abertura velar, uma para as consoantes nasais e outro para as vogais nasais. Amelot & Rossato (2006) e Rossato *et alii* (2003, 2006) observam diferenças similares para o francês.

Relativamente às consoantes, verificou-se que as nasais são produzidas com o véu palatino na posição mais baixa. Em alguns contextos, as laterais e vibrantes podem ser realizadas com o velo numa posição similar às consoantes nasais, à semelhança do que foi observado por Amelot & Rossato (2006), para o francês. O mesmo acontece com as oclusivas vozeadas em contexto de vogais nasais, que, por oposição às oclusivas surdas, parecem tolerar um certo grau de abertura vélica. Este facto, já atestado noutros estudos (Basset *et alii*, 2001; Ohala & Ohala, 1991), poderá encontrar explicação em factores aerodinâmicos (Solé, 2007): "nasal leakage during the oral constriction for the stop contributes to keeping a low oropharyngeal pressure which favors transgottal flow for voicing".

Contrariamente à tendência geral - e até a estudos realizados anteriormente para o PE (Rossato *et alii*, 2006) - que dá conta de uma posição vélica mais baixa para as vogais orais mais baixas (Ohala, 1975; Clumeck, 1976), os resultados obtidos mostraram que o velo está mais baixo na vogal

[u] do que nas vogais [i a] [40].

Também em relação às vogais nasais, os dados coligidos não permitem suportar uma correlação entre a altura da vogal nasal e a amplitude do velo. Com efeito, as análises estatísticas indicam que as diferenças na altura do velo para as várias vogais nasais não são estatisticamente significativas. Isto significa que não haverá necessidade de assumir diferentes *targets* para o gesto velar das várias vogais nasais.

Finalmente, cumpre notar que o método usado para medição da altura do velo tem uma influência determinante nos resultados. Seguindo a metodologia adoptada pela equipa de investigação francesa, não foi possível detectar diferenças na altura do velo entre as vogais e as consoantes nasais do PE. Uma observação mais cuidada dos dados permitiu, contudo, concluir que este método não era adequado para lidar com as vogais nasais do PE, já que, nestes segmentos, o ponto de máxima amplitude não ocorre na zona central, como parece acontecer no francês, mas mais perto do final da vogal nasal. Neste sentido, optou-se por seguir uma abordagem mais articulatória, medindo a amplitude das vogais nasais no ponto em que o sensor do velo atinge o valor mais baixo. Esta metodologia permitiu identificar diferenças significativas na altura do velo entre vogais e consoantes nasais.

5.4.4 Duração dos gestos do velo

Tendo já em mente o estudo da coordenação gestual, nesta fase da experiência, foi nosso objectivo específico obter dados acerca da duração do movimento do velo, nomeadamente do gesto de abertura, da fase estável (*plateau*) e do gesto de fecho. Esta problemática é contemplada na questão 3 do estudo, que antevê, igualmente, uma eventual influência de variáveis contextuais e da taxa de elocução sobre a duração do movimento do velo.

5.4.4.1 Metodologia

Quanto ao material linguístico utilizado, recorreu-se à secção 3 do *corpus*, descrita sumariamente no ponto 5.4.2. Recordamos que estão em causa sequências simétricas, em que a vogal nasal surge em posição inicial, medial, ou final, flanqueada por uma consoante oclusiva ([p b d t]) ou fricativa ([f]). Cada uma das sequências foi produzida duas vezes, pelos informantes AT e LF, usando duas taxas de elocução distintas (normal e rápida). Mais uma vez lembramos, que os resultados apresentados na secção subsequente se reportam apenas à informante feminina.

Os valores de duração foram obtidos automaticamente, com base na anotação da trajectória do sensor do velo (5.4.2). Depois de identificadas e assinaladas as *landmarks* (Gafos, 2002) - início do movimento (S), *target* do movimento (T), *release* do movimento (R) e fim do movimento (T2) - os valores foram extraídos automaticamente e exportados para o SPSS. Antes de se proceder ao cálculo

[40] As análises preliminares dos dados do segundo informante não corroboram, contudo, este resultado, confirmando a tendência verificada anteriormente, noutros estudos para o PE (Rossato *et alii*, 2006).

da duração das várias fases, os dados foram cuidadosamente verificados, mediante inspecção visual das pautas gestuais desenhadas automaticamente.

Todos os cálculos conducentes à obtenção da duração dos movimentos foram realizados directamente no SPSS. A variável *duração da abertura* resulta da diferença entre o T e o S; a *duração da zona estável* corresponde ao intervalo entre T e R; e a *duração do fecho* foi definida como a subtracção entre T2 e R.

5.4.4.2 Resultados

Os resultados globais relativos à duração do movimento do velo são apresentados na tabela 5.4. Esta inclui os valores médios obtidos para a duração do ciclo completo de abertura e fecho (*duração total*) e de cada uma das fases do movimento (*abertura, zona estável, fecho*), em função da taxa de elocução (*normal* ou *rápida*) e da posição lexical (*inicial, medial, final*) da vogal nasal.

Tabela 5.4: Valores médios de duração (em ms) dos vários movimentos do velo (duração total, duração da abertura, duração da zona estável e duração do fecho), em função da taxa de elocução (normal ou rápida) e da posição lexical da vogal nasal (inicial, medial ou final).

		Duração			
Taxa	Posição	Total	Abertura	Z. Estável	Fecho
Normal	INICIAL	340	133	42	122
	MEDIAL	300	143	24	112
	FINAL		141		
Rápida	INICIAL	200	74	15	97
	MEDIAL	220	106	15	79
	FINAL		165		

Como se pode verificar através da análise da tabela, quando passamos de uma taxa de elocução normal para uma taxa de elocução mais rápida, a duração total do movimento do velo diminui consideravelmente, independentemente da posição lexical ocupada pela vogal nasal. Enquanto no primeiro caso os valores médios rondam os 300 ms, para a taxa rápida a duração média apurada fica-se pelos 200 ms.

O mesmo acontece em relação à duração da abertura, da zona estável e do fecho do velo: a duração das diferentes fases é menor em taxa de elocução rápida do que em taxa normal. A única excepção prende-se com a duração do ciclo de abertura em posição final, em que a tendência é inversa, isto é, a duração, em vez de diminuir, aumenta, quando a taxa de elocução sobe. Esta diferença poderá estar relacionada com o próprio método de detecção e anotação dos movimentos, mais sujeito a imprecisões em situação de fala rápida. Com efeito, verificou-se que, no caso específico da posição final, o processo automático nem sempre foi capaz de identificar e assinalar o T do velo, confundindo-o com *target* da respiração, o que se traduz num aumento aparente da duração do ciclo

de abertura. Do mesmo modo, em posição inicial, detectaram-se alguns problemas na marcação do início do movimento do véu palatino.

Comparando a fase de abertura com a fase de fecho, observa-se que a primeira é, de um modo geral, mais longa do que a segunda, não importa qual a taxa de elocução utilizada. Esta tendência é apenas contrariada no caso da posição inicial (taxa rápida), em que, de acordo com os valores apurados, o ciclo de fecho tende a ser maior do que o de abertura (97 ms e 74 ms, respectivamente). Uma vez mais, admitimos a hipótese desta singularidade poder estar relacionada com problemas na detecção automática dos gestos articulatórios, já que um dos casos identificados como particularmente difíceis e, consequentemente, mais sujeito a imprecisões foi - conforme mencionado anteriormente - precisamente a posição inicial.

O gráfico que se segue (5.15) ilustra os resultados da tabela anterior (5.4), relacionando-os com a vogal nasal.

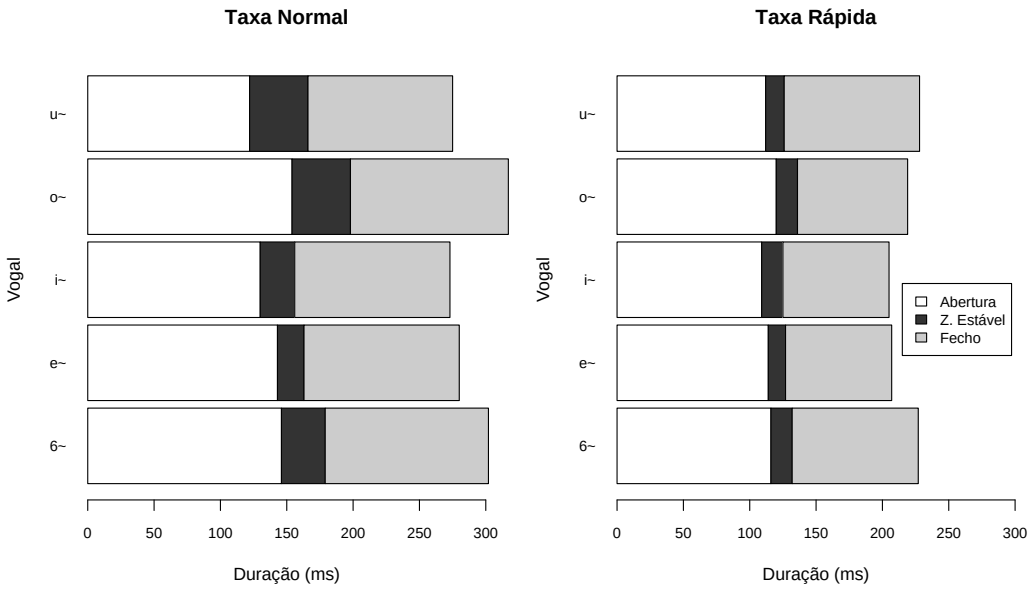


Figura 5.15: Valores médios da duração (em ms) da fase de abertura, parte estável e fecho do velo, para as cinco vogais nasais do PE ([ɐ̃], [ẽ], [ĩ], [õ], [ũ]), nas duas taxas de elocução consideradas (normal e rápida).

Da observação do gráfico, resulta, mais uma vez, claro o contraste de duração média - quer total, quer das várias fases - entre a taxa normal e a taxa rápida, embora essa diferença não se traduza de igual modo para todas as vogais.

Com o intuito de aprofundar estes dados e avaliar o grau de significância das diferenças acima mencionadas, foram realizados alguns testes estatísticos, nomeadamente uma ANOVA com três factores (taxa de elocução, posição e vogal), seguida de testes *post-hoc*, considerando quer a duração total, quer a duração dos vários movimentos do velo (abertura, zona estável, fecho).

No tocante à duração total, a ANOVA confirmou que a taxa de elocução teve um efeito estatisticamente significativo sobre a duração total do movimento do velo [$F(1,77)=152.88$, $p<0.001$]. Quanto ao tipo de vogal, embora a ANOVA tenha determinado como estisticamente significativas as diferenças de duração entre as várias vogais nasais [$F(4,77)=2.65$, $p=0.039$], os testes *post-hoc* de Tamhane não se revelaram conclusivos, não sendo, por isso, possível, averiguar a natureza dessas diferenças. Este tipo de situação em que a ANOVA e os testes de comparação múltipla chegam a conclusões diferentes, se bem que pouco provável, é possível. Como solução para este problema, recomenda-se, geralmente, a repetição do estudo com uma amostra de maior dimensão (cf. Maroco, 2007). Já a posição lexical não influenciou significativamente a duração total do ciclo de abertura e fecho do velo ($p>0.05$).

Em relação à fase de abertura, limitamo-nos aqui a referir os resultados gerais da ANOVA, que demonstram como estatisticamente significativos os três factores (taxa de elocução, vogal, posição). Os problemas relacionados com anotação dos movimentos do velo, na taxa de elocução rápida (posição inicial e final), já mencionados em parágrafos anteriores, impediram-nos de desenvolver análises estatísticas mais detalhadas, que a terem lugar poderiam induzir conclusões erradas.

A duração da zona estável é, segundo os resultados da ANOVA, significativamente afectada pela taxa de elocução [$F(1,77)=20.89$, $p<0.001$] e pela posição lexical da vogal nasal [$F(1,77)=4.93$, $p=0.029$]. As diferenças de duração decorrentes da posição ficam a dever-se não tanto à taxa rápida, mas à taxa normal, o que se traduz, em termos de resultados da ANOVA, numa interacção significativa entre taxa de elocução e posição [$F(1,77)=4.92$, $p=0.029$]. Retomando os dados da tabela 5.4, na taxa normal, a duração média da fase estável é maior em posição inicial do que em posição medial. O mesmo não acontece na taxa rápida, onde o valor de duração se mantém inalterado.

De modo semelhante, também a duração do fecho é significativamente influenciada pelos factores taxa de elocução [$F(1,78)=40.23$, $p<0.001$] e posição [$F(1,78)=9.40$, $p=0.003$].

5.4.4.3 Comentários aos resultados

De acordo com os resultados apurados, a duração média de um ciclo completo do velo (abertura e fecho) ronda os 300 ms, um valor perfeitamente compatível com o intervalo de tempo estimado por Stevens (1998, p.43), para o inglês, ou por Amelot (2004) e Basset *et alii* (2006), para o francês. Para além disso, os dados vêm confirmar os resultados de um estudo piloto anterior (Oliveira & Teixeira, 2007b), também baseado num *corpus* EMA, com características similares.

Neste mesmo trabalho (Oliveira & Teixeira, 2007b), foram ainda apresentados dados relativos à duração média da abertura, do *plateau* e do fecho do velo: 152 (std = 31) ms, 41 (std = 26) ms e 114 (std = 33) ms, respectivamente.

Na análise agora realizada, encontrámos também confirmação destes dados, com durações médias a rondar os 130-140 ms para a abertura, os 30-40 ms para a fase estável e os 110-120 ms para

o fecho (taxa normal). A partir destes resultados, é ainda possível inferir que o gesto de abertura tende a ser mais longo do que o gesto de fecho, um aspecto que sobressai de outros dados publicados na literatura, como, por exemplo, os apresentados por Solé (1995) num estudo comparativo entre o espanhol e o inglês americano. Nesta última língua, a duração do movimento parece estar intimamente relacionada com a velocidade de abertura e fecho do velo. Com efeito, no caso específico dos informantes americanos, por oposição aos falantes espanhóis, as diferenças de duração entre a fase de abertura e fecho fazem-se acompanhar de diferenças significativas de velocidade. Independentemente da taxa de elocução, a velocidade de abertura é sempre inferior à de fecho, o que se traduz numa abertura mais longa e num fecho mais curto.

Já no francês, conforme sugerem os resultados de Amelot (2004), obtidos através de fibroscopia, o movimento de abertura e de fecho do velo tem sensivelmente a mesma duração (logátomos, frases e *corpus* espontâneo) ou o primeiro é mais curto do que o segundo (vogais nasais isoladas).

Adicionalmente, no material recolhido e analisado, foi possível verificar que todas as medidas de duração realizadas (duração total, duração da abertura, duração da zona estável e duração do fecho) são significativamente afectadas pela taxa de elocução. No que concerne à duração total, observou-se que o gesto articulatório sofre um encurtamento de cerca de 100 ms, quando produzido a uma taxa de elocução mais rápida. Também em relação à abertura, ao *plateau* e ao fecho, ficou patente uma diminuição significativa da duração média, em consequência do aumento da taxa de elocução. Este efeito da taxa de elocução sobre a duração dos gestos articulatórios não é, de modo algum, uma surpresa e está em linha com a literatura sobre a matéria (para uma revisão, cf. Krakow, 1993). A respeito da interferência da taxa de elocução no movimento do velo, convém, no entanto, frisar que a redução da extensão do gesto não é a única estratégia usada pelos falantes, no sentido de diminuir o tempo de produção do enunciado (Krakow, 1993). Segundo os estudos que tratam desta interferência, é possível optar, por exemplo, por um aumento da velocidade do movimento ou, simplesmente, combinar as duas estratégias (Krakow, 1993). Segundo os resultados de Solé (1995), já referidos, nos três informantes americanos, uma taxa de elocução mais rápida está sempre associada a uma redução significativa da fase de abertura do velo, sendo que a altura do articulador se mantém constante. Para que tal aconteça, há um aumento significativo da velocidade do movimento para todos os informantes.

Finalmente, os nossos dados parecem também sugerir uma influência da posição lexical na duração média da fase de abertura, *plateau* e fecho. Contudo, os problemas detectados na anotação automática dos gestos não nos permitem formular observações conclusivas a respeito desta matéria, tornando-se pois necessário prosseguir a investigação, nomeadamente aumentando o número de dados e, eventualmente, de informantes, garantindo, desta forma, uma maior sustentabilidade à análise estatística.

5.4.5 *Stiffness* do velo

Nesta parte do estudo, de acordo com o definido na questão 3, pretendemos reunir alguns dados sobre o *stiffness* do velo, um parâmetro que procura caracterizar a velocidade do movimento articulatório. Como tivemos já ocasião de frisar, este parâmetro foi formalmente incorporado ao modelo da dinâmica da tarefa (*task-dynamics*), usado pela Fonologia Articulatória, e tem efeitos importantes ao nível da duração do movimento dos articuladores. Foi nosso objectivo calcular os valores de *stiffness*, associados ao gesto de abertura e ao gesto de fecho do velo, para além de determinar possíveis influências de outros factores (e.g. taxa de elocução) sobre esses mesmos valores.

5.4.5.1 Metodologia

Como base para esta experiência, foi usado o mesmo material linguístico do estudo precedente, relativo à duração dos gestos, i.e. a secção 3 do *corpus*, cuja apresentação mais detalhada teve lugar no ponto 5.4.2. Recordamos que para a produção das sequências foram utilizadas duas taxas de elocução diferentes, uma normal e outra mais rápida.

Na linha do proposto por Roon *et alii* (2007), para obter o *stiffness*, a velocidade máxima foi dividida pelo deslocamento máximo (vd. 5.1):

$$\textit{Stiffness} \equiv \frac{\text{Velocidade máxima } (cm/seg)}{\text{Deslocamento máximo } (cm)} \tag{5.1}$$

5.4.5.2 Resultados

Tabela 5.5: Valores médios do *stiffness* para a abertura e fecho do velo, em função da posição lexical da vogal nasal (inicial, medial e final) e da taxa de elocução (normal e rápida).

		Stiffness	
Taxa	Posição	Abertura	Fecho
Normal	INICIAL	10.2	12.5
	MEDIAL	9.6	13.1
	FINAL	9.7	—
Rápida	INICIAL	14.6	15.1
	MEDIAL	12.4	16.8
	FINAL	9.4	—

Os resultados da análise do *stiffness* para a abertura e fecho do velo, por taxa de elocução (normal e rápida), constam da tabela 5.5.

Segundo os dados aí apresentados, o *stiffness* do fecho é sempre mais elevado do que o *stiffness* de abertura, não importa qual a posição lexical ou a taxa de elocução considerada.

Para além disso, importa ainda sublinhar o efeito da taxa de elocução: à medida que esta última se torna mais rápida, os valores do *stiffness* tendem também a aumentar, com excepção da posição final.

Tendo em conta os problemas detectados na anotação do movimento do velo, nesta posição e também em posição inicial, quando se trata de taxa rápida, optámos por utilizar, na análise inferencial, apenas os dados da posição medial.

Efectuou-se uma análise ANOVA de três factores para avaliar o efeito da taxa de elocução, da vogal e da posição sobre o *stiffness* de abertura. Segundo os resultados, é possível afirmar que a taxa de elocução [$F(1,77)=23.03$, $p<0.001$] e a vogal [$F(4,77)=2.58$, $p=0.044$] tiveram um efeito significativo no *stiffness* de abertura do velo. Apesar disso, os testes *post-hoc* de Tamhane não se revelaram capazes de detectar diferenças entre as várias vogais, o que sugere a necessidade de repetir os testes com uma amostra de maior dimensão.

Para o *stiffness* do fecho, a ANOVA de 3 factores (taxa de elocução, posição e vogal) comprovou como estatisticamente significativo apenas o efeito da taxa de elocução [$F(1,59)=56.51$, $p<0.001$].

Finalmente, a relação e possível diferença entre os valores do *stiffness* de abertura e de fecho do velo foram investigados através de um teste t de medidas repetidas. Os resultados indicaram como significativamente menor o valor para o *stiffness* de abertura [$t(71)=-10.90$, $p<0.001$]. A diferença média situa-se entre -3.94 e -2.72, para um nível de confiança de 95 %. Os valores das duas medidas de *stiffness* revelaram-se significativamente correlacionados [$r=0.45$, $p<0.001$].

5.4.5.3 Comentários aos resultados

No estudo exploratório acima apresentado foram estimados os valores de *stiffness* do velo, enquanto "context-independent measurement of the velocity profile of articulatory movement" (Roon *et alii*, 2007).

Uma dos aspectos principais a reter da análise dos resultados está relacionado com as diferenças de *stiffness* entre a fase de abertura e a fase de fecho do velo: independentemente da taxa de elocução usada, esta última está associada a valores de *stiffness* mais elevados do que a primeira, o que implica que o velo seja mais lento a abrir do que a fechar. Em geral, estes dados estão de acordo, quer com as observações prévias acerca da velocidade de abertura e fecho do velo (Benguerel *et alii*, 1977; Horiguchi & Bell-Berti, 1987; Solé, 1995), quer com os resultados obtidos para os movimentos da mandíbula (Kelso *et alii*, 1985) e do dorso da língua (Kelso *et alii*, 1985). No sentido de explicar as diferenças de velocidade de abertura e de fecho do velo, Solé (1995) admite a possibilidade de os dois gestos estarem associados a diferentes graus de *stiffness*, explicação esta que encontra eco nos nossos dados.

Quanto ao papel das várias variáveis analisadas (taxa de elocução, posição e vogal), verificou-

se que a taxa de elocução tem um efeito significativo sobre o *stiffness* de abertura e fecho do velo. Este resultado coincide, ainda que parcialmente, com as observações experimentais de Solé (1995), para os falantes americanos, mas não para os sujeitos espanhóis. Com efeito, enquanto no espanhol nenhum dos gestos do velo é afectado pelo aumento da taxa de elocução, no inglês americano, a velocidade do gesto de abertura - mas não o de fecho - ajusta-se às variações na taxa de elocução. Estas diferenças entre as duas línguas, levam a investigadora a concluir que, no espanhol, o gesto de abertura do velo é biomecanicamente controlado, ao passo que no americano, ele é fonologicamente especificado. No PE, tanto o *stiffness* da abertura como o de fecho são afectados pelo aumento da taxa de elocução.

Cumpre sublinhar que, tal como nas experiências de Roon *et alii* (2007), em nenhuma da análises efectuadas, a posição lexical se revelou um factor significativo. Este resultado é perfeitamente coincidente com a própria definição de *stiffness* como parâmetro que pretende medir o comportamento dos movimentos articulatórios em termos de velocidade, independentemente do contexto.

5.4.6 Coordenação intergestual

O problema - levantado na questão 4 - que nos ocupa nesta fase do estudo experimental está relacionado com a organização dos gestos nasais, orais e glotais, durante a produção das vogais nasais, em diferentes contextos. Num primeiro momento, procurámos analisar a relação do gesto de abertura (e fecho) do velo com o gesto oral seguinte, cuja sobreposição - a par da existência de vozeamento - poderá habilitar a emergência da chamada consoante nasal intrusiva. Paralelamente, foram consideradas outras questões como 1) a coordenação com C1, isto é a consoante que precede a vogal nasal, em contextos [CṼ.CV]; 2) a sincronização entre o gesto de fecho do velo e o gesto de abertura da glote, em contexto de consoante oral surda.

5.4.6.1 Metodologia

Para este estudo, recorremos ao mesmo material linguístico - secção 3 do *corpus* (vd. 5.4.2) - utilizado nas duas experiências anteriores, relativas à duração e *stiffness* do movimento do velo.

A coordenação entre dois gestos é, normalmente, quantificada como a distância temporal entre dois eventos articulatórios (*landmarks*), algoritmicamente determinados a partir da velocidade dos sensores em causa. Para o estudo da coordenação entre o gesto do velo e o gesto oral (dos lábios ou da ponta língua), o evento articulatório seleccionado foi o *target* (cf. Oliveira & Marin, 2005).

O intervalo temporal entre o *target* de abertura do velo e o *target* de fecho oral designa-se de TTL (do inglês, *Target-to-Target Lag*) (Oliveira & Marin, 2005) e encontra-se ilustrado em 5.16.

Considerando não o *target* de abertura do velo, mas o *target* de fecho (T2), obteve-se o T2TL. Este é definido como a distância temporal entre o *target* de fecho do velo e o *target* de fecho oral, conforme ilustrado na figura 5.16.

Beneficiando de experiências anteriores (Oliveira & Teixeira, 2007b), para a sincronização

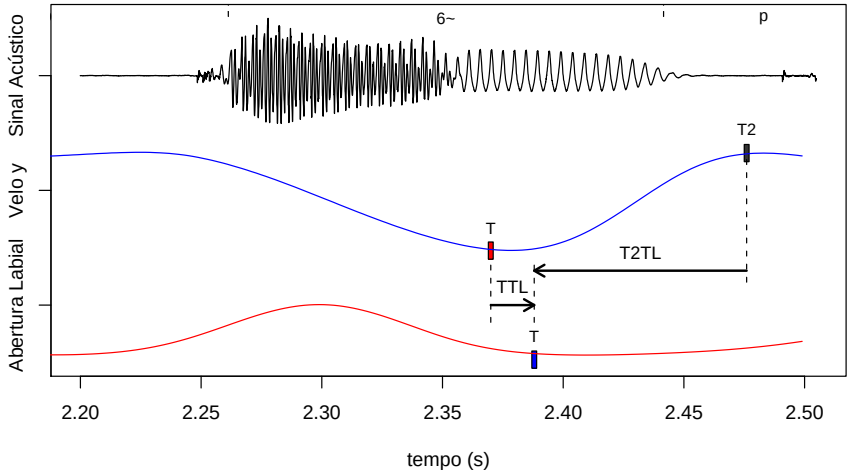


Figura 5.16: Medidas de sincronização entre o gesto do velo e o gesto oral: intervalo temporal entre *target* de abertura do velo e o *target* de fecho oral (TTL) e intervalo temporal entre o *target* de fecho do velo e o *target* de fecho oral (T2TL).

temporal do velo com C1, foi considerada a distância temporal entre o início do gesto de abertura do velo (S) e a *release* (R) da consoante anterior (*Start-to-Release Lag* (SRL)) [41].

Finalmente, investigámos a coordenação do gesto de fecho do velo com o gesto de abertura da glote, em contextos de consoante surda. Para tal, foi medida a distância temporal entre o final do movimento de fecho do velo e o gesto de abertura da glote. Este último gesto foi determinado com base no sinal acústico.

5.4.6.2 Resultados

Coordenação entre a abertura do velo e o fecho oral

Os resultados relativos à distância temporal entre o *target* de abertura do velo e o *target* oral (*Target-to-Target Lag* (TTL)) encontram-se na tabela 5.6.

Em primeiro lugar, importa referir que os valores médios de TTL são todos negativos, independentemente da taxa de elocução e da posição lexical, o que sugere uma coordenação sequencial, ou seja, o velo atinge o *target* antes do articulador oral fechar.

Observando os resultados obtidos para as duas taxas de elocução (normal e rápida), vemos que os valores médios do TTL são inferiores para a taxa rápida. O mesmo é dizer que, quando a taxa de elocução aumenta, a distância temporal entre os dois *targets* (nasal e oral) tende a diminuir.

[41] Em Oliveira & Teixeira (2007b), esta medida foi designada de ORL.

Tabela 5.6: Valores médios (em ms) e intervalo de confiança (a 95%) da distância temporal entre o *target* de abertura do velo e o *target* do gesto oral (TTL), em função da taxa de elocução (normal e rápida) e da posição lexical (inicial e medial).

Taxa	Posição	Média	CI 95 %
Normal	INICIAL	-57.1	[-74.6 .. -39.6]
	MEDIAL	-25.8	[-31.9 .. -19.8]
Rápida	INICIAL	-38.8	[-42.6 .. -35.0]
	MEDIAL	-21.5	[-27.6 .. -15.5]

Do mesmo modo, regista-se também uma diferença nos valores médios do TTL, em função da posição lexical: em posição inicial, a distância entre o *target* de abertura do velo e o *target* oral é consideravelmente maior do que em posição medial. Esta tendência é comum às duas taxas de elocução analisadas, embora a diferença seja mais acentuada em taxa normal do que em taxa rápida. Cumpre, no entanto, notar que, a posição inicial (taxa normal) está sujeita a uma grande variabilidade, o que se traduz num intervalo de confiança grande.

A significância estatística das diferenças acima assinaladas foi testada com uma ANOVA de 3 factores: à taxa de elocução e posição lexical, acrescentámos o articulador oral (lábios ou ponta da língua). O resultado deste teste demonstrou como estatisticamente significativos os efeitos da posição lexical [$F(1,81)=25.99$, $p<0.001$] e do articulador oral [$F(1,81)=7.83$, $p=0.006$].

O gráfico 5.17 - onde todos estes dados são comparados - permite-nos concluir que, para além das diferenças já referenciadas, o articulador lábios está, de um modo geral, associado a valores de TTL superiores, à excepção da posição medial (taxa rápida). Por outras palavras, a distância temporal entre o target do velo e o target oral é maior, sempre que está em causa um gesto dos lábios. Em termos médios, este intervalo cifra-se nos -36.64 ms para os contextos com gestos labiais e nos -26.93 ms para os casos que envolvem uma articulação da ponta da língua.

Coordenação entre o fecho do velo e o fecho oral

Depois de quantificado o TTL, neste ponto da experiência centramo-nos na análise do designado *Target2-to-Target Lag* (T2TL), cujos valores médios são apresentados na tabela 5.7. Recordamos que esta medida diz respeito ao intervalo de tempo entre o *target* de fecho do velo e o *target* de fecho do articulador oral.

A principal observação a respeito desta tabela prende-se com as diferenças nos valores médios do T2TL entre as duas taxas de elocução consideradas. Efectivamente, a distância temporal entre os dois *targets* diminui consideravelmente quando se passa de uma taxa normal para uma taxa mais rápida, embora se registe alguma variabilidade nos valores obtidos, como se depreende dos intervalos de confiança (a 95 %).

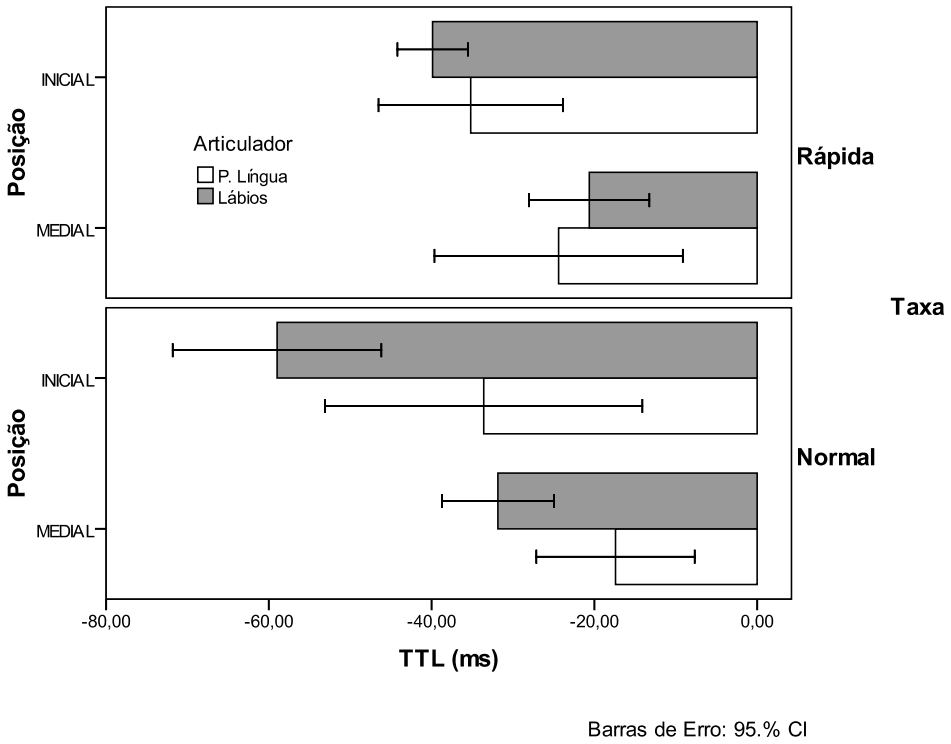


Figura 5.17: Desfasamento entre a abertura do velo e o fecho oral (TTL), em função das duas posições lexicais (inicial e medial), da taxa de elocução (normal e rápida) e do articulador oral (ponta da língua ou lábios).

Tabela 5.7: Valores médios (em ms) e intervalo de confiança (a 95%) da distância temporal entre o *target* de fecho do velo e o *target* do gesto oral (TTL), em função da taxa de elocução (normal e rápida) e da posição lexical (inicial e medial).

Taxa	Posição	Média	CI 95 %
Normal	INICIAL	106.6	[95.5 .. 117.6]
	MEDIAL	110.8	[102.4 .. 119.1]
Rápida	INICIAL	61.6	[56.7 .. 66.4]
	MEDIAL	72.8	[66.3 .. 79.2]

Para além disso, verifica-se também um aumento médio do T2TL em posição medial, comparativamente à posição inicial, nas duas taxas de elocução (normal e rápida).

Das diferenças supra mencionadas - relacionadas com a taxa de elocução, por um lado, e com a posição lexical, por outro - apenas as referentes à taxa de elocução são estatisticamente significativas [F(1,84)=99.58, p<0.001], de acordo com os resultados da ANOVA. O mesmo teste - onde se consideraram para além taxa de elocução e posição lexical, o articulador oral - revelou ainda a significância estatística do articulador oral [F(1,84)=10.42, p=0.002].

O gráfico 5.18 ilustra a influência destes dois factores (taxa de elocução e articulador oral) no intervalo temporal entre os dois *targets* (T2TL).

Para além das diferenças decorrentes da taxa de elocução, da observação do gráfico, sobressai que o articulador ponta da língua está, em média, associado a valores de T2TL superiores aos obtidos para os lábios, sobretudo em taxa normal.

A análise detalhada destes resultados, apresentada no gráfico 5.19, evidencia a influência de uma nova variável nos valores do T2TL: o vozeamento da consoante que sucede à vogal nasal (C2).

Segundo o gráfico [42], as consoantes sonoras ([b d]) implicam valores de T2TL mais elevados do que as consoantes surdas ([p t f]). Esta tendência é quantificada na tabela 5.8, onde são mostrados os valores médios do T2TL para as consoantes surdas e sonoras, nas duas taxas de elocução.

Coordenação entre o fecho do velo e a abertura da glote

Finalmente, foi investigada a coordenação entre o movimento de fecho do velo e o gesto de abertura da glote para a consoante surda subsequente. A par do desfasamento entre o gesto de fecho do velo e o gesto oral, o vozeamento desempenha um papel fundamental na criação da chamada consoante nasal intrusiva.

Os valores médios para este parâmetro encontram-se na tabela 5.9.

[42]Em consequência de danos no sinal do sensor da ponta da língua, não nos foi possível analisar alguns contextos, nomeadamente os que envolvem a consoante [d] após a vogal nasal, pelo que esta consoante não consta do gráfico 5.19.

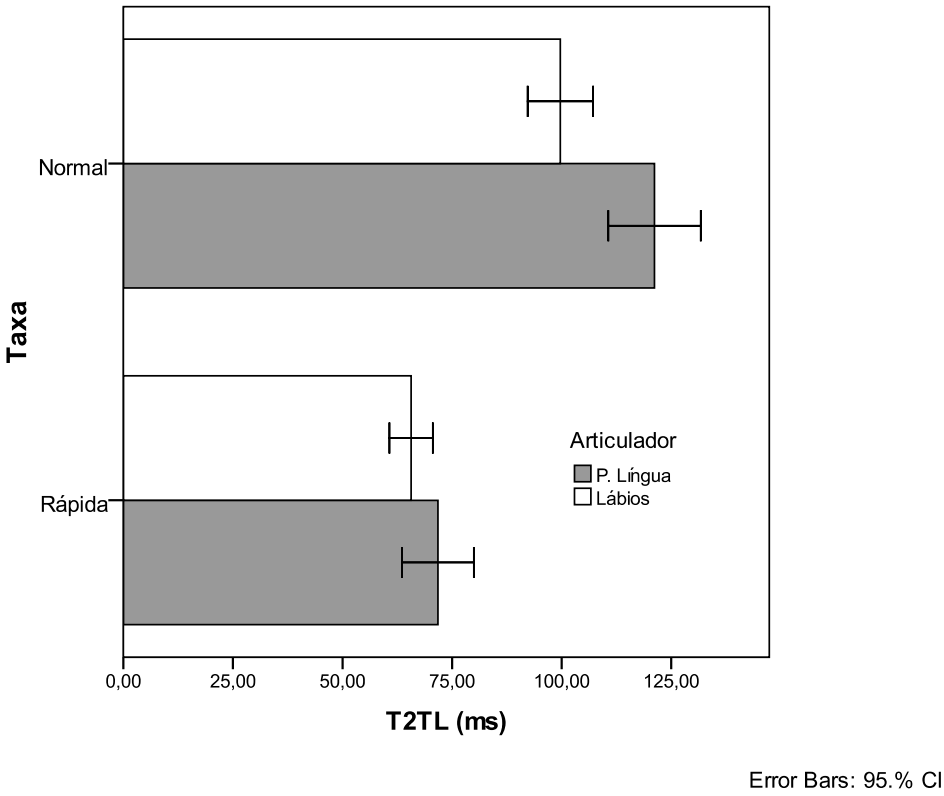


Figura 5.18: Desfasamento entre o fecho do velo e o fecho oral (T2TL), em função do articulador oral (ponta da língua ou lábios) e da taxa de elocução (normal e rápida).

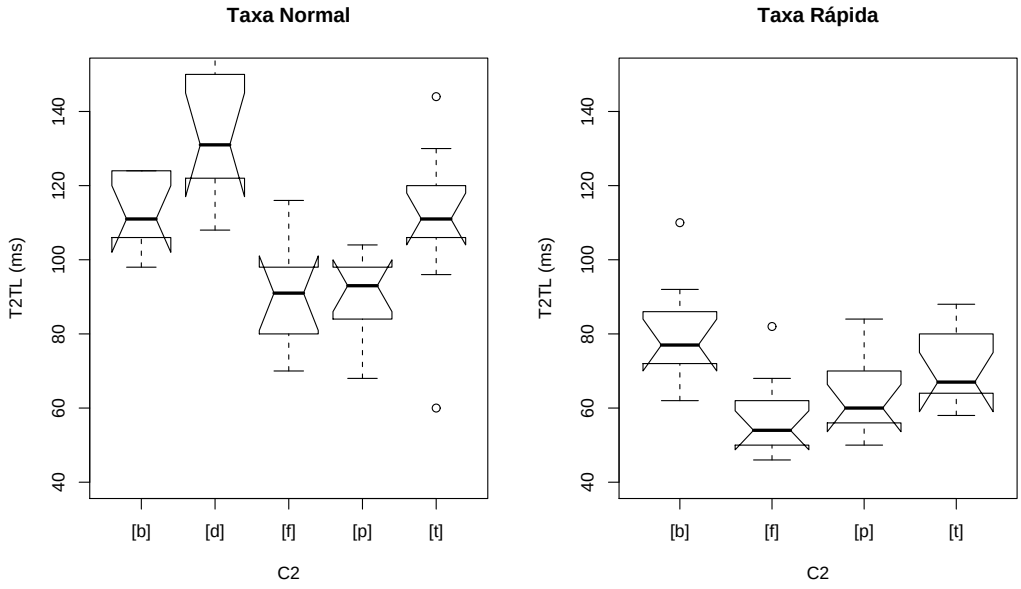


Figura 5.19: Desfasamento entre o fecho do velo e o fecho do oral (T2TL), considerando as várias consoantes que sucedem à vogal nasal (C2), nas duas taxas de elocução (normal e rápida).

Tabela 5.8: Valores médios e intervalos de confiança (a 95%) do T2TL, considerando o vozeamento da consoante (sonoras ou surdas) que se segue à vogal nasal (C2), nas duas taxas de elocução analisadas (normal e rápida).

Taxa	Vozeamento (C2)	Média	CI 95 %
Normal	Sonoras	124.4	[115.3.. 133.5]
	Surdas	97.4	[90.2 .. 104.6]
Rápida	Sonoras	80.0	[70.1 .. 89.9]
	Surdas	63.1	[59.1 .. 67.0]

Tabela 5.9: Valores médios (em ms) e intervalos de confiança (a 95%) da distância temporal entre o fecho do velo e a abertura da glote, nas duas posições lexicais (inicial e medial) e nas duas taxas de elocução (normal e rápida) estudadas.

Taxa	Posição	Média	CI 95 %
Normal	INICIAL	58.5	[49.7 .. 66.4]
	MEDIAL	53.4	[46.3 .. 60.5]
Rápida	INICIAL	46.4	[41.4 .. 51.4]
	MEDIAL	50.2	[45.6 .. 54.7]

Analisando os dados obtidos, observa-se que os valores são todos positivos, o que indica que a glote abre antes do fim do gesto de fecho do velo.

Comparando as duas taxas de elocução, verifica-se que o intervalo entre os dois movimentos articulatórios diminui para a taxa rápida, sendo esta tendência mais acentuada em posição inicial.

O resultado da ANOVA - com três factores (taxa de elocução, posição e C2) - indica que esta diferença entre a taxa normal e a taxa rápida é significativa [$F(1,55)=6.05$, $p=0.017$], assim como também é significativo o efeito de C2 [$F(2,55)=11.37$, $p<0.001$].

De acordo com o teste *post-hoc* de Tamhane, as diferenças estatisticamente significativas para este factor ocorrem entre a consoante alveolar ([t]) e as consoantes labiais ([p f]). Entre estas últimas não há diferenças significativas.

O efeito dos factores taxa de elocução e C2 nos valores médios da distância entre o fecho do velo e a abertura da glote é ilustrado na figura 5.20.

Coordenação entre a abertura do velo e a consoante precedente

No tocante à sincronização entre gestos, foi também analisada - ainda que a um nível que consideramos secundário, comparativamente com as as duas dimensões exploradas anteriormente - a coordenação temporal entre o velo e a consoante precedente (C1). Esta foi quantificada através da distância

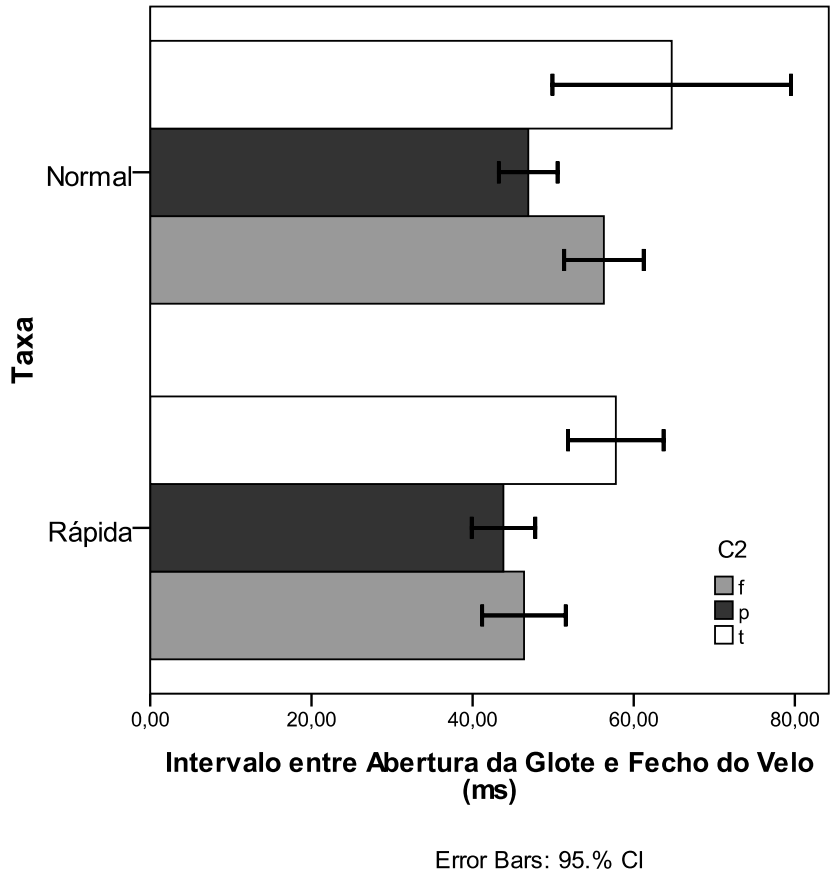


Figura 5.20: Intervalo entre a abertura da glote e o fecho do velo (em ms), considerando as três consoantes surdas após a vogal nasal [f p t], nas duas taxas de elocução (normal e rápida).

entre o início da abertura do velo (S) e a *release* (R) da consoante (SRL) (Oliveira & Teixeira, 2007b). Devido a dificuldades na detecção e marcação dos movimentos articulatórios da ponta da língua, apenas foram consideradas as consoantes labiais. Os valores médios, por taxa de elocução e posição, encontram-se na tabela 5.10.

Tabela 5.10: Valores médios (em ms) e intervalos de confiança (a 95%) da distância temporal entre o início do movimento do velo e a *release* da consoante precedente (SRL), nas duas posições lexicais (medial e final) e nas duas taxas de elocução (normal e rápida) estudadas.

Taxa	Posição	Média	CI 95 %
Normal	FINAL	21.8	[10.42 .. 33.15]
	MEDIAL	32.9	[22.70 .. 43.01]
Rápida	FINAL	-5.2	[-19.35 .. 9.02]
	MEDIAL	-1.6	[-12.57 .. 9.42]

De acordo com os resultados da tabela, o início do movimento do velo tem início após a *release* da consoante anterior, pelo menos no caso da taxa normal. À primeira vista, os valores obtidos

para a taxa rápida indiciam que a abertura do velo poderá ter início antes da *release* da consoante, contudo os testes estatísticos (*t*-Student) não comprovam como significativamente diferentes de 0 (zero) os valores obtidos para a taxa rápida ($p>0.05$). Isto significa que, em taxa rápida, o gesto de *release* pode estar em total sincronismo com o início da abertura do velo.

Os testes estatísticos ANOVA com três factores (taxa de elocução, posição e C1), apenas revelaram como significativo o efeito da taxa de elocução [$F(1,48)=27.65$, $p<0.001$].

5.4.6.3 Comentários aos resultados

A partir dos valores médios obtidos para os vários intervalos temporais, obtiveram-se as representações gráficas da figura 5.21. Estas ilustram a coordenação entre o gesto oral (lábios) [43] e o nasal - em posição inicial ou medial, nas duas velocidades consideradas (normal e rápida) - e servirão de base à discussão parcial dos resultados empreendida a seguir.

De acordo com a FA, existem dois modos de coordenação básicos - coordenação síncrona (em fase) e coordenação sequencial (anti-fase) - que estão na base da estrutura silábica e caracterizam o Ataque e a Coda, respectivamente (capítulo 2, secção 2.2.4.2). Este pressuposto teórico encontra fundamento nos resultados experimentais de estudos como o de Krakow (1993) (para o inglês) e o de Oliveira & Marin (2005) (para o PB), a propósito da coordenação entre o velo e o gesto oral. Segundo estes estudos, em posição de Coda, o gesto oral sucede ao gesto de abertura do velo (coordenação anti-fase), enquanto, em Ataque, os dois gestos têm lugar simultaneamente (coordenação em fase). O aumento da taxa de elocução tem como consequência a reorganização temporal dos gestos de Coda, ou seja, uma mudança espontânea de uma coordenação sequencial para uma coordenação síncrona (capítulo 2, secção 2.2.4.2). Assim se explica o fenómeno - frequentemente atestado no PB (Lipski, 1975) - de redução de grupos do tipo [N.d] (e.g. “partindo”) para [n] (“partino”) (Oliveira & Marin, 2005).

A um primeiro olhar, os resultados obtidos - neste e num outro estudo similar com base em EMA (Oliveira & Teixeira, 2007b) - parecem confirmar, até certo ponto, as constatações teóricas e experimentais referidas. Com efeito, os valores médios do TTL são todos negativos, o que sugere uma coordenação sequencial entre o gesto de abertura do velo e o gesto oral, típica da posição de Coda. Quando a taxa de elocução aumenta, o TTL continua negativo, mas o intervalo entre os dois *targets* (nasal e oral) diminui e os valores aproximam-se mais de zero (cf. figura 5.21). Isto significa que a coordenação sequencial tende a tornar-se síncrona - caracterizada por valores positivos ou próximos de zero (Oliveira & Marin, 2005) - o que chega mesmo a acontecer em alguns casos particulares (vd. figura 5.22).

Neste sentido, parece-nos lícito afirmar que os nossos dados apoiam parcialmente os pressupostos teóricos demonstrados noutros estudos experimentais. Admitimos, todavia, a necessidade de

[43] Os valores de duração referentes aos gestos dos lábios foram obtidos com base nos dados da informante LF, seguindo a mesma metodologia usada para o velo.

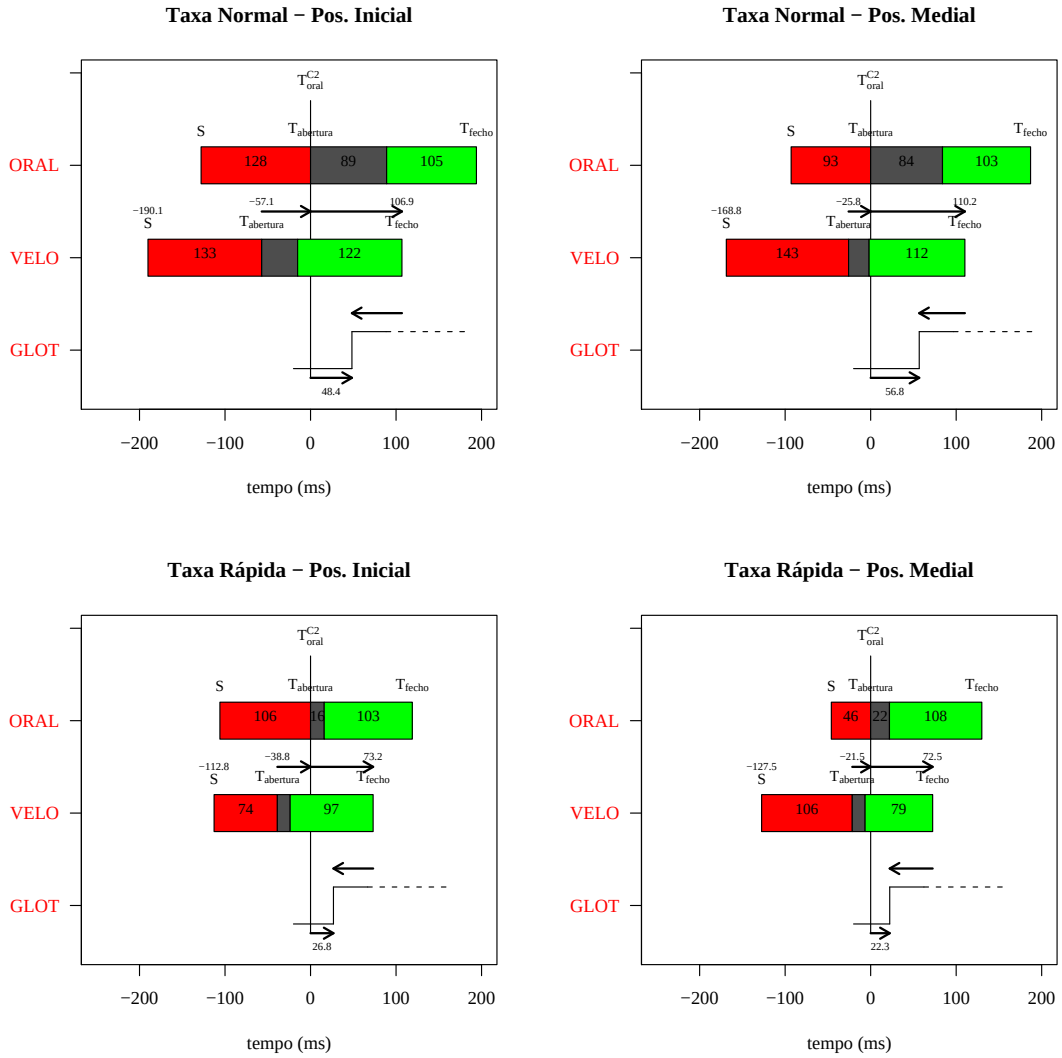


Figura 5.21: Representação gráficas gerais da coordenação entre o gesto do velo, o gesto dos lábios e o gesto de abertura da glote, nas duas taxas de elocução (normal e rápida), em posição inicial e medial.

aprofundar esta questão num estudo ulterior, em que a taxa de elocução, para além de rigorosamente controlada, possa ser, ao mesmo tempo, manipulada em vários níveis, à semelhança do realizado por Oliveira & Marin (2005).

Pese embora a grande variabilidade de valores obtidos, verificou-se também que em posição inicial, a distância entre o *target* de abertura do velo e o *target* oral é consideravelmente maior do que em posição medial, independentemente da taxa de elocução considerada. Conforme se depreende da análise da figura 5.21, esta diferença decorre do facto de, em posição inicial, o movimento de abertura do velo ter início mais cedo (cf. Lovatto *et alii*, 2007) - em média 21.3 ms antes, para a taxa normal, e 14.7 ms antes, para a taxa rápida - presumivelmente por não se encontrar sujeito a qualquer tipo de constrangimento aerodinâmico, imposto pela presença de uma consoante. Consequentemente, e

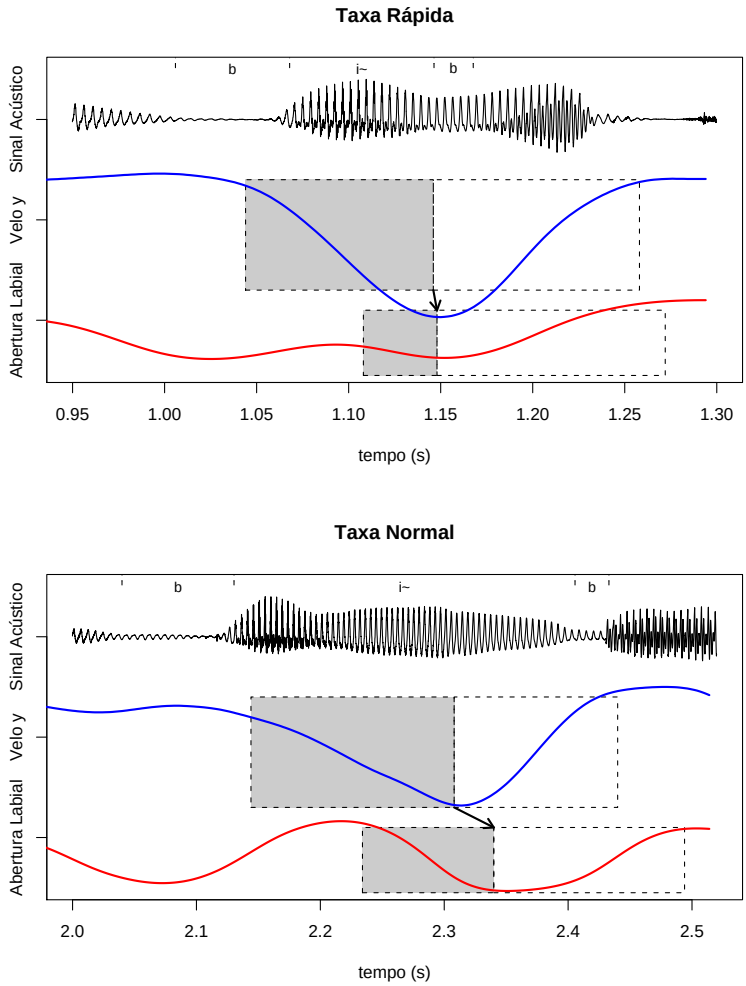


Figura 5.22: Exemplos de coordenação síncrona (em cima) e de coordenação sequencial (em baixo) entre o gesto do velo e o gesto labial, durante a produção da sequência [bĩb], em taxa rápida e em taxa normal.

assumindo uma velocidade de abertura idêntica nas duas posições lexicais, o *target* é também atingido antes.

Já os resultados relativos à influência do articulador oral nos valores médios do TTL - que sugerem que os lábios estão, de um modo geral, associados a valores de TTL mais elevados do que a ponta da língua - são absolutamente contrários aos obtidos num estudo anterior, para o PE, seguindo uma metodologia em tudo semelhante (Oliveira & Teixeira, 2007b). A questão da interferência do articulador no TTL fica, portanto, por esclarecer, até que se disponha de mais dados.

Considerando agora o T2TL, vimos que o *target* oral ocorre antes do *target* nasal. Este dado é importante, pois este desfasamento é o primeiro responsável pela criação da chamada consoante nasal intrusiva. O T2TL diminui significativamente em consequência do aumento da taxa de elocução, sendo também afectado pelo tipo de articulador oral: a ponta da língua implica sempre - mas sobretudo em taxa normal - valores de T2TL superiores. A observação atenta destes últimos dados veio revelar ainda um efeito nítido do vozeamento de C2: constatou-se que as consoantes sonoras estão associadas a valores médios de T2TL superiores, o que significa consoantes nasais intrusivas (N) potencialmente mais longas. Este resultado coaduna-se, pelo menos em parte, com os dados de estudos anteriores (Beddor, 2007), que estabelecem uma relação entre a variação temporal dos gestos orais e nasais e o vozeamento da consoante pós-nasal, embora só para algumas línguas (e.g. inglês americano). O que os dados de Beddor (2007) sugerem é uma co-variação entre a nasalização da vogal e a duração da consoante nasal: a duração do N está inversamente relacionada com a extensão da nasalidade vocálica. Assim, em comparação com as consoantes sonoras, as consoantes surdas pós-nasais estão associadas a uma nasalização mais extensiva da vogal e murmúrios nasais substancialmente mais curtos.

Paralelamente ao desfasamento entre *target* nasal e oral, a existência de uma consoante nasal intrusiva pressupõe vozeamento. Com o intuito de determinar a duração articulatória deste segmento consonântico, foi determinada a distância entre o gesto de abertura da glote e o *target* de fecho do velo, em contextos não-vozeados. Segundo os nossos resultados, em termos médios, este intervalo ronda os 50 ms, sendo significativamente influenciado pela taxa de elocução e pelo ponto de articulação da consoante pós-nasal. Conjugando as informações relativas à distância entre o gesto de abertura da glote e o fecho do velo e o T2TL, obtém-se uma duração média para a consoante nasal intrusiva de 48.4 ms, para a posição inicial, e 56.8 ms, para a posição medial (vd. figura 5.21). Quando a taxa de elocução aumenta, estes valores sofrem uma redução considerável, para os 26.8 ms e os 21.3 ms, respectivamente. Neste último caso, há ainda registo de vários exemplos em que a consoante nasal simplesmente não existe ou tem uma duração residual.

Uma consequência óbvia deste tipo de coordenação reside na diminuição do intervalo acústico correspondente à consoante oral, em resultado de uma espécie de pré-nasalização da oclusiva, conforme é alegado em vários estudos (Sousa, 1994; Moraes & Wetzels, 1992; Seara, 2000; Medeiros *et alii*, 2008). Segundo estes, comparando as durações de sequências de vogal oral+consoante oclusiva e vogal nasal+consoante oclusiva, verifica-se que o alongamento da vogal nasal é sempre compensado pela diminuição da duração da consoante subsequente, de tal forma que as duas sequências apresen-

tam durações totais praticamente equivalentes. A ideia de que a duração acústica da consoante oral [C] (como em "pa**t**a") tende a ser igual à soma do segmento nasal com a consoante oral [N.C] (como em "pa**nt**a") é-nos sugerida também por Barbosa (2006, pp.329-334).

Finalmente, os resultados do SRL sugerem que, em taxa de elocução considerada normal, o movimento de abertura do velo tem início cerca de 20 a 30 ms após a *release* da consoante anterior. Os valores médios obtidos num estudo preliminar (Oliveira & Teixeira, 2007b) são um pouco mais elevados (60 ms), o que poderá ficar a dever-se a diferenças na taxa de elocução utilizada nas duas situações de teste. Em taxa rápida, o velo inicia o movimento de abertura mais cedo, mais concretamente durante a *release* da consoante precedente.

5.5 Teste perceptivo

Os dados EMA analisados nas secções anteriores revelaram diferenças articulatórias significativas entre as vogais e as consoantes nasais, no que se refere à altura do velo. Resta agora avaliar se tal distinção é ou não relevante do ponto de vista perceptivo.

A questão é equacionada experimentalmente, no estudo perceptivo que a seguir se descreve, e deverá ter implicações directas ao nível da atribuição de um segundo grau de amplitude vélica às vogais nasais, distinto do que caracteriza actualmente as consoantes nasais no TADA. Se as diferenças de altura não forem pertinentes do ponto de vista perceptivo e se a variação deste parâmetro não se traduzir numa melhoria da qualidade e naturalidade da vogal nasal sintetizada, as diferenças articulatórias entre vogais e consoantes nasais não deverão ser tidas em conta.

Um outro problema de partida para esta investigação perceptiva relaciona-se com a coordenação entre os gestos nasal, glotal e oral. Segundo os dados articulatórios, o desfasamento entre o gesto de fecho do velo e o gesto oral seguinte, a par com um retardamento da abertura da glote, é responsável pela emergência de uma consoante nasal intrusiva antes da consoante oclusiva. O que aqui se procurará investigar é a importância e necessidade desta última para a percepção da nasalidade em português.

Tendo em mente estas duas questões centrais, o SAPWindows foi usado para sintetizar um conjunto de estímulos, sistematicamente manipulados - como adiante se explicará - quanto à 1) altura do velo e 2) coordenação temporal entre os três gestos referidos. Quanto à altura do velo, foram considerados apenas dois valores distintos; já a coordenação foi modificada gradualmente, em cinco níveis diferentes, com intervalos de 20 ms entre si. A manipulação da sincronização temporal entre os gestos tem como consequências a criação de uma consoante nasal de duração variável e variações na extensão da nasalidade vocálica.

Os estímulos foram apresentados aos ouvintes em grupos de dois e a tarefa solicitada passou pela escolha de um dos estímulos do par (A ou B), em função da sua proximidade com uma vogal nasal natural [44].

A secção que se segue é dedicada à descrição dos procedimentos conducentes à obtenção dos estímulos, caracterização dos sujeitos-ouvintes, construção e aplicação do teste. Os resultados do mesmo constam do ponto 5.5.2.

[44] Estando garantida a inteligibilidade dos estímulos, como foi possível comprovar através de testes informais, foi já possível optar por um teste de preferência, ao contrário do capítulo anterior.

5.5.1 Metodologia

5.5.1.1 Estímulos

Para não tornar o teste de percepção demasiado extenso, na construção dos estímulos, foram consideradas apenas três vogais nasais do PE - [ɐ̃], [ĩ], [ũ] - que inseridas num contexto de oclusivas bilabiais surdas, formam as seguintes sequências dissilábicas: [pɐ̃pu], [pĩpu], [pũpu].

Estes estímulos foram sujeitos a manipulação, com modificações na altura do velo da vogal nasal e alterações na coordenação intergestual.

No que respeita à altura do velo, foram consideradas duas possibilidades distintas: 1) o valor por defeito do TADA, que caracteriza a variável do tracto VEL, associada às consoantes nasais (WIDE=0.2 mm); e 2) um segundo grau de amplitude, com cerca do dobro do valor do primeiro (WIDE=0.4 mm).

Quanto à coordenação intergestual, foram utilizados cinco níveis diferentes de desfasamento entre o gesto de fecho do velo e a oclusão oral seguinte, sendo que o gesto de abertura glotal foi também deslocado na mesma proporção. Distintos graus de sobreposição entre gestos consecutivos têm como consequência a emergência de consoantes nasais intrusivas de duração distinta, antes da consoante oclusiva, como pode ser comprovado pelos espectrogramas da figura 5.25.

Para a construção das várias versões dos estímulos, foi necessário efectuar um conjunto de procedimentos, que passamos a descrever:

1. introdução das três sequências no dicionário - [pɐ̃pu], [pĩpu], [pũpu] - associando-lhes a respectiva representação fonética e silábica, de acordo com os requisitos do TADA;

2. criação das pautas gestuais base para os estímulos, especificando apenas os gestos orais associados às três vogais nasais;

3. introdução manual, na pauta gestual, de um gesto de abertura vélica, de duração similar à estimada a partir dos dados articulatórios (cerca de 140 ms), e movimentação dos gestos de abertura glotal e fecho do velo, de modo a promover, por um lado, a sobreposição entre este último gesto e a oclusão oral seguinte, e, por outro, um atraso da abertura da glote (vd. figura 5.23). Este padrão de coordenação é semelhante ao encontrado nos dados EMA e traduz-se na criação de um intervalo articulatório de 50 a 60 ms entre o fecho oral e abertura da glote;

4. modificação dos valores de *stiffness* associados ao gesto de abertura e fecho do velo, tornando-os mais compatíveis com os dados articulatórios. Isto significou assumir um valor de *stiffness* diferenciado para o gesto de abertura e o gesto de fecho, sendo o primeiro mais baixo do que o segundo;

5. alteração da primeira versão da pauta gestual, através da movimentação progressiva dos gestos articulatórios (abertura do velo, fecho do velo e abertura glotal), com intervalos de 20 ms. Esta

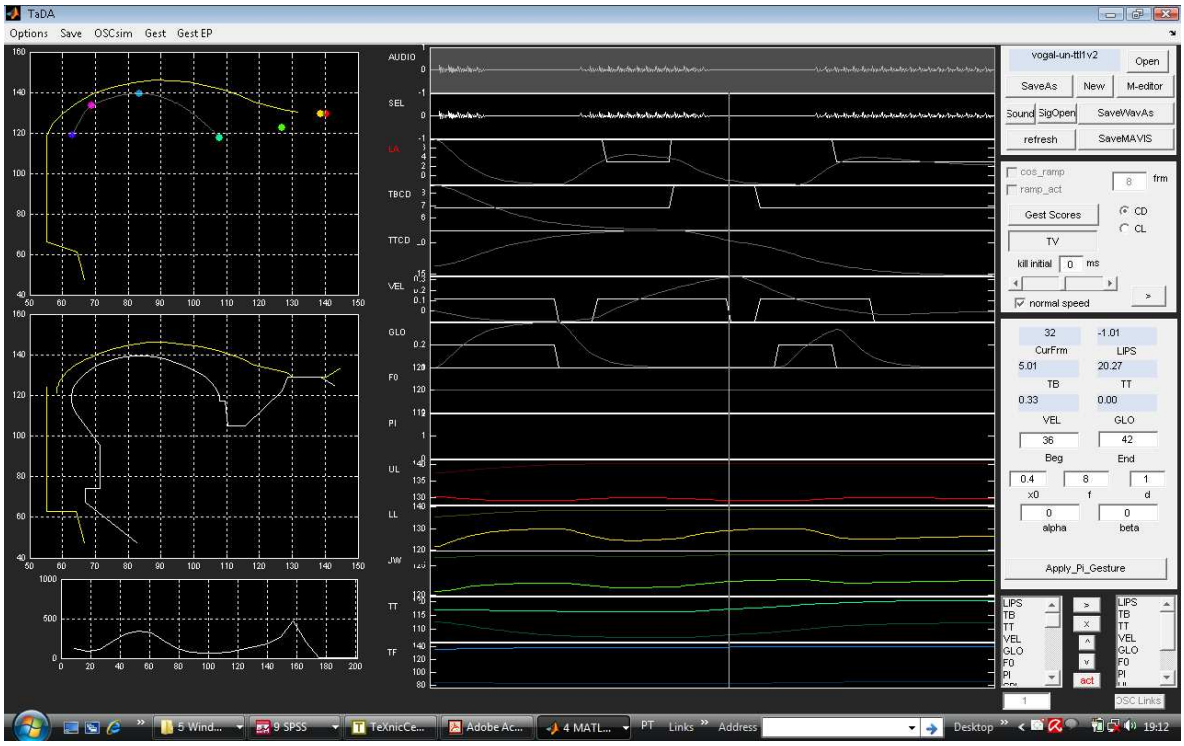


Figura 5.23: Janela do TADA referente à sequência [pũpu]. À direita, encontram-se representados o tracto vocal (dois painéis superiores) e função de área (painel inferior) correspondentes ao posicionamento do cursor. Ao centro, mostra-se o sinal acústico, a pauta gestual editada (introdução de um gesto de abertura do velo), com as trajectórias das variáveis do tracto sobrepostas e as trajectórias dos articuladores (com várias cores). À esquerda, podem ver-se quatro painéis, que permitem (de cima para baixo): 1) controlar os ficheiros (abrir, gravar, etc.); 2) correr o modelo *task-dynamics* e o CASY; 3) visualizar os valores dos parâmetros dinâmicos que controlam os gestos; e 4) editar e controlar variáveis.

operação foi realizada com base nos valores apresentados na tabela 5.11 e permitiu a criação de cinco pautas gestuais distintas para cada uma das vogais nasais;

Tabela 5.11: Valores de activação temporal (em ms) dos gestos (abertura do velo, fecho do velo e abertura da glote) para cada um dos cinco estímulos criados e duração acústica da consoante nasal intrusiva (N), resultante do desfasamento entre os três gestos.

Estímulo	Abertura do Velo	Fecho do Velo	Abertura Glote	Duração Acúst. do N
1	17 - 31	34 - 46	36 - 42	98 ms
2	15 - 29	32 - 44	34 - 40	80 ms
3	13 - 27	30 - 40	32 - 38	53 ms
4	11 - 25	28 - 38	30 - 36	40 ms
5	9 - 23	26 - 36	28 - 34	0 ms

6. repetição de todo o processo de criação das pautas gestuais, para a segunda amplitude do velo

considerada (+- 0.4 mm);

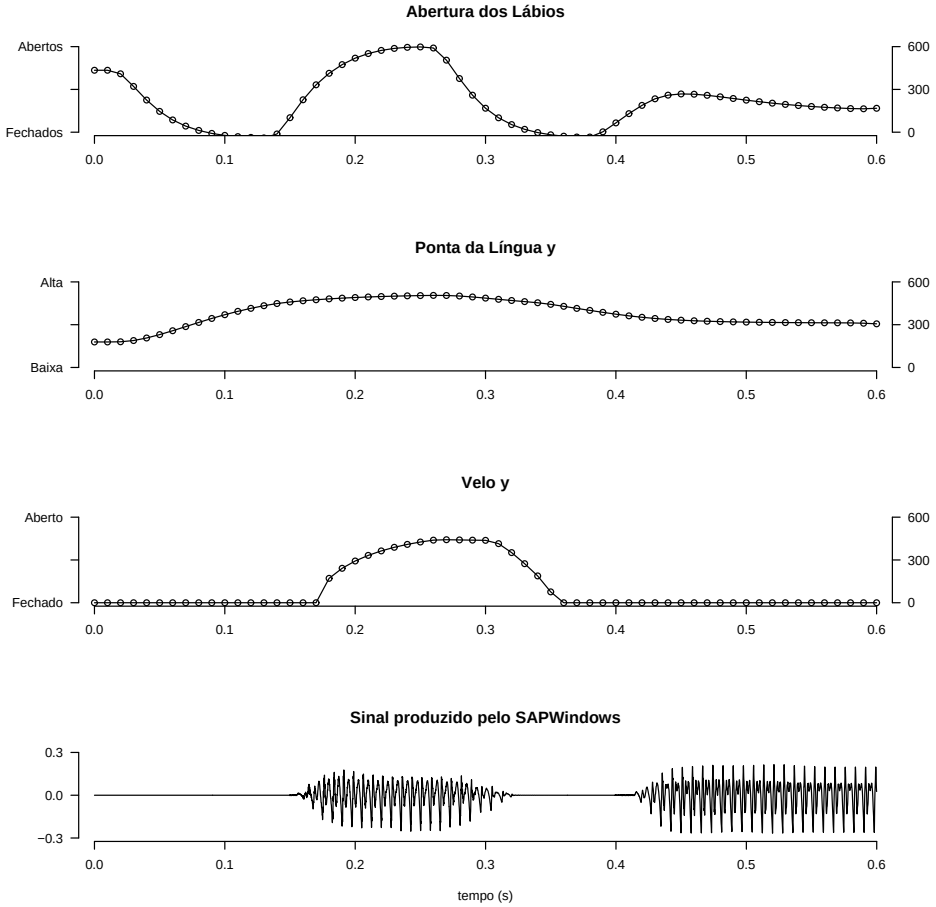


Figura 5.24: Trajectórias dos articuladores lábios, ápice da língua e velo, gerados pelo TADA (três primeiros gráficos a contar do topo) e respectivo sinal acústico produzido pelo sintetizador SAPWindows (gráfico inferior).

7. geração das trajectórias dos articuladores (vd. figura 5.24) e respectivos *outputs* para o sintetizador SAPWindows;

8. adição, manual, de uma trajectória estilizada de F0 a cada um dos estímulos;

9. síntese dos estímulos com o SAPWindows. Na Figura 5.25, encontram-se representados os espectrogramas para as cinco versões da sequência [pĩpu].

No total, foram sintetizados trinta e seis estímulos (3 vogais X 2 amplitudes X 5 coordenações).

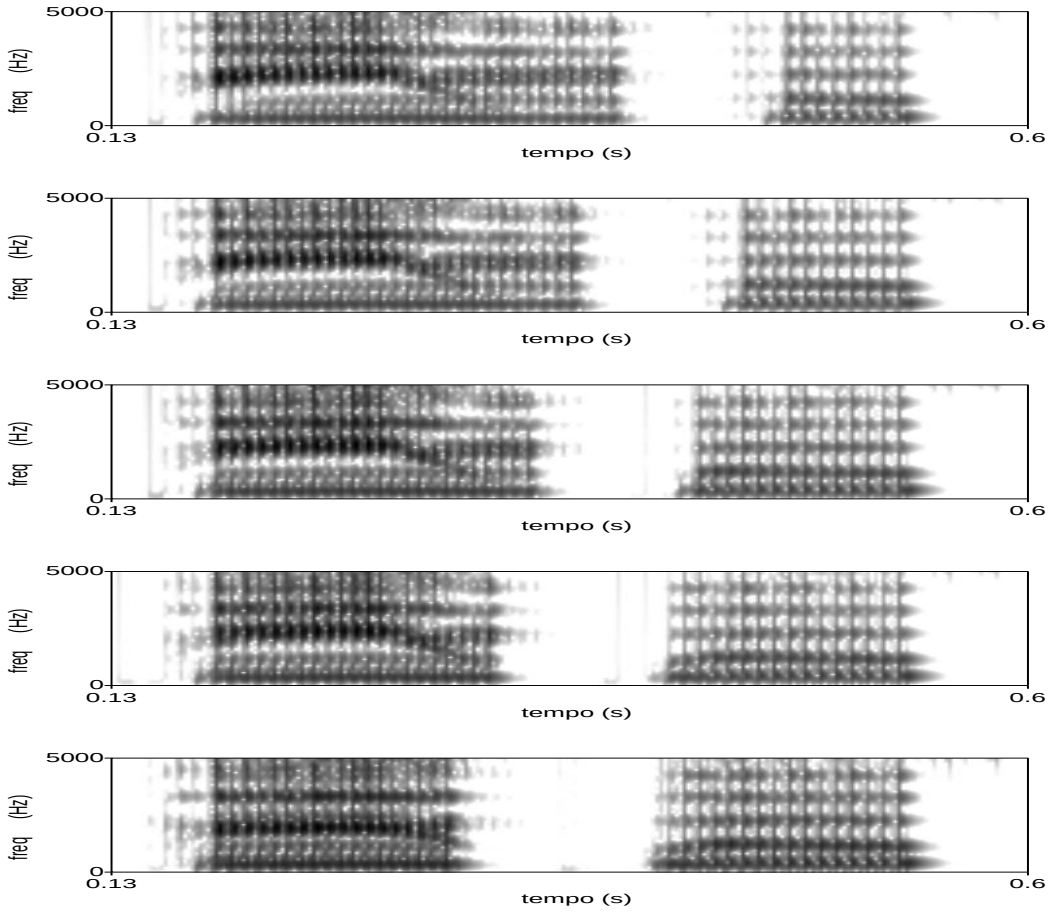


Figura 5.25: Espectrogramas (gerados em Praat), resultantes da síntese, com o SAPWindows, dos cinco estímulos criados (sequência [pĩpu]).

5.5.1.2 Construção do teste

Os procedimentos adoptados na construção do teste AB foram idênticos aos já explicitados no capítulo anterior, para o teste de identificação. Mais uma vez, recorreu-se a um programa, em linguagem Tcl/Tk (Teixeira & Vaz, 2000), para criar automaticamente os pares de estímulos (135 pares) e as repetições necessárias - neste caso duas, correspondentes à inversão da ordem dos estímulos (AB ou BA) - baralhar a ordem de apresentação dos mesmos e guardar os dados, num formato adequado à realização de testes de consistência das respostas e tratamento estatístico dos resultados.

Também a interface gráfica foi configurada, de acordo com as especificidades do teste. Para a pergunta formulada ("Que vogal nasal prefere?"), foram disponibilizadas duas hipóteses de resposta ("Primeira" e "Segunda"), a seleccionar através de um clique com o rato do computador. Tal como aconteceu no teste de identificação, cada um dos sujeitos teve a possibilidade de ouvir os estímulos mais do que uma vez, bastando para isso, pressionar o botão "Ouvir novamente". Para além destas funcionalidades, a interface concebida - vd. figura 5.26 - permitiu também a identificação do participante no teste e a monitorização da progressão do mesmo, através da apresentação de uma barra

móvel com a percentagem dos estímulos já avaliados.

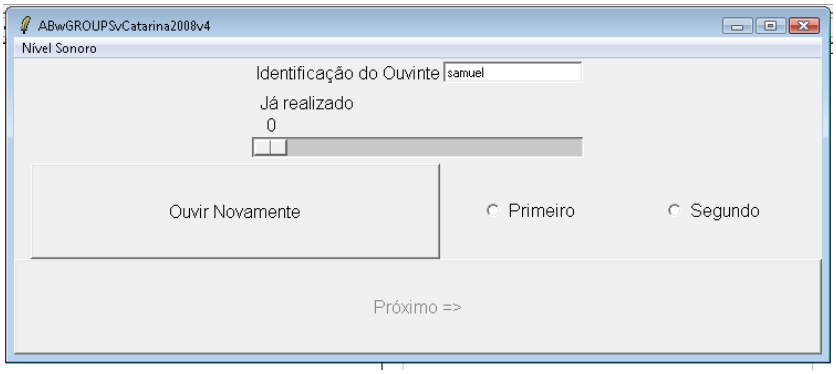


Figura 5.26: Interface gráfica do teste perceptivo AB. Para além da identificação do ouvinte (em cima), a interface dispõe de uma barra móvel ("Já realizado") para monitorização dos estímulos avaliados (ao centro), um botão para ouvir novamente os estímulos (em baixo à esquerda) e duas opções ("Primeiro" ou "Segundo") para seleccionar (em baixo à direita).

5.5.1.3 Ouvintes

Participaram no teste de percepção vinte e um indivíduos (dez do sexo masculino e onze do sexo feminino), falantes nativos do PE, com idades compreendidas entre os 20 e os 33 anos (média de idades 24.9 anos), provenientes de várias regiões, com especial incidência na zona norte-centro do país. À excepção de um - detentor de um curso técnico de electromecânica - todos os ouvintes recrutados possuem formação universitária, sendo que um é doutorado, estando, neste momento a frequentar o pós-doutoramento; quatro são alunos de doutoramento; quatro são licenciados, estando dois a frequentar um segundo curso; e onze são estudantes universitários. Dos vinte e um ouvintes, apenas dez possuem conhecimentos explícitos de Linguística e Fonética, adquiridos ao nível da licenciatura. Nenhum deles sofria de problemas auditivos. Todos os elementos aceitaram participar gratuitamente na experiência.

Tendo em vista o apuramento da consistência das respostas dos ouvintes, foi calculada a percentagem de casos em que este seleccionou o mesmo estímulo, nas duas situações de teste (AB e BA).

Os resultados desta análise preliminar dos dados são apresentados na figura 5.27.

Entre os ouvintes mais consistentes contam-se os sujeitos 6, 16, 17. No extremo inverso, com as percentagens de consistência mais baixas, estão os ouvintes 8 e 11. A opção de excluir ou não um determinado ouvinte baseou-se na proximidade entre a percentagem de consistência obtida e o valor que, em média, se atingiria, se o sujeito tivesse respondido ao acaso. Este valor corresponde a 67.5 de respostas iguais, o que determinou a exclusão dos ouvintes 12 e 13.

Para além da consistência para cada um dos ouvintes, foi ainda avaliada a concordância entre os vários ouvintes (*judge-to-judge*). Esta foi calculada segundo um método bastante simples, já

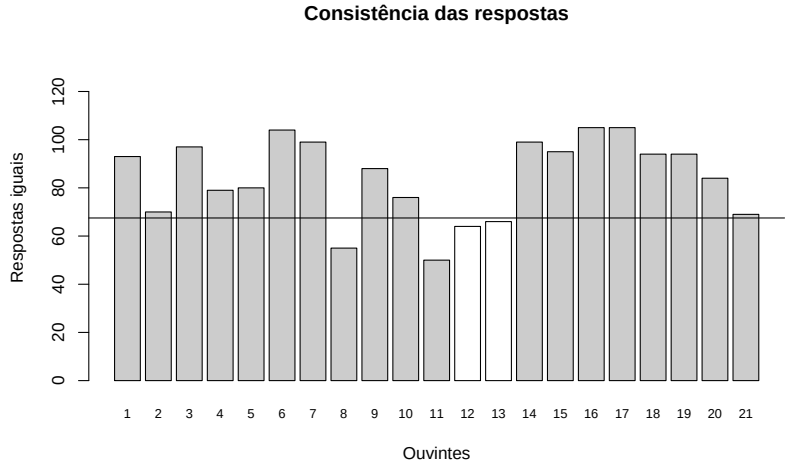


Figura 5.27: Resultados da consistência das respostas dos 21 ouvintes. Número de casos em que o ouvinte seleccionou o mesmo estímulo, nas duas situações de teste (AB e BA), e ouvintes eliminados (barras a branco).

usado por Teixeira (2000): o número de vezes em que os ouvintes concordam entre si é dividido pelo total de possibilidades e multiplicado por cem, de modo a obter uma percentagem de concordância inter-ouvintes (Schweigert, 1994, p.87). Chegámos, assim, aos valores da tabela 5.12.

Conforme se pode verificar através da análise da linha inferior da tabela, os valores médios de concordância inter-ouvintes oscilam entre os 56.6% e os 74.17%. Foi definido um patamar de 60% de percentagem de concordância, abaixo do qual os ouvintes foram excluídos. Segundo este critério, os sujeitos 8 e 21 foram eliminados.

Em suma, depois de avaliada a consistência intra e inter ouvintes, a população do teste ficou reduzida a 17 ouvintes.

5.5.1.4 Aplicação do teste

O teste foi realizado individualmente, em salas com um nível de ruído baixo a moderado, na Escola Superior de Saúde da Universidade de Aveiro (ESSUA) (4 ouvintes) e no IEETA (restantes ouvintes).

Imediatamente antes de cada sessão, os sujeitos foram instruídos oralmente no sentido de escolher uma das vogais nasais (A ou B) das pseudopalavras usadas como estímulos, apresentadas através de auscultadores e com um intervalo de 600 ms entre elas.

Depois de ouvidos os estímulos - uma ou várias vezes, de acordo com as necessidades de cada ouvinte (mediante recurso ao comando “Ouvir novamente” da interface gráfica) - a preferência foi indicada, através de um clique, com o botão do rato, junto da primeira ou da segunda opção.

Não foi imposto nenhum limite de tempo para a realização do teste perceptivo, que, em

o1	o2	o3	o4	o5	o6	o7	o8	o9	o10	o11	o12	o13	o14	o15	o16	o17	o18	o19	o20	o21
-	62.6	70.4	68.1	77.4	73.0	74.1	**54.1**	64.8	68.5	**57.4**	63.3	64.1	68.9	68.9	77.0	75.6	71.5	69.3	63.3	**52.6**
62.6	-	64.1	63.3	71.1	64.4	69.3	**55.2**	64.4	60.7	**59.3**	**55.6**	60.7	61.9	61.1	65.6	67.0	64.4	61.5	**57.8**	**55.2**
70.4	64.1	-	65.2	72.2	72.2	71.9	**53.3**	67.8	63.3	**50.7**	60.4	**58.9**	68.1	**56.3**	71.9	75.6	74.4	65.6	68.5	**48.1**
68.1	63.3	65.2	-	69.3	69.3	71.9	**45.9**	64.8	61.9	**58.9**	**52.2**	**59.6**	68.1	60.0	65.9	68.1	73.7	63.3	61.1	**50.4**
77.4	71.1	72.2	69.3	-	74.1	77.4	**53.7**	69.6	72.6	63.0	63.0	68.1	71.5	64.1	77.4	78.1	73.3	69.6	61.5	**55.2**
73.0	64.4	72.2	69.3	74.1	-	73.7	**50.7**	68.1	66.7	**54.1**	**54.8**	64.4	75.2	62.6	75.2	76.7	73.3	68.9	68.1	**47.8**
74.1	69.3	71.9	71.9	77.4	73.7	-	**48.9**	70.0	64.8	**58.1**	**54.4**	64.8	74.8	60.7	71.1	75.6	74.4	73.0	67.0	**48.1**
54.1	**55.2**	**53.3**	**45.9**	**53.7**	**50.7**	**48.9**	-	**52.2**	**55.9**	**57.4**	**50.0**	**50.0**	**52.6**	**49.6**	**54.1**	**51.1**	**50.7**	**49.3**	**47.8**	**53.3**
64.8	64.4	67.8	64.8	69.6	68.1	70.0	**52.2**	-	60.7	**52.6**	**57.8**	64.4	67.8	**56.7**	65.6	70.7	68.1	63.7	**57.0**	**52.2**
68.5	60.7	63.3	61.9	72.6	66.7	64.8	**55.9**	60.7	-	60.0	62.2	62.2	61.9	**59.6**	64.8	64.8	63.0	65.9	60.0	**55.2**
57.4	**59.3**	**50.7**	**58.9**	63.0	**54.1**	**58.1**	**57.4**	**52.6**	60.0	-	**58.5**	60.0	**56.7**	**51.5**	**56.7**	**53.7**	60.0	**58.5**	**48.9**	**55.9**
63.3	**55.6**	60.4	**52.2**	63.0	**54.8**	**54.4**	**50.0**	**57.8**	62.2	**58.5**	-	**56.3**	**54.4**	**59.6**	**58.9**	**57.4**	**55.6**	**58.5**	**51.1**	**53.7**
64.1	60.7	**58.9**	**59.6**	68.1	64.4	64.8	**50.0**	64.4	62.2	60.0	**56.3**	-	61.9	**56.7**	64.8	66.3	63.0	**59.3**	**54.1**	**58.1**
68.9	61.9	68.1	68.1	71.5	75.2	74.8	**52.6**	67.8	61.9	**56.7**	**54.4**	61.9	-	62.2	70.4	74.8	75.2	70.0	67.0	**48.9**
68.9	61.1	**56.3**	60.0	64.1	62.6	60.7	**49.6**	**56.7**	**59.6**	**51.5**	**59.6**	**56.7**	62.2	-	63.0	61.5	61.1	**55.2**	**56.7**	**53.3**
77.0	65.6	71.9	65.9	77.4	75.2	71.1	**54.1**	65.6	64.8	**56.7**	**58.9**	64.8	70.4	63.0	-	77.8	72.2	67.8	69.3	**52.6**
75.6	67.0	75.6	68.1	78.1	76.7	75.6	**51.1**	70.7	64.8	**53.7**	**57.4**	66.3	74.8	61.5	77.8	-	75.2	72.2	66.3	**51.9**
71.5	64.4	74.4	73.7	73.3	73.3	74.4	**50.7**	68.1	63.0	60.0	**55.6**	63.0	75.2	61.1	72.2	75.2	-	71.1	63.7	**47.8**
69.3	61.5	65.6	63.3	69.6	68.9	73.0	**49.3**	63.7	65.9	**58.5**	**58.5**	**59.3**	70.0	**55.2**	67.8	72.2	71.1	-	63.7	**50.7**
63.3	**57.8**	68.5	61.1	61.5	68.1	67.0	**47.8**	**57.0**	60.0	**48.9**	**51.1**	**54.1**	67.0	**56.7**	69.3	66.3	63.7	63.7	-	**41.9**
52.6	**55.2**	**48.1**	**50.4**	**55.2**	**47.8**	**48.1**	**53.3**	**52.2**	**55.2**	**55.9**	**53.7**	**58.1**	**48.9**	**53.3**	**52.6**	**51.9**	**47.8**	**50.7**	**41.9**	-
72.2	67.2	69.9	68.0	74.1	71.6	72.2	**56.7**	67.9	67.7	61.5	61.8	65.8	70.6	64.0	72.0	73.0	71.5	68.8	64.7	**56.6**

Tabela 5.12: Percentagem de concordância entre os vários ouvintes participantes no teste perceptivo (*judge to judge*). Na linha inferior apresenta-se a média de concordâncias para um determinado ouvinte. A negrito encontram-se assinalados os valores inferiores a 60 %.

média, durou cerca de 30 minutos, com alguns ouvintes a demorar quase 1 hora e outros apenas 20 minutos.

Os primeiros quatro pares de estímulos (usando duas coordenações diferentes para duas das vogais) apresentados tiveram como objectivo familiarizar os sujeitos, quer com a tarefa a desempenhar, quer com a voz sintética, pelo que não foram tidos em consideração na análise dos resultados.

5.5.2 Resultados

Para efeitos de análise, retiveram-se apenas as contagens do número de vezes em que uma determinada amplitude ou coordenação foi escolhida pelos ouvintes.

A média das preferências dos ouvintes é apresentada na figura 5.28, sob a forma de dois gráficos de barras. O primeiro contempla as preferências dos ouvintes, quando a altura do velo dos estímulos era distinta. No segundo constam as respostas dos ouvintes, quando em causa esteve um par de estímulos com a mesma altura do velo.

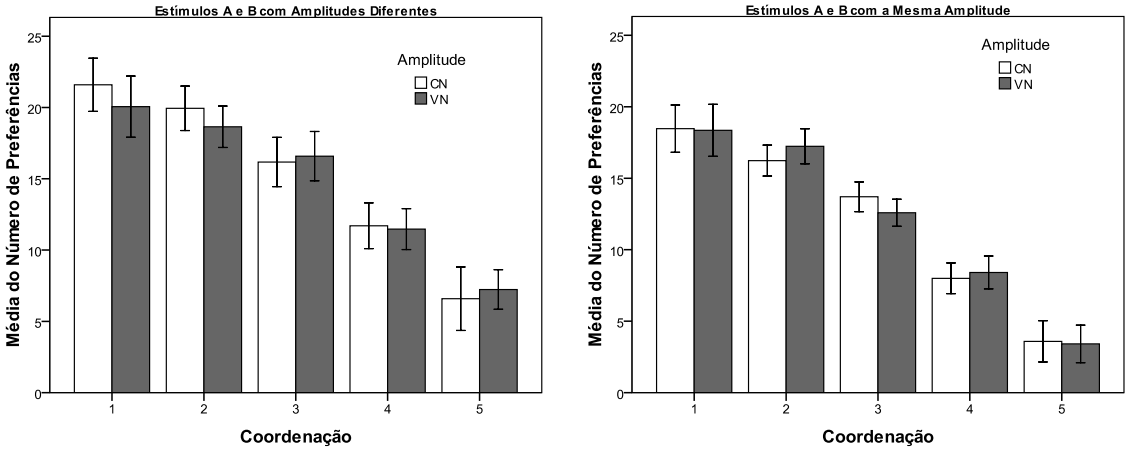


Figura 5.28: Média do número de preferências dos ouvintes, quando a altura do velo era distinta (gráfico à esquerda) e quando a altura do velo era similar (gráfico à direita), para as cinco coordenações e duas amplitudes (CN e VN) testadas.

Independentemente da amplitude dos estímulos, a preferência dos ouvintes vai para as coordenações 1 e 2, ou seja, aquelas em que o intervalo entre *target* de fecho oral e a abertura da glote ronda valores próximos dos dados articulatórios, traduzindo-se na criação de uma consoante oral intrusiva com uma duração acústica aproximada de 98 ms e 80 ms, respectivamente.

Considerando o gráfico em que a amplitude dos estímulos é diferente, verifica-se, em termos gerais, que a média de preferências dos ouvintes é muito similar para as duas amplitudes (CN e VN). É, contudo, possível observar pequenas diferenças nas escolhas dos intervenientes no teste, em função da coordenação associada ao estímulo: em face de estímulos associados a uma coordenação 1 ou 2, os ouvintes parecem preferir os que têm uma amplitude típica das consoantes nasais, invertendo-se esta tendência a partir da coordenação 3.

Segundo os resultados do teste de variância de medida repetida, com dois factores (amplitude e coordenação), as diferenças decorrentes da coordenação são significativas ($p<0.001$), ao contrário do efeito da amplitude ou da interacção entre os dois factores.

Os testes *post-hoc* de Bonferroni revelaram como significativas as diferenças entre todas as coordenações, excepto entre a coordenação 1 e 2.

5.5.3 Comentários aos Resultados

Para a realização do teste perceptual acima descrito, foi criada, usando o SAPWindows, uma série de trinta e seis estímulos, sistematicamente modificados quanto à altura do velo e sincronização temporal entre os gestos. Um conjunto de vinte e um ouvintes, falantes nativos do PE, foi chamado a julgar, através de um teste de preferência, a naturalidade e qualidade das vogais nasais assim produzidas.

Segundo os resultados obtidos, os ouvintes portugueses preferem um padrão de sincronização temporal que favoreça a nasalização da parte final da vogal e a emergência de uma consoante nasal longa (com uma duração acústica de cerca de 80 a 100 ms) entre a vogal e a consoante oclusiva. Este tipo de coordenação temporal é compatível com os dados articulatórios apresentados.

Este resultado está também em linha com estudos perceptuais anteriores, nomeadamente o de Teixeira (2000) e o de Stevens *et alii* (1987), que evidenciam, através de testes perceptuais com estímulos sintéticos, a preferência dos ouvintes portugueses por vogais nasais seguidas de um murmúrio nasal. Pelo contrário, os ouvintes franceses são praticamente indiferentes à presença deste murmúrio, preferindo antes uma nasalização forte e longa da vogal (Stevens *et alii*, 1987).

Já o efeito da variação da altura do velo na percepção das vogais nasais do português é, de acordo com os resultados obtidos, praticamente nulo. Embora os dois estudos não sejam directamente comparáveis - testando parâmetros distintos (altura do velo e *velopharyngeal port opening* - *VPOQ*), mas com alguma relação entre si - estes dados contrariam, de certa forma, os resultados perceptivos de Hajek & Watson (2007), que indicam que os ouvintes portugueses são altamente sensíveis às variações no VPOQ, sendo os estímulos com maior VPOQ percebidos como mais nasais. Uma possível explicação para estas diferenças poderá residir nos níveis de amplitude testados: em conformidade com o objectivo a que nos propusémos, no presente estudo, foram considerados apenas dois valores distintos de amplitude, sendo o segundo o dobro do primeiro. Para que se esclareça se a altura do velo é ou não uma pista importante para a percepção da nasalidade, poderá ser necessário contemplar mais graus de amplitude, eventualmente com uma maior diferença entre eles.

5.6 Discussão Final

Na presente secção, proceder-se-á à discussão global dos resultados, em função do objectivo enunciado na Introdução a este capítulo: a descrição e análise do fenómeno da nasalidade vocálica à luz dos

princípios dinâmicos da FA. Adicionalmente, procuraremos reflectir sobre algumas questões suscitadas pelos resultados, pertinentes para o aprofundamento dos conhecimentos acerca do mecanismo de produção das vogais nasais.

No quadro teórico da FA, o tratamento das vogais nasais implica uma referência a, pelo menos, duas variáveis do tracto: o corpo da língua (TB) e o velo (VEL).

Para a caracterização da primeira, contámos, à semelhança de capítulos anteriores, com a informação obtida através de ressonância magnética. Os dados recolhidos, através da sobreposição de contornos articulatórios, evidenciaram pequenas diferenças entre as vogais nasais e as vogais orais, ao nível do posicionamento da língua. As mudanças mais acentuadas registaram-se para as vogais nasais mais posteriores, sobretudo o [ɐ̃]. Em face destes resultados - que revelam que a nasalidade vocálica não envolve, no PE, ajustes articulatórios significativos, para além da abertura do velo - enveredámos por uma descrição gestual da dimensão oral das vogais nasais baseada na caracterização gestual das vogais orais correspondentes. De significativo, há apenas a registar uma possível alteração do *target* do local de contrição do [ɐ̃] - de faríngeo para palatal - em consequência da anteriorização do corpo da língua. Apesar disso, a apreciação informal da qualidade do som gerado a partir da configuração palatal - mais próxima de um [ɛ̃] do que de um [ɐ̃] - ditou a manutenção do CL oral, definido como faríngeo.

Uma outra questão que se levanta a propósito destes resultados - e que tivemos já ocasião de sublinhar - prende-se com as diferenças entre o português e o francês, ao nível da postura assumida pelos articuladores orais, durante a produção das vogais nasais. No francês, a abertura do velo é acompanhada por um conjunto de manobras articulatórias - nomeadamente recuo do dorso da língua e arredondamento dos lábios (Zerling, 1984; Demolin *et alii*, 1998; Delvaux *et alii*, 2002; Demolin *et alii*, 2003) - que desempenham um papel fundamental na implementação da nasalidade (Maeda, 1993), mas que, no português, simplesmente não se verificam.

Estes dados vêm, assim, confirmar as predições de Delattre (1968), que distingue duas formas de nasalização, por abaixamento do velo e por dupla articulação (abaixamento do velo conjugado com ajustes ao volume da cavidade faríngea), que caracterizariam, respectivamente, o português (e o inglês americano) e o francês [45].

Já a caracterização do gesto do velo - quer em termos de parâmetros dinâmicos (grau de constrição, *stiffness*), quer na sua relação temporal com os demais gestos orais - assentou num estudo EMA. Os resultados recolhidos serão aqui sujeitos a uma apreciação crítica em função das questões enunciadas no ponto 5.4.1.

[45] A afirmação de Delattre acerca da nasalidade vocálica por abaixamento do velo diz assim: "Quant à la nasalisation vocalique par simple abaissement du voile du palais, sans ajustement du volume pharyngal, elle ne produit acoustiquement que l'amortissement du premier formant, ce qui donne une impression de nasalité moin forte que dans les voyelles nasales françaises. Cette forme de nasalisation est fréquente dans une langue comme l'anglais d'Amérique (...) On trouve la même forme de nasalisation par simple abaissement du voile du palais dans une langue comme le portugais où les voyelles nasales sont restées relativemente fermées, /ĩ/ /ə̃/ /ẽ/ /õ/ /ũ/, c'est-à-dire avec grande cavité pharyngale non ajustée au volume de lá cavité vélique." (Delattre, 1968, p.70).

A primeira questão contemplada prendia-se com a análise da altura do velo durante a produção dos sons nasais, tendo em vista a definição de um grau de constrição (CD) para a variável do tracto VEL. Com base nos resultados experimentais de Rossato *et alii* (2003), Amelot & Rossato (2006), Rossato *et alii* (2006) e Amelot & Rossato (2007), previa-se uma maior amplitude vélica para as vogais nasais do que para as consoantes nasais.

Os dados articulatórios reunidos demonstraram que, no PE, a altura do velo varia de acordo com a seguinte progressão: C<V<N<Vn. Para além disso, as diferenças entre consoantes nasais e vogais nasais revelaram-se estatisticamente significativas, confirmando, assim, a hipótese inicial, levantada na questão 1.

Não obstante as diferenças articulatórias entre vogais e consoantes nasais, no que toca à altura do velo, segundo os resultados do teste de preferência, tal distinção não parece assumir qualquer relevância perceptiva. Neste sentido, e até que novos dados estejam disponíveis, não haverá qualquer necessidade de atribuir um terceiro nível de abertura velar às vogais nasais, distinto do WIDE, que, na FA, caracteriza as consoantes nasais.

Um outro aspecto que consideramos digno de nota prende-se com o método usado para determinar a altura do velo. Nos estudos experimentais anteriores (Rossato *et alii*, 2003; Amelot & Rossato, 2006; Rossato *et alii*, 2006; Amelot & Rossato, 2007), realizados, na sua maioria para o francês, a altura do velo foi obtida no ponto médio dos fones, determinado a partir do sinal acústico. Embora esta abordagem permita estimar a altura do velo para qualquer tipo de som, independentemente do contexto, ela revelou-se totalmente ineficaz para as vogais nasais do PE, mascarando até - como tivemos ocasião de comprovar - as diferenças de amplitude entre vogais e consoantes nasais. Com efeito, ao contrário do francês (e.g. Amelot, 2004), no PE, o ponto de máxima amplitude não ocorre a meio da vogal, mas mais próximo do fim desta, imediatamente antes da consoante nasal intrusiva, ou mesmo durante esta última (cf. Lovatto *et alii*, 2007). Por isso, a amplitude do velo nas vogais nasais foi calculada a partir do ponto mais baixo da trajectória do sensor do velo, e não do ponto médio acústico, usado apenas para os restantes sons.

A segunda questão analisada decorre directamente da primeira e previa uma relação entre a altura da língua e a altura do velo, mais concretamente uma tendência para as vogais baixas serem produzidas com o velo numa posição mais baixa do que as vogais altas.

No respeitante às vogais orais, verificou-se que o velo está mais baixo na vogal [u] do que nas vogais [i a], o que contraria os resultados de estudos anteriores (e.g. Ohala, 1975; Rossato *et alii*, 2006; Clumeck, 1976), que reportam mesmo casos de vogais orais baixas produzidas com a passagem velo-faríngea aberta. Embora a importância desta questão para os objectivos centrais do presente estudo seja relativamente lateral, é possível já adiantar que os dados de amplitude recolhidos para o segundo informante, entretanto analisados, seguem a tendência geral, com a vogal [a] a apresentar uma amplitude vélica inferior à das vogais altas. Recordamos que este mesmo informante havia já participado num outro estudo articulatório, baseado em EMA, cujos resultados foram muito similares

(Rossato *et alii*, 2006). O que estes dados parecem sugerir é uma variação condicionada por estratégias individuais, mas esta explicação precisa, evidentemente, de ser aprofundada e validada em estudos ulteriores.

Os dados coligidos também não permitiram suportar a hipótese de uma correlação entre a altura da vogal nasal e a amplitude do velo, já que as diferenças de amplitude do velo entre as várias vogais nasais não se revelaram estatisticamente significativas [$F(4,195)=1.12$, $p>0.5$]. Isto significa que não haverá necessidade de assumir diferentes graus de constrição para as várias vogais nasais.

A questão 3, relativa à duração dos movimentos do velo, visou, essencialmente, possibilitar os estudos de coordenação desenvolvidos posteriormente. Os resultados da duração do ciclo completo do velo (cerca de 300 ms) coadunam-se perfeitamente com os valores obtidos em estudos prévios (Stevens, 1998; Teixeira *et alii*, 2001; Amelot, 2004; Basset *et alii*, 2006; Oliveira & Teixeira, 2007b). Apesar da duração total ser similar nas várias línguas analisadas nesses trabalhos, a duração das fases de abertura e fecho parece variar com a língua estudada, com possíveis consequências ao nível da coordenação entre o gesto do velo e o gesto oral seguinte e da presença ou não de nasal intrusiva. No PE e no inglês americano (Solé, 1995), o movimento de abertura tende a ser mais longo do que o de fecho, ao passo que no francês (Amelot, 2004) as durações são similares.

Esta mesma questão previa ainda uma eventual influência de variáveis contextuais (Amelot, 2004) e/ou prosódicas (Solé, 1995) sobre a duração do movimento do velo. Os nossos resultados confirmam um efeito significativo da taxa de elocução (cf. Krakow, 1993) sobre todas as medidas de duração efectuadas, com todos os movimentos a sofrer um encurtamento considerável.

Quanto à possível influência da posição lexical na duração das diferentes fases do movimento do velo, os problemas decorrentes da anotação automática dos gestos impedem-nos, como já dissémos, de formular conclusões definitivas.

A duração dos gestos articulatórios está intimamente relacionada com o *stiffness* desses mesmos gestos. Este último parâmetro, formalmente incorporado pela FA (e pelo TADA), procura caracterizar a velocidade dos movimentos articulatórios e tem consequências importantes ao nível da duração: os gestos com *stiffness* mais elevado resultam em movimentos de menor duração.

O problema foi contemplado na questão 4, que teve como propósito essencial estimar os valores de *stiffness* para os movimentos de abertura e fecho do velo, para além de determinar possíveis influências de vários factores (e.g. taxa de elocução) sobre esses mesmos valores. Os resultados obtidos indicam que o movimento de abertura está associado a valores de *stiffness* mais baixos do que o movimento de fecho, o que implica, por um lado, que o velo seja mais lento a abrir do que a fechar - o que está de acordo com observações prévias para línguas como o francês (Benguerel *et alii*, 1977; Horiguchi & Bell-Berti, 1987) [46]- e, por outro, que o gesto de abertura seja mais longo do que o de

[46] As observações de Benguerel *et alii* (1977), para o francês, são contrárias aos dados publicados por Rossato *et alii* (2006), que sugerem que o velo desce e sobe a uma velocidade similar. Esta seria, aliás, uma das diferenças entre a nasalidade do francês e do PE, língua na qual, segundo o mesmo estudo, o movimento de abertura é mais lento do que o de fecho. Articulando os dois aspectos (velocidade e duração), os resultados de Rossato *et alii* (2006) parecem mais

fecho, o que efectivamente se verifica nos nossos dados.

Daqui decorre a necessidade contemplar, no TADA, dois valores distintos para o gesto de abertura e fecho do velo, sendo o primeiro mais baixo do que o segundo.

Ainda de acordo com nossos resultados, uma variável que afecta significativamente os valores de *stiffness* é a taxa de elocução. Estes aumentam consideravelmente, em razão da mudança na taxa de elocução, embora as diferenças entre o *stiffness* de abertura e fecho se mantenham. Isto significa que a velocidade dos referidos gestos irá também aumentar, enquanto a duração diminui, o que faz todo o sentido, à luz dos nossos dados.

Finalmente, o problema da coordenação entre os gestos nasais, orais e glotais - uma questão absolutamente fundamental do ponto de vista dos objectivos a atingir neste trabalho - foi equacionado na pergunta 5. Partindo da proposta de Albano (1999) - delineada com base nos postulados teóricos da FA - antecipava-se que o gesto de abertura vélica pudesse começar depois do início do gesto vocálico, terminando depois do fim deste e sobrepondo-se ao gesto consonantal seguinte (no caso de o haver).

Como se depreende da análise dos resultados do SRL - que implicam a existência de uma consoante antes da vogal nasal - em circunstâncias normais, o movimento de abertura do velo tem início cerca de 20 a 30 ms após a distensão da consoante (oclusiva ou fricativa) [47], ou seja, depois do início do gesto vocálico. Com efeito, apesar de não dispormos de medidas directas de coordenação entre o velo e a língua - muito difíceis de obter, por causa da detecção dos movimentos do dorso da língua - é sabido que os movimentos articulatórios das vogais e consoantes se sobrepõem (e.g. Öhman, 1966) e, em casos de sílabas com Ataques simples, o início do gesto vocálico é aproximadamente síncrono com o início do gesto consonântico (Goldstein *et alii*, 2006) [48]. Isto significaria que, pelo menos em contextos CV, o gesto vocálico já estaria activo, quando o velo começa a abrir, depois da *release* da consoante (oclusivas e fricativas).

Em posição inicial, como vimos, o movimento do velo, não estando sujeito a qualquer tipo de constrangimentos aerodinâmicos, tem início um pouco antes, presumivelmente quase em simultâneo como o início da vogal.

Em qualquer dos casos, no início da vogal, o véu palatino encontra-se ainda numa posição elevada ou apenas levemente abaixada (cf. Gregio, 2006), o que poderá traduzir-se, em termos acústicos, numa parte inicial caracterizada por uma configuração formântica próxima da vogal oral (Drenska, 1988; Sousa, 1994; Seara, 2000). Esta fase oral inicial não é exclusiva das vogais nasais do

compatíveis com as durações obtidas por Amelot (2004): uma velocidade similar para a abertura e fecho do velo implicaria uma duração similar.

[47] Embora não disponhamos de dados, prevê-se que, no caso das laterais ou vibrantes, o velo possa abrir um pouco antes, conforme sugerem os dados articulatórios de Amelot (2004) e os dados acústicos de Montagu (2007), para o francês. Efectivamente, como já referimos, as consoantes obstruintes envolvem grandes constrangimentos articulatórios e são, portanto, mais resistentes à coarticulação nasal, causando um "retardamento" do abaixamento do velo, pelo que a vogal nasal seguinte pode ser produzida como parcialmente oral.

[48] Cf. a este propósito as afirmações de Browman & Goldstein (1990b, p.352): "The X-ray data we have analyzed (...) have consistently supported the contention (...) that consonant articulations are superimposed on continuous vowel articulations, which themselves minimally overlap.".

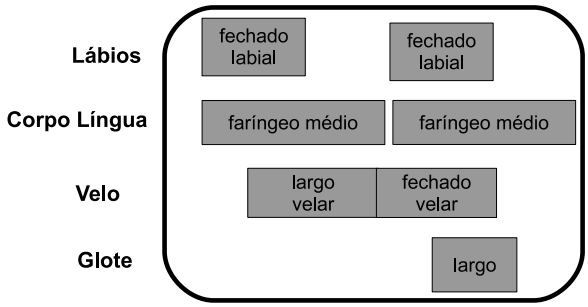


Figura 5.29: Pauta gestual das vogais nasais do PE. Nesta ilustração, o gesto dos lábios especifica a consoante anterior e seguinte (uma oclusiva bilabial surda) e o gesto de corpo da língua corresponde a um [ɐ̃].

português, tendo sido atestada, acustica e perceptualmente, também para as nasais do francês (Montagu, 2007).

Contrariando os pressupostos de Albano (1999), segundo os dados articulatórios recolhidos, o gesto de abertura vélica não se sobrepõe ao gesto consonantal seguinte. A presença de uma consoante nasal intrusiva resulta, antes, de um desfasamento entre o *target* oral, o movimento de fecho do velo e o gesto de abertura da glote. Por outras palavras, o articulador oral atinge o *target* antes do velo completar o movimento de fecho e da glote abrir, induzindo a chamada terceira fase da vogal nasal, correspondente a um segmento de natureza consonântica.

Esta análise é formalizada na pauta gestual da figura 5.29.

O grau de sobreposição entre estes três gestos pode, contudo, variar em função de vários factores - como o vozeamento da consoante seguinte, o articulador oral e a taxa de elocução - pelo que a duração do N vai também variar. Em última análise, este pode nem sequer estar presente, como foi já atestado anteriormente por Sousa (1994), Seara (2000) e Lovatto *et alii* (2007) [49].

Uma das principais interrogações que se levanta a propósito destes dados diz respeito à importância e necessidade da consoante nasal intrusiva para a implementação da nasalidade em português. O facto desta nem sempre estar presente - sobretudo em posição final, em que o velo permanece aberto e não há movimentos articulatórios orais a assinalar [50], mas também em contexto [(C)Ṽ.CV] - e da sua duração estar dependente de condicionalismos vários (vozeamento da consoante seguinte, ponto de articulação de C2, taxa de elocução) parece sugerir que o N não é uma condição fundamental à produção da vogal nasal.

Do ponto de vista perceptivo, contudo, a presença do chamado "*nasal tail*" parece ser importante, senão fundamental, com os ouvintes portugueses a preferirem invariavelmente padrões de sincronização que favorecem a sua emergência (cf. Stevens *et alii*, 1987; Teixeira, 2000). O mesmo

[49] A presença e duração da consoante nasal intrusiva poderá estar ainda dependente de factores dialectais e/ou individuais, como sugere Sousa (1994), o que apenas poderá ser comprovado através de estudos adicionais.

[50] Os movimentos articulatórios do dorso da língua não foram quantificados para este estudo, em virtude do posicionamento pouco favorável dos próprios sensores - um na ponta da língua e um outro no médio-dorso, para não interferir com o sensor do velo - e também devido à dificuldade em detectar os próprios gestos, como já referimos.

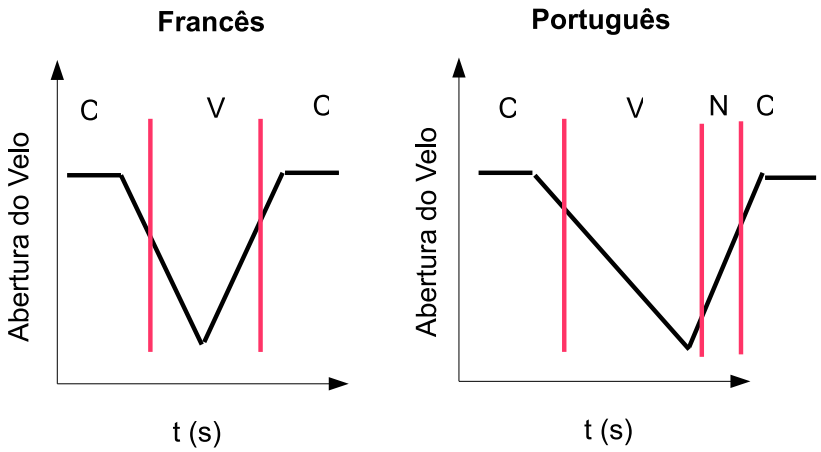


Figura 5.30: Esquema idealizado do movimento do velo durante a produção das vogais nasais do francês (à esquerda), com base na proposta de Amelot (2004), e do português (à direita), segundo o estudo por nós realizado.

não se verificará em relação aos ouvintes franceses, que, segundo os resultados de estudos perceptivos anteriores (Stevens *et alii*, 1987), são praticamente indiferentes à presença do murmúrio nasal. Esta diferença - corroborada pela análise de exemplos acústicos e articulatórios do francês - aponta para estratégias diferenciadas na organização temporal dos eventos implicados na nasalização das vogais do francês (pelo menos na parte norte do país [51]) e do português. Com efeito, segundo os dados articulatórios disponíveis - nomeadamente os adquiridos através de fibroscopia (Amelot, 2004) -, no francês, o ponto de máxima abertura do velo é atingido sensivelmente a meio da vogal nasal e o gesto de fecho está praticamente completo quando o *target* oral acontece, pelo que não é previsível a presença de qualquer murmúrio nasal. Pelo contrário, no português, não obstante alguma variabilidade, o velo atinge a máxima amplitude já no final da vogal nasal (cf. Lovatto *et alii*, 2007) e o gesto de fecho do velo é posterior ao gesto oral, habilitando a emergência de uma consoante nasal intrusiva relativamente longa (vd. figura 5.30).

Tendo em conta a natureza dinâmica da nasalidade vocálica em português e o seu carácter inegavelmente linguístico (i.e. não mecânico) - revelada nesta e noutras análises fonéticas anteriores - a representação da FA é bem mais satisfatória e eficaz do que as descrições mais tradicionais. Através da alteração dos parâmetros dinâmicos do modelo - nomeadamente duração, *stiffness* e sobreposição entre gestos consecutivos - ela é capaz de expressar uma realidade complexa e, nas palavras de Albano (1999), "gradiente". Se a vogal nasal apresenta uma configuração inicial próxima da vogal oral e o murmúrio nasal desaparece em alguns casos e noutros é perfeitamente detectável, isso é representado através de uma maior ou menor sobreposição entre os gestos.

[51] No norte do país as vogais nasais são articuladas sem elementos consonânticos, enquanto a pronúncia do sul se caracteriza pela nasalização apenas parcial das vogais nasais, que são seguidas de um apêndice nasal (Mareüil *et alii*, 2007) claramente audível.

Capítulo 6

Conclusões

Hic cursus fuit.

Sérvio

Science is always wrong, it never solves a problem without creating ten more.

George Bernard Shaw

Este último capítulo é dedicado à apresentação global dos principais resultados e conclusões permitidas pelo estudo desenvolvido.

Tendo em mente os objectivos inicialmente propostos, estas observações finais retomarão apenas os aspectos que consideramos mais relevantes, alguns deles já referenciados nas discussões empreendidas no final de cada capítulo. Como a grande maioria das investigações desta natureza, também esta tem limitações, que procuraremos igualmente realçar, ao longo deste capítulo. Apesar dessas limitações, este é, até onde nos é dado conhecer, o primeiro trabalho centrado na descrição dinâmica do PE. Fica, pois, aberto o caminho para que mais investigadores se interessem por conhecer e utilizar modelos dinâmicos como a FA e para que seja possível - através do recurso a uma metodologia instrumental, similar à utilizada no tratamento das vogais nasais - analisar outros fenómenos linguísticos do PE à luz de teorias que não negligenciam a temporalidade da fala e procuram conciliar os aspectos abstractos com os aspectos fonéticos da linguagem. Uma vez que se encontra apenas no início, as possibilidades de investigação nesta área de estudos são muitas e variadas. É neste sentido que terminamos esta dissertação com a referência a um conjunto de tópicos que merecem ser desenvolvidos em investigações futuras que pretendam dar continuidade ao nosso trabalho.

6.1 Principais resultados e conclusões

Motivados pelo propósito central de contribuir para a construção de um sistema completo de conversão de texto para fala, baseado em síntese articulatória - dando, assim, continuidade ao trabalho de Teixeira (2000) - desenvolvemos um primeiro modelo articulatório para o PE, que visou transformar, automaticamente, o texto de entrada na trajectória dos articuladores. A estratégia seguida passou pela adaptação do sistema TADA ao PE, através da criação de uma base de dados gestual, específica para esta língua.

Tendo em conta os resultados da avaliação perceptiva preliminar, consideramos que o objectivo a que nos propusémos foi, em termos globais, cumprido. Os segmentos sintetizados a partir das configurações gestuais propostas são já minimamente inteligíveis, embora ainda pouco naturais devido a um conjunto de problemas (e.g. prosódia), que vão muito para além da configuração gestual subjacente, e que não está dentro dos objectivos desta tese resolver.

Não obstante a viabilidade de recorrer à FA (e ao TADA) para modelar dinamicamente o léxico do PE, as dificuldades enfrentadas na representação de alguns segmentos da língua (e.g. vibrantes e laterais) evidenciam a necessidade de revisão de alguns parâmetros do modelo. Para além disso, muitas das propostas gestuais aqui apresentadas poderão ser aperfeiçoadas e/ou revistas, se os dados articulatórios entretanto disponibilizados assim o justificarem.

O desenvolvimento do referido modelo dinâmico ditou a realização de um conjunto de tarefas relacionadas com a silabificação automática e a transcrição fonética das palavras de entrada do TADA. Uma vez que funcionam de forma totalmente independente, estes módulos poderão servir outras aplicações como dicionários ou *corpora* electrónicos, *softwares* educativos, etc., para além da natural integração num qualquer sistema TTS.

No respeitante à divisão silábica, os dois algoritmos desenvolvidos fizeram apelo a conhecimentos de ordem fonológica sobre a estrutura da sílaba. O sucesso desta abordagem - sobretudo no caso da aplicação do algoritmo originalmente proposto por Mateus & d'Andrade (2000), cujo desempenho ronda os 99.77% - poderá encontrar explicação na relativa simplicidade da estrutura silábica do PE. Este é também um exemplo em como as propostas fonológicas podem e devem ser implementadas e testadas, o que poderá contribuir para o seu desenvolvimento e/ou modificação.

O sistemas de conversão grafema-fone desenvolvidos basearam-se 1) na combinação de um conjunto de regras com um método de aprendizagem automática e 2) na aplicação de métodos automáticos, nomeadamente o MBL e o TBL.

Como seria de esperar - tendo em conta as experiências anteriores de conversão grafema-fone do PE e a própria natureza do sistema ortográfico, essencialmente de base fonológica - a primeira

solução fundada na aplicação de regras linguísticas, veio a revelar-se bastante eficaz, apesar de continuar a apresentar margem para desenvolvimentos, seja através da definição de novas regras, seja através do aumento do tamanho do *corpus* para treino de TBL.

A entrada em vigor do Novo Acordo Ortográfico terá como corolário natural a revisão de algumas das regras propostas, como, por exemplo, as que dizem respeito à leitura das chamadas consoantes mudas <p> e <c>. O impacto no sistema das várias mudanças impostas pelo Acordo está, contudo, ainda em avaliação, sendo já certo que a supressão e/ ou alteração de algumas regras irá certamente dar origem a erros de transcrição.

Apesar dos resultados claramente positivos da avaliação do sistema baseado em regras linguísticas, não quisémos deixar de explorar outras metodologias, menos usadas no processamento do português, mas com créditos firmados na conversão grafema-fone de outras línguas, como, por exemplo, o inglês. Uma das principais conclusões a retirar da investigação desenvolvida nesta área diz respeito às vantagens da integração de informação silábica nos sistemas G2P. De acordo com os resultados obtidos, este tipo de informação tem um impacto determinante no desempenho global dos sistemas automáticos, a par de outras condicionantes como a dimensão do *corpus* de treino. Como já tivemos ocasião de frisar por diversas vezes, o tamanho do *corpus* é ainda um grande problema para quem, como nós, pretenda desenvolver (e avaliar) sistemas de conversão grafema-fone, baseados em métodos *data-driven*. Não obstante esta limitação - que esperamos que venha a ser colmatada em breve - somos de opinião que as técnicas automáticas têm grandes potencialidades, podendo mesmo vir a superar os desempenhos actualmente atingidos através da aplicação de regras manuais.

Com o intuito de ilustrar as potencialidades de uma representação dinâmica dos fenómenos linguísticos, encerrámos esta dissertação com uma nova proposta de análise da nasalidade vocálica, à luz dos princípios da FA.

Os dados articulatórios obtidos através de EMA evidenciaram uma realidade complexa e gradativa - já salientada por outros autores, referidos no enquadramento teórico do capítulo 5 - com o gesto de abertura nasal a começar depois do gesto vocálico (pelo menos em contexto [CṼ.CV]) e o gesto de fecho do velo a sobrepor-se, na maioria das vezes, ao gesto consonantal seguinte, o que a par com um atraso na abertura do glote, dá origem a uma consoante nasal intrusiva. As várias medidas articulatórias efectuadas demonstraram que a presença e duração desta consoante é influenciada por vários factores, nomeadamente a taxa de elocução. O aumento desta última tem como corolário natural um encurtamento significativo dos gestos articulatórios e uma diminuição da distância entre estes. Além disso, verificámos distintos graus de sobreposição entre gestos nasais e orais de acordo com o contexto segmental e a posição da vogal na palavra. Admitimos ainda a influência de factores individuais (Sousa, 1994), embora não tenha sido possível averiguar esta relação.

Contrariamente aos modelos fonológicos tradicionais, a FA permite lidar intuitivamente com esta variabilidade, uma vez que ela é inerente ao próprio paradigma de investigação, e a dicotomia

clássica entre fonética e fonologia (discreto e contínuo) não existe. Segundo este modelo, não há aqui inserção de qualquer segmento [1] e a presença (e duração) da consoante nasal depende, antes, da maior ou menor sobreposição entre os gestos nasal, consonantal e glotal.

Em consequência de condicionalismos vários - de que avultam a necessidade de responder a outros objectivos, que não os definidos para esta tese, e a impossibilidade de gravar os dados em Portugal - o estudo EMA do comportamento do velo circunscreveu-se a um único sujeito e a um número limitado de dados. Embora tenhamos tentado compensar estas limitações com um estudo perceptivo, as conclusões retiradas devem, contudo, continuar a ser encaradas como preliminares, observações de partida para investigações futuras. Enquanto realidade complexa, o fenómeno da nasalidade deve ser abordado do ponto de vista da produção (análises acústicas e articulatórias) e da percepção (testes perceptivos) (Hajek, 2008). Ambos os níveis têm também um papel crítico no que concerne ao desenvolvimento da síntese de fala.

No nosso caso particular, os dados perceptivos validaram os resultados articulatórios relativos à coordenação entre gestos, mas não os respeitantes à altura do velo. Com efeito, ou ouvintes portugueses testados preferiram um padrão de sincronização temporal entre os gestos, que favorece a nasalização da parte final da vogal e a emergência de uma consoante nasal relativamente longa entre a vogal e a consoante oclusiva seguinte, mas mostraram-se insensíveis às variações na altura do velo. Estes resultados foram determinantes na selecção dos parâmetros dinâmicos a considerar, tendo em vista a inserção das vogais nasais no TADA e a sua síntese através do SAPWindows.

No que respeita ao gesto do corpo da língua, verificámos que as diferenças entre vogais orais e nasais são muito pequenas. Também ao nível dos lábios, não existem ajustes articulatórios dignos de nota. Estes dados permitem-nos concluir que, ao contrário do que acontece no francês, a nasalidade vocálica não envolve, no PE, ajustes articulatórios significativos, para além do movimento de abertura do velo.

[1] Lembramos que os gestos, na concepção da FA, têm extensão temporal e espacial e, portanto, estão sempre presentes. A variação linguística resulta de: 1) diminuição da magnitude dos gestos e 2) aumento (redução) da sobreposição entre gestos.

6.2 Desenvolvimentos futuros

Para terminar, apontam-se algumas sugestões para aprofundamento e desenvolvimento do trabalho por nós efectuado.

Obtenção de dados

Um dos maiores problemas enfrentados ao longo da realização desta dissertação prendeu-se, como foi já sobejamente referido, com a escassez de dados articulatórios para os sons do PE. A recolha e análise de dados de produção é absolutamente essencial para a validação e evolução das propostas de representação gestual dos vários segmentos do PE. O problema é particularmente premente em relação a algumas classes de sons, como as laterais alveolares ou as vibrantes, cujas configurações gestuais carecem reconhecidamente de ser aferidas e, se necessário, revistas.

No sentido de esclarecer a questão da manifestação fonética da lateral no PE, procedemos já à aquisição de alguns dados, usando EMA, e planeamos para breve a recolha de mais informação, através de ressonância magnética. Para além de um conhecimento mais assertivo dos mecanismos de produção desta consoante, em diferentes contextos, está também em causa a verificação das hipóteses formuladas acerca do modo como os gestos que compõem a lateral se organizam entre si, em função da filiação silábica. Os estudos articulatórios poderão ser complementados com outros métodos de observação indirecta do comportamento dos articuladores como a análise acústica. Face às possibilidades desta última - em termos portabilidade e quantidade de informantes que é possível gravar - esta será talvez a via mais adequada para averiguar a hipótese da variação dialectal do /l/.

O caso das vibrantes é ainda mais problemático, já que, até onde nos foi dado conhecer, não existem estudos, articulatórios ou acústicos sobre esta classe de sons. Este vazio de informação - a par com outros problemas relacionados com as limitações do próprio TADA - condicionou fortemente o nosso objectivo de caracterização gestual das vibrantes. As imagens de RM de que dispomos actualmente são em tempo real, não sendo a sua resolução temporal a mais desejável. Para além disso, a sua análise encontra-se condicionada ao desenvolvimento de novas técnicas de segmentação, que permitam a obtenção de informação útil, um problema que esperamos que venha a ser resolvido a muito curto prazo. Também a descrição fonética das vibrantes teria muito a ganhar com um estudo acústico, que pudesse sustentar, ainda que de forma indirecta, as observações impressionistas dos linguistas portugueses.

Desenvolvimento de outros sistemas de silabificação automática

Os módulos de silabificação automática implementados e testados baseiam-se na aplicação de um conjunto de regras. Esta é considerada a melhor abordagem para lidar com línguas de baixa complexidade silábica (e.g. português, espanhol, italiano, grego), mas estudos recentes (Marchand *et alii*,

2007) indicam que, em termos de desempenho, os métodos automáticos são capazes de ultrapassar os algoritmos de base linguística, mesmo em línguas de estrutura silábica relativamente simples e regular (Adsett & Marchand, no prelo). Tendo em mente os resultados obtidos - para o inglês (Marchand *et alii*, 2007), mas sobretudo para o italiano (Adsett & Marchand, no prelo) - valeria a pena averiguar até que ponto esta realidade se confirma para o português. A viabilidade deste tipo de estudo está, contudo, condicionada à existência de um *corpus* com informação silábica e de dimensão considerável, que actualmente não está disponível para o português. Isto significa que não só não podemos aplicar métodos automáticos à silabificação automática do PE - nem a outras tarefas de processamento de linguagem natural - como não podemos comparar o desempenho dos nossos sistemas com outros já existentes, uma vez que não existe um *corpus* de teste comum.

Realização de mais experiências sobre a nasalidade

Devido a motivos técnicos, o estudo articulatório sobre a nasalidade vocálica, apresentado no capítulo 5, remeteu-se a um único informante. Aferida a metodologia de análise - sobretudo, de detecção automática dos movimentos articulatórios - estamos agora em condições de analisar os dados relativos ao segundo informante e, assim, confirmar alguns dos resultados já obtidos.

Para além disso, grande parte do *corpus* adquirido não foi ainda alvo de análise pormenorizada. À nossa disposição, está ainda um grande manancial de dados EMA, que carecem de tratamento, e hão-de viabilizar outras pesquisas - nomeadamente sobre a nasalidade consonântica e a nasalização de vogais orais por efeito do contexto - e estudos comparativos entre o francês e o português. Um outro contexto, constante do *corpus*, que valerá a pena explorar, diz respeito às vogais nasais precedidas de consoante nasal.

As potencialidades do SAPWindows enquanto ferramenta de investigação e de criação de estímulos "in specific and controlled ways, in order to test theoretical hypotheses" (Keller & Keller, 2000a, p.135) haviam já sido demonstradas por Teixeira (2000). Da integração do sintetizador SAPWindows com o sistema TADA nascem novas possibilidades, no que às experiências de percepção diz respeito, nomeadamente no campo da nasalidade. O teste de percepção desenvolvido no último capítulo desta dissertação veio levantar novas questões acerca da sensibilidade perceptiva dos ouvintes portugueses às variações na altura do velo. O papel relativamente secundário deste último parâmetro em relação à dinâmica temporal dos gestos, sugerido pelos resultados preliminares, precisa de ser comprovado por via da realização de um novo teste perceptivo. Este deverá oferecer a possibilidade de manipular apenas a amplitude do velo, em, pelo menos, três ou quatro níveis, com valores superiores aos considerados no teste anterior, sem quaisquer modificações ao nível da coordenação entre os gestos.

Uma outra interrogação que subsiste diz respeito à dita fase oral das vogais nasais. É nossa intenção averiguar, também através de um teste de percepção, mas desta feita com estímulos naturais e, usando vários contextos, se, tal como no francês (Amelot, 2004; Montagu, 2007), os ouvintes do PE

identificam ou não uma fase oral e em que momento da articulação a vogal é percebida pelos mesmos como nasal.

Bibliografia

Observação: *A formatação da bibliografia obedece, no essencial, às recomendações de um grupo de editores de revistas da área da Linguística (aprovadas em Janeiro de 2007) e disponíveis online no documento "Unified style sheet for linguistics".*

Acero, Alex. 1995. The role of phoneticians in speech technologies. In G. Bloothooft, V. Hazan, D. Huber & J. Llisterri (eds.), *European Studies in Phonetics and Speech Communication*, 170–175. Utrecht: OTS Publications.

Adsett, Connie & Marchand, Yannick. no prelo. Syllabification rules versus data-driven methods in a low syllabic complexity language? The case of Italian. *Computer Speech and Language* .

Al-Bamerni, Ameen. 1983. *Oral, Velic and Laryngeal Coarticulation across Languages*. PhD, Oxford University.

Albano, Eleonora. 1999. O português brasileiro e as controvérsias da fonética atual: pelo aperfeiçoamento da fonologia articulatória. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada* 15, 23–51.

Albano, Eleonora. 2001. *O Gesto e suas Bordas: Esboço de Fonologia Acústico-Articulatória do Português Brasileiro*. Campinas: Mercado de Letras.

Albano, Eleonora & Aquino, Patrícia. 1997. Linguistic criteria for building and recording units for concatenative speech synthesis in Brazilian Portuguese. In *European Conference on Speech Communication and Technology (Eurospeech)*, 725–728. Rhodes, Greece.

Albano, Eleonora & Moreira, Agnaldo. 1996. Archisegment-based letter-to-phone conversion for concatenative speech synthesis in Portuguese. In *International Conference on Spoken Language Processing (ICSLP)*, 1708–1711. Philadelphia.

Allen, Donald & Strong, William. 1985. A model for the synthesis of natural sounding vowels. *Journal of the Acoustical Society of America* 78(1), 58–69.

Allen, Jonathan, Hunnicutt, Sharon, Klatt, Dennis, Armstrong, Robert & Pisoni, David. 1987. *From Text to Speech: The MITalk System*. Cambridge: Cambridge University Press.

Almeida, António. 1976. The Portuguese nasal vowels: Phonetics and phonemics. In J. Schmidt-Radefelt (ed.), *Readings in Portuguese Linguistics*, 348–396. New York: North Holland.

Almeida, João & Simões, Alberto. 2001. Text-to-speech - "A rewriting system approach". *Procesamiento del Lenguaje Natural* 27, 247–255.

Almeida, João, Simões, Alberto & Rocha, Paulo. 2003. Lingua-PT-PLN-0.06. URL `http://cpan.org`.

Amelot, Angélique. 2004. *Étude Aérodynamique, Fibroscopique, Acoustique et Perceptive des Voyelles Nasales du Français*. Thèse de doctorat, Université Paris III- Sorbonne Nouvelle.

Amelot, Angélique & Rossato, Solange. 2006. Velar movements for the feature [+nasal] for two French speakers. In *International Seminar on Speech Production*, 459–467. Ubatuba, Brasil.

Amelot, Angélique & Rossato, Solange. 2007. Velar movements for two French speakers. In *International Congress of Phonetic Sciences (ICPhS)*. Saarbrücken, Germany.

Andrade, Amália. 1996. Reflexões sobre o "e mudo" em português europeu. In Inês Duarte & Isabel Leiria (eds.), *Congresso Internacional sobre o Português*, 303–344. Lisboa: Colibri/APL.

Andrade, Amália. 1998. Variação fonética de /l/ em ataque silábico em português europeu. In *Encontro da Associação Portuguesa de Linguística (APL)*, 55–76. Lisboa.

Andrade, Amália. 1999. On /l/ velarization in European Portuguese. In *International Congress of Phonetic Sciences (ICPhS)*, 543–546. San Francisco.

Andrade, Amália & Viana, Maria do Céu. 1996. Fonética. In I. H. Faria, E. R. Pedro, I. Duarte & C. A. Gouveia (eds.), *Introdução à Linguística Geral e Portuguesa*, 115–167. Lisboa: Caminho.

Bagemihl, Bruce. 1995. Language games and related areas. In J. A. Goldsmith (ed.), *The Handbook of Phonological Theory*, 697–712. Cambridge/Oxford: Blackwell.

Bagshaw, Paul. 1998. Unsupervised training of phone duration and energy models for text-to-speech synthesis. In *International Conference on Spoken Language Processing (ICSLP)*, 17–20. Sydney, Australia.

Bailly, Gérard, Benoît, Chris & Sawallis, Thomas (eds.). 1992. *Talking Machines - Theories, Models, and Designs*. Amsterdam: North-Holland, Elsevier Science Publishers B. V.

Bailly, Gérard, Campbell, Nick & Möbius, Bernd. 2003. ISCA special session: Hot Topics in Speech Synthesis. In *European Conference on Speech Communication and Technology (Eurospeech)*, 37–40. Geneva, Switzerland.

Bailly, Gérard, Revéret, Lionel, Baciu, Monica, Segebarth, Christoph & Savariaux, Christophe. 2002. Three-dimensional articulatory modeling of tongue, lips and face, based on MRI and video images. *Journal of Phonetics* 30(3), 533–553.

Baken, Ronald & Orlikoff, Robert. 1999. *Clinical Measurement of Speech and Voice*. Singular, 2a. edn..

Barbeiro, L. 1986. *Estrutura Silábica do Português. O Papel da Sílaba na Análise dos Processos Fonológicos e Fonéticos*. Tese de mestrado, Faculdade de Letras da Universidade de Lisboa.

Barbosa, Filipe, Ferrari, Lilian & Resende, Fernando. 2003a. A methodology to analyse homographs for a Brazilian Portuguese TTS system. In Nuno Mamede, Jorge Baptista, Isabel Trancoso & Maria das Graças Volpe Nunes (eds.), *Computacional Processing of the Portuguese Language (PROPOR)*, 57–61. Springer.

Barbosa, Filipe, Pinto, Guilherme, Resende, Fernando, Gonçalves, Carlos, Monserrat, Ruth & Rosa, Maria Carlota. 2003b. Grapheme- phone transcription algorithm for a Brazilian Portuguese TTS. In Nuno Mamede, Jorge Baptista, Isabel Trancoso & Maria das Graças Volpe Nunes (eds.), *Computacional Processing of the Portuguese Language (PROPOR)*, 23–30. Springer.

Barbosa, Filipe, Rosa, Maria Carlota, Gonçalves, Carlos & Resende, Fernando. 2003c. Algoritmo para leitura de siglas em um sintetizador de voz. In *Simpósio Brasileiro de Telecomunicações*, 672–675. Rio de Janeiro: IME e PUC-RJ.

Barbosa, Jorge Morais. 1961. Les voyelles nasales portugaises: Interpretation phonologique. In Antti Sovijärvi & Pentti Aalto (eds.), *International Congress of Phonetic Sciences (ICPhS)*, 691–709. The Hague: Mouton & Co.

Barbosa, Jorge Morais. 1965. *Études de Phonologie Portugaise*. Lisboa: Junta de Investigações do Ultramar, Centro de Estudos Políticos e Sociais.

Barbosa, Jorge Morais. 1994a. *Introdução ao Estudo da Fonologia e Morfologia do Português*. Coimbra: Livraria Almedina.

Barbosa, Plínio. 1994b. *Caractérisation et Génération Automatique de la Structuration Rythmique du Français*. Thèse de doctorat, Institut de la Communication Parlée, ICP/INPG, França.

Barbosa, Plínio. 1997. A model of segment (and pause) duration generation for Brazilian Portuguese text-to-speech synthesis. In *European Conference on Speech Communication and Technology (Eurospeech)*, 2655–2658. Rhodes, Grécia.

Barbosa, Plínio. 2001. Máquinas falantes como instrumentos lingüísticos: por um humanismo éclairé. *Línguas e Instrumentos Lingüísticos* 8, 51–99.

Barbosa, Plínio. 2005. On the defense of von Kempelen as the predecessor of experimental phonetics and speech synthesis research. In Eduardo J. Guimarães & Diana P. de Barros (eds.), *History of Linguistics. Selected Papers from the Ninth International Conference on the History of Language Sciences in Brazil*. Amsterdam: John Benjamins.

Barbosa, Plínio. 2006. *Incursões em torno do Ritmo da Fala*. S. Paulo: Pontes Editores.

Barbosa, Plínio & Albano, Eleonora. 2004. Brazilian Portuguese. Illustrations of the IPA. *Journal of the International Phonetic Association* 34(2), 227–232.

Barker, Lecia J. 2003. Computer-assisted vocabulary acquisition: the CSLU vocabulary tutor in oral-deaf education. *Journal of Deaf Studies and Deaf Education* 8(2), 187–198.

Barros, João de. 1540. *Grammatica da Língua Portuguesa*. Faculdade de Letras da Universidade de Lisboa.

Barros, Maria João & Weiss, Christian. 2006. Maximum entropy motivated grapheme-to-phoneme, stress and syllable boundary prediction for Portuguese text-to-speech. In *Jornadas en Tecnologia del Habla*, 177–182. Zaragoza, España.

Barroso, Henrique. 1999. *Forma e Substância da Expressão da Língua Portuguesa*. Coimbra: Almedina.

Barry, William. 1997. Another r-tickle. *Journal of the International Phonetic Association* 27(1-2), 35–45.

Bartkova, Katarina, Haffner, P. & Larreur, Danielle. 1993. Intensity prediction for speech synthesis in French. In *ESCA Workshop on Prosody*, 280–283. Lund, Sweden.

Bartkova, Katarina & Sorin, Christel. 1987. A model of segmental duration for speech synthesis in French. *Speech Communication* 6, 245–260.

Bartlett, Susan, Kondrak, Grzegorz & Cherry, Colin. 2008. Automatic syllabification with structured SVMs for letter-to-phoneme conversion. In *Association for Computational Linguistics (ACL)*, 568–576. Columbus, USA.

Basset, Patricia, Amelot, Angélique & Crevier-Buchman, Lise. 2006. Etude multiparamétrique des consonnes nasales du français: Prise de données simultanées aérodynamiques et fibroscopiques. *Revue Parole* 39-40, 113–135.

Basset, Patricia, Amelot, Angélique, Vaissière, Jacqueline, & Roubeau, Bernard. 2001. Nasal flow in spontaneous speech. *Journal of the Internacional Phonetic Association* 31, 87–99.

Beaulieu, Kathy. 2001. La structure interne de la syllable: Ce qu'en disent les lapsus. In *Colloque des Etudiants en Sciences du Langage*. Université de Québec, Montréal.

Beddor, Patrice Speeter. 1993. The perception of nasal vowels. In Marie K. Huffman & Rena A. Krakow (eds.), *Phonetics and Phonology, Volume 5: Nasals, Nasalization, and the Velum*, 171–196. Academic Press.

Beddor, Patrice Speeter. 2007. Nasals and nasalization: The relations between segmental and coarticulatory timing. In *International Congress of Phonetic Sciences (ICPhS)*. Saarbrücken, Germany.

Bell-Berti, Fredericka. 1993. Understanding velic motor control: Studies of segmental context. In Marie K. Huffman & Rena A. Krakow (eds.), *Phonetics and Phonology, Volume 5: Nasals, Nasalization, and the Velum*, 63–85. Academic Press.

Bell-Berti, Fredericka, Baer, Thomas, Harris, Katherine & Niimi, Seiji. 1979. Coarticulatory effects of vowel quality on velar function. *Phonetica* 36(3), 187–193.

Benguerel, André-Pierre, Hirose, Hajime, Sawashima, Masayuki & Ushijima, Tatsujiro. 1977. Velar coarticulation in French: fiberscopic study. *Journal of Phonetics* 5, 149–158.

Benoît, Christian & Pols, Louis. 1995. Speech synthesis: Present and future. In G. Bloothooft, V. Hazan, D. Huber & J. Llisterri (eds.), *European Studies in Phonetic and Speech Communication*, 119–123. Utrecht, Netherlands: OTS Publications.

Beringer, Nicole. 2004. *How to Integrate Phonetic and Linguistic Knowledge in a Text-to-Phoneme Conversion Task: A Syllabic TPC Tool for French*. Rel. Tec. IDSIA-18-04, IDSIA /USI-SUPSI, Dalle Molle Institute for Artificial Intelligence.

Bianchi, Olivier. 2005. *Vox Rediuiua: Synthèse de la Parole et Métrique Latine*. Doctoral thesis, Faculté des Lettres, University of Lausanne.

Birkholz, Peter. 2007a. Articulatory synthesis of singing. In *Annual Conference of the International Speech Communication Association (Interspeech)*. Antwerp, Belgium.

Birkholz, Peter. 2007b. Vocal tract lab. `http://www.vocaltractlab.de` (2 Novembro, 2007).

Birkholz, Peter, Jackèl, Dietmar & Kröger, Bernd. 2006. Construction and control of a three-dimensional vocal tract model. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 873–876. Toulouse, France.

Birkholz, Peter & Kröger, Bernd. 2007. Simulation of vocal tract growth for articulatory speech synthesis. In *International Congress of Phonetic Sciences (ICPhS)*, 377–380. Saarbrücken, Germany.

Birkholz, Peter, Steiner, Ingmar & Breuer, Stefan. 2007. Control concepts for articulatory speech synthesis. In *ISCA Workshop on Speech Synthesis*, 5–10. Bonn, Germany.

Bisol, Leda. 1989. O ditongo na perspectiva da fonologia atual. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada* 10(2), 185–224.

Bisol, Leda. 1994. Ditongos derivados. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada* 5(2), 123–140.

Bisol, Leda (ed.). 2001. *Introdução a Estudos de Fonologia do Português Brasileiro*. Porto Alegre: EDIPUCRS, 3a. edn..

Björk, Lars. 1961. Velopharyngeal function in connected speech. *Acta Radiologica* Supplement 202, 1–94.

Black, Alan & Hunt, Andrew. 1996. Generating F0 contours from toBI labels using linear regression. In *International Conference on Spoken Language Processing (ICSLP)*, vol. 3, 1385–1388. Philadelphia.

Bladon, Anthony & Al-Bamerni, Ameen. 1976. Coarticulatory resistance in English /l/. *Journal of Phonetics* 4, 137–150.

Blecua, Beatriz. 1999. Caracteristicas acusticas de la vibrante multiple del espanol en habla espontanea. In *Congress of Experimental Phonetics*, 119–126. Tarragona.

Blecua, Beatriz & Acín, Vanessa. 1995. Propuesta de un modelo de intensidad vocálica del castellano y el catalán aplicable a un sistema de conversión de texto a habla. *Procesamiento del lenguaje natural* 17, 257–271.

Blevins, Juliette. 1995. The syllable in phonological theory. In J. A. Goldsmith (ed.), *The Handbook of Phonological Theory*, 206–244. Cambridge/Oxford: Blackwell.

Boersma, Paul. 1995. Interaction between glottal and vocal-tract aerodynamics in a comprehensive model of the speech apparatus. In *International Congress of Phonetic Sciences (ICPhS)*, vol. 2, 430–433. Stockholm.

Boersma, Paul & Weenink, David (2009). 2009. Praat: doing phonetics by computer (version 5.1) [computer program]. `http://www.praat.org/`.

Bohlenius, Jonas. 2005. *Vox ex Machina*. Master thesis, Göteborg University.

Boléo, Manuel de Paiva & Silva, Maria Helena. 1962. O mapa dos dialectos e falares de Portugal continental. In *Congresso Internacional de Linguística Românica*, 85–115.

Bosseler, Alexis & Massaro, Dominic. 2003. Development and evaluation of a computer-animated tutor for vocabulary and language learning in children with autism. *Journal of Autism and Developmental Disorders* 33(6), 653–672.

Bouma, Gosse. 2000. A finite state and data-oriented method for grapheme-to-phoneme conversion. In *Conference on North American chapter of the Association for Computational Linguistics*, 303–310. Seattle, Washington: Morgan Kaufmann Publishers Inc.

Bouma, Gosse. 2002. Finite state methods for hyphenation. *Journal of Natural Language Engineering* 1(1), 1–16. Special Issue on Finite State Methods in NLP.

Brackhane, Fabian & Trouvain, Juergen. 2008. What makes “mama” and “papa” acceptable? - Experiments with a replica of von Kempelen’s speaking machine. In *International Speech Production Seminar*, 329–332. Strasbourg, France.

Bradley, Travis. 2001. *The Phonetics and Phonology of Rhotic Duration Contrast and Neutralization*. PhD dissertation, The Pennsylvania State University.

Braga, Daniela. 2006. Grapheme-to-phone transcription algorithm for text-to-speech sytems in European Portuguese. *POLISSEMA - Revista de Letras do ISCAP (Instituto Superior de Contabilidade e Administração do Porto)* 6.

Braga, Daniela. 2007. Desambiguação de homógrafos para sistemas de conversão texto-fala em português. *Diacrítica* 21(1), 25–50.

Braga, Daniela. 2008. *Algoritmos de Processamento da Linguagem Natural para sistemas de conversão texto-fala em português*. Tese de doutoramento, Facultade de Filoloxía da Universidade da Coruña.

Braga, Daniela & Coelho, Luís. 2006. Letter-to-sound conversion for Galician TTS systems. In *Jornadas en Tecnologías del Habla*. Zaragoza, Espanha.

Braga, Daniela, Coelho, Luís & Resende, Fernando. 2006. A rule-based grapheme-phone converter for TTS Systems in European Portuguese. In *International Telecommunications Symposium*, 976–981. Fortaleza, Brasil.

Braga, Daniela, Coelho, Luís & Resende, Fernando. 2007. Homograph ambiguity resolution in front-end design for Portuguese TTS systems. In *Annual Conference of the International Speech Communication Association (Interspeech)*, 1761–1764. Antwerp.

Breen, Gavan & Pensalfini, Rob. 1999. Arrernte: a language with no syllable onsets. *Linguistic Inquiry* 30(1), 1–25.

Brill, Eric. 1995. Transformation-based error-driven learning and natural language processing: a case study in part-of-speech tagging. *Computacional Linguistics* 21, 543–566.

Broecke, M. van den. 1983. Wolfgang von Kempelen's Speaking Machine as a Performer. In Marcet van den Broecke (ed.), *Sound Structures*. Dordrecht: Foris Publications.

Browman, Catherine. 1994. Lip aperture and consonant releases. In Patricia Keating (ed.), *Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form*, 331–353. Cambridge: Cambridge University Press.

Browman, Catherine & Goldstein, Louis. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3, 219–252.

Browman, Catherine & Goldstein, Louis. 1988. Some notes on syllable structure in articulatory phonology. *Phonetica* 45, 140–155.

Browman, Catherine & Goldstein, Louis. 1989. Articulatory gestures as phonological units. *Phonology* 6, 201–251.

Browman, Catherine & Goldstein, Louis. 1990a. Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics* 18, 299–320.

Browman, Catherine & Goldstein, Louis. 1990b. Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*, 341–376. Cambridge: Cambridge University Press.

Browman, Catherine & Goldstein, Louis. 1992. Articulatory phonology: An overview. *Phonetica* 49, 155–180.

Browman, Catherine & Goldstein, Louis. 1995a. Dynamics and articulatory phonology. In Robert F. Port & Tim van Gelder (eds.), *Mind as Motion: Explorations in the Dynamics of Cognition*, 175–193. Cambridge, MA, USA: Massachusetts Institute of Technology.

Browman, Catherine & Goldstein, Louis. 1995b. Gestural syllable position effects in American English. In Fredericka Bell-Berti & Lawrence J. Raphael (eds.), *Producing Speech: Contemporary Issues, for Katherine Safford Harris*, 19–33. New York: American Institute of Physics (AIP) Press.

Browman, Catherine & Goldstein, Louis. 2000. Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée* 5, 25–34.

Browman, Catherine, Goldstein, Louis, Nam, Hosung, Rubin, Philip, Proctor, Michael & Saltzman, Elliot. 2001-2006. *TADA (TAsk Dynamics Application) Manual*. Haskins Laboratories.

Brown, Alan. 1991. A review of the tip-of-the-tongue experience. *Phychological Bulletin* 109, 204–223.

Burke, Deborah, MacKay, Donald, Worthley, Joanna & Wade, E. 1991. On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language* 30, 237–246.

Busà, Maria Grazia. 2003. Vowel nasalization and nasal loss in Italian. In *International Congress of Phonetic Sciences (ICPhS)*, 711–714. Barcelona.

Byrd, Dani. 1994. *Articulatory Timing in English Consonant Sequences*. PhD dissertation, UCLA.

Byrd, Dani. 1995. C-centers revisited. *Phonetica* 285–306.

Byrd, Dani & Saltzman, Elliot. 1998. Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics* 26(2), 173–199.

Byrd, Dani & Saltzman, Elliot. 2003. The elastic phrase: dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31, 149–180.

Byrd, Dani, Tobin, Stephen, Bresch, Erik & Narayanan, Shrikanth. no prelo. Timing effects of syllable structure and stress on nasals: A real-time MRI examination. *Journal of Phonetics* .

Cagliari, Luiz. 1977. *An Experimental Study of Nasality with Particular Reference to Brazilian Portuguese*. PhD thesis, University of Edinburgh.

Campbell, Nick. 1992a. *Multi-Level Timing in Speech*. PhD thesis, University of Sussex.

Campbell, Nick. 1992b. Syllable-based segmental duration. In Bailly *et alii* (1992), 211–224.

Campbell, Nick & Isard, Steve. 1991. Segmental duration in syllable frames. *Journal of Phonetics* 19, 37–47.

Carlson, Rolf. 1994. Models of speech synthesis. In David B. Roe & Jay G. Wilpon (eds.), *Voice Communication between Humans and Machines*. Washington DC: National Academy Press.

Carnegie Mellon University. 1993. The CMU Pronouncing Dictionary (v 0.1). School of Computer Science - Carnegie Mellon University.

Carter, Paul. 2002. *Structured Variation in British English Liquids: the Role of Resonance*. PhD dissertation, University of York.

Carvalho, Pedro, Oliveira, Luís, Trancoso, Isabel & Viana, Maria do Céu. 1998. Concatenative speech synthesis for European Portuguese. In *ESCA/COCOSDA International Workshop on Speech Synthesis*, 159–164. Jenolan Caves, Australia.

Caseiro, Diamantino, Trancoso, Isabel, do Céu Viana, Maria & Barros, Manuela. 2003. A comparative description of GtoP modules for Portuguese and Mirandese using finite-state transducers. In *International Congress of Phonetic Sciences (ICPhS)*, 2605–2608. Barcelona, Spain.

Caseiro, Diamantino, Trancoso, Isabel, Oliveira, Luís & Viana, Maria do Céu. 2002. Grapheme-to-phone using finite-state transducers. In *IEEE Workshop on Speech Synthesis*, 1349–1360.

Castejón, Federico, Escalada, Gregorio, Monzón, Luis, Rodríguez, Miguel & Velasco, P. Sanz. 1994. Un conversor texto-voz para el español. *Comunicaciones de Telefónica I+D* 5(2), 114–131.

Catford, John C. 1977. *Fundamental Problems in Phonetics*. Bloomington: Indiana University Press.

Chen, Marilyn. 1997. Acoustic correlates of English and French nasalized vowels. *Journal of the Acoustical Society of America* 102(4), 2360–2370.

Chitoran, Ioana, Goldstein, Louis & Byrd, Dani. 2002. Gestural overlap and recoverability: Articulatory evidence from Georgian. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory Phonology 7*, 419–447. Mouton de Gruyter.

Chomsky, Noam & Halle, Morris. 1968. *Sound Pattern of English*. New York: Harper & Row, Publishers.

Chung, Hyunsong. 2002. Segment duration in spoken Korean. In *International Conference on Spoken Language Processing (ICSLP)*. Denver, Colorado, USA.

Clements, George & Keyser, Samuel. 1999. From CV phonology: a generative theory of the syllable. In J. A. Goldsmith (ed.), *Phonological Theory. The Essential Readings*, 328–350. Wiley-Blackwell.

Clements, George & Osu, Sylvester. 2003. Ikwere nasal harmony in typological perspective. In Patrick Sauzet & Anne Zribi-Hertz (eds.), *Typologie des Langues d'Afrique et Universaux de la Grammaire*, 70–95. Paris: L'Harmattan.

Clumeck, Harold. 1976. Patterns of soft palate movements in six languages. *Journal of Phonetics* 4, 337–351.

Câmara, J. Mattoso. 1971. *Problemas de Lingüistica Descritiva*. Petrópolis: Editora Vozes Limitada, 5a. edn..

Câmara, J. Mattoso. 1973. *Estrutura da Língua Portuguesa*. Petrópolis: Editora Vozes Limitada, 4a. edn..

Câmara, J. Mattoso. 1977. *Para o Estudo da Fonêmica Portuguesa*. Rio de Janeiro: Padrão - Livraria Editora, 2a. edn..

Cohn, Abigail. 1990. *Phonetic and Phonological Rules of Nasalization*. PhD thesis, UCLA.

Coker, Cecil. 1967. Synthesis by rule from articulatory parameters. In *Conference Speech Communication Processes*, 52–53.

Coker, Cecil, Umeda, Noriko & Browman, Catherine. 1973. Automatic synthesis from ordinary English text. *IEEE Transactions on Audio and Electroacoustics* 21(3), 293–298.

Conejo, José & van Coile, Bert. 1991. Desarrollo de un conversor de texto a voz en español dentro de una arquitectura multilingüe. *Boletín de la Sociedad Española para el Procesamiento del Lenguaje Natural* 11, 221–227.

Cook, Perry, Wang, Ge, Misra, Ananya & Daly, Mark. 2006. Voice Synthesis Technology. Historical Construction Experiment. `http://voce.cs.princeton.edu` (16 Fevereiro, 2006).

Correia, Susana. 2003. A aquisição de consoantes em final de sílaba no português europeu. In *Encontro da Associação Portuguesa de Linguística (APL)*. Lisboa.

Correia, Susana. 2004a. A aquisição da rima em português europeu - ditongos e consoantes em final de sílaba. In *Encontro da Associação Portuguesa de Linguística (APL)*. Lisboa.

Correia, Susana. 2004b. *A Aquisição da Rima em Português Europeu - Ditongos e Consoantes em Final de Sílaba*. Dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa, Lisboa, Portugal.

Córdoba, R., Vallejo, J. A., Montero, J. M., Gutierrez-Arriola, J., López, M. A. & Pardo, J. M. 1999. Automatic modeling of duration in a Spanish text-to-speech system using neural networks. In *European Conference on Speech Communication and Technology (Eurospeech)*, 1619–1622. Budapest, Hungary.

Cristo, Albert Di, Cristo, Philippe Di & Campione, Estelle. 2000. A prosodic model for text-to-speech synthesis in French. In A. Botinis (ed.), *Intonation: Models and Theories*, 321–356. Dordrecht: Kluwer Academic Publishers.

Cruz-Ferreira, Madalena. 1995. European Portuguese. Illustrations of the IPA. *Journal of the International Phonetic Association* 25(2), 90–94.

Cruz-Ferreira, Madalena. 1999a. Illustrations of the IPA. Portuguese (European). In International Phonetic Association (ed.), *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*, 126–130. Cambridge: Cambridge University Press.

Cruz-Ferreira, Madalena. 1999b. Intonation in European Portuguese. In Daniel Hirst & Albert Di Cristo (eds.), *Intonation Systems: A Survey of Twenty Languages*. Cambridge: Cambridge University Press.

Cunha, Celso & Cintra, Lindley. 1997. *Nova Gramática do Português Contemporâneo*. Lisboa: Edições João Sá da Costa, 13a. edn..

Daelemans, Walter & Bosch, Antal Van Den. 1992. Generalization performance of backpropagation learning on a syllabification task. In M. Drossaers & A. Nijholt (eds.), *Twente Workshop on Language Technology 3: Connectionism and Natural Language Processing*, 27–38. The Netherlands: Twente University: Enschede.

Daelemans, Walter & Bosch, Antal Van Den. 1997. Language-independent data-oriented grapheme-to-phoneme conversion. In Jan P.H. van Santen, Richard Sproat, Joseph Olive & Julia Hirschberg (eds.), *Progress in Speech Synthesis*. Springer.

Daelemans, Walter, Zavrel, Jakub, van der Sloot, Ko & van den Bosch, Antal. 2004. *TiMBL: Tilburg Memory Based Learner, version 5.1*. Reference Guide ILK-0402, Tilburg University. ILK Research Group Technical Report Series no. 04-02.

Damper, Robert, Marchand, Yannick, Adamson, M. J. & Gustafson, Kjell. 1998. Comparative evaluation of letter-to-sound conversion techniques for English text-to-speech synthesis. In *ESCA/COCOSDA International Workshop on Speech Synthesis*, 53–58. Jenolan Caves, Australia.

Damper, Robert, Marchand, Yannick, Adamson, M. J. & Gustafson, Kjell. 1999. Evaluating the pronunciation component of text-to-speech systems for English: a performance comparison of different approaches. *Computer Speech and Language* 13(2), 155–176.

d'Andrade, Ernesto. 1977. *Aspects de la Phonologie (Générative) du Portugais*. Lisboa: INIC.

d'Andrade, Ernesto. 1994. *Temas de Fonologia*. Lisboa: Colibri.

d'Andrade, Ernesto & Viana, Maria do Céu. 1985. *CORSO I: um Conversor de Texto Ortográfico em Código Fonético para o Português*. Relatórios do Grupo de Fonética e Fonologia 6, CLUL.

d'Andrade, Ernesto & Viana, Maria do Céu. 1993a. As sobrodas da translineação. In *Encontro sobre o Processamento da Língua Portuguesa Escrita e Falada (EPLP)*, 209–213. Lisboa.

d'Andrade, Ernesto & Viana, Maria do Céu. 1993b. Sinérese, diérese e estrutura silábica. In *Encontro da Associação Portuguesa de Linguística (APL)*. Coimbra.

Davidson, Lisa. 2003. *The Atoms of Phonological Representation: Gestures, Coordination and Perceptual Features in Consonant Cluster Phonotactics*. PhD, John Hopkins University.

Delattre, Pierre. 1968. Divergences entre nasalités vocalique et consonantique en Français. *Word* 24, 64–73.

Delgado-Martins, Maria Raquel. 1994. Relação fonética-fonologia: a propósito do sistema vocálico do português. In *Congresso Internacional sobre o Português*, vol. 1. Lisboa.

Delvaux, Véronique, Metens, Thierry & Soquet, Alain. 2002. Propriétés acoustiques e articulatoires des voyelles nasales du français. In *Journées d'Étude sur la Parole*, 348–352. Nancy.

Demolin, Didier. 1991. L'analyse des segments, de la syllable et des tons dans un jeu de langue Mangbetu. *Langages* 101, 30–50.

Demolin, Didier, Delvaux, Véronique, Metens, Thierry & Soquet, Alain. 2003. Determination of velum opening for French nasal vowels by magnetic resonance imaging. *Journal of Voice* 17(4), 454–467.

Demolin, Didier, Lecuit, Véronique, Metens, Thierry, Nazarian, Bruno & Soquet, Alain. 1998. Magnetic resonance measurements of the velum port opening. In *International Conference on Spoken Language Processing (ICSLP)*. Sydney, Australia.

Diamond, Kevin. 1994. L'oralisation des sigles en anglais. *Linx* 30, 109–132.

Divay, Michel & Vitale, Anthony. 1997. Algorithms for grapheme-phoneme translation for English and French: applications for database searches and speech synthesis. *Computational Linguistics* 23(4), 495 – 523.

Dohalská, Marie, Mejvaldová, Jana & Dubeda, Tomás. 2002. Prosodic parameters of synthetic Czech: Developing rules for duration and intensity. In E. Keller, G. Bailly, A. Monaghan, J. Terken & M. Huckvale (eds.), *Improvements in Speech Synthesis. Cost 258: The Naturalness of Synthetic Speech*, cap. 12, 129–133. Chichester: John Wiley & Sons.

Draper, M. H., Ladefoged, Peter & Whitteridge, D. 1959. Respiratory muscles in speech. *Journal of speech and Hearing Research* 2, 16–27.

Drenska, Margarita. 1986. Existem ditongos crescentes em posição final de palavra em português? In *Encontro da Associação Portuguesa de Linguística (APL)*, 53–65.

Drenska, Margarita. 1988. Análise acústica das vogais nasais em português e búlgaro. In *Encontro da Associação Portuguesa de Linguística (APL)*, 139–165.

Dunn, H. K. 1950. The calculation of vowel ressonances, and an electrical vocal tract. *Journal of the Acoustical Society of America* 22, 740–753. Reimpresso em Flanagan & Rabiner (1973).

Durand, Marguerite. 1953. De la formation des voyelles nasales. *Studia Linguistica* 7, 33–53.

Dutoit, Thierry. 1997. *An Introduction to Text-to-Speech Synthesis*. The Netherlands: Kluwer Academic Publisher.

Dutoit, Thierry & Stylianou, Yannis. 2003. Text-to-speech synthesis. In R. Miktov (ed.), *The Oxford Handbook of Computational Linguistics*, cap. 17, 323–338. Oxford University Press.

Emiliano, António. 2006. Convenções gerais de transcrição fonética do português europeu (de acordo com a pronunciação de Lisboa). `http://www.fcsh.unl.pt/docentes/aemiliano/documentos_diversos/transcricao_fonetica_do_PE.pdf` (8 Maio, 2007).

Engwall, Olov. 1999. Modelling of the vocal tract in three dimensions. In Géza Gordos & Géza Németh (eds.), *European Conference on Speech Communication and Technology (Eurospeech)*, vol. 1, 113–116. Budapest, Hungary.

Engwall, Olov. 2004. Speaker adaptation of a three-dimensional tongue model. In *International Conference on Spoken Language Processing (ICSLP)*, 465–468. Jeju Island, Korea.

Epsy-Wilson, Carol. 1992. Acoustic measures for linguistic features distinguishing the semivowels /wjrl/ in American English. *Journal of the Acoustical Society of America* 92, 736–757.

Fallside, Frank & Young, Stephen. 1979. Speech synthesis from concept: A method for speech output from information systems. *Journal of the Acoustical Society of America* 66(3), 685–695.

Fant, Gunnar. 1960. *Acoustic Theory of Speech Production*. Gravenhage, The Netherlands: Mouton & Co.

Faria, Isabel Hub, Pedro, Emília Ribeiro, Duarte, Inês & Gouveia, Carlos A. M. (eds.). 1996. *Introdução à Linguística Geral e Portuguesa*. Lisboa: Caminho.

Feldman, David. 1972. On utterance-final [l] and [u] in Portuguese. In A. Valdman (ed.), *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre*, 129–141. Mouton: The Hague.

Ferreira, Manuela Barros, Carrilho, Ernestina, Lobo, Maria, Saramago, João & da Cruz, Luísa Segura. 1996. Variação linguística: perspectiva dialectológica. In Isabel Faria, Emília Ribeiro Pedro, Inês Duarte & Carlos Gouveia (eds.), *Introdução à Linguística Geral e Portuguesa*, 479–502. Lisboa: Caminho.

Fisher, William. 1996. A C implementation of Daniel Kahn's theory of English syllable structure. `ftp://jaguar.ncsl.nist.gov/pub/tsylb2-1.1.tar.Z`.

Flanagan, J. L., Ishizaka, K. & Shipley, K. L. 1975. Synthesis of speech from a dynamic model of the vocal cords and vocal tract. *The Bell System Technical Journal* 54(3), 485–506.

Flanagan, James. 1972. *Speech Analysis, Synthesis and Perception*. New York: Springer-Verlag, 2a. edn..

Flanagan, James & Rabiner, Lawrence (eds.). 1973. *Speech synthesis*. Benchmark Papers in Acoustics. Dowden Hutchinson & Ross.

Florian, Radu & Ngai, Grace. 2001. *Fast Transformation-Based Learning*.

Fowler, Carol. 1977. *Timing Control in Speech Production*. Bloomington: Indiana University Linguistic Club.

Frantz, Gene & Wiggins, Richard. 1982. Design case history: Speak & Spell learns to talk. *IEEE Spectrum* 19(2), 45–49.

Freitas, Maria João. 1997. *Aquisição da Estrutura Silábica do Português Europeu*. Tese de Doutoramento, Universidade de Lisboa, Lisboa.

Freitas, Maria João. 1998. Estatutos das consoantes que fecham sílabas no português europeu. In *Encontro da Associação Portuguesa de Linguística (APL)*, 541–555.

Freitas, Maria João. 2000. Os ping[w]ins são diferentes dos c[w]elhos? Questões sobre oclusivas velares, semivogais e arredondamentos na aquisição do português europeu. In *Encontro da Associação Portuguesa de Linguística (APL)*, 213–235.

Freitas, Maria João. 2002. Pratos, patos e p[ö]ratos: o caso da aquisição dos ataques complexos em português europeu. In Maria Helena Mateus e Clara Nunes Correia (ed.), *Saberes no Tempo. Homenagem a Maria Henriqueta Costa Campos*, 299–314. Lisboa: Edições Colibri.

Freitas, Maria João. 2003. The acquisition of onset clusters in European Portuguese. *Probus. International Journal of Latin and Romance Linguistics* 15(1), 27–46.

Freitas, Maria João & Santos, Ana Lúcia. 2001. *Contar (histórias) de sílabas. Descrição e Implicações para o Ensino do Português como Língua Materna*. Lisboa: Edições Colibri.

Frota, Sónia. 2000. *Prosody and focus in European Portuguese. Phonological phrasing and intonation*. New York: Garland Publishing.

Fudge, Eric. 1999. Syllables. In J. A. Goldsmith (ed.), *Phonological Theory. The Essential Readings*, cap. 19, 370–391. Blackwell Publishers.

Fujimura, Osamu & Lovins, Julie. 1978. Syllables as concatenative phonetic elements. In A. Bell & J. B. Hooper (eds.), *Syllables and Segments*, 107–120. New York: North Holland.

Fujisaki, Hiroya & Nagashima, S. 1969. A model for the synthesis of pitch contours of connected speech. *Annual Report of Engineering Research Institute* 28, 53–60.

Fujisaki, Hiroya, Ohno, Sumio & Yagi, Takashi. 1997. Analysis and modeling of fundamental frequency contours of Greek utterances. In *European Conference on Speech Communication and Technology (Eurospeech)*, 465–468. Rhodes, Greece.

Gabioud, Bernard. 1994. Articulatory models in speech synthesis. In Keller (1994), cap. 10, 215–270.

Gafos, Adamantios. 2002. A grammar of gestural coordination. *Natural Language and Linguistic Theory* 20(2), 269–337.

Gartenberg, Robert. 1984. An electropalatographic investigation of allophonic variation in English /l/ articulations. *Speech Research Laboratory Work in Progress* 4, 135–157.

Gick, Brian. 2003. Articulatory correlates of ambisyllabicity in English glides and liquids. In J. Local, R. Ogden & R. Temple (eds.), *Papers in Laboratory Phonology VI: Constrains on Phonetic Interpretation*, 222–236. Cambridge: Cambridge University Press.

Gick, Bryan. 1999a. *The Articulatory Basis of Syllable Structure: A study of English glides and liquids*. PhD thesis, Yale University, Department of Linguistics.

Gick, Bryan. 1999b. A gesture-based account of intrusive consonants in English. *Phonology* 16, 29–54.

Gick, Bryan, Campbel, Fiona, Oh, Sunyoung & Tamburri-Watt, Linda. 2006. Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics* 34, 49–72.

Giles, Stephen & Moll, Kenneth. 1975. Cinefluorographic study of selected allophones of English /l/. *Phonetica* 31, 206–227.

Goldsmith, John. 1990. *Autosegmental and Metrical Phonology*. Cambridge: Blackwell.

Goldstein, Louis. 2005. Syllable structure and modes of coupled dynamical systems. In *PaPI-Phonetics and Phonology in Iberia*. Barcelona.

Goldstein, Louis, Byrd, Dani & Saltzman, Elliot. 2006. The role of vocal tract gestural action units in understanding the evolution of phonology. In Michael Arbib (ed.), *Action to Language via the Mirror Neuron System*, 215–249. Cambridge, UK: Cambridge University Press.

Goldstein, Louis, Chitoran, Ioana & Selkirk, Elisabeth. 2007. Syllable structure as coupled oscillator modes: Evidence from Georgian vs. Tashlhiyt Berber. In *International Congress of Phonetic Sciences (ICPhS)*, 241–243. Saarbrücken, Germany.

Gouveia, Paulo, Teixeira, João Paulo & Freitas, Diamantino. 2000. Divisão silábica automática do texto escrito e falado. In *Encontro para o Processamento Computacional da Língua Portuguesa Escrita e Falada (PROPOR)*. Atibaia, S. Paulo, Brasil.

Gregio, Fabiana Nogueira. 2006. *Configuração do trato vocal supraglótico na produção das vogais do português brasileiro: dados de imagens de ressonância magnética*. Tese de mestrado, Pontifícia Universidade Católica de São Paulo, PUC/SP, Brasil.

Hajek, John. 1997. *Universals of Sound Change in Nasalization*. Oxford: Blackwell.

Hajek, John. 2008. Vowel nasalization. In Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie (eds.), *The World Atlas of Language Structures Online*, cap. 10. Munich: Max Planck Digital Library. C`http://wals.info/feature/10` (5 Dezembro, 2008).

Hajek, John & Watson, Ian. 2007. Prosodic conditioning of Portuguese subjects' perception of vowel nasality. In *International Congress of Phonetic Sciences (ICPhS)*. Saarbrücken, Germany.

Haken, H., Kelso, J. A. S. & Bunz, H. 1985. A theoretical model of phase transitions in human hand movements. *Biological Cybernetics* 51(5), 347–356.

Hall, Robert. 1972. The place of rules in linguistic analysis. In *Studies in Linguistics in Honour of George L. Trager*, 41–43. The Hague: Mouton.

Handley, Zöe & Hamel, Marie-Josée. 2005. Establishing a methodology for benchmarking Speech Synthesis for Computer-Assisted Language Learning (CALL). *Language Learning & Technology* 9(3), 99–120.

Hardcastle, William & Barry, William. 1989. Articulatory and perceptual factors in /l/ vocalisations in English. *Journal of the Internacional Phonetic Association* 15, 3–17.

Hawkins, Sarah. 1992. An introduction to task dynamics. In G.J. Docherty & D.R. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, cap. I, 9–25. Cambridge: Cambridge University Press.

Head, Brian F. 1964. *A comparison of the segmental phonology of Lisbon and Rio de Janeiro*. PhD dissertation, University of Texas, Austin.

Henderson, Janette. 1984. *Velopharyngeal function in oral and nasal vowels: a cross-language study*. PhD thesis, University of Connecticut, Storrs.

Hermes, Anne, Grice, Martine, Mücke, Doris & Niemann, Henrik. 2008. Articulatory indicators of syllable affiliation in word initial consonant clusters in Italian. In *International Seminar on Speech Production*. Strasbourg, France.

Hirst, Daniel. 1994. The symbolic coding of fundamental frequency curves: from acoustics to phonology. In *International Symposium on Prosody, Satellite Workshop of ICSLP 94*. Yokohama.

Holmes, John & Holmes, Wendy. 2001. *Speech Synthesis and Recognition*. Taylor & Francis, 2a. edn..

Hombert, Jean-Marie. 1986. Word games: Some implications for analysis of tone and other phonological constructs. In J. J. Ohala & J. J. Jaeger (eds.), *Experimental Phonology*, 27–37. Orlando: Academic Press.

Honorof, Douglas & Browman, Catherine. 1995. The center of edge: How are consonant clusters organized with respect to the vowel? In *International Congress of Phonetic Sciences (ICPhS)*, 552–555. Stockholm.

Hoole, Philip. 1996. Issues in the acquisition, processing, reduction and parametrization of articulographic data. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM)* 34, 158–173.

Hoole, Philip & Nguyen, Noel. 1999. Electromagnetic articulography. In William J. Hardcastle & Nigel HewlettVers une phonémisation automatique des sigles (eds.), *Coarticulation: Theory, Data and Techniques*, 260–269. Cambridge: Cambridge University Press.

Hooper, Joan B. 1972. The syllable in phonological theory. *Language* 48, 525–540.

Horiguchi, Satoshi & Bell-Berti, Fredericka. 1987. The Velotrace: A device for monitoring velar position. *The Cleft Palate Journal* 24(2), 104–111.

Howard, Harry & Goldman, Robert P. 1994. From text to syllable in Castilian. *Procesamiento del lenguaje natural* 15.

Hualde, José Ignacio. 1991. On Spanish syllabification. In Héctor Campos y Fernando Martínez-Gil (ed.), *Current Studies in Spanish Linguistics*, 475–493. Washington, DC: Georgetown University Press.

Huang, Xuedong, Acero, Alex & Hon, Hsiao-Wuen. 2001. *Spoken Language Processing: A Guide to Theory, Algorithm and System Development*. USA: Prentice Hall PTR, 1a. edn..

Huffman, Marie. 1989. *Implementation of Nasal Timing and Articulatory Landmarks*. PhD thesis, UCLA.

Hunnicut, Sharon. 1980. Grapheme-to-phoneme rules: a review. *STL-QPSR* 21(2-3), 38–60.

International Phonetic Association. 1999. *Handbook of the International Phonetic Association: A Guide of the Use of the International Phonetic Alphabet*. Cambridge University Press.

Iskarous, Khalil, Goldstein, Louis, Whalen, Douglas, Tiede, Mark & Rubin, Philip. 2003. CASY: The Haskins configurable articulatory synthesizer. In *International Congress of Phonetic Sciences (ICPhS)*, vol. 1, 185–188. Barcelona, Spain.

Jackson, Philip. 2005. Mama and papa: the ancestors of modern-day speech science. In C.U.M. Smith & R.G. Arnott (eds.), *The Genius of Erasmus Darwin*, 217–236. Aldershot, UK: Ashgate.

Jakobson, Roman. 1969. *Langage enfantin et aphasie*. Paris: Editions Minuit.

Jesus, Luis, Almeida, Carolina & Araújo, Luísa. 2007. IPA and SAMPA charts. Universidade de Aveiro.

Jesus, Luis & Shadle, Christine. 2005. Acoustic analysis of European Portuguese uvular [ʁ χ] and voiceless tapped alveolar [ɾ̥] fricatives. *Journal of the International Phonetic Association* 35(1), 27–44.

Jilka, Matthias, Möhler, Gregor & Dogil, Grzegorz. 1999. Rules for the generation of ToBI-based American English intonation. *Speech Communication* 28(2), 83–108.

Jurafsky, Daniel & Martin, James H. 2008. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall (draft).

Kahn, Daniel. 1976. *Syllable-based generalizations in English Phonology*. PhD, Cambridge:MIT.

Keller, Eric (ed.). 1994. *Fundamentals of Speech Synthesis and Speech Recognition - Basic Concepts, State-of-the-art and Future Challenges*. John Wiley & Sons.

Keller, Eric & Bianchi, Olivier. 2002. Virtual Historic Reconstruction with Speech Synthesis. In A. Braun & H.R Masthoff (eds.), *Phonetics and its Applications. Festschrift for Jens-Peter Köster on the Occasion of his 60th Birthday*, 465–484. Stuttgart: Steiner.

Keller, Eric & Keller, Brigitte. 2000a. New Uses for Speech Synthesis. *The Phonetican* 81(1), 35–40.

Keller, Eric & Keller, Brigitte. 2000b. Speech Synthesis in Language Learning: Challenges and opportunities. In *InSTIL Conference*, 109–116. Dundee, Scotland.

Kelso, J. A. Scott, Vatikiotis-Bateson, Eric, Saltzman, Elliot & Kay, Bruce. 1985. A qualitative dynamic analysis of reiterant speech production: phase portraits, kinematics and dynamic modelling. *Journal Acoustical Society of America* 77, 266–280.

Kiraz, George & Möbius, Bernd. 1998. Multilingual syllabification using weighted finite-state transducers. In *ESCA Workshop on Speech Synthesis*, 61–64. Jenolan Caves, Australia.

Klatt, Dennis. 1979. Synthesis by rule of segmental durations in English sentences. In B. Lindblom & S. Öhman (eds.), *Frontiers of Speech Communication Research*, 287–299. New York: Academic Press.

Klatt, Dennis. 1980. Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America* 67(3), 971–995.

Klatt, Dennis. 1987. Review of Text-to-Speech Conversion for English. *Journal of the Acoustical Society of America* 82(3), 737–793.

Krakow, Rena. 1989. *The articulatory organization of syllables: a kinematic analysis of labial and velar gestures*. PhD dissertation, Yale University.

Krakow, Rena. 1993. Nonsegmental influences on velum movement patterns: Syllables, sentences, stress, and speaking rate. In Marie K. Huffman & Rena A. Krakow (eds.), *Nasals, Nasalization, and the Velum*, Phonetics and Phonology (vol. 5), 87–116. Academic Press Inc.

Krakow, Rena. 1999. Physiological organization of syllables: a review. *Journal of Phonetics* 27, 23–54.

Kröger, Bernd & Birkholz, Peter. 2007. A gesture-based concept for speech movement control in articulatory speech synthesis. In *Verbal and Nonverbal Communication Behaviours (COST Action 2102 International Workshop, Vietri sul Mare, Italy, March 29-31, 2007, Revised Selected and Invited Papers)*, 174–189. Springer Berlin / Heidelberg.

Kugler, P. N., Kelso, J. A. Scott & Turvey, M. T. 1982. On the control and coordination of naturally developing systems. In J. A. S. Kelso & J. E. Clark (eds.), *The Development of Movement Control and Coordination*, AIP Series in Modern Acoustics and Signal Processing. Cichester: Wiley.

Lacerda, Armando & Hammarström, Göran. 1952. Transcrição fonética do português normal. *Revista do Laboratório de Fonética Experimental (de Coimbra)* I, 119–135.

Lacerda, Armando & Head, Brian. 1966. Análise de sons nasais e sons nasalizados do português. *Revista do Laboratório de Fonética Experimental (de Coimbra)* 6, 5–70.

Lacerda, Armando & Strevens, Peter. 1956. Some phonetic observations using a speech-stretcher. *Revista do Laboratório de Fonética Experimental (de Coimbra)* 3, 5–16.

Ladefoged, Peter. 1975. *A Course in Phonetics*. Harcourt Brace Jovanovich, Inc.

Ladefoged, Peter & Maddieson, Ian. 1995. *The Sounds of the World's Languages*. Oxford: Blackwell Publishers.

Lajoie, Stephane, Carr, Oliver, Hashemi-Sakhtsari, Ahmad & Coleman, Michael. 2000. Application of Language Technology to Lotus Notes Based Messaging for Command and Control. In *International Command and Control Research and Technology Symposium*. Canberra, Austrália.

Laporte, Eric. 1988. *Méthodes algoritmiques et lexicales de phonetization de textes*. Tesis doctoral, Université Paris 7. Centre d'Études et de Recherches en Informatique Linguistique.

Lüdtke, Helmut. 1952. Fonemática portuguesa. *Boletim de Filologia* 13-14, 273–288,197–217.

Lehiste, Ilse. 1964. *Acoustical characteristics of selected English consonants*. Indiana: Indiana University.

Lehiste, Ilse. 1996. Suprasegmentals features of speech. In *Principles of Experimental Phonetics*, 226–245. Mosby.

Lemmetty, Sami. 1999. *Review of Speech Synthesis Technology*. Master thesis, Helsinki University of Technology.

Levinson, Stephen, Olive, Joseph & Tschirgi, Judith. 1993. Speech Synthesis in Telecommunications. *IEEE Communications Magazine* 31(11), 46–53.

Liberman, Mark & Church, Kenneth. 1992. Text analysis in word pronounciation. In Sadaoki Furui & M. Mohan Sondhi (eds.), *Advances in Speech Signal Processing*, cap. 24, 791–831. New York, USA: Marcel Dekker.

Libossek, Marion & Schiel, Florian. 2000. Syllable-based text-to-phoneme conversion for German. In *International Conference on Spoken Language Processing (ICSLP)*, vol. 2, 283–286. Beijing, China.

Liénard, Jean-Sylvain. 1967. Reconstitution de la machine parlante de Kempelen. In *Conférence Internacionale d'Acoustique*. Budapest.

Liénard, Jean-Sylvain. 1991. From Speaking Machines to Speech Synthesis. In *Congrès International des Sciences Phonétiques*, 18–27. Aix-en-Provence, France.

Lipski, John. 1975. Brazilian Portuguese vowel nasalization: secondary aspects. *Canadian Journal of Linguistics* 20, 59–77.

Llisterri, Joaquim, Carbó, Carme, Machuca, Maria Jesús, de la Mota, Carme, Riera, Montserrat & Ríos, Antonio. 2004. La conversión de texto en habla: aspectos lingüísticos. In M. A. Martí & J. Llisterri (eds.), *Tecnologías del texto y del habla*, 145–186. Barcelona: Edicions de la Universitat de Barcelona - Fundación Duques de Soria.

Llisterri, Joaquim, Machuca, Maria Jesús, de la Mota, Carme, Riera, Montserrat & Ríos, Antonio. 2003. Entonación y tecnologías del habla. In P. Prieto (ed.), *Teorías de la entonación*, 209–243. Barcelona: Ariel.

Lopez-Gonzalo, Eduardo, Olaszy, Gabor & Nemeth, Geza. 1993. Improvements of the Spanish version of the MULTIVOX text-to-speech system. In *European Conference on Speech Communication and Technology (Eurospeech)*, vol. 2, 869–872. Berlin, Germany.

Louro, José Inês. 1954-1955. Estudo e classificação das vogais. *Boletim de Filologia* Tomo XV, 215–248.

Lovatto, Liane, Amelot, Angélique, Crevier-Buchman, Lise, Basset, Patricia & Vaissière, Jacqueline. 2007. A fiberscopic analysis of nasal vowels in Brazilian Portuguese. In *International Congress of Phonetic Sciences (ICPhS)*, 549–552. Saarbrücken, Germany.

MacKay, Donald. 1978. Speech errors inside the syllable. In A.Bell & J. Hooper (eds.), *Syllables and segments*, 201–212. New York: North Holland.

MacNeilage, Peter. 1998. The frame/ content theory of evolution of speech production. *Brain and Behavioral Science* 21, 499–546.

Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.

Maddieson, Ian. 2007. Areal distribution of nasalized vowels. In *International Congress of Phonetic Sciences (ICPhS)*, 1381–1384. Saarbrücken, Germany.

Maeda, Shinji. 1993. Acoustics of vowel nasalization and articulatory shifts in French nasal vowels. In Marie K. Huffman & Rena A. Krakow (eds.), *Nasals, Nasalization, and the Velum*, Phonetics and Phonology (vol. 5), 147–167. Academic Press Inc.

Marchand, Yannick, Adsett, Connie & Damper, Robert. 2007. Evaluation of automatic syllabification algorithms for English. In *International Speech Communication Association (ISCA) Workshop on Speech Synthesis*. Bonn, Germany.

Marchand, Yannick & Damper, Robert. 2000. A multi-strategy approach to improving pronunciation by analogy. *Computational Linguistics* 26(2), 195–219.

Marchand, Yannick & Damper, Robert. 2007. Can syllabification improve pronunciation by analogy of English? *Natural Language Engineering* 13(1), 1–24.

Mareüil, Phillipe Boula de. 1994. Vers une phonémisation automatique des sigles. In *Journées d'Étude sur la Parole*, 95–100. Trégastel.

Mareüil, Phillipe Boula de. 1995. Vers la phonématisation automatique des sigles. *La Linguistique: revue de la Société internationale de linguistique fonctionnelle* 31(1), 93–103.

Mareüil, Phillipe Boula de, Adda-Decker, Martine & Woehrling, Cécile. 2007. Analysis of oral and nasal vowel realization in northern and southern French varieties. In *International Congress of Phonetic Sciences (ICPhS)*. Saarbrücken, Germany.

Mareüil, Phillipe Boula de & Floricic, Franck. 2001. On the pronunciation of acronyms in French and in Italian. In *European Conference on Speech Communication and Technology (Eurospeech)*. Aalborg.

Marin, Stefania. 2005. Complex nuclei in articulatory phonology: The case of Romanian diphthongs. In R. S. Gess & E. Rubin (eds.), *Selected papers of the 34th LSRL*, 161–177. Amsterdam, Philadelphia: John Benjamins.

Marin, Stefania. 2007. An articulatory modeling of Romanian diphthong alternations. In *International Congress of Phonetic Sciences (ICPhS)*. Saarbrücken, Germany.

Marin, Stefania & Pouplier, Marianne. 2008. Organization of complex onsets and codas in American English: Evidence for a competitive coupling model. In *International Seminar on Speech Production*. Strasbourg, France.

Maroco, João. 2007. *Análise Estatística com Utilização do SPSS*. Edições Sílabo, 3a. edn..

Martí, Josep & Niñerola, Daniel. 1987. SINCAS: un conversor texto-voz en castellano. In *Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural*, 112–122. Tarragona, Spain.

Martins, Ana Maria. 1995. A evolução das vogais nasais finais -ã, -õ, -ẽ no português. In Cilene da Cunha Pereira & Paulo R. D. Pereira (eds.), *Miscelânea de Estudos Lingüísticos, Filológicos e Literários In Memoriam Celso Cunha*, 617–646. Rio de Janeiro: Editora Nova Fronteira.

Martins, Maria Raquel Delgado. 1977. *Caderno de Fonética do Português*. Laboratório de Fonética da Universidade de Lisboa.

Martins, Paula. 2007. *Ressonância Magnética no Estudo da Produção do Português Europeu*. Dissertação de Mestrado, Universidade de Aveiro, Aveiro, Portugal.

Martins, Paula, Carbone, Inês, Pinto, Alda, Silva, Augusto & Teixeira, António. 2008a. European Portuguese MRI based speech production studies. *Speech Communication* 50(11-12), 925–952.

Martins, Paula, Carbone, Inês, Silva, Augusto & Teixeira, António. 2007. An MRI study of European Portuguese nasals. In *Annual Conference of the International Speech Communication Association (Interspeech)*. Antwerp.

Martins, Paula, Carbone, Inês, Silva, Augusto & Teixeira, António. 2008b. Coarticulatory effects on European Portuguese: A first MRI study. In M. Embarki & C. Dodane (eds.), *La Coarticulation : Indices, Direction et Représentation*. L' Harmattan.

Massaro, Dominic. 2002. The psychology and technology of talking heads in human-machine interaction. In *International CLASS Workshop on Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems*, 106–119. Copenhagen.

Massaro, Dominic & Light, Joanna. 2003. Read my tongue movements: Bimodal learning to perceive and produce non-native speech /r/ and /l/. In *European Conference on Speech Communication and Technology (Eurospeech)*, 2249–2252. Geneva, Switzerland.

Massaro, Dominic & Light, Joanna. 2004. Using visible speech to train perception and production of speech for individuals with hearing loss. *Journal of Speech Language and Hearing Research* 47(2), 304–320.

Master, Suely, Pontes, Paulo & Behlau, Mara. 1991. Configuração do trato vocal na emissão das vogais nasais do português brasileiro. *Acta AWHO* 10(2), 67–71.

Mateescu, Dan. 2003. *English Phonetics and Phonological Theory: 20th Century Approaches*. Bucareste: Editora da Universidade d Bucareste.

Mateus, Maria Helena Mira. 1975. *Aspectos de Fonologia Portuguesa*. Lisboa: Publicações do Centro de Estudos Filológicos.

Mateus, Maria Helena Mira. 1989. Prosódia e estratégias do discurso. In *Congresso da Faculdade de Letras do Rio de Janeiro- Discurso e Ideologia*, 237–249. Rio de Janeiro.

Mateus, Maria Helena Mira. 1993. Onset of Portuguese syllables and rising diphtongs. In *Workshop on Phonology*, 93–104. Coimbra.

Mateus, Maria Helena Mira. 1994. Silabificação de base em Português. In *Encontro da Associação Portuguesa de Linguística (APL)*, 289–300. Lisboa.

Mateus, Maria Helena Mira. 2006. Sobre a natureza fonológica da ortografia portuguesa. In *Estudos da Linguagem: Questões de Fonética e Fonologia: uma Homenagem a Luís Carlos Cagliari*, 159–180. Bahia: Universidade estadual do Sudoeste da Bahia.

Mateus, Maria Helena Mira, Andrade, Amália, do Céu Viana, Maria & Villalva, Alina. 1990. *Fonética, Fonologia e Morfologia do Português*. Lisboa: Universidade Aberta.

Mateus, Maria Helena Mira, Brito, Ana Maria & Duarte, Inês. 2003. *Gramática da Língua Portuguesa*. Lisboa: Caminho, 5a. edn..

Mateus, Maria Helena Mira & d'Andrade, Ernesto. 1998. The syllable structure in European Portuguese. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada* 14(1), 13–32.

Mateus, Maria Helena Mira & d'Andrade, Ernesto. 2000. *Phonology of Portuguese*. Oxford: Oxford University Press.

Mateus, Maria Helena Mira, Falé, Isabel & Freitas, Maria João. 2005. *Fonética e Fonologia do Português*. Lisboa: Universidade Aberta.

Mateus, Maria Helena Mira & Rodrigues, Celeste. 2003. A vibrante em coda no português. In Dermeval da Hora & Gisela Collischonn (eds.), *Teoria Lingüística. Fonologia e outros temas*, 181–199. João Pessoa, Brasil: Editora Universitária.

Matta Machado, Miriam Therezinha da. 1993. Fenômenos de Nasalização Vocálica em Português. Estudo Cine-radiográfico. *Caderno de Estudos Linguísticos* 25, 113–117.

Mattingly, Ignatius. 1974. Speech Synthesis for Phonetic and Phonological Models. In T. S. Sebeok (ed.), *Current Trends in Linguistics*, vol. 12, 2451–2487. The Hague: Mouton.

Medeiros, Beatriz Raposo de, D'Imperio, Mariapaola & Espesser, Robert. 2008. La voyelle nasale en portugais brésilien et son appendice nasal: étude acoustique et aérodynamique. In *Journées d'Etude sur la Parole*, 285–288. Avignon: Editions Universitaires d'Avignon.

Medeiros, José Carlos Dinis. 1995. *Processamento Morfológico e Correcção Ortográfica do Português*. Tese de mestrado, Instituto Superior Técnico, Universidade de Lisboa.

Meinedo, Hugo & Neto, João Paulo. 2000. The use of syllable segmentation information in continuous speech recognition hybrid systems applied to the Portuguese language. In *International Conference on Spoken Language Processing (ICSLP)*. Beijing, China.

Meinedo, Hugo, Neto, João Paulo & Almeida, Luís. 1999. Syllable onset detection applied to the Portuguese language. In *European Conference on Speech Communication and Technology (Eurospeech)*. Budapest, Hungary.

Meinedo, Hugo Daniel dos Santos. 2000. *Utilização de Informação Silábica no Reconhecimento de Fala Contínua*. Tese de mestrado, Instituto Superior Técnico.

Meireles, Alexsandro & Barbosa, Plínio. 2008. Lexical reorganization in Brazilian Portuguese: An articulatory study. *Speech Communication* 50(11-12), 916–924.

Mendes, Helena Margarida, Oliveira, Catarina & Teixeira, António. 2003. PLE: uma sigla para ler ou soletrar? *Cadernos de PLE* 3, 121–139.

Mermelstein, Paul. 1973. Articulatory model for the study of speech production. *Journal of the Acoustical Society of America* 53(4), 1070–1082.

Mestre, Antonio Ríos. 1998. *La transcripción fonética automática del Diccionario Electrónico de Formas Simples Flexivas del Español: un estudio fonológico en el léxico*. Tese de doutoramento, Universidad Autónoma Barcelona.

Mixdorff, Hansjörg & Fujisaki, Hiroya. 1994. Analysis of voice fundamental frequency contours of German utterances using a quantitative model. In *International Conference on Spoken Language Processing (ICSLP)*, 2231–2234. Yokohama, Japan.

Müller, Karin. 2001. Automatic detection of syllable boundaries combining the advantages of treebank and bracketed corpora training. In *Annual Meeting on Association for Computational Linguistics (ACL)*, 410–417. Toulouse, France.

Müller, Karin, Möbius, Bernd & Prescher, Detlef. 2000. Inducing probabilistic syllable classes using multivariate clustering. In *Annual Meeting on Association for Computational Linguistics (ACL)*, 225–32. Hong Kong, China.

Moll, Kenneth. 1962. Velopharyngeal closure on vowels. *Journal of Speech and Hearing Research* 5(1), 30–37.

Moll, Kenneth & Daniloff, Raymond. 1971. An investigation of the timing of velar movements during speech. *Journal of the Acoustical Society of America* 50(2), 678–684.

Montagu, Julie. 2007. *Analyse Acoustique et Perceptive des voyelles nasales et nasalisées du français parisien*. Thèse de doctorat, Université Paris III - Nouvelle Sourbonne.

Montero, J.M., Arriola, J. Gutiérrez, Colás, J., Guarasa, J. Macías, Enríquez, E. & Pardo, J.M. 1999. Development of an emotional speech synthesiser in Spanish. In *European Conference on Speech Communication and Technology (Eurospeech)*, 2099–2102. Budapest, Hungary.

Moraes, João Antônio de. 1997. Vowel nasalization in Brazilian Portuguese: An articulatory investigation. In *European Conference on Speech Communication and Technology (Eurospeech)*. Rhodes, Greece.

Moraes, João Antônio de. 2003. A nasalidade vocálica no português do Brasil e no português de Portugal. In *Congreso Internacional de Lingüística y Filología Románica*. Max Niemeyer Verlag.

Moraes, João Antônio de & Wetzels, Leo. 1992. Sobre a duração dos segmentos vocálicos nasais e nasalizados em português. um exercício de fonologia experimental. *Cadernos de Estudos Lingüísticos* 23, 153–166.

Mota, Carme de la. 1991a. Caracterización acústica de los sonidos vibrantes del español en distintos estilos de habla. In *Congreso de la Sociedad Española de Lingüistica*. Granada.

Mota, Carme de la. 1991b. A study of [r] and [ɾ] in spontaneous speech. In *Congres International des Sciences Phonetiques*, vol. vol. 4, 386–389. Aix-en-Provence, France.

Moutinho, Lurdes de Castro. 2000. *Uma introdução ao Estudo da Fonética e Fonologia do Português*. Plátano Edições Técnicas, 1a. edn..

Nam, Hosung. 2007. Articulatory modeling of consonant release gesture. In *International Congress of Phonetic Sciences (ICPhS)*, 625–628. Saarbrücken, Germany.

Nam, Hosung, Goldstein, Louis & Saltzman, Elliot. 2006. Dynamical modeling of supragestural timing. In *Laboratory Phonology Conference*. Paris, France.

Nam, Hosung, Goldstein, Louis, Saltzman, Elliot & Byrd, Dani. 2004. TADA: An enhanced, portable task dynamics model in MATLAB. *Journal of the Acoustical Society of America* 115(5,2), 2430.

Nam, Hosung & Saltzman, Elliot. 2003. A competitive, coupled oscillator model of syllable structure. In *International Congress of Phonetic Sciences (ICPhS)*, 2253–2256. Barcelona.

Narayanan, Shrikanth, Alwan, Abeer & Haker, Katherine. 1997. Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. part I. The laterals. *Journal of the Acoustical Society of America* 101, 1064–2007.

Nascimento, Maria Fernanda, Marques, Lúcia & Segura, Luísa. 1987. *Português Fundamental: Métodos e Documentos*. Lisboa: INIC-CLUL.

Nikléczy, Péter & Olaszy, Gábor. 2003. A reconstruction of Farkas Kempelen's speaking machine. In *European Conference on Speech Communication and Technology (Eurospeech)*, 2453–2456. Geneva, Switzerland.

Nobiling, Oskar. 1974. As vogais nasais em português. *Littera* 12, 80–109.

Ogden, Richard, Hawkins, Sarah, House, Jill, Huckvale, Mark, Local, John, Carter, Paul, Dankovicova, Jana & Heid, Sebastian. 2000. Prosynth: An integrated prosodic approach to device-independent, natural-sounding speech synthesis. *Computer Speech and Language* 14, 177–210.

Ohala, John & Busà, Maria Gracia. 1995. Nasal loss before voiceless fricatives: a perceptually-based sound change. *Rivista di Linguistica* 7, 125–144.

Ohala, John & Ohala, Manjari. 1991. Nasal epenthesis in Hindi. *Phonetica* 48, 207–220.

Ohala, John & Ohala, Manjari. 1993. The phonetics of nasal phonology: Theorems and data. In Marie K. Huffman & Rena A. Krakow (eds.), *Nasals, Nasalization, and the Velum*, Phonetics and Phonology (vol. 5), 225–249. Academic Press Inc.

Ohala, John, Solé, Maria-Josep & Ying, Goangshiuan. 1998a. Aerodynamic characteristics of trills. In *Meeting of the ISCA/Acoustical Society of America*, 2923–2924. Seattle, Washington, USA.

Ohala, John, Sole, Maria-Josep & Ying, Goangshiuan. 1998b. Do nasalized fricatives exist? In Patricia K. Kuhl & Lawrence A. Crum (eds.), *International Congress on Acoustics (ICA) and the Meeting of the Acoustical Society of America*, vol. IV, 2921–2922. Seattle, Washington, USA.

Ohala, Manjari. 1975. Phonetic explanations for nasal sound patterns. In Charles A. Ferguson, Larry M. Hyman & John J. Ohala (eds.), *Nasálfest - Papers from a Symposium on Nasals and Nasalization*, 289–316. Stanford, CA, USA: Language Universals Project, Department of Linguistics, Stanford University.

Öhman, Sven. 1966. Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America* 39, 151–168.

Olive, Joseph. 1996. "The talking computer": Text to speech synthesis. In David G. Stork (ed.), *HAL's Legacy: 2001'S Computer As Dream and Reality*, cap. 6, 101–129. The MIT Press.

Oliveira, Catarina & Teixeira, António. 2006. *Base de Dados EMMA com Anotação Automática de Gestos*. Relatório técnico, Dep. Electrónica e Telecomunicações / IEETA, Universidade de Aveiro.

Oliveira, Catarina & Teixeira, António. 2007a. *Nova Base de Dados EMMA relativa às Nasais e Laterais do Português Europeu*. Relatório técnico, IEETA, Universidade de Aveiro.

Oliveira, Catarina & Teixeira, António. 2007b. On gestures timing in European Portuguese nasals. In *International Congress of Phonetic Sciences (ICPhS)*. Saarbrücken, Germany.

Oliveira, Fernão de. 1536. *Grammatica da Linguagem Portugueza*. INCM. Edição de 1975.

Oliveira, Leonardo & Marin, Stefania. 2005. Patterns of velum coordination in Brazilian Portuguese. In *Phonetics and Phonology in Iberia (PAPI)*. Barcelona, Spain.

Oliveira, Luís Caldas, Viana, Maria do Céu & Trancoso, Isabel. 1991. DIXI: Portuguese Text-to-Speech System. In *European Conference on Speech Communication and Technology (Eurospeech)*, 1239–1242. Genoa.

Oliveira, Luís Caldas, Viana, Maria do Céu & Trancoso, Isabel. 1992. A rule-based text-to-speech system for Portuguese. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 73–76. San Francisco, CA, USA.

Oliveira, Luís Caldas, Viana, Maria do Céu & Trancoso, Isabel. 1993. DIXI: Sistema de síntese de fala a partir do texto para o português. In *Encontro sobre o Processamento da Língua Portuguesa Escrita e Falada (EPLP)*. Lisboa.

Oliveira, Luís Caldas. 1996. *Síntese de Fala a Partir do Texto*. Dissertação de doutoramento, Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisboa.

Oliveira, Luís Caldas, Viana, Maria do Céu, Mata, Ana & Trancoso, Isabel. 2001. *Progress report of project DIXI+: A Portuguese Text-to-Speech Synthesizer for Alternative and Augmentative Communication*. Relatório técnico, FCT.

O'Shaughnessy, Douglas, Barbeau, Louis, Bernardi, David & Archanbault, Daniele. 1988. Diphone speech synthesis. *Speech Communication* 7(1), 55–65.

Pardo, José Manuel. 2004. Nuevas fronteras de la tecnología del habla: Reconocimiento de voz y síntesis de voces y emociones. In *Foro Complutense: Jornada Tecnología y Discapacidad Visual*. Madrid: Facultad de Informática de la UCM.

Parkinson, Stephen. 1983. Portuguese nasal vowels as phonological diphthongs. *Lingua* 61, 157–177.

Parkinson, Stephen. 1996. Aspectos teóricos da história das vogais nasais portuguesas. In *Encontro da Associação Portuguesa de Linguística (APL)*, 253–272. Braga.

Paulo, Sérgio, Oliveira, Luís Caldas, Mendes, Carlos, Figueira, Luís, Cassaca, Renato, Viana, Maria do Céu & Moniz, Helena. 2008. DIXI - a generic text-to-speech system for European Portuguese. In António Teixeira, Vera Lúcia Strube de Lima, Luís Caldas Oliveira & Paulo Quaresma (eds.), *Computational Processing of the Portuguese Language (PROPOR 2008)*, 91–100. Springer.

Perkell, Joseph, Cohen, Marc, Svirsky, Mario, Natthies, Melanie, naki Garabieta, I & Jackson, Michell. 1992. Electromagnetic midsagital articulometer systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America* 92(6), 3078–3096.

Peterson, Gordon, Wang, William & Sivertsen, Eva. 1958. Segmentation techniques in speech synthesis. *Journal of the Acoustical Society of America* 30(8), 739–742.

Pettorino, Massimo. 1999. Memmon, the vocal statue. In John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville & Ashlee C. Bailey (eds.), *International Congress of Phonetic Sciences (ICPhS)*, 184–187. San Francisco: University of California, Berkeley.

Pettorino, Massimo & Giannini, Antonella. 1999. *Le teste parlanti*. Palermo: Sellerio.

Pierrehumbert, Janet. 1981. Synthesizing intonation. *Journal of the Acoustical Society of America* 70(4), 985–995.

Plénat, Marc. 1993. Observations sur le mot minimal français. L'oralisation des sigles. In B. Laks & M. Plénat (eds.), *De Natura sonorum. Essais de phonologie*, 143–172. Saint-Denis: Presses Universitaires de Vincennes.

Plénat, Marc. 1998. De quelques paramètres intervenant dans l'oralisation des sigles en français. *Cahiers d'Etudes Romanes (CERCLID)* 9, 27–52.

Pompino-Marschall, Bernd. 2005. Von Kempelen et al. - remarks on the history of articulatory-acoustic modelling. *Zas Papers in Linguistics* (40), 145–159.

Pontes, Eunice. 1972. *Estrutura do Verbo no Português Coloquial*. Petrópolis: Vozes.

Portele, Thomas, Steffan, Birgit, PreuSS, Rainer, Sendlmeier, Walter & Hess, Wolfgang. 1992. HADIFIX - A speech synthesis system for German. In *International Conference on Spoken Language Processing (ICSLP)*. Banff, Alberta, Canada.

Pouplier, Marianne. 2008. The articulatory organization of onset and coda clusters in German. In *Workshop: Consonant Clusters and Structural Complexity*. Munich.

Prado, Pedro. 1991. *A Target-Based Articulatory Synthesizer*. Tese de Doutoramento, University of Florida.

Pérez, Juan Carlos & Vidal, Enrique. 1991. Un sistema de conversión de texto a voz para castellano. *Boletín de la Sociedad Española para el Procesamiento del Lenguaje Natural* 11, 197–207.

Prince, Alan & Smolensky, Paul. 2004. *Optimality Theory. Constraint Interaction in Generative Grammar*. Malden MA: Blackwell.

Ramalho, José Carlos & Henriques, Pedro. 2002. *XML /&/ XSL: Da teoria à prática*. Lisboa: FCA - Editora de Informática.

Recasens, Daniel. 1991. On the production characteristics of apicoalveolar taps and trills. *Journal of Phonetics* 19, 267–280.

Recasens, Daniel. 2004. Darkness in [l] as a scalar phonetic property: implications for phonology and articulatory control. *Clinical Linguistics & Phonetics* 18(6-8), 593–603.

Recasens, Daniel & Espinosa, Aina. 2005. Articulatory, positional and coarticulatory characteristics for clear /l/ and dark /l/: evidence from two Catalan dialects. *Journal of the International Phonetic Association* 35(1), 1–25.

Recasens, Daniel & Espinosa, Aina. 2006. Articulatory, positional and contextual characteristics of palatal consonants: Evidence from majorcan Catalan. *Journal of Phonetics* 34, 295–318.

Recasens, Daniel, Fontdevila, Jordi & Pallares, Maria Dolors. 1995. Velarization degree and coarticulatory resistance for /l/ in Catalan and German. *Journal of Phonetics* 23, 37–52.

Recasens, Daniel, Fontdevila, Jordi & Pallarès, Maria Dolors. 1996. Linguopalatal coarticulation and alveolar-palatal correlations for velarized and non-velarized /l/. *Journal of Phonetics* 24, 165–185.

Recasens, Daniel & Pallarès, Maria Dolors. 1999. A study of /r/ and /rr/ in the light of the "DAC" coarticulation model. *Journal of Phonetics* 27, 143–170.

Redford, Melissa. 1999. *An articulatory basis for the syllable*. PhD dissertation, University of Texas, Austin.

Reichel, Uwe & Schiel, Florian. 2005. Using morphology and phoneme history to improve grapheme-to-phoneme conversion. In *Annual Conference of the International Speech Communication Association (Interspeech)*, 1937–1940. Lisboa.

Ribeiro, Ricardo, Oliveira, Luís & Trancoso, Isabel. 2002. Morphossyntactic disambiguation for TTS systems. In *International Conference on Language Resources and Evaluation (LREC)*. Las Palmas, Spain.

Ribeiro, Ricardo, Oliveira, Luís & Trancoso, Isabel. 2003. Using morphossyntactic information in TTS systems: Comparing strategies for European Portuguese. In *Computacional Processing of the Portuguese Language (PROPOR)*, 143–150. Faro, Portugal.

Riedi, Marcel Plazi. 1998. *Controlling segmental duration in speech synthesis systems*. Thesis doctoral, Swiss Federal Institute of Technology Zurich.

Riley, Michael. 1992. Tree-based modelling for speech synthesis. In Bailly *et alii* (1992), 265–273.

Riskin, Jessica. 2003. Eighteenth-century wetware. *Representations* Summer(83), 97–125.

Rodríguez, Miguel, Escalada, José, Macarrón, Alejandro & Monzón, Luis. 1993. AMIGO: un conversor texto-voz para español. *Boletín de la Sociedad Española para el Procesamiento del Lenguaje Natural* 13, 388–400.

Roon, Kevin, Gafos, Adamantios, Hoole, Phil & Zeroual, Chakir. 2007. Influence of articulator and manner on stiffness. In *International Congress of Phonetic Sciences (ICPhS)*, 409–412. Saarbrücken, Germany.

Roon, Kevin, Gafos, Adamantios, Hoole, Phil & Zeroual, Chakir. 2008. Obligatory release and stiffness modulation in Moroccan Arabic. In *Laboratory Phonology 11*. Wellington, New Zealand.

Rosen, George. 1958. Dynamic analog speech synthesizer. *Journal of the Acoustical Society of America* 30(3), 201–209.

Rossato, Solange, Badin, Pierre & Bouaouni, F. 2003. Velar movements in French: An articulatory and acoustical analysis of coarticulation. In *International Congress of Phonetic Sciences (ICPhS)*. Barcelona.

Rossato, Solange, Teixeira, António & Ferreira, Liliana. 2006. Les nasales du portugais et du français: une étude comparative sur les données EMMA. In *Journées d'Études sur la Parole (JEP)*, 143–146. Dinard, France.

Rousset, Isabelle. 2004. *Structures Syllabiques et lexicales des langues du monde. Données, typologies, tendances universelles et contraintes substantielles*. PhD thesis, Institut de la Communication Parlée, Université Grenoble III.

Rubin, Philip, Baer, Thomas & Mermelstein, Paul. 1981. An articulatory synthesizer for perceptual research. *Journal of the Acoustical Society of America* 70(2), 321–328.

Rubin, Philip, Saltzman, Elliot, Goldstein, Louis, McGowan, Richard, Tiede, Mark & Browman, Catherine. 1996. CASY and extensions to the task-dynamic model. In *ETRW on Speech Production Modeling: From Control Strategies to Acoustics*, 125–128. Autrans, France.

Rubin, Philip & Vatikiotis-Bateson, Eric. 2006. Talking heads: Simulacra . The Early History of Talking Machines. `http://www.haskins.yale.edu/featured/heads/simulacra.html` (5 Janeiro, 2007).

Sá Nogueira, Rodrigo de. 1938. *Elementos para um tratado de Fonética Portuguesa*. Lisboa: Imprensa Nacional de Lisboa.

Saltzman, Elliot. 1986. Task dynamic coordination of the speech articulators: A preliminary model. In H. Heuer & C. Fromm (eds.), *Generation and Modulation of Action Patterns*, Experimental Brain Research Series 15, 129–144. Springer-Verlag.

Saltzman, Elliot & Byrd, Dani. 2000. Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science* 19, 499–526.

Saltzman, Elliot & Munhall, Kevin. 1989. A dynamic approach to gestural patterning in speech production. *Ecological Psychology* 1(4), 333–382.

Saltzman, Elliot, Nam, Hosung, Krivokapic, Jelena & Goldstein, Louis. 2008. A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. In P. A. Barbosa, S. Madureira & C. Reis (eds.), *Speech Prosody*. Associação Luso-Brasileira de Ciências da Fala.

Sampson, Rodney. 1999. *Nasal Vowel Evolution in Romance*. Oxford University Press.

Santen, Jan van. 1994. Assignment of segmental duration in text-to-speech synthesis. *Computer Speech and Language* 8, 95–128.

Santen, Jan van, Shih, Chilin, Möbius, Bernd, Tzoukermann, Evelyne & Tanenblatt, Michael. 1997. Multilingual duration modeling. In G. Kokkinakis, N. Fakotakis & E. Dermatas (eds.), *European Conference on Speech Communication and Technology (Eurospeech)*, 2651–2654. Rhodes, Greece.

Santen, Jan van & Sproat, Richard. 1998. Introduction. In J. van Santen (ed.), *Multilingual Text-to-Speech: The Bell Labs Approach*, cap. 1, 1–6. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Savino, Michelina, Refice, Mario & Mitaritonna, Massimo. 2005. Which Italian do current systems speak? A first step towards pronunciation modelling of Italian varieties. In *Annual Conference of the International Speech Communication Association (Interspeech)*, 1961–1964. Lisbon, Portugal.

Schiller, Niels. 1998. The effect of visually masked syllables primes on the naming latencies of words and pictures. *Journal of Memory and Language* 39, 484–507.

Schourup, Lawrence. 1972. Characteristics of vowel nasalization. *Papers in Linguistics* 5, 530–548.

Schröder, Manfred & Trouvain, Jürgen. 2003. The German text-to-speech synthesis system Mary: a tool for research, development and teaching. *International Journal of Speech Technology* 6(4), 365–377.

Schröeder, Manfred. 1993. A brief history of speech synthesis. *Speech Communication* 13(1–2), 231–237.

Schweigert, Wendy A. 1994. *Research Methods & Statistics for Psychology*. Pacific Grove, California: Brooks/Cole Publishing Company.

Seara, Christine. 2000. *Estudo acústico-perceptual da nasalidade das vogais do português brasileiro*. Tese de Doutoramento, Universidade Federal de Santa Catarina.

Selkirk, Elisabeth. 1984. On the major class features and syllable theory. In M. Aronoff & R. T. Oehrle (eds.), *Language, Sound, and Structure. Studies Presented to Morris Halle by His Teacher and Students*, 107–136. Cambridge: The MIT Press.

Selkirk, Elisabeth. 1999. The syllable. In J. A. Goldsmith (ed.), *Phonological Theory. The Essential Readings*, cap. 17, 328–350. Blackwell Publishers.

Shadle, Christine & Damper, Robert. 2001. Prospects for Articulatory Synthesis: A position paper. In *ISCA Workshop on Speech Synthesis*. Pitlochry, Scotland.

Shaw, Jason & Gafos, Adamantios. 2008. C-center and syllabification in Moroccan Arabic. In *CUNY conference on the syllable*. City University of New York Graduate Center, New York, NY.

Shosted, Ryan. 2003. Nasal coda restoration in Brazilian Portuguese. In M. J. Solé, D. Recasens & J. Romero (eds.), *International Congress of Phonetic Sciences (ICPhS)*, 3037–3040. Barcelona: Universitat Autonoma.

Shosted, Ryan. 2006a. *The Aeroacoustics of Nasalized Fricatives*. PhD dissertation, University of California.

Shosted, Ryan. 2006b. Vocalic context as a condition for nasal coda emergence: aerodynamic evidence. *Journal of the International Phonetic Association* 36(1), 39–58.

Siciliano, Catherine, Williams, Geoff, Beskow, Jonas & Faulkner, Andrew. 2003. Evaluation of a multilingual synthetic talking face as a communication aid for the hearing impaired. In *European Conference on Speech Communication and Technology (Eurospeech)*, 131–134. Geneva, Switzerland.

Silva, Ana Isabel Mata da. 1987. Ditongos crescentes do português: análise acústica. In *Encontro da Associação Portuguesa de Linguística (APL)*, 379–400.

Silva, António Ricardo Trindade Vieira. 1995. *Análise de Fonemas Nasais da Língua Portuguesa*. Tese de mestrado, Universidade de Aveiro.

Silva, Denilson, Lima, Amaro de, Maia, Ranniery, Braga, Daniela, Moraes, João F. de, Moraes, João A. de & Resende, Fernando. 2006. A rule-based grapheme-phone converter and stress determination for Brazilian Portuguese natural language processing. In *International Telecommunications Symposium (ITS)*. Fortaleza, Brasil.

Silva, Luís Nuno Oliveira Rodrigues. 2001. *Desenvolvimento de um Sintetizador Articulatório*. Tese de Mestrado, Universidade de Aveiro, Aveiro, Portugal.

Silverman, Kim, Beckman, Mary, Pitrelli, John, Ostendorf, Mori, Wightman, Colin, Price, Patti, Pierrehumbert, Janet & Hirschberg, Julia. 1992. TOBI: A standard for labeling English prosody. In *International Conference on Spoken Language Processing (ICSLP)*, 867–870. Banff, Canada.

Simões, António. 1990. Predicting sound segment duration in connected speech: an acoustical study of Brazilian Portuguese. In *Workshop on Speech Synthesis*, 173–176. Autrans.

Simões, Flávio Olmos. 1999. *Implementação de um Sistema de Conversão Texto-Fala para o Português do Brasil*. Tese de mestrado, Universidade Estadual de Campinas - Faculdade de Engenharia Eléctrica e de Computação.

Solé, Maria-Josep. 1995. Spatio-temporal patterns of velopharyngeal action in phonetic and phonological nasalization. *Language and Speech* 38(1), 1–23.

Solé, Maria-Josep. 1999. The phonetic basis of phonological structure: The role of aerodynamic factors. In *Congress of Experimental Phonetics*, 77–94. Tarragona, Spain.

Solé, Maria-Josep. 2002. Aerodynamic characteristics of trills and phonological patterning. *Journal of Phonetics* 30, 655–688.

Solé, Maria-Josep. 2007. Compatibility of features and phonetic content. the case of nasalization. In *International Congress of Phonetic Sciences (ICPhS)*. Saarbrücken, Germany.

Solé, Maria-Josep & Ohala, John. 1991. Differentiating between phonetic and phonological processes: The case of nasalization. In *International Congress of Phonetic Sciences (ICPhS)*, 110–113. Aix-en-Provence.

Sondhi, Man Mohan & Schroeter, Juergen. 1987. A hybrid time-frequency domain articulatory speech synthesizer. *IEEE Trans. on Acoustics, Speech, and Signal Processing* ASSP-35(7), 955–967.

Sondhi, Mohan & Sinder, Daniel. 2005. Articulatory modeling: a role in concatenative text to speech synthesis. In S. Narayanan & Abeer Alwan (eds.), *Text to speech synthesis: New Paradigms and Advances*, 63–87. New Jersey: Pearson Education.

Sousa, Elizabeth Maria de. 1994. *Para a Caracterização Fonético-Acústica da Nasalidade no Português do Brasil*. Dissertação de mestrado em linguística, Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Brasil.

Sproat, R., Ostendorf, M. & Hunt, A. 1999. *The need for increased Speech Synthesis research*. Rel. Tec., NSF Speech Synthesis Workshop 1998.

Sproat, Richard (ed.). 1998. *Multilingual Text-to-Speech Synthesis: the Bell Labs Approach*. Kluwer.

Sproat, Richard & Fujimura, Osamu. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21, 291–311.

Stetson, R. H. 1951. *Motor Phonetics: A study of speech movement in action*. Amsterdam: North-Holland Publishing Company.

Stevens, Kenneth, Andrade, Amália & Viana, Maria do Céu. 1987. Perception of vowel nasalization in VC contexts: A cross-language study. *Journal of the Acoustical Society of America* 82, SI, S119 (A).

Stevens, Kenneth & Hanson, H. M. 2003. Production of consonants with a quasi-articulatory synthesizer. In *International Congress of Phonetic Sciences (ICPhS)*, vol. 1, 199–202. Barcelona, Spain.

Stevens, Kenneth, Kasowski, S. & Fant, Gunnar. 1953. An electrical analog of the vocal tract. *Journal of the Acoustical Society of America* 25, 734–742.

Stevens, Kenneth N. 1998. *Acoustic Phonetics*. Current Studies in Linguistics. MIT Press.

Stone, Maureen. 1999. Laboratory techniques for investigating speech articulation. In William J. Hardcastle & John Laver (eds.), *The Handbook of Phonetic Sciences*, 11–32. Blackwell.

Strevens, Peter. 1954. Some observations on the phonetics and pronunciation of modern Portuguese. *Revista do Laboratório de Fonética Experimental (de Coimbra)* II, 5–29.

Styger, Thomas & Keller, Eric. 1994. Formant synthesis. In Keller (1994), cap. 6, 109–128.

Taylor, Paul. 2005. Hidden markov models for grapheme to phoneme conversion. In *Annual Conference of the International Speech Communication Association (Interspeech)*. Lisbon, Portugal.

Taylor, Paul. no prelo. *Text-to-Speech Synthesis*. Cambridge University Press.

Teixeira, António. 2000. *Síntese Articulatória das Vogais Nasais do Português Europeu*. Tese de Doutoramento, Universidade de Aveiro.

Teixeira, António. 2001. *Base de Dados EMMA dos Sons Nasais do Português Europeu*. Relatório Técnico do Projecto PLP/PP/11222/1998, Síntese Articulatória do Português (SAP), financiado pela Fundação para a Ciência e Tecnologia através do programa PRAXIS XXI No. SAP 4/2001, Instituto de Engenharia Electrónica e Telemática de Aveiro (IEETA).

Teixeira, António. 2005. Speech Synthesis. In S. Pigeon, C. Swail, E. Geoffrois, C. Bruckner, D. van Leeuwen, C. Teixeira, O. Orman, Paul Collins, T. Anderson, J. Grieco & M. Zissman (eds.), *Use of Speech and Language Technology in Military Environments*. Research and Technology Organisation, North Atlantic Treaty Organisation (NATO).

Teixeira, António, Martinez, Roberto, Silva, Luís Nuno, Jesus, Luís, Príncipe, José & Vaz, Francisco. 2005. Simulation of human speech production applied to the study and synthesis of European Portuguese. *EURASIP Journal on Applied Signal Processing* 9, 1435–1448.

Teixeira, António, Oliveira, Catarina & Moutinho, Lurdes. 2006. A síntese de voz aplicada ao ensino das línguas. In *Encontro Internacional de Linguística Aplicada*, 199–213. Aveiro.

Teixeira, António & Vaz, Francisco. 2000. *A Suite of Tcl/Tk Programs for Perceptual Tests*. Relatório Técnico do Projecto PLP/PP/11222/1998, Síntese Articulatória do Português (SAP), financiado pela Fundação para a Ciência e Tecnologia através do programa PRAXIS XXI No. SAP 2/2000, Instituto de Engenharia Electrónica e Telemática de Aveiro (IEETA).

Teixeira, António, Vaz, Francisco & Príncipe, José. 1999. Influence of dynamics in the perceived naturalness of Portuguese nasal vowels. In John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel

Granville & Ashlee C. Bailey (eds.), *International Congress of Phonetic Sciences (ICPhS)*. San Francisco: University of California, Berkeley.

Teixeira, João Paulo, Freitas, Diamantino, Braga, Daniela, Barros, Maria João & Latsch, Vagner. 2001. Phonetic events from the labelling of the European Portuguese database for speech synthesis, FEUP/IPB-DB. In *European Conference on Speech Communication and Technology (Eurospeech)*, 1707–1710. Aalborg, Denmark.

Teixeira, João Paulo, Freitas, Diamantino, Gouveia, Paulo, Olaszy, Gábor & Németh, Géza. 1998. MULTIVOX - Conversor texto fala para português. In *Encontro para o Processamento Computacional da Língua Portuguesa Escrita e Falada (PROPOR)*. Porto Alegre, Brasil.

Teixeira, João Paulo Ramos. 2004. *A Prosody Model to TTS Systems*. Tese de Doutoramento, Faculdade de Engenharia da Universidade do Porto.

Teranishi, Ryunen & Umeda, Noriko. 1968. Use of pronouncing dictionary in speech synthesis experiments. In *International Congress on Acoustics*, B155–B158. Tokyo.

Teyssier, Paul. 1980. *Histoire de La Langue Portuguaise*. Paris: Presses Universitaires de France.

Tian, Jilei. 2004. Data-driven approaches for automatic detection of syllable boundaries. In *International Conference on Spoken Language Processing (ICSLP)*, 61–64. Jeju Island, Korea.

Trancoso, Isabel, Oliveira, Luís, Neto, João & Viana, Maria do Céu. 2000. Da escrita à fala - da fala à escrita. In *A Escrita das Escritas*. Lisboa: Fundação Portuguesa de Comunicações.

Trancoso, Isabel & Viana, Maria do Céu. 1995. Issues in the pronunciation of proper names. In *Workshop on Integration of Language and Speech*. Moscow, Russia.

Trancoso, Isabel & Viana, Maria do Céu. 1997. On the pronunciation mode acronyms in several european languages. In *European Conference on Speech Communication and Technology (Eurospeech)*, 573–576. Rhodes, Greece.

Trancoso, Isabel, Viana, Maria do Céu & Mascarenhas, Isabel. 1995. Léxicos de pronúncia: a experiência do projecto Onomastica. In *Encontro da Associação Portuguesa de Linguística (APL)*, vol. I, 241–263. Lisboa.

Trancoso, Isabel, Viana, Maria do Céu & Silva, Fernando. 1994. On the pronunciation of common lexica and proper names in European Portuguese. In *Onomastica Research Colloquium*. London.

Turvey, Michael. 1977. Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (eds.), *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*. Hillsdale: Lawrence Erlbaum Associates.

Turvey, Michael. 1990. Coordination. *American Psychologist* 45, 938–953.

Vaissière, Jacqueline. 1988. Prediction of velum movement from phonological specifications. *Phonetica* 45, 122–139.

Vaissière, Jacqueline. 1995. Nasalité et phonétique. In *Colloque sur le voile pathologique*. Lyon.

Veloso, João. 1994. Algumas notas sobre a classificação de /t/ e /d/ em português. dinâmica articulatória e funcionalidade linguística. *Revista da Faculdade de Letras da Universidade do Porto-Línguas e Literaturas* XI, 131–146.

Veloso, João. 1999. *Na Ponta da Língua. Exercícios de Fonética do Português*. Porto: Granito, Editores e Livreiros.

Veloso, João. 2006. Reavaliando o estatuto silábico das sequências obstruinte+lateral em português europeu. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada* 22(1), 127–158.

Veloso, João. 2008. Coda-avoiding: Some evidence from Portuguese. *Romanitas, Lenguas e Literaturas Romances* 3(1).

Veloso, João Manuel Pires Silva Almeida. 2003. *Da Influência do conhecimento ortográfico sobre o conhecimento fonológico. Estudo Longitudinal de um grupo de crianças falantes nativas do Português Europeu*. Tese de doutoramento, Faculdade de Letras da Universidade do Porto.

Viana, Aniceto dos Reis Gonçalves. 1903. Portugais. phonétique et phonologie. morphologie. textes. In W. Vietor (ed.), *Skizzen lebender Sprachen*. Leipzig: Teubner.

Viana, Aniceto dos Reis Gonçalves. 1973a. Essai de phonétique et de phonologie de la langue portuguaise d'aprés le dialecte actuel de lisbonne. In Luís F. Lindley Cintra & José A. Peral Ribeiro (eds.), *Estudos de Fonética Portuguesa*, 153–250. Lisboa: Imprensa Nacional.

Viana, Aniceto dos Reis Gonçalves. 1973b. Exposição da pronúncia normal portuguesa para uso de nacionais e estrangeiros. In Luís F. Lindley Cintra & José A. Peral Ribeiro (eds.), *Estudos de Fonética Portuguesa*, 153–250. Lisboa: Imprensa Nacional.

Viana, Maria do Céu, d'Andrade, Ernesto, Oliveira, Luís & Trancoso, Isabel. 1991. Ler_PE: Um utensílio para o estudo da ortografia do Português. In *Encontro da Associação Portuguesa de Linguística (APL)*, 474–489. Lisboa.

Viana, Maria do Céu, Trancoso, Isabel, Ribeiro, Carlos, Andrade, Amália & d'Andrade, Ernesto. 1993. The relationship between spelled and spoken Portuguese: Implications for speech synthesis and recognition. In *European Conference on Speech Communication and Technology (Eurospeech)*. Berlim.

Viana, Maria do Céu, Trancoso, Isabel, Silva, Fernando, Marques, G. C., d'Andrade, Ernesto & Oliveira, Luís Caldas. 1996. Sobre a pronúncia de nomes próprios, siglas e acrónimos em português europeu. In *Congresso Internacional sobre o Português*, 481–517. Lisboa.

Vigário, Marina. 2003. *The Prosodic Word in European Portuguese*. Berlim: Mouton de Gruyter.

Vigário, Marina & Falé, Isabel. 1993. A sílaba no português fundamental: uma descrição e algumas considerações de ordem teórica. In *Encontro da Associação Portuguesa de Linguística (APL)*, 465–478. Coimbra.

Vigário, Marina, Martins, Fernando & Frota, Sónia. 2005. Frequências no português: a ferramenta FreP. In *Encontro da Associação Portuguesa de Linguística (APL)*. Lisboa.

Vigário, Marina, Martins, Fernando & Frota, Sónia. 2006. A ferramenta FreP e a frequência de tipos silábicos e classes de segmentos no português. In *Encontro da Associação Portuguesa de Linguística (APL)*, 675–687. Porto.

Vihman, Marilyn May. 1996. *Phonological development : the origins of language in the child*. Cambridge: Blackwell.

Véronis, Jean, Cristo, Philippe Di, Courtois, Fabienne & Chaumette, Cédric. 1998. A stochastic model of intonation for text-to-speech synthesis. *Speech Communication* 26(4), 233–244.

Walker, Rachel. 2000. *Nasalization, Neutral Segments and Opacity Effects*. New York: Routledge.

Warren, Donald, Dalston, Rodger & Mayo, Robert. 1994. Hypernasality and velopharyngeal impairment. *The Cleft Palate-Craniofacial Journal* 31(4), 257–262.

Weerasinghe, Ruvan, Wasala, Asanka & Gamage, Kumudu. 2005. A rule based syllabification algorithm for Sinhala. In Robert Dale, Kam-Fai Wong, Jian Su & Oi Yee Kwong (eds.), *Natural Language Processing - IJCNLP*, 438–449. Berlin / Heidelberg: Springer.

West, Paula. 1999. The extent of coarticulation of English liquids: An acoustic & articulatory study. In John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville & Ashlee C. Bailey (eds.), *International Congress of Phonetic Sciences (ICPhS)*. San Francisco: University of California, Berkeley.

West, Paula. 2000. Long-distance coarticulatory effects of British English /l/ and /r/: An EMA, EPG and acoustic study. In *Seminar on Speech Production: Models and Data*, 105–108. Seeon.

Westbury, John, Turner, Greg & Dembowsky, Jim. 1994. *X-ray microbeam speech production database user's handbook. Version 1.0*. Rel. Tec., Waisman Center on Mental Retardation & Human Development, University of Wisconsin.

Wetzels, Leo. 1997. The lexical representation of nasality in Brazilian Portuguese. *Probus* 9(2), 203–232.

Wrench, Alan. 1999. An investigation of sagittal velar movement and its correlation with lip, tongue and jaw movement. In John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville & Ashlee C. Bailey (eds.), *International Congress of Phonetic Sciences (ICPhS)*, 435–438. San Francisco: University of California, Berkeley.

Wrench, Alan & Scobbie, James. 2003a. An articulatory investigation of word-final /l/ and /l/-sandhi in three dialects of English. In M. J. Solé, D. Recasens & J. Romero (eds.), *International Congress of Phonetic Sciences (ICPhS)*, 1871–1874. Barcelona.

Wrench, Alan & Scobbie, James. 2003b. Categorising vocalisation of English /l/ using EPG, EMA and ultrasound. In *International Seminar on Speech Production*, 314–319. Sydney.

Yarowsky, David. 1997. Homograph disambiguation in text-to-speech synthesis. In J. P. H. van Santen, R. W. Sproat, J. P. Olive & J. Hirschberg (eds.), *Progress in Speech Synthesis*, 157–172. New York: Springer.

Yvon, Francois. 1994. Règles de transcription graphème-phonème pour la prononciation automatique de sigles. *Lynx* 30, 153–166.

Zellner, Brigitte. 1998. *Caractérisation et prédiction du débit de parole en français. Une étude de cas*. Thèse de doctorat, Faculté des Lettres, Université de Lausanne.

Zerling, Jean Pierre. 1984. Phénomenes de nasalité et de nasalization vocaliques: Étude cinéradiographique pour deux locuteurs. *Travaux de l'Institut de Phonétique de Strasbourg* 16, 241–266.

Zerling, Jean Pierre. 2000. Structure syllabic et morphologique des mots à caractères onomatopéique et répétitif en français. *Travaux de l'Institut de Phonétique de Strasbourg* 30, 115–162.

Zeroual, Chakir, Gafos, Adamantios & Hoole, Phil. 2008. Degree of temporal overlap within Moroccan Arabic stop-stop clusters and its relation to place order and voicing. In *Workshop Consonant Clusters and Structural Complexity*. Munich, Germany.

Apêndice A

Especificação dos gestos no TADA

Tabela A.1: Especificação dos gestos no TADA. Cada uma das colunas representa: a variável do tracto (TV), o ponto e grau de constrição (Constr), o alvo (*target*), o *alpha*, o peso (*weights*) dos articuladores (colunas 5-14) e os segmentos do inglês (em ARPABET).

TV	Constr	Target	Alpha	LX	JA	UH	LH	CL	CA	TL	TA	NA	GW	Applicable Segments
LA	CLO	-2	100	.	8	5	1	.	.	.	.	.	.	B, P, M
LA	CRIT	1	10	.	8	5	1	.	.	.	.	.	.	V, F
LA	NAR	2	1	.	8	5	1	.	.	.	.	.	.	W, R, AO, UW, UH, OW, ER
LA	V	8	1	.	8	5	1	.	.	.	.	.	.	IY
LA	REL	11	1	.	8	5	1	.	.	.	.	.	.	B, P, M, V, F, W, R
LP	DENT	8	1	1	.	.	.	.	.	.	.	.	.	V, F
LP	PRO	12	1	1	.	.	.	.	.	.	.	.	.	AO, UW, UH, OW
LP	REL	9.11	1	1	.	.	.	.	.	.	.	.	.	V, F
TTCL	DENT	40	1	.	32	.	.	32	32	1	1	.	.	DH, TH
TTCL	ALV	56	1	.	32	.	.	32	32	1	1	.	.	D, T, N, Z, S, L
TTCL	ALVPAL	60	1	.	32	.	.	32	32	1	1	.	.	ZH, SH, JH, CH
TTCL	PAL	80	1	.	32	.	.	32	32	1	1	.	.	R, ER
TTCL	REL	24	1	.	32	.	.	32	32	1	1	.	.	D, T, N, DH, TH, Z, S, ZH, SH, JH, CH, L
TTCD	CLO	-2	100	.	32	.	.	32	32	1	1	.	.	D, T, N, JH, CH
TTCD	CRIT	1	10	.	32	.	.	32	32	1	1	.	.	DH, TH, Z, S, ZH, SH
TTCD	NAR	2	1	.	32	.	.	32	32	1	1	.	.	R, L, ER
TTCD	REL	11	1	.	32	.	.	32	32	1	1	.	.	D, T, N, DH, TH, Z, S, ZH, SH, JH, CH, L
TBCL	PAL	95	100	.	10	.	.	1	1	.	.	.	.	ZH, SH, JH, CH, Y, IY, IH, EY, EH
TBCL	VEL	100	10	.	10	.	.	1	1	.	.	.	.	G, K, NX, Z, S
TBCL	UVU	125	10	.	10	.	.	1	1	.	.	.	.	W, L, UW, UH, AH, AX
TBCL	UVUPHAR	150	1	.	1	.	.	1	1	.	.	.	.	OW
TBCL	PHAR	180	1	.	1	.	.	1	1	.	.	.	.	AE, AA, AO
TBCD	CLO	-2	100	.	10	.	.	1	1	.	.	.	.	G, K, NX,
TBCD	CRIT	1	100	.	10	.	.	1	1	.	.	.	.	
TBCD	NAR	2	100	.	10	.	.	1	1	.	.	.	.	ZH, SH, JH, CH, Y, W, L
TBCD	WIDE	10	10	.	10	.	.	1	1	.	.	.	.	Z, S
TBCD	REL	6	1	.	10	.	.	1	1	.	.	.	.	G, K, NX
TBCD	V	10	1	.	1	.	.	1	1	.	.	.	.	IY, IH, EY, EH, AE, AA, AO, UW, UH, OW, AH, AX
VEL	CLO	-0.1	0	.	.	.	.	.	.	.	.	1	.	B, P, D, T, G, K, V, F, DH, TH, Z, S, ZH, SH,
VEL	WIDE	0.2	0	.	.	.	.	.	.	.	.	1	.	M, N, NX
GLO	CLO	-0.5	100	.	.	.	.	.	.	.	.	.	1	Q
GLO	WIDE	0.4	0	.	.	.	.	.	.	.	.	.	1	P, T, K, F, TH, S, SH, CH, HH

Apêndice B

Coordenação dos osciladores no TADA

```
/coupling/
% C = (clo | crt | nar | voc)
% CNS = (clo | crt | nar)
% OBS = (clo | crt)
% onset
% C-C coupling
ONS_OBS ONS_CNS 1 1 180 % anti-phase relation in onset clusters
% C-V coupling
ONS_CNS* V 1 1 0 % all CNS gestures synchronous with V
ONS_H V 1 1 0 % GLO synchronous with V, if not coupled to CNS
% within-C coupling
ONS_CNS ONS_REL 1 1 180 % REL is anti-phase
                        % with respect to Constriction
ONS_CRT ONS_H 1 1 0 % GLO gesture is synchronous with frics
ONS_CLO ONS_H 1 1 90 % else GLO gesture is delayed for stops
ONS_CLO ONS_N 1 1 0 % VEL gesture synchronous wih oral constr.
ONS_VOC ONS_NAR 1 1 0 % VOC gesture of /r/, /l/ sychronous
                      % with primary NAR constriction
% vowel
V_RND V 1 1 0 %rounding synchronous with V tongue constr.
% coda
% C-C coupling
COD_CNS COD_CNS 1 1 180 % C in coda are phased 180 degrees
% V-C coupling
```

```
V COD_CNS 1 1 180 % first coda CNS anti-phase to V

% within-C coupling
COD_CNS COD_REL 1 1 180 % REL is anti-phase
                        % with respect to Constriction
COD_CLO COD_H 1 1 90 % GLO gesture is delayed for stops
COD_CRT COD_H 1 1 0 % else GLO gesture is synchronous with frics
COD_N COD_CNS 1 1 180 % VEL gesture anti-phase to oral constr.
COD_VOC COD_NAR 1 1 180 % VOC gesture anti-phase to NAR constr.
/cross-syllable/
COD_CNS ONS_CNS 1 1 180 % applies if boundary is C$C
V ONS_CNS 1 1 180 % applies if boundary is V$C
COD_CNS V 1 1 0 % applies if boundary is C$V
V V 1 1 360 % applies if boundary is V$V
/cross-word/
COD_REL ONS_CNS 1 1 0 % applies if boundary is C#C
V ONS_CNS 1 1 180 % applies if boundary is V#C
COD_CNS V 1 1 0 % applies if boundary is C#V
V V 1 1 360 % applies if boundary is V#V
```

Apêndice C

Corpus EMMA de sons nasais

C.1 Vogais (orais e nasais) isoladas - taxa normal

[i e ɛ u o ɔ ɐ ɨ a]
[ɐ̃ ẽ ĩ õ ũ]

C.2 Sequências VCV - taxa normal

C= [p t k b d g f s ʃ v z ʒ l ʎ m n ɾ r]
V= [ɐ̃ ẽ ĩ õ ũ a i u]

C.3 Vogais nasais em diferentes posições na palavra - taxa normal e rápida

[ɐ̃pɐ] três vezes ... diz [pɐ̃pɐ] três vezes ... diz [pɐ̃]
[ẽpɐ] três vezes ... diz [pẽpɐ] três vezes ... diz [pẽ]
[õpɐ] três vezes ... diz [põpɐ] três vezes ... diz [põ]
[ũpɐ] três vezes ... diz [pũpɐ] três vezes ... diz [pũ]
[ĩpɐ] três vezes ... diz [pĩpɐ] três vezes ... diz [pĩ]
[ɐ̃bɐ] três vezes ... diz [bɐ̃bɐ] três vezes ... diz [bɐ̃]
[ẽbɐ] três vezes ... diz [bẽbɐ] três vezes ... diz [bẽ]
[õbɐ] três vezes ... diz [bõbɐ] três vezes ... diz [bõ]
[ũbɐ] três vezes ... diz [bũbɐ] três vezes ... diz [bũ]
[ĩbɐ] três vezes ... diz [bĩbɐ] três vezes ... diz [bĩ]
[ɐ̃tɐ] três vezes ... diz [tɐ̃tɐ] três vezes ... diz [tɐ̃]
[ẽtɐ] três vezes ... diz [tẽtɐ] três vezes ... diz [tẽ]

[õtɐ] três vezes ... diz [tõtɐ] três vezes ... diz [tõ]
[ũtɐ] três vezes ... diz [tũtɐ] três vezes ... diz [tũ]
[ĩtɐ] três vezes ... diz [tĩtɐ] três vezes ... diz [tĩ]
[ẽdɐ] três vezes ... diz [dẽdɐ] três vezes ... diz [dẽ]
[ẽdɐ] três vezes ... diz [dẽdɐ] três vezes ... diz [dẽ]
[õdɐ] três vezes ... diz [dõdɐ] três vezes ... diz [dõ]
[ũdɐ] três vezes ... diz [dũdɐ] três vezes ... diz [dũ]
[ĩdɐ] três vezes ... diz [dĩdɐ] três vezes ... diz [dĩ]
[ẽfɐ] três vezes ... diz [fẽfɐ] três vezes ... diz [fẽ]
[ẽfɐ] três vezes ... diz [fẽfɐ] três vezes ... diz [fẽ]
[õfɐ] três vezes ... diz [fõfɐ] três vezes ... diz [fõ]
[ũfɐ] três vezes ... diz [fũfɐ] três vezes ... diz [fũ]
[ĩfɐ] três vezes ... diz [fĩfɐ] três vezes ... diz [fĩ]

C.4 NVN - taxa normal

diz [nin] três vezes ... diz [mim] três vezes
diz [nĩn] três vezes ... diz [mĩm] três vezes
diz [nan] três vezes ... diz [mam] três vezes
diz [nãn] três vezes ... diz [mãm] três vezes
diz [nɐn] três vezes ... diz [mɐm] três vezes
diz [nẽn] três vezes ... diz [mẽm] três vezes
diz [nun] três vezes ... diz [mum] três vezes
diz [nũn] três vezes ... diz [mũm] três vezes

C.5 Vogal nasal vs consoante nasal vs oclusiva - taxa normal e rápida

[kẽpɐ] [kɐmɐ] [kapɐ]
[kẽtɐ] [kɐnɐ] [katɐ]
[bẽku] [bɐnu] [baku]
[pẽtɐ] [penɐ] [petɐ]
[tẽtɐ] [tenɐ] [tetɐ]
[sĩtu] [sĩnu] [sĩtu]
[põpɐ] [pomo] [popɐ]
[ʃũbu] [ʃumɐ] [ʃubɐ]

C.6 NṼ versus NV - taxa normal e rápida

diz [mẽtu] três vezes ... diz [matu] três vezes ...

diz [mẽtu] três vezes ... diz [metu] três vezes ...
diz [mĩtu] três vezes ... diz [mitu] três vezes ...
diz [mũtu] três vezes ... diz [mutu] três vezes ...
diz [mõtu] três vezes ... diz [motu] três vezes ...
diz [ɐmẽtɨ] três vezes ... diz [suɲẽdu] três vezes ... diz [fɪrnẽdu] três vezes ...
diz [fumẽtu] três vezes ... diz [kiɲẽtuʃ] três vezes ... diz [kõtinẽtɨ] três vezes ...
diz [fɐmĩtu] três vezes ... diz [unĩdu] três vezes ...
diz [imũdu] três vezes ... diz [inũdu] três vezes ...
diz [diʃmõtu] três vezes ...

C.7 Palavras - taxa normal e rápida

diz [amʃtɛr] três vezes ... diz [aknɨ] três vezes
diz [tɛkniku] três vezes ... diz [knɛmidɨ] três vezes
diz [pnɐw] três vezes ... diz [ipnɔzɨ] três vezes
diz [ɐpnɐjɐ] três vezes ... diz [pnewmatiku] três vezes
diz [tmɛzɨ] três vezes ... diz [ʁɪtmu] três vezes
diz [ɐtmuʃfɛrɐ] três vezes ... diz [pɐlĩpseʃtu] três vezes

C.8 Extra - taxa normal

[dɐmɐ] [pumɐ] [pomu] [simu]
[pɐnu] [sonu] [senɐ] [pinu] [tunɐ]
[bɐɲu] [poɲu] [puɲu] [tiɲɐ]

Apêndice D

Consentimento informado (EMMA)

DOCUMENT D'INFORMATION et CONSENTEMENT DE PARTICIPATION

Remis aux personnes sollicitées pour participer à une recherche

Investigateur principal :
Responsable scientifique :
Intervenants :

Titre du projet de recherche :
Dynamique de la nasalité.
Émergence et phonologisation des voyelles nasales.

Lieu de l'enregistrement :
Institut de la Communication Parlée (ICP)
Université Stendhal, 1180 avenue Centrale BP 25
38040 GRENOBLE CEDEX 9

Objectif scientifique général :
Nous effectuons cette recherche dans le but d'obtenir des informations relatives aux mouvements des articulateurs durant la production de parole pour des personnes parlant français et/ou portugais, en utilisant la technique de l'articulographe électromagnétique. Cette technique permet d'enregistrer les mouvements des articulateurs : lèvres, langue, mâchoire inférieure et velum, à l'aide de petites bobines collées sur la muqueuse. A l'aide d'un casque fixé sur la tête, les coordonnées dans le plan médio-sagittal de chaque bobine sont extraites et permettent de mesurer le mouvement (trajectoire, vitesse, amplitude) des articulateurs de la parole (la fréquence d'échantillonnage peut aller jusqu'à 1 kHz). Nous nous proposons d'utiliser cette technique de mesure pour suivre les mouvements du velum lors de la production de voyelles et de consonnes nasales.

But de l'expérimentation :
De nombreuses études attribuent l'émergence des voyelles nasales à la propagation, le plus souvent régressive, de la nasalité depuis une consonne nasale. Cette assimilation, réalisée par un abaissement anticipé du velum, serait favorisée par certaines caractéristiques de la voyelle comme la durée, le contexte segmental ou la position dans le mot, la phrase. Une étude préliminaire a montré que le geste articulatoire était d'une amplitude beaucoup plus grande pour les voyelles nasales que pour les consonnes nasales. Une simple anticipation du geste articulatoire de la consonne nasale ne suffirait pas à produire suffisamment de nasalité dans la voyelle pour être perçue comme nasale. Les objectifs de ce projet sont d'étudier les conditions qui permettent de favoriser non seulement l'anticipation mais également l'amplification du geste articulatoire de la nasalité de la consonne vers la voyelle.

Méthodologie :
L'enregistrement avec l'articulographe électromagnétique (EMA) se déroulera à l'Institut de la Communication Parlée (ICP) et durera environ 1h30. Vous serez assis avec un casque maintenu sur la tête, casque sur lequel sont montés trois solénoïdes émettant chacune un champ électromagnétique d'une intensité moyenne de 35µT. Le champ magnétique émis par chaque solénoïde induit un faible courant électrique dans les bobines collés sur les articulateurs, permettant ainsi d'en déduire sa distance. Par un processus de triangularisation, les coordonnées de la bobine dans le plan médio-sagittal sont donc extraites après analyse du courant traversant la bobine en question.

Neuf bobines sont collées à des positions précises sur le sujet. La colle utilisée est de la famille des cyanoacrylates qui est une colle médicale utilisée par les dentistes. Pour cette expérimentation, nous avons choisi de fixer les bobines de la façon suivante :

- une bobine est collée sur l'arrête du nez
- une bobine est collée sur la gencive supérieure au milieu des deux incisives
- une bobine est collée sur la mâchoire inférieure, au niveau de deux incisives
- trois bobines sont collées sur la langue respectivement sur le bout de la langue (+ 1cm), le milieu et la partie arrière
- une bobine est collée sur la partie inférieure du velum.

Une fois les bobines collées, l'expérimentateur positionne le casque sur la tête du sujet et les enregistrements peuvent commencer.
Dans un premier temps, des enregistrements de références sont demandés : incisives jointes, plan médian, position de repos,…
Puis l'enregistrement du corpus peut commencer. Vous aurez un certain nombre de mots (ou non-mots) et de phrases à prononcer dans un ordre qui vous sera indiqué par un intervenant, parfois avec des débits d'élocution différent. Si une bobine se décolle durant l'enregistrement, on peut arrêter là les enregistrements ou la recoller pour poursuivre. Cette décision sera prise en temps voulu avec vous.
Vous pourrez à tout moment arrêter l'enregistrement si vous éprouvez une gêne quelconque et on vous enlèvera tout l'appareillage.

Caractéristiques de la technique d'EMA :

Avantages : L'articulographe électromagnétique permet d'obtenir une grande quantité de mesures sur les mouvements articulatoires de façon non-invasive, aucune substance ne sera injectée dans votre corps.
Contraintes : Vos mouvements sont très limités durant toute la phase de l'enregistrement, le casque vous maintient la tête dans la même position durant toute la période de l'enregistrement.
Contre-indications :
* intolérance à la colle, pour minimiser le risque, un test sera fait sur un endroit localisé à l'intérieur de la bouche plusieurs heures avant l'enregistrement final
* personne enceinte
* existence d'une affection sévère sur le plan général : cardiaque, respiratoire, hématologique, rénale, hépatique, cancéreuse
* présence intracorporelle d'objets ferromagnétiques (implants métalliques, pacemaker, éclats d'obus, plombs de chasse…)
* sujet présentant une pathologie psychiatrique patente

CONSENTEMENT à l'étude

Je soussigné…Nuno da Silva Teixeira…donne mon accord pour participer à l'étude