

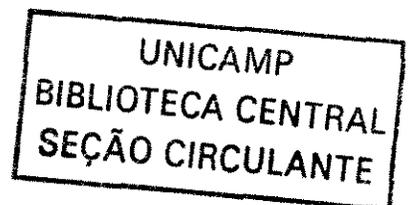
Beatriz Raposo de Medeiros

**Descrição comparativa de aspectos fonético-acústicos  
selecionados da fala e do canto em português brasileiro.**

Tese apresentada ao Curso de Lingüística do  
Instituto de Estudos da Linguagem da  
Universidade Estadual de Campinas como  
requisito parcial para a obtenção do título de  
Doutor em Lingüística.

Orientadora:  
Profª. Dra. Eleonora Cavalcante Albano

Instituto de Estudos da Linguagem  
UNICAMP  
Campinas  
2002



20030/575

UNIDADE	BC
Nº CHAMADA	UNICAMP M467d
V	EX
TOMBO BC/	52082
PROC.	16-124103
C	<input type="checkbox"/>
D	<input checked="" type="checkbox"/>
PREÇO	R\$ 11,00
DATA	12/02/03
Nº CPD	

CM0017B006-7

BIB ID 27623P

## FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA IEL - UNICAMP

M467d

Medeiros, Beatriz Raposo de

Descrição comparativa de aspectos fonético-acústicos selecionados da fala e do canto em português brasileiro / Beatriz Raposo de Medeiros. - - Campinas, SP: [s.n.], 2002.

Orientador: Eleonora Cavalcante Albano

Tese (doutorado) - Universidade Estadual de Campinas, Instituto de Estudos da Linguagem.

1. Fonética. 2. Acústica. 3. Fala. 4. Canto. I. Albano, Eleonora Cavalcante. II. Universidade Estadual de Campinas. Instituto de Estudos da Linguagem. III. Título.

Este exemplar e a redação final da tese  
defendida por Beatriz Raposo de  
Medeiros

e aprovada em \_\_\_\_\_  
25 / 11 / 2002. \_\_\_\_\_  
\_\_\_\_\_

Banca Examinadora

Prof. Dr. César Reis

Prof. Dr. Leonardo Fuks

Prof. Dr. Plínio Almeida Barbosa

Prof. Dr. Wilmar da Rocha D'Angelis

Suplentes

Prof. Dra. Ana Luiza Pinto Navas

Prof. Dr. Jonatas Manzolli

Para **Catarina**.  
pulsou pequenina em mim,  
abrigou-se nas minhas águas,  
transbordou.  
E ainda transbordando, permanece o mais profundo  
e maior amor da minha vida.

## AGRADECIMENTOS

À Eleonora Cavalcante Albano, minha orientadora, que indicou, no caminho dos sons, passagens antes desconhecidas para mim; fascinantes umas, outras ainda por explorar, e por isso instigantes.

Ao Leonardo Couto F. de Oliveira, colega e amigo, que sem dúvida, me ajudou imensamente por todo o percurso, e a quem dedico uma admiração e um carinho especiais.

À minha mãe Mariza e meus irmãos, Marta, Renato, André e Sergio que, - cada qual a seu modo - adotaram minha maternidade e me ajudaram a chegar ao fim deste trabalho.

Ao CNPq, pela bolsa de doutorado, processo 142707/98-8.

Aos meus **muito queridos** colegas do Lafape, pela solidariedade, troca de idéias, cafés e por todo apoio que sempre me deram.

Aos demais colegas do IEL, inesquecíveis, como Claudia Wanderley e Marina Augusto.

Ao Prof. Edson Françaço, por me apresentar a uma rede neural e por todo apoio com as dúvidas microsófticas.

Ao Prof. Plinio Almeida Barbosa, por sugestões sempre precisas e valiosas para este trabalho.

Ao Prof. Wilmar R. D'Angelis, por suas leituras atentas de meus trabalhos.

À Profa. Ana Luiza Navas, pela doçura e competência reunidas numa só pessoa.

Ao Paulo Rehder, pela estatística e por me ensinar o que dizem as curvas matemáticas.

À Wilma F. de Souza e à Carmen Gambetta, secretárias do Lafape, cujo apoio foi para além do aspecto burocrático, pelo carinho com que sempre me trataram.

Ao Luís Antônio Santos, pelo apoio na diagramação dos originais.

Ao Prof. César Reis, por sua disponibilidade em participar da banca de defesa da tese.

Ao Prof. Leonardo Fuks, que acompanhou com entusiasmo o trabalho desde seu início, reunindo sempre a recepção calorosa e o rigor científico na sua interlocução comigo.

Ao pessoal do NICS (Núcleo Interdisciplinar de Comunicação Sonora), da Unicamp, pelas imagens em vídeo.

Ao pessoal do Setor de Informática do IEL, Unicamp.

À Josana N. Badaró Leme e a todo pessoal da escola Mamãe Canguru, pelos cuidados diários com a Catarina.

À Maria Flávia Barbosa, pela agudeza e capacidade de compreender o mundo.

À Kátia Rossini, pela sensibilidade.

À Patrícia Salles, pela amizade longa e sólida.

À Doutora Lígia Capellari, pelas cervejas e pelos antibióticos.

À Sueli Cusato, amiga certa nas horas incertas.

Às cantoras Adriana Kayama, Carla Mendes, Cláudia Haberman, Márcia Guimarães e Sandra Gebram, vozes cheias de luz.

À Nair C. P. Correa Silva, pela sabedoria.

Ao Dante Ozzetti, por sua música.

E por fim, um agradecimento, que pode tomar a forma de uma homenagem, ao meu pai Marcos R. de Oliveira Raposo de Medeiros, já falecido, cujo amor pela música foi um legado insuperável, tendo sido, com a voz materna, o primeiro indicativo no caminho dos sons.

## Homenagem a Mário de Andrade

Acho que homenagear Mário de Andrade com essa tese é uma coisa pertinente. Assim, escolhi para epígrafe de cada capítulo, trechos do texto *Os Compositores e a Língua Nacional*, dos *Anais do Primeiro Congresso da Língua Nacional Cantada*, (1938) como forma de homenagear alguém cujo gênio permite dizer, assim, como quem estivesse conversando, aquilo que os outros ainda estão por descobrir no rigor da ciência. Acho que se hoje vivo fosse, e lesse meu trabalho, Mário de Andrade diria: “Eu já sabia que ia dar nisso”.

## RESUMO

Este trabalho, assim como diz o título da tese, é uma comparação de aspectos fonético-acústicos entre fala e canto. Os aspectos selecionados são: estrutura temporal, padrão formântico e co-produção entre consoantes e vogais. Para cada aspecto elaboramos um estudo, constituindo-se, assim, a tese, de três estudos.

Inicialmente questionamos o mito da má dicção existente no meio musical de que os cantores e, sobretudo, as cantoras cantam sem proporcionar inteligibilidade ao texto cantado. Propomos verificar as semelhanças e diferenças existentes entre o que chamamos de duas modalidades de fala, a fala *stricto sensu* e a fala cantada, à luz da descoberta de Sundberg (1977), a fim, de levantar hipóteses do ponto de vista da produção sobre a inteligibilidade. A descoberta de Sundberg, que focou seu experimento em notas agudas produzidas por um soprano, é a de que, quanto mais elevada é a frequência fundamental da nota musical, por exemplo 700 Hz, tanto mais a cantora abaixa a mandíbula, fazendo com que o primeiro formante da vogal cantada pareie com o primeiro harmônico da nota musical ou frequência de fonação, como diz o autor sueco.

Para a elaboração do experimento, levou-se em conta a canção erudita brasileira e uma característica sua própria, que é a de ser composta, de modo geral, em regiões de frequência média, ou mesmo baixa. Assim, escolhemos uma canção de Francisco Mignone, a Cantiga de Ninar, para soprano, e destacamos uma frase para servir de frase-veículo, de tessitura grave-média. Introduzimos o logatoma /la<sup>1</sup>CV/ na frase, sem alterar sua organização rítmica e melódica, como se pode ver com detalhes na Introdução, e a partir da gravação e digitalização das frases cantadas por cinco sopranos brasileiros, pudemos realizar as medidas acústicas necessárias. Foram elas: medida da duração dos segmentos do logatoma e medida dos três primeiros formantes da vogal-alvo da sílaba CV. Nesta sílaba, a consoante era [p,t,k] e as vogais eram [a,e,i,o,u], orais, em posição acentual tônica.

Para realizar nossas hipóteses explicativas a respeito das diferenças encontradas entre fala e canto, partimos dos resultados acústicos e, a partir deles, respaldados pela Teoria Acústica da Produção de Fala (Fant, 1960) interpretamos as manobras articulatórias do canto e chegamos à conclusão geral de que estas estão submetidas a pelo menos duas restrições musicais que são: a afinação e o volume da nota. No entanto, como esta nota é produzida sobre uma vogal, para que essa última não seja totalmente descaracterizada, há uma negociação entre a fala e o canto, com o fim de manter certa inteligibilidade do texto. Nesta negociação verifica-se – esse é nosso achado – o seguinte: as consoantes são encurtadas no canto, os três primeiros formantes pareiam com os harmônicos da nota e não há fenômeno de coarticulação no canto. Propomos que esta negociação, que implica capacidades cognitivas lingüísticas e musicais, pode ser estudada – num outro momento – à luz da Fonologia Acústico-Articulatória, a FAAR, (Albano, 2001).

Adiamos propositalmente estudos perceptuais para o futuro, por sua complexidade e morosidade. Nesta tese, nosso objetivo era uma descrição fonético-acústica, do ponto de vista da produção das duas modalidades investigadas, a fim lançar uma luz, apenas inicial, na questão da inteligibilidade da canção erudita.

## ABSTRACT

In this dissertation three acoustic phonetic aspects were selected in order to compare speech to singing. The selected aspects were temporal structure, formant pattern and coproduction between consonants and vowels; and for each aspect a study has been elaborated, thus consisting the thesis the sum of this three studies.

First we discuss the bad diction myth, widespread among musicians, which propagates that singers and especially female singers sing in an unintelligible way. Our purpose in this investigation is to verify similarities and differences between what we call the two speech types, speech *stricto sensu* and sung speech, following Sundberg's finding (1977) in order to raise hypotheses from the production's point of view. Sundberg's finding, which focused treble notes produced by a soprano in his experiment, is that the higher the fundamental frequency of the musical note, the more the mandible is lowered, and as a result the first formant of the sung vowel matches the first harmonic of phonation frequency.

To elaborate the experiment we focused on the Brazilian classical song and one of its characteristics that is to be composed, in general, in middle frequency regions, or even low frequency. A song by Francisco Mignone, *Cantiga de Ninar*, for soprano was chosen, from which we extracted a phrase to be the carrier-phrase in a low mean frequency region. The nonsense word /la<sup>1</sup>CV/ was introduced in the carrier phrase, without altering its rhythmic and melodic structure. Once the spoken and sung phrases produced by five Brazilian sopranos were recorded and digitized, we were able to perform the acoustic measurements which were the nonsense word segment duration and the first three formants of the vowel target in the CV syllable. In this syllable the consonant was [p,t,k] and vowels were [a, ε, e, i, ɔ, o, u], oral, in stressed position.

To raise our explaining hypotheses concerning the differences between speech and singing we took into account the acoustic results and interpreted the articulatory singing manoeuvres based on the Acoustic Theory of Speech Production of Fant (1960). Our general conclusion was that these manoeuvres are subject to at least to two musical constraints: tuning and volume. However, as the note is produced on a vowel it means there must be a trade-off between speech and singing to avoid the loss of vowel colour, as well as to avoid unintelligibility. Our finding has to do with how this trade-off happens: consonants are shortened in singing, the first three formants match the harmonics and instead of coarticulation there is coproduction in singing. We propose that this trade-off implying musical and linguistic cognitive abilities, can be examined later, on the basis of the FAAR Acoustic Articulatory Phonology (Albano, 2001).

## Sumário

RESUMO .....	11
ABSTRACT .....	13
SUMÁRIO .....	15

<b>Introdução</b> .....	19
0.1 O objeto (ou objetos) a ser(em) investigado(s) .....	20
0.1.1 O estudo piloto .....	26
0.1.2 A hipótese explicativa .....	32
0.1.3 O recorte do objeto .....	35
0.2 Metodologia .....	35
0.2.1 A Fonética Acústica .....	35
0.2.2 <i>Design</i> experimental .....	37
0.2.2.1 A frase-veículo .....	37
0.2.3 A gravação .....	40
0.2.4 A análise dos dados .....	41
0.2.5 A análise estatística .....	44
0.2.6 Teoria da Produção de Fala .....	44
0.3. Uma canção e três estudos .....	48

### Capítulo 1: Duração de consoantes e vogais e estruturação temporal da fala e do canto

1.1 Introdução .....	51
1.2 Medindo as durações de /la <sup>1</sup> CV/ .....	56
1.3 Resultados .....	62
1.3.1 Padrão geral .....	64
1.3.2 Padrão por consoante oclusiva .....	74
1.4 Discussão final .....	77

### Capítulo 2: Padrão formântico de vogais na fala e no canto

2.1 Introdução .....	81
2.2 Estudos sobre vogais: aspectos articulatórios, acústicos e perceptuais .....	82
2.3 Resultados .....	90
2.3.1 Padrão geral .....	90
2.3.1.1 Movimentos de $F_1$ .....	94
2.3.1.2 Movimentos de $F_2$ .....	96
2.3.1.3 Movimentos de $F_3$ .....	99
2.3.1.4 A distância $F_2 - F_1$ .....	101
2.3.2 Explicação mais consistente para o padrão formântico do canto .....	102
2.3.3 Padrão por informante .....	108
2.3.3.1 Informante CM .....	109
2.3.3.2 Informante CH .....	115
2.3.3.3 Informante MG .....	119
2.3.3.4 Informante SS .....	123
2.3.3.5 Informante AK .....	127
2.4 Discussão final .....	130

### **Capítulo 3: Efeitos de co-produção entre consoantes e vogais na fala e no canto**

3.1 Introdução .....	135
3.2 O que estamos chamando de co-produção? .....	136
3.2.1 Estudos envolvendo coarticulação .....	138
3.3 Resultados .....	141
3.3.1 Padrão geral da fala .....	141
3.3.1.1 Resultados da influência da consoante sobre a vogal na fala concernentes a $F_2$ ...	144
3.3.1.2 Resultados da influência da consoante sobre a vogal na fala concernentes a $F_3$ .	146
3.3.2 Padrão geral do canto .....	147
3.3.2.1 Resultados da influência da consoante sobre a vogal concernentes a $F_1$ .....	151
3.3.2.2 Resultados da influência da consoante sobre a vogal concernentes a $F_2$ .....	152
3.3.2.3 Resultados da influência da consoante sobre a vogal concernentes a $F_3$ .....	154
3.4 Discussão final .....	154

<b>Conclusão</b> .....	157
4.1. Uma perspectiva diferente para os estudos lingüísticos: a do aspecto dinâmico da produção de fala .....	158
4.2. Conclusões finais .....	163
<b>Referências bibliográficas</b> .....	167
<b>Apêndice</b> .....	175
I. Estudo piloto .....	175
II. As notas musicais e suas frequências em <i>hertz</i> .....	176
III. Extensão da voz falada e cantada .....	177
IV. Tipos vocais de vozes femininas .....	177
V. Conteúdo do cedê da tese .....	177

## Introdução

Quando damos início a um trabalho de investigação, parece-nos impossível fugir de duas tarefas básicas que estão imbricadas (obviamente depois de explicitadas as questões iniciais, que servem de guias para a busca): eleger os aspectos do objeto maior que nos chamou a atenção - no meu caso o canto com texto - sobre os quais se quer dizer algo, e maximizar os fenômenos que partem do objeto recortado, como se estivéssemos com uma lupa. O curioso é que aquilo que parecia não existir, ou que parecia, mas não se revelava, (nem ao certo sabíamos o que era) perde aquelas pequeninas dimensões e se avulta de tal forma, que, primeira coisa: ficamos pequenos diante delas. Perguntamo-nos: e agora, o que fazer? Recortar mais o objeto a ser investigado? Delimitá-lo mais, com ajuda da literatura existente? Claro. Daí, pequena nau em meio a informações, a tarefa é seguir aquelas que levem à terra firme. Ao longo do tempo e aos poucos, elas nos fazem crescer. Respiramos e tomamos mais fôlego para nova empreitada, em que saímos em busca de mais informações e realizamos mais análises cujo fruto são mais discussões, sempre com a pergunta ecoando: Será que estamos no caminho certo?

E queremos crer que estamos. É bem verdade que titubeamos, às vezes. Fazemos pequenas incursões em estradas vicinais, ralentamos o passo, mas tudo para chegar ao fim com o máximo de respostas possíveis para as perguntas iniciais e para aquelas que não cessam de surgir ao longo do trabalho.

Esses dois parágrafos iniciais são, assim, uma espécie de justificativa dos aspectos selecionados da fala e do canto que estão anunciados no título desta tese de doutorado. Inicialmente preocupada com uma questão do mito da dicção no canto, que no meio musical sempre esteve ligada à pronúncia deste último (v. Andrade, 1975), imaginei poder falar de todos os segmentos fônicos da nossa língua. Acabei ficando com todas as vogais canônicas, é verdade. Mas dispensei, a par de oclusivas surdas, todas as consoantes. Mesmo quanto aos sons vocálicos, tive que escolher aqueles em posição acentual tônica. E assim foi, para que pudéssemos aprofundar os achados.

### **0.1 O objeto (ou objetos) a ser(em) investigado(s)**

O canto é um modo de expressão humana de origem já longínqua no tempo. Quando o homem começou a cantar? Não se sabe ao certo. A resposta é mais ou menos igual àquela que serve para responder a outras questões parecidas: Quando o homem começou a falar? Quando o homem começou a escrever? E a tocar instrumentos? E por aí poderíamos desfiar um rol de perguntas sobre as habilidades humanas.

Ao cumprir a tarefa de definir o objeto desta tese, as perguntas, algumas especulativas, não cessam de aparecer e adiam a definição. E o canto com texto, desde quando existiria? E se o homem cantou antes de falar? Ou se a fala nas suas variações de frequência fundamental foi repetindo melodias e germinou o canto, ou se o canto já existia antes das palavras e abrigou-as servindo de substância para que elas pudessem se materializar?

À primeira pergunta do parágrafo anterior podemos responder com base na história. Segundo a teoria da origem da música do sociólogo Karl Bücher (apud Nef, 1948), o canto nasceu da necessidade do homem de aliviar o esforço do trabalho corporal, marcado este

também pelo ritmo das ações repetidas: “Chez les primitifs, la mélodie seule est fixée, avec son rythme; le texte s’improvise.” (p.20)

Os registros escritos, ainda que sob a forma de fragmentos, remontam à Grécia antiga e entre eles, só para dar exemplo, podemos citar um coro para o *Orestes*, de Eurípedes (408 a.C.) e dois hinos délficos a Apolo (138 ou 128 a.C.), datadas estas transcritas de Reaney (1960). No início da era cristã até a Idade Média, há também poucos documentos musicais de canto com texto, dos quais nos dá notícia Reaney (op.cit); entre eles: alguns fragmentos da *Eneida* e lamentos para personagens como o rei visigótico Chindasvintus e a rainha Reciberga, estes últimos datando do século VII. Às outras perguntas, precisaríamos de mais tempo e de outro fórum de discussão que não este para a resposta, pois trata-se de especulação de origens, o que não é assunto da tese.

O texto no canto é, assim, perpetuado na nossa cultura ocidental, desde os hinos gregos até a canção de hoje, existindo tanto na cultura erudita, como na popular, sob as mais variadas formas. No Brasil<sup>1</sup>, além dos hinos cristãos largamente ensinados aos índios pelos jesuítas, canções trazidas d’além-mar começaram a criar cor local na colônia, fundindo-se com o lundu<sup>2</sup>, o que acabou por criar a canção brasileira. É nesta fusão, em que não se vislumbrava uma fronteira nítida entre o popular e o erudito, que nasceu o que podemos chamar de canção erudita brasileira, a partir do século XIX<sup>3</sup>.

---

<sup>1</sup> Sobre a música e em especial a canção no Brasil, recomendo as leituras da Pequena História da Música de Mário de Andrade (1977) e A canção brasileira: popular e erudita (que deveria haver às pencas nas bibliotecas das faculdades de Música) de Vasco Mariz, (1985).

<sup>2</sup> Inicialmente dança trazida pelos negros africanos ao Brasil, acompanhada de canto e de caráter bastante sensual (Andrade, 1977). Segundo Aurélio Buarque de Holanda, *teve seu esplendor em Brasil em fins do século XVIII e começos do século XIX. Dos meados do século XIX em diante, canção solista, influenciados pelo lirismo da modinha e freqüentemente de caráter cômico.* (Caldas, 1985)

<sup>3</sup> As canções de Villa-Lobos, tido como um compositor erudito, são consideradas *cross-over*, na Europa: ou seja, não são nem populares nem eruditas.

Este tipo de canção era e é cantado – de modo geral - com o acompanhamento do piano, segundo as partituras originais, e em alguns casos, da orquestra. O cantor ou cantora que executa a canção erudita, devem ter um treinamento vocal semelhante àquele do belcanto para poder cumprir as exigências de cantar com volume de voz, a fim do som não ser abafado seja pelo piano, seja pela orquestra, sem amplificação; o que só pode ser conseguido com manobras articulatórias (Sundberg,1977) que acabam por comprometer a inteligibilidade do texto cantado.

A descoberta de Sundberg (1977)<sup>4</sup> - de que uma manobra como a do acentuado abaixamento da mandíbula para notas cantadas em notas agudas é responsável pela deformação da vogal - é importante para explicar, em parte, porque vários trechos do texto cantado nos soam ininteligíveis, no entanto, não se propagou para além dos meios científicos.

No Brasil, mesmo a genialidade de Mário de Andrade, que foi também crítico da pronúncia no canto erudito, não poderia ir de encontro à visão, no meio musical, de que a ininteligibilidade era uma falha de quem cantava, pois, ainda que baseadas em intuições interessantes, suas críticas não forneciam explicações científicas e acabavam por resultar um tanto impressionísticas.

A sua contribuição poderia ter sido maior na questão que se chamava de “diferenciações fonéticas regionais”, se não fosse o resultado normativo, além do quê, pouco divulgado, a que se chegou depois do Congresso da Língua Nacional Cantada realizado em 1937, sobre o qual cabe agora uma breve discussão.

De autoria de um filólogo, Antenor Nascentes, de um musicólogo, Luís Heitor e de Mário de Andrade, este último representando o Departamento de Cultura do Estado de São

Paulo, o texto das “Normas para a boa pronúncia da língua nacional no canto erudito ditadas pelo 1º Congresso da Língua Nacional Cantada” argumenta em favor de uma língua padrão para o teatro e, especificamente, para a canção brasileira. Uma justificativa, no entanto, revela o preconceito lingüístico que dá o tom do texto. Ao defender que a arte erudita precisa de uma língua padronizada, os autores dizem:

E o teatro, essa arma poderosíssima de lições e de exemplos, largado ao léu pelos governos, manejado às pressas por artistas acoitados pela precisão de viver, tornou-se um eterno exemplo de desleixo e impureza lingüística, que erigiu a linguagem mal falada em norma de erudição nacional (in Mariz, 1985:332)

A “impureza” lingüística de que se fala aí, na verdade, é a convivência de vários falares no texto artístico, em vez do uso de uma língua padrão, que se convencionasse utilizar para fins artísticos, e que fosse, portanto, uma variedade comprometida com estilo. Percebe-se aí uma falta de clareza, quanto à definição do que se fala: se de uma língua inexistente, ou se de uma má língua que se quer tornar em boa para os fins artísticos. Há imprecisão de conceitos lingüísticos – natural para a época . Ao que um lingüista chamaria hoje de variedade, os autores das “Normas” chamam de “diferenciações fonéticas regionais”. Há também conceitos obscuros, como por exemplo “irregularidade e anormalidade de pronúncia” o que “afeta perigosamente as artes do dizer e do canto”, ao que se contrapõe a pronúncia carioca do Rio de Janeiro por ser a “mais **evolucionada** dentre as pronúncias regionais do Brasil”. ( Normas para a boa pronúncia da língua nacional no canto erudito, ditadas pelo Primeiro Congresso da Língua Nacional Cantada, realizado em S. Paulo, em 1937, in Mariz, 1985: pp 329-357. Grifo meu)

---

<sup>4</sup> A descoberta deste foneticista sueco será explicada com detalhes no segundo capítulo dessa tese.

Assim, do início das “Normas”, até a resolução de se eleger a pronúncia carioca como a mais adequada para as artes de declamar e cantar, revelam-se dificuldades em abordar a problemática adequadamente, preconceitos lingüísticos e imprecisões conceituais. Trata-se de uma escolha realmente delicada, que envolve julgamentos de melhor ou pior na língua, o que seria razoável tratar no âmbito da fonostilística, a qual os autores parecem desconhecer. O fato é que os autores de tais normas, diante da tarefa de eleger um dialeto padrão, acabaram por desenvolver um estudo sobre as vogais e consoantes do português do Brasil. Ainda que contendo falhas, tal estudo não deve ser menosprezado, bem como não deve ser desprezada a iniciativa de se discutir qual a língua padrão para o canto no Brasil. Pelo que nos conta Mariz (1948), citando Mário de Andrade (em itálico), esta tentativa foi um tanto frustrada:

*A minha conclusão particular e aflita é que os nossos compositores, quase todos, jamais se preocuparam com o problema, jamais se lançaram na árdua pesquisa estética de acomodar às exigências do canto, as exigências da palavra nacional. Passaram-se dez anos desde que Mário de Andrade escreveu essas linhas. Valeu a pena o esforço despendido?* (op.cit.: 25)

Parece que valeu, mas em um âmbito muito restrito. Na biblioteca do Centro de Ciências Letras e Artes de Campinas, de acervo já bastante desatualizado, é possível encontrar o livro *Dicção Lírica* de Pedro Moreira Lopes (1944). O autor demonstra um conhecimento avançado para a época, mas que, assim como as Normas de 1937 – às quais se refere - , parece não ter se difundido no meio musical:

A boa dicção no canto não pode ser considerada de maneira absoluta, como querem os leigos que não levam em conta a tessitura em que está cantando o cantor. Músicas de folclore têm tessitura amável, curta. Estas, sim, facultam a articulação das palavras. As músicas de câmara são de tessitura um pouco mais extensas, de maneira que certas notas podem perturbar o rigor da articulação e pronúncia.

Cinqüenta anos depois de Lopes, Duarte (1994), ao tratar do que chama de dois modelos de emissão vocal, não deixa de se referir às Normas de 1937, tomando-as como parâmetro para a escolha de pronúncia para o canto erudito. Estas últimas são, até hoje, única referência, ainda que restrita ao mundo acadêmico, de um trabalho de monta a respeito de pronúncia da voz cantada no Brasil.

Em vista da não continuidade do trabalho do Congresso de 1937 e da pouca divulgação das suas discussões que são muito mais ricas do que os resultados das Normas<sup>5</sup>, atribuía-se e atribui-se até hoje, entre cantores, outros músicos e maestros, a falta de inteligibilidade do texto cantado à má dicção.

Finda a discussão acima, retomamos o que foi anunciado no título da subseção: é o **canto** com texto, ou este canto que acolhe a fala, fazendo dela uma fala cantada, que constitui uma contraparte do objeto mais amplo deste trabalho e da qual deriva o problema da inteligibilidade. A outra parte do objeto deste estudo é a fala falada, ou apenas **fala**.

E a fala o que é? Respostas diferentes – e em grande número - podem ser dadas a essa mesma pergunta. Podemos citar duas, à guisa de ilustração: “Speech is movement made audible” (Stetson, 1928, apud Kent & Read, 1992, p.1). Ou ainda:

---

<sup>5</sup> Várias comunicações interessantes sobre a pronúncia das diversas regiões brasileiras e até peculiaridades do folclore encontram-se reunidas nos Anais do Primeiro Congresso da Língua Nacional Cantada.

Ou que é, então, a fala, essa forma de realização histórica e estruturalmente privilegiada da linguagem natural? Para responder a essa pergunta, muitas ciências têm somado seus esforços, pois, apesar de ser apenas uma faceta da linguagem, a fala é, ela mesma, um objeto multifacetado. (Motta Maia, 1991, p.9).

A fala é mesmo um objeto fascinante: é movimento mutante desde sua produção fisiológica, deixando os lábios, propagando-se no ar (movimento acústico) chegando ou não aos ouvidos (movimento perceptual); é forma de realização de uma das faculdades humanas mais difíceis de explicar: a linguagem natural. Por ser assim objeto de tantas faces, difícil apreendê-lo e resolver sua definição de uma hora para outra. É grande: o tipo da coisa que se pega de um lado e não se consegue pegar do outro.

Assim é que, diante do objeto escolhido para nossa investigação, a fala, e cientes de sua dimensão, tivemos que optar pelo primeiro recorte que foi tratar a questão da inteligibilidade do canto à luz da Fonética Acústica, a ciência do sinal acústico dos sons de fala e das suas transduções eletro-eletrônicas.

Outros recortes foram então feitos a partir do estudo piloto, o qual descrevemos brevemente, a seguir.

### **0.1.1 O estudo piloto**

Quando cantarolamos qualquer trecho do cancionário da música popular brasileira – por exemplo, podemos pensar em Aquarela do Brasil – a inteligibilidade do texto cantado parece ser uma questão tranqüila. Há dois fatores que contribuem para isto: primeiro, as manobras articulatórias para este estilo de canto, o popular, não necessitam ser muito diferentes das da fala e, segundo, a tessitura do canto popular é, em geral, estreita e realiza-se na mesma região de freqüência da fala.

O mesmo não acontece com o canto erudito. A inteligibilidade é e sempre foi um problema para o *lied* ou para a ópera. Por que, nestes estilos, as notas musicais parecem roubar o sentido à fala?<sup>6</sup> Estudos sobre inteligibilidade do canto (Scotto di Carlo, 1978,1985,1991 e 1993) e estudos acústicos da voz cantada (Sundberg, 1975,1977 e 1987) explicam, em parte, por que esta é comprometida no canto erudito. Exploram, sobretudo, os fatores musicais envolvidos: *pitch* (a nota em que a vogal é entoada) e as manobras articulatórias do canto e os resultados acústicos destas manobras.

Assim foi que, da necessidade de responder à pergunta do parágrafo anterior, com base na bibliografia citada, nasceu o estudo preliminar que possibilitou delimitar o assunto desse trabalho de tese. Passamos a resumi-lo (para versão mais detalhada, ver Raposo de Medeiros, 2000).

Coletados de forma espontânea<sup>7</sup>, os dados do piloto foram extraídos do cancionário erudito brasileiro (v. Apêndice I) com o cuidado inicial de se respeitar a região de frequência (grave, média ou aguda)<sup>8</sup> com base no estudo de Scotto di Carlo e Germain (1985) e de obter as vogais-alvo ([i,e,ɛ,a,ɔ,o,u])<sup>9</sup> na posição acentual tônica. Assim

---

<sup>6</sup> É interessante notar que esta problemática, na ópera clássica de Mozart ou de Rossini, por exemplo, era contornada com os recitativos. Os recitativos ficam num meio termo entre fala e canto e são acompanhados apenas pelo cravo, ficando a orquestra em silêncio. Os intervalos entre as notas são estreitos (segundas, terças e quartas) e encontrados também na fala. Além disso, os recitativos trazem marcas entoacionais da fala: uma pergunta termina numa quarta ascendente e uma declaração termina num intervalo estreito, descendente. Além deste exemplo podemos reproduzir, a seguir, o que Wisnik (1999) diz a respeito da preocupação dos autores da música com o texto: “Os madrigais e o melodrama barrocos assumirão um estilo expressivo, declamatório, climatizando os recursos melódicos e harmônicos, as consonâncias e dissonâncias, com uma gesticulação entoativa a serviço da ênfase nas palavras. (p. 127)

<sup>7</sup> O termo espontâneo aqui deve ser entendido como oposto a condições controladas de um experimento: por exemplo, vários contextos foram escolhidos, desde que contendo a vogal-alvo. Não foi observada rigidez para escolha da frequência fundamental da nota, desde que pertencesse ao que se classificou inicialmente como registro baixo, médio e alto.

<sup>8</sup> No estudo preliminar para região de frequência, utilizamos os termos *registro baixo, médio e/ou alto*.

<sup>9</sup> Ao longo do texto desta tese os colchetes ([ ]) serão usados na transcrição das vogais e consoantes investigadas. As barras (/ /) serão utilizadas para a transcrição larga de sílabas e logatomas. No caso de estarmos resumindo outros trabalhos, manter-se-á o sinal de representação escolhido pelos autores de tais trabalhos.

obtivemos, tanto *clusters* CV, VC ou VV, em posição aleatoriamente paroxítona ou oxítona.

O objetivo inicial do piloto era discutir a questão da inteligibilidade à luz de uma comparação de padrões formânticos, o da fala e o do canto. Em determinado momento, elaborou-se e aplicou-se um teste de percepção, que mostrou-se pouco útil e foi descartado – mesmo porque alargava o escopo da pesquisa – embora tenha servido para delinear os estudos relacionados à estrutura temporal e à coarticulação, conforme veremos oportunamente.

Os resultados que se apresentarão a seguir, sob a forma de gráfico (ver gráfico 1, p. 31) e tabelas são aqueles cujas vogais foram entoadas em uma região de frequência grave para a voz de soprano, ou seja, de 300 Hz a 510 Hz. Os dados referentes às regiões média e aguda foram descartados por razões que vêm explicadas no capítulo 2. Diante dos resultados estatísticos que atestaram diferença significativa <sup>10</sup> de padrão formântico sobretudo junto às vogais [i], [e], [ɛ] e [a] e de uma diminuição do valor de F<sub>2</sub> para essas vogais no canto (ver tabela 1), tendemos a dividir as vogais cantadas em duas classes : as não-arredondadas e as arredondadas, levantando a explicação de que aquelas primeiras tendem a se centralizar e as últimas permanecem articulatoriamente com as mesmas características da fala. A manobra típica do canto, então, seria relaxar a constrição do tubo, que sabemos é necessária para a produção das anteriores. Quanto às posteriores, consideramos que a protrusão labial encarregava-se de compensar a constrição posterior e, assim, manter a uniformidade do tubo requerida pela modalidade cantada.

---

<sup>10</sup> A análise estatística contou com a análise variância via GLM (*General Linear Model*) com alfa a 0.05. Isto quer dizer que há significância ou diferença significativa quando P (*p-value*) for igual ou menor a 0.05.

Tabela 1. Estudo piloto: padrão formântico da fala comparado ao do canto.

Vogal	Padrão Formântico: Fala x Canto								
	F <sub>0</sub>			F <sub>1</sub>			F <sub>2</sub>		
	Fala	Canto	Sig	Fala	Canto	Sig	Fala	Canto	Sig
i	256	394	s	336	414	s	2668	2306	s
e	238	320	s	490	451	ns	2589	2268	s
ε	222	370	s	652	774	s	2183	1872	s
a	205	302	s	844	669	s	1688	1268	s
ɔ	210	510	s	648	573	s	1195	1066	ns
o	242	386	s	464	495	s	962	1015	ns
u	233	367	s	398	420	s	969	961	ns

A partir da tabela acima podemos traçar um padrão formântico do canto em relação à fala, indicando, por meio de setas, o que se eleva, o que diminui e o que se mantém. Obviamente os valores de F<sub>0</sub> não serão registrados nesta tabela por se tratar da exigência da partitura, e não resultado involuntário da informante.

Tabela 2: Movimento de aumento ou diminuição dos valores de formantes do canto em relação aos formantes da fala

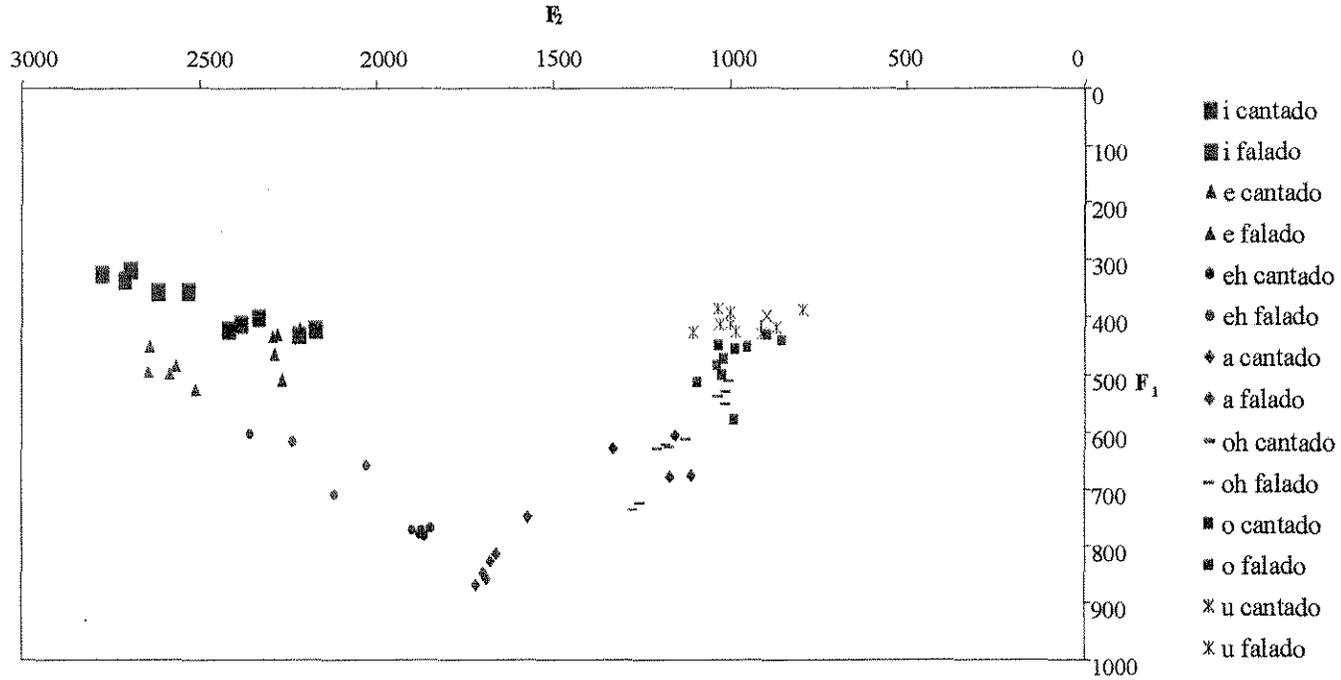
Vogais	i	e	ε	a	ɔ	o	u
F <sub>1</sub>	↑	=	↑	↓	=	=	↑
F <sub>2</sub>	↓	↓	↓	↓	=	=	=

Por esta tabela vemos que as vogais anteriores mais o [a], sempre têm seus F<sub>2</sub> diminuídos, o que nos levou a colocar o [a] do canto no mesmo grupo daquelas primeiras, formando assim o grupo das não arredondadas. No gráfico 1 (p. 31), porém, é possível ver que o [a] tem um comportamento diferente das demais vogais no canto, inclusive, as repetições mostram que ele está mais disperso no espaço do triângulo vocálico do que as

outras vogais. Esse fato foi melhor discutido no estudo da tese, no qual os agrupamentos arredondado e não-arredondado foram revistos.

A conclusão a que chegamos no estudo piloto, cujos dados aqui apresentados já são um recorte do estágio inicial do próprio piloto, tomou a forma de uma hipótese explicativa sobre a qual se baseou todo o trabalho de investigação da tese.

Gráfico 1: Vogais faladas x cantadas/ estudo piloto- contextos variados:  $F_1$  x  $F_2$  em Hz



### 0.1.2 A hipótese explicativa

O formato *schwalike* das vogais anteriores – ver no gráfico um movimento de [i], [e] e [ɛ], em direção ao centro do triângulo – pode ser explicado pelo fato de que, ao configurar o trato vocal para vogais mais centralizadas, a cantora cria um tubo com menor constrição, dentro do possível, a fim de preservar o sinal sonoro da fonte, responsável pela frequência de fonação. Essa última é uma exigência musical: é a própria nota que tem de estabelecer uma relação de altura com outras notas para que o ouvinte a identifique como parte de um todo, no nosso caso, a canção. Daí a necessidade da afinação. Evitando constrição demasiada, a cantora estaria assegurando a afinação e o volume, ainda que distorcendo a cor de algumas vogais, explicação que foi aceita como uma hipótese a ser explorada no estudo da tese e a que chamamos de uniformização do tubo.

Estávamos diante de uma hipótese que contemplava o aspecto de produção: o padrão formântico do canto mostrou-se diferente do padrão formântico da fala, mas o quanto isto interfere no aspecto perceptual? A pergunta que nos fazíamos, naquele momento, era se o que se mostrou semelhante (as vogais posteriores) eram identificado adequadamente, e se o que se mostrou diferente não era inteligível?

Para uma resposta baseamo-nos em Scotto di Carlo e Germain (1985) que investigaram a questão da inteligibilidade do canto em francês, submetendo quatro sujeitos não músicos à identificação de 15 vogais em 17 *pitches* (frequência de fonação) diferentes. A conclusão das autoras foi de que a inteligibilidade da vogal é inversamente proporcional ao *pitch*. Ou seja, quanto mais aguda a nota em que a vogal é cantada, mais a vogal é entendida como [a], comprometendo-se assim a inteligibilidade do texto cantado.

Elaboramos então um teste de percepção<sup>11</sup> em que sete sujeitos ouviram, ao todo, 42 vogais – 18 isoladas, somadas a mais 12 em contexto consonantal e essas mesmas 12, também isoladas – com a instrução de identificá-las, uma após outra, com uma pausa de quatro segundos entre um estímulo e outro, sem se fixar no que já havia passado. Assim como em Scotto di Carlo e Germain (1985), nossos informantes ouviram as vogais cantadas em diferentes *pitches*, agrupados em três diferentes regiões de frequência a grave, média e aguda. Mas, diferentemente do estudo das autoras francesas, o nosso contemplou as variáveis vogal isolada e vogal contextualizada, ao passo que o dessas autoras envolveu apenas vogais isoladas.

Nossos resultados confirmaram a hipótese relativa à variável região de frequência: quanto mais se distanciava da região grave, mais a vogal perdia em inteligibilidade. O acerto na identificação, levando-se em conta todas as vogais, é de 60%, caindo para 28% na região média e para 25% na aguda. Cabe lembrar agora que não houve controle rígido do *pitch* (como por exemplo em um *design* experimental em que se determina que todas as vogais sejam cantadas numa mesma frequência de fonação), a não ser o agrupamento por regiões de frequência, isto quer dizer que o [a] agudo, era um lá 4 (880 Hz) e o [ɔ] agudo podia ser um fá 4 (700 Hz).

Quanto às variáveis vogal isolada e vogal contextualizada, mantendo-se a divisão entre as regiões de frequência, os resultados não diferiram muito dos apresentados acima. Na região grave, a vogal isolada é um pouco melhor reconhecida que a contextualizada, recebendo 65 % de acerto contra 55% desta última. Para as regiões média e aguda, as duas condições da vogal recebem identificação acertada em torno de 25%, o que corrobora os

---

<sup>11</sup> Na verdade foram dois testes de percepção, tendo sido o segundo aperfeiçoado. No primeiro, as vogais

resultados de Scotto di Carlo, segundo os quais a inteligibilidade do texto cantado decresce na medida em que o *pitch* se eleva; mas que acrescenta pouco à nossa suspeita de que vogais contextualizadas talvez oferecessem melhores condições de inteligibilidade.

No entanto, um fato decorrente da construção do teste levou-nos a introduzir na tese um estudo sobre coarticulação. Para isolarmos a vogal, por exemplo, num contexto como /venturas/, vimos que, ao buscar o *onset* da vogal, tendo estabelecido que este viria imediatamente após a explosão da oclusiva, tivemos dificuldade de identificar visualmente onde estava esta explosão, deduzindo que, no canto, essa parecia diminuta, o que nos levou a pensar a questão da inteligibilidade do canto no âmbito da estrutura temporal, e não só circunscrita a fenômenos de frequência. O suposto era de que as consoantes, de uma maneira geral, eram rapidamente articuladas e isso afetaria o fenômeno da coarticulação na fala cantada. Desta constatação, decidimos que mereciam atenção, além do padrão formântico, o aspecto de duração dos segmentos no canto e o de coarticulação.

A partir destes resultados, reconhecemos que as condições dos estímulos variadas como estavam – por exemplo, aceitou-se segmentos contextualizando as vogais que tanto podiam ser CV, VC, como silêncio mais vogal, por se tratar de dados espontâneos, sem modificação do texto da partitura – impediriam chegarmos a explicações conclusivas em torno da questão da inteligibilidade. Reconhecemos também, que, apesar do teste de percepção nos ter aberto os olhos para fenômenos interessantes da fala cantada, era preciso delimitar o estudo no âmbito da teoria de produção de fala, que por seu estatuto, seria uma base mais sólida para nossa limpeza de terreno.

---

escolhidas foram [i], [e] e [ɛ] e foram ouvidas isoladamente. No segundo, as vogais eram [a], [ɔ], [o] e [u].

### 0.1.3 O recorte do objeto

Esperamos ter, na subseção acima, elucidado, então, o modo como recortamos o objeto de estudo. O recorte feito – a partir de uma problemática do canto com texto – contemplou os seguintes aspectos fonético-acústicos da fala:

1. o padrão formântico
2. a estrutura temporal
3. a coarticulação

aspectos a serem descritos comparativamente, tendo-se, de uma lado a fala falada e de outro a cantada, a que chamamos de modalidades. De agora em diante referir-nos-emos a elas como *fala* e *canto*, simplesmente.

## 0.2 Metodologia

### 0.2.1 A Fonética Acústica

Este é o caso em que a natureza dos aspectos focados, ou que se deseja focar, elege a área da ciência que poderá trabalhar a hipótese levantada. Com os aspectos definidos, acima citados, a Fonética Acústica foi eleita para desvendar o mistério de diferentes modalidades, produzidas, a um tempo, por um só aparelho. A disciplina sobre a qual falamos agora redimensiona-se a partir dos anos 40 do século XX, com o desenvolvimento do espectrógrafo<sup>12</sup>.

O nascimento do espectrógrafo, filho das técnicas de gravação magnética do som com o desenvolvimento de técnicas para tornar a fala visível, redimensiona, não só para as ciências chamadas duras, o estudo do som da fala: a informação acústica passa a ser

---

<sup>12</sup> Para uma história do espectrógrafo, ver Fujimura, in Hardcastle and Laver (1997).

relevante para a descrição dos sons (veja-se Jakobson, Fant e Halle, 1952) e para teorias lingüísticas que, embora discutam a necessidade da Fonética, têm seus adeptos (veja-se Pierrehumbert & Beckman, 1988). O espectrógrafo nos dá o espectrograma, o qual nos permite ver, ao longo do tempo, as mudanças de frequência e de concentração de energia na cadeia sonora da fala. É como se fosse um raio X da fala (acho que nada há de original nessa comparação, mas ela é boa, pois dá essa idéia de desvendar o que parecia impossível de se ver).

Nesses sessenta anos, a ciência da computação proporcionou-nos a informatização de análises - antes feitas por meios analógicos - em um *hardware* cada vez menor, viabilizando um *software* de análise de fala, rico em possibilidades, que roda tranqüilamente num micro-computador. Um *software* como o CSL (*Computerized Speech Laboratory*, da *Kay Elemetric Corporation*), realiza em poucos segundos, a digitalização de um som, a geração de sua forma de onda, de seu espectrograma e as análises FFT (*Fast Fourier Transform*) e LPC (*Linear Predictive Coding*), só para começar. As últimas análises mencionadas, só para dar um exemplo, fornecem estimativas dos valores dos formantes, tão importantes na identificação e caracterização das vogais.

A Fonética Acústica, no entanto, não se resume a resultados de análises geradas por *softwares* de base matemática. Fazer Fonética Acústica é conversar o tempo todo com a área fisiológica e com a área perceptual. A primeira nos dá a conhecer os articuladores da fala humana, e na condição de responsáveis por sua produção, fazem o elo de saída com a Fonética Acústica e a outra, nos ensina sobre os receptores desta fala, fazendo com a disciplina em questão, o elo de chegada. O que nos interessa nessa tese de doutorado é explicar o objeto com base nos estudos de produção da fala. **Fica portanto, relegada a segundo plano, a questão perceptual.**

Explicar, comparativamente, os fenômenos acústicos encontrados na fala e no canto evidenciando seus correlatos articulatórios será o meio de ancorarmos nossas conclusões.

Por estarem assim, definidos e recortados, objeto e área do saber: fala e Fonética Acústica, deu-se início, então, ao *design* experimental, a fim de que se realizassem as medidas acústicas necessárias.

### 0.2.2 *Design experimental*

Para comparar a fala e o canto em PB, escolhemos as sete vogais orais [i,e,ɛ,a,ɔ,o,u], da posição acentual tônica; e as consoantes oclusivas surdas [p,t,k], que foram realizadas no logatoma /la<sup>1</sup>CV/, contido na frase veículo: “Canto /la<sup>1</sup>CV/ baixinho numa velha canção de ninar”, sobre a qual falaremos com maiores detalhes na próxima subseção.

A frase veículo foi falada e cantada em cinco repetições por cinco cantoras (sopranos de câmara), com idades que variavam de 23 a 40 anos. Multiplicando o número de informantes, as cantoras, com o de repetições, com o de modalidades (canto e fala), com o de consoantes e, finalmente, com o de vogais, obtivemos um total de 1050 ocorrências.

$$5 \times 5 \times 2 \times 3 \times 7 = 1050$$

A partir dessas ocorrências, foi-nos possível, então, medir os três primeiros formantes de todas as vogais (para os estudos de padrão formântico e de coarticulação) e a duração de cada segmento do logatoma (para o estudo da estrutura temporal).

#### 0.2.2.1 A frase-veículo

Há e sempre deve haver um cuidado com a frase veículo em que inserimos a palavra ou logatoma que queremos analisar num experimento fonético-acústico. A frase-veículo

age como uma proteção evitando o efeito lista no contorno entoacional da palavra<sup>13</sup>. Um bom exemplo é *Digo (palavra/logatoma) baixinho*, em que a vogal final do verbo, /o/ faz uma fronteira confortável com uma palavra/logatoma iniciados por consoante (o nosso caso) e em que a consoante sonora /b/ do adjunto adverbial evita um corte abrupto no vozeamento da vogal final da palavra/logatoma. Mas como o presente experimento tinha outra modalidade a ser analisada, o canto, acabamos optando por uma frase que parecesse mais natural para as nossas informantes. Não foi difícil encontrarmos a frase veículo, que felizmente, está numa das canções, cujo compositor é um dos mais produtivos e reverenciados pela crítica musical brasileira<sup>14</sup>. Trata-se da Cantiga de Ninar de Francisco Mignone.

Tanto o texto da canção, como a tessitura em que esta foi composta, eram perfeitamente compatíveis com as necessidades do experimento. Uma pequena modificação teve de ser feita: o logatoma foi inserido na frase da canção, sem alterar altura, de forma alguma, nem ritmo de forma substancial. A frase original é a transcrita abaixo, numa partitura cuja tonalidade é mi bemol maior e o tempo é 2/4, um compasso binário.

## Cantiga de Ninar

Moderato e calmo

Francisco Mignone (1925)

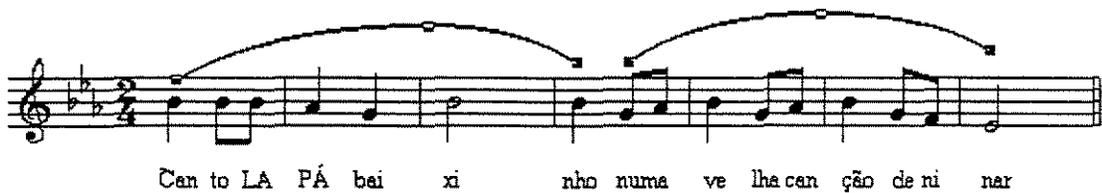
*p* *dolcissimo*

Can to bai xi nho u ma ve lhacã o de ni nar

<sup>13</sup> O chamado “efeito lista” é o efeito entoacional que resulta justamente de uma leitura em voz alta de itens numa lista, em que há um contorno que pode ser assim descrito: eleva-se a frequência fundamental na sílaba tônica, para abaixar em seguida, antes da pausa, ao fim de cada item.

<sup>14</sup> Para uma história bastante completa da canção brasileira, ver Mariz (1985).

O compasso binário é aquele que agrupa duas batidas, as quais podem ser organizadas da seguinte forma: ou subjazem à realização de uma nota que dura por todo o compasso, portanto uma nota relativamente longa, ou agrupam uma ou mais notas. No caso da Cantiga de Ninar, a nota longa é a mínima, e a nota que corresponde a uma batida é a semínima. Para inserir o logatoma na frase da canção, foi necessário encurtar a duração da sílaba /can/ de /canto/ a uma semínima e a sílaba /to/ a uma colcheia; assim, ainda sobrava meio tempo a ser preenchido, o que foi feito com a sílaba /la/ e, no compasso seguinte, o primeiro tempo foi ocupado pela sílaba /CV/, mantendo-se o resto deste compasso – assim como o resto da frase – conforme a partitura original. O resultado obtido foi este



A extensão da canção é de uma oitava, começa em um mi 3 e vai até um mi 4, sendo a tonalidade, mi bemol maior. A escolha de tal tessitura, foi proposital, pois estávamos evitando, exatamente, notas médias e agudas, as quais forçosamente sobrepõem as restrições musicais sobre as lingüísticas, o que será discutido em detalhes no capítulo 2.

Observe-se que a nota sobre a qual se entoa a vogal alvo é um lá bemol, ou seja, no canto, é a frequência fundamental desta vogal, que fica em torno de 410 Hz (v. apêndice para um quadro das frequências fundamentais das notas musicais). Pode-se pensar em uma contra-indicação ao se escolher uma nota que, em termos absolutos (se assim podemos dizer) não é considerada grave (está na oitava central do piano), mas em se tratando de

informantes com voz de soprano, uma nota mais grave do que isso, implica numa região desconfortável para a cantora.

Para resolver o que seria um problema em termos de taxa de elocução<sup>15</sup>, esta ficou estipulada em 92 batidas por minuto, tanto para fala como para o canto<sup>16</sup>. Isto significou uma alteração de andamento em relação à partitura original, que era de 72 batidas por minuto e, portanto, um tanto quanto lento para ser produzido na fala. O andamento escolhido neste experimento pareceu-nos bastante adequado, pois era natural na fala e não desvirtuava a composição musical, que pedia um andamento calmo.

Tentamos, assim, nos limites de um experimento laboratorial, respeitar as características das diferentes modalidades estudadas: a fala e o canto, possibilitando às informantes desempenharem a tarefa sem maiores estranhamentos. Todas já possuíam experiência com gravação, e a repetição é algo bastante familiar para o músico que repassa sempre e muitas vezes o mesmo trecho em um ensaio.

### 0.2.3 A gravação

Familiarizadas, então, as informantes, com o instrumental de uma gravação (cabine, microfone, mesa de som) e com os ajustes deste, a tarefa de instrução limitou-se ao seguinte :

1. A informantes deveriam produzir as frases cantadas antes de falar, para evitar cansaço vocal.

---

<sup>15</sup> Taxa de elocução tinha o mesmo número de batidas que o andamento, sendo este último termo próprio da Música. Assim, as frases cantadas foram realizadas no andamento *Andante*, ou seja, aquele em que realizam-se 92 batidas por minuto.

<sup>16</sup> Isto significou uma frase falada de 4 segundos e uma frase cantada de 8 segundos, aproximadamente.

2. Deveriam, inicialmente, ler a partitura já modificada (as cantoras já conheciam a canção) e manter a afinação e o andamento, para o quê, sempre tinham à mão o diapasão de garfo e um metrônomo luminoso, Matrix, modelo MR-500.
3. Uma vez decorada a partitura, as informantes deviam ler os logotomas, apresentados em seqüência, com intervalo de tempo de dois segundos entre uma frase e outra. Os logotomas foram escritos em tiras de papel e apresentados do lado de fora da cabine.

Para gravar as frases faladas, procedeu-se como o descrito no item 3, obviamente, decorando-se apenas a frase-veículo.

A gravação foi realizada a 44 kHz, em cabine insonorizada do Laboratório de Fonética e Psicolinguística, LAFAPE; utilizando-se um microfone unidirecional Shure, SM 48, uma Fita DAT Sony DG 90P, um gravador TASCAM DA 30 e um áudio mixer profissional MXM 6.

#### **0.2.4 A análise dos dados**

Uma vez terminada a gravação, estava disponível, então, o conjunto de dados – portanto nosso *corpus* – do qual viemos a obter as medidas acústicas de duração e de frequência.

Digitalizaram-se os dados gravados a 26500 Hz , utilizando-se o *software* de análise de fala CSL, modelo 4300b da *Kay Elemetrics*. Ao digitalizar os dados, obtinha-se a forma da onda de toda frase-veículo, da qual selecionávamos a porção de interesse – o logotoma – para o qual se gerava um espectrograma de um filtro de 375 Hz, este usado para medidas de duração.

Para as medidas de duração, começamos pelo logatoma. Este era medido do *onset* do /l/ até o *offset* da vogal-alvo. Em seguida medimos a consoante, do início do silêncio até o *onset* da vogal-alvo; depois a vogal-alvo até o *onset* da oclusão da palavra *baixinho*. Por último, medimos a vogal pré-tônica e o /l/, sempre com o auxílio visual da forma da onda e do espectrograma.

As medidas de frequência compreenderam as medidas dos três primeiros formantes,  $F_1$ ,  $F_2$  e  $F_3$  das vogais, em sua porção estacionária. O algoritmo utilizado foi o LPC (*Linear Predictive Code*), sobreposto ao FFT (*Fast Fourier Transform*)<sup>17</sup>. Quando o resultado numérico do formante não correspondia a um valor esperado para um dos formantes em questão, por exemplo,  $F_2$  e  $F_3$  muito próximos, no caso de /i/, ou  $F_1$  e  $F_2$  também muito próximos, em /u/, tanto o espectrograma quanto o espectro de FFT eram consultados. Muitas vezes, sobretudo na fala, era impossível visualizar os formantes vizinhos, assim diante do “erro” do LPC e da ausência de pistas visuais, acabávamos por não atribuir nenhum valor ao formante, tornando-se este, um valor faltante.

Felizmente, o número de valores faltantes foi pequeno e não representou um empecilho para a análise estatística, sobre a qual falamos a seguir.

Antes de concluir esta seção, cabe agora uma breve explicação sobre a escolha de determinados parâmetros tanto para a digitalização como para a geração do espectrograma e para as análises FFT e LPC.

A taxa de amostragem para a digitalização, como se sabe, deve ser o dobro da frequência mais alta esperada num sinal de fala. Assim, se o sinal a ser analisado não contiver fricativas, por exemplo, pode ser digitalizado a 10 kHz (10.000 Hz), levando-se

---

<sup>17</sup> A fixação de parâmetros para, respectivamente FFT e LPC, foi a seguinte: 1024 pontos para tamanho da amostra, 20 para tamanho do quadro e 24 a ordem de filtro.

em conta que a frequência máxima não ultrapassará 5 kHz (5000 Hz). Em geral, considera-se que 22.000 Hz de taxa de amostragem é suficiente para uma boa análise do sinal de fala (para detalhes, em uma leitura elementar, ver Kent e Read 1992) . Como nossa frase veículo apresenta uma variedade de sons da fala, incluindo as fricativas (embora estas não sejam focadas em nossa investigação), e considerando ainda que trabalhamos com a fala cantada, cuja frequência fundamental excederia aquela da fala, optamos por uma taxa de amostragem maior do que a de 22.000 Hz. Assim, 25.600 Hz pareceu-nos razoável.

Para FFT, ajustamos o tamanho da amostra (*analysis size*) para 1024 pontos, pois as vogais-alvo consistiam, tanto na fala como no canto, sons consideravelmente estáveis (vogais tônicas, produzidas numa taxa de elocução controlada), portanto, ainda que uma janela de 1024 pontos pudesse representar algo que excedesse o tempo em que os fenômenos da fala normalmente ocorrem, não era esse o caso de nossos dados; mesmo porque, com uma taxa de amostragem a 25600 Hz, uma janela de 1024 pontos tem 40 milissegundos.

Para LPC escolhemos ordem de filtro 24. Segundo Ladefoged (1996), para se calcular a ordem de filtro a ser utilizada, basta considerar o número da taxa de amostragem em kHz e acrescentar-lhe 2. No nosso caso atingiríamos um filtro de 28 coeficientes ( $25.6 + 2 = 27.6$ ), o qual não nos foi possível utilizar pela seguinte razão: dado que um algoritmo como LPC tem o poder de não só determinar os formantes, mas também detectar efeitos da radiação, e tanto na fala, como no canto de nossas informantes a energia começava a se perder por volta dos 10 kHz, para em seguida cair abruptamente, um filtro de mais de 24 coeficientes poderia dar resultados errôneos, como indicar dois formantes próximos ( $F_2$  e  $F_3$  por exemplo), quando na verdade tratava-se de apenas um, o  $F_2$  (para explicações detalhadas de FFT e LPC, ver Ladefoged 1996).

Por fim, esclarecemos que os espectrogramas gerados tinham uma banda de 375 Hz de largura, proporcionando boa visualização dos segmentos da fala e do canto, o que nos possibilitou localizar com precisão o início e o fim das consoantes e vogais dos logatomas para as medidas de duração.

### **0.2.5 A análise estatística**

O *software* de análise estatística utilizado foi o SAS. Tanto para os dados de duração como para aqueles de padrão formântico, o procedimento de análise utilizado foi o GLM (General Linear Model) para medidas repetidas e o Teste de Duncan, para comparação aos pares.

Os dados de duração apresentavam-se não paramétricos entre as duas modalidades a serem comparadas. Como vimos na sub-seção 3.2.1, o mesmo andamento de fala e canto significa que aquela se realiza na metade do tempo deste último. Assim, extraímos razões dos segmentos em relação ao logatoma.

### **0.2.6 Teoria da Produção de Fala**

Para explicar o que chama de teoria geral, no primeiro capítulo de *Acoustic Theory of Speech Production*, Fant (1960) toma como pressuposto o conhecimento do leitor em relação à teoria fonte-filtro e é a partir de sua descrição que vai desafiando os conceitos de acústica, com o cuidado de traçar, por vezes, uma diferença entre conceitos que chama de técnicos e fonéticos. Por exemplo: o termo harmônico e periódico seriam termos próprios da acústica, sendo o último inadequado para a fonética, uma vez que a fala não possui sons periódicos, mas sim quasi-periódicos.

A preocupação terminológica já não é uma característica de manuais introdutórios da acústica da fala, como por exemplo podemos citar Kent e Read (1992) em que a questão da quase-periodicidade já é um conceito assimilado pela fonética acústica. Para explicarem a teoria fonte-filtro – fundamental para a teoria de produção de fala – os autores preparam o terreno numa descrição do aparelho fonador – sistema respiratório, laríngeo e articulatório – e do aspecto acústico da fala.

Parece que, reportando-nos aos manuais de fonética acústica, fica claro por que Fant fala inicialmente da teoria fonte-filtro que resume, a partir de cálculos matemáticos, o funcionamento dos sistemas laríngeo (fonte) e supra-laríngeo (filtro) supracitados, permitindo a compreensão da produção de fala como fenómeno físico.

Compreendemos melhor a importância da teoria de Fant, tanto mais ficamos atentos à preocupação que ele devota ao fato de uma teoria construída sobre conceitos acústicos dar respostas quanto a fenómenos articulatórios, com o cuidado de ver e rever as bases de segmentação da fala para fins de análise. Embora vista, hoje, como limitada para explicar todos os fenómenos da produção de fala por ser uma teoria linear, implicando, só para dar um exemplo, a independência da fonte e do trato vocal – que na verdade interagem – a teoria fonte-filtro é responsável, ao longo de mais de três décadas, pelos progressos alcançados no âmbito da fonética acústica e dá conta, com mais de 90% de rigor das análises envolvidas na área. É somente quando a qualidade de voz interfere crucialmente com as análises, que se faz necessário recorrer a versões mais complexas dessa teoria, em que é revisto o pressuposto de linearidade.

Também é de interesse específico dessa tese destacar o que Fant chama de *F-pattern* ou padrão formântico, visível a partir de uma envoltória suavizada que contempla os picos de frequência formados a partir dos ressoadores do trato vocal (filtro). A partir do

F-pattern, que compreende os quatro primeiros formantes,  $F_1$ ,  $F_2$ ,  $F_3$  e  $F_4$ , é possível inferir a articulação dos sons falados, o que, segundo o próprio Fant é razão de interesse para a análise acústica da fala:

The possibility of inferring the articulation of any given sound given the evidence of the spectrum envelope and in particular the frequencies of those formants that have a continuity with the formants of an adjacent vowel. (Fant, op.cit., p.25)

Cabe explicar agora o seguinte: a teoria fonte-filtro implica a produção de um som a partir da vibração das cordas vocais e que passa “livremente” pelo trato vocal, produzindo um som vocálico. Assim, as palavras *any given sound* empregadas por Fant, devem ser entendidas, a priori, como qualquer som vocálico.

Enquanto a teoria fonte-filtro nos dá respostas matemáticas, a teoria da Perturbação (Chiba e Kajiyama, 1941) é qualitativa e explica – também usando um modelo de tubo para vogais - que constrições no tubo ressoador (tubo vocal) podem elevar ou diminuir a frequência dos formantes.

O trato vocal pode ser dividido em regiões de velocidade máxima, os nós, e velocidade mínima, os anti-nós. Quando a constrição realizada no trato vocal se dá sobre o nó, as frequências de formantes diminuem, em relação ao *schwa* é o caso de  $F_1$  e  $F_2$  de /o/ e de /u/. Se há constrição numa região velar, a frequência do formante se eleva, com relação ao padrão do tubo uniforme que é o *schwa*, cujos formantes se espaçam a intervalos de aproximadamente 1 kHz, a partir do primeiro, que tem o valor de 500 Hz para um trato vocal de 17,5 cm, e velocidade do som de 350 m/s. Esse é o caso de todos os formantes de /a/ (para detalhes ver Kent e Read, 1992), cuja maior constrição é na região velar.

A teoria acústica nos permite explicar o comportamento de determinado padrão formântico com base nas manobras articulatórias incluindo desde o abaixamento da laringe até a protrusão labial – manobras que diminuem os valores de todos os formantes – ou o levantamento da laringe e a retração dos lábios que elevam todas as frequências.

O comportamento das consoantes, fricativas e oclusivas recebe um tratamento minucioso de Fant , tratamento esse possível a partir de cálculos baseados em dados de raio-X. Fant descreve as consoantes como sons da fala cuja fonte é uma fonte de ruído , formada na constrição, maior ou menor, em determinado ponto do trato vocal. As classes de consoantes descritas são as seguintes:

- labiais
- dentais
- todos outros sons produzidos com um ponto de articulação mais retraído.

No último grupo inclui-se a velar [k] que, juntamente com a labial [p] e a dental [t], faz parte dos sons consonânticos contemplados por essa tese.

O que Fant chama de um modelo idealizado de produção das consoantes obviamente não tem o mesmo *status* do modelo fonte-filtro na teoria de produção de fala, mas é importante por sua riqueza de detalhes e por explorar a questão das mudanças no trato vocal e suas contribuições para a fonte de sons consonânticos, o que define características acústicas de cada classe.

Embora o estudo que contemple as consoantes [p,t,k] seja temporal, não envolvendo fenômenos de frequência, o modelo idealizado de Fant (op.cit) e textos clássicos sobre oclusivas como o de Halle, Hughes and Raddley (1957) serão pontos de referência importantes para a análise de fala realizada nesta tese. Também o modelo fonte-filtro e a

teoria da perturbação serão evocados, sempre que necessário, para explicar os desvios de comportamento da fala cantada em relação à fala falada.

### 0.3 Uma canção e três estudos: os capítulos da tese

Como vimos na seção, 2.2.1 a *Cantiga de Ninar* de Francisco Mignone proporcionou-nos a frase-veículo a partir da qual erigiram-se os estudos desta tese<sup>18</sup>. Cada capítulo da tese corresponde a um estudo, conforme se expõe a seguir.

O capítulo 1, primeiro estudo, *Duração de consoantes e vogais e estruturação temporal da fala e do canto*, trata da estrutura temporal das diferentes modalidades da fala estudadas. Comparou-se a duração dos segmentos do logatoma /la<sup>1</sup>CV/ produzido nas duas modalidades, com especial atenção para as consoantes oclusivas, atentando-se para o fato de que fala e canto, ainda que realizadas no mesmo andamento, têm uma organização temporal diferente, devido ao alongamento da nota musical. O capítulo 2, segundo estudo, *Padrão formântico de vogais na fala e no canto*, versa sobre fenômenos no domínio da frequência, que oferecem especial desafio para o canto, no que diz respeito a identificação das vogais. O capítulo 3, terceiro estudo, *Efeitos de co-produção entre consoantes e vogais na fala e no canto*, discute o fenômeno da co-produção no trecho CV, contemplando o fenômeno da coarticulação como lingüístico na fala e, raro, porém verificável no canto.

Os três estudos que constituem esta tese de doutorado não foram executados a duas mãos, apenas; daí o texto ser redigido na primeira pessoa do plural. Pode-se comparar o trabalho dessa tese à execução de uma sinfonia concertante<sup>19</sup>. Eventualmente, nas

<sup>18</sup> Ver, ao final da tese, encarte com cedê, cujo conteúdo está descrito no anexo V.

<sup>19</sup> No dicionário Oxford de música, temos a seguinte definição de *sinfonia concertante*: termo preferido a concerto por Haydn, Mozart e outros para composições com mais de um instrumento solista e orquestra, por exemplo as de Mozart para violino e viola. No século XX, Walton Williamson e outros usaram o termo,

passagens solo, assumo a primeira pessoa do singular, para isentar os outros executantes de responsabilidade por possíveis falhas, as quais assumo.

---

mesmo quando havia um único solista, para assinalar que a parte de solo deveria estar mais profundamente integrada na orquestra do que num concerto de virtuosismo espetacular.

## Capítulo 1

### Duração de consoantes e vogais e estruturação temporal da fala e do canto

Como o arco que vibra tanto pra lançar longe a flecha como para lançar perto o som: a voz humana tanto vibra para lançar perto a palavra como lançar longe o som musical. (M. de Andrade)

#### 1.1 Introdução

O aspecto fonético acústico selecionado para esse capítulo é o da duração de segmentos da fala, ou melhor, das falas com que trabalhamos nesta pesquisa: a fala *strictu senso* e a fala cantada. Convém inicialmente ressaltar que, apesar do título conter a palavra *vogais*, são as consoantes oclusivas que recebem maior atenção nessa parte do estudo comparativo entre fala e canto.

Retomando o que já foi dito na Introdução, a questão da estruturação temporal do canto mostrou-se relevante para o trabalho. Ao tentar isolar a vogal da consoante oclusiva, para um experimento do estudo piloto, verificou-se que, no canto, a explosão parecia diminuta. O fenômeno que se esboçou, ali, chamou a atenção para o fato de que a inteligibilidade do canto não estava apenas relacionada com os fenômenos de frequência, mas também relacionava-se a fenômenos de duração de segmentos, portanto, estruturando temporalmente os gestos daquele último de modo diferente dos da fala.

Estudos sobre os sons da fala envolvendo duração, como os de Halle et al. (1957), Blumstein e Stevens (1979), Kessinger e Blumstein (1998) e Stevens (2000), tentam

encontrar uma característica acústica invariante, que permite ao ouvinte reconhecer a consoante, apesar da variabilidade do sinal sonoro. Dentre os estudos acima citados – todos focando a seqüência CV – o de Kessinger e Blumstein (op.cit) discute o efeito da taxa de elocução na produção do VOT<sup>1</sup>, duração da vogal e duração do trecho CV para as oclusivas surdas do inglês. O resultado é que quanto mais lenta a taxa de elocução tanto maior a duração do VOT e da vogal, achado que implica no aperfeiçoamento de testes de percepção e na caracterização acústica dos sons da fala.

Stevens (2000) aponta para o fato de existirem várias pistas acústicas no que ele chama de bordas (*landmarks*) das consoantes. Tais pistas são recuperadas pelo ouvinte a partir de uma descontinuidade do sinal sonoro introduzida por uma mudança abrupta atribuída à ação de três articuladores: os lábios; e a ponta e o corpo da língua. Stevens explica a produção de cada classe de consoantes (as fricativas e oclusivas, as nasais e as laterais) no nível articulatorio e desenvolve o texto sobre a explicação pormenorizada das pistas acústicas das mencionadas classes. A discussão final enfatiza que a diversidade de pistas localizadas nas duas bordas da consoante implica numa identificação robusta.

O que nos interessa não é discutir no interior desse capítulo, as pistas acústicas, mas considerar, a partir do texto de Stevens que a consoante (cuja principal característica é a descontinuidade acústica) é importante para o ouvinte identificar o texto no canto e deve ser realizada num tempo mínimo suficiente para ser percebida como tal.

Os trabalhos acima resumidos apresentam achados mais recentes envolvendo duração e traços acústicos das consoantes, mas tais aspectos fonéticos já recebem atenção

---

<sup>1</sup> Segundo Fujimura e Erickson (in Hardcastle e Laver, 1997), “The time interval between the articulatory release of a stop consonant and voice onset, typically in a CV syllable context, is called voice onset time (VOT)” (p.76).

há cinco décadas. Embora não focando centralmente o aspecto perceptual, os trabalhos de Kessinger e Blumstein (op.cit) e o de Stevens (op.cit) em suas discussões finais acabam referindo-se a ele e trazem sugestões interessantes para a nossa comparação fala/canto que serão destacados ao final do capítulo.

Em relação a estudos da voz cantada uma autora e um autor se destacam: Scotto di Carlo (1977, 1978, 1991, 1993 e 1994) e Sundberg (1987). A primeira desenvolve estudos sobre a voz cantada, os mais variados, desde dicção, passando por inteligibilidade, problemas de respiração, de expressão facial, até o envelhecimento vocal. É junto a essa autora que encontramos dois trabalhos sobre a consoante no canto, que resumimos a seguir. O segundo autor, Sundberg, restringe e aprofunda seus estudos sobre a voz cantada no âmbito da acústica, a partir de um instrumental poderoso que lhe possibilita explicações precisas de vários fenômenos acústicos do canto.

Não fosse pelos trabalhos de Scotto di Carlo de 1977 e 1991, poderíamos dizer que as consoantes recebem ou vêm recebendo pouca atenção dos estudiosos da voz cantada. O “*Étude acoustique et statistique de l’influence des consonnes sur la justesse des voyelles subsequentes en voix chantée*” de 1977, trata da manutenção ou não da frequência fundamental no ataque da vogal, sendo esta precedida por consoante. Dezesesseis consoantes do francês foram utilizadas e a vogal do experimento era apenas uma, o [a]. Não houve comparação fala/canto, ou seja, o trecho CV cantado nas regiões grave, média e aguda, não teve sua contraparte falada realizada. A comparação com a fala se dá via literatura, ou seja, a autora refere-se, na conclusão, a estudos sobre a fala como de Di Cristo (1976) e Chistovitch (1969): tanto a fala como o canto nas regiões grave e média-grave, apresentam o mesmo efeito microprosódico no tocante a frequência fundamental das vogais: precedidas de consoantes sonoras, têm a frequência fundamental diminuída e precedidas de surdas, têm

a frequência fundamental elevada. A perturbação da vogal – que no canto deve ser percebida como desafinação – fica clara no registro agudo, quando há diminuição da frequência fundamental, qualquer que seja a consoante.

O trabalho de 1991 “*L’organisation temporelle de la syllabe dans la parole et dans le chant*” (Scotto di Carlo et Autesserre) aproxima-se mais do estudo deste capítulo e chega a conclusão semelhante à nossa. A primeira semelhança com nosso trabalho é que, diferentemente do trabalho com seqüência CV, o experimento de Scotto e Autesserre contempla trechos de óperas, e embora os autores não explicitem quais, imaginamos que sejam ambientes que reproduzam bem a concretização da fala cantada. Em segundo lugar, existe a comparação fala/canto, o trecho cantado é também lido em voz alta. A taxa de elocução é controlada, mas de modo diverso do nosso. Para a fala, não se utiliza o metrônomo, somente pedindo ao informante que leia o texto normalmente, depois o mais rápido possível, depois o mais lento possível. A essas três taxas de elocução correspondem os seguintes andamentos musicais, respectivamente: 80 batidas por minuto ( $\downarrow = 80$ ), 160 batidas por minuto ( $\downarrow = 160$ ) e 40 batidas por minuto ( $\downarrow = 40$ ). Usando três taxas de locução diferentes, Scotto e Autesserre puderam chegar a um resultado que mostra o que eles chamaram de elasticidade dos segmentos, na fala e no canto: na fala a duração média das consoantes é de 75 a 134 ms, enquanto no canto, essa duração fica entre 52 e 137 ms (os autores não explicitam, mas devem estar se referindo aos segmentos produzidos na taxa de elocução normal). Em termos de elasticidade, então, fala e canto parecem ser semelhantes,

no tocante às consoantes. Em termos de organização temporal da sílaba, no entanto, a conclusão é a seguinte:

En résumé, on peut dire que dans le chant, l'organisation temporelle de la syllable est différente de celle qui existe dans la parole et qu'il n'y a pas de tendance à l'isochronisme syllabique. D'autre part, l'allongement syllabique très important entraîne une modification de la durée relative des consonnes et des voyelles à l'intérieur de la syllabe, ce qui a pour effet d'augmenter le pourcentage des voyelles et de diminuer celui des consonnes. Enfin, l'élasticité des consonnes chantées est pratiquement identique à celles des consonnes parlées, em raison de contraintes physiologiques très puissantes. (p.4)

Podemos dizer agora, como o leitor poderá constatar ao fim do capítulo, que a conclusão a que chegam Scotto di Carlo e Autesserre é basicamente a mesma a que chegamos. É prudente e necessário, no entanto, evidenciar as diferenças entre os trabalhos em questão, a começar pela língua investigada, a taxa de elocução, o controle da frequência fundamental e a escolha das consoantes. Ou seja, o tipo de controle de nosso experimento é diferente do experimento de Scotto, o que esperamos, nos levem a conclusões mais precisas, dentro mesmo do escopo do nosso trabalho, que propõe estudos de aspectos fonético-acústicos co-relacionados.

Já, os estudos, sobre a voz cantada, desenvolvidos por Sundberg (1987), autor sem dúvida representativo na área da acústica da voz cantada, não tratam da questão. Nos nove capítulos de seu livro *The Science of the Singing Voice*, não há sequer uma seção dedicada aos sons consonantais no canto, ao passo que as questões relacionadas às vogais estão presentes nas diversas discussões do capítulo 2, *The Voice Organ*, assim como estão presentes no capítulo 4, *The Voice Source*; nº 5, *Articulation*; nº 6, *Choral Voice* e nº 8, *A Rhapsody on Perception*.

O estudo que ora se apresenta volta-se para o fenômeno da estruturação temporal das consoantes oclusivas surdas [p,t,k] no canto, comparando-o à fala. O que nos baliza, na investigação, pois, são os estudos fonético-acústicos sobre a fala e canto, como os brevemente resumidos acima. Nosso objetivo é integrar o estudo deste capítulo aos outros dois dos próximos capítulos, a fim de fechar o círculo aberto em nossas buscas iniciais, as quais indicaram uma relação entre coordenação temporal dos gestos e as manobras próprias do canto.

## 1.2 Medindo as durações de /la'CV/

Como já foi dito na Introdução, uma mesma frase serviu para todos os propósitos dessa tese de doutorado, os quais necessitaram de medidas de formantes e de duração de segmentos tônicos.

Também constam da Introdução, na seção 3.2.1, detalhes importantes do *design* experimental no tocante a este capítulo, explicitando como resolvemos as questões de taxa de elocução (andamento) e adaptação da frase-veículo ao ritmo da canção. É bom, agora, lembrar que a frase veículo contendo o logatoma / la'CV/ , “Canto /la'CV/ baixinho numa velha canção de ninar” foi repetida cinco vezes por cinco cantoras (sopranos) em duas modalidades diferentes: fala e canto. Obtivemos, no total, 1050 ocorrências, nas quais encontram-se as sete vogais do PB em posição acentual tônica ([a, ε, e, i, o, o e u]) e as consoantes [p,t,k], faladas e cantadas.

Para esse capítulo, embora interessem-nos, principalmente, as consoantes oclusivas [p,t,k], todos os outros segmentos do logatoma /la'CV/ foram medidos: o [l], a vogal pré-tônica [a], a consoante oclusiva e a vogal-alvo. A seguir, expomos 10 figuras da forma da

onda e seus respectivos espectrogramas, cada uma mostrando o posicionamento dos cursores a cada porção medida.

Figura 1: logatoma falado /la'ka/ entre cursores

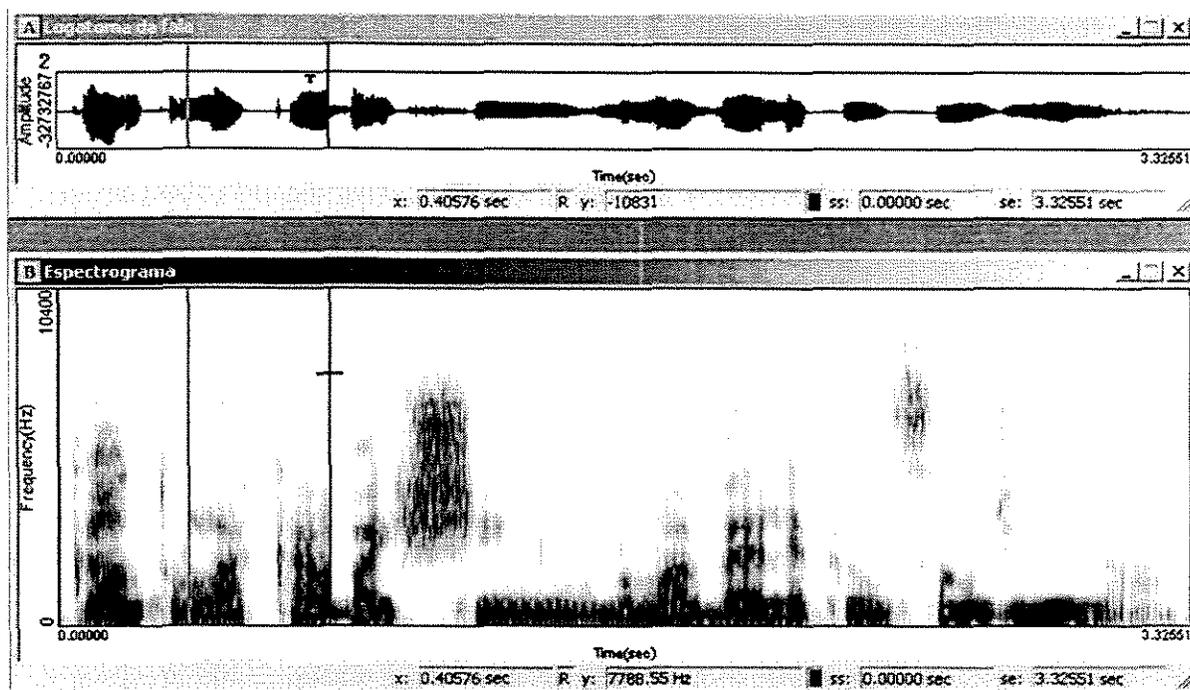
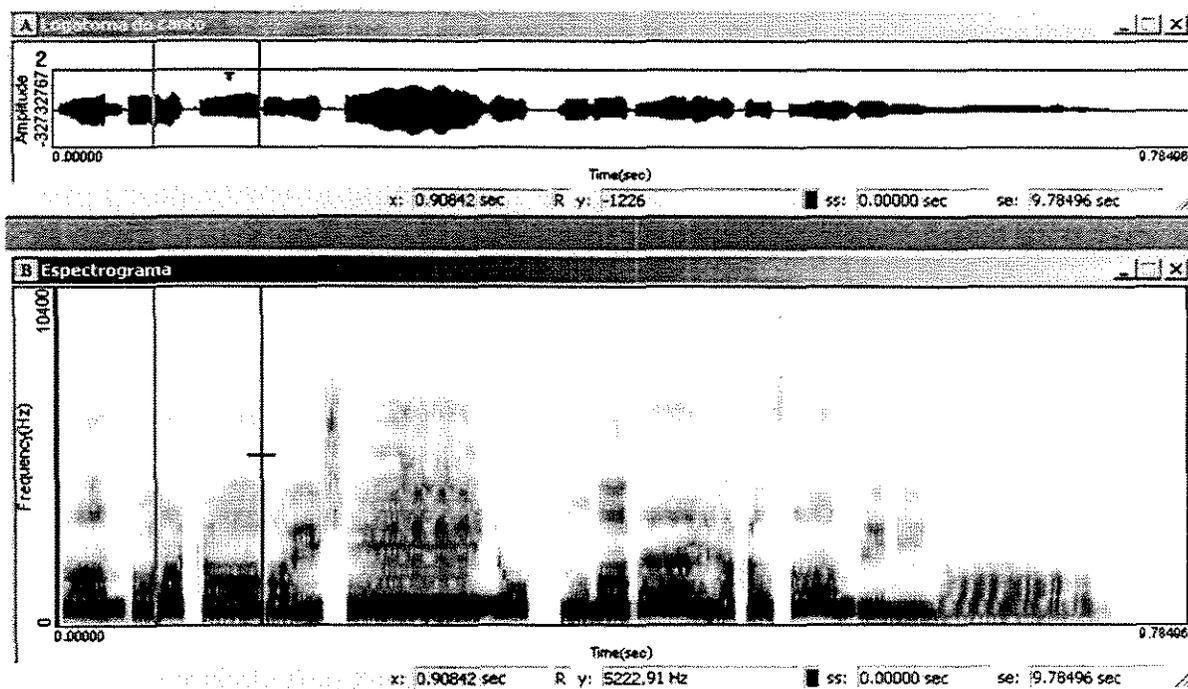


Figura 2: logatoma /la'ka/ cantado entre cursores



As figuras de 3 a 10, a seguir, com o mesmo logatoma /la'ka/, falado e cantado pela informante AK, mostram o posicionamento dos cursores - ora na fala, ora no canto – ao se medir a consoante [k], a vogal-alvo [a], a vogal pré-tônica [a] e a soante [l].

Figura 3: consoante [k] falada entre cursores

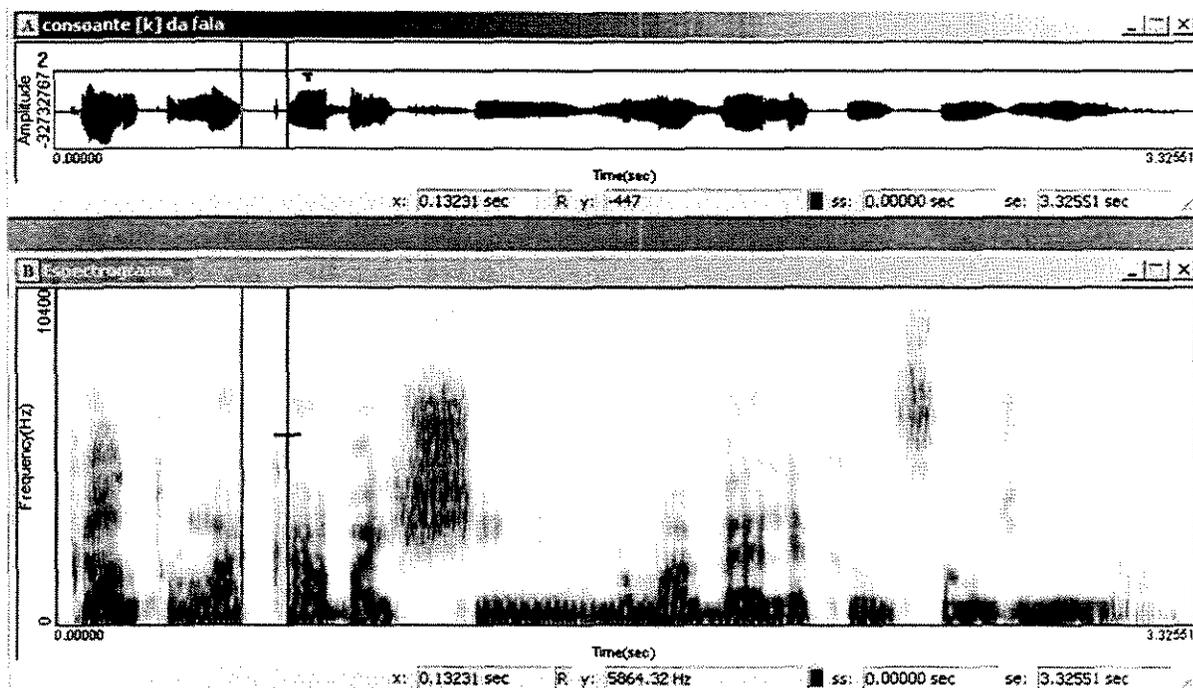


Figura 4: consoante [k] cantada entre cursores

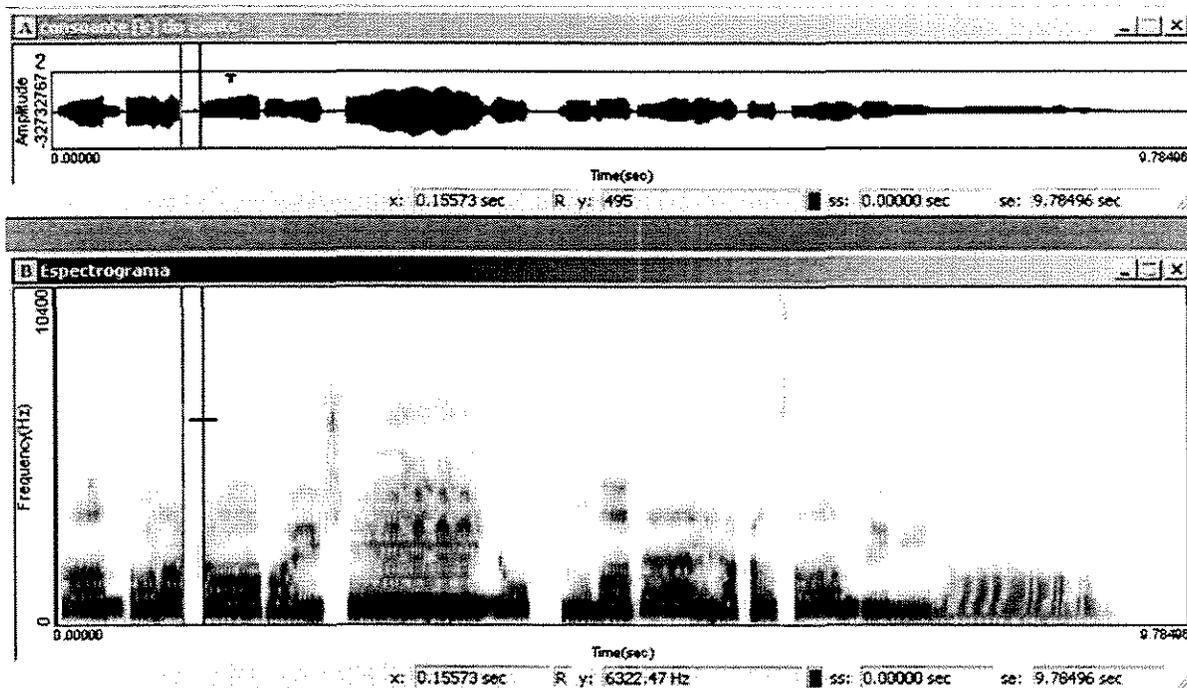


Figura 5: vogal [a] falada entre cursores

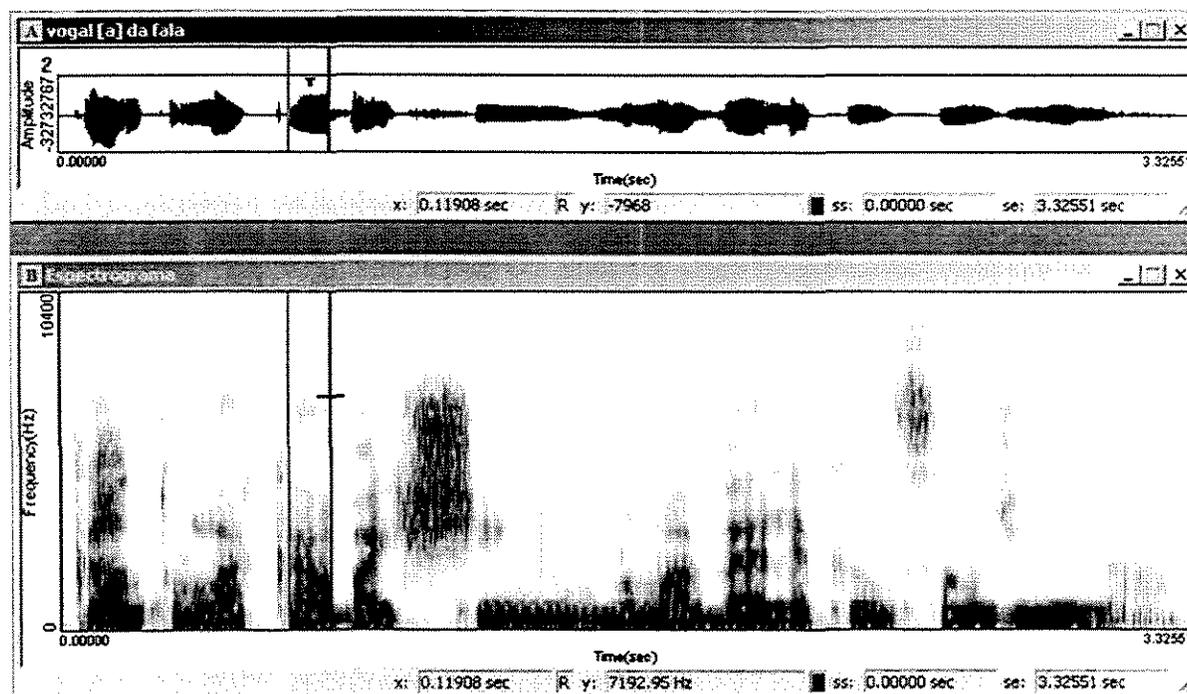


Figura 6 : vogal [a] cantada entre cursores

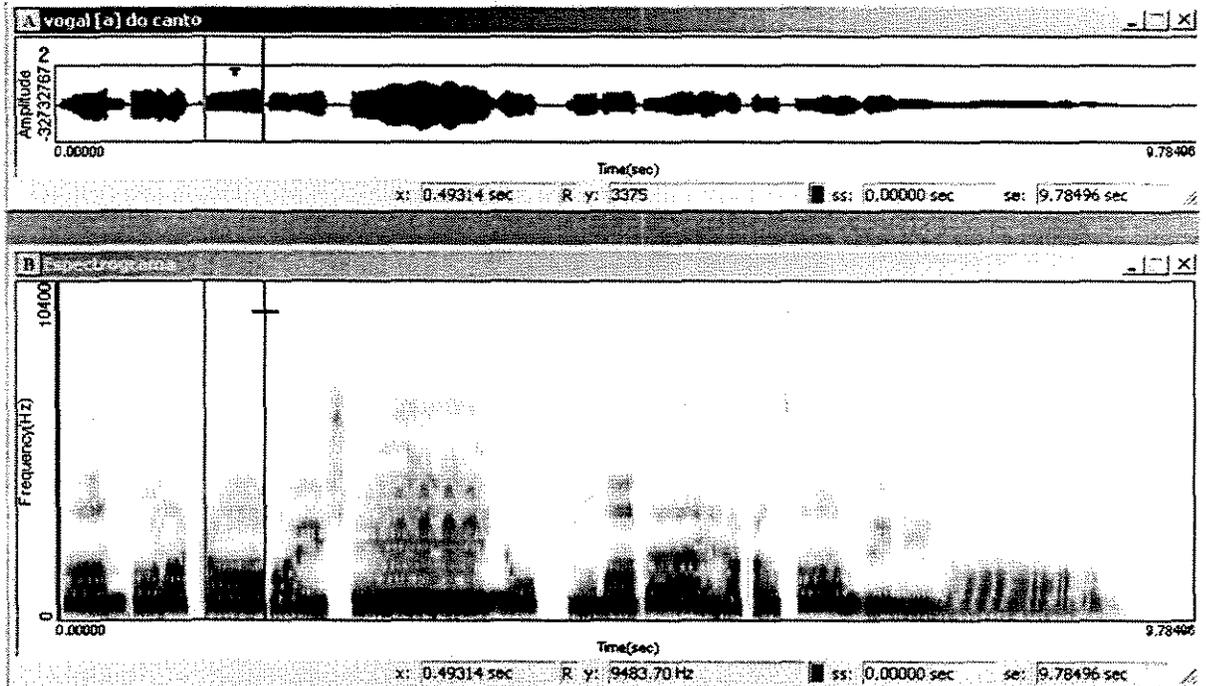


Figura 7 : vogal pré-tônica falada entre cursores

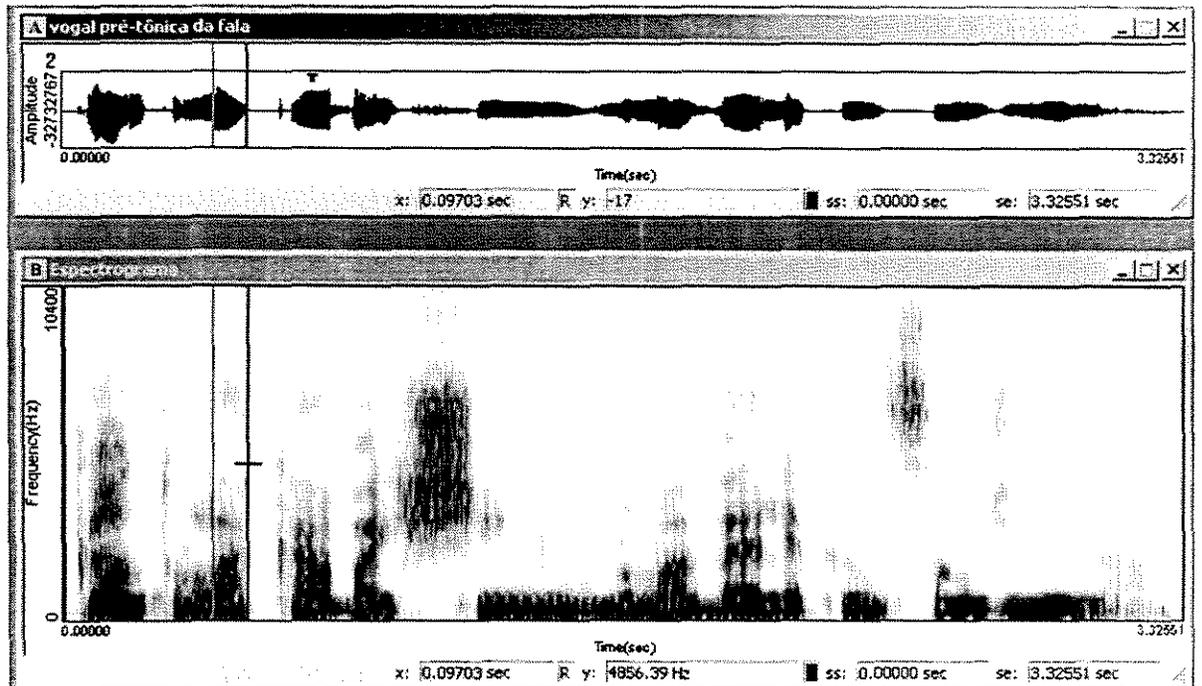


Figura 8: vogal pré-tônica cantada entre cursores

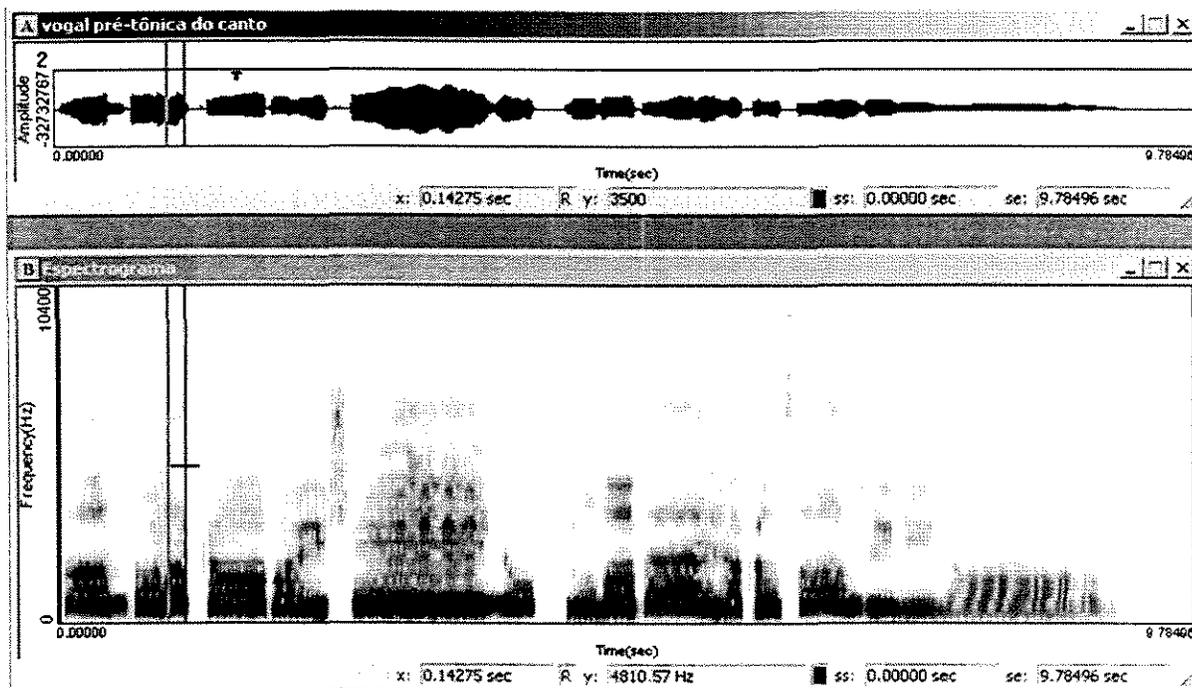


Figura 9: [l] falado entre cursores

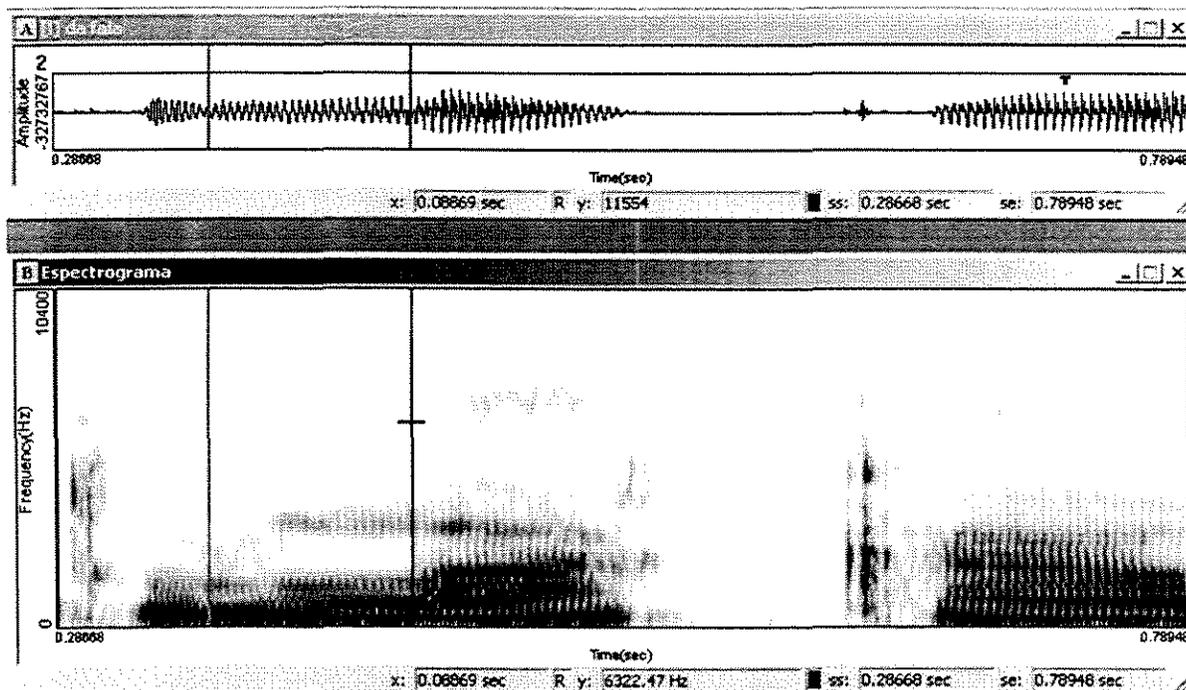
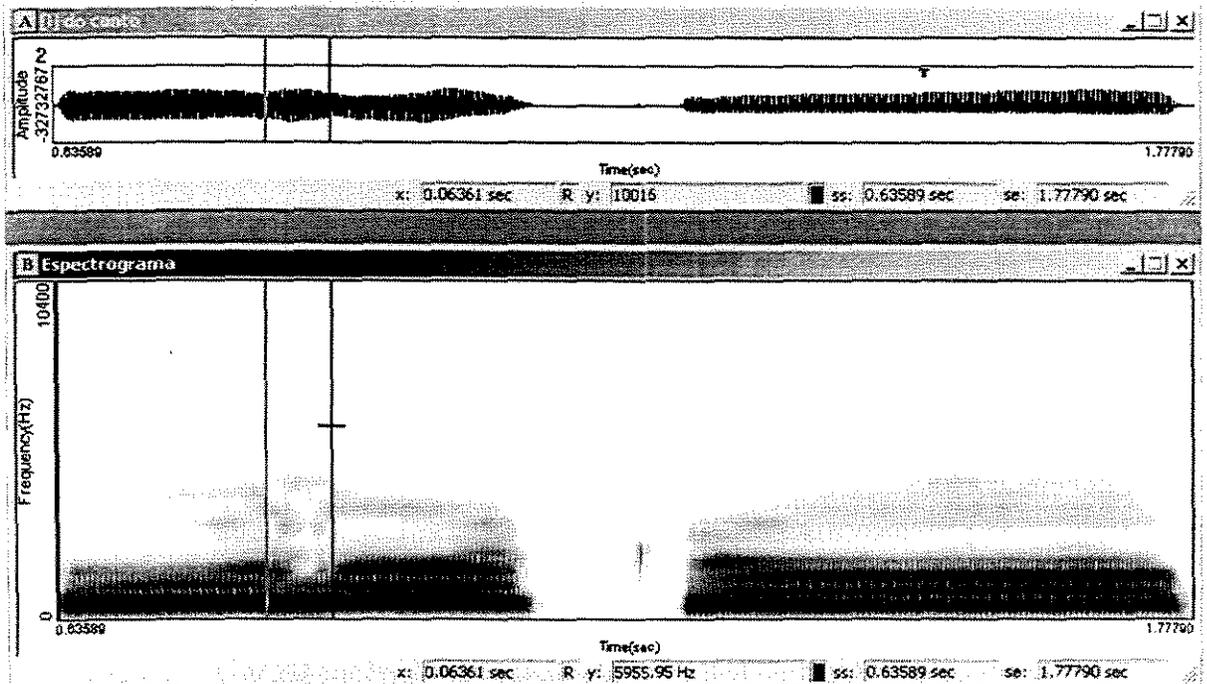


Figura 10: [l] cantado entre cursores



### 1.3 Resultados

Ao olharmos com atenção as figuras 1 e 2 (na página 57), observamos, ao ler o resultado do campo x, localizado na parte inferior direita da figura, que o logatoma do canto é bem mais longo (duração em milissegundos) do que o da fala, praticamente o dobro. Esse dobro é exato em quantidade de batidas, uma vez que a taxa de elocução da fala e do canto era a mesma e as informantes realizaram-na rigorosamente a tempo. Por exemplo: na fala temos sete batidas, contando-se uma pausa ou ralentando ao final de *baixinho* que marca uma fronteira prosódica forte (Fujimura, 1994). Vejamos:

Canto lacá baixinho numa velha canção de ninar

x x x x x x x

No canto temos catorze batidas, correspondentes obviamente às duas batidas de cada um dos sete compassos que formam a frase veículo cantada em 2/4:

Canto lacá baixi      nho numa velha canção de ninar

| V | | | | V | V | V |

• • • • • • • • • • • • • •

Atente-se, primeiramente, para o fato de as batidas da fala obedecerem a uma unidade rítmica que é a palavra fonológica, a qual estaria para a fala como o compasso está para o canto. A palavra fonológica contém a sílaba acentuada em torno da qual se agrupam sílabas fracas. Isto fica evidente quando analisamos *numa velha* como uma palavra fonológica em que *numa* é um clítico e não recebe acento, que recai sobre /vɛ/ de *velha*. É preciso ressaltar aqui o caso das fronteiras que podem conter pausa silenciosa.

Já o canto, para o qual utilizou-se uma camada de traços que indicam quais sílabas foram agrupadas numa batida musical e que também correspondem ao número de notas musicais no trecho em questão, a unidade rítmica é o compasso, que, segundo exige a partitura, deve ter duas batidas. As bolinhas indicam tais batidas, que na fala estão indicadas por *x*. É então a arquitetura rítmica musical - que pode agrupar duas sílabas numa só batida ou tempo, ou alongar uma sílaba como /xi/ em dois tempos - e o andamento pedido (andante) que alongam a frase-veículo do canto em relação à da fala. Daí haver uma correspondência apenas relativa entre a frase falada e a cantada, e não absoluta, em termos de duração.

Diante do exposto, a simples comparação dos valores absolutos de duração entre uma modalidade e outra não nos levaria a nenhuma conclusão. Uma vez realizadas todas as medidas, começamos por extrair as razões do logatoma e seus segmentos intra-fala e intracanto. Assim obtivemos as porções relativas de quanto daqueles 1000 milissegundos do logatoma cantado e quanto dos 450 milissegundos do logatoma falado eram ocupados pelo [l], pela vogal pré-tônica [a], pela consoante oclusiva e pela vogal-alvo, respectivamente. Para ilustrar o que estamos falando, seguem abaixo as tabelas com os valores médios absolutos e as razões médias.

### 1.3.1 Padrão geral

Chamamos de padrão geral o resultado envolvendo os valores médios de cada porção medida de todas as informantes. Esse padrão é apresentado nas tabelas de 1 a 9 abaixo, que contêm as médias e os resultados estatísticos (valores absolutos e razões) de significância (s) ou não significância (ns), respectivamente com  $P$  menor que 0.05 e maior do que 0.05.

Tabela 1: Padrão geral do logatoma. Todas as informantes, vogais e consoantes.

Duração do logatoma (ms)			
Fala	Canto	P	Sig.
451	1005	0.000	s

Tabela 2: Padrão geral das consoantes oclusivas [p,t,k] Todas as informantes, vogais e consoantes.

Duração da consoante oclusiva (ms)			
Fala	Canto	P	Sig.
155	151	0.3385	ns

Tabela 3: Padrão geral das vogais [a, ε, e, i, o, o e u]. Todas as informantes, vogais e consoantes.

Duração da vogal alvo (ms)			
Fala	Canto	P	Sig.
114	534	0.000	s

Tabela 4: Padrão geral do [l]. Todas as informantes, vogais e consoantes.

Duração do [l]			
Fala	Canto	P	Sig.
77	87	0.000	s

Tabela 5: Padrão geral da vogal pré-tônica [a]. Todas as informantes, vogais e consoantes.

Duração da pré-tônica (ms)			
Fala	Canto	P	Sig.
107	233	0.000	s

As tabelas acima resumem, então, a comparação que envolveu 525 ocorrências da fala e as outras 525 do canto, tendo sido submetidas as medidas à análise estatística, empregando-se o procedimento de Análise de Variância para Medidas Repetidas, o de Modelo Linear Geral e o Teste de *Duncan*, para comparações aos pares com alfa fixado em 0.05. Obtiveram-se valores de *P* inferiores a  $10^{-4}$ , exceto no caso das consoantes oclusivas, que, em termos absolutos, apresentam duração igual na fala e no canto.

O logatoma falado tem cerca de metade da duração do logatoma cantado (v. Gráfico 1, página seguinte). A consoante oclusiva falada tem duração semelhante à cantada, sendo o resultado estatístico de não-significância (ver gráfico 2, p. 67). A vogal-alvo cantada é

quase cinco vezes mais longa do que a falada (ver gráfico 3, p. 67). A soante [l] é alongada em dez milissegundos no canto e a vogal pré-tônica é duas vezes maior nesta modalidade do que na falada (ver gráficos 4 e 5, p.68). Este resultado já era esperado, dado o alongamento da vogal no canto, que é uma exigência musical. O que mais nos interessava investigar, exatamente, é o que acontece com a consoante oclusiva no canto (ver exemplo p. 64). A partir da comparação dos valores brutos, a nossa expectativa era de que, para a razão, o resultado seria significativo, uma vez que o logatoma cantado tem o dobro da duração do falado.

Gráfico 1: Duração do logatoma: fala x canto

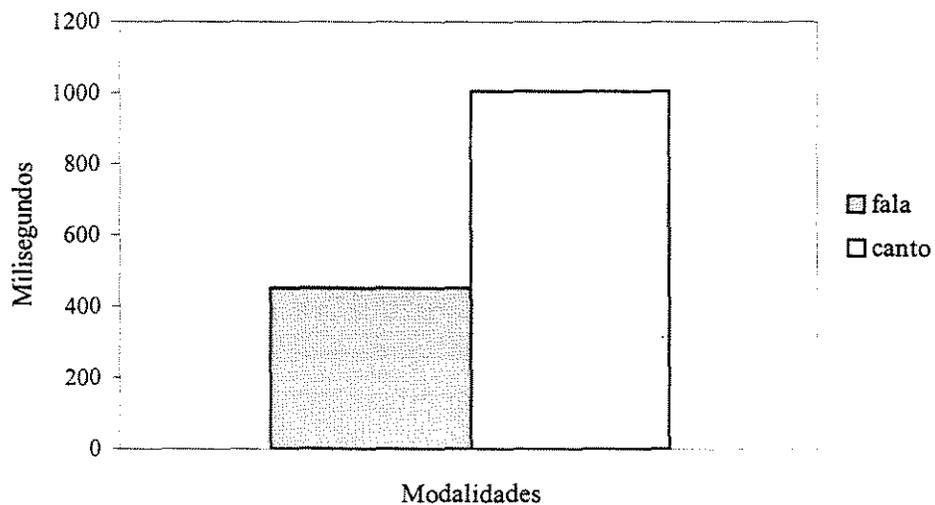


Gráfico 2: Duração da consoante oclusiva: fala x canto

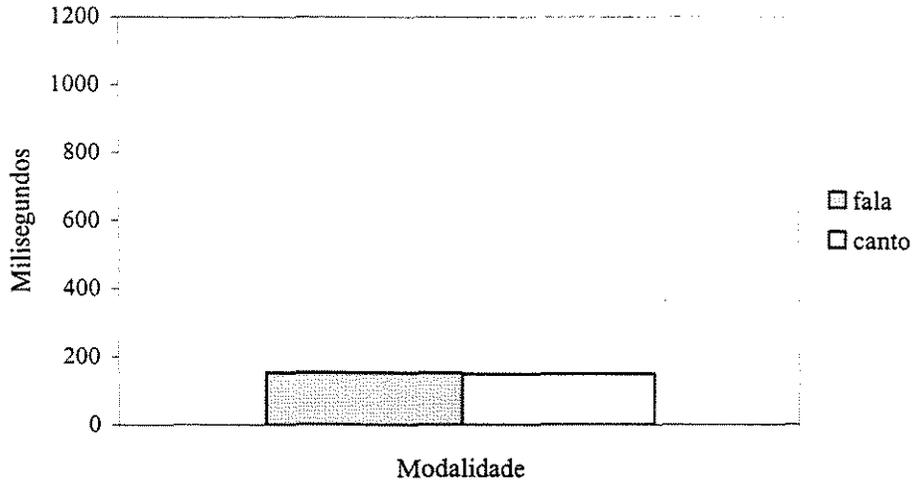


Gráfico 3: Duração da vogal alvo: fala x canto

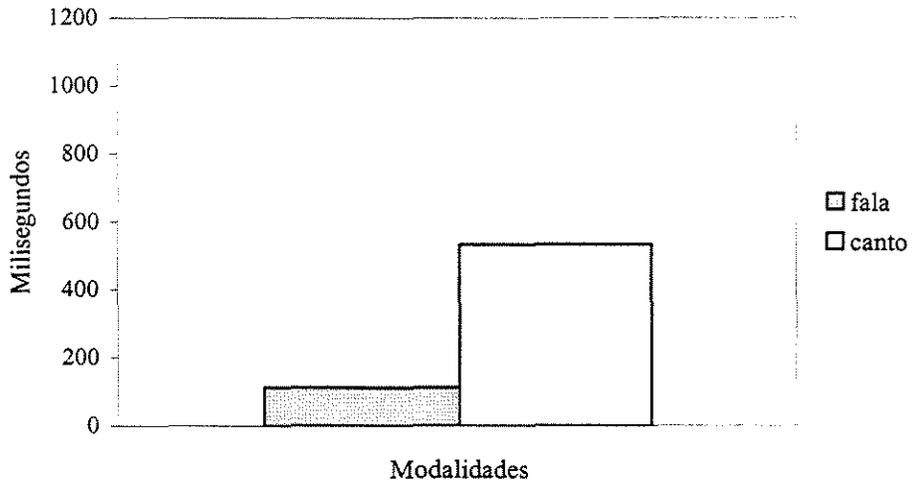


Gráfico 4: Duração do [l]: fala x canto

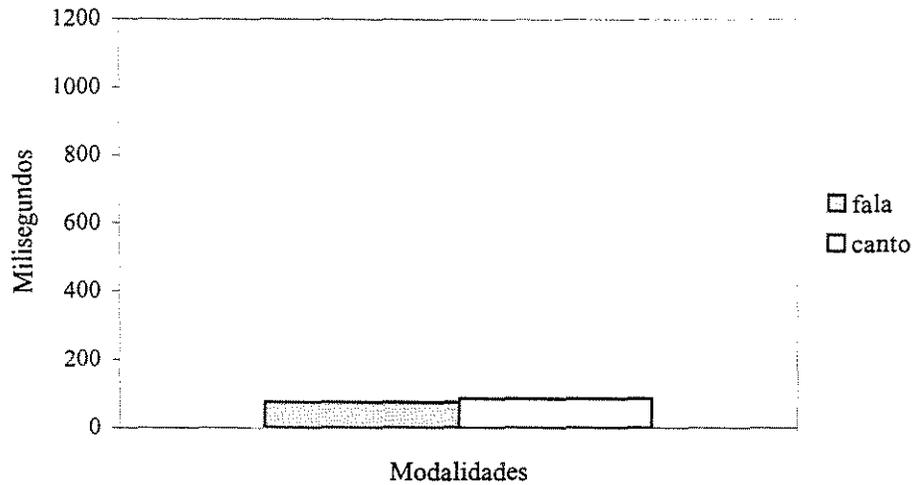
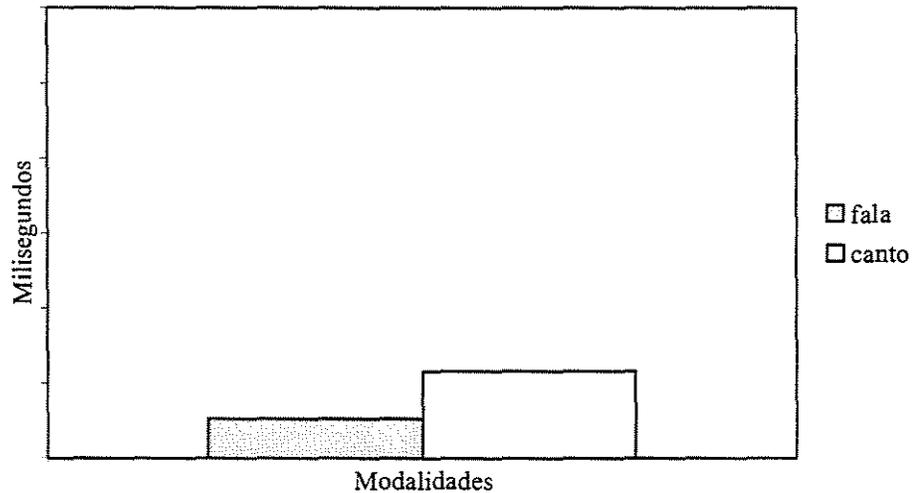


Gráfico 5: Duração da pré-tônica: fala x canto



Da tabela 6 a 9, abaixo, temos os resultados estatísticos e os valores médios das razões extraídas entre logatoma e cada porção dele. A tabela abaixo nos revela, afinal, quanto a consoante ocupa do logatoma na fala e no canto.

Tabela 6: Padrão geral da razão consoante/logatoma oclusiva. Todas as informantes, vogais e consoantes.

Razão consoante/logatoma			
Fala	Canto	P	Sig.
34%	15%	0.000	s

Tabela 7: Padrão geral da razão vogal alvo/logatoma. Todas as informantes, vogais e consoantes.

Razão vogal alvo/logatoma			
Fala	Canto	P	Sig.
25%	53%	0.000	s

Tabela 8: Padrão geral da razão vogal pré-tônica/logatoma [a]. Todas as informantes, vogais e consoantes.

Razão vogal pré-tônica/logatoma			
Fala	Canto	P	Sig.
23%	23%	0.4646	ns

Tabela 9: Padrão geral da razão [l]/logatoma Todas as informantes, vogais e consoantes.

Razão [l]/logatoma			
Fala	Canto	P	Sig.
17%	8%	0.000	s

Em termos relativos, então, vemos que a consoante oclusiva tem importância diferente na fala comparada ao canto, ocupando 34% do logatoma, ou seja, o segmento de maior duração deste último. No canto, a oclusiva ocupa 15% (ver gráficos 6 e 7, página seguinte) do total do logatoma, sendo o segundo segmento menor em duração, pois vem

depois do [l] que ocupa 8% do total. Os dois gráficos (6 e 7, a seguir) que ilustram a razão, a seguir, são constituídos de formas

Gráfico 6: Razão logatoma/consoante oclusiva na fala

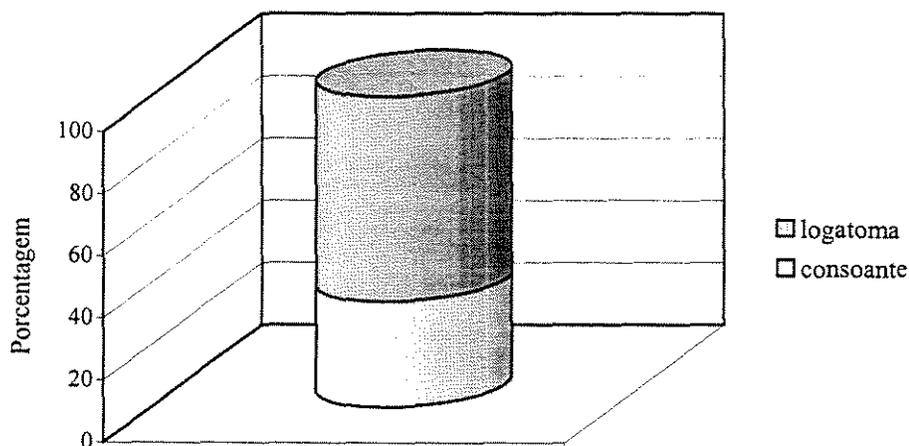
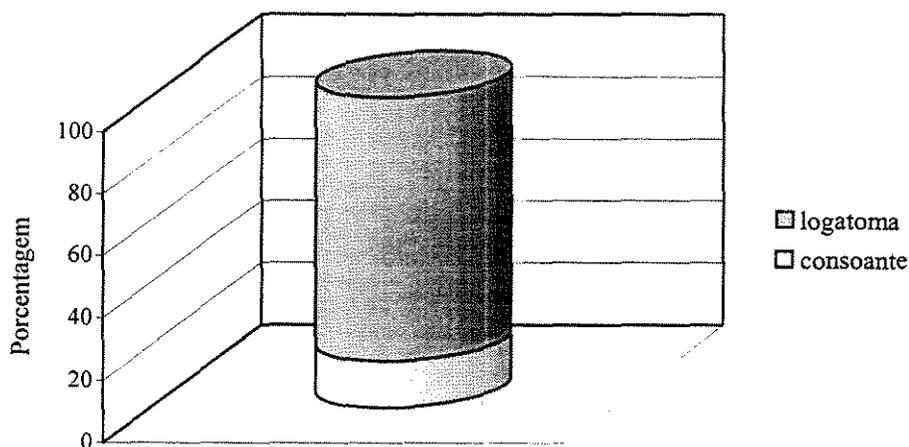


Gráfico 7: Razão logatoma/consoante oclusiva no canto



cilíndricas cujo total (100%) representa o logatoma, sendo as porções medidas, tanto as da fala como as do canto, representadas por cinza claro.

A vogal-alvo, no canto, é inequivocamente alongada e a pré-tônica [a] é o único segmento que parece desempenhar o mesmo papel nas duas modalidades. Tanto na fala como no canto, esta se realiza – para utilizar um termo musical – no contratempo, ou seja, na batida fraca (ver gráficos 8, 9, 10 e 11).

Gráfico 8: Razão logatoma/vogal-alvo na fala

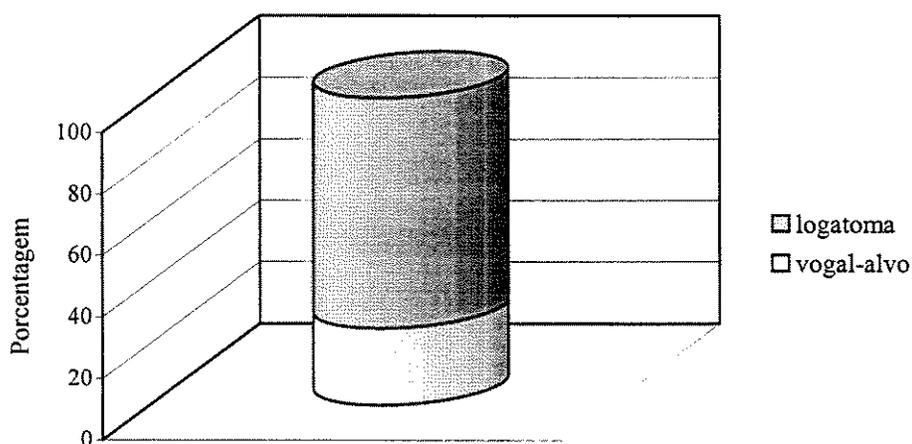


Gráfico 9: Razão logatoma/vogal-alvo no canto

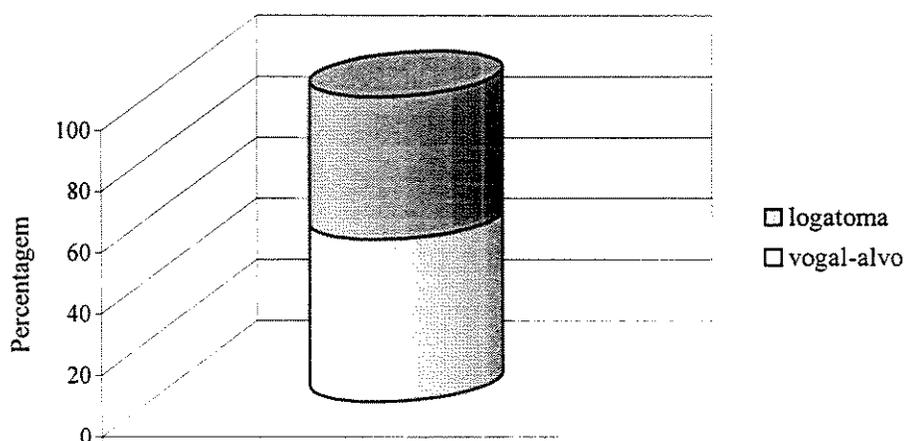


Gráfico 10: Razão logatoma/pré-tônica na fala

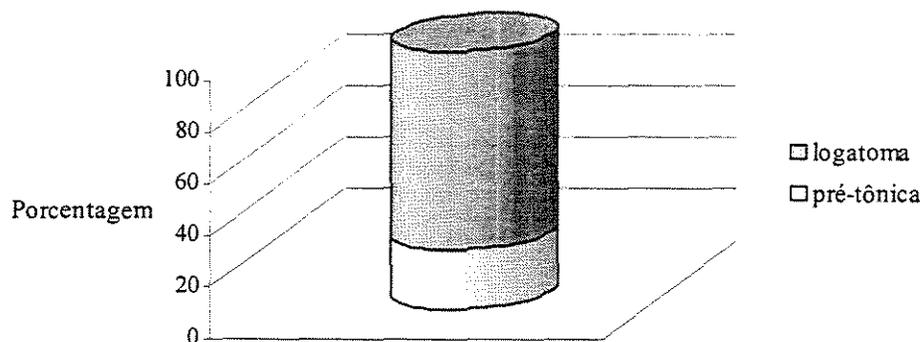
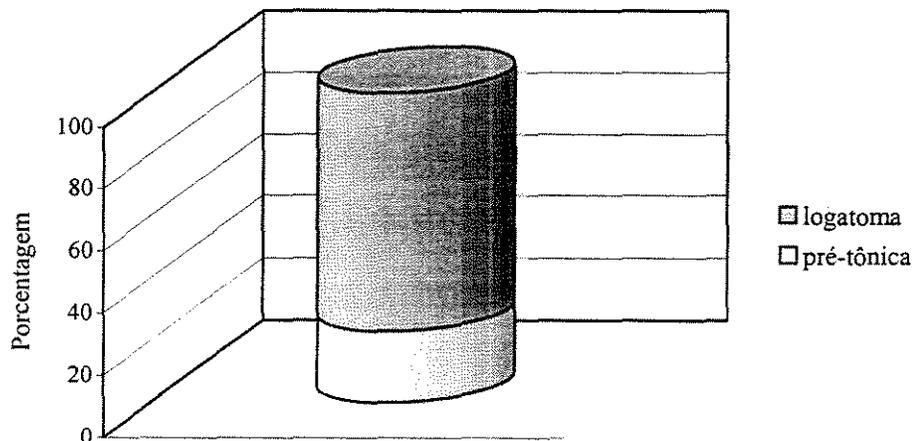


Gráfico 11: Razão logatoma/pré-tônica no canto



O [l] tem um comportamento curioso. Apesar de se alongar no canto em termos absolutos, representa apenas metade dos 17% que ocupa no logatoma da fala (ver gráficos 12 e 13, a seguir). Seguem, portanto, em termos relativos, as tendências gerais das consoantes:  $34/15$  (oclusivas)  $\approx 17/8$  ([l]).

Gráfico 12: Razão logatoma/[l] na fala

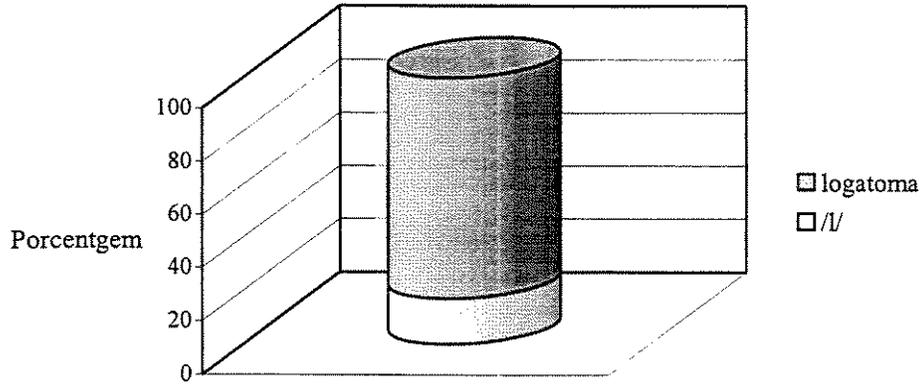
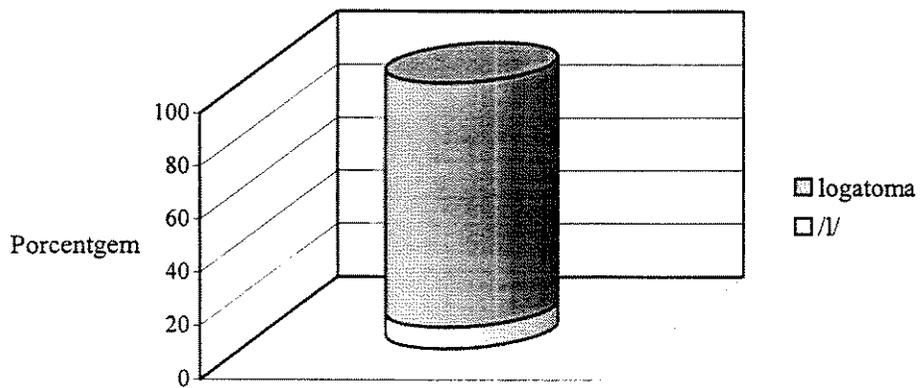


Gráfico 13: Razão logatoma/[l] no canto



A partir dos resultados acima, podemos dizer que a fala, em comparação ao canto, dá importância às consoantes, enquanto este último destaca a importância das vogais. A

importância atribuída a cada classe dos sons da fala explica-se pelas restrições próprias de cada modalidade, o que será discutido ao final do capítulo.

### 1.3.2 Padrão por consoante oclusiva

É importante visualizar agora o que acontece com cada consoante oclusiva em particular, ao que chamamos de *padrão da consoante oclusiva*, e que ilustramos com as tabelas abaixo.

Tabela 10: Padrão da duração. Todas as informantes e todas as vogais

Consoante oclusiva [k]				
	Fala	Canto	P	Sig.
Dur. [k]	162	155	0.0005	s
Dur. log.	455	1014	0.0001	s
Dur. vogal	110	532	0.0001	s
Dur [l]	79	87	0.0001	s
Dur pre to.	104	240	0.0001	s

Tabela 11: Padrão da duração. Todas as informantes e todas as vogais

Consoante oclusiva [p]				
	Fala	Canto	P	Sig.
Dur. [p]	147	147	0.8752	ns
Dur. Log.	439	999	0.0001	s
Dur. vogal	117	538	0.0001	s
Dur [l]	76	87	0.0001	s
Dur pre to.	99	227	0.0001	s

Tabela 12: Padrão da duração. Todas as informantes e todas as vogais

Consoante oclusiva [t]				
	Fala	Canto	P	Sig.
Dur. [t]	156	150	0.0036	s
Dur. Log.	458	1000	0.0001	s
Dur. vogal	116	533	0.0001	s
Dur [l]	77	86	0.0001	s
Dur pre to.	109	231	0.0001	s

Comparando as tabelas acima com as tabelas do padrão geral, é possível ver que o padrão de cada consoante –sobretudo a consoante [p] - repete aquele último padrão. A consoante apresenta durações semelhantes nas duas modalidades, o logatoma dura mais ou menos o dobro, as vogais-alvo são quase cinco vezes mais longas, o [l] é alongado, absoluta, mas não relativamente; e a pré-tônica é mais do que o dobro no canto, mantendo-se, no entanto, constante em termos relativos.

Quanto ao [k] temos que é sete milissegundos mais longo na fala do que o canto e, quanto ao [t], esse alongamento é de seis milissegundos. Ambas as diferenças são mínimas, mas se revelaram significativas. Isso significa que a dispersão das medidas foi muito pequena, o que aponta para uma relativa precisão das manobras articulatórias envolvidas na produção de consoantes. Tal fenômeno confirma o que havíamos dito sobre a importância da fala atribuída às consoantes. As duas consoantes em questão envolvem movimentos de língua, a primeira envolvendo o dorso e a segunda, a ponta da língua, com velocidade talvez menor do que aquela com que se produz [p], tanto no canto como na fala.

Quanto à vogal-alvo, sua duração, na fala é inversamente proporcional à da consoante oclusiva: quanto mais longa a consoante, menos longa a vogal. Isto acontece no canto, guardadas as devidas proporções, ou seja: a vogal cantada realizada no ambiente [k] (532 ms), de modo geral, que é o que nos indica o valor médio, é mais curta do que a vogal que se realiza no ambiente [p] (538 ms), por exemplo.

Segue, também, o padrão geral, o comportamento dos demais segmentos tônicos. Verificamos a seguir, vendo as tabelas com os resultados das razões, que estes também se repetem em relação ao padrão geral.

Tabela 13: Padrão da razão de [k]. Todas as informantes e todas as vogais

Consoante oclusiva [k]				
	Fala	Canto	P	Sig.
R. Cons/L	35%	15%	0.0001	s
R. Vo/L	23%	52%	0.0001	s
R. [l] / L	17%	8%	0.0001	s
R. Pre t./L	23%	22%	0.0115	s

Tabela 14: Padrão da razão de [p]. Todas as informantes e todas as vogais

Consoante oclusiva [p]				
	Fala	Canto	P	Sig.
R. Cons/L	33%	14%	0.0001	s
R. Vo/L	26%	53%	0.0001	s
R. [l] / L	17%	8%	0.0001	s
R. Pre t./L	22%	22%	0.4487	ns

Tabela 15: Padrão da razão de [t]. Todas as informantes e todas as vogais

Consoante oclusiva [t]				
	Fala	Canto	P	Sig.
R. Cons/L	34%	15%	0.0001	s
R. Vo/L	25%	53%	0.0001	s
R. [l] / L	16%	8%	0.0001	s
R. Pre t/L	24%	23%	0.0172	s

As vogais pré-tônicas faladas dos logotomas com [k] e [t] são significativamente diferentes das pré-tônicas cantadas nos logotomas correspondentes. Mas, percentualmente, a diferença é de apenas 1%. Isso, de novo, aponta para uma relativa precisão do “contratempo”, i. e., dos segmentos fônicos, consoantes e vogais, que ocupam posições fracas. Essa precisão de que falamos aqui corrobora a hipótese de um relógio interno conforme proposto por Barbosa e Bailly (1994) à luz de comparações do ritmo da fala com o de outras atividades motoras, tais como estudadas por Kugler e Turvey (1987).

No que diz respeito aos demais segmentos fônicos, o padrão é o mesmo já apresentado nas tabelas do padrão geral.

#### 1.4 Discussão final

Intuitivamente, ou com conhecimento da natureza da distinção básica entre os sons da fala (vogais, sons quase-periódicos e obstruintes, sons aperiódicos, em maioria freqüentemente ruidosos), ensina-se a cantar dizendo que o som mais importante do canto é a vogal. O aquecimento do instrumento musical dos cantores é feito por meio de vocalises,

que envolvem a produção de diferentes vogais – ainda que prevaleça a vogal [ɔ] - apoiadas, em geral, nas consoantes bilabiais [m]<sup>2</sup> ou [p].

O fato de a vogal ser o som da fala eleito pela música não deve causar nenhuma surpresa. Os sons musicais, para serem considerados como tais, devem possuir uma estabilidade que não se encontra em sons não musicais. No capítulo 5 de Kent & Read (1992), *The Acoustic Characteristics of Vowels and Diphtongs*, temos a seguinte definição de vogais:

(...) vowels are associated with a steady-state articulatory configuration and a steady-state acoustic pattern. Supposedly, then, a vowel can be indefinitely prolonged as an articulatory or acoustic phenomenon. (p.87)

Ora, o prolongamento indefinido de que nos falamos Kent e Read (op.cit.) é também o que acontece com uma nota musical. Assim, a vogal cantada segundo as partituras, tal como as conhecemos na música ocidental, deve obedecer à duração da nota, pois, no canto apenas vocalizado ou com texto, ela é a própria realização desta nota.

Assim é que, pela natureza mesma da vogal e pelas exigências temporais da realização de sons musicais, a vogal é alongada no canto em relação à fala. O que acontece, então com o som da fala despojado de características musicais, a consoante?

---

<sup>2</sup> A soante nasal tem um papel importante nestes vocalises, segundo os professores de canto, por “abrir” o ressoador nasal. A soante investigada neste capítulo, o [l], apresenta-se mais longa no canto que na fala em termos absolutos, indicando que ela tem um comportamento diferente das obstruintes na primeira modalidade. Talvez tal comportamento pudesse ser realmente diferenciado caso a soante pertencesse a uma sílaba tônica, mas isso foge ao escopo do presente estudo. Por ora, basta destacar que, na classe das consoantes, as obstruintes são em maior número, merecendo mais atenção deste estudo do que as soantes, o que não implica, necessariamente, o desinteresse de um estudo sobre estas últimas.

Responderemos à essa questão utilizando como exemplo as três oclusivas surdas que foram o foco da investigação nesta parte do estudo. Vimos que [p, t, k] ocupam uma porção importante do logatoma (35%) na fala e que essa importância cai para a metade no canto, muito embora, em termos absolutos, tenham durações semelhantes. A vogal, por sua vez, em termos relativos ou absolutos é sempre mais longa no canto que na fala, ocupando 53% do logatoma cantado e 23% do falado. Daí concluímos que as consoantes oclusivas surdas cantadas são encurtadas no canto (à exceção do [p] em termos absolutos), como pudemos aventar inicialmente, e têm uma importância relativa numa e noutra modalidade.

As duas modalidades investigadas apresentam-se bastante diferentes no tocante à organização da estrutura temporal dos gestos da fala. A coordenação dos gestos cantados minimiza as características acústicas de silêncio e ruído que marcam a consoante, pelo fato de a voz cantada ser instrumento musical de alturas e não rítmico e assim necessitar das vogais, sons da fala que possuem uma estrutura harmônica.

A coordenação dos gestos falados, comparados aos cantados, privilegia a consoante; vimos em Kessinger e Blumstein (1998), que quando a vogal se alonga, o VOT também se alonga; isto também deve ocorrer quando há foco ou acento frasal, provavelmente aumentando a duração da consoante. Por outro lado vimos em Stevens (2000) que a descontinuidade do sinal da fala marcado por uma mudança repentina cria pistas acústicas importantes em ambos os lados da consoante, que são fundamentais para sua identificação. Tais pistas parecem se manter no canto, dado que há um tempo mínimo de duração para as oclusivas investigadas.

Temos então, de um lado, a fala, em que as consoantes oclusivas são realizadas num tempo considerável, e, de outro, o canto, em que tais consoantes realizam-se num tempo

que poderíamos chamar de tempo suficiente, cuja função é ancorar a textualidade cantada, numa espécie de negociação entre as duas modalidades.

## Capítulo 2

### Padrão formântico de vogais na fala e no canto

É a boca. É a boca também a exigir que o *bâijo* portuga se transfigurasse num *bêjo* porventura de lábios mais grossos. As linguagens crescem e se transformam, não por “vícios de linguagem”, mas pelas exigências psicofísicas das gentes. São estas exigências que fazem as variações dos fonemas, as variantes de timbre e movimento, as diferenças sintáxicas de ritmo. (M. de Andrade)

#### 2.1 Introdução

Deixamos o domínio do tempo – tratado no capítulo anterior – para entrar no domínio das frequências. Para traçar o padrão formântico no canto tomamos muitos cuidados – assim como fomos cautelosos também para medir as durações – como, por exemplo, manter todas as vogais investigadas no mesmo *pitch* ou frequência fundamental.

A frequência fundamental da nota musical sobre a qual são produzidas todas as vogais-alvo desse experimento fica em torno de 415 Hz, sendo este um lá bemol, portanto, pertencente a uma região grave-média para o soprano. A frase-veículo, “Canto /la'CV/ baixinho” como já foi explicitado na introdução, abrigou o logatoma – que por sua vez abrigou a vogal-alvo – com os cuidados necessários para não haver interrupção abrupta da vogal. A posição acentual tônica também foi uma escolha baseada no bom senso: era preciso partir de uma análise mais robusta, já que o trabalho se propõe a fazer uma limpeza de terreno, e por isso as vogais átonas foram deixadas de lado, por possivelmente causarem

dificuldades nas medições acústicas. Investigar mais de uma posição acentual implicaria aumentar as variáveis, que já não eram poucas, dado que o estudo envolve duas modalidades; três consoantes, combinadas com sete vogais e cinco informantes. Os sopranos envolvidos na pesquisa foram escolhidos por possuírem vozes leves<sup>1</sup> e repertórios semelhantes. Todas as cantoras têm experiência em música antiga – por exemplo o barroco europeu e o colonial brasileiro – e na música de câmara brasileira, além de serem falantes nativas do português do Brasil, mais precisamente, dos dialetos paulistas de regiões de Jundiaí e Campinas.

Falar de vogais no âmbito da fonética acústica parece fácil à primeira vista, devido aos bem sucedidos e propagados modelos da produção de vogais que conhecemos, o de Fant (1960) e de Chiba e Kajiyama (1941, apud Kent e Read, (1992)). No entanto, há uma literatura extensa referente a estudos de vogais a começar, por exemplo, pelo estudo da vogais do inglês americano de Peterson e Barney (1952).

A fim de contextualizar a investigação do padrão formântico das vogais cantadas em português do Brasil e dar uma noção das questões que ocupam estudiosos a respeito das características acústicas das vogais, resumo, a seguir, alguns artigos relevantes para o estudo do presente capítulo.

## 2.2 Estudos sobre vogais: aspectos articulatórios, acústicos e perceptuais.

No seu clássico artigo de 1952, *Control methods used in a study of vowels*, G. E. Peterson e H. E. Barney reuniram 76 informantes para produzirem as vogais, [ɪ, i, ε, æ, ɛ̃, a, ɔ, ʌ, u, u] do inglês americano num contexto h\_d, a fim de rodar um teste de percepção,

---

<sup>1</sup> Três das informantes; CH, SS e AK são sopranos-ligeiro. MG é lírico-ligeiro; e apenas uma informante, CM, não poderia ser classificada como soprano lírico-ligeiro, mas sim como lírico.

para o qual dos 70 ouvintes, 32 tinham sido infomantes-falantes. Uma das inquietações de Peterson e Barney era de como falantes tão diferentes entre si (33 eram homens, 28, mulheres e 15 crianças) apresentando uma grande variação de produção de vogais, davam conta de identificá-las.

Para interpretar, ou iniciar a interpretação desse fenômeno (outros experimentos foram feitos com base nos dados de Peterson e Barney 1952) os autores mediram os primeiros formantes das vogais do experimento, plotando, a partir destas medidas, um gráfico de  $F_1 \times F_2$ , a fim de visualizarem o que acontecia em termos de produção. A sobreposição era inevitável entre vogais como [ɛ] e [ɜ] e entre [ɑ] e [ɔ], por exemplo. Um resultado interessante a que chegam os autores, é de que o falante que tem uma produção mais variada de um [ɑ], por exemplo, tende também a variar seu julgamento: ora ouve-o como [ɑ], ora como [ɔ]. Uma “confusão” como esta era computada como um erro, uma vez que a vogal ouvida não correspondia à vogal pretendida. A isso, os autores chamaram de experiência prévia da linguagem que exerce influência na produção e na percepção de fala. As vogais que apresentaram o maior número de discordâncias no momento de serem identificadas foram: [ɪ], [ɛ], [ɑ], [ɔ], [ʌ]; sendo que [i] e [u] foram identificadas unanimemente como tais, por 109 vezes das 152 que foram apresentadas aos ouvintes.

Embora se trate de um experimento perceptual, ele nos dá bases para falar de sobreposição de vogais, de posição limite dos mecanismos articulatórios e diferenças entre falantes, por isso nos interessa.

Desobedecendo à ordem cronológica dos textos escolhidos para serem resumidos nessa seção, vamos direto a Syrdal e Gopal (1986), pois o experimento deles toma emprestado o *corpus* de Peterson e Barney (op. cit). Transformando frequências

fundamentais e formantes das vogais em *barks*, ou seja, numa escala psicofisicamente motivada (para detalhes ver os próprios Syrdal e Gopal, 1986), Syrdal e Gopal chegaram à conclusão de que a normalização realizada pelo ouvinte diante da grande variação acústica das vogais, é uma normalização inerente a alguns traços da própria vogal, por isso independente do falante. Por exemplo: uma vez transformados os formantes em *barks*, a variabilidade da distância  $F_2-F_3$  era significativamente menor para as vogais anteriores do que para as posteriores; e a variabilidade da distância  $F_1-F_0$  era significativamente menor para as vogais altas; ora isso implica que a sobreposição que se verificava em *hertz*, não era a mesma verificada em *barks*, o que explica o sucesso de identificação de vogais como [i] e [u].

Voltando à ordem cronológica, resumimos, a seguir, Lindblom e Sundberg (1971), que percorrem o caminho oposto de Syrdal e Gopal. Se, nestes dois últimos autores, temos a preocupação com um modelo perceptual de vogal que dê conta robustamente da ligação entre os traços fonéticos e os co-relatos acústicos, junto aos autores suecos temos um modelo dos articuladores para prever um padrão formântico. Tal modelo constituiu-se com base em medidas feitas a partir de fotografias laterais de raio-X de um informante ao produzir vogais longas. A abertura da boca (lábios) e o abaixamento da mandíbula foram medidos em milímetros, ao passo que, para determinar o “quanto” a língua muda de forma atribuíram-se alguns valores numéricos, conforme a direção da língua no trato. Assim, o valor -1 aplica-se ao formato da língua em direção ao palato duro; 0 significa o formato da língua em direção ao véu, e 1 significa o formato da língua em direção à parte inferior da parede posterior da faringe. Esses valores corresponderiam às vogais [i], [u] e [a], respectivamente. Investigou-se, com base nas medidas feitas do trato vocal, as

conseqüências acústicas dos movimentos dos seguintes articuladores: lábios, mandíbula, língua (ponta e corpo) e laringe. Os efeitos produzidos por tais articuladores são:

- Lábios: quando arredondados, abaixam todos os formantes, quando estirados tendem a encurtar a constrição palatal, o que afeta  $F_3$ .
- Mandíbula: alterações sensíveis sobre  $F_1$ , já conhecidas: quanto mais a mandíbula se abaixa, mais  $F_1$  se eleva.
- Língua: os autores constataram que, para as vogais, a ponta da língua fica em repouso, quase que invariavelmente. A posição e a forma do corpo da língua é que influenciam bastante o comportamento de  $F_2$ .
- Laringe: esta pode ser alta, elevando todos os formantes, ou baixa, diminuindo os valores dos formantes. Os autores não chegaram a uma conclusão de que formante seria mais sensível aos movimentos desse articulador.

O último estudo a ser resumido sobre as vogais da fala é o de Johnson, Ladefoged e Lindau (1993) que aborda as diferenças individuais na produção de vogais a partir de um experimento com cinco informantes falantes do inglês americano do meio oeste. As vogais produzidas foram /i ɪ e<sup>1</sup> ε æ a ɔ o<sup>u</sup> u /, abrigadas pelos contextos /d\_t/ ; /d\_d/; /b\_b/ e /s\_s/. Com sensores de ouro grudadas nos lábios, dentes e língua de cada informante, foi possível rastrear, por meio de raio-x, a trajetória dos articuladores. A conclusão a que chegaram os autores é de que há diferenças articulatorias individuais, devido, basicamente, ao fato de que a geometria do trato vocal difere de falante para falante<sup>2</sup>. Isto implica que os

---

<sup>2</sup> Outras diferenças individuais, que se devem ou não à geometria, são resumidas ao final do texto e são de vários tipos: "The individual differences which we found are interesting because they were of several different types. Speakers differed in terms of characteristic speaking rate, which in Saltzman and Munhall's (1989) task dynamic model is associated with gestural stiffness. There were also differences in how speakers

falantes de uma mesma língua podem não partilhar os mesmos gestos articulatórios, mas partilham o produto acústico destes gestos. Assim, descarta-se uma fonética articulatória universal - hipótese que se queria refutar inicialmente – e defende-se a validade de uma *teoria auditiva da produção de fala*, pelo menos para as vogais.

Quanto a estudos das vogais cantadas, temos em Sundberg (1969, 1975, 1977) os trabalhos mais expressivos. O único que faz uma comparação fala/canto é o de 1969, em que as vogais alvo são nove vogais longas do sueco, num contexto / em dV:t/. O autor conclui que há diferenças entre as frequências formantes da fala e aquelas do canto devido a diferenças articulatórias, como por exemplo o abaixamento da laringe, que diminui  $F_2$  nas vogais anteriores, e uma alteração da forma da língua que causaria a elevação de  $F_3$  nas vogais posteriores. Os resultados a que chega Sundberg, quanto à mudança de formantes do canto em relação à fala são os seguintes:  $F_2$  diminui em todas as vogais, exceto em /u/, /o/, /a/ e /æ/,  $F_3$  se eleva nas vogais posteriores e diminui nas outras,  $F_4$  diminui em todas as vogais e a distância entre  $F_3$  e  $F_4$  reduz-se em todas as vogais. Veremos, na seção a seguir, que estes resultados não diferem muito dos nossos (com a ressalva de que não investigamos o quarto formante), embora sejam os informantes baixos suecos, entoando uma nota grave.

Sundberg (1975) discute a influência do vibrato na identificação da vogal cantada e sua conclusão é de que, embora altere levemente a qualidade da vogal, por tender a ocultar as frequências dos formantes, esta influência é ínfima e ocorre apenas em algumas determinadas condições.

---

produced tense/lax vowel distinctions which appeared to reflect variations in interarticulatory coordination in the model. Speakers differed in how they produced vowel height distinctions, which could also be due to differences in how speakers maintained the distinction between front and back vowels, which seemed to indicate a variation in gestural targets.”

Sundberg (1977), ao qual já nos referimos na introdução, revela-nos um achado de grande relevância para repensar a questão da inteligibilidade no canto e atacar o mito da má dicção: ao investigar a acústica da voz cantada, o autor descobre duas manobras articulatórias, uma dos cantores e outra das cantoras, para conseguirem afinação e volume de voz, a fim de furarem a massa de som da orquestra – aqui, claro, ele está falando de cantores de ópera – sem realizar um esforço extenuante para as pregas vocais.

Sundberg descobre, em torno de uma frequência de 2500 a 3000 Hz, um pico extra no canto – ou seja, um pico inexistente na fala – que chama de formante do cantor (*singing formant*). É este pico extra que possibilita ao cantor fazer com que sua voz se destaque em meio à massa de som da orquestra, cujo volume maior de som se concentra abaixo 450 hertz. A manobra articulatória responsável pelo formante do cantor é o abaixamento da laringe, que, defende Sundberg (op.cit), propicia o escurecimento<sup>3</sup> das vogais, além de oferecer uma posição mais relaxada para os tecidos. Mas como é possível afirmar que o abaixamento da laringe é responsável pela origem do *singing formant*? Vejamos a explicação do próprio Sundberg:

I have estimated on the basis of X-ray pictures of a lowered larynx that this lowered larynx resonance frequency should be between 2500 e 3000 hertz, that is, between the frequencies of the normal third and fourth formants and just where the singing formant appears. The lowering of the larynx, in other words, seems to explain the singing formant peak (p.86).

As cantoras realizam uma manobra diferente para obter maior energia nos agudos. Sundberg constata que tanto mais a nota deixa uma região grave (sol 3, por exemplo), para uma aguda (fá 4), maior é abertura da boca, ou seja, mais a mandíbula abaixa. Como  $F_1$  é

---

<sup>3</sup> Escuro: é um termo utilizado entre cantores para se referir a qualidade de sons que tendem para o grave, ou timbres sonoros mais graves.

sensível ao abaixamento da mandíbula, o resultado acústico é de que, qualquer que seja a vogal, seu primeiro formante tende a parear com a frequência fundamental da nota, ou, como diz Sundberg, com a frequência de fonação. Isto acontece, em geral, a partir de um fá 4 que tem aproximadamente 700 Hz, o que vai resultar na alteração da cor da vogal cantada: por exemplo, se a vogal for um /i/ tenderá a ser ouvida como um /ε/, devido à elevação de  $F_1$ . As vogais, no canto, perdem em inteligibilidade, mas ganham em energia, pois, para ajustá-las às exigências musicais de afinação e volume, a manobra do abaixamento da mandíbula é uma solução do soprano para esforço vocal desnecessário ao cantar, como explica Sundberg:

The soprano's solution is to move the first formant up in frequency to match the frequency of the fundamental, thus allowing the formant always to enhance the amplitude of the fundamental. The result is that there is minimal variation in loudness from pitch to pitch and from vowel to vowel. Moreover, changing the size of the jaw opening in this way provides maximum loudness at the lowest possible cost in vocal effort. The strategy is probably resorted to not only by sopranos but also by other singers whose pitch range includes frequencies higher than those of the first formants of ordinary speech: contraltos, tenors and occasionally even baritones. (p. 90)

A manobra do abaixamento da mandíbula e seu efeito acústico de parear frequência fundamental com  $F_1$  da vogal é importante para esta tese, pois, desde o início, guiou-nos na busca de hipóteses válidas para uma discussão inicial quanto à questão da inteligibilidade da voz cantada. Em trabalho mais recente, Sundberg (1999) persegue a existência ou não do formante do cantor nas vozes femininas, mas acena que este seria típico apenas de vozes graves, especialmente as masculinas.

Quanto à questão da deformação das vogais, temos em Scotto di Carlo e Germain (1985) um estudo bastante ilustrativo. As autoras francesas investigam qual a influência do *pitch* sobre a inteligibilidade de vogais cantadas, utilizando 17 notas musicais em cinco regiões de frequência diferentes: grave, grave-média, média, aguda-média, aguda. Do lá 3 (220 Hz) ao dó 6 (1046 Hz), a soprano ligeiro M. Mesplé cantou 15 vogais ( [i, e, ε, a, a, y, ø, œ, u, o, ɔ, ĩ, œ̃, ã, ã ] do francês em cada uma das 17 notas musicais, para serem identificadas por quatro ouvintes foneticistas, porém não músicos. A hipótese inicial era a de que tanto a natureza da vogal como a frequência em que era cantada teriam influência na inteligibilidade. Quanto à natureza da vogal, os resultados apontaram que as labializadas são as vogais que atestam menos inteligibilidade da região média para alta; as vogais fechadas – assim as nomeia as autoras quanto a abertura da mandíbula – não resistem à região de frequência média e apresentam 59% de inteligibilidade, contra os 88% na região grave - média; as vogais abertas apresentam a percentagem mais alta na região alta: 19%, contra 1,5% e 4% das labiais e fechadas, respectivamente. Quanto à frequência fundamental em que foram entoadas as vogais, verificou-se que a inteligibilidade cai abruptamente, de modo geral, a partir da região média, passando de 38% para 18% na região alta-média. Isto quer dizer que quanto mais alta é a nota musical cantada, menos inteligível é a vogal, o que nos remete obrigatoriamente ao achado de Sundberg (1977): uma exigência articulatória do canto, o abaixamento da mandíbula, tem consequência acústica desejada que é o aumento de energia do sinal sonoro em detrimento da inteligibilidade do texto, uma vez que a vogal é deformada.

Os estudos referentes às vogais faladas aqui resumidos podem ser assim classificados: os acústico-perceptuais de Peterson & Barney (1952) e Syrdal & Gopal

(1986) e os acústico articulatórios de Lindblom e Sundberg (1971) e Johnson, Ladefoged e Lindau (1993). A pergunta subjacente a estes quatro trabalhos *Onde está a invariabilidade do som da fala: no efeito acústico ou na produção fisiológica?* parece ser inerente à natureza da busca lingüística cuja questão é explicar, a partir da aparente diversidade, uma estrutura invariável.

Por seu lado, os estudos que envolvem as vogais cantadas se perguntam *o que causa a "variabilidade"?* . Viu-se que a frequência fundamental ou frequência de fonação é em grande parte responsável pela variabilidade. Como cantar significa produzir alturas musicais diferentes e significa, ainda, na nossa cultura ocidental, sobretudo, cantar com palavras, acaba-se "variando" a fala cantada em excesso (altera-se a vogal para se manter a frequência fundamental) o que nossos ouvidos aceitam em favor da música. Há, porém, uma busca do texto, e ao nos preocuparmos com a ininteligibilidade e investigarmos a estrutura da fala cantada, estamos mapeando suas tendências ou preferências.

## **2.3 Resultados**

### **2.3.1 Padrão geral**

Assim, como no primeiro capítulo, apresentamos um padrão geral envolvendo todas as informantes. O padrão geral desse capítulo refere-se às médias dos valores formantes de cada vogal [i, e, ε, a, ɔ, o, u ], reunindo todas as informantes do experimento. Os gráficos 1 e 2, que seguem, ilustram a localização das vogais no espaço vocálico plotado a partir dos valores de  $F_1$  e  $F_2$ . Também ilustrarão as próximas seções, uma tabela em que setas para cima ou para baixo indicam a elevação ou diminuição do formante no canto em relação à fala e outra tabela apresentando os valores médios dos formantes e os resultados de significância ou não significância.

Gráfico 1: vogais faladas: F1 x F2

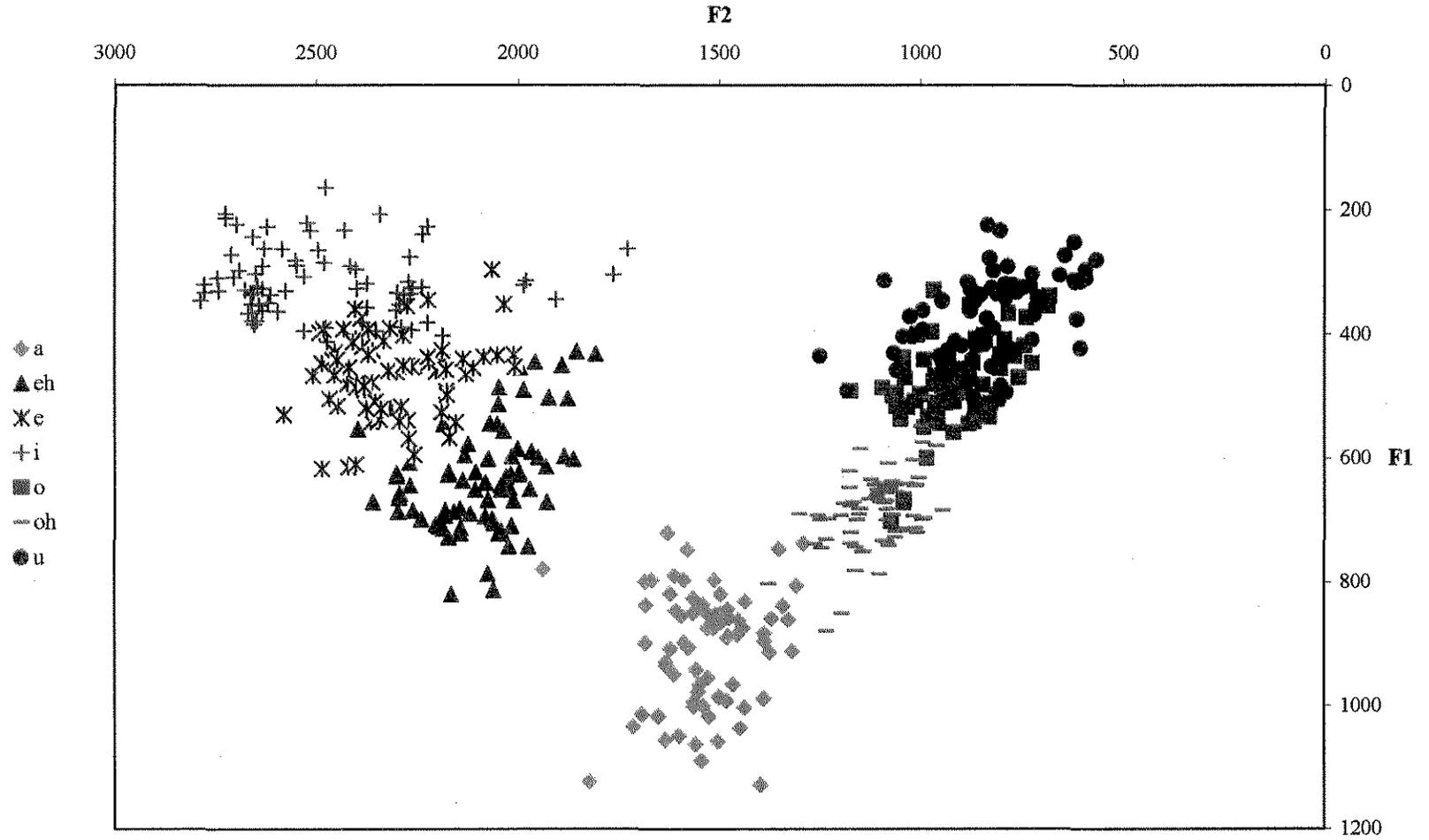
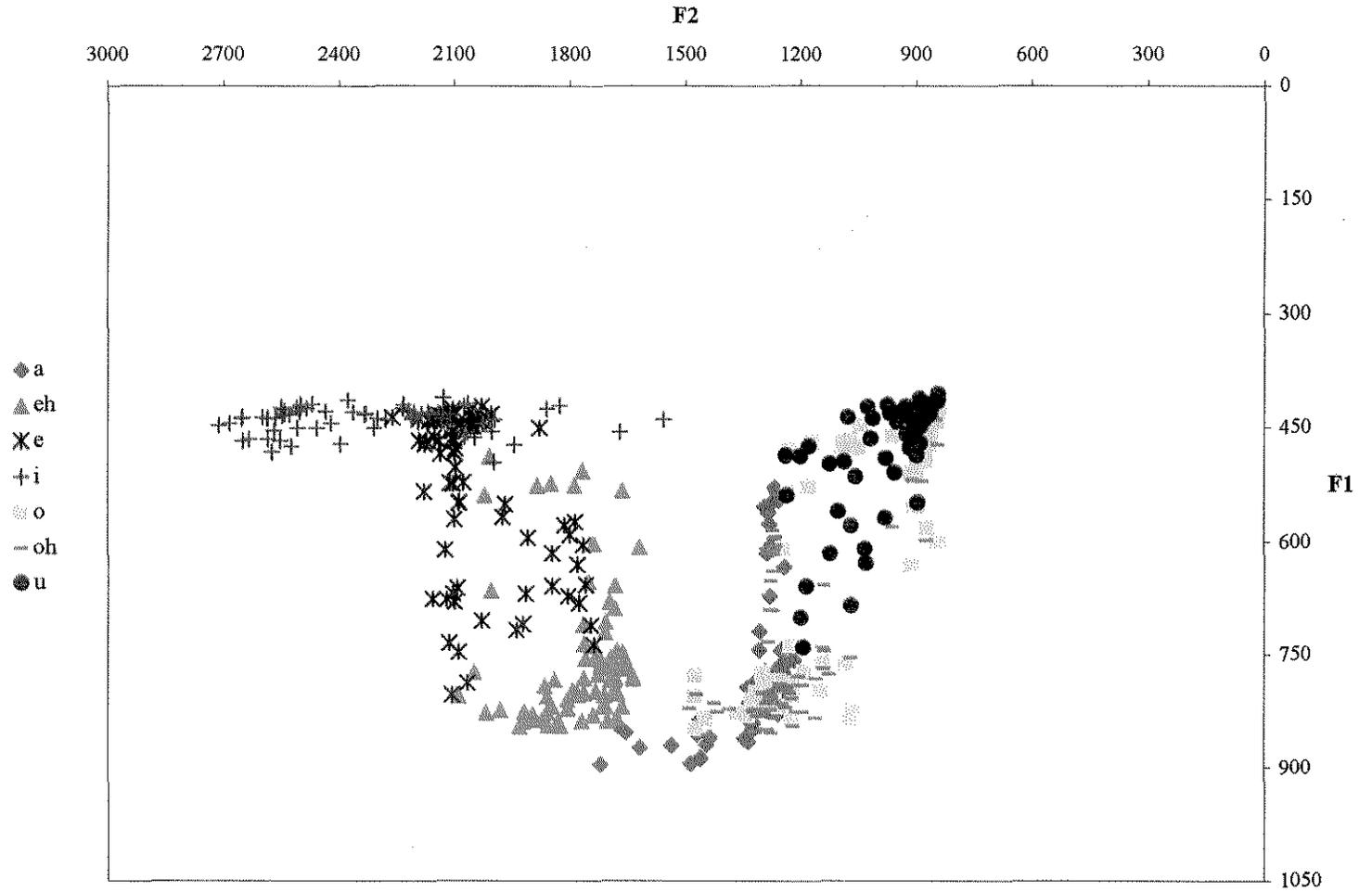


Gráfico 2: vogais cantadas: F1 x F2



Comparando os dois gráficos dos valores de  $F_1 \times F_2$ , vê-se que o gráfico das vogais cantadas redesenha o triângulo vocálico. Poderíamos dizer que as pontas do triângulo são achatadas, num movimento em que as vogais cantadas, a não ser pelo isolamento do [i], parecem querer formar um retângulo vocálico. O gráfico do canto ilustra bem, também, uma sobreposição de vogais bem maior do que a da fala.

Para explicar essa diferença de padrão formântico entre fala e canto, interpretaremos os movimentos dos dois primeiros formantes,  $F_1$  e  $F_2$ , consagrando uma seção para cada um a seguir e posteriormente discutindo, também numa seção à parte,  $F_3$ . Discutiremos os movimentos de elevação, abaixamento ou manutenção dos valores de formantes do canto em relação aos valores da fala, levando-se em conta o seguinte agrupamento das vogais: as anteriores: [i, e, ε], o [a] (central, destacado com linhas pontilhadas por apresentar um comportamento diferente das demais vogais, apresentando-se mais posterior no canto) e as posteriores: [ɔ, o, u]. Ou falando de cada uma individualmente, quando for necessário. A tabela com setas, a seguir, ilustra os movimentos dos formantes em questão.

Tabela 1: Movimentos dos formantes do canto em relação aos da fala.

vogais ⇒								
Padrão geral		[i]	[e]	[ε]	[a]	[ɔ]	[o]	[u]
↓								
PG	$F_1$	↑	↑	↑	↓	↑	↑	↑
	$F_2$	↓	↓	↓	↓	↑	↑	↑
	$F_3$	↓	=	↑	↑	↑	↑	↑

### 2.3.1.1 Movimentos de $F_1$

O  $F_1$  se eleva para todas as vogais cantadas, em comparação às faladas, exceto para a vogal [a]. Isto pode ser verificado junto à tabela 2 do padrão geral dos valores médios de formantes. A Tabela 1 ilustra os movimentos com setas.

Tabela 2: Valores médios da vogais faladas e cantadas por todas as informantes

Vogais	$F_1$			$F_2$			$F_3$		
	fala	canto	Sig	fala	Canto	Sig	fala	canto	Sig
i	314	440	s	2476	2314	s	3078	3004	s
e	465	540	s	2310	2039	s	2952	2991	ns
ε	634	755	s	2087	1779	s	2878	2984	s
a	906	767	s	1529	1317	s	2742	3018	s
o	682	752	s	1108	1234	s	2749	3009	s
o	475	614	s	918	1077	s	2864	3052	s
u	377	471	s	855	958	s	2937	3060	s

A elevação de  $F_1$  se deve ao abaixamento da mandíbula descrito na literatura (Sundberg, 1977) como uma manobra típica do canto. Ao abaixar a mandíbula, o cantor busca parear a frequência fundamental da nota musical com o primeiro formante. O pareamento completo se verifica junto a notas agudas, ou seja, de frequência fundamental bem maior do que a nota em que as vogais de nosso experimento foram entoadas. Defendemos, então, a explicação de que, numa região grave-média, o abaixamento da mandíbula é importante, mas não funciona da mesma forma que nas notas agudas. Segundo Sundberg, como vimos no início da seção, o pareamento de  $F_1$  com o  $F_0$  da nota busca maior amplitude do sinal, o que é uma exigência do canto erudito, já que não é ampliado

eletr(on)icamente, e além disso, no caso da ópera e de alguns casos do *lied* europeu, pode ser acompanhado da orquestra<sup>4</sup>. Não é esse, porém, o caso de manobras realizadas numa região grave-média, uma vez que o  $F_1$  de [ɔ] e [ɛ] se elevam em relação ao  $F_1$  da fala e em relação à frequência fundamental da nota (ver tabela 2). O que explica o abaixamento em [i,e,ɛ,ɔ,o,u] é a busca da redução de constrição no trato. No caso do [a], dá-se o inverso. O  $F_1$  do canto diminui, o que explicamos pelo fato de se tentar diminuir a constrição faríngea, típica do [a] falado, atenuando-se o abaixamento da mandíbula, no canto, para esta vogal, em particular.

Uma interpretação das manobras articulatórias e seus efeitos acústicos, encontrada em Lindblom e Sundberg (1971), permite-nos concluir, a partir de nossos dados acústicos de  $F_1$ , que o canto produz suas vogais com a mandíbula mais abaixada do que na fala, de um modo geral, alterando significativamente o primeiro formante. Quando esse abaixamento já existe na fala, caso do [a], o canto tende a amenizá-lo para criar um trato vocal com menos constrições.

Isso traz um problema para interpretarmos a manobra articulatória do pareamento descrita por Sundberg. Se o canto precisa de um trato vocal sem maiores constrições na região grave-média, porque nas notas agudas o abaixamento tão acentuado? A resposta que nos é possível neste momento é a de que existem manobras do canto específicas para cada região de frequência. No caso de nosso experimento, que investiga expressamente o canto

---

<sup>4</sup> Há canções de G. Mahler e R. Strauss que são acompanhadas pela orquestra. Como foi explicado na introdução, a canção camerística brasileira, é na grande maioria, produzida para canto e piano. Em menor número há as canções compostas com acompanhamento da orquestra, como as quatro canções que compõem a Floresta Amazônica de H.Villalobos.

numa região grave-média, a descoberta de Sundberg é mais um guia do que propriamente uma explicação para o que encontramos.

### 2.3.1.2 Movimentos de $F_2$

No canto, o segundo formante diminui em todas as vogais anteriores [i, e, ε], e na vogal [a]; e se eleva nas posteriores (ver tabelas 1 e 2, pp. 93 e 94). Essa diminuição dos valores de  $F_2$  nas anteriores explica-se pelo fato de a mandíbula estar mais abaixada, fazendo com que o corpo da língua também se abaixe, o que atenua a constrição palatal. Segundo Lindblom e Sundberg (1971), o segundo formante é “extremamente sensível” à posição do corpo da língua, e diminui consideravelmente, uma vez que o articulador em questão, além de deixar a região palatal, é deslocado em direção à faringe. Este deslocamento para a faringe, no canto, entretanto, não é uma manobra que deva acontecer.

Se por um lado é fácil explicar a diminuição de  $F_2$  nas anteriores, a elevação de  $F_2$  nas posteriores não tem uma explicação muito evidente. Estudos reportados por Sundberg (1987), como o de Shipp e Izdebski (1975), dizem que, a partir de uma frequência fundamental de 200 Hz, a laringe tende a se abaixar, alongando o trato vocal, o que se sabe, causa a diminuição dos valores dos formantes. Estudos concernentes à fala, como o da Teoria da Perturbação (Chiba e Kajiyama, 1946, apud Kent e Read, 1992), demonstram que a protrusão dos lábios – o que é típico de vogais como [o] e [u] – faz com que os valores formantes diminuam. Nossos resultados indicam que os formantes das vogais posteriores, tanto  $F_1$ , como  $F_2$  como  $F_3$ , se elevam.

Estamos diante de um quebra-cabeça. Para poder montar esse quebra-cabeça vamos retomar alguns pontos do parágrafo anterior:

*Posição da laringe:* embora Sundberg refira-se ao abaixamento da laringe como uma manobra típica do canto erudito, não parece claro que seja uma manobra obrigatória. As nossas informantes parecem levantar a laringe, senão para cantar todas as vogais, pelo menos para cantar as vogais posteriores. O próprio Sundberg, juntamente com Nordström (1983, apud Sundberg, 1987) desenvolveu um experimento onde sujeitos cantavam com a laringe ora levantada, ora abaixada e os resultados mostraram que, naquela primeira posição, havia elevação dos valores dos formantes, enquanto, na segunda, havia diminuição de valores formantes. Também em Sundberg (1987) encontramos o seguinte a respeito da posição da laringe e sua relação com vozes agudas:

It also supports the assumption that tenors have shorter pharynxes than basses and that a shift of voice category from baritone to tenor involves, among other things, a habitually higher larynx position. (p.115)

*Protrusão labial:* Outra explicação, menos provável, é de que as cantoras desse experimento não realizem a protrusão labial ao cantar as vogais [o] e [u]. Essa explicação foi descartada, depois de alguns registros de imagem feitos<sup>5</sup> com a informante MG. A gravação em vídeo mostra que a informante realiza [ɔ,o,u] com arredondamento e protrusão. Essa informação visual mantém nossa hipótese de que, para produzir as vogais arredondadas, as cantoras levantam a laringe, daí todos os formantes destas vogais se elevarem. Por outro lado, para o [a], cujo abaixamento da mandíbula não é tão acentuado,

---

<sup>5</sup> Esse registro não faz parte do *design experimental* inicial e foi realizado para que pudéssemos ter um outro ponto de partida para inferir os movimentos articulatorios do canto, tão difíceis de serem interpretados apenas à luz da Teoria Acústica da Fala.

aventamos a hipótese de um abaixamento laríngeo compensatório, o que faz com que  $F_1$  diminua. Abaixo, ilustramos, com fotogramas obtidos das imagens filmadas, o que discutimos a respeito do arredondamento dos lábios para vogais como [u]. Há também fotogramas da vogal [a] sobre a qual destacamos o menor abaixamento da mandíbula no canto, se comparado a outras vogais cantadas; e de [i] cujo abaixamento da mandíbula, é relativamente acentuado devido à natureza mesma da vogal, e conta com o abaixamento do lábio inferior.



/la'tu/ falado



/la'tu/ cantado



/la'pa/ falado



/la'pa/ cantado



/la'pi/ falado



/la'pi/ cantado

### 2.3.1.3 Movimentos de F<sub>3</sub>

O F<sub>3</sub> se eleva em [ɛ, a, ɔ, o, u] , diminui em [i] e permanece igual em [e]. Considerando que o canto busca atenuar as constrictões do trato vocal encontradas na fala, a elevação do F<sub>3</sub> pode ser explicada pelo abaixamento da mandíbula. Segundo Lindblom e Sundberg (op.cit), o abaixamento da mandíbula têm como efeito a elevação de F<sub>3</sub>. Os autores fizeram medidas em milímetros dos movimentos da mandíbula (de 5 a 25 mm) Quanto maior fosse a abertura da boca, 25 mm, segundo os gráficos plotados – ou seja, mandíbula mais abaixada - mais o F<sub>3</sub> das três vogais [i, a, u] se elevava, na condição arredondada, assim chamada pelos autores referindo-se à forma dos lábios. Esta condição foi comparada, no experimento em questão, com a de lábios estirados (spread).

O [i] do canto apresenta uma diminuição do valor de F<sub>3</sub>, comparado ao [i] da fala. O [i] é a vogal de maior constrictão palatal, na fala, o que explica elevação do segundo formante, que, sabe-se, distancia-se pouco do terceiro formante, que por sua vez, segundo Fant (1970), é também alto quando a região articulatória é prepalatal ou dental. Assim, no canto, é plausível dizer que mais do que o abaixamento da mandíbula – que não pode ser exagerado nesta vogal para não deformá-la – a acomodação da língua numa posição de

descanso para produzir o [i] cantado atenua a constrição prepalatal de que nos fala Fant (1970), abaixando assim o  $F_3$ . Abaixar o  $F_3$  pode ser um recurso para manter a qualidade do [i], diferenciando-o das duas outras vogais altas. Lembremos que  $F_2$  desta vogal é também abaixado no canto, o que possibilita trazer para esta modalidade o mesmo efeito de centro de gravidade (Carlson et al. apud Stevens 2000):

When two formants among  $F_2$ ,  $F_3$  and  $F_4$  are relatively close together (such as  $F_2$  and  $F_3$  for the left spectrum and  $F_3$  and  $F_4$  for the right spectrum) the corresponding peaks are of high amplitude and dominate the spectrum in this frequency region. There is evidence that the perceived vowel quality is determined more by the “center of gravity” of this multi-peaked high-frequency spectrum than by the frequency of one of the formants (Carlson et al., 1970)

O [e] mantém, no canto, o  $F_3$  da fala (2952 Hz na fala e 2991 Hz no canto), mesmo tendo abaixado o  $F_2$ . Talvez seja o que o diferencia de [i] e [ε], pois a distância entre  $F_2$ - $F_3$  de [e] cantado é maior que a de [i] e menor que a de [ε] (ver tabela 2). Com o conceito de vogais de altura intermediária, (Stevens, 2000) talvez possamos ter uma idéia do que acontece com as vogais altas no canto:

Because the constriction is not as narrow as for the high vowel, there is more acoustic coupling between the narrow and wide portions of the tube, and consequently  $F_2$  does not approach as close to  $F_3$  (nor  $F_3$  to  $F_4$ ) and thus the maximum  $F_2$  is not as high as it is for a high vowel. Nevertheless,  $F_2$  and  $F_3$  are close enough that their effects are combined auditorily, and there is an effective high-frequency center of gravity for the spectral prominence formed by these two formants. The frequency of this center of gravity is probably intermediate between  $F_2$  and  $F_3$ . Whereas the tongue blade is raised to form the

narrow constriction for the high front vowel /i/, it is less involved in forming the configuration for /e/.

#### 2.3.1.4 A distância $F_2 - F_1$

A distância entre  $F_1$  e  $F_2$  do canto comparada à distância dos mesmos formantes da fala é menor para as vogais anteriores e para o [a]; é um pouco maior para a vogal [ɔ] e igual para [o] e [u], como se pode ver na tabela abaixo:

Tabela 3: Valores da diferença entre  $F_1$  e  $F_2$  em Hertz

Vogais	Distância $F_2 - F_1$		
	Fala	Canto	P-value
i	2148	1876	0.0001
e	1844	1499	0.0001
ɛ	1453	1024	0.0001
a	624	551	0.0001
ɔ	426	488	0.0001
o	442	468	0.0846
u	482	487	0.786

A tendência do canto de diminuir as distâncias entre  $F_1$  e  $F_2$  para as vogais que podemos chamar de não arredondadas [i,e,ɛ,a], sobretudo no que diz respeito às anteriores, em que a distância  $F_1 - F_2$  é bastante grande na fala, indica manobras que buscam um formato *schwalike*, assegurando um tubo mais uniforme para o canto.

Para a vogal [ɔ], há um aumento de distância entre  $F_1 - F_2$  no canto, em comparação à fala, o que é plausível ser interpretado à luz do conceito de vogal de altura intermediária, em que os formantes do lado esquerdo do espectro não estão próximos o bastante para

reforçarem o centro de gravidade, o que ajuda a diferenciá-lo de seus pares, as arredondadas [o] e [u]. Estas duas vogais, por sua vez, elevam  $F_1$  e  $F_2$  no canto, de tal sorte que mantêm no canto a mesma distância que apresentam na fala.

### 2.3.2 Explicação mais consistente para o padrão formântico do canto

A partir das pistas acústicas, procuramos descobrir o que estava acontecendo articulatoriamente no canto. Entretanto, nossas explicações poderiam perder em consistência se insistíssemos apenas em revelar as manobras articulatórias mais prováveis da modalidade em questão.

Descrever o que acontecia no trato vocal a partir da região supraglotal até os lábios não era suficiente para explicar as vogais cantadas e seu padrão formântico. Passamos a levar em consideração uma tendência das vogais do canto de se sobreporem por estarem buscando um ponto de atração, sobretudo para o segundo formante. No entanto, havia, mesmo dentro do retângulo no padrão geral, uma dispersão, principalmente de [e, ε, o] que nos impedia de postular uma explicação nesses termos. Para visualizar melhor a concentração dos valores dos formantes que pareciam estar em torno de um mesmo ponto, transformamos os formantes em hertz para *bark* (ver Gráfico 3, próxima página).

Plotando, então, um gráfico com os formantes do canto, em *barks*, vimos que há um ponto de atração para  $F_1$  e  $F_2$ , dependendo da vogal. Por exemplo, no caso de  $F_2$ : olhando o gráfico 3, na próxima página, vemos que [i, e, ε] tendem a se agrupar em torno de 13,41 *barks* (2100 Hz), [a, ɔ, o] tendem a se agrupar em torno de 10,02 *barks* (1260Hz), e que [o,u] tendem a se agrupar em torno de 7,43 *barks* (840 Hz) (ver o Gráfico 4, p. 104).

Gráfico 3: vogais do canto: valores em Bark

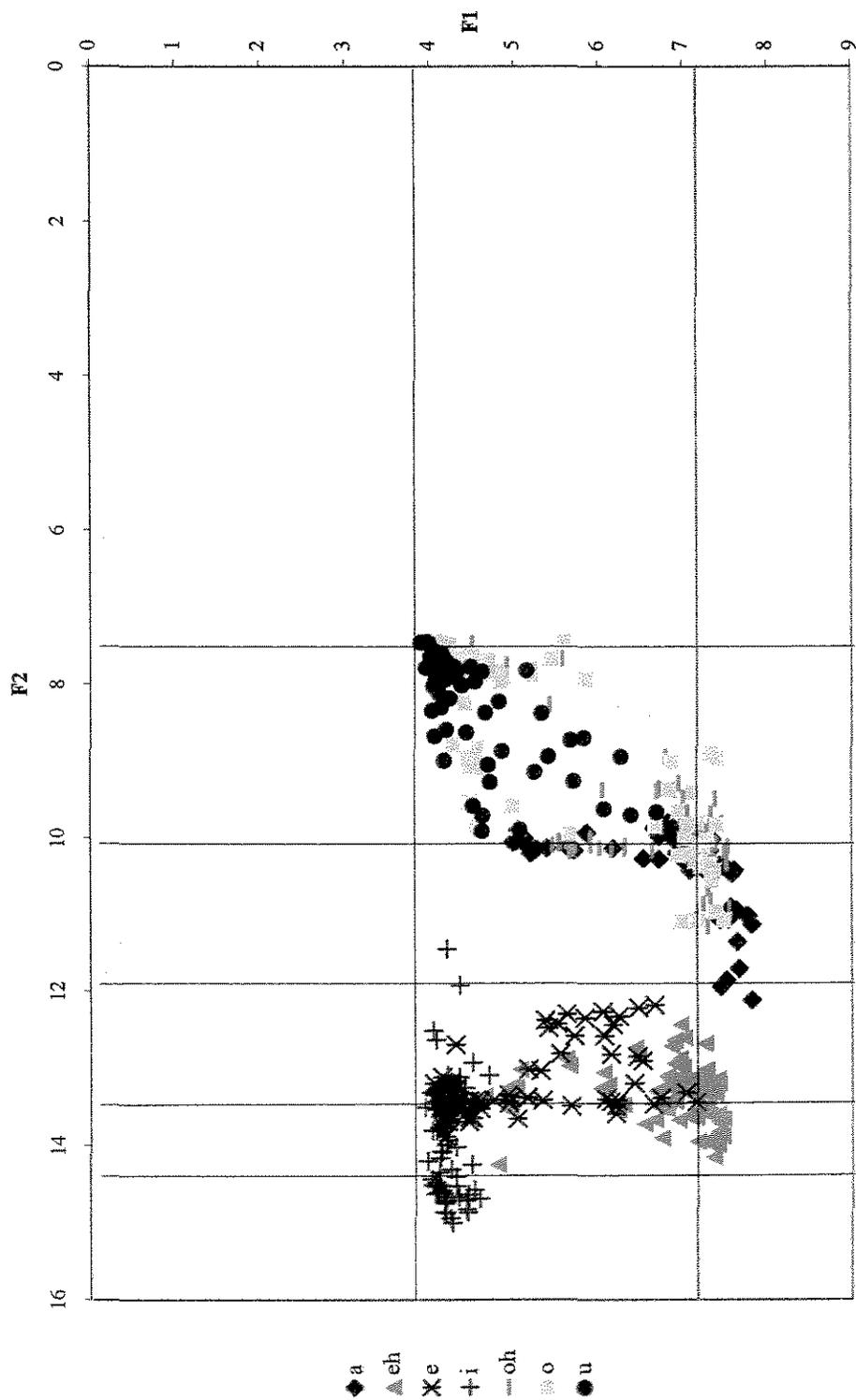
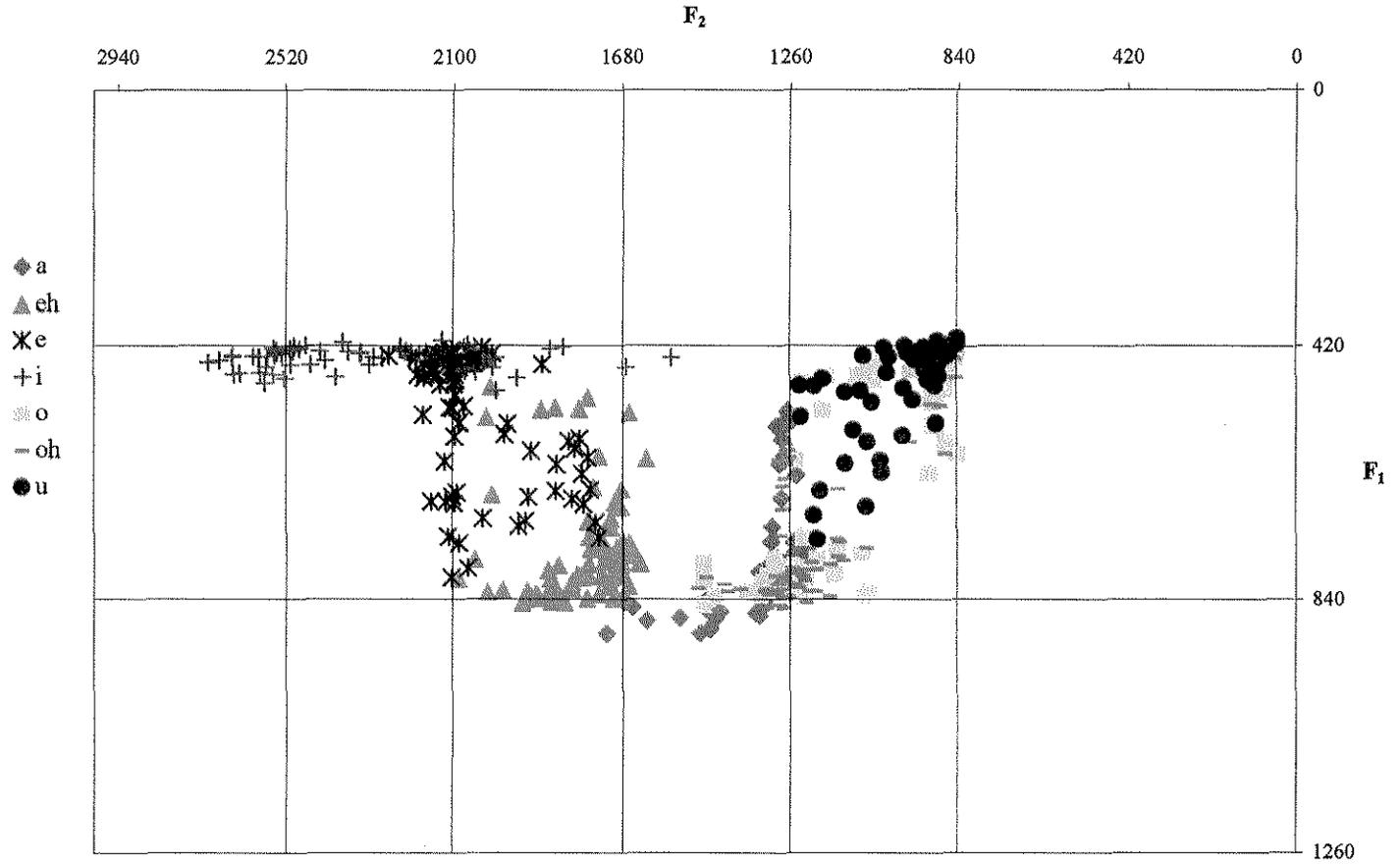


Gráfico 4: vogais cantadas e seus harmônicos:  $F_1 \times F_2$



Os valores de  $F_2$  do canto correspondem a harmônicos da nota sobre a qual todas as vogais-alvo foram entoadas. Vejamos, se tomarmos 420 Hz como um valor em torno do qual as informantes afinam o lá bemol, temos em 840 Hz o segundo harmônico da nota, ou seja, uma oitava acima. O terceiro harmônico é 1260 Hz, o quarto, 1680 Hz, o quinto 2100 Hz e o sexto, 2520 Hz. Ainda temos o sétimo harmônico em 2940 Hz, conforme aparece no gráfico 4, e embora isto não seja visualizado neste gráfico, veremos, no capítulo 3 que há pareamento de  $H_7$  com o terceiro formante.

Cabe, agora, uma explicação a respeito da frequência fundamental das vogais deste experimento: as informantes não afinavam o lá bemol precisamente em 415 Hz, levando-se em conta dois fatores importantes: primeiro, a tendência de algumas cantoras de afinar para cima, chegando a 426 Hz, e de outras de afinar para baixo 413 Hz; segundo, a realização do vibrato. Em Wood, (1950) temos, sobre o vibrato o seguinte:

It is a well-known fact that many singers do not attempt to hold a sustained note at constant pitch, but produce a periodic variation of the pitch known as vibrato. A good vibrato is said to consist of 'a pulsation of pitch usually accompanied by synchronous pulsations of loudness and timbre of such extent and rate as to give a pleasing flexibility, tenderness, and richness to the tone'. (p.56).

Segundo Wood, a taxa de variação é de 6,5 ciclos por segundo e a faixa de variação é de aproximadamente meio-tom. Outra definição de vibrato é a Schultz-Coulon e Battmer (1981) apud Sundberg (1987):

Physically, the vibrato corresponds to a periodic, rather sinusoidal modulation of the phonation frequency. (p.163)

Finalmente, sobre a taxa e faixa de variação, Sundberg (op.cit), em congruência com Wood (op.cit), diz o seguinte:

As the vibrato corresponds to a regular undulation of phonation frequency, it is characterized by two parameters: the rate and the extent of the undulations. (...) Generally a vibrato of less than 5.5 undulations per second sounds unacceptably slow, and vibrato rates exceeding 7.5 undulations per second tend to sound nervous. (pp. 163 e 164).

Dada, então, a esta variação de afinação, produzida, seja pela afinação de cada cantora, seja pelo vibrato, consideramos 420 Hz um valor plausível para representar a frequência fundamental em que foram entoadas as vogais.

Empregamos o termo harmônico tal como ele é utilizado na música, uma vez que a frequência fundamental da vogal é uma nota musical. O que vemos, assim, no gráfico 4 é a correspondência entre formantes e harmônicos, ou seja, formantes – fenômeno acústico da fala - ajustando-se em torno de múltiplos da frequência fundamental, os harmônicos – fenômeno acústico do canto.

O gráfico em *bark* (Gráfico 3) ilustra melhor o ponto de atração dos formantes, como já expusemos acima, mas parece-nos necessário ilustrar o fenômeno descoberto com um gráfico em que os valores em *hertz* correspondam à frequência fundamental (doravante  $H_1$ ) e a seus harmônicos. Embora diante da dispersão de algumas vogais, devido, sobretudo, a diferenças individuais entre as informantes, não nos resta mais dúvida de que se trata de um fenômeno de atração: formantes das vogais atraídos pelos harmônicos da nota, ou simplesmente, pareando-se com tais harmônicos.

A esse fenômeno<sup>6</sup>, então, damos o nome de pareamento. Usamos a palavra pareamento com base no uso de *match*, termo utilizado em Sundberg (1977), conforme vimos na seção 2, para explicar que o  $F_1$  da vogal cantada busca ajustar-se ao *pitch* ou frequência fundamental da nota. Há duas importantes diferenças entre o pareamento descoberto por Sundberg e o pareamento de que ora tratamos:

- a) O formante atraído na manobra articulatória do soprano operístico, informante de Sundberg, é o primeiro formante, uma vez que a o enfoque é sobre vogais entoadas em notas agudas. Já, no nosso experimento, propositalmente realizado com notas grave-médias, **o formante nitidamente atraído, é o segundo.**
- b) O ponto de atração em Sundberg (op. cit) é a frequência fundamental da nota. Quanto mais aguda a nota, mais a cantora abaixa a mandíbula, provocando a elevação de  $F_1$ . Nossa descoberta diz que há **mais de um ponto de atração**: além de  $F_1$  para vogais como [i, u] (ver o gráfico 4) , temos os harmônicos da nota, ou seja, múltiplos de 420 Hz que atraem  $F_2$  e também  $F_1$  no caso de [ε,a,ɔ,o].

As diferenças entre os nossos achados e os da literatura não indicam conflito, mas complementaridade, dado que a intenção deste trabalho de tese é investigar as vogais cantadas numa região grave-média, ao passo que Sundberg volta sua atenção para uma manobra realizada com notas agudas.

Conseqüência do pareamento entre formantes e harmônicos é a sobreposição de algumas vogais, que passaremos a chamar de **sobreposição desejável na modalidade**

---

<sup>6</sup> Em Carlsson e Sundberg (1992), fala-se de “afinação” (tunning) dos dois primeiros formantes com as parciais (harmônicos) da nota musical; mas a conclusão a que se chega é de que não se trata de um fenômeno extensivo a todos os cantores.

**cantada.** As vogais que mais se sobrepõem umas às outras no canto são [a,ɔ,o], o que é esperado, pois [a] tem seu  $F_1$  e  $F_2$  consistentemente diminuídos, enquanto os dois primeiros formantes de [ɔ,o] se elevam. Para cada vogal há uma manobra articulatória efetuada pelas cantoras para moldar o trato vocal à modalidade cantada – conforme se discutiu nas seções anteriores -, mas o resultado acústico tende a ser o mesmo, ou seja, o ajuste dos formantes em torno da frequência fundamental e de seus harmônicos.

Algumas realizações de [o] tendem a se sobrepor a [u], o que acontece também na fala. O mesmo pode se dizer de [i, e]. Por sua vez, o [ɛ] cantado parece ser a vogal que apresenta o comportamento mais diferenciado, do ponto de vista que embora eleve bastante seu  $F_1$ , não se sobrepõe a [a], sendo sua tendência sobrepor-se a [e] em algumas realizações.

Obviamente, os resultados acústicos indicam que se trata de uma restrição musical impondo-se a restrições da fala. Sem dúvida, a cor da vogal sofre alterações e isso vai ter conseqüências para a inteligibilidade do texto. A manobra articulatória responsável pelas diferenças encontradas entre canto e fala será discutida ao final deste capítulo.

### 2.3.3 Padrão por informante

A seguir, apresentamos os movimentos das vogais de cada informante, ilustrando-os com gráficos de  $F_1$  x  $F_2$  da fala e do canto, este com seus harmônicos; tabela de valores médios e tabela com setas. Também explicamos como cada informante realiza o pareamento dos formantes da vogal com os harmônicos da nota musical.

### 2.3.3.1 Informante CM

O padrão formântico da fala desta informante é bastante diferente do seu padrão do canto. É a única informante que apresenta, sem exceção, significância para todos os valores na comparação do primeiro e segundo formantes (ver tabela 5, p. 115). Há pouca sobreposição de vogais na fala (ver gráficos 5 e 6, nas páginas seguintes) e no canto pode-se dizer que há a sobreposição mínima possível.

Gráfico 5: informante CM, vogais faladas

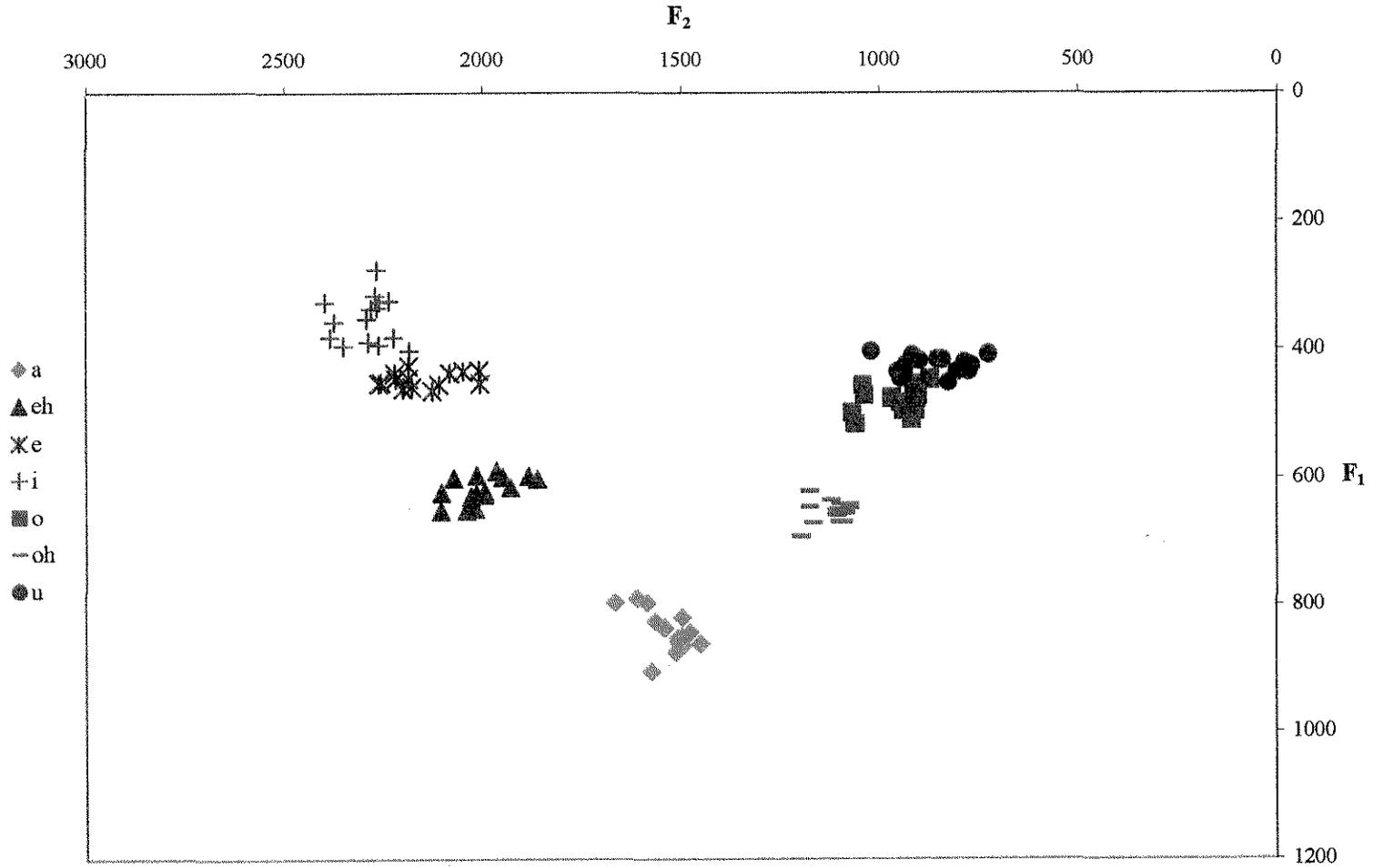
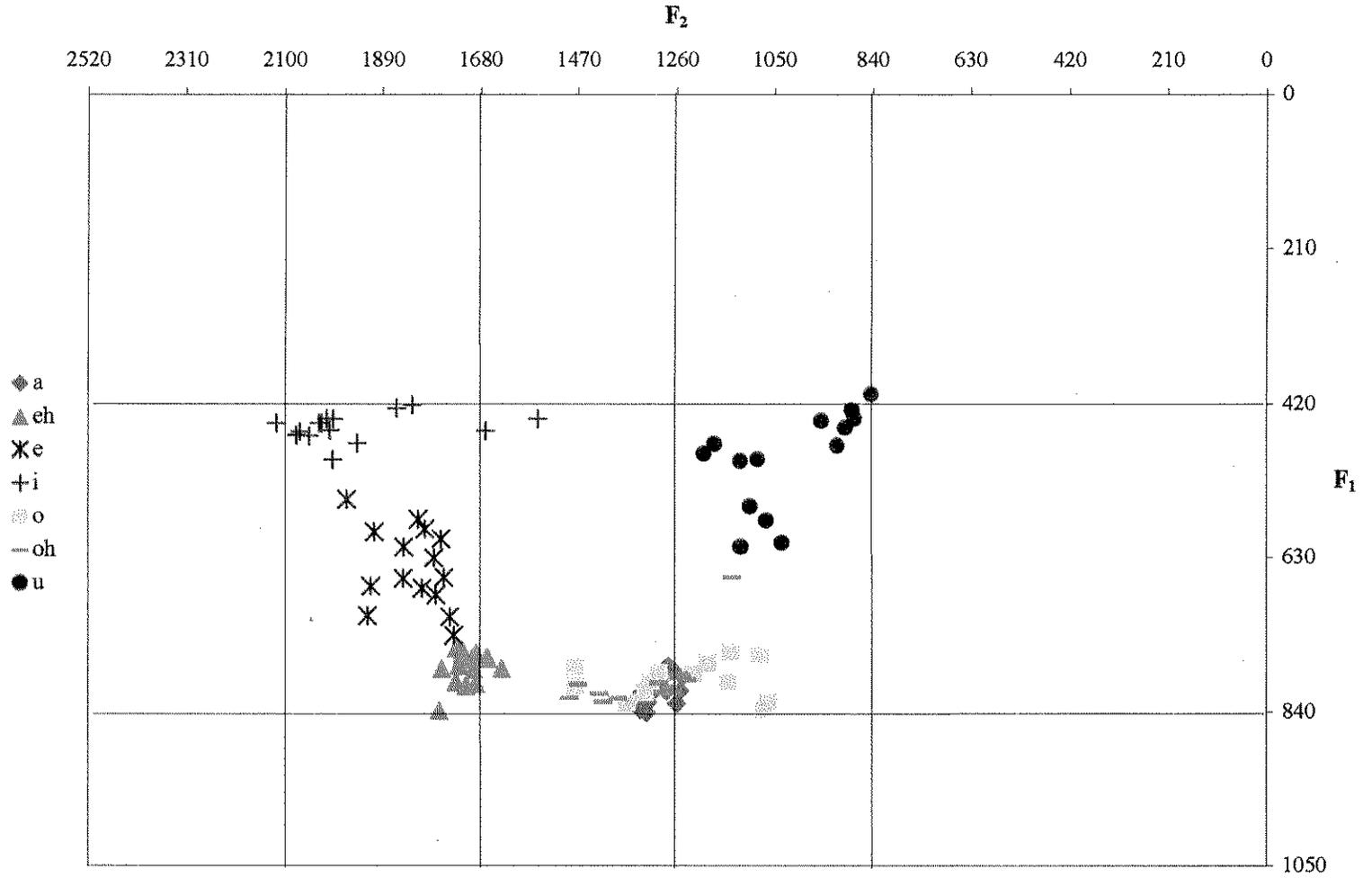


Gráfico 6: informante CM, vogais cantadas



Na fala [o e u] se sobrepõem; já, no canto, [o] migra para se sobrepor a [ɔ e a]. No canto, [u] tende a fazer o mesmo movimento que [o], mas mantém-se na parte superior do retângulo.

Quanto às anteriores: as realizações de [ɛ] mantêm-se, como na fala, próximas umas das outras, ao passo, que [e] se espraia, tendendo ora mais para [i], ora mais para [ɛ]. A vogal [i], em algumas realizações, tende a diminuir o  $F_2$ , mantendo o  $F_1$  em torno de 450 Hz, o que parece ser uma explicação para sua diferença em relação às duas outras vogais anteriores.

Passamos agora a explicar como a informante CM realiza o pareamento, ou seja o fenômeno em que há pontos para os quais os dois primeiros formantes são atraídos: o primeiro harmônico e os harmônicos subseqüentes. Tomando o eixo de  $F_2$ , temos a seguinte localização das vogais quanto aos sucessivos harmônicos da frequência fundamental:

O segundo harmônico, 840 Hz, portanto um harmônico acima de 420 Hz atrai [u], que se espraia em direção ao terceiro harmônico.

O terceiro harmônico, 1260 Hz, atrai [a, ɔ, o], sendo que esta última vogal tende a se espraiair, enquanto as outras duas são mais fiéis ao ponto de atração, 1260 Hz.

O quarto harmônico, 1680 Hz atrai [ɛ, e]; e o quinto, 2100 Hz atrai [i].

Olhando para o eixo de  $F_1$ , vemos que as vogais parecem se enquadrar nos limites entre 420 e 840 Hz, dada a tendência, no canto, às vogais terem seus primeiros formantes elevados, exceto no caso de [a]. Há, então, o pareamento de  $F_1$  com o primeiro harmônico nas vogais [u, i]; enquanto, [ɛ, a, ɔ, o] tendem a parear o primeiro formante com o segundo harmônico.

Para esta informante, em especial, [e] não segue a tendência do pareamento.

Tabela 4. Movimento de formantes do canto em relação à fala: informante CM.

vogais ⇒							
PI ↓	i	e	ε	a	ɔ	o	u
F1	↑	↑	↑	↓	↑	↑	↑
F2	↓	↓	↓	↓	↑	↑	↑
F3	=	=	=	↑	↑	↑	↑

Tabela 5. Padrão formântico da fala comparado ao do canto, informante CM

Vogal	Padrão formântico fala x canto: informante CM								
	F <sub>1</sub>			F <sub>2</sub>			F <sub>3</sub>		
	Fala	Canto	Sig	Fala	Canto	Sig	Fala	Canto	Sig
i	352	452	s	2290	2008	s	2894	2874	ns
e	445	644	s	2150	1825	s	2883	2859	ns
ε	619	780	s	1999	1707	s	2755	2862	ns
a	841	810	s	1534	1293	s	2498	2839	s
ɔ	654	796	s	1117	1314	s	2303	2852	s
o	479	798	s	953	1252	s	2371	2888	s
u	426	495	s	864	1024	s	2550	2920	s

### 2.3.3.2 Informante CH

Esta informante apresenta muita sobreposição, tanto na fala como no canto, com tendências semelhantes: por exemplo, [ε, e], sobrepõem-se na fala e no canto, assim também ocorrendo com [o, ɔ]. Na fala [a] se isola, apresentando o que consideramos pouca sobreposição no canto, já que, nessa modalidade, tal vogal costuma mesmo até se confundir com [ɔ].

Tabela 6. Movimento de formantes do canto em relação à fala: informante CH.

Vogais	i	e	ε	a	ɔ	o	u
F <sub>1</sub>	↑	↑	↑	↓	↑	↑	↑
F <sub>2</sub>	=	↓	↓	↓	↑	↑	=
F <sub>3</sub>	↓	=	=	=	↑	=	=

Tabela 7. Padrão formântico da fala comparado ao do canto, informante CH.

Vogal	Padrão formântico fala x canto: informante CH								
	F <sub>1</sub>			F <sub>2</sub>			F <sub>3</sub>		
	Fala	Canto	Sig	Fala	Canto	Sig	Fala	Canto	Sig
i	328	457	s	2545	2554	ns	3231	3101	s
e	474	667	s	2394	2058	s	3073	3050	ns
ε	684	805	s	2151	1865	s	3039	3094	ns
a	956	868	s	1641	1495	s	2834	2993	ns
ɔ	701	823	s	1118	1303	s	2714	2974	s
o	510	813	s	972	1304	s	2926	2966	ns
u	386	562	s	935	1053	ns	2992	3029	ns

Quanto ao pareamento realizado pela informante CH, temos o seguinte: as vogais que tendem a parear o F<sub>2</sub> com as harmônicos subseqüentes a 420 Hz são [ɔ, o], em torno de 1260 Hz e [e], em torno de 2100 Hz. O [i] fica em torno de 2520 Hz. Outras duas vogais, porém, não são bons exemplos de pareamento: [u] isola-se, espreado-se entre 840 a 1260 Hz. O [ε] também espalha-se, só que de 1680 a 2100 Hz. O pareamento de F<sub>1</sub> com o primeiro e segundo harmônicos não está muito bem delineado, embora [a] e [i] correspondam bem aos valores em torno de 450 e 868 Hz, valores mais próximos da nota entoada por essa informante, em particular, que ajustava seu lá bemol para cima.

Gráfico 7: informante CH, vogais faladas

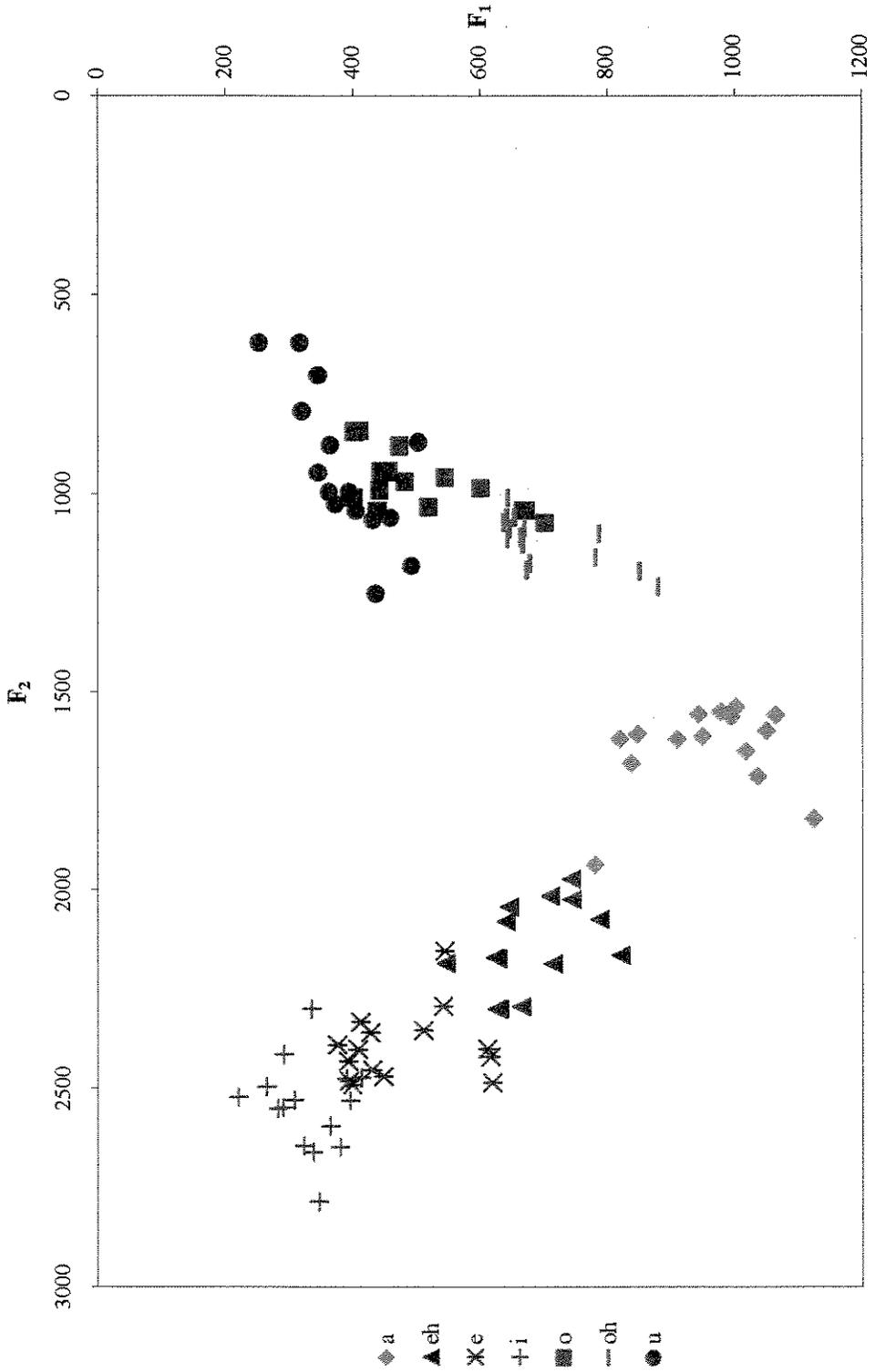
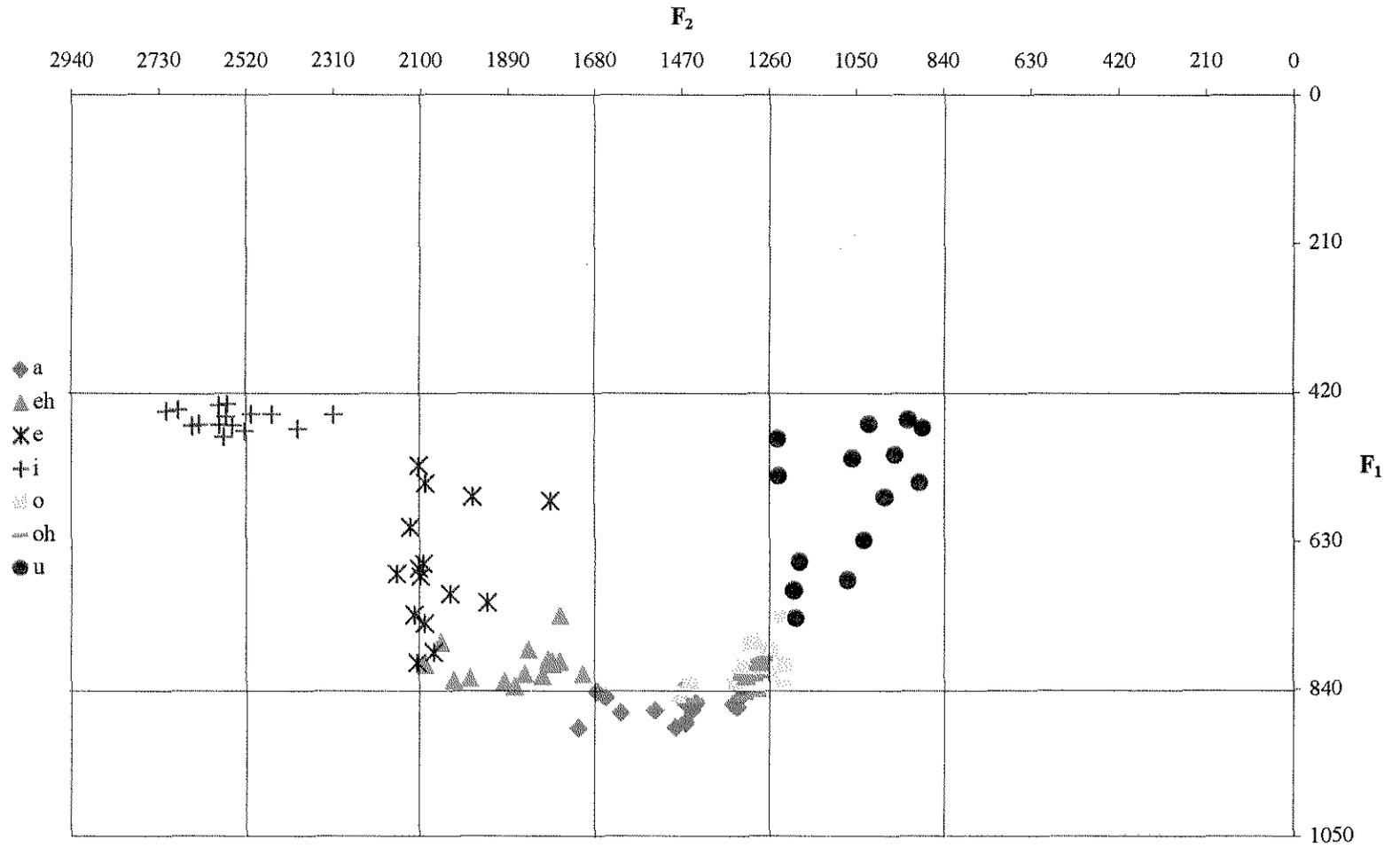


Gráfico 8: informante CH, vogais cantadas



As informantes CM e CH apresentam os gráficos menos ilustrativos do ponto de atração que queremos apontar como o responsável pelo padrão formântico do canto, mas teremos, nas próximas três informantes, bons exemplos do fenômeno descoberto. Não podemos, no entanto, deixar de levantar a questão das diferenças individuais, e alertar para o fato de que as manobras do canto que tentamos elucidar neste trabalho de tese não são obrigatórias, mas constituem tendências muito fortes. A dúvida em torno do caráter não obrigatório da realização das manobras – sobre a qual outros estudos se fazem necessários - é se (i) ela está relacionada à morfologia do trato de cada pessoa; ou (ii) ao estilo que a cantora desenvolveu em sua formação.

### **2.3.3.3 Informante MG**

A informante MG, como se pode ver pelo Gráfico 10 (p. 121), fornece-nos um bom exemplo do ponto de atração dos formantes, que acaba resultando no pareamento formante/ harmônicos da nota musical.

Há sobreposição na fala, que acontece mais fortemente no canto, entre os pares [o, u] e [e, i].

Gráfico 9: informante MG, vogais faladas

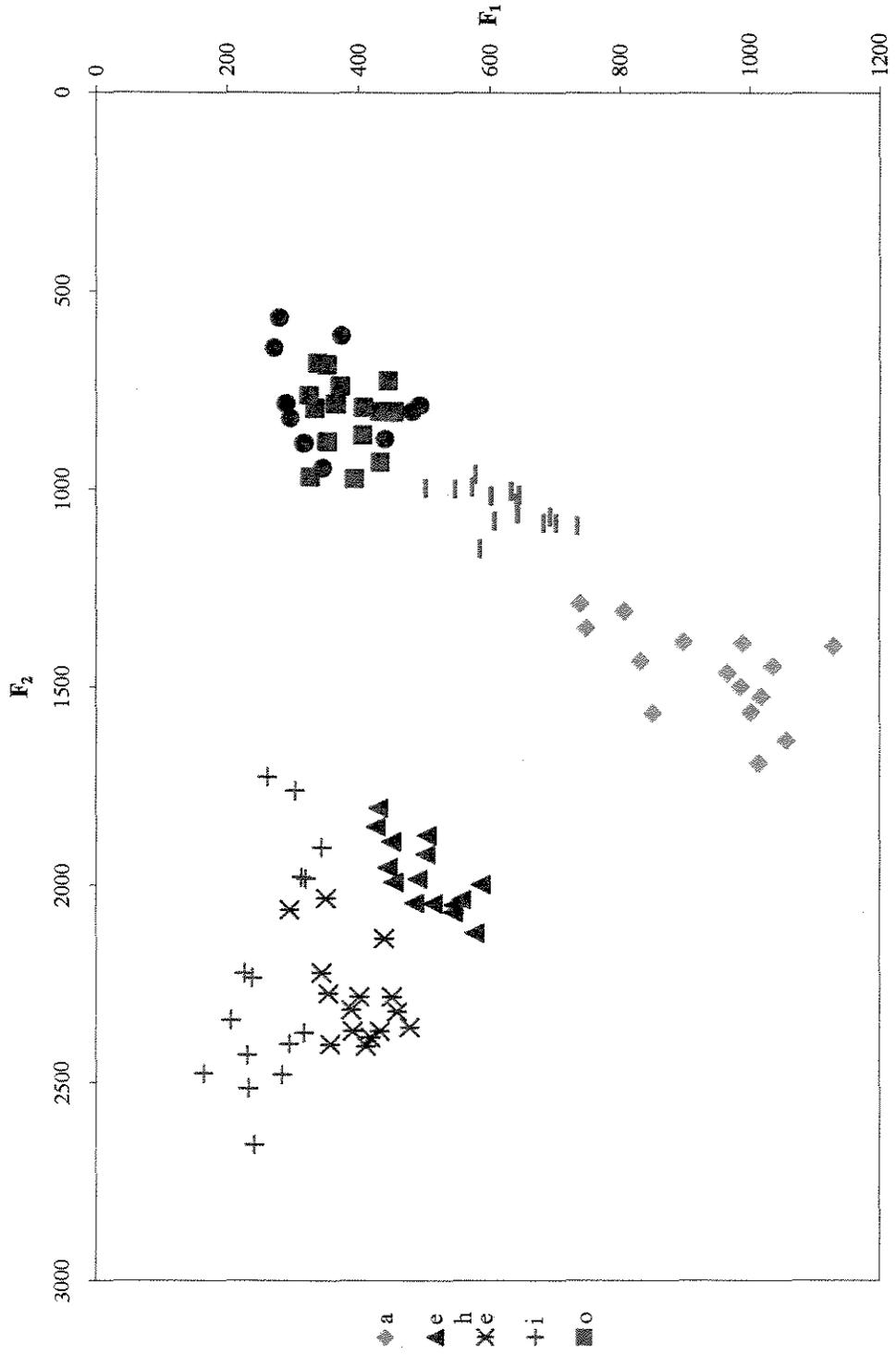


Gráfico 10: informante MG, vogais cantadas

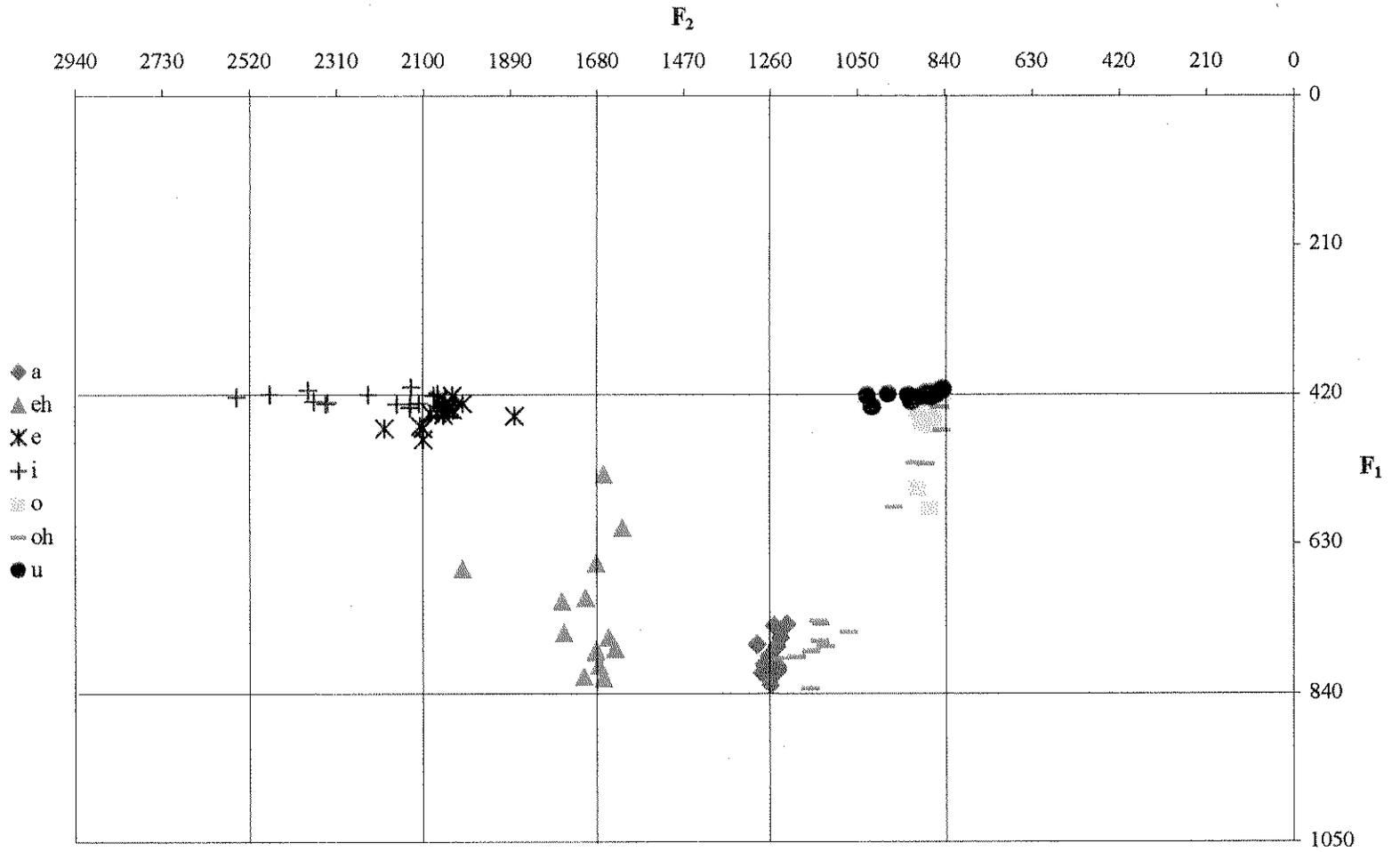


Tabela 8. Movimentos de formantes do canto em relação à fala: informante MG

Vogais	i	e	ε	a	ɔ	o	u
F1	↑	↑	↑	↓	=	↑	=
F2	=	↓	↓	↓	=	↑	↑
F3	↑	↑	↑	↑	↑	↑	=

Tabela 9. Padrão formântico da fala comparado ao do canto, informante MG.

Vogal	Padrão formântico fala x canto: informante MG								
	F <sub>1</sub>			F <sub>2</sub>			F <sub>3</sub>		
	Fala	Canto	Sig	Fala	Canto	Sig	Fala	Canto	Sig
i	266	426	s	2233	2234	ns	2819	2973	s
e	398	448	s	2282	2055	s	2831	2952	s
ε	500	733	s	1976	1705	s	2733	2934	s
a	937	791	s	1461	1255	s	2684	2976	s
ɔ	624	689	ns	1042	1059	ns	2824	2983	s
o	383	461	s	811	868	s	2908	3067	s
u	387	423	Ns	771	906	s	2947	3017	ns

O pareamento de F<sub>2</sub> com o segundo harmônico, 840 Hz, está claro em [u, o, ɔ]. Já, com as vogais [a, ɔ], F<sub>2</sub> atrai-se pelo terceiro harmônico da nota, ou seja, 1260 Hz. Na seqüência, temos o [ε], pareando seu F<sub>2</sub> em 1680 Hz. E por fim o [e, i], atraindo-se pelo quinto harmônico, ou 2100 Hz.

No caso desta informante, não é [e], mas sim [ε] que tende a se espriar, traindo o ponto de atração. Ainda assim, a maioria de [ε] realizados fica em torno de 1680 Hz.

Já o pareamento de  $F_1$ , como parece começar a ficar patente, é melhor realizado nas vogais cardeais, [i, a, u]. As vogais [ɛ, ɔ, o] ficam indecisas no eixo de  $F_1$ , ora aproximando-se de 420 Hz, ora de 840 Hz, sendo [o], entre essas três vogais, mais fiel ao pareamento com o segundo harmônico.

#### 2.3.3.4 Informante SS

As vogais faladas desta informante ocupam lugares bem demarcados no triângulo vocálico, e é possível dizer que não há nenhuma sobreposição. No canto, mantém-se essa tendência, a não ser pela sobreposição desejável. Sobrepõem-se [o] e [u]; [ɔ] e [a].

O pareamento se dá, claramente, da seguinte maneira:

O  $F_2$  de [u,o] são atraídos pelo primeiro harmônico, 840 Hz; de [a, ɔ], pelo harmônico seguinte. A vogal [ɛ], desloca-se um pouco do esperado, em direção a 1890 Hz. Por sua vez, [e] atrai-se em 2100 Hz e [i] em 2520; portanto, nos quinto e sexto harmônicos, respectivamente.

Gráfico 11: informante SS, vogais faladas

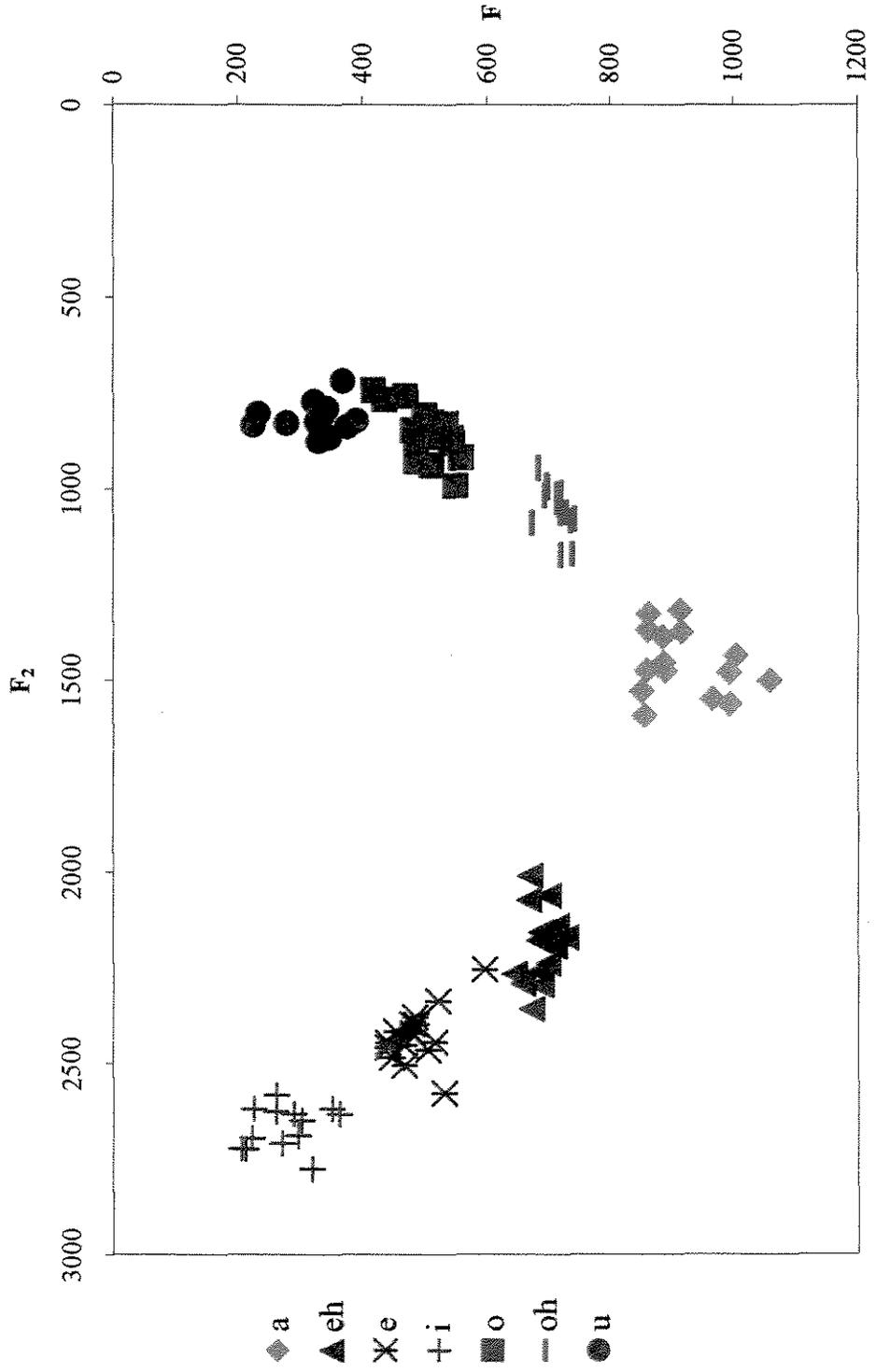
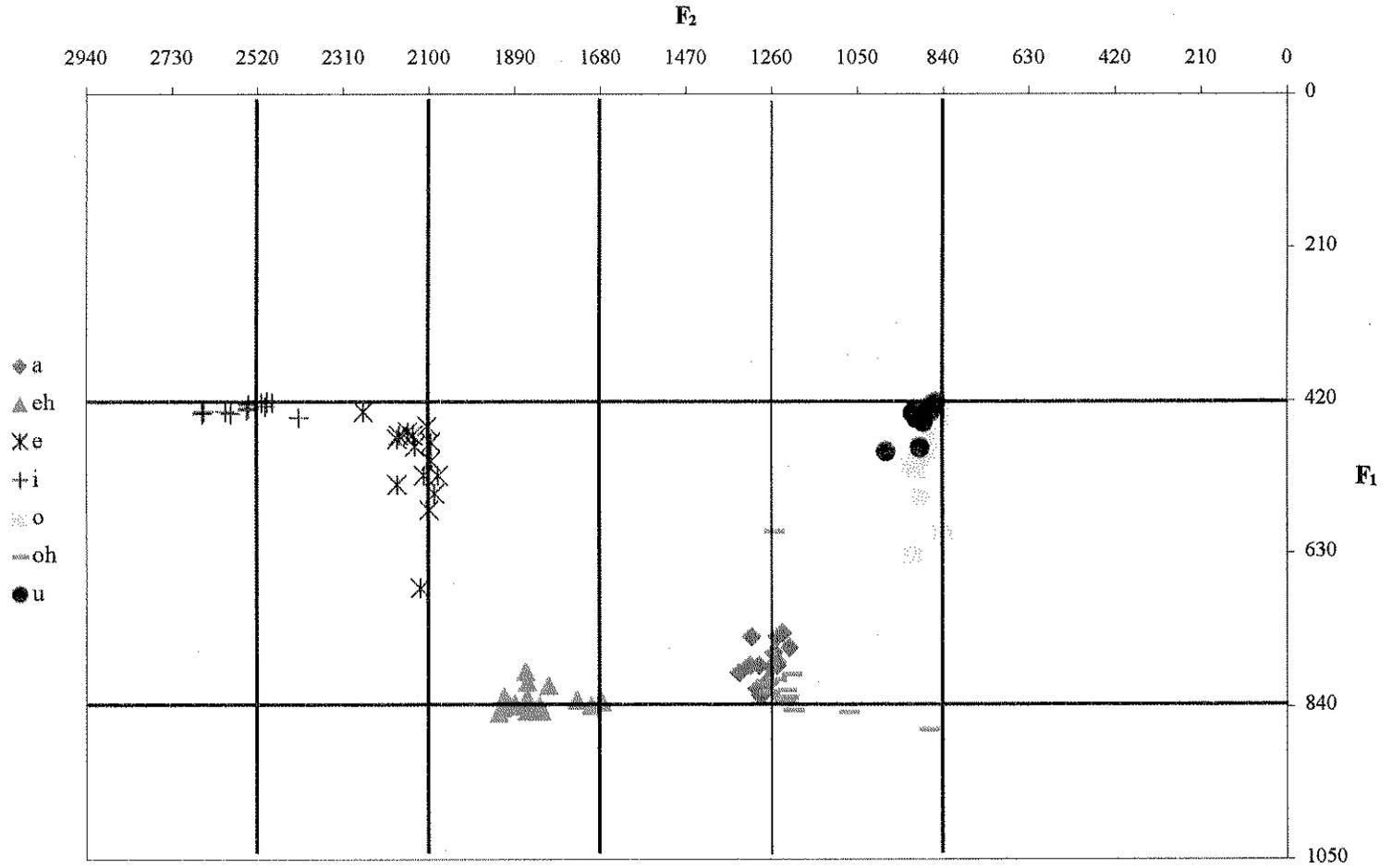


Gráfico 12: informante SS, vogais cantadas



A informante SS é a que tem o  $F_1$  mais nitidamente atraído pelo primeiro harmônico, no caso das vogais [i, o, u] e pelo segundo harmônico no caso de [ε, a, ɔ].

Podemos dizer que a informante SS é, juntamente com MG, bom exemplo do pareamento.

Tabela 10. Movimento de formantes do canto em relação à fala: informante SS

Vogais	i	e	ε	a	ɔ	o	u
F1	↑	=	↑	↓	↑	=	↑
F2	↓	↓	↓	↓	↑	=	=
F3	=	=	↑	↑	↑	=	=

Tabela 11. Padrão formântico da fala comparado ao do canto, informante SS.

Vogal	Padrão formântico fala x canto: informante SS								
	F <sub>1</sub>			F <sub>2</sub>			F <sub>3</sub>		
	Fala	Canto	Sig	Fala	Canto	Sig	Fala	Canto	Sig
i	275	431	s	2669	2541	s	3265	3127	ns
e	487	507	ns	2431	2134	s	2974	3029	ns
ε	687	829	s	2193	1839	s	2851	2983	s
a	919	780	s	1455	1271	s	2746	2985	s
ɔ	713	813	s	1050	1205	s	2896	3050	s
o	503	503	ns	856	887	ns	2991	2995	ns
u	354	443	s	951	893	ns	3073	3071	ns

### 2.3.3.5 Informante AK

A não ser por [ɛ, e], esta informante não apresenta sobreposição na fala, ao passo que o fará no canto. Nesta modalidade, [a] e [ɔ] praticamente não se distinguem, o mesmo acontecendo com [i] e [e].

O pareamento de  $F_2$  de [u] se dá no segundo harmônico; o de [a] e [ɔ] fica claramente no terceiro harmônico, 1260 Hz; e o de [e] e [i] no quinto 2100 Hz. A vogal [ɛ] tende a ter o  $F_2$  pareado em 1680 Hz, portanto, no quarto harmônico, enquanto [o] espalha-se bastante, tendendo ora para o segundo harmônico, ora para o terceiro.

Como [a] tem um comportamento diferente de [a] das outras informantes, que em geral fica mais próximo à base do retângulo do canto, o pareamento de  $F_1$  se dá com as vogais cardiais [i] e [u] e com a vogal [e], a qual fica praticamente indistinta de [i] no canto.

Gráfico 13: informante AK, vogais faladas

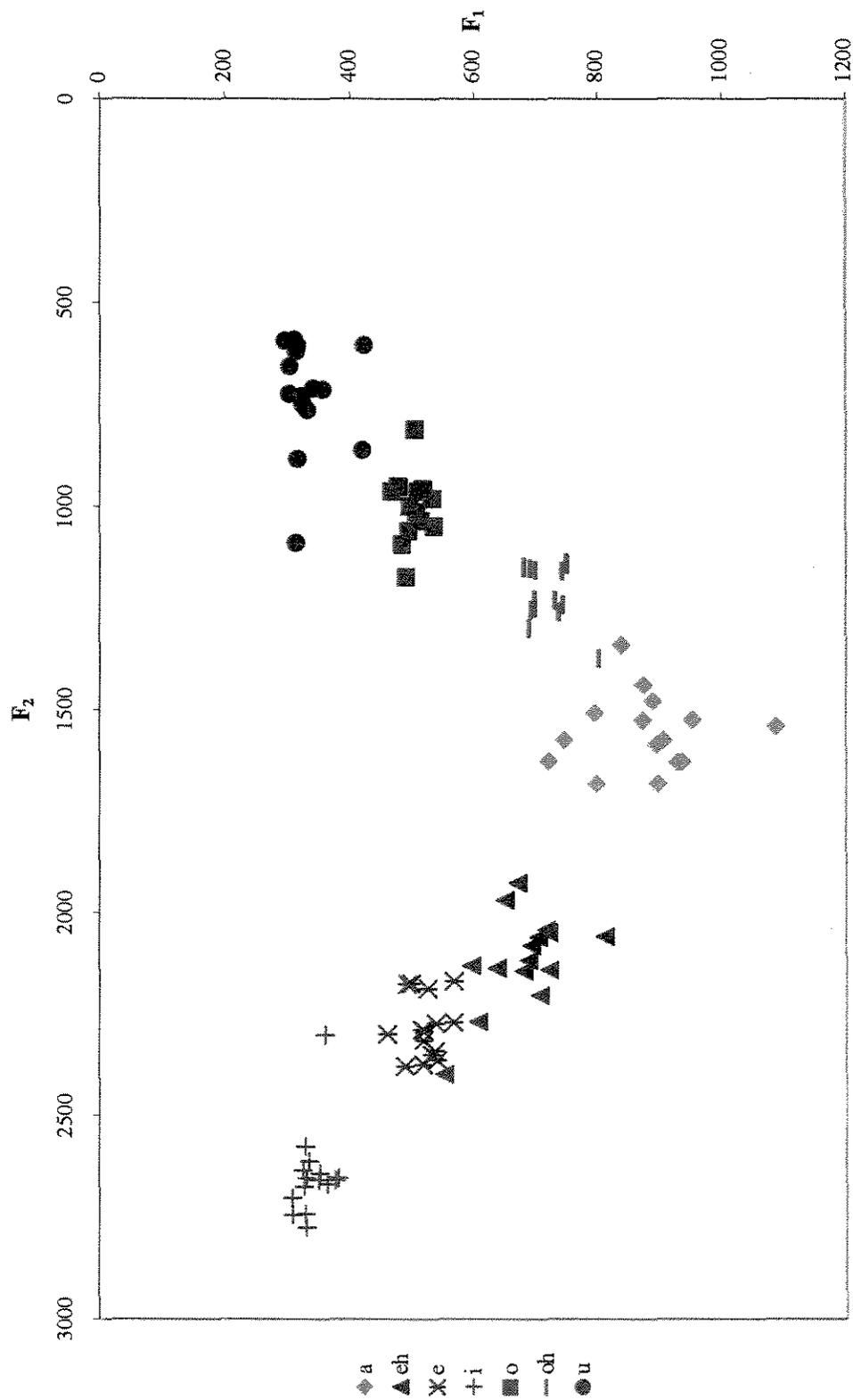


Gráfico 14: informante AK, vogais cantadas

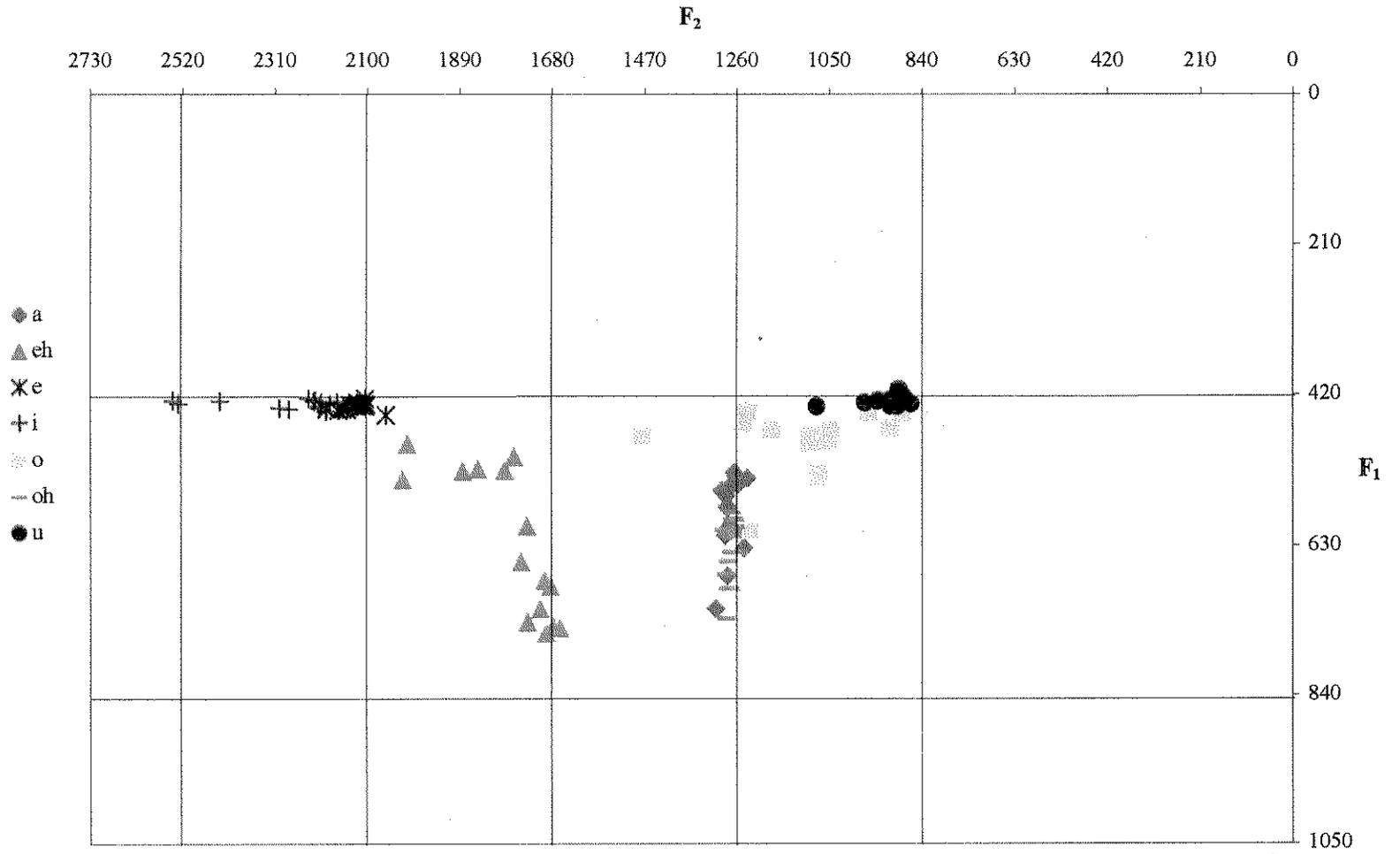


Tabela 12 Movimento de formantes do canto: informante AK.

Vogais	i	e	ε	a	ɔ	o	u
F <sub>1</sub>	↑	↓	=	↓	↓	=	↑
F <sub>2</sub>	↓	↓	↓	↓	↑	↑	↑
F <sub>3</sub>	↓	↑	=	↑	↑	↑	↑

Tabela 13. Padrão formântico da fala comparado ao do canto, informante AK.

Vogal	Padrão formântico fala x canto: informante AK								
	F <sub>1</sub>			F <sub>2</sub>			F <sub>3</sub>		
	Fala	Canto	Sig	Fala	Canto	Sig	Fala	Canto	Sig
i	342	433	s	2647	2272	s	3201	2894	s
e	522	436	s	2284	2126	s	2980	3011	s
ε	678	629	ns	2115	1778	s	3003	3023	ns
a	877	585	s	1557	1272	s	2901	3295	s
ɔ	720	637	s	1212	1274	s	2976	3186	s
o	503	486	ns	1088	997	s	3089	3365	s
u	333	429	s	725	913	s	3101	3387	s

## 2.4 Discussão final

Analisando os gráficos de F<sub>1</sub> x F<sub>2</sub> das vogais cantadas, seja no padrão global, seja por informante, foi-nos possível verificar pontos de atração para os dois primeiros formantes das vogais cantadas. Ao fenômeno de atração demos o nome de **pareamento**, pois o formante tende a parar-se aos harmônicos da nota.

Mas qual a manobra articulatória que propicia o pareamento? Para responder a esta pergunta, voltamos à proposta inicial de recuperar os movimentos articulatórios através das medidas acústicas dos três primeiros formantes.

Comparando fala e canto, tendo aquela primeira como ponto de partida, chegamos à seguinte movimentação dos formantes das vogais cantadas:  $F_1$  se eleva em todas as vogais, exceto em [a], em que diminui.  $F_2$  diminui em todas as anteriores, no [a] e eleva-se nas posteriores.  $F_3$  diminui em [i], permanece igual em [e], eleva-se em [ɛ], [a] e nas posteriores. Com base na literatura clássica, (Sundberg e Lindblom, 1971) explicamos a movimentação de cada formante (diminuição e elevação) como efeito dos articuladores do trato vocal: lábios, língua, mandíbula e laringe. No entanto, embora tentássemos entender os movimentos dos formantes como efeito da ação de diferentes articuladores, talvez um apenas seja responsável pela manobra que resulta no pareamento.

Trata-se de uma observação especular – num primeiro momento - no sentido mais amplo desta palavra. Inicialmente, restou-me observar no espelho o que acontecia. Nesta observação descartei um a um os articuladores, em vista de que apresentavam-se semelhantes fosse na fala, fosse no canto. Os lábios, por exemplo, permaneciam, sobretudo nas vogais arredondadas, como na fala. A mandíbula, por se tratar de uma região de frequência grave-média, mantinha um abaixamento discreto ainda que acentuado, comparado com o da fala; a língua e a laringe não podiam ser visualizadas. A laringe, no entanto, sabemos que é um articulador sobre o qual não temos controle. O domínio que o cantor tem sobre a laringe é relativo e adquirido depois de algum tempo de prática.

Esta observação especular contou, posteriormente, com a confirmação de imagens de vídeo, através das quais era possível ver a movimentação dos lábios e da mandíbula, na produção da frase-veículo, em suas versões falada e cantada. A movimentação dos articuladores visíveis confirma o que já falamos sobre a elevação de  $F_1$  em todas as vogais cantadas, exceto para [a], e sobretudo elimina a dúvida a respeito de um possível não

arredondamento das posteriores cantadas, que de fato são produzidas com arredondamento. Restou especular sobre a língua, mais precisamente, seu formato ou forma.

Para sustentar o fato de que a língua é responsável pelo ajuste a que se submetem as vogais cantadas para criar um tubo mais livre de constrictões, apoiar-me-ei, basicamente em dois trabalhos, o de Sundberg (1987) e o de Johnson, Ladefoged e Lindau (1993). O primeiro trata a língua como um articulador que pode tomar várias formas, dependendo da direção em que se avulta:

From figure 5.3 we conclude that the tongue contour may assume very different shapes; it may bulge in various directions – such as toward the hard palate, toward the velum, and toward the back pharynx wall – and this bulging may be slight or extreme. (p.97)

Digamos que, no canto, esse avultamento da língua seria pequeno, a fim de permitir um trato vocal com as restrições **apenas suficientes** para criar uma cor vocálica. Tendemos a considerar a língua como um articulador importante do canto, sobretudo na região de frequência investigada, a grave-média, região em que a manobra de abaixamento da mandíbula é relativizada, e outra manobra é necessária para manter a energia do sinal. A língua é responsável por marcadas diferenças articulatórias, também na fala, segundo reportado em J. L. e Lindau (1993):

As with jaw/tongue coordination in tense/lax vowels discussed in the previous section, these data suggest that the speakers differ in the articulators which they recruit to carry out a particular gesture. In Saltzman's and Munhall's (1989) framework, this is a difference in interarticulator coordination. Notice also in Fig 9 that the tongue body constriction during high vowels (those with a positive loading in factor 1, the filled circles) is more narrowly constrained to a small area on the roof of the mouth for speaker AM, while the constriction is more distributed for speaker AO. Perhaps the targets are the same, but the difference in vocal tract area function which results from the changes in interarticulator coordination

indicates that different means of raising the tongue body to the roof of the mouth may produce different area functions.

Diferente do que acontece na fala, em que cada falante realiza um formato de língua diferente, no canto, as cantoras tenderiam a buscar sempre um mesmo formato, talvez menos avultado. Este formato de língua, por sua vez, seria responsável pelas manobras articulatórias que tornam o tubo mais livre de constrictões, o que propicia o pareamento dos dois primeiros formantes com os harmônicos da nota<sup>7</sup>.

---

<sup>7</sup> Veremos no próximo capítulo, em que  $F_3$  é focalizado, que este formante realiza pareamento com o sétimo harmônico (2940 Hz).

## Capítulo 3

### Efeitos de co-produção entre consoantes e vogais na fala e no canto

Música e poesia têm exigências e destinos diferentes, que põem em novo e igualmente inconciliável conflito a voz falada e a voz cantada. A voz cantada quer a pureza e a imediata intensidade fisiológica do som musical. A voz falada quer a inteligibilidade e a imediata intensidade psicológica da palavra oral. (M. de Andrade.)

#### 3.1 Introdução

Inicialmente falamos do aspecto temporal das consoantes e das vogais, dando destaque para as primeiras. Depois, deixamos de lado as consoantes para entendermos como são articuladas as vogais cantadas. Por fim, com base nas medidas de formantes e na análise estatística, investigamos o fenômeno de co-produção entre consoantes e vogais na fala e no canto.

As medidas de formantes de que falamos neste capítulo são as medidas de  $F_1$ ,  $F_2$  e  $F_3$  de cada uma das vogais do experimento, [a,ε,e,i,ɔ,o,u], agrupadas segundo a consoante precedente [p,t,k] para, via análise estatística, verificar-se o efeito da consoante sobre a vogal, através do teste de *Duncan*, que compara o efeito das variáveis independentes agrupadas em pares.

Antes de apresentarmos os resultados na seção 3, discutimos, na próxima seção, o termo co-produção, procurando estabelecer um limite mais preciso entre o emprego desse termo e o do termo coarticulação.

### **3.2 O que estamos chamando de co-produção?**

Adotamos o termo co-produção, no lugar de coarticulação, por entendermos que o primeiro é mais abrangente que o segundo, embora seja este mais usado na literatura até Fowler (1980), o que será melhor discutido adiante. Duas vertentes lingüísticas importantes, as fonologias não lineares, herdeiras da fonologia gerativa (Chomsky & Halle, 1968) e uma fonologia calcada nos modelos dinâmicos (Browman & Goldstein, 1989, 1992) vêem o fenômeno da coarticulação de dois modos distintos. Para a primeira, a coarticulação é um fenômeno puramente mecânico, não podendo ser compreendido como cognitivo e, portanto, não pertencente ao conjunto de fenômenos fonológicos de uma língua. Já para a segunda, que não separa fonética e fonologia em módulos distintos, o fenômeno de coarticulação, sobretudo quando se fala de coarticulação antecipatória<sup>1</sup>, deve ser estudado levando-se em conta aspectos como tempo intrínseco dos gestos articulatorios que participam da produção de fala, aspecto que não é levado em conta por fonologias gerativas. Para uma literatura em torno dos dois tipos de modelo, ver Albano (2001), especialmente o capítulo 1, páginas de 14 a 23 e o capítulo 2, páginas de 52 a 60.

Mas afinal como se define coarticulação? O termo coarticulação, conforme vem sendo empregado nos estudos fonéticos desde Öhman (1966) até Whalen (1990) - apenas à guisa de exemplo e sem nos atermos à discussão de que modelo ou teoria dá conta do

---

<sup>1</sup> Há uma defesa contundente do aspecto fonológico da coarticulação em Whalen 1990, abordando que fenômenos de coarticulação antecipatórios são planejados.

fenômeno - refere-se à influência de um segmento sobre o outro, implicando fenômenos de assimilação, por exemplo.

Outra boa definição de coarticulação pode ser encontrada em Laver (1995), cuja visão parece coadunar com a visão da segunda vertente lingüística mencionada acima. A coarticulação não é um mero fenômeno físico, segundo o que é possível verificar na Parte V de seu livro *Principles of Phonetics*, ao falar de “co-ordenação” articulatória e o que chama de *phonetic settings*<sup>2</sup>:

The co-ordinatory phenomenon of accommodatory spreading of articulatory features of this sort has come to be known as co-articulation. This term was first introduced by Menzerath and Lacerda (1933), but its use became widespread in phonetics only after the publication of two influential papers by the Swedish phonetician Sven Öhman (1966, 1967). The phenomenon has also been called ‘similitude’ (Jones 1918, 9<sup>th</sup> edn 1960: 217) and ‘articulatory smoothing’ (Fujimura and Lovins, 1978: 108).

Embora sugerindo que o fenômeno é fonológico ao adotar a visão do “espalhamento de traços”, a definição de Laver ainda apresenta uma diferença fundamental em relação à de Fowler (1980), que não acredita em apenas um sistema articulatório para dar conta de duas classes diferentes dos sons da fala: a classe das vogais e a das consoantes.

Para falarmos de co-produção segundo a concepção de Fowler (op.cit), devemos entender, inicialmente, que há duas classes distintas de som na fala, pertencentes, cada qual a um conjunto de estruturas coordenativas diferentes:

---

<sup>2</sup> Não se encontrou uma boa tradução para esse termo: poderia ser cenário fonético.

[...] the capacity for coproduction derives from an adaptive property of speech that the two classes of articulatory gestures, consonants and vowels, are products of different (coordinate) neuromuscular systems. ( p. 129)

A discussão de Fowler representou o marco inicial do que hoje é referido como teoria de co-produção (Harrington, Fletcher e Roberts, 1995) e inseriu-se posteriormente no âmbito da fonologia que elege modelos dinâmicos, conforme preconizado por Kelso, Saltzman e Tuller (1986), não só porque leva em conta a questão do tempo intrínseco, mas por rejeitar a dicotomia entre o plano e a execução, considerando as diferentes naturezas de ordem biomecânica da classe das vogais e da classe das consoantes. Para Fowler, a co-produção existe na fala para coordenar dois sistemas articulatorios diferentes, ligados a dois tipos de gestos diferentes por serem produzidos por conjuntos musculares diferentes.

À luz dos resultados obtidos, sobretudo no canto, em que a influência mútua entre segmentos não apontava para um fenômeno que pudesse ser descrito como coarticulação, e acolhendo com entusiasmo a visão de Fowler (1980) a que nos referimos acima, adotamos os termo co-produção a fim de discutirmos a interação de consoantes e vogais no canto.

### **3.2.1 Estudos envolvendo coarticulação**

Os estudos resumidos a seguir, voltados para investigar os aspectos dinâmicos de organização da fala (*connected speech*) fazem parte de uma linhagem de estudos coarticulatórios, como clássicos da literatura; são eles: Stevens & House, (1963), Öhman (1966), Schouten & Pols (1979).

Stevens & House (1963) realizaram um experimento contemplando oito vogais e todas as consoantes do inglês americano que pudessem ocorrer no início e no final de

sílaba, formando portanto, uma sílaba CVC, no contexto de um logatoma iniciado por /hə/. Os autores questionam a variação das frequências formantes (i) de falante para falante, (ii) dependendo do contexto, (iii) ao longo do tempo; e apontam para um possível equívoco quanto à falta de um procedimento geral para atestar os valores das frequências formantes de forma confiável. A conclusão a que chegam é de que não só o contexto consonantal, mas também a própria natureza dinâmica do mecanismo articulatorio, são responsáveis pelas variações de padrão formântico. Diante de tanta variação, a percepção tem, segundo os autores, um papel importantíssimo na recuperação de regras para dar conta dos efeitos contextuais e que

...must somehow be learned by listeners during the acquisition of language through constant exposure to their own speech and to the speech of others. (p. 127)

Öhman (1966) realiza medidas espectrográficas de transições dos formantes de palavras VCV, com as consoantes /b, d, g/ e as vogais /Y, ø, α, u/ do sueco, realizando paralelamente, e com a mesma estrutura silábica, estudos com o inglês americano e com o russo. Sua investigação aponta para a coarticulação antecipatória, uma vez que ele verifica forte influência dos valores de transição da consoante para a vogal inicial, a partir do que propõe que se considere a coarticulação um efeito fonêmico; portanto, um fenômeno lingüístico. Ao final do trabalho, Öhman discute a Teoria de *Locus* de Delattre, Liberman e Couper (1955), destacando que os resultados obtidos não reconhecem um ponto fixo para o valor de *locus* das consoantes investigadas, variando muito conforme a vogal.

Schouten & Pols (1979) investigaram as transições CV e VC em palavras do holandês com estrutura silábica CVC, em que as consoantes iniciais eram /t,d,n,χ,p,r/, as

vogais /i,e,a,u/, e as consoantes finais eram as mesmas que as iniciais, exceto /d/. Estas palavras foram lidas em duas condições diferentes: isoladas numa lista e inseridas num texto. Tais, condições não têm influência sobre a regularidade do fenômeno de coarticulação, possibilitando definir a área de *locus* no espaço da vogal para a maioria das consoantes, segundo concluem os autores.

Oliveira (1999, 2001) investigando a coarticulação CV com logatomas idênticos aos nossos ( nossa escolha foi proposital, justamente para fazer, com a investigação da fala, um certo paralelo com Oliveira 2001) chega a resultados satisfatórios para defender a coarticulação perseveratória como um fenômeno não mecânico, na medida que específico a uma língua; portanto, bom candidato a fonológico. Oliveira investigou as vogais /i,a,u/, nos seguintes contextos /la<sup>1</sup>CV/ ; /<sup>1</sup>laCV/ e CV, em que C era /p,t,k/. Diferente do nosso trabalho, Oliveira contemplou, ainda, as vogais em posição acentual pós-tônica e monossilábica. Quanto ao logatoma /la<sup>1</sup>CV/, portanto o mesmo que usamos na fala e no canto, conseguimos replicar seus resultados.

Oliveira (1999) ainda compara seus resultados com os de Stevens & House (1963) e Schouten & Pols (1979) verificando serem coincidentes com os seus:

Schouten & Pols 1979 encontram um aumento de  $F_2$  em /a/ e /u/ no contexto de /t/ em relação a /p/, o que coincide em parte com os nossos resultados (tabelas 1 a 3 e 7 a 9), mas ressaltamos que a consoante /k/ não estava no corpus do estudo dos autores e portanto a comparação deve ser relativizada . Reproduzimos na citação abaixo do artigo de Schouten & Pols (1979), o trecho que expõe os resultados sobre coarticulação perseveratória.

*“/u/ tends to be pulled more to the left by preceding /t/, /d/ and /n/, and more to the right by preceding /p/, /x/ and /r/. The same may be true for /ε/ and /a/ in the text word condition,*

*but the effect is not totally convincing. Since the first dimension is roughly equivalent to F<sub>2</sub>, there seems to be agreement here with a finding reported by Chollet (1976), namely that F<sub>2</sub> of /u/ and /a/ is consistently higher after /t, /d/ and /n/, than after /p/.”(p.12)*

Stevens & House (1963) também encontraram resultados bastante parecidos com os nossos e os de Schouten & Pols, mas diferentemente de Schouten & Pols, detectaram um efeito mais convincente para a vogal /u/. (p.25)

Os resultados do trabalho de Oliveira (2001) não nos autorizam inteiramente a afirmar que a coarticulação perseveratória é um fenômeno lingüístico, já que um dos sujeitos investigados não apresenta efeito coarticulatório para F<sub>2</sub> em /u/ pós-tônico, por exemplo, mas ilumina, no âmbito dos estudos fonéticos sobre o PB, uma questão cara à Lingüística que é de indagar se os fenômenos considerados puramente mecânicos não são lingüísticos.

### **3.3 Resultados**

#### **3.3.1 Padrão geral da fala**

Primeiramente, apresentamos os resultados referentes à fala, obtidos na análise estatística, comparando-os com os resultados de Oliveira (2001). Neste capítulo, diferente do anterior em que havia mais variáveis, apresentamos apenas o padrão geral, ou seja, o padrão que engloba todas as informantes.

As tabelas de 1 a 7 apresentam, por vogal, os valores dos três primeiros formantes para cada uma das consoantes [p,t,k] e o valor de *P*, que indica significância ou não significância. Quando há efeito de co-produção sobre a vogal, toda a coluna é destacada em cinza-claro para F<sub>2</sub> e em cinza escuro para F<sub>3</sub>.

Tendo encontrado resultados que se explicam pelo processo de assimilação e por estarmos tratando de uma modalidade mais tradicional dos estudos fonético acústicos, a fala, o termo *coarticulação* será empregado. Como era esperado, não encontramos efeitos da consoante junto a F<sub>1</sub>, mas sim em F<sub>2</sub> – claramente – e também, mas timidamente, em F<sub>3</sub>. Iniciamos, a seguir, as explicações detalhadas do fenômeno de coarticulação, dedicando uma seção a F<sub>2</sub> e outra a F<sub>3</sub>.

Tabela 1: efeito das consoantes sobre a vogal [a]

Vogal [a] falada		
F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>
[p] 932	[k] 1584	[t] 2882
[k] 902	[t] 1545	[p] 2802
[t] 884	[p] 1460	[k] 2560
p = k = t	k = t > p	t = p > k
P 0.1318	P 0.0001	P 0.0001

Tabela 2: efeito das consoantes sobre a vogal [ε]

Vogal [ε] falada		
F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>
[t] 638	[p] 2114	[t] 2918
[k] 633	[k] 2104	[p] 2881
[p] 631	[t] 2043	[k] 2834
t = k = p	p = k > t	t = p p = k t > k
P 0.8972	P 0.0220	P 0.0498

Tabela 3: efeito das consoantes sobre a vogal [e]

Vogal [e] falada		
F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>
[k] 477 [t] 461 [p] 458 k = t = p	[k] 2334 [p] 2305 [t] 2292 k = p = t	[p] 2968 [t] 2962 [k] 2927 p = t = k
P 0.3738	P 0.2492	P 0.4423

Tabela 4: efeito das consoantes sobre a vogal [i]

Vogal [i] falada		
F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>
[t] 320 [p] 313 [k] 308 t = p = k	[p] 2502 [k] 2496 [t] 2432 p = k = t	[p] 3169 [k] 3073 [t] 2989 p > k = t
P 0.6568	P 0.2093	P 0.0005

Tabela 5: efeito das consoantes sobre a vogal [ɔ]

Vogal [ɔ] falada		
F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>
[k] 695 [p] 677 [t] 675 k = p = t	[t] 1128 [k] 1124 [p] 1073 t = k > p	[p] 2806 [t] 2757 [k] 2680 p = t > k
P 0.2960	P 0.0005	P 0.0001

Tabela 6: efeito das consoantes sobre a vogal [o]

Vogal [o] falada		
F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>
[t] 481 [k] 473 [p] 473 t = k = p	[t] 976 [k] 904 [p] 873 t > k = p	[p] 2907 [t] 2902 [k] 2784 p = t > k
P 0.8451	P 0.0001	P 0.0001

Tabela 7: efeitos das consoantes sobre a vogal [u]

Vogal [u] falada		
F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>
[k] 389 [t] 370 [p] 357 k = t = p	[t] 912 [k] 819 [p] 743 t > k > p	[t] 2951 [p] 2920 [k] 2906 t = p = k
P 0.1473	P 0.0001	P 0.3699

### 3.3.1.1 Resultados da influência da consoante sobre a vogal na fala concernentes a F<sub>2</sub>

As vogais que apresentaram os efeitos de coarticulação junto a F<sub>2</sub> foram: [a,ε,ɔ,o,u], ou seja, a maioria delas, sendo que [e] e [i] não apresentaram significância quanto ao segundo formante. Apresentar um F<sub>2</sub> ou F<sub>3</sub> mais elevado em um ambiente consonantal e diminuído em outro significa que a vogal é suscetível ao lugar de produção, já a alteração de F<sub>1</sub>, indica que a vogal é suscetível ao modo de produção (Kent e Read, 1992).

De todas as vogais, a única que não apresenta diferença de valor formântico, qualquer que seja a consoante, é [e], e então, pode-se concluir que, por se tratar de uma

vogal produzida com uma configuração do trato vocal mais próxima da neutra, não é suscetível a grandes mudanças articulatórias, seja quanto à abertura do trato, seja quanto à posição da língua. Assim, a não ser para a vogal [e], para todas as outras vogais desse experimento, teceremos alguns comentários explicativos e apontaremos para os fenômenos que podem ser considerados coarticulatórios *strictu sensu* e os que não podem.

A vogal [a] de nosso experimento apresenta o seguinte efeito: nos ambientes [k] e [t],  $F_2$  se eleva em relação a [p]. Esse resultado é semelhante ao de Oliveira (2001) que obteve, para três de cinco de seus informantes o resultado  $k > t > p$  e para dois, resultado igual ao nosso  $k = t > p$ , no mesmo contexto, ou seja em /la'ka/. A explicação mais plausível que podemos dar sobre o aumento de  $F_2$  de [a] no ambiente [k] é de que o avanço do corpo da língua parece produzir, nas vogais mais baixas como [a] e [ɔ], uma menor constrição faríngea que faz com  $F_2$  se eleve e  $F_3$  abaixe. O ambiente [t] mantém  $F_2$  elevado, e o ambiente [p] diminui o  $F_2$ , o que já é esperado. Assim, pode-se falar em coarticulação no caso desta vogal, na modalidade falada.

A vogal [ɛ] tem  $F_2$  mais elevado em ambiente [p] e [k] e menor em ambiente [t]. Isso pode talvez ser explicado pelo efeito do chamado *locus* de t, que estabelece um valor de  $F_2$  comum a todas as vogais produzidas depois de [t] e que seria algo em torno de 1800 Hz, indicando coarticulação. Também indica coarticulação, a manutenção de  $F_2$  elevado em ambiente [k], dado que o *locus* das velares pode ficar em torno de 2000 Hz. Já a consoante [p] que juntamente com o ambiente [k], apresenta um  $F_2$  mais elevado, parece propiciar um ambiente dissimilatório, pois seu *locus* fica em torno de 1100 – 1500 Hz, ou seja, bem abaixo dos 2114 Hz, valor médio de  $F_2$  para /pɛ/.

A vogal [ɔ], assim como a vogal [a], apresenta um efeito em que a bilabial tem o valor mais baixo de  $F_2$ :  $t = k > p$ . Assim seria possível repetir, para aquela primeira vogal, a mesma explicação dada para o que ocorre com [a].

A vogal [o] apresenta efeito de coarticulação no ambiente [t], elevando  $F_2$  em relação a [k] e [p], podendo receber a mesma interpretação dada aos efeitos das consoantes sobre a vogal [u], a seguir.

Por fim a vogal [u] apresenta resultado idêntico ao de Oliveira (2001), em que  $t > k > p$ . Nesta vogal, os efeitos de coarticulação são facilmente verificados. Seu  $F_2$ , o mais baixo dentre as posteriores, apresenta um valor próximo ao do *locus* da bilabial, tende a se elevar quando em ambiente de consoante velar, [k], ainda que em direção ao *locus* mais baixo e eleva-se ainda mais em ambiente [t].

### 3.3.1.2 Resultados da influência da consoante sobre a vogal na fala concernentes a $F_3$

Para  $F_3$ , continuamos com um número expressivo de vogais do experimento que apresentam efeito de coarticulação. A não ser pela vogal [ɛ], cujo resultado de  $P$  é um valor muito próximo do limiar para a diferença marginal, as vogais [a,i,ɔ,o] apresentam diferença significativa, dependendo da consoante precedente. Para falar das mudanças de  $F_3$  basearemos em Fant (1960) por nos dar algumas pistas a respeito do correlato articulatorio deste formante, o que é difícil encontrar nos estudos da área. Assim, é preciso ressaltar, de antemão, que as explicações dadas não são conclusivas, visto que não dispomos de dados articulatorios para respaldá-las, mas antes especulativas.

A vogal [a] tem o menor  $F_3$  no ambiente [k], sendo elevado em [t] e [p]. O mesmo ocorre com [ɔ, o], assim é possível, a um tempo, explicar o que acontece com essas três vogais, quando precedidas por uma velar. As vogais em questão têm, no ambiente [k], o  $F_3$  mais baixo, o que pode ser explicado pelo fato de  $F_3$  estar suscetível, no caso da central e das posteriores, a alterações que ocorrem nas partes anteriores da constrição da língua (Fant, 1960). Comparando a produção de [k] com a de [p] e [t], pode-se dizer que a região da cavidade anterior sofre com a seguinte mudança: há avanço do corpo da língua ocasionando, como já falamos na seção anterior,  $F_2$  alto e  $F_3$  baixo.

Já para [i], [p] se destaca de [k] e [t] que apresentam  $F_3$  iguais entre si e mais baixos do que o do ambiente [p]. Segundo os *loci* estimados por Kewley-Port (1983)<sup>3</sup>, [t] deveria estar acima de [k] e [p], nesta ordem, no entanto, [p] parece elevar o  $F_3$ , o que interpretamos como dissimilação. Assim, mantém-se a distância  $F_2-F_3$  (ver tabela 4), o que possibilita uma melhor identificação da vogal.

### 3.3.2 Padrão geral do canto

Na fala, encontramos resultados gerais bastante semelhantes àqueles de investigações, algumas já clássicas (Öhman, 1966, Steven e House, 1963), sobre a co-produção CV (Oliveira, 2001). A maior parte dos efeitos ocorrem em  $F_2$ , ao passo que nenhum efeito ocorre em  $F_1$ . Já no canto, três vogais apresentam efeito da consoante sobre

<sup>3</sup> Abaixo, reescrevemos a tabela de Kewley-Port (1983) apud Kent e Read (1992)

	$F_1$	$F_2$	$F_3$
Bilabial	0	1100-1500	2200-2400
Alveolar	0	1800	2500- 2700
Velar	0	1500-2500	2200-3000

a vogal em  $F_1$ . Assim, a próxima seção será dedicada a explicar este fenômeno. Há também efeitos encontrados em  $F_2$  e algo que deve receber nossa atenção em  $F_3$ .

Os fenômenos de co-produção no canto relacionam-se ao fenômeno de pareamento do formante com a frequência fundamental e/ou com o harmônico da nota. Este último fenômeno foi apresentado e discutido no capítulo anterior. Para facilitar as discussões que se seguem, apresentamos, abaixo, a tabela 8, que ilustra a correspondência entre a frequência fundamental da nota e seus harmônicos com os três primeiros formantes e as vogais.

Tabela 8: Valores da Frequência Fundamental e seus Harmônicos em Hertz

$F_0$ e Harmônicos	Formantes	Vogal
$H_1$ 420	$F_1$	[i, u]
$H_2$ 840	$F_1$	[ε, a, ɔ, o]
$H_2$ 840	$F_2$	[u]
$H_3$ 1260	$F_2$	[a, ɔ, o]
$H_4$ 1680	$F_2$	[ε]
$H_5$ 2100	$F_2$	[e, i]
$H_6$ 2520	$F_2$	[i]
$H_7$ 2940	$F_3$	todas as vogais

Também auxiliam na “visualização” da relação co-produção/pareamento, a inserção de uma coluna, nas tabelas de 9 a 15, com o valor da frequência fundamental da nota ou do harmônico desta que estaria pareando com o formante.

Tabela 9: efeitos da consoante sobre a vogal [a]

Vogal [a] cantada					
F <sub>1</sub>	H <sub>2</sub>	F <sub>2</sub>	H <sub>3</sub>	F <sub>3</sub>	H <sub>7</sub>
[p] 781 [t] 766 [k] 755 p = t t = k p > k	840	[t] 1328 [k] 1316 [p] 1308 t = k = p	1260	[t] 3040 [p] 3026 [k] 2988 t = p = k	2940
P 0.0148		P 0.5013		P 0.01958	

Tabela 10: efeitos da consoante sobre a vogal [ɛ]

Vogal [ɛ] cantada					
F <sub>1</sub>	H <sub>2</sub>	F <sub>2</sub>	H <sub>4</sub>	F <sub>3</sub>	H <sub>7</sub>
[t] 775 [p] 752 [k] 738 t = p p = k t > k	840	[k] 1801 [p] 1796 [t] 1739 k = p > t	1680	[p] 3022 [t] 2966 [k] 2964 p = t = k	2940
P 0.0970		P 0.0285		P 0.2546	

Tabela 11: efeitos da consoante sobre a vogal [e]

Vogal [e] cantada					
F <sub>1</sub>	H <sub>1</sub>	F <sub>2</sub>	H <sub>5</sub>	F <sub>3</sub>	H <sub>7</sub>
[k] 562 [p] 533 [t] 527 k > p = t	420	[p] 2054 [k] 2041 [t] 2024 p = k = t	2100	[k] 3000 [p] 2991 [t] 2983 k = p = t	2940
P 0.0407		P 0.2846		P 0.8878	

Tabela 12: efeitos da consoante sobre a vogal [i]

Vogal [i] cantada					
F <sub>1</sub>	H <sub>1</sub>	F <sub>2</sub>	H <sub>3</sub>	F <sub>3</sub>	H <sub>4</sub>
[t] 443 [p] 440 [k] 436 t = p = k	420	[p] 2353 [k] 2324 [t] 2227 p = k > t	2100	[p] 3107 [t] 2960 [k] 2938 p > t = k	2940
P 0.1112		P 0.0019		P 0.0043	

Tabela 13: efeitos da consoante sobre a vogal [ɔ]

Vogal [ɔ] cantada					
F <sub>1</sub>	H <sub>2</sub>	F <sub>2</sub>	H <sub>4</sub>	F <sub>3</sub>	H <sub>7</sub>
[k] 756 [p] 750 [t] 749 k = p = t	840	[p] 1241 [k] 1232 [t] 1230 p = k = t	1260	[t] 3022 [k] 3004 [p] 3002 t = k = p	2940
P 0.9385		0.9120		P 0.6619	

Tabela 14: efeitos da consoante sobre a vogal [o]

Vogal [o] cantada					
F <sub>1</sub>	H <sub>1</sub> ou H <sub>2</sub>	F <sub>2</sub>	H <sub>3</sub>	F <sub>3</sub>	H <sub>4</sub>
[p] 625 [t] 609 [k] 607 p = t = k	420 ou 840	[t] 1097 [k] 1074 [p] 1054 t = k = p	1260	[t] 3091 [p] 3039 [k] 3025 t = p p = k t > k	2940
P 0.2484		P 0.3090		P 0.0398	

Tabela 15: efeitos da consoante sobre a vogal [u]

Vogal [u] cantada					
F <sub>1</sub>	H <sub>1</sub>	F <sub>2</sub>	H <sub>2</sub>	F <sub>3</sub>	H <sub>3</sub>
[p] 480 [k] 474 [t] 457  p = k = t	420	[t] 999 [p] 935 [k] 932  t > p = k	840	[t] 3094 [k] 3058 [p] 3038 t = k k = p t > p	2940
P 0.2965		P 0.0075		P 0.0830	

### 3.3.2.1 Resultados da influência da consoante sobre a vogal no canto concernentes a F<sub>1</sub>

As vogais que apresentam diferença significativa entre valores de F<sub>1</sub> são: [a, e], diferente de [ε] que apresenta apenas diferença marginal (ver tabela 10).

A vogal [a] apresenta uma queda de F<sub>1</sub> quando o ambiente [k] é comparado ao ambiente [p]. Isso se deve ao fato de que, na produção de [k], a elevação da língua acaba ocasionando menor abaixamento da mandíbula, ao passo que fica este mais livre em ambiente [p]. Não se pode falar aqui de coarticulação, pois o que acontece é que a fala impõe uma restrição (produção de [k]) ao canto, atenuando uma restrição deste último, ou seja, o abaixamento da mandíbula.

Além da explicação articulatória, destaca-se a relação dos efeitos de co-produção com o pareamento entre formantes das vogais e os harmônicos da nota. Pela tabela 8 e pelas tabelas de 9 a 15, pode-se ver que F<sub>1</sub> de /pa/, /ta/ e /te/, /pε/ estão mais próximos de 840 Hz (segundo harmônico ou H<sub>2</sub>) que os valores de /ka/ e /kε/, corroborando a explicação de que, para a produção de [k] cantado, a restrição musical submete-se à da fala.

A vogal [ɛ] apresenta também um  $F_1$  menor em [k], mas, agora, comparado a [t]. Embora a diferença seja marginal, ( $P= 0.0970$ ) parece de novo estar ocorrendo, aí, em ambiente [k], o mesmo que ocorre com a vogal [a].

A vogal [e], ao contrário das outras duas vogais analisadas nesta seção, apresenta o  $F_1$  mais elevado em ambiente [k]. Sendo uma vogal média alta, e, portanto, de  $F_1$  mais baixo que os de [a,ɛ], vê-se, na tabela 11, que o pareamento não se dá nitidamente, ainda que vá em direção ao  $H_1$  da nota (420 Hz). O ambiente [k] mostra-se novamente – só que agora com uma vogal média alta – menos afeito às necessidades articulatórias da língua para o pareamento. O que parece haver é que, a fim de manter a constrição de [e] mais relaxada, para cumprir as exigências do canto, abaixa-se a mandíbula, o que resulta em  $F_1$  mais elevado que o de /kɛ/ e /ka/. Perceptualmente, o resultado deve ser [e] com uma cor mais própria de [ɛ].

### 3.3.2.2 Resultados da influência da consoante sobre a vogal concernentes a $F_2$

A vogais que apresentam resultado de significância para valor de  $F_2$  são [ɛ, i, u].

A vogal [ɛ] apresenta no canto ( $k = p > t$ ) resultado semelhante ao da fala ( $p = k > t$ ), mas que, provavelmente, não pode receber a mesma explicação, uma vez que a contraparte cantada de tal vogal apresenta valores formânticos que a aproximam de uma vogal ainda mais baixa e centralizada. Veja-se o valor médio (padrão geral, capítulo 2) dos dois primeiros formantes de [ɛ]:  $F_1$  da fala é 634 Hz,  $F_1$  do canto é 755 Hz;  $F_2$  da fala 2087 é Hz,  $F_2$  do canto é 1739 Hz. A explicação que podemos dar para o comportamento da

vogal cantada no ambiente [t] é que este é sensível ao abaixamento da mandíbula e à manobra articulatória que causa a centralização, e, assim, o  $F_2$  em 1739 Hz revela-se mais próximo do quarto harmônico ou  $H_4$  (1680 Hz).

Para a vogal [i] o ambiente [t] apresenta o  $F_2$  mais baixo, um efeito de dissimilação, sendo neste caso, também como acontece com [ε], uma sensibilidade deste ambiente à centralização. Daí o resultado ser, novamente, o pareamento do segundo formante, em 2227 Hz, com o  $H_5$  (2100 Hz).

Há semelhanças entre o que acontece com [u] falado e o [u] cantado. Podemos ver, no capítulo 2, que esta vogal não apresenta, em três informantes, diferença de valores para alguns formantes. As informantes 2 e 4 apresentam  $F_2$  e  $F_3$  do [u] cantado igual aos valores desses formantes para [u] falado; e a informante 3 apresenta os mesmos valores de  $F_1$  e  $F_3$  nas diferentes modalidades para a vogal em questão. Dado que  $t > p = k$ , o efeito de  $F_2$  mais elevado em [t] do que em [p] não é contrário à coarticulação perseveratória clássica. Sendo este resultado muito parecido com o da fala, podemos mais uma vez hipotetizar que a restrição da fala impõe-se à do canto, que seria o pareamento de  $F_2$  com  $H_1$ . O que se vê é que  $F_2$  de /tu/ em 999 Hz, distancia-se do  $H_2$  (840 Hz), se comparado aos  $F_2$  de /pu/ e /ku/ (ver tabela 15). Estudos perceptuais futuros, levando em conta o efeito de [t] sobre [u], seriam bastante úteis para verificar se tal efeito não está relacionado com a inteligibilidade do texto cantado.

### 3.3.2.3 Resultados da influência da consoante sobre a vogal concernentes a $F_3$

As vogais cantadas que apresentam resultado de significância para os valores de  $F_3$  são [i, o]. Já a vogal [u] apresenta resultado de diferença marginal (P 0.0830).

O  $F_3$  do [i] cantado se eleva em ambiente [p] em relação a [t] e [k], semelhante ao resultado de [i] falado que é:  $p > k = t$ . Sabe-se que as oclusivas [t, k] tendem a palatalizar-se diante de [i], a primeira, por sua homorganicidade em relação a [i], o que anula um efeito coarticulatório perseveratório, fazendo predominar o antecipatório; e a segunda por elevar o dorso da língua em direção ao palato. O canto parece manter isso, a partir do que podemos dizer que há um entendimento entre as manobras das duas modalidades. Assim o canto acaba promovendo o pareamento de  $F_3$  com o sétimo harmônico ( $H_7$ ), conforme está assinalado na tabela 12.

Quanto à vogal [o] não é possível traçar um paralelo com sua contraparte falada, o que torna ainda mais obscura a análise que já se ressentia de correlatos articulatorios precisos para  $F_3$ . O que se aplica também a [u]. Por outro lado, levando em conta o fenômeno do pareamento, vê-se que os ambientes [p] e [k] buscam parear  $F_3$  com  $H_7$ , ao passo que o ambiente [t] tende a buscar um valor mais elevado para  $F_3$ , distanciando-se de  $H_6$ .

## 3.4 Discussão final

Na fala, pelos resultados expressivos em  $F_2$ , é possível dizer que há efeitos da consoante sobre a vogal que podemos chamar de efeitos de coarticulação perseveratória, o que é notado em [a, o], sobretudo no ambiente [p]; e em [ε] e [o] no ambiente [t]. A vogal

[u] é a melhor representante da coarticulação na fala, em todos os três ambientes investigados, [p, t, k].

Quanto a  $F_3$ , ainda não estamos certos de que os efeitos da consoante [k] sobre as vogais [a, ɔ, o] e de [p] sobre [i] venham a ser um fenômeno coarticulatório.

No canto, [a, ε, e] apresentam efeito de co-produção no ambiente [k] para  $F_1$ , por uma restrição da própria modalidade - ou seja, o abaixamento acentuado da mandíbula - envolvendo o aspecto modo de articulação. Vimos que, para as vogais baixas, o [k] não propicia o abaixamento necessário na modalidade cantada. Isso faz com que, neste ambiente, o valor de  $F_1$  distancie-se do segundo harmônico (840 Hz). O inverso acontece com a vogal alta no que tange o abaixamento da mandíbula; e, mesmo assim,  $F_1$  distancia-se do primeiro harmônico, 420 Hz (ver tabela 11).

Para  $F_2$  de [i, ε] o ambiente [t] parece ser, ao contrário do [k], um ambiente propício para a manobra do canto que centraliza a vogal, pois ajusta o segundo formante em torno de  $H_4$  e  $H_5$ , respectivamente (ver tabelas 10 e 12). Por fim, o  $F_2$  de [u] no ambiente [t] se distancia mais do segundo harmônico (840 Hz), a partir do que concluímos que este ambiente, diferentemente do que acontece com as vogais anteriores, não favorece a manobra do canto, que é a de tornar o tubo o mais uniforme possível, segundo o que vimos na discussão final do capítulo anterior.

O que ocorre com  $F_3$  de [i] e [o] é de difícil interpretação. Em /pi/,  $F_3$  se distancia mais de  $H_7$ , e em /to/ este formante se aproxima mais desse harmônico. Limitamo-nos a interpretar isso como a manutenção da distância  $F_2$ - $F_3$  (ver tabelas 12 e 14), sobretudo no

caso de [i], em que esta distância deve ser menor para garantir o timbre característico da vogal.

Podemos concluir este estudo dizendo que os ambiente [k] e [t], este último seguido de [u], evidenciam efeitos de co-produção no canto, pois competem, em favor da fala, com o fenômeno do pareamento.

Em vista disso, podemos dizer, ainda, que, embora a manobra do canto para uniformização do tubo seja uma restrição musical importante, não seria possível introduzir consoantes no canto, se restrições da fala – como a da produção do [k], que envolve um movimento importante da língua – não estivessem presentes; trazendo à tona os efeitos de co-produção e possibilitando, assim, suficientemente, a presença do texto na canção.

Cabe agora reafirmar o que já foi dito ao final da sessão 3.3.2.2, quanto a estudos perceptuais futuros. As hipóteses levantadas neste capítulo estão a exigir tais estudos, que foram deliberadamente adiados neste trabalho, por se tratar de investigação que requer elaboração complexa e demorada.

## Conclusão

Sendo este um estudo comparativo de aspectos fonético-acústicos da fala e do canto, cabe apontar agora as diferenças entre estas duas modalidades. Vimos, no fim de cada capítulo, que o canto mantém suficiente ou minimamente os traços da fala. Fazendo assim é que o homem tem produzido fala e canto, ao mesmo tempo, com o mesmo instrumento. Este fazer implica num saber musical e arriscarei dizer também lingüístico, se não fizermos a dicotomia saber/fazer, presente em teorias reconhecidamente lingüísticas, que tentam dar conta dos universais lingüísticos e apontar, a partir deles, a competência do falante, deixando de discutir os dados variáveis do nível da performance.

Ora, o presente estudo, que investigou aspectos fonético-acústicos da fala e do canto, e que poderia, no âmbito da ciência da linguagem ser classificado como um estudo voltado para fenômenos tão somente físicos da fala, acaba se vendo, no capítulo 3, quando trata da co-produção, diante do pressuposto teórico da vertente lingüística que não faz dicotomia entre o plano e ação e tem sua melhor representante na Fonologia Articulatória (doravante FAR), calcada em um modelo dinâmico de produção de fala, conforme exporemos a seguir.

#### 4.1. Uma perspectiva diferente para os estudos lingüísticos: a do aspecto dinâmico da produção de fala.

No capítulo 3 desta tese, ao compararmos aspectos de co-produção entre a fala e canto em português brasileiro, adotamos a visão de Fowler (1980), segundo a qual a dimensão do tempo e a não distinção entre o planejamento de uma ação e seu executor<sup>1</sup>, devem ser levadas em conta num modelo de tempo extrínseco da coarticulação. Essa visão ocupa, com destaque, um lugar no modelo teórico da perspectiva dinâmica da produção de fala, defendido por Kelso, Saltzman e Tuller (1986), cujo desafio é partir de dados de fala variáveis, devido a fatores tais como taxa de elocução, acento lexical ou frasal, para tentar chegar a uma estrutura coordenativa invariável. Ou seja, estes autores propõem um modelo de produção de fala, acolhendo todos os aspectos da fala que as fonologias gerativas fazem questão de deixar de lado. Sobre as duas linhagens opostas convivendo no interior da ciência da linguagem, Pierrehumbert & Pierrehumbert (1990) dizem o seguinte:

Among schools of thought on this subject, at one extreme is the position that phonological theory should deal directly with the dynamics, the surface regularity emerging more or less automatically as macroscopic property of the system much as classical bulk thermodynamics emerges from the statistical mechanics of interacting particles. At the other extreme is the position that the continuous dynamics is so hopelessly far from the deep regularities of language, that it is pointless to do anything other than deal with the discrete representation directly. (p. 466)

---

<sup>1</sup> Como evidência empírica para defender a não distinção, a autora cita o resultado de um experimento em que sujeitos mantinham um tempo de latência tão longo quanto longo fosse o enunciado lingüístico que tivessem de produzir a um dado sinal, mesmo sabendo de antemão o que teria de produzir, e mesmo sendo esse enunciado uma frase feita. Sobre o que a autora conclui: "... the reason why subjects fail to 'take advantage' of the time before the presentation of the signal to devise a plan for their utterance is that **plans are self-executing**". (p.121, grifo meu).

Parece que estamos diante de um embate: de um lado uma vertente tradicional da lingüística que erige seus modelos teóricos a partir do conceito de que a unidade lingüística é simbólica e abstrata e, de outro, uma vertente que vem desafiar essa visão tradicional, dizendo ser impossível dar conta dos fenômenos fonológicos sem olhar para detalhes de sua realização física. Assim é que Kelso, Saltzman e Tuller propõem o seguinte:

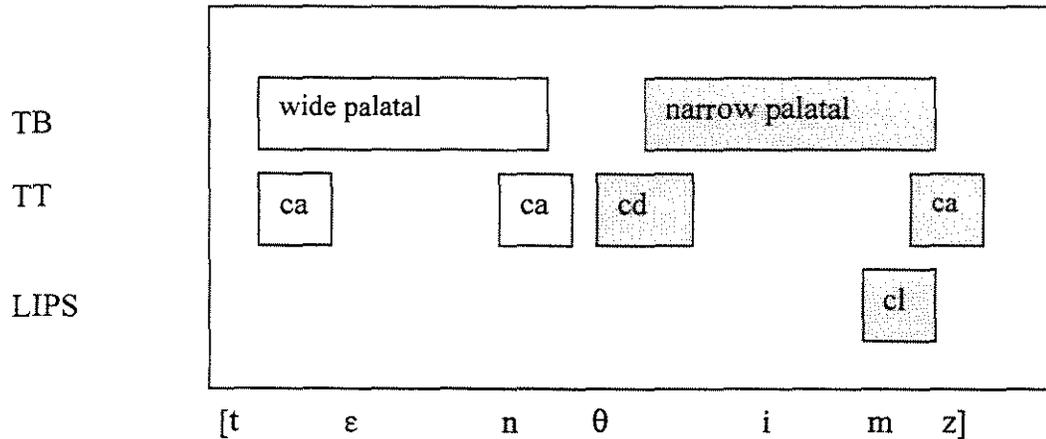
We outlined a dynamical account of speech production that differs radically from views that characterized speech as planned sequence of static linguistic/symbolic units that are different in kind from the physical processes involved in the execution of such a plan. Rather, we hypothesize that the coordinative structures for speech are dynamically defined in a unitary way across both abstract “planning” and concrete articulatory “production” levels. These units are not timeless, but rather incorporate time in an intrinsic manner (cf. Fowler, 1980; Bell-Berti & Harris, 1981).

Browman & Goldstein estenderam a visão dinâmica à fonologia e desenvolveram a Fonologia Articulatória (1989, 1990, 1992), propondo o gesto como unidade fonológica, - em lugar do traço ou do segmento - cuja representação requer apenas um nível, ao mesmo tempo fonético e fonológico, uma vez que este modelo teórico não faz distinção entre esses dois componentes gramaticais<sup>2</sup>. Para representar o gesto, que, conforme frisam Browman & Goldstein é constituído por eventos físicos que se desdobram durante a produção de fala, constituídos pela formação de constrições no trato vocal, através do tempo, os autores propõem a pauta gestual (*gestural score*)<sup>3</sup>. Um exemplo de pauta gestual (Browman e

<sup>2</sup> Este modelo fundamenta-se na chamada dinâmica de tarefa (*task dynamics*) - esta por sua vez fundamentada em princípios biológicos e físicos gerais de movimento coordenado - que segundo Browman & Goldstein (1992) pode modelar ações de multi- articuladores coordenadas. Sugiro fortemente a leitura de Hawkins (1992) para entender os construtos básicos da dinâmica de tarefa.

<sup>3</sup> A pauta gestual organiza-se a partir dos articuladores, a saber: TB, *tongue body*, TT, *tongue tip*, LIPS, *lips*. As caixas mais curtas ou mais longas representam o tempo e trazem inscritos o ponto e o modo de articulação. Na reprodução a seguir, na linha de TT, *ca* das caixas brancas é *closure alveolar* e *ca* da caixa sombreada é *critical alveolar*, *cd* é *critical dental*; na linha LIPS *cl* é *closure labial*.

Goldstein, 1989, p. 218), mostrando a pausa entre as palavras *ten* (caixas brancas) e *themes* (caixas sombreadas), vem abaixo para ilustrar o que estamos dizendo.



Farnettani (1997) destaca algumas vantagens de tratar as variações coarticulatórias no modelo gestual. Em seu desdobramento no tempo e no espaço, os gestos podem ou não partilhar os mesmos articuladores, derivando daí maior ou menor sobreposição destes gestos. O resultado pode ser a maior ou menor influência entre esses gestos sobrepostos, o que se pode verificar diretamente na pauta gestual, sem ter que se postular outros níveis de análise, que é o que acontece em outros modelos teóricos. No entanto, em sua resenha de estudos coarticulatórios, Farnettani trata apenas da coarticulação antecipatória, o que nos obriga a voltar ao estudo de Oliveira (2001) para falar da coarticulação perseveratória sobre a qual obtivemos resultados semelhantes ao deste autor, no último estudo desta tese. Oliveira, apesar de verificar dados que apontam para a universalidade do fenômeno que investiga – já que para  $F_2$  encontra dados semelhantes entre o holandês, o inglês americano e o português do Brasil – e diante disso não poder afirmar que a coarticulação perseveratória não é puramente mecânica, reconhece que o modelo dinâmico possibilita-lhe

explicar outros aspectos do fenômeno, como a gradiência (nem todas as posições acentuais apresentam o mesmo resultado, o que é aceitável num modelo gestual, uma vez que ele contempla a magnitude e a redução do gesto).

Assim, se podemos continuar buscando evidências lingüísticas junto a fenômenos que em princípio parecem ser mecânicos, é possível encontrar, dadas as restrições musicais de afinação e volume, e da diferença de padrão articulatorio, evidências de que, ao se produzir a fala cantada, esta-se coordenando planos e execuções lingüísticos e musicais, e não apenas reproduzindo um enunciado estereotipado com alturas determinadas por uma tonalidade musical. Contudo, para tanto, seriam necessários outros experimentos delineados num estudo não apenas descritivo, como é o caso desta tese. Um modelo que adota a FAR e a amplia, defendido por Albano (2001), mostra-se promissor para se realizar um estudo aprofundado a respeito da relação de possíveis planos lingüísticos na execução da fala cantada. Ao tratar das constrições que definem gestos articulatorios, a FAAR (Fonologia Acústico-Articularória) desloca a ênfase da FAR, do articulador para regiões articulatorias associadas a efeitos acústicos (cf. a Teoria Acústica de Fant). As manobras articulatorias hipotetizadas a partir dos resultados acústicos desta tese são consistentes com essa visão mais flexível da realização das constrições que constituem os gestos articulatorios.

Uma forte razão para advogar isso, sobretudo em se tratando de estudos que implicam a capacidade musical, é que os modelos que se inserem na teoria de dinâmica de tarefa aspiram a dar conta de capacidades cognitivas mais gerais, enraizadas na motricidade. Isso é facilmente reconhecido nos trechos transcritos abaixo, falando, o primeiro, da fundamentação a teoria (daí eu ter julgado importante uma tradução) e outro de ritmo.

A dinâmica de tarefa é um modelo geral de controle do movimento que foi originalmente desenvolvido para explicar tarefas tais como alcançar objetos e ficar ereto, e só mais recentemente vem sendo aplicado à fala. Baseia-se em princípios biológicos e físicos gerais de movimentos coordenados, mas fundamenta-se em termos dinâmicos, mais do que em termos anatômicos ou fisiológicos. Envolve uma abordagem relativamente radical, mais abstrata do que muitos sistemas tradicionais, e vem se mostrando um meio particularmente útil de análise de produção de fala, em parte porque decompõe movimentos complexos num conjunto de tarefas funcionalmente independentes. (Hawkins, 1992, p. 9; tradução minha).

Outra proposta na mesma linha é a de Port, Cummins e Gasser (1995), de que um estudo aprofundado sobre ritmo lingüístico é mais do que justificável - dado o caráter rítmico das línguas. Tal estudo ganha uma abordagem dinâmica desses autores, que destacam o caráter temporal dos eventos lingüísticos.

In this paper, we will look at evidence for speech rhythm, where rhythm is defined by the repetition of similar events after similar time intervals. We will need to study both the temporal structure of physical speech signals and also develop a psychological model which can perform the right kind of measurements on the physical signal and make the appropriate predictions (about the similar events and similar intervals). We need to measure linguistically relevant durations: voice-onset time, vowel durations, interstress intervals etc. (Port, 1986; Port, 1990; Port et al., 1995). These measures are language-specific, and so we anticipate that a general perceptual model for metrical systems will be individually parameterized for each language.

We are hopeful that eventually many of the phenomena currently described using traditional linguistic concepts such as 'stress', 'metrical hierarchy', 'foot', 'short vowel', 'stress timing', etc. can be reinterpreted as manifestations of a general perceptual model.  
(p.4)

Por ora, limitamo-nos a apresentar os resultados da comparação descritiva que já têm o mérito de desvendar alguns aspectos do texto cantado no plano fonético-acústico, fundamentando estudos futuros que contemplem aspectos lingüísticos de fala e canto, que podem se valer do modelo dinâmico para explicar a complexa negociação realizada a todo momento pelas cantoras entre restrições que garantem inteligibilidade mínima ao texto cantado e restrições que garantem afinação e volume.

#### 4.2 Conclusões finais

Retomando o que foi dito no início deste capítulo dedicado às conclusões, podemos dizer, então, ao cabo de três estudos voltados para a fala cantada, contida em uma das nossas canções de câmara, comparando-a com a fala falada, o seguinte:

Quanto à *Estrutura Temporal*, **o canto alonga as vogais**, responsáveis, nessa modalidade, pela entonação da nota musical, **produzindo as consoantes num tempo mínimo**. Ainda que encurtadas em termos absolutos e relativos, as consoantes são responsáveis pela manutenção do texto da canção.

Quanto ao *Padrão Formântico*, podemos dizer o seguinte: no longo tempo em que as vogais são produzidas no canto, a manutenção da frequência fundamental para garantir afinação é propiciada por duas manobras articulatórias: abaixamento do maxilar (Sundberg, 1977) e **menor avultamento da língua** (exceto para [a]).

As conseqüências acústicas destas duas manobras articulatórias são: a **sobreposição desejável** de [i,e]; [e,ε]; [a,ɔ] e [o,u], **o pareamento de F<sub>1</sub>** com a frequência fundamental ou H<sub>1</sub>(420 Hz) e com o segundo harmônico ou H<sub>2</sub> (840 Hz) e **o pareamento de F<sub>2</sub>** com os harmônicos de 2 a 7. Vimos, também, no capítulo 3, ao desvendar o que acontecia com o

terceiro formante em termos de co-produção consoante/vogal, que ocorre o **pareamento de  $F_3$  com o sétimo harmônico  $H_7$  (2940 Hz)**.

Quanto à *Co-produção*, podemos concluir que **os efeitos das consoantes – sobretudo [k] – são sentidos sobre a vogal quando a produção da consoante requer um trato vocal com constrictões que não podem fugir muito do padrão da fala**, impedindo o fenômeno do pareamento. Defendemos que a coarticulação, que pode ser encontrada na fala, não pode aparecer sempre no canto devido: i) à longa duração da vogal cantada e ii) à manobra articulatória que possibilita um tubo mais uniforme no canto.

Pelo exposto acima, o canto parece apresentar mais diferenças do que semelhanças com a fala; diferenças essas que se devem às exigências musicais. Como vimos no capítulo 1, a consoante é importante na fala, ocupando 34% do logatoma, ao passo que a vogal ocupa 25%. Já no canto, a consoante é encurtada. No capítulo 2, descobrimos um triângulo vocálico redesenhado para o canto, em que as pontas parecem se dobrar para seu interior, num movimento de centralização das vogais, o que indica que o canto requer um tubo mais uniforme, ao passo que a fala, por excelência, requer as constrictões desse tubo. No capítulo 3 aprendemos que não parece haver coarticulação no canto, mas antes, há efeitos de co-produção entre consoantes e vogais, efeitos esses que estão sob a influência do pareamento.

Cientes dessas diferenças, devemos dizer agora algo sobre o mito da má dicção, sobre o qual falamos na Introdução. Se o padrão articulatório do canto na região de frequência grave-média é diferente do da fala, sabemos, baseados em estudos como os de Scotto di Carlo (1985) que a tendência é de haver ainda diferença maior entre os padrões formânticos, quanto mais alta for a região de frequência<sup>4</sup>, impossibilitando cantores e ouvintes de perseguirem a inteligibilidade do texto. Assim, devemos dizer que a má dicção

---

<sup>4</sup> Ou seja, cantar uma frase inteira numa tessitura estreita de notas agudas.

não é uma explicação razoável para a deformação de vogais e consoantes no canto, e, portanto, deve permanecer mítica, porém mito desvendado. Não se pode esperar formas canônicas dos enunciados lingüísticos sob a pressão de um cenário articulatório voltado para as exigências musicais.

Se há inteligibilidade do texto, –que segundo Scotto di Carlo (op.cit) fica em torno de 47% no registro médio- há restrições lingüísticas agindo, indicando que os traços da fala minimamente mantidos no canto são suficientes para garantir à outra ponta do processo, na qual está o ouvinte, a recuperação do texto.

## Referências Bibliográficas

ALBANO, E. *O gesto e suas bordas*, Campinas, Mercado de Letras, 2001.

ANAIS do Primeiro Congresso da Língua Nacional Cantada, São Paulo, Departamento de Cultura do Estado de São Paulo, 1938.

ANDRADE, M. de *Aspectos da música brasileira*, Martins, Brasília, 1975.

\_\_\_\_\_. Pequena história da música, 8ª ed. São Paulo, Livraria Martins Editora, 1977.

\_\_\_\_\_. Anais do primeiro congresso da língua nacional cantada, São Paulo, Departamento de Cultura, 1938.

BARBOSA, P.A. e BAILLY, G. Characterisation of rhythmic patterns for text to – speech synthesis, *Speech Communication*, 15 (1-2), 1994, pp.127-137.

BLUMSTEIN, S. e STEVENS, K. Acoustic invariance in speech production: evidence from measurements of the spectral characteristics of stop consonants, in *The Journal of the Acoustical Society of America*, 66, 1979, pp.1001-1017.

BROWMAN, C e GOLDSTEIN, L. Articulatory gestures as phonological units, in *Phonology Yearbook*, 6, 1989, pp.201-251.

\_\_\_\_\_. Representation and reality: physical systems and phonological structure, in *J. of Phonetics*, 18, 1990, pp.411-424.

- \_\_\_\_\_. Articulatory Phonology: An Overview, in *Phonetica* 49: pp 155-180, 1992.
- CARLSSON, G. e SUNDBERG, J. 1992 Formant frequency tuning in singing, in *Journal of Voice*, Vol 6, No. 3, 1992, pp 256-260.
- CHOMSKY, N. e HALLE, M. The sound pattern of English, Cambridge, The MIT Press, 1968.
- DUARTE, F.J.C. A fala e o canto no Brasil: dois modelos de emissão vocal, in *ARTEunesp*, São Paulo, 10:87-97. 1994.
- FANT, G. (1960) *Acoustic Theory of Acoustic Perception*, 2a. edição, The Hague-Paris, Mouton, 1970.
- FARNETANI, E. Coarticulation and Connected Speech in HARDCASTLE and LAVER (orgs) *The Handbook fo Phonetic Sciences*, 1997.
- FOWLER. C. Coarticulation and theories of extrinsic timing, in *J. of Phonetics* 8, 1980, pp.113-133.
- FUJIMURA, Osamu. The syllable: its internal structure and role in the prosodic organisation. Palek, B. (ed.) *Proceedings of the Linguistics and Phonetics Conferences*, Praga, Charles University Press, 1994, pp 53-93.
- HALLE, M., HUGHES, G., e RADDLEY, J. P. Acoustic properties of stop consonants, in *The Journal of the Acoustical Society of America*, 29, 1957, pp. 107-116.
- HARDCASTLE, W. J e LAVER, J. *The handbook of phonetic sciences*, Oxford, Blackwell Publishers, 1997.

- HARRINGTON, J. FLETCHER, J. e ROBERTS, C. Coarticulation and the accented/unaccented distinction: evidence from jaw movement data, in *Journal of Phonetics*, 23, 1995, pp.305-322.
- HAWKINS, S. An introduction to task dynamics, in Docherty, L. *Papers in laboratory phonology II*, Cambridge, CUP, 1992, pp.9-25.
- JAKOBSON, R., FANT, G. e HALLE, M. *Preliminaries to speech analysis: the distinctive features and their correlates*, Cambridge, The MIT Press, 1952.
- JOHNSON, K., LADEFOGED, P., e LINDAU, M. Individual differences in vowel production, in *The Journal of the Acoustical Society of America*, vol. 94, no. 2, 1993, pp 701-714
- KELSO, J.; SALTZMAN, E. e TULLER, B. The dynamical perspective on speech production: data and theory, in *Journal of Phonetics*, 14, 1986, pp.29-59.
- KENT, R.D. e READ, C. *The acoustic analysis of speech*, San Diego, Singular Publishing Group, 1992.
- KESSINGER e BLUMSTEIN, S. Effects of speaking rate on voice-onset time and vowel production: some implications for perception studies, in *Journal of Phonetics*, 26, 1998, pp 117-128.
- KUGLER, P. e TURVEY, M. *Information, natural law and the self-assembly of rhythmic movement*. Hillsdale, Lawrence Erlbaum, 1987.
- LADEFOGED, Peter. *Elements of acoustic phonetics* (2a. edição), Chicago e Londres, The University of Chicago Press, 1996.
- LAVIER, J. *Principles of Phonetics*. Cambridge University Press, 1994.

LINDBLOM, B. e SUNDBERG, J. Acoustical consequences of Lip, tongue, Jaw and Larynx movement, in *The Journal of the Acoustical Society of America*, Vol. 50, no. 4, 1971, pp 1166-1179.

LOPES, P. M. *Dicção Lírica*, Edição própria, Rio de Janeiro, 1944.

MARIZ, V. *A canção de câmara no Brasil*, Porto, Editora-Livraria Progredior, 1948.

\_\_\_\_\_. *A canção brasileira: popular e erudita*, 5ª. Edição, Rio de Janeiro, Nova Fronteira, 1985.

MOTTA MAIA, E. *No reino da fala: a linguagem e seus sons*, São Paulo, Ática, 1991.

NEF, C. *Histoire de la Musique*, Paris, Payot, 1948.

ÖHMAN, S. E. G., Coarticulation in VCV utterances: spectrographic measurements in *The Journal of the Acoustic Society of America*, 39, 1966, pp 151-168.

OLIVEIRA, L.C.F. de, *Relatório Iniciação Científica à Fapesp*, 1999.

OLIVEIRA, L. C. F. de, Por uma abordagem dinâmica dos processos fônicos, in SILVA et al., *Revista Letras*, Curitiba, 55, jan/jun., Ed. da UFPR 2001, pp. 93 – 113.

PETERSON, G. E. e BARNEY, H. E. Control methods used in a study of vowels, in *The Journal of the Acoustical Society of America*, 24, 1952, pp. 175-184.

PIERREHUMBERT, J. e BECKMAN, M. E. *Japanese tone structure*, Cambridge, The MIT Press, 1988.

- PIERREHUMBERT, J. T. e PIERREHUMBERT R. T., On attributing grammars to dynamical systems, in *Journal of Phonetics*, 18, 1990, pp. 465-477.
- PORT, R.; CUMMINS, F. e GASSER, M. *A dynamic approach to rhythm in language: toward a temporal phonology*. Manuscrito inédito, Indiana University. Disponível em <http://www.cs.indiana.edu/~port/>, 1998.
- RAPOSO DE MEDEIROS, B. Estudo preliminar da inteligibilidade das vogais cantadas no português brasileiro, in *Estudos Lingüísticos XXIX*, 2000, pp. 657-662.
- REANEY, G. The middle ages [(1960)] in STEVENS, D. *A history of song*, WW Norton & Company, Nova Yorque, 1970, pp 15-66.
- SCHOUTEN, M. e POLS, L. Vowel segments in consonantal contexts: a spectral study of coarticulation – part I, in *J. of Phonetics* 7, 1979, pp.1-23.
- SCOTTO DI CARLO, Nicole Étude acoustique e statistique de l'influence des consonnes sua la justesse des voyelles subsequente em voix chantée, in *Travaux de l'Institut de Phonétique d'Aix*, V. 4, 1977, pp 237-260.
- \_\_\_\_\_, Pour quoi ne comprend-on pas les chanteurs d'opéras?, in *La Recherche*, IX, 1978, 89, pp. 495-497.
- \_\_\_\_\_. e GERMAIN, A. A perceptual study of the influence of pitch in the intelligibility of sung vowels. *Phonetica*, 42, 4, 1985, pp.188-197.
- \_\_\_\_\_. L'organisation temporelle de la syllabe dans la parole e dans le chant: étude préliminaire, Communication au *XII Congrès des Sciences Phonétiques Aix-en-Provence*, Août, 1991.
- \_\_\_\_\_. Sons et sens dans l'opéra, in *Médecine des Arts*, 4 1993, pp. 4-11.

- \_\_\_\_\_. L'intelligibilité de la voix chantée, in *Médecine des Arts*, 10, 1994, pp 2-15
- STEVENS, K. e HOUSE, A. Perturbation of vowel articulations by consonantal context: an acoustical study, in *Journal of Speech and Hearing Research*, 6, n° 2, 1963, pp. 111-128.
- STEVENS, K. Diverse acoustic cues at consonantal landmarks in *Phonetica*, 57, 2000, pp139-151.
- SUNDBERG, J. Articulatory differences between spoken and sung vowels in singers, *STL-QPSR*, 1/1969, pp 33-46.
- \_\_\_\_\_. Vibrato and vowel identification, *STL-QPSR* 2-3/1975, pp 49-60.
- \_\_\_\_\_. The acoustics of the singing voice, *Scientific American*, Vol 236, no. 3, 1977, pp 82-91.
- \_\_\_\_\_. *The science of the singing voice*, Northern Illinois University Press, Dekalb, 1987.
- \_\_\_\_\_. Level and center frequency of the singer's formant in *TMH Quarterly Progress Status Report*, Nos. 3-4, 1999, pp 87-94.
- SYRDAL, A. K. & GOPAL, H.S. A perceptual model of vowel recognition based on the auditory representation of American English vowels, in *The Journal of the Acoustical Society of America* vol. 79, no. 4, 1986, pp 1086 – 1100.
- WHALEN, D.H. Coarticulation is largely planned, *Journal of Phonetics* 18, 1990, pp.3-35.

WISNIK, J. M. *O som e o sentido*, 2ª. edição, São Paulo, Companhia das Letras, 1999.

WOOD, A. *The physics of music*, 5ª edição, Londres, Methuen & Co. Ltd, 1950.

## Apêndice

### I. Estudo piloto

Contextos extraídos das canções brasileiras em que se inserem as vogais tônicas em foco do estudo piloto e suas respectivas regiões de frequência (grave, média e/ou aguda):

#### *Bachiana número 5* de Heitor Villa-Lobos

- Prá lembrar o Cariri (vogal em foco: [i], aguda )
- Que tua flauta do sertão quando assobia (vogal em foco: [i], média)
- Irerê meu passarinho (vogal em foco: [e], grave)
- Bem-te-vi, eh! Sabiá (vogal em foco [a], média)
- E sabiá da mata cantadô (vogal em foco [o], média)
- E sabiá da mata sofrédô (vogal em foco [o], aguda)

#### *Canção do Amor* de H.V.Lobos

- Vencida pela dor (vogal em foco: [i], grave)
- Tão doce aquela hora ( vogais em foco [o] e [ɔ] graves)

#### *Seresta “Saudade da Minha Vida”* de H.V. Lobos

- Aquela inocência feliz sem saber (vogal em foco: [ɛ], média)

#### *Essa negra Fulô* de Oscar Lorenzo Fernandez

- Ah! Essa negra Fulô (vogal em foco [ɛ], aguda)
- Essa negra Fulô (idem, grave)

#### *Modinha “Quem sabe...”* de Carlos Gomes

- Se esqueceste, se esqueceste, se esqueceste o juramento (vogal em foco: [e], o primeiro, agudo e o segundo médio)
- Da saudade agro tormento (vogal em foco [a], aguda)
- Quisera saber agora ([ɔ], média)

#### *Cantiga de Ninar* de Francisco Mignone

- Sobre teus olhos a mão devagar (vogal em foco [a], grave)

*Trovas 1 e 2* de Alberto Nepomuceno

- Ao escutar a minha voz (vogal em foco [ɔ], aguda)

*Seresta "Serenata"* de H. Villalobos

- Violões suplicando promessas de venturas (vogal em foco [u], grave)
- Que o violão traduz (vogal em foco [u], média)

*Melodia Sentimental* de H. Villalobos

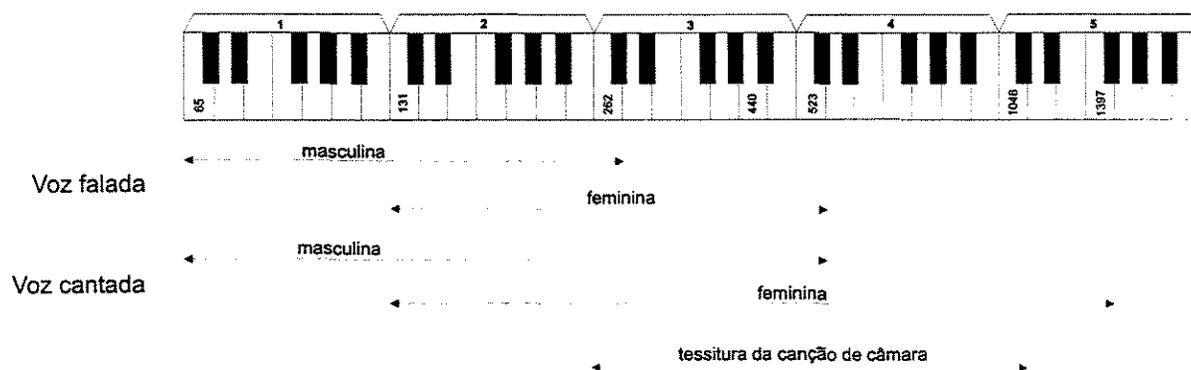
- As asas da noite que surgem (vogal em foco [u], aguda)

**II. As notas musicais<sup>1</sup> e suas freqüências em hertz.**

Nota	Hertz	Região de freqüência
Dó	261,6	Grave
Ré	293,7	
Mi	329,6	
Fá	349,2	
Fá#	370	
Sol	392	
Lá	440	
Si	493,9	Média
Dó	523,3	
Ré	584,7	
Mi	652,2	
Fá	698,4	Aguda
Fá#	740	
Sol	784	
Lá	880	
Si	987,8	
Dó	1046,6	

<sup>1</sup> O primeiro dó da tabela, de 261, 6 Hz, corresponde ao dó central ou dó 3. Na literatura, este mesmo dó é também chamado de dó 4.

### III. Extensão da voz falada e cantada.<sup>2</sup>



### IV. Tipos vocais de vozes femininas<sup>3</sup>

Contralto: voz que prolonga o registro médio em direção ao grave.

Mezzo soprano: voz intermediária entre soprano e contralto.

Soprano:

- soprano dramático: além da extensão de soprano, emite notas graves.
- soprano lírico: voz brilhante e extensa.
- soprano lírico-ligeiro: voz brilhante, sua extensão é em direção aos agudos.
- soprano ligeiro (ou coloratura): grande extensão de registro agudo.

### V. Conteúdo do cedê da tese

1. Apresentação

2. Frase da Cantiga de Ninar de Francisco Mignone :

*Canto baixinho uma velha cação de ninar.*

Exemplos da frase-veículo do experimento da tese.

 frase cantada

frase falada

<sup>2</sup> Figura baseada em Scotto di Carlo, 1993.

<sup>3</sup> O texto da classificação das vozes femininas baseou-se nos dados disponíveis no sítio [www.berlioz.com.br](http://www.berlioz.com.br).

3.  Canto /la'ki/ baixinho numa velha canção de ninar.

4. /la'ki/

5.  /la'pe/

6. /la'pe/

7.  /la'te/

8. /la'te/

9.  /la'ka/

10. /la'ka/

11.  /la'tɔ/

12. /la'tɔ/

13.  /la'po/

14. /la'po/

15.  /la'pu/

16. /la'pu/