

“As cores da voz”: expressão das emoções no timbre da voz cantada*

The “voice colours”: expressing emotions through the singing voice timbre

Ana Leonor Pereira

Resumo

A expressão das emoções na voz cantada faz-se segundo um código com vários parâmetros tais como: tempo, vibrato, ritmo, afinação, amplitude e timbre. O objectivo deste estudo é compreender como é que o timbre expressa as emoções do *performer*. Doze cantores profissionais foram gravados a cantar peças musicais eruditas com um conteúdo emocional específico – Tristeza, Alegria, Raiva e Medo – e um vocalizo (Neutro). Os dados consistiram em 432 vogais extraídas das gravações aceites perceptivamente por um júri composto por sete estudantes de canto. Os resultados mostraram que as frequências dos formantes mudam de acordo com as emoções: Raiva e Alegria têm frequência de ocorrência dos formantes mais altas do que Tristeza e Medo. Esta investigação mostrou, também, que os formantes têm funções especializadas: F2 e F1 discriminam vogais, F4, F3 e F5 discriminam indivíduos e F5 e F1 discriminam emoções. Os resultados apontaram para o facto de, independentemente do tipo de voz, cada emoção ter a sua própria assinatura tímbrica, visível no espectro, com uma conexão específica com um formante: Raiva/F5, Tristeza/F4 e Alegria/F3. Este último ponto necessita de investigação futura.

Palavras-chave: timbre, voz cantada, formantes, expressão emocional. ◀◀

Introdução

A comunicação emocional em palco é determinante para a qualidade da *performance* e determinante para o grau de talento atribuído ao intérprete. A questão do talento é uma questão central, uma vez que determina se um aprendiz de música poderá ser um concertista ou não. Enquanto que se supõe que a técnica é adquirida ao longo dos anos de

* Trabalho realizado no âmbito do Mestrado em Ciências da Fala do Instituto de Ciências da Saúde, Universidade Católica Portuguesa
aleonorpereira@gmail.com

Abstract

Expression of emotions in singing is connected to a code with several parameters such as tempo, vibrato, rhythm, pitch, amplitude and timbre. The aim of this study is to understand how timbre expresses the performer's emotions. Twelve professional singers were recorded while singing pieces of music with specific emotional content – Sadness, Joy, Anger and Fear – and a vocalise (Neutral). The data consisted of 432 vowels extracted from the perceptually accepted singing recordings judge by seven singing students. The results showed that formant frequencies changes are dependent on the emotions: Anger and Joy show higher formant frequency than Sadness and Fear. This investigation showed, also, that formants have specific functions: vowel discrimination is dependent on F2 and F1, individual discrimination is done by F4, F3 and F5, and emotional discrimination by F5 and F1. The results pointed out for the fact that, independently of the voice type, each emotion has its own specificity, visible in spectra, with a particular connection to a formant: Anger /F5, Sadness/ F4, Joy/F3. Further research needs to be considered on this last issue.

Keywords: timbre, singing voice, formants, emotional expression. ◀◀

estudo de um músico, supõe-se o talento como inato. Deste modo, considera-se que não se pode ensinar o talento. Ora, podendo aferir, em parte, o talento de um músico pela sua capacidade de comunicar as emoções através do som, urge perguntar o que significa possuir esta capacidade. Por outro lado, se houver fundamentos científicos que permitam determinar de que modo tal comunicação é conseguida, então pode ser que aquilo que se tem atribuído ao talento também possa, em certa medida, ser ensinado. Quis-se, assim, averiguar quais os correlatos no sinal acústico da expressão das emoções na voz cantada.

Que a comunicação das emoções na voz cantada depende de um conjunto diversificado de factores quer na produção, quer na percepção, é já matéria consensual. Se um destes factores, especificamente o tímbrico, permite, por si só, determinar e definir emoções é a questão que se põe. Há estudos de caso que sugerem que há espectros-tipo relativos a determinadas emoções, mas a cartografia destes espectros-tipo não está delineada [1, 2, 3]. A maior parte da investigação que pretende analisar esta relação fá-lo com o intuito de compreender a percepção da emoção por parte do ouvinte e não com a intenção de descrever o modo de produção da expressão da emoção por parte do cantor [4, 5, 6].

Tal como para a voz falada [7], a primeira e mais premente preocupação foi a de saber quais os diversos parâmetros vocais afectados pelas emoções. Morozov (1996) concluiu que são os seguintes os parâmetros da voz cantada afectados pelas emoções: tempo e ritmo, dinâmica, duração das sílabas e das micropausas, características do vibrato, afinação, dicção, pronúncia e, finalmente, timbre. Este autor considera que as manipulações do timbre são macro e micro estruturais referindo-se, as primeiras, a modificações da amplitude e frequência dos formantes, e as segundas, a mudanças dos harmónicos das suas posições esperadas. Está, pois, particularmente interessado em observar se há deslocação das frequências de ocorrência no formante do cantor concluindo que, efectivamente, este sobe quando as emoções visadas são a alegria ou a raiva, e desce, quando as emoções pretendidas são a tristeza ou o medo. Para além destas modificações espectrais macro estruturais, o autor refere que há, também, modificações micro estruturais significativas na expressão emocional da voz cantada, especialmente no caso do medo e da raiva. Nestes casos, há desvios de mais de 7% para cima das frequências dos harmónicos das suas posições ideais harmónicas relativas à frequência fundamental (Fo). O autor observa, contudo, que a natureza de tais desvios pode ser acidental.

Posteriormente, Sundberg e seus colaboradores [2, 3] fizeram também um levantamento dos diversos parâmetros afectados pelas emoções na voz cantada: tempo, duração vogal-a-vogal (desvios de ritmo), amplitude, padrões de Fo (vibrato) e frequências dos formantes. No que respeita à análise acústica chegaram aos seguintes resultados: quanto ao uso do parâmetro ‘tempo’, os resultados não mostraram grandes variações entre as versões neutras e expressivas (dezassete excertos foram executados *a capella* por um cantor, em versão neutra e versão

expressiva); quanto ao parâmetro ‘duração vogal-a-vogal’, a relação entre versões neutras e expressivas foi também de grande semelhança obedecendo a dois princípios – tornar mais lentas as notas curtas e marcação da estrutura da frase com o desenho lento-acelera-desacelera –, nas versões expressivas o alongamento das notas curtas foi maior; quanto ao parâmetro ‘volume geral de som’ também a diferença entre versões neutras e expressivas não foi significativa; no parâmetro ‘variabilidade de volume a curto termo’ todos os excertos agitados mostraram grande variabilidade por oposição aos excertos não agitados; quanto ao parâmetro ‘frequência fundamental’ nos excertos agitados a ondulação do vibrato era maior do que nas versões neutras e nos excertos não agitados; quanto ao parâmetro ‘frequência dos formantes’ os investigadores observaram que em todas as versões expressivas havia manipulação da frequência dos formantes.

Estes resultados mostram que a variabilidade do nível do som a curto termo, os padrões da frequência fundamental e as frequências dos formantes são parâmetros a que o cantor recorre para se exprimir emocionalmente. O cantor manipula a pronúncia das consoantes, provocando grandes amplitudes de modulação, manipula a extensão do vibrato e o timbre da voz para comunicar emocionalmente. Esta listagem das diversas manipulações efectuadas no intuito de resultados acústicos que sejam percebidos como veiculando determinadas emoções vem confirmar estudos anteriores [2]. A preponderância de determinados recursos acústicos sobre outros necessita, contudo, de investigação futura que corrobore ou não este estudo de caso. A limitação mais premente diz respeito, naturalmente, ao facto de ser um estudo de caso cujos resultados não podem ser generalizados.

Acusticamente falando, a cada emoção corresponde um tecido específico de propriedades sonoras cujas características são resultado dos gestos específicos do intérprete tendo por objectivo modificações subtis do som que estão para além da música escrita; uma espécie de metalinguagem musical [8].

A relação entre estes parâmetros acústicos e os seus correlatos causais, no plano fisiológico, está por sistematizar se bem que haja alguns estudos que visam averiguar algumas destas relações: o efeito das emoções sobre a respiração é de tal modo preponderante que os resultados técnicos da *performance* são afectados por esta relação [9]. Globalmente constata-se que, tendo as emoções padrões fisiológicos específicos, estes, tal como

na voz falada, perturbam quer a respiração, quer a fonte sonora, quer a articulação, quer a ressonância; ou seja, o aparelho fonador, no seu todo, é afectado, se bem que não igualmente, por todas as emoções [10, 11, 3]. Efectivamente, a ideia de a voz ser um microcosmo da fisiologia global das emoções verifica-se tanto na voz cantada, como na voz falada; contudo, a análise específica de todas estas relações acústico-fisiológicas não está efectuada para a voz cantada, não havendo, portanto, uma descrição dos diferentes mapas emocionais nessas diversas perspectivas. Rapoport (1996) sublinha, inclusivamente, que os cantores fazem uso de um verdadeiro alfabeto de gestos expressivos de modo a utilizar as cores emocionais necessárias apropriadas à *performance* de determinada obra.

Se há dominância de alguns destes factores sobre os outros também ainda não foi averiguado, se bem que a produção, a comunicação e a percepção das emoções exija um conjunto multifactorial de aspectos para ser convenientemente caracterizada. A origem desta metalinguagem não verbal, e não estritamente musical, é uma questão em aberto; Sundberg avança a hipótese de este código de expressão emocional ser importado da prosódia da voz falada mas de não se esgotar nesta importação já que, por exemplo, a macro-entoação¹ está dada pela música escrita [3, 11]. Estas questões inter cruzam a psicologia, a linguística, a biologia humana, a sociologia, a *performance* musical e a pedagogia vocal, mostrando que o corpo de conhecimento é de uma multifacetada interdisciplinaridade e transdisciplinaridade.

Estes investigadores e suas equipas tiveram, também, por objectivo, o estudo da relação emoções/timbre, mas, ambos o fizeram apenas em indivíduos singulares, como já referido: Sundberg e colaboradores no estudo de caso em que analisam a voz de Hagegård [2]; Morozov e colaboradores no estudo em que analisam a voz de Shalyapin [1]. Estes estudos sugerem que há invariâncias espectrais respectivas a determinadas emoções, mas o mapeamento destas invariâncias, e as eventuais estruturas espectrais das emoções não chega a ser levado a cabo.

Mostrar que as emoções se exprimem com mapas tímbricos diferentes, tendo, em todos os indivíduos padrões idênticos independentemente do seu tipo

vocal, do contexto musical expresso pelo compositor e do contexto linguístico em que ocorrem é a intenção deste trabalho. Procura-se, pois, o código tímbrico da meta-linguagem emocional produzida pelo cantor na *performance* musical. Ao visar analisar as eventuais manipulações do timbre efectuadas pela expressão das emoções, pretende observar-se quais as possíveis modificações espectrais do som, independentemente do timbre individual do cantor e do timbre particular de cada vogal através da qual tais modificações estão a ocorrer: se estão a ocorrer transformações “para além” das especificidades individuais tímbricas e “para além” das propriedades inerentes a cada vogal, uma vez que se considera, naturalmente, que, ao comunicar determinada emoção, o cantor não pode, nem fazê-lo fora do seu próprio timbre, nem deturpando o contexto linguístico e, portanto, tê-lo-á que fazer sem ser à custa da distinção das diferentes vogais e sem ser à custa da sua própria individualidade tímbrica.

Delineou-se o seguinte plano “experimental”:
1. Gravaram-se cantores profissionais em situação performativa emocional específica; 2. Seleccionaram-se perceptivamente os excertos musicais cuja emoção era assim corroborada; 3. Analisaram-se espectrograficamente vogais dos excertos confirmados e, finalmente, 4. Procedeu-se à análise estatística dos dados assim obtidos. De acordo com este plano testou-se se, da análise exclusiva do timbre da voz cantada, se poderiam inferir os quatro estados emocionais (Alegria, Tristeza, Raiva e Medo) visados na produção acústica dos cantores.

Metodologia

O presente estudo é um estudo transversal e prospectivo. São utilizadas cinco variáveis independentes: quatro emoções básicas, ou primárias, a saber: alegria (A), tristeza (T), raiva (R) e medo (M); e um vocalizo, que nos dá uma “emoção” neutra (N) e que serve como emoção de controlo. A variável dependente é o timbre, na sua definição acústica, visando observar as possíveis consequências das variáveis independentes na amplitude dos primeiros formantes (F1, F2, F3, F4 e F5) e na frequência a que ocorrem nas vogais [a], [i] e [u].

Pretende fazer-se uma comparação intrassujeitos e intersujeitos. Por exemplo, com o indivíduo 1: 1A; 1T; 1R; 1M e 1N observando em cada uma das situações emocionais os formantes do timbre vocálico. Mais especificamente, comparar: 1A média das vogais [a] com 1T média das vogais [a] com 1R média das

¹ Sundberg distingue macro-entoação de micro-entoação: a primeira, diz respeito à melodia e ao ritmo escritos pelo compositor e que o intérprete tem que respeitar, a segunda, diz respeito às *nuanças* interpretativas do performer (*rubato*, micro-pausas, duração das consoantes, etc), que mais não são do que desvios relativamente à composição escrita.

vogais [a] com 1M média das vogais [a] e com 1N média das vogais [a]; e depois, sucessivamente, com as vogais [i] e [u]. Esta comparação intrassujeito é feita relativamente a cada uma das emoções e com cada um dos participantes no estudo. Após esta primeira comparação analítica que visa averiguar se há uma relação entre a emoção expressa e o timbre, e quais os mapas formânticos de cada emoção estudada, passa-se a um segundo momento de comparação intersujeitos. Por exemplo, com os indivíduos 1 e 2: 1A,2A; 1T,2T; 1R,2R; 1M,2M e 1N,2N; isto entre todos os sujeitos participantes no estudo. Esta comparação permite averiguar se dada intenção emocional se repercute acusticamente na estrutura formancial de modo similar em todos os sujeitos. A haver uma resposta positiva a esta questão poderemos dizer se há ou não padrões generalizados na comunicação emocional da voz cantada.

Objectivos

O objectivo deste estudo é mostrar que os diferentes mapas espectrais das quatro emoções básicas aqui estudadas sofrem, relativamente uns aos outros, mudanças idênticas em todos os indivíduos aquando da *performance* cantada.

Pretende-se, pois, responder a quatro questões fundamentais:

1. Existe uma relação entre a expressão das emoções e o timbre da voz cantada?
2. A existir uma relação, que tipo de relação é esta?
3. Há padrões típicos nesta relação emoções/timbre?
4. Havendo padrões, estes são independentes do tipo de voz e do sexo dos indivíduos?

No decurso da investigação surgiram, a partir destas questões fundantes, outras questões secundárias, igualmente importantes:

1. A expressão tímbrica das emoções afecta igualmente todos os formantes, quer na sua frequência, quer na sua amplitude?
2. F1 e F2, por serem formantes particularmente importantes para a discriminação das vogais,

são menos manipuláveis para a expressão emocional do que F3, F4 e F5?

3. A frequência e amplitude de Fo afectam a frequência de ocorrência e a amplitude dos formantes?
4. A expressão tímbrica das emoções varia com as vogais? Há vogais mais permeáveis às emoções do que outras?
5. Há emoções cuja expressão na voz cantada é mais exequível do que outras? Se há, quais?

Amostra

A amostra é de conveniência. Não existindo em Portugal uma associação de cantores profissionais é muito difícil saber qual o tamanho e composição da população e, com esta informação, proceder a uma recolha aleatória. Assim, é feita uma amostragem de conveniência por ser a única exequível.

A amostra é constituída por cantores profissionais solistas, isto é, cuja actividade profissional performativa é executada a solo e não em conjunto como é o caso dos coralistas; isto, porque se pensa que a expressão das emoções individuais só é executada na *performance* a solo e não na *performance* coral, a qual, eventualmente, torna patente as emoções do maestro que dirige o grupo, e não dos indivíduos que a constituem.

A idade dos indivíduos está compreendida entre os vinte e nove e os trinta e nove anos. Esta faixa etária inclui cantores com alguma experiência, pois todos eles já têm alguns anos de carreira, e, simultaneamente, possibilita homogeneidade no que respeita à idade vocal³ dos indivíduos. São cantores que possuem entre três e quinze anos de experiência profissional.

Há pessoas de ambos os sexos, cinco grupos: vozes femininas agudas (Soprano 1 e Soprano 2) e graves (Mezzo 1 e Mezzo 2); e vozes masculinas agudas (Contratenor, Tenor 1, Tenor 2, Tenor 3 e Tenor 4) e graves (Barítono 1, Barítono 2 e Barítono 3). Deste modo, abrangem-se os diversos tipos vocais e pode observar-se se a expressão das emoções nos seus correlatos acústicos tímbricos é, ou não, dependente da tipologia das vozes ou do sexo dos indivíduos.

³ Idade vocal do indivíduo é aquela que resulta da sua maturação neuro-fisiológica. Há, pois, uma qualidade vocal correspondente a uma determinada idade vocal de acordo com as variações anatomofisiológicas decorrentes da idade dos indivíduos [12].

Equipamentos, instrumentos e materiais

As gravações foram efectuadas num gravador Sony TCD-D8 DAT que permite seleccionar frequências de amostragem de 48 khz e uma quantização de 16-bits. A vantagem deste gravador DAT é que as gravações por ele efectuadas são verdadeiros clones do original, isto é, não há qualquer compactação ou filtragem. O microfone utilizado foi um Beyer Dynamic M 69 TG que tem uma frequência de resposta de 50 Hz a 16 kHz, um nível de pressão de som (SPL) máximo de 120 dB e um padrão polar hypercardióidico. Por ter uma larga extensão de frequência de resposta, uma grande sensibilidade e um baixo *feed-back* é considerado um bom microfone para a gravação da voz cantada. A cassette utilizada foi uma Sony DAT (Digital Audio Tape) de 180 minutos, também ela permitindo uma gravação digital de frequência de amostragem de 48 kHz e uma quantização de 16 bits. As gravações efectuadas foram depois introduzidas no software SpeechStation2 num computador LG LW20-22DP Intel Centrino Mobile Technology, Intel Pentium Intel M Processor 740 (1.73GHz), com uma placa gráfica Intel GMA 900 Graphics e memória de 512 MB DDR2. Para a análise formântica o *software* utilizado foi o SpeechStation2 da Sensimetrics, para a estimativa de Fo e sua amplitude foi utilizado o Praat.

Foi utilizada uma folha de registo criada pela investigadora para aferir se as emoções que os cantores visavam expressar eram perceptivamente confirmadas ou não. Neste folha, os indivíduos constituídos como júri para o efeito, após ouvirem cada excerto, tinham que marcar com um 'X' na coluna da emoção (Alegria, Tristeza, Raiva ou Medo) que consideravam estar a ser expressa ou, caso considerassem que nenhuma dessas emoções estava especificamente definida, marcavam 'Indefinida'.

A amostra foi inicialmente constituída por um conjunto de doze cantores profissionais, tendo posteriormente um sido eliminado na selecção perceptiva, ficando 11 indivíduos efectivamente na amostra analisada (um dos tenores foi excluído, pois nenhuma das emoções por ele visadas nas peças executadas foi confirmada perceptivamente). Cada um deles executou quatro excertos de repertório e um vocalizo perfazendo um total de sessenta excertos. Destes, foram posteriormente seleccionados na avaliação perceptiva os quarenta e oito excertos relativos às emoções específicas que foram confirmadas perceptivamente.

As vogais escolhidas para analisar foram as vogais [a], [i] e [u]. A razão desta escolha remete para o facto de estas serem as vogais correspondentes aos três extremos de constricção do tracto oral, pelas suas posições extremas de articulação representam o triângulo das vogais (todas as outras vogais são produzidas com posições articulatorias intermediárias destas). Assim, as estruturas formânticas por elas exibidas serão paradigmáticas, isto é, se houver manipulação nestas estruturas para a expressão emocional, possivelmente também o haverá nas outras vogais que neste estudo não foram analisadas.

Procedimentos

Requisitos éticos

Os indivíduos que participaram neste estudo fizeram-no em regime de voluntariado, isto é, após terem sido contactados e questionados acerca da sua disponibilidade para participarem nesta investigação acederam sem quaisquer reservas. Foi-lhes garantido o anonimato. Tanto os cantores, após as gravações, como os estudantes de canto que integraram o júri de avaliação perceptiva das vozes, após o preenchimento do inquérito; autorizaram verbalmente a utilização dos dados para a investigação em curso.

Gravações

As gravações foram todas efectuadas no mesmo espaço físico durante duas semanas. Os cantores executaram as suas *performances a cappella*, isto é, sem qualquer acompanhamento musical. Esta situação foi artificial relativamente à situação performativa habitual de um intérprete de música erudita: normalmente a execução não é *a cappella*. Também artificial foi o facto de o executante estar simplesmente a cantar sem público. Estes dois factores foram apontados pelos indivíduos como perturbadores, mas nenhum considerou que a situação impossibilitasse a comunicação emocional. É de notar que estes músicos são artistas habituados a situações performativas que nem sempre são as ideais.

O facto de os cantores terem sido gravados sem acompanhamento foi uma exigência que teve que ser cumprida para poder garantir que não havia no espectro analisado nenhuma interferência de outro espectro dado por um ou vários outros instrumentos, isto é, foi uma exigência imposta pela análise espectrográfica que se visava realizar. Não se terem gravado os cantores em situação de concerto deveu-se, também, a uma necessidade

² Entende-se aqui "timbre da voz cantada" num sentido lato: se bem que seja sempre timbre das vogais, uma vez que é nas vogais que este timbre se está sempre a dar é, também, timbre de algo que está para além das vogais. É, pois, timbre das vogais, timbre individual e timbre "emocional". Contudo, cada uma destas instâncias tímbricas está erigida sobre a anterior, ou sobre as anteriores.

espectrográfica: era necessário que o sinal gravado tivesse a menor interferência de ruído possível; ora, em situação concertística evitar o ruído do público é quase uma impossibilidade. Contudo, preservou-se a possibilidade de gravar numa sala com alguma reverberação, pois o auto-controlo da fonação num cantor de música erudita depende, em parte, do *feed-back* por ele ouvido na sala onde se encontra.

A cada participante no estudo foi pedido que executasse quatro excertos de música previamente escolhidos do seu repertório, representativos, no seu parecer, das quatro emoções visadas e um vocalizo. A cada excerto, ou peça completa, correspondia uma das emoções pretendidas. Foi-lhe pedido que indicasse qual o extracto, dos que executou, que associava a alegria, tristeza, raiva e medo. Cada excerto incluiu, pelo menos três vogais [a], três vogais [i] e três vogais [u]. A ordem pela qual as peças foi executada foi escolhida pelo cantor. A maioria dos cantores referiu dificuldade em encontrar no seu repertório peças que expressassem medo. No entanto, todos trouxeram excertos que, de algum modo, consideravam responder aos requisitos pedidos. Também foi referida a dificuldade de encontrar extractos de música onde a expressão de cada uma das emoções fosse pura, ou seja, cada emoção parecia-lhes aparecer em determinado contexto emocional mais amplo, por exemplo: a Alegria devia ser “Alegria exultante” ou “Alegria contemplativa”? A investigadora nunca interferiu no repertório escolhido pelos intérpretes, mesmo quando instigada a fazê-lo. Assim sendo, logo na escolha do repertório a cantar houve *nuanças* diferentes nas interpretações de Alegria, Tristeza, Medo ou Raiva.

Todos os cantores foram gravados a uma distância de três metros do microfone. A sala foi fechada. Foi pedido que, se possível, não se deslocassem no espaço. Os excertos foram cantados duas vezes e duas vezes gravados. As gravações de cada indivíduo duraram entre 15 e 25 minutos.

O nível de gravação no gravador foi estável, sendo, por vezes, ajustado de acordo com o volume da voz do cantor e com a peça a executar, no sentido de garantir níveis de entrada bons e sem distorção. Os excertos foram regravados no programa em duas frequências de amostragem: 11025 Hz e 44100Hz, no sentido de posteriormente se escolher o que exibisse melhor qualidade para análise.

Aplicação da folha de registo para avaliação perceptiva

Com o intuito de potenciar a eventual expressão emocional exibida nas amostras, procedeu-se à aplicação de um inquérito de avaliação perceptiva das mesmas. Deste modo, os excertos seleccionados, não só teriam sido produzidos com uma determinada intenção emocional, como também esta intenção teria sido corroborada por um júri. Pretendeu-se, assim, salvaguardar, que os excertos analisados tinham de facto um dado conteúdo emocional, quer este fosse, ou não, visível no espectro. O júri de avaliação perceptiva foi constituído por 7 estudantes de canto de idade compreendida entre os 23 e os 29 anos, tendo entre 8 e 21 anos de estudo de música e entre 5 e 12 anos de estudo de canto. Optou-se, assim, por um painel de indivíduos que tivessem vivência da música erudita, quer ao nível da escuta, quer ao nível da execução vocal propriamente dita. Foi esta a escolha por se considerar que, embora a comunicação emocional através da música seja dirigida a todos, aqueles que estão familiarizados com este código específico, terão, eventualmente, melhor capacidade para o decifrar⁴.

Após ter gravado os excertos em DAT e estes terem sido transferidos para computador procedeu-se a uma combinação aleatória dos mesmos, fazendo-se uma sequência não previsível das emoções visadas pelos intérpretes. Foi nesta sequência que os excertos foram posteriormente submetidos à avaliação do júri. Cada membro do júri tinha que decidir, perante o inquérito atrás referido, qual a emoção que efectivamente escutava a partir da audição individual de cada excerto. Era-lhe pedido que escolhesse entre ‘tristeza’, ‘alegria’, ‘raiva’, ‘medo’ ou indefinido. O item ‘indefinido’ foi explicitado previamente pela investigadora. Cada excerto foi ouvido três vezes.

Análise e tratamentos de dados

Análise espectrográfica

A análise espectrográfica foi efectuada com o programa *SpeechStation2* como referido atrás. De cada excerto seleccionaram-se três vogais [a], três

⁴ Buekers (1998, citado por Guimarães, 2007) refere que, embora não haja uma diferença significativa entre a avaliação perceptiva de profissionais experientes e não experientes, a avaliação dos profissionais era menos variável. Compreende-se, assim, que um júri especialista poderá melhor avaliar perceptivamente.

vogais [i] e três vogais [u]. Esta selecção obedeceu aos seguintes critérios: qualidade perceptiva e visual da vogal, clareza do sinal e F_0 abaixo de 600 Hz. Preferiu-se sempre as vogais coincidentes com sílabas tónicas por se considerarem estas mais expressivas e mais pertinentes do ponto de vista linguístico e, também, por serem vogais longas (superiores a 100 ms). Dos 48 excertos seleccionados foram analisadas 432 vogais (144 vogais [a], 144 vogais [i] e 144 vogais [u]). Procedeu-se posteriormente à medição de cada vogal em três pontos: ponto médio (2/4), 1/4 e 3/4. Obtiveram-se, assim, três medições (três réplicas) de cada vogal. Em cada um destes pontos mediu-se, manualmente, os cinco primeiros formantes, F1, F2, F3, F4 e F5, retirando-se o valor de ocorrência da frequência e o valor da amplitude. Esta medição manual fez-se observando, também, a estimativa dos formantes dada pelo programa (análise pico a pico no espectro de predição linear, LPC). Contudo, por se considerar que esta estimativa não é fiável, a medição final não foi com base nesta estimativa. Pretendeu-se, sobretudo, definir os formantes por um lado, procurando no espectrograma frequência/tempo as linhas horizontais (harmónicos) de maior amplitude visualizadas pela maior intensidade luminosa; por outro, procurando a coincidência com o pico formântico no espectro amplitude/frequência oferecido pela funcionalidade *Spectrum viewer* do programa *Speech Station2*.

O espectro frequência/tempo visualizado foi o baseado na transformada rápida de Fourier (FFT). A frequência de amostragem escolhida aquando da gravação para o programa foi de 11025 Hz (a visualização efectiva exibida foi de 0 a 5500 Hz), por se considerar que a qualidade do espectro assim apresentada era a melhor para a análise espectrográfica. O espectro assim visualizado tem pouca compressão o que permite uma melhor acuidade na medição. A janela de análise utilizada foi de tipo Hanning, de tamanho 128.

Os formantes que sistematicamente foram mais difíceis de medir foram F1 e F5: o primeiro, porque muitas vezes é difícil decidir se o pico ressonancial se encontra no primeiro harmónico ou no segundo; o segundo, porque, por vezes, tem muito pouca intensidade, especialmente na vogal [a]⁵.

⁵ Diferentemente do que acontece na fala, a vogal [a] não é uma vogal “forte” no canto, pois o facto de, do ponto de vista funcional, haver necessidade de converter a pressão alta na faringe típica da vogal [a] em pressão na região palatal anterior, globalmente procurada para todas as vogais cantadas, torna a aquisição tímbrica do [a] cantado uma aquisição difícil

Também foram mais difíceis de mensurar os espectros das mulheres, por ser mais difícil distinguir harmónicos de picos de ressonância, isto é, a ênfase, do ponto de vista espectral, de quase todos os harmónicos torna complicado decidir, com verdadeira objectividade, quais os harmónicos coincidentes com os formantes. Quanto mais alta a frequência fundamental cantada mais complexa esta decisão. Na tentativa de, de algum modo, contornar parte desta dificuldade, procuraram-se extrair vogais com frequência fundamental mais baixa (nunca se excedeu os 600 Hz). Assim, evitou-se sempre valores iguais ou superiores à frequência fundamental crítica que é de 880Hz, pois, neste caso, a leitura formântica não seria viável.

A medição, sempre dupla, da amplitude formancial e da frequência a que ocorrem os formantes deveu-se ao facto de considerar que a fidedignidade do valor encontrado para o pico ressonancial era tanto maior, quanto maior o valor da amplitude exibida, ou seja, como a determinação do formante é feita através da amplitude observada em determinada frequência, o valor da amplitude era relevante. Por outro lado, a manipulação das amplitudes dos formantes poderia concorrer também para a expressão das emoções no timbre da voz cantada, objecto do nosso estudo.

Para cada vogal foi ainda encontrada a frequência fundamental média e sua amplitude com o programa PRAAT. Escolheu-se este programa para encontrar a estimativa da F_0 por ser este programa considerado um dos mais fiáveis na leitura da frequência fundamental. A razão pela qual se retirou o valor médio da frequência fundamental de cada vogal analisada remete ainda para o facto de garantir que, eventuais manipulações de F1 nas vozes agudas femininas (a existirem na amostra estudada), por fazerem parte das estratégias utilizadas por estas vozes para a projecção do som, não faziam parte das estratégias usadas para a expressão das emoções. Por outro lado, visava-se também saber se o valor de F_0 afectava a expressão emocional nos formantes ou não.

A razão pela qual as vogais escolhidas para analisar foram as vogais [a], [i] e [u] remete para o facto de estas serem as vogais correspondentes aos três extremos de constricção do tracto oral, pelas suas posições extremas de articulação representam o triângulo das vogais (todas as outras vogais são produzidas com

e morosa. A vogal [a] é sentida por todos os cantores como a vogal mais difícil de “timbrar”. Assim se entende, que o espectro exibido nas vogais [a] cantadas, nem sempre seja facilmente legível.

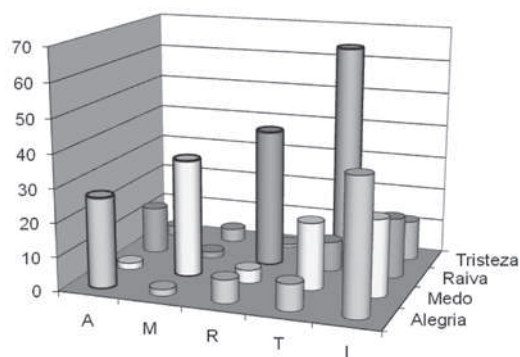
posições articulatórias intermédias destas). Assim, as estruturas formânticas por elas exibidas poderão ser paradigmáticas, isto é, se houver manipulação nestas estruturas para a expressão emocional, possivelmente também o haverá nas outras vogais que neste estudo não foram analisadas.

Estatística utilizada

Neste trabalho considerou-se que H0 se rejeita a partir de um valor de significância de $p < 0,01$. No presente trabalho, várias análises permitem descrever (análise descritivas: normalidade, variância, tabelas, médias, correlações e distâncias), explorar (análise de componentes principais) e testar hipóteses (análise de variância paramétrica, análise de variância não paramétrica). Os programas utilizados foram SPSS 15 (ANOVA, KRUSKALL-WALLIS), Minitab (ANOVA nested, testes de normalidade e de homogeneidade de variância) e Canoco 4.5 (Análise de componentes principais – ACP).

Resultados

Seleção perceptiva das amostras



| | A | M | R | T | I |
|----------|----|----|----|----|----|
| Alegria | 27 | 2 | 7 | 8 | 40 |
| Medo | 2 | 35 | 4 | 20 | 23 |
| Raiva | 14 | 2 | 41 | 9 | 18 |
| Tristeza | 3 | 4 | 1 | 64 | 12 |

Figura 1 – Emoções visadas pelos cantores versus emoções apreendidas perceptivamente

Nesta análise preliminar apresentam-se, na Figura 1, os resultados relativos à avaliação perceptiva. Após ter as respostas do júri, fez-se um levantamento das percentagens de excertos cuja emoção foi confirmada ou infirmada. Foi considerado confirmado aquele excerto cuja corroboração emocional perceptiva foi igual, ou superior, a 57,1%, isto é, confirmado por

quatro ou mais indivíduos. Foi considerado infirmado aquele excerto cuja corroboração emocional perceptiva foi igual, ou inferior, a 42,8%, isto é, apenas confirmado por três ou menos indivíduos.

Tendo em conta que o júri de avaliação perceptiva foi constituído por sete elementos, cada emoção, a ser corroborada em absoluto, nos doze excertos dos cantores, sê-lo-ia 84 vezes. Os resultados dos inquéritos expressos pelo gráfico da Figura 1 mostram que a emoção percebida com maior sucesso foi a Tristeza (embora apenas 64 vezes confirmada) e que a emoção mais difícil de ser corroborada perceptivamente foi a Alegria (apenas 27 vezes).

No sentido de averiguar as razões pelas quais determinadas emoções visadas tinham tido menos sucesso do que outras, pretendeu-se tornar explícito de que modo se efectivava a incompreensão: se simplesmente o ouvinte não conseguira definir a emoção que ouvira, ou se a tinha trocado por outra, e neste caso, por qual ou quais. A Figura 1 torna patentes algumas confusões curiosas entre emoção visada/emoção apreendida: um número significativo de excertos que visavam exprimir Medo foram ouvidos como Tristeza (20 vezes), e um número importante de excertos que pretendiam expressar Raiva foram apreendidos como Alegria (14 vezes). A emoção cuja expressão foi mais difícil de ser reconhecida foi a Alegria, que, por 40 vezes foi considerada como indefinida.

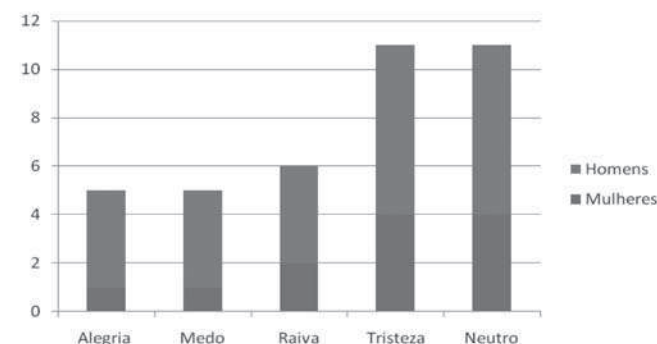


Figura 2 – Número de excertos analisados por emoção e género

Na Figura 2 mostra-se o número de excertos que vieram a ser efectivamente analisados. Neste gráfico pode ainda observar-se que, apesar de um homem ter sido excluído (Tenor 4), pelo facto de nenhum dos seus excertos ter sido corroborado do ponto de vista emocional, ainda assim, o peso dos homens na amostra continua a ser maior (tal como acontecia na amostra inicial – 8 homens e 4 mulheres). Observe-se, ainda, que os excertos

relativos a Tristeza e Neutro foram analisados em onze indivíduos enquanto que os relativos a Alegria e Medo ficaram substancialmente reduzidos.

Validação das premissas estatísticas

Normalidade

Para cada formante, de cada vogal, o teste de normalidade de Ruan-Joiner (Rj), que varia entre 0 e 1, (sendo mais próximo de 1 quando a distribuição se aproxima da normal) deu resultados significativos para os cinco formantes. [13]

Homogeneidade de variância

Das 15 combinações possíveis entre Factores e Formantes, apenas duas não são significativas para o Teste de Bartlett χ^2 [13], Vogais F1 e Vogais F4. No entanto, como foi realizada uma análise não paramétrica livre das restrições da ANOVA, poder-se-á verificar, à posteriori, se estas duas exceções afectaram ou não os resultados.

Feitos os testes de homogeneidade de variância [13], determinou-se que os dados cumpriam, globalmente, os requisitos exigidos pela ANOVA. Os requisitos da ANOVA permitem a utilização da média em tabelas, gráficos, como *input* da análise multivariada. Por isso, considerar a média como um bom estimador da população para cada factor (distribuições análogas e variância essencialmente homogénea) é fundamental para a validade de toda a análise estatística que a utiliza.

Correlações e distâncias

As correlações observadas [13] justificam a utilização da ANCOVA, de forma a eliminar o efeito da Fo nos restantes formantes.

Primeira instância do timbre: vogais

A Figura 3, que apresenta os diferentes gráficos de bigodes relativos às médias de cada formante, para cada uma das vogais, mostra que a variação mais significativa na diferenciação das vogais se dá ao nível dos formantes F1 e F2. F1 é mais alto em média de frequência de ocorrência para a vogal [a], mais baixo em [i] e ligeiramente mais alto do que [i], em [u]. F2 está próximo de F1 na vogal [a], está afastado de F1 na vogal [i] e de novo próximo de F1 na vogal [u], se bem que mais baixo do que em [a]. No que respeita aos formantes F3, F4 e F5

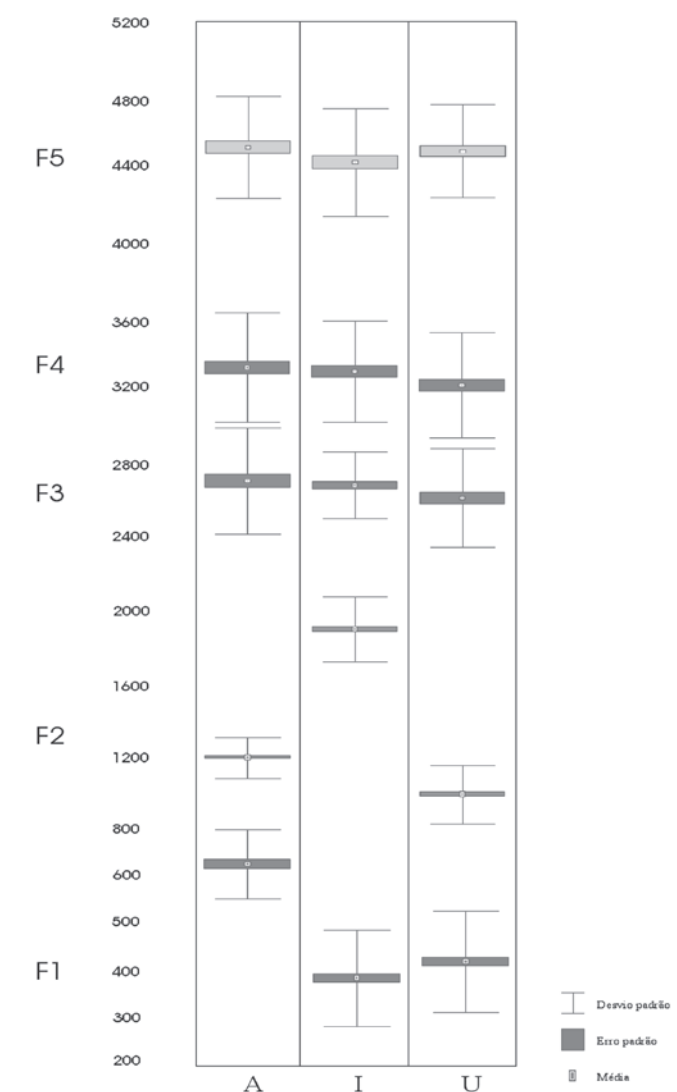


Figura 3 – Médias, erro e desvio padrão para as três vogais relativamente aos cinco formantes

a sua variação nas diferentes vogais não é muito manifesta, se bem que F3 e F4 em [u], se encontrem ligeiramente mais baixos do que em [a] e [i]; e F5 se encontre ligeiramente mais baixo em [i], do que em [a] e [u]. Assim, F3 e F4 discriminam [a] e [i] de [u], enquanto que F5 separa o [i] das restantes vogais. É ainda de salientar que os padrões das médias das três vogais são completamente diferentes para os formantes: F1 e F2 têm comportamentos opostos ([i] tem o valor mínimo para F1 e máximo para F2); F3 e F4 têm comportamentos similares e F5 é semelhante a F1 (mais alto em [a], mais baixo em [i] e um pouco mais alto que [i] em [u]), mas a magnitude das diferenças é muito menor. Finalmente, o erro padrão reduzido, e a diferença entre as médias observados, evidencia que a resposta dos formantes às vogais é clara, com erros muito inferiores às

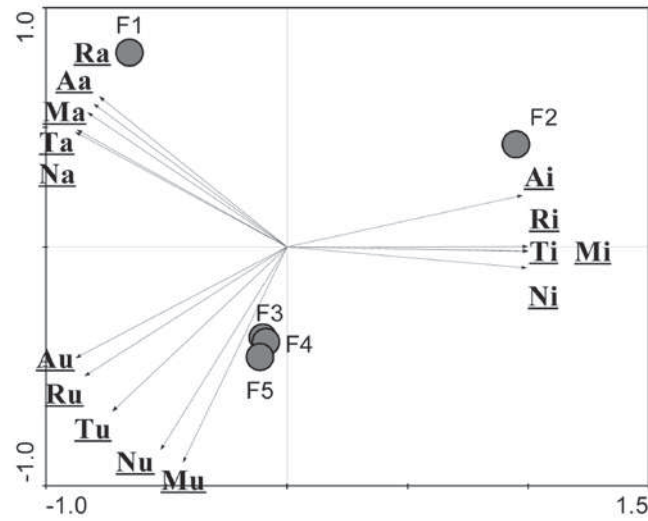


Figura 4 – ACP para Vogais versus Emoções (todos os cantores)

diferenças das médias, sendo F2 e F1 os formantes onde este fenómeno é mais evidente.

Na Análise de Componentes Principais (ACP) – Figura 4 – o 1º eixo explica entre 62% a 81% da variância da amostra, e o 2º eixo explica entre 17% e 36% da variância da amostra. Nesta figura pode observar-se que as três vogais são muito bem discriminadas em qualquer um dos casos analisados (com algumas excepções), com o F2 a separar o [i] do [a] e do [u] e com o F1 a separar o [a] do [i] e do [u]. Este resultado confirma o gráfico de bigodes para as vogais (Figura 3), onde as médias das vogais para cada formante davam exactamente o mesmo padrão. É por estas razões evidente que bastam dois formantes para discriminar com clareza as três vogais, sendo a informação retida nos F3, F4 e F5 pouco significativa para as vogais. É possível observar que a separação das vogais é muito clara, com muito pouca dispersão de emoções, sendo esta menor para [a], intermédia para [i] e maior para [u]; este resultado pode indicar que existem capacidades diferentes de expressar emoções de acordo com as vogais cantadas, sendo a vogal [u] mais expressiva pois é aquela na qual as emoções se apresentam mais distanciadas umas das outras, estando desse modo mais diferenciadas e melhor definidas. Ainda neste gráfico (Figura 4), pode observar-se que Raiva e Alegria estão sempre lado a lado, e que Alegria é sempre a emoção mais afastada.

No Quadro 1 pode confirmar-se o resultado da análise exploratória (Figura 3), que nos indicava que era o F2 o formante mais afectado pelas 3 vogais. O valor de 3307 para o F da ANOVA é muito grande (o valor de F é muito pequeno quando

Quadro 1 – Tabela ANOVA e MANOVA para as vogais

| ANOVA (DF=2) | F | p-level |
|--------------|---------|---------|
| F1 | 511 | 0 |
| F2 | 3307 | 0 |
| F3 | 13 | 3,1 E-6 |
| F4 | 11 | 0,00002 |
| F5 | 8 | 0,00032 |
| MANOVA | Rao's R | p-level |
| Todos F | 651 | 0 |

H0 se verifica, ou seja, quando não há diferenças significativas entre as médias), indicando que as diferenças observadas no gráfico de bigodes (Figura 3) são, sem qualquer dúvida, significativas, e, como se observou naquele gráfico, discrimina [i] de [a] e [u]. F1 vem logo a seguir em termos de importância, com um valor de F de 511. Mais uma vez, este resultado confirma o gráfico de bigodes, onde F1 era tido como muito importante para a discriminação das vogais. F3 e F4 têm valores muito inferiores de F, apesar de continuarem significativos. F5 é o formante menos importante na discriminação das vogais. Quando se consideram todos o formantes, o valor de F é muito elevado, contribuindo para isso as diferenças observadas para F1 e F2. Todas as análises são significativas, sendo o valor de F (ANOVA) ou de Rao (MANOVA) proporcional à robustez do resultado.

Segunda instância do timbre: individualidade tímbrica dos cantores

A Figura 5 mostra as médias de frequência de cada um dos cinco formantes nos onze cantores discriminados através do gráfico de bigodes. A proximidade nos valores apresentados em F1 e F2 é muito maior do que a que ocorre em F3, F4 e F5. F1 e F2 apresentam padrões análogos enquanto que F3 e F4 formam outro par. F5 foge à lógica geral. Para F1 e F2 é visível que as diferenças observadas são pequenas, e que o erro (caixa em volta da média) é por vezes superior à diferença entre as médias, sendo este efeito mais importante para F2. F3 e F4 são, sem dúvida, os formantes onde as médias apresentam maiores diferenças, correspondendo provavelmente às características tímbricas dos cantores. F5 é menos variante que F3 e F4, mas mais que F1 e F2. Há, contudo, diversidade nas diferenças encontradas entre cantores podem também dever-se

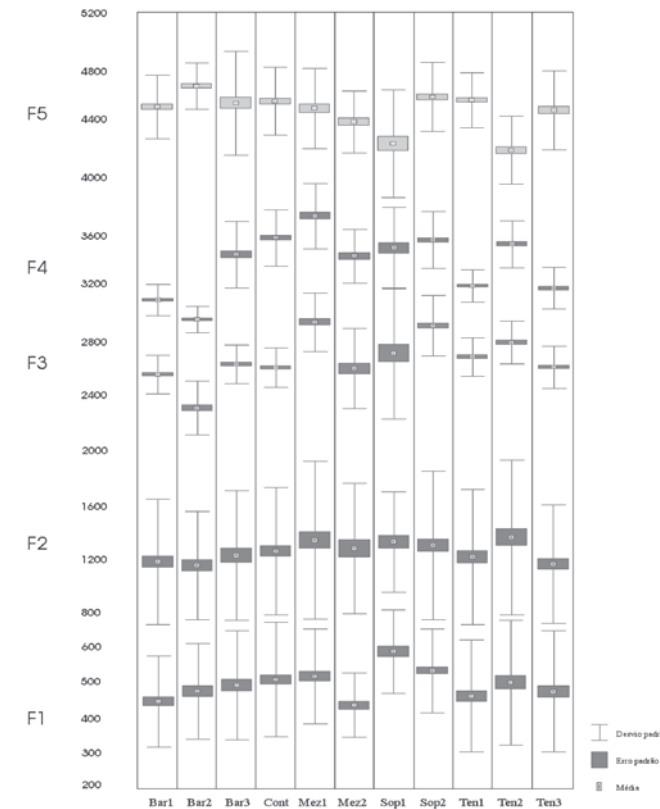


Figura 5 – Médias, erro e desvio padrão dos cinco formantes nos onze cantores

ao facto de não se ter sempre o mesmo número de emoções a ser consideradas. Por exemplo, no Mezzo 2 (com duas emoções) e no Tenor 2 (com 3 emoções). Aparentemente, não existe um padrão comum ao mesmo naipe de vozes, mas esse resultado pode estar enviesado pelo número de emoções que entram na média.

Na Figura 6 o primeiro eixo explica 63% de variância, o segundo eixo 22% e o 3º eixo 13%. Da análise desta ACP constata-se que a discriminação dos cantores se efectua pelos formantes F5, F4 e

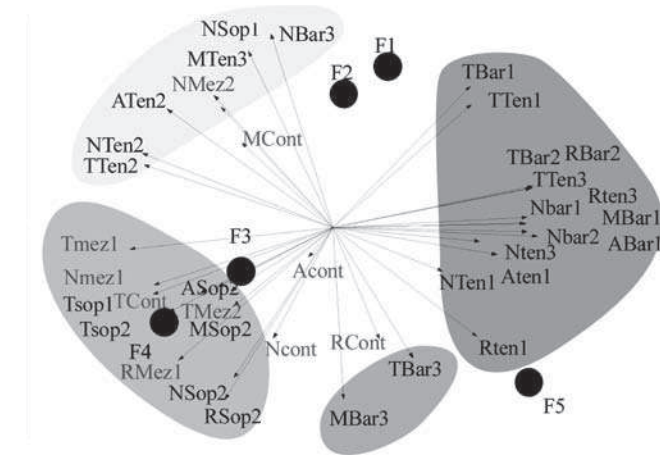


Figura 6 – ACP para Cantores versus Emoções

F3 (valores mais distantes da origem relativamente ao 1º eixo) e, finalmente, F1 e F2 (sendo F2 aquele que menos contribui).

O primeiro eixo divide claramente as características tímbricas da voz, com a maioria dos homens agrupados no lado direito do gráfico e as mulheres agrupadas no lado esquerdo do gráfico. O eixo dois está relacionado com outras características da voz e, possivelmente, com os próprios trechos musicais, sendo difícil visualizar qualquer padrão relativamente às emoções. Muitas emoções estão agrupadas para o mesmo cantor, evidenciando a supremacia das características tímbricas da voz em detrimento da emoção expressa. No entanto, nesta análise, esta derivação está provavelmente mais associada ao excerto que cantaram, já que as emoções se encontram dispersas. Analisando apenas o primeiro eixo (Figura 6), observa-se a existência de uma relação de F3 e F4 com as características tímbricas das vozes femininas e de F5 com as vozes masculinas.

Quadro 2 – Tabela ANOVA e MANOVA para cantores

| ANOVA (DF=10) | F | p-level |
|---------------|---------|---------|
| F1 | 6,5 | 0 |
| F2 | 1,9 | 0,037 |
| F3 | 65,9 | 0 |
| F4 | 138 | 0 |
| F5 | 22 | 0 |
| MANOVA | Rao's R | p-level |
| Todos F | 37,4 | 0 |

Na tabela do Quadro 2 vemos que, efectivamente, é o formante F4, seguido de F3, que mais contribui para a discriminação dos cantores, confirmando os resultados das análises anteriores (Figura 5 e Figura 6), que já evidenciavam este traço. O F2, muito importante na discriminação de vogais, deixa de ser significativo (p = 0,03) reforçando que os formantes têm uma função bastante especializada relativamente a cada factor (vogais, cantores ou emoções). O F1 tem um valor de F um pouco maior. F5 encontra-se entre F1 e F2 por um lado, e F3 e F4, por outro. Estes resultados, confirmam na íntegra a informação observada no gráfico de bigodes (Figura 5).

Terceira instância do timbre: emoções

Na Figura 7, pela análise das médias nos gráficos de bigodes, pode observar-se um padrão consistente em todos formantes relativamente às emoções. O Neutro é aquele que apresenta valores médios

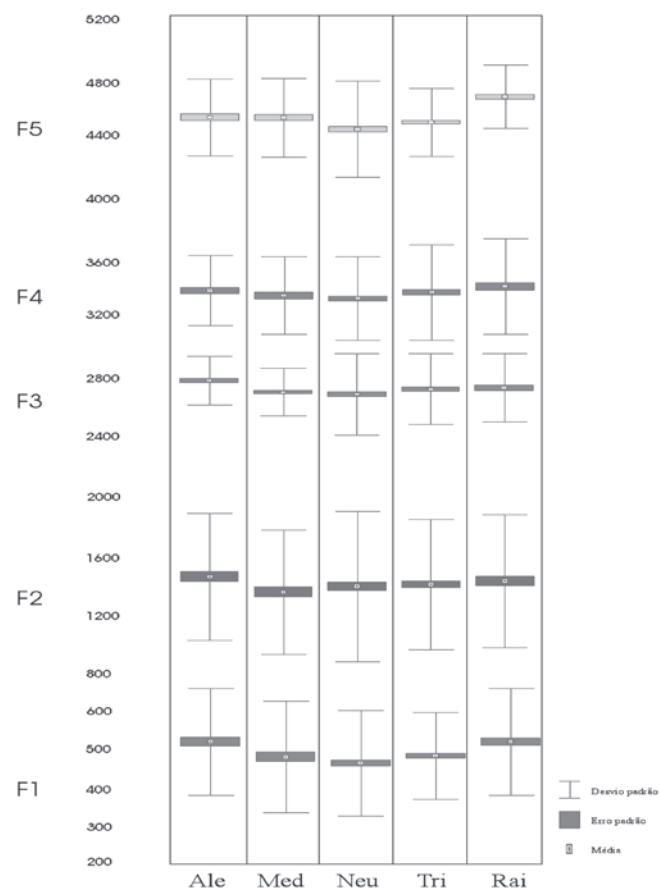


Figura 7 – Médias, erro e desvio padrão das emoções nos cinco formantes

mais baixos na frequência de todos os formantes (excepto em F2); Medo e Tristeza apresentam médias de valores próximos e Alegria e Raiva possuem também valores próximos, excepto em F5, em que, claramente Raiva apresenta as médias de valores mais altas de todos. É já visível que cada emoção aponta para um desenho distinto de frequências de formantes.

O mais interessante de observar na Figura 7 é que as emoções, ao contrário dos factores precedentes, apresentam o mesmo padrão para todos os formantes (as médias das frequências decrescem de Alegria para Neutro, e voltam a crescer para a Raiva), sendo F2 o que tem menos variância, e F5 o que tem maior variância nas descidas e subidas de frequência. Pelo contrário, as diferenças observadas são sempre muito pequenas relativamente aos cantores ou às vogais.

De seguida procedeu-se à decomposição das emoções pelas três vogais de modo a confirmar o padrão observado no gráfico de bigodes (consultar os resultados em [13]). Verificou-se, ainda, se havia diferenças entre as relações das médias da frequência

dos formantes (para cada uma das vogais e para as cinco emoções) entre os cantores do sexo masculino e os cantores do sexo feminino (consultar resultados em [13]) observando-se, por exemplo, que os cantores de sexo feminino, diferentemente do que acontece nos cantores do sexo masculino, dissociam Neutro de Tristeza.

Observando o gráfico da ACP para as vogais (Figuras 4), pôde ver-se que a discriminação das vogais é feita pelo primeiro eixo e F2 (separando as vogais [i] das vogais [a] e [u]) e pelo 2º eixo e F1 (separando o [a] do [u]). Por isso, covariando com F1 e F2, poderemos ter os efeitos tímbricos das emoções quase puros. As análises seguintes são covariações de F3, F4 e F5 com F1 e F2, retirando assim o efeito das vogais da análise.

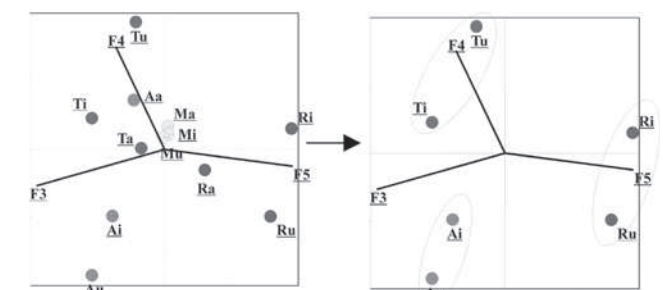


Figura 8 – Análises de componentes principais (ACP) para emoções (sem vogais): todos

O primeiro resultado que ressalta da Figura 8 (ACP sem vogais – todos) é que, se eliminarmos a região menos significativa do gráfico (região central), as emoções ficam razoavelmente agrupadas, separando-se claramente umas das outras. Lembrando o gráfico da Vogais versus Emoções (Figura 4), onde as vogais eram claramente determinantes no padrão observado, vemos que este resultado é bastante surpreendente, evidenciando que mesmo que as diferenças entre as médias seja pequeno (gráfico de bigodes das emoções – Figura 7), existem, no entanto, padrões que emergem quando separamos as emoções dos restantes factores.

Outro resultado interessante é que, sistematicamente, são as vogais [i] e [u] que estão significativamente distantes do centro, indicando que são as melhores vogais no que diz respeito à discriminação das emoções. F5 está claramente relacionado com a Raiva, o F4 com a Tristeza e o F3 ligeiramente com a Alegria.

Para as 5 emoções, foram feitas análises de variância para cada um dos formantes (ANOVA) e todos simultaneamente (MANOVA) para todas as vogais (coluna 1 do Quadro 3). Nas análises seguintes –

Quadro 3 – Tabela de análise de variância para as emoções, para cada um dos formantes (ANOVA), covariada com Fo (ANCOVA), covariada com Fo para cada uma das vogais (ANCOVA [a], ANCOVA [i] e ANCOVA [u]) e para todos os formantes (MANOVA), covariados com Fo (MANCOVA, MANCOVA [a], MANCOVA [i] e MANCOVA [u])

| DF= 4 | ANOVA | | ANCOVA | | ANCOVA [a] | | ANCOVA [i] | | ANCOVA [u] | |
|-------|---------|---------|---------|---------|-------------|---------|-------------|---------|-------------|---------|
| | F | p-level | F | p-level | F | p-level | F | p-level | F | p-level |
| f1 | 6,94 | 1,7 E-5 | 4,11 | 0,003 | 6,94 | 2,2E-05 | 2,84 | 0,0245 | 6,74 | 3,1E-05 |
| f2 | 1,06 | 0,38 | 0,89 | 0,47 | 4,81 | 0,0009 | 7,7 | 6E-06 | 5,88 | 0,0001 |
| f3 | 4,04 | 0,003 | 3,41 | 0,009 | 1,85 | 0,118 | 4,86 | 0,0008 | 1,46 | 0,213 |
| f4 | 2,31 | 0,056 | 1,58 | 0,177 | 4,08 | 0,003 | 2,56 | 0,039 | 1,29 | 0,275 |
| f5 | 18,5 | 1,1E-14 | 18,3 | 1,6E-14 | 3,3 | 0,0107 | 15,1 | 2,2E-11 | 6,4 | 5,9E-05 |
| | MANOVA | | MANCOVA | | MANCOVA [a] | | MANCOVA [i] | | MANCOVA [u] | |
| Todos | Rao's R | p-level | Rao's R | p-level | Rao's R | p-level | Rao's R | p-level | Rao's R | p-level |
| F | 6,3 | 6,5E-17 | 6,4 | 1,8E-17 | 4,2 | 2,1E-9 | 5,9 | 5,4E-15 | 4,07 | 5,3E-9 |

2ª a 5ª colunas – Fo entrou sempre como covariável, para eliminar o efeito da frequência fundamental da análise das emoções. Na 3ª, 4ª e 5ª colunas analisaram-se as vogais isoladamente, para investigar a importância de cada vogal na discriminação das emoções.

Na primeira coluna do Quadro 3 é evidente que o F5 é o formante que melhor discrimina as emoções, sendo o valor de p muito significativo (p = 1,1E-14). Seguidamente, com grau de significância menor, F1 e F3. F2 e F4 não são significativos (com p = 0,3 e p = 0,05 respectivamente). Este resultado confirma o gráfico de bigodes (Figura 7), onde F5 evidenciava maior variância. Na segunda coluna do Quadro 3 pode ainda verificar-se que os resultados são idênticos aos da primeira coluna, com F5 mais significativo seguido de F1 e F3. Da terceira à quinta colunas (ANCOVA [a],[i] e [u], respectivamente) pode ainda observar-se que a vogal [i] é, efectivamente, aquela que melhor discrimina as emoções em todos os formantes exceptuando F4 (com p = 0,03). Relativamente a F2, a vogal [i] é a única capaz de discriminar emoções. As vogais [a] e [u] não são significativas para as emoções nos formantes F3[a], F3[u] e F4 [u], respectivamente.

Ao utilizar todos os formantes na MANOVA e MANCOVA (Quadro 3), os resultados ficam altamente significativos, o que indica que a utilização do complexo de todos os formantes permite uma descrição mais clara da expressão das emoções no timbre da voz cantada. Em termos gerais, F2 e F1 discriminam vogais, F5 e F1 emoções e F4 e F3 cantores.

No intuito de averiguar se haveria padrões específicos das emoções, construíram-se, a título exemplificativo (uma vez que a amostra é heterogénea no que respeita às emoções executadas pelos cantores),

os gráficos relativos à visualização da emoção Raiva (consultar os resultados em [13]).

Comparações das três instâncias

Comparações Vogais/Pessoas/Emoções

Os valores de significância na análise paramétrica, dado a artificialidade de considerar que a amostra tem mais de 1000 graus de liberdade, são muito mais pequenos, isto é, são muito mais significativos. A análise oferecida pelo teste Kruskal-Wallis χ^2 (Quadro 4), basicamente, confirma os valores obtidos da ANOVA para os três factores estudados (Quadros 1, 2 e 3), mas os valores bastante mais baixos de p permitem uma comparação efectiva dos três factores.

Quadro 4 – Teste do Kruskal-Wallis χ^2 : Vogais/ Pessoas/ Emoções e Réplicas

| | Vogais (DF=2) | | Pessoas (DF=10) | | Emoções (DF=4) | | Réplica (DF=2) | |
|----|---------------|---------|-----------------|---------|----------------|---------|----------------|---------|
| | χ^2 | p-level | χ^2 | p-level | χ^2 | p-level | χ^2 | p-level |
| f1 | 505 | 0 | 69 | 0 | 31 | 0 | 0,042 | 0,98 |
| f2 | 808 | 0 | 22 | 0,015 | 10 | 0,040 | 0,056 | 0,97 |
| f3 | 29 | 0 | 422 | 0 | 15 | 0,005 | 0,29 | 0,87 |
| f4 | 21 | 0 | 625 | 0 | 7,6 | 0,108 | 0,08 | 0,96 |
| f5 | 17 | 0,0002 | 160 | 0 | 64 | 0 | 0,58 | 0,75 |

No Quadro 4 e na Figura 9 observa-se que, para as vogais, os formantes mais significativos são o F2 e F1 (com valores do χ^2 na ordem dos 808 e 505, respectivamente); para os cantores, apenas F2 é não significativo sendo F4 e F3 os mais significativos (com valores do χ^2 na ordem dos 625 e 422, respectivamente), e, para as emoções, F5 é muito significativo (com valor de χ^2 de 64), F1 menos

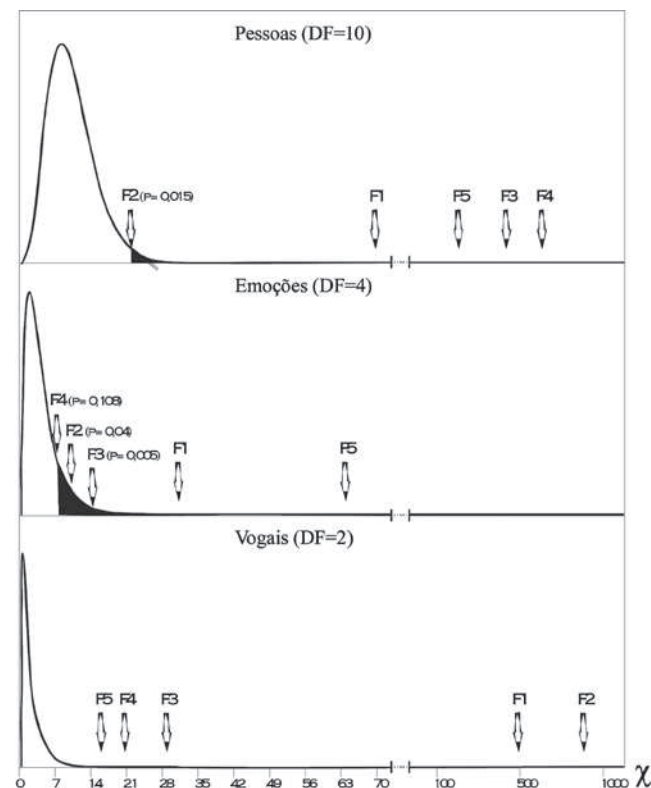


Figura 9 – Valores do Kruskal-Wallis χ^2 dos cinco formantes para as pessoas, para as emoções e para as vogais, respectivamente

(com valor de χ^2 de 31), F3 pouco significativo (com valor de χ^2 de 15) e F2 e F4 não são significativos. Na Figura 9 note-se que a probabilidade de H0 se verificar é proporcional à área assinalada pelo que, quanto mais longe desta área o formante se encontrar, mais significativo ele é no que concerne o factor mencionado.

Relativamente à quantidade de variância explicada por cada componente, a análise paramétrica de variância tem duas medidas proporcionais a este valor, o F para a ANOVA e o R para a MANOVA. No

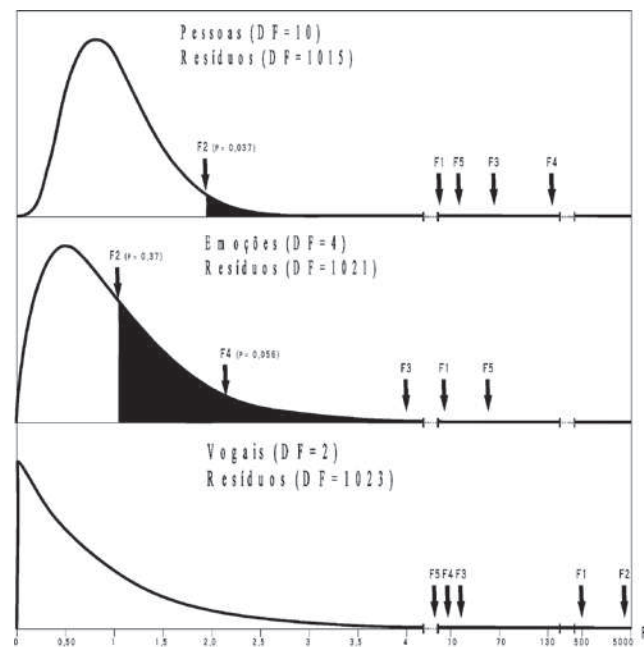


Figura 10 – Valores da ANOVA F relativamente aos cinco formantes para as pessoas, para as emoções e para as vogais, respectivamente

entanto, podemos comparar o F (Anova) ou o Rao (MANOVA) entre análises, já que a amostra é sempre a mesma. O Quadro 5 resume estes resultados.

Pela análise do Quadro 5 e da Figura 10 pode observar-se que para as vogais os formantes F2 e F1 são os mais significativos com valores de F muitíssimo altos (3307 e 511, respectivamente); para a discriminação tímbrica das pessoas os formantes F4 seguido de F3 são os mais significativos (com valores de F de 138 e 66, respectivamente) sendo apenas F2 não significativo (com $p = 0,037$); para as emoções os formantes F5 e F1 são os mais significativos (com valores de F de 18,5 e 6,9, respectivamente), não sendo significativos os formantes F2 e F4 (com

Quadro 5 – Tabela ANOVA e MANOVA comparativa para quantidade de variância entre Vogal/Emoção/Cantor/Réplica

| | Vogais (DF=2) Resíduos (DF=1023) | | Pessoas (DF=10) Resíduos (DF=1015) | | Emoções (DF=4) Resíduos (DF=1021) | | Réplica (DF=1) Resíduos (DF=1024) | |
|--------------------|-------------------------------------|---------|---------------------------------------|---------|--------------------------------------|----------|--------------------------------------|---------|
| | F | p-level | F | p-level | F | p-level | F | p-level |
| f1 | 511 | 0 | 6,45 | 1,09e-9 | 6,93 | 0,000017 | 0,0004 | 0,98 |
| f2 | 3307 | 0 | 1,93 | 0,037 | 1,1 | 0,37 | 0,009 | 0,92 |
| f3 | 12,8 | 3,1E-6 | 66 | 0 | 4 | 0,003 | 0,187 | 0,67 |
| f4 | 10,8 | 0,00002 | 138 | 0 | 2,3 | 0,056 | 0,19 | 0,66 |
| f5 | 8 | 0,00032 | 22 | 0 | 18,5 | 1,05E-14 | 0,26 | 0,55 |
| Manova | Rao's R | p-level | Rao's R | p-level | Rao's R | p-level | Rao's R | p-level |
| Todos os formantes | 630,8 | 0 | 37,4 | 0 | 6,27 | 0 | 0,18 | 0,997 |

p-level de 0,37 e 0,056, respectivamente). Note-se, mais uma vez, que a probabilidade de H0 se verificar é proporcional à área assinalada (Figura 10) pelo que, quanto mais longe desta área o formante se encontrar, mais significativo ele é relativamente ao factor mencionado.

Pode concluir-se, deste modo, que os resultados obtidos pela análise paramétrica (Quadro 5/ Figura 10) e pela análise não paramétrica (Quadro 4/Figura 9) são idênticos.

De seguida analisou-se a importância das amplitudes dos formantes na discriminação entre vogais, cantores e emoções. A tabela do Quadro 6 resume estes resultados.

Quadro 6: Tabela ANOVA para amplitude dos formantes em Vogais, Emoções, Cantores e Réplicas

| Formantes | Vogal | Emoção | Cantores | Réplica |
|-----------|-------|--------|----------|---------|
| F1 | 34,3 | 32,6 | 45,5 | 0,16 |
| F2 | 65,4 | 43,1 | 30,6 | 1,54 |
| F3 | 84,4 | 38,7 | 35,6 | 3,1 |
| F4 | 56,8 | 34,2 | 57,2 | 1,8 |
| F5 | 23,1 | 16,4 | 16,4 | 2,14 |
| Todos | 39,7 | 13,4 | 29,7 | 1,21 |

O resultado mais interessante é que, neste caso, apesar da ordem de importância ser a mesma (Vogais

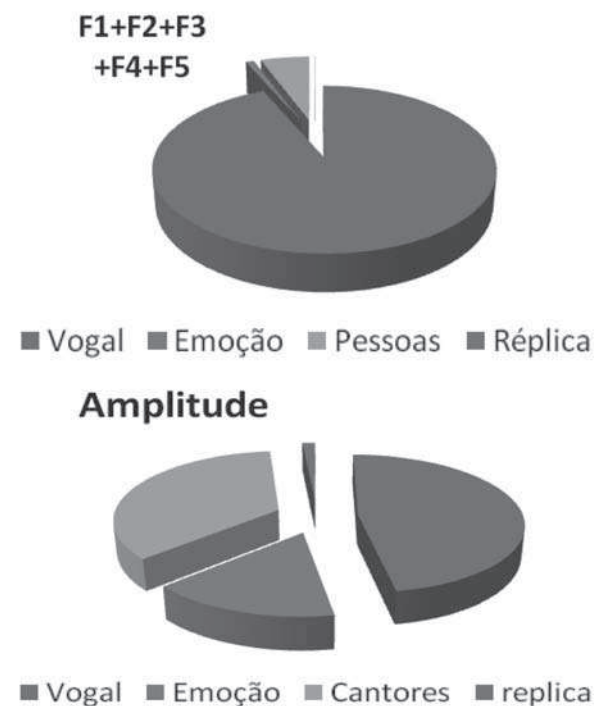


Figura 11 – Gráficos de queijos mostrando a partição da variância na frequência dos formantes e na amplitude dos mesmos

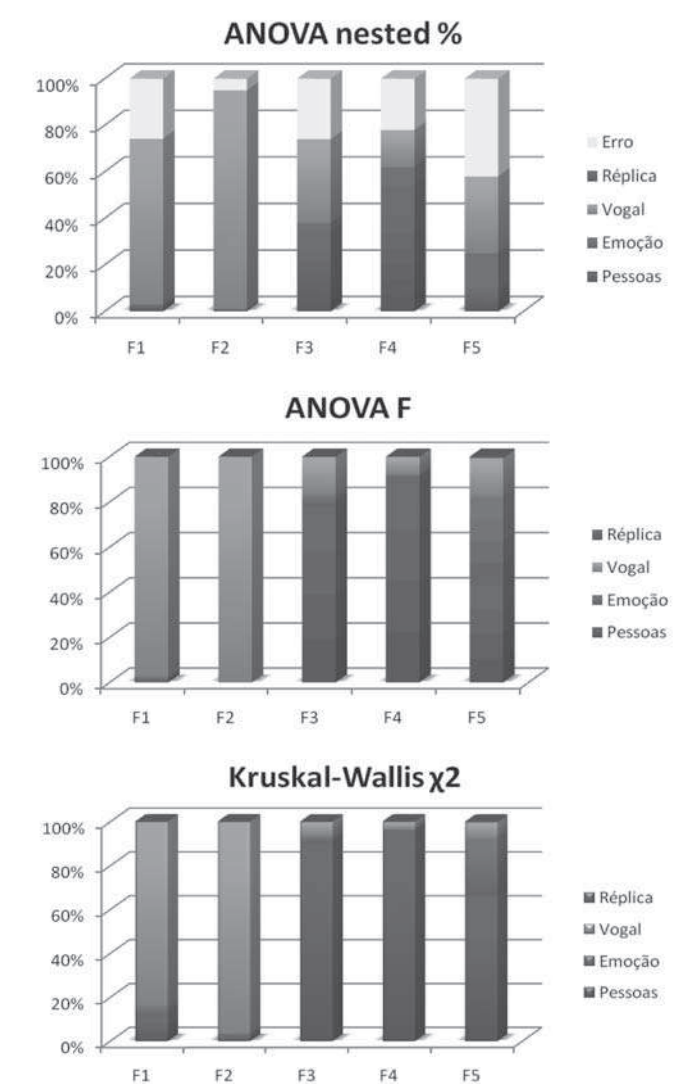


Figura 12 – Gráficos comparativos de Nested ANOVA, ANOVA F e Kruskal-Wallis χ^2 para Pessoas, Emoções Vogais e Réplica

> Cantores > Emoções), os valores são muito mais próximos, como o ilustra o gráfico da Figura 11.

A Figura 11 mostra-nos que a variação da amplitude dos formantes é mais significativa do que a frequência dos mesmos na discriminação das emoções, uma vez que, na partição de variação de amplitude dos formantes os valores de variância entre vogal, emoção e cantor estão mais próximos.

Finalmente (Figura 12), realizou-se uma ANOVA *nested*, onde a ordem dos factores foi a da natureza das variáveis: Pessoas (que cantam emoções), Emoções (onde se discrimina cada vogal), Vogais (onde são feitas três medições) e Réplica (medição). Esta análise integra todos os factores, sendo a ordem dos factores hierárquica. Tem ainda a particularidade de fornecer uma medida do erro. O mais surpreendente

quando se comparam as três análises de variância (ANOVA *nested*, ANOVA parcial e Kruskal-Wallis – Figura 12) é que é evidente a semelhança entre os gráficos, repetindo-se o mesmo padrão de partição da variância em cada análise.

Decomposição da variância (Vogais, Pessoas, Emoções)

No sentido de observar como é que a variância entre os factores estudados se decompôs sucessivamente em cada formante, fizeram-se árvores de regressão para cada um dos formantes. Estes resultados podem consultar-se em [13].

Discussão e conclusões

Resposta às questões de investigação

Quanto à primeira questão de investigação “*Existe uma relação entre a expressão das emoções e o timbre da voz cantada?*”. Os resultados desta investigação permitem responder afirmativamente a esta questão. Pelo facto de se terem efectuado diversas análises os resultados são robustos. Embora a variância tímbrica para as emoções seja, com efeito subtil, não é negligenciável.

Para se poder chegar à eventual variância tímbrica da exclusiva responsabilidade das emoções foi necessário compreender a variância formancial relativa quer às vogais, quer à individualidade tímbrica dos cantores. Após as duas primeiras instâncias estarem analisadas foi, então, possível observar o que se passava na frequência de ocorrência dos formantes que pudesse ser atribuído ao conteúdo emocional visado pelos cantores. Os resultados mostraram, em primeiro lugar, que há manipulação da frequência dos formantes para as diferentes emoções, o que corrobora os estudos de Morozov e Sundberg [1, 2 e 3]; em segundo lugar, que Alegria e Raiva apresentam na globalidade frequências formânticas mais elevadas do que Tristeza e Medo – Morozov [1] havia concluído que o formante do cantor subia na sua frequência de ocorrência para Raiva e Alegria e descia para Tristeza e Medo –.

Quanto à segunda questão de investigação: “*A existir uma relação, que tipo de relação é esta?*” Este estudo permitiu mostrar que há uma relação directa exibida entre a amplitude e a frequência de ocorrência dos formantes e as emoções expressas. No que respeita às frequências dos formantes pôde constatar-se a existência de valores mais altos de frequência dos formantes nas emoções Raiva e Alegria, e de valores mais baixos nas emoções Tristeza e

Medo. Nos resultados da avaliação perceptiva houve, como foi referido, uma confusão perceptiva assinalável entre as emoções Alegria/Raiva e Medo/Tristeza. Ora, após os resultados pode, agora, compreender-se a razão desta incompreensão perceptiva: os quadros formânticos de Alegria e Raiva estão próximos e os de Tristeza e Medo também.

Com o resultado da ACP para as emoções (Figura 8), no qual se excluiu a variância da responsabilidade de F1 e F2 (uma vez que, covariando com F1 e F2, ou com as três vogais, se obtiveram os mesmos resultados) pôde observar-se que há uma relação de Raiva com F5, de Tristeza com F4 e, de Alegria com F3 (se bem que menos acentuada). Pode, assim, constatar-se uma possível relação específica de cada emoção com cada formante. Provavelmente, para Medo deve faltar, nesta análise, o formante com o qual possa estar mais directamente relacionado (possivelmente um formante acima de F5).

Quanto à terceira questão de investigação: “*Há padrões típicos nesta relação emoções/timbre?*” Ainda que os resultados desta investigação não sejam suficientes para afirmar categoricamente que há padrões específicos das emoções visíveis no timbre, nem qual a tipologia destes padrões em cada emoção, estes resultados permitem afirmar que cada emoção tem uma assinatura individualizada no timbre da voz cantada (pois que esta evidência foi recorrente nos vários resultados): em geral, o mapa das frequências dos formantes em ordem descendente é Raiva, Alegria e Tristeza/Medo. Na intenção de melhor compreender o padrão Raiva fizeram-se gráficos, a título exemplificativo, que permitem apontar para o facto de haver padrões-tipo relativamente a cada emoção no timbre exibido a partir das frequências formanciais [13]. É preciso, contudo, mais investigação que permita melhor delinear os padrões tímbricos das emoções.

Quanto à quarta questão de investigação: “*Havendo padrões, estes são independentes do tipo de voz e do sexo dos indivíduos?*”. No que respeita a esta questão, ainda que se não tenha podido responder totalmente, os resultados apontam para o facto de, por um lado, os possíveis padrões das emoções serem independentes do tipo de voz dos indivíduos (uma vez que, naqueles em que foi possível averiguar, por terem um maior número de emoções, a tipologia da voz não influenciou o padrão geral observado), por outro, que há *nuances* nestes padrões relacionadas com o sexo dos indivíduos [13]. Pôde observar-se uma diferença no que respeita ao modo de se expressarem os cantores do sexo feminino, relativamente aos

cantores do sexo masculino [13]: nos primeiros a discriminação emocional é mais acentuada havendo uma clara distinção entre o modo Neutro e a Tristeza; nos segundos a discriminação emocional é menos acentuada havendo um paralelismo entre Neutro e Tristeza. Estes resultados necessitam de mais investigação futura.

É, também, de salientar o gradiente claro visível na ACP da Figura 6 (a qual, por ser de Cantores *versus* Emoções exibe resultados robustos relativamente ao desequilíbrio da amostra), onde, na parte positiva do primeiro eixo se localizam a maioria dos timbres masculinos e na parte negativa os timbres femininos. O desajuste de alguns cantores, deve-se, provavelmente, à sua qualidade técnica e ao facto de possuírem menos anos como profissionais. É, por isso, possível, que a melhoria na qualidade técnica do canto, adquirida ao longo da vida profissional, esteja associada à manipulação de F3 e F4. Isto não é surpreendente uma vez que uma das aprendizagens técnicas perseguidas pela pedagogia vocal é a do abaixamento da laringe, cuja consequência geral é o abaixamento da frequência dos formantes. Nos cantores do sexo masculino de menor qualidade técnica há sempre desvios, no caso das cantoras do sexo feminino, não há dados suficientemente conclusivos, mas o padrão parece ser o inverso. Com esta informação pode compreender-se melhor a Figura 6 uma vez que a progressão na excelência implica uma especialização diferencial que maximiza as diferenças tímbricas entre homens e mulheres bem como a capacidade de sucessivamente alinhar as emoções com o eixo representado pelos vectores F3 e F4.

Quanto à quinta questão de investigação: “*A expressão tímbrica das emoções afecta igualmente todos os formantes? Quer na sua frequência, quer na sua amplitude?*” A resposta à primeira pergunta desta questão é negativa. Os resultados mostram, claramente, que há formantes mais importantes para a expressão tímbrica das emoções, nomeadamente F5 é o mais significativo, seguido de F1, e que há formantes não significativos para a expressão emocional, nomeadamente F2 e F4.

Nos diversos resultados apresentados são F4, F3, e menos significativamente F5, os principais responsáveis pela discriminação da individualidade tímbrica dos cantores. Estes resultados corroboram a literatura existente [11 e 14] que afirma estar a qualidade da voz relacionada com os formantes acima de F2. Sundberg aponta o quarto formante como o principal indicador da individualidade tímbrica, por este estar muito

relacionado com as dimensões do tubo laríngeo e com o comprimento do tracto vocal. Curiosamente, F4 e F3 são também os formantes responsáveis pelo formante do cantor, pelo que há uma sobreposição das funções de F3 e F4: diferenciar individualmente os cantores (o que foi estudado nesta investigação) e, simultaneamente, permitir a projecção da voz cantada. A relação íntima de F3 e F4 é visível nas Figuras 3, 5 e 7 em que estão mais próximos do que os restantes formantes. Este facto poderá estar relacionado com a manipulação que os cantores fazem destes formantes para melhor projectar a voz. Esta poderia, também, ser a explicação possível para diferenças tão significativas para os cantores relativamente a F3 e F4, já que cada cantor terá o seu caminho muito particular no domínio e controlo da projecção da voz (exactamente ao contrário das vogais, que sendo um padrão comum a todos os cantores dificilmente os formantes responsáveis pela sua discriminação poderiam separar timbricamente os cantores).

Na globalidade, a amplitude dos formantes é afectada pela expressão das emoções se bem que não se tenha chegado a averiguar como é que ela é discriminada para cada emoção. No que respeita à amplitude dos formantes apenas se analisou se havia uma relação significativa entre a amplitude exibida pelos formantes e as emoções, e os resultados foram confirmativos, inclusivamente, mostraram que o grau de variância das amplitudes dos formantes para as emoções é maior do que a variância da frequência dos formantes para as emoções.

Quanto à sexta questão de investigação: “*F1 e F2, por serem formantes particularmente importantes para a discriminação das vogais, são menos manipuláveis para a expressão emocional do que F3, F4 e F5?*” Os resultados respeitantes à manipulação tímbrica efectuada para a discriminação das vogais confirmam a literatura existente, quer para a voz falada [15, 16, 17, 18], quer para a voz cantada [11,14, 19, 20, 21, 22], – tendo em conta que no presente estudo Fo esteve sempre abaixo de 600 Hz pelo que as posições relativas de F1/F2 não se alteraram –. Neste estudo, todos os resultados mostraram F1 e F2 como principais responsáveis pela distinção tímbrica das vogais e mostraram que é a sua posição relativa que permite distinguir as vogais [a], [i] e [u], embora a movimentação frequencial do segundo formante seja mais significativa para estas vogais do que a variância de F1. Este formante separa claramente [a] de [i] e de [u], enquanto que F2 discrimina com absoluta clareza [i] de [a] e [u]; a combinação de F1 e F2 é por

isso completamente determinante na discriminação das três vogais, muito provavelmente de todas as vogais. Relativamente a F2 pode dizer-se, à luz dos resultados, que, efectivamente, é pouco manipulável na expressão e discriminação das emoções. Relativamente a F1, embora a sua preponderância na diferenciação das vogais seja eminente, este é também um formante importante para as emoções, apresentando-se como o segundo formante com maior variância nas emoções. O que leva a supor, portanto, que F1 não tem uma função tão importante como F2 na diferenciação vocálica e que, de qualquer modo, não é esta função que inviabiliza a movimentação da frequência dos formantes na expressão emocional. Este estudo permitiu, também, averiguar que a maioria da variância observada no timbre diz respeito à diferenciação das vogais.

Quanto à sétima questão de investigação: “A frequência e amplitude de Fo afecta a frequência de ocorrência e a amplitude dos formantes?” A resposta a esta questão é negativa de acordo com os procedimentos metodológicos que se usaram: a escolha de vogais a analisar cuja Fo fosse baixa de modo a garantir que a movimentação frequencial de F1 devida a Fo não fosse significativa. Verificou-se, posteriormente (com os resultados patentes no Quadro 3), que este procedimento possibilitara a não significância de Fo na frequência de ocorrência dos formantes o que garantiu, neste estudo, a neutralização desta variância na dependência de Fo. A influência da amplitude de Fo na amplitude dos formantes não chegou a ser investigada se bem que a literatura existente quer para a voz falada [18], quer para a voz cantada [20], indique haver uma relação significativa entre as duas instâncias. Haveria, ainda, que saber até que ponto as diferenças de amplitude observadas nos formantes se deviam a Fo ou à expressão emocional, isto é, seria necessário efectuar as mesmas análises de covariância efectuadas para a frequência dos formantes relativamente às amplitudes. A literatura existente para a voz cantada [1] afirma uma relação efectiva entre a diversidade das amplitudes dos formantes e a expressão emocional.

Quanto à oitava questão de investigação: “A expressão tímbrica das emoções varia com as vogais? Há vogais mais permeáveis às emoções do que outras?” Pelos resultados podemos afirmar, por um lado, que sim pois, recorrentemente, as vogais [i] e [u] apareceram como melhores na discriminação das emoções do que [a]. Contudo, pode também afirmar-

-se que, quando a qualidade expressiva do cantor é melhor [13] as diferenças da qualidade discriminatória das vogais não é tão acentuada aparecendo já [a] a par de [i] e [u]. Tendo em conta que a vogal [a] é a vogal cantada mais difícil de produzir com qualidade pelo cantor haveria que averiguar se a qualidade técnica dos indivíduos interfere na sua capacidade expressiva. Por outro lado, pode supor-se que, se todos os cantores analisados tivessem uma qualidade técnica e expressiva elevada, então não haveria diferenças na expressão tímbrica exibida pelas vogais. É preciso mais investigação para responder a estas questões. Pode, no entanto, afirmar-se que há vogais nas quais a expressão emocional é mais fácil do que noutras (não que elas sejam mais ou menos permeáveis), por ordem: [i], [u] e [a].

Quanto à nona questão de investigação: “Há emoções cuja expressão na voz cantada é mais executável do que outras? Se há, quais?” Os resultados expressos nesta investigação não permitem responder com clareza a esta questão se bem que possamos observar que a emoção Raiva é aquela que é mais claramente visível no timbre da voz cantada, e a emoção Medo a mais dificilmente reconhecível. Isto talvez possa ser explicado pela dificuldade sentida pelos cantores relativamente a esta emoção.

Conclusões

1. Com esta investigação pode afirmar-se que, exclusivamente pela análise tímbrica, se podem averiguar conteúdos emocionais específicos. A literatura existente [1, 2 e 3] afirma que as emoções da voz cantada são legíveis acusticamente a partir de um conjunto multifactorial de parâmetros, aqui pudemos constatar que a análise exclusiva do timbre permite aferir das emoções veiculadas. A variância exibida pela frequência de ocorrência dos formantes para as emoções é, quando comparada com a mesma variância para as vogais ou para a individualidade dos cantores, subtil, no entanto, é claramente consistente.

2. Pode também afirmar-se que há quadros formânticos para cada emoção, se bem que o padrão-tipo tímbrico de cada emoção não tenha sido delineado. Pelos resultados não generalizáveis aqui obtidos, a assinatura específica das emoções é independente da tipologia das vozes mas não é independente do sexo dos indivíduos. Há contudo, que esperar por mais investigação.

3. Esta investigação permitiu mostrar que os formantes têm funções especializadas: F2 e F1 para

a discriminação das vogais, F4, F3 e F5 para a diferenciação tímbrica individual das pessoas, F5 e F1 para a variância dos formantes atribuída às emoções.

4. Dos resultados obtidos nesta investigação pode afirmar-se que, na amostra estudada, há uma hierarquia das vogais para a expressão das emoções: das vogais aqui analisadas a expressão emocional é mais executável em [i], a seguir em [u] e só depois [a].

Limitações do estudo

A primeira limitação deste estudo diz respeito ao tamanho da amostra estudada que, se bem que tenha inicialmente incluído 12 indivíduos (60 extractos musicais com conteúdo emocional específico), após os excertos por eles cantados terem sido seleccionados perceptivamente com o intuito de garantir uma efectiva expressão emocional posteriormente analisável, ficou consideravelmente reduzida (11 indivíduos e 38 excertos). Esta redução teve como consequência que das emoções Alegria e Medo apenas cinco excertos tenham sido analisados e da emoção Raiva, seis; de Tristeza foram analisados onze excertos correspondentes a onze dos doze indivíduos da amostra total. Por não se ter tido todas as emoções em todos os indivíduos, não se pôde comparar todos os cantores em todas as emoções, por um lado, nem avaliar da sua maior ou menor expressividade, por outro. Também não se pôde aferir, com exactidão, se a expressão emocional visível no timbre é independente ou não da tipologia das vozes ou do sexo dos indivíduos.

A segunda limitação decorre da primeira: pelo facto de não se ter tido todas as emoções em todos os indivíduos também não foi possível averiguar até que ponto os padrões individuais das emoções são idênticos entre si ou não, embora se tenha delineado a assinatura global de cada emoção.

Implicações teóricas e práticas e Perspectivas futuras

A existência de uma relação directa entre a expressão emocional e o timbre permite inferir, inversamente, o estado emocional do indivíduo a partir da análise do timbre da sua voz cantada. Isto explica, do ponto de vista teórico, como é que o cantor comunica emocionalmente e porque é que esta comunicação musical é metalinguística e universal. Ainda do ponto de vista teórico, sabendo que cada emoção se produz debaixo de um quadro

específico tímbrico pode produzir-se artificialmente a “voz” daquela emoção, pode comunicar-se artificialmente aquela emoção e, enfim, pode manipular-se, e mesmo ludibriar o interlocutor, a partir deste código metalinguístico. A vantagem de tal saber é óbvia para as mais diversas áreas (por exemplo, para as editoras discográficas, para a indústria cinematográfica, para a investigação forense, para a robótica, entre outras) e não apenas para a artística (*performance* e pedagogia do canto).

No plano prático, sabendo exactamente que as emoções se repercutem no timbre do indivíduo, e que o fazem de modo específico, os gestos precisos conducentes àqueles mapas tímbricos podem ser ensinados e mimados para a *performance* artística. Há que, naturalmente, saber, como é que no plano fisiológico se dão as manipulações para a produção das emoções que se repercutem acusticamente no timbre, isto é, compreender de que modo é que os quadros acústicos emocionais foram produzidos fisiologicamente. Esta é uma investigação futura com alcance evidente.

Aceitando como robustos os resultados desta investigação que mostram os formantes como entidades com funções especializadas, seria interessante em investigações futuras verificar se a análise formântica do espectro acima de F5 produz resultados. De qualquer modo, tudo parece apontar para o facto de que a compreensão das emoções no timbre, e a compreensão dos seus padrões típicos segundo a emoção, precisa de uma análise mais alargada do espectro. Este é, também um campo profícuo, que necessita de mais investigação futura (ainda que as limitações técnicas permaneçam, por ora, como um obstáculo).

No início desta dissertação perguntava-se se, a limite, o talento artístico, e neste caso específico, o talento musical veiculado pela voz, seria ensinável. A investigação aqui levada a cabo apontou para o facto de se poder responder a esta questão afirmativamente: naturalmente que não se pode ensinar um indivíduo a ter emoções, mas, sabendo exactamente o que se produz tímbricamente para cada emoção pode ensinar-se a mimar o gesto dessa emoção, pelo que, em parte, o talento também é ensinável.

Compreendendo que o timbre da voz é um microcosmo do universo emocional da pessoa, compreendendo o timbre, compreende-se, também, a pessoa. Compreendendo que o timbre das emoções é a metalinguagem universal da música compreende-se a música e o seu propósito.

Bibliografia

1. Morozov, V. (1996). Emotional expressiveness of the singing voice: The role of macrostructural and microstructural modifications of spectra. *Logopedics Phoniatics Vocology*, 21: 49-58.
2. Sundberg, J; Iwarsson, J. & Hagegard, H. (1995). A singer's expression of emotions in sung performance. In O. Fujimura & M. Hirano (Eds.), *Vocal fold physiology* (pp. 217-231). San Diego: Singular Publishing Group.
3. Sundberg, J. (1998) Expressivity in singing. A review of some recent investigations. *Logopedics Phoniatics Vocology*, 23:121-127.
4. Jansens, S., Bloothoof, G., Krom, G. de (1997). Perception and acoustics of emotions in singing. *Proceedings of the 5th Eurospeech*, Rhodes, IV: 2155-2158.
5. Scherer, K.R. (1995). Expression of emotion in voice and music. *Journal of Voice*. 9(3): 235-248.
6. Siegwart, H. & Scherer, K. (1995). Acoustic concomitants of emotional expression in operatic singing: the case of Lucia in *Ardi gli incensi*. *Journal of Voice*, 9(3): 249-260.
7. Laukkanen, A.; Vilkmán, E.; Alku, P. & Oksanen, H. (1997). On the perception of emotions in speech: The role of voice quality. *Logopedics Phoniatics Vocology*, 22, 157-168.
8. Rapoport, E. (1996). Emotional expression code in Opera and Lied singing. *Journal New Music Res.* 25, 109-149.
9. Foulds-Elliot, S.D.; Thorpe, C.W.; Cala, S.J. & Davis, P.J. (2000). Respiratory function in operatic singing: Effects of emotional connection. *Logopedics Phoniatics Vocology* 25.
10. Fonagy, I. (1991). *La vive voix. Essai de psycho-phonétique*. Paris : Ed.Payot.
11. Sundberg, J. (1987). *The science of the singing voice*. Illinois: Northern Illinois University Press.
12. Guimarães, I. (2007). *A ciência e a arte da voz humana*. Alcabideche: Escola Superior de Saúde do Alcoitão.
13. Pereira, A. L. (2007). *As cores da voz: Expressão das emoções no timbre da voz cantada*. Dissertação apresentada à Universidade Católica Portuguesa para a obtenção do grau de mestre em Ciências da Fala.
14. Sundberg, J. (1994). Perceptual aspects of singing. *Journal of Voice* 8(2): 106-122
15. Delgado-Martins, M.R. (1973). Análise acústica das vogais tónicas do português. *Boletim de Filologia XXII* (3/4): 303-314.
16. Kent, R.D. & Read, C. (1992). *The acoustic analysis of speech*. San Diego: Singular Publishing Group.
17. Ladefoged, P. (1996). *Elements of acoustic phonetics*. Chicago: The University of Chicago Press.
18. Mateus, M. H. M., Andrade, A, Viana, M.C. & Villalva, A. (1990). *Fonética, fonologia e morfologia do português*. Lisboa : Universidade Aberta.
19. Appelman, D.R. (1967). *The science of vocal pedagogy*. Bloomington: Indiana University Press.
20. Milhouse, T.J. & Clermont, F. (2004). Systematic comparison of spoken and sung vowels using perceptual linear-prediction analysis. *Proceedings of the 10th Australian International Conference on Speech Science & Technology*. Sydney: Australian Speech Science & Technology Association Inc. December, 8-10.
21. Scotto Di Carlo, N. (1972). Étude acoustique et auditive des facteurs d'intelligibilité de la voix chantée. *Proceedings of the VIIIth International Congress of Phonetic Sciences Mouton*, La Haye 1017-1023.
22. Scotto Di Carlo, N. (2005) Contraintes de production et intelligibilité de la voix chantée, *Travaux interdisciplinaires du Laboratoire Parole et Langage* 24: 159-179.