

Instituto Politécnico do Porto  
Escola Superior de Música, Artes e Espectáculo



**Tecnologia de Apoio em Tempo-Real ao Canto -  
Relação entre parâmetros perceptivos da voz  
cantada com fenómenos acústicos objectivos.**

João Filipe Terleira de Sá Ferreira

Mestrado em Música - Interpretação Artística

Área de especialização Canto

Orientador: Professora Doutora Sofia Lourenço

Co-Orientador: Doutor Aníbal Ferreira

Professor Associado do Departamento de Engenharia Eletrotécnica e Computadores da  
Faculdade de Engenharia da Universidade do Porto

Novembro de 2012

---

©João Terleira, 2012



---

## Resumo

A avaliação da qualidade de voz de um cantor ou de um estudante de canto, está normalmente associada a fenómenos de percepção e à interpretação de um som pelo orientador ou professor. Estes fenómenos de natureza auditiva são frequentemente subjectivos e dependentes da interpretação de cada orientador. Embora existam certos parâmetros de avaliação e classificação de cada tipo de voz relativamente consensuais, estes são de natureza subjectiva e por vezes difíceis de explicar. Uma das grandes batalhas dos orientadores/professores de canto é de facto, saber interpretar e descodificar os vícios e o porquê das dificuldades de cada cantor, com base única e exclusivamente na sua audição.

A tecnologia carece, até ao momento, de qualquer dispositivo ou software capaz de interpretar e comprovar o que distingue uma voz com maior ou menor qualidade.

Esta dissertação, insere-se num projecto apoiado pela Fundação para a Ciência e Tecnologia, que visa a elaboração de um apoio ao ensino do canto, tanto num contexto de aula como fora desta.

Nesta dissertação foram estudados vários parâmetros perceptivos normalmente utilizados no contexto de uma aula de Canto, mas desconhecidos ou difíceis de interpretar para o público leigo e muitas vezes pelos próprios estudantes quando estes se encontram numa fase inicial da sua aprendizagem. Seguidamente, foram efectuados diversos testes baseados na área da engenharia de processamento de sinal onde as gravações efectuadas foram analisadas e delas retiradas características acústicas representativas cuja correspondência se pretende estabelecer com parâmetros perceptivos.

O objectivo deste trabalho de investigação é então identificar, caracterizar e definir, da forma mais objetiva possível, os atributos musicais/estilísticos/expressivos mais importantes da voz cantada, em correspondência com parâmetros perceptivos (e.g., altura, brilho, etc.) e estabelecer uma conexão entre esses parâmetros e características acústicas objetivas obtidas na análise das gravações.

Este projecto de investigação será articulado com investigação na área da engenharia de processamento de sinal visando apurar características acústicas da voz que

---

correlacionam e comprovam os parâmetros perceptivos seleccionados através da existência de tais fenómenos acústicos ou não.

---

## Abstract

The quality measurement of a singer's voice is normally associated with perceptual phenomenon and sound interpretation by the singing teacher. Although there are there are certain parameters of evaluation and classification of each type of voice, these auditory perceptual parameters are frequently subjective and sometimes with a difficult explanation. One of the great challenges to singing teachers is to know how to decode and how to solve the student's difficulties and vocal limitations only with the audition and recognition of these perceptual attributes.

Technology, as we know it, lacks of any software device in order to interpret or to prove what distinguishes a voice with less or better quality.

This dissertation is part of a project supported by "Fundação para a Ciência e Tecnologia" with the objective to develop several tools for teaching and learning singing in the class or in home. Is not the objective of this investigation to limit or to suppress an orientation by the singing teacher although to create and develop some tools to complete and support the orientation.

In this dissertation, several perceptual parameters that are used in the context of a singing class were studied and clarified as well as possible. After that, several engineering tests based on the signal recognition and processing where the recordings were analyzed.

One of the tasks of this work is then to identify and to define as objective as possible, the musical/expressive attributes of the singing voice in correspondence with perceptual parameters (ex. Brightness, highness, etc.) and establish a connection between these parameters and the results obtained by the analysis of the recording samples.

This investigation project is articulated with investigation on electrical engineering of sign processing in order to order to determine the acoustic characteristics of the voice that correlate and confirm the perceptive parameters selected by the existence of such acoustic phenomena or not.



---

## Agradecimentos

Gostaria de agradecer à professora Sofia Lourenço pela orientação e apoio durante este trabalho. Ao mentor deste projecto, o professor Aníbal Ferreira, não só pela orientação mas também pela oportunidade de participar neste trabalho tão ambicioso e ao mesmo tempo tão interessante.

A todos os meus colegas que trabalharam comigo neste projecto destacando obviamente o Vítor, Tiago e Ricardo por toda a cooperação e ajuda tão necessária a um leigo na área da engenharia.

À Lígia e á minha família pela compreensão, apoio e sobretudo paciência, não só no meu percurso académico mas também durante toda uma vida. Ao Duarte Leitão pelo investimento de tempo nas gravações efectuadas e ao Pedro Alves pela cedência do seu estúdio privado sem o qual grande parte deste trabalho não era realizado.

Por fim ao professor Rui Taveira por toda a importância que teve na minha caminhada como estudante e pela influencia que exerce diariamente no meu percurso profissional e artístico.



---

*“Em arte, procurar não significa nada. O que importa é encontrar.”*

Pablo Picasso



---

# Índice

## 1. Introdução

1.1.	Motivação	23
1.2.	Enquadramento	24
1.3.	Objectivos	25
1.4.	Estrutura	26

## 2. Estado da Arte

2.1.	Introdução	29
2.2.	Voz – O fenómeno de Fonação	30
2.3.	O Trato Vocal	31
2.4.	Modelo Fonte- Filtro	32
2.5.	Formantes	33
2.6.	Frequência fundamental	36
2.7.	Voz Cantada vs. Voz Falada: Principais Diferenças	37
2.8.	Formante de Cantor	37
2.9.	Vozeamento	39
2.10.	Percepção da voz cantada	40
2.11.	Parâmetros perceptivos e qualitativos	
2.11.1.	Introdução	42
2.11.2.	Afinação	42
2.11.3.	Tessitura	43
2.11.4.	Timbre	45
2.11.4.1.	Claro/Escuro	46
2.11.4.2.	Voz na frente/ Voz recuada	47

---

2.11.4.3.	Voz leve/ Voz pesada (ou Repertório leve/ Repertório pesado)	47
2.11.4.4.	Limpidez/ Soprosidade	48
2.11.4.5.	Suavidade/ Aspereza	48
2.11.5.	Falsete	48
2.11.6.	Ataque	49
2.11.7.	<i>Vibrato</i>	50
2.11.8.	<i>Legato</i>	51
2.11.9.	<i>Staccato</i>	51
2.11.10.	Microdinâmicas e destreza vocal	52
2.11.11.	Portamento	53
3.	Análise dos parâmetros qualitativos/perceptivos	
3.1.	Introdução	55
3.2.	Base de Dado	56
3.3.	Parâmetros Escolhidos	56
3.3.1.	Portamento	57
3.3.2.	Falsete/ Voz de Cabeça (Falsete vs. Registo Modal)	59
3.3.3.	Limpidez vs Soprosidade	61
3.3.4.	Formante de Cantor vs. Voz Plana	62
3.3.5.	Vibrato	64
4.	Análise Estatística – Resultados obtidos	
4.1.	Introdução	67
4.2.	Parâmetros acústicos	68
4.2.1.	<i>Jitter e Shimmer</i>	68
4.2.2.	Autocorrelação	70
4.2.3.	<i>Harmonics-to-noise Ratio (HNR)</i>	70

---

4.2.4.	<i>Noise-to-harmonics Ratio (NHR)</i>	71
4.2.5.	<i>Short-time Energy</i>	71
4.2.6.	<i>Spectral Flux</i>	71
4.2.7.	<i>Spectral Centroid</i>	71
4.2.8.	<i>Spectral Entropy</i>	72
4.2.9.	<i>Spectral Rolloff</i>	72
4.2.10.	<i>Zero-crossing rate (ZCR)</i>	72
4.3.	Resultados obtidos	
4.3.1.	Vibrato	73
4.3.1.1.	Frequência do vibrato	73
4.3.1.2.	Extensão do vibrato ( <i>pitch</i> )	75
4.3.1.3.	Pureza Sinusoidal	76
4.3.2.	Falsete	77
4.3.2.1.	<i>Spectra Entropy</i>	79
4.3.2.2.	Autocorrelação média	80
4.3.2.3.	<i>NHR (Noise-to-harmonics Ratio)</i>	81
4.3.3.	Formante de cantor	82
4.3.3.1.	<i>Jitter</i>	84
4.3.3.2.	<i>Shimmer</i>	85
4.3.3.3.	<i>HNR (Harmonics-to-noise Ratio)</i>	86
4.3.4.	Limpidez/Soprosidade	87
4.3.4.1.	Jitter (rap)	89
4.3.4.2.	<i>Spectral Flux</i>	90
4.3.4.3.	<i>Zero-cross Rate</i>	91
4.3.4.4.	O caso especial da análise do par Limpidez/Soprosidade	92
4.3.5.	Portamento	92
4.3.5.1.	Tempo da transição	94
4.3.5.2.	Tempo Médio por Nota	95
4.3.5.3.	Declive da transição	96
4.4.	O caso especial de Teresa Salgueiro	97

---

4.4.1. Vibrato	97
4.4.2. Existência de Voz Plana	98
4.4.3. Possível existência de falsete	100
4.5. Conclusões	103
5. Conclusão e Futuras aplicações	105
Bibliografia	108

---

## Lista de Figuras

1. 2.3 Diferentes configurações do tracto vocal durante as emissões das vogais “i”, “a” e “u”.
2. 2.3. O Tracto Vocal
3. 2.5. Configuração do tracto vocal e distribuição de formantes para as vogais <i> e <a>.
4. 2.5. Representação gráfica do fenómeno Sintonia F0-F1 também chamado de “afinação dos formantes”.
5. 2.8. Espectro de longa duração do som de uma orquestra com e sem cantor e de conversação normal.
6. 3.3.1. Transição entre as notas Mi 3 e Si 3 feita através de portamento.
7. 3.3.1. Transição entre as notas Mi 3, Si 3 e Sol# 3 feitas sem portamento.
8. 3.3.1. Representação espectral da frequência fundamental e parciais harmónicos na transição com uso de portamento.
9. 3.3.1. Representação espectral da frequência fundamental e parciais harmónicos na transição sem uso de portamento.
10. 3.3.2. Comparação entre parciais harmónicos nos registos de voz modal e registo de falsete para voz masculina.
11. 3.3.2. Comparação entre parciais harmónicos nos registos de voz modal e registo de falsete para voz feminina.
12. 3.3.3. Representação espectral de voz limpa e voz soprosa para voz masculina.
13. Representação espectral de voz limpa e voz soprosa para voz feminina.
14. 3.3.4. Análise espectral dos parciais harmónicos numa voz masculina no registo voz plana e voz com formante de cantor.
15. 3.3.4. Análise espectral dos parciais harmónicos numa voz feminina no registo voz plana e voz com formante de cantor.
16. 3.3.5. Detecção de *frame* contendo vibrato no programa SingingStudio
17. 3.3.5. Valores extraídos do vibrato detectado anteriormente.

---

## Lista de Tabelas

1. 2.11.3 Tabela de catalogação de tipos de voz masculinas consoante a tessitura (*Fach*).
2. 2.11.3. Tabela de catalogação de tipos de voz femininas consoante a tessitura (*Fach*).
3. 2.11.4.1 Exemplos de vozes masculinas e femininas para os parâmetros Claro e Escuro.
4. 4.3.1.1 Valores médios e desvio padrão das amostras recolhidas para o parâmetro Frequência do vibrato para as vozes de Baixo, Barítono, Tenor, Contralto, Mezzo-Soprano e Soprano.  $N$  – número de amostras.
5. 4.3.1.2 Valores médios e desvio padrão das amostras recolhidas para o parâmetro Extensão do vibrato (*pitch*) para as vozes de Baixo, Barítono, Tenor, Contralto, Mezzo-Soprano e Soprano.  $N$  – número de amostras.
6. 4.3.1.3 Valores médios e desvio padrão das amostras recolhidas para o parâmetro Pureza Sinusoidal para as vozes de Baixo, Barítono, Tenor, Contralto, Mezzo-Soprano e Soprano.  $N$  – número de amostras.
7. 4.3.2. Valores dos testes de Kruskal-Wallis e Mann-Whitney U para os parâmetros acústicos nas amostras de voz em falsete e em registo modal.
8. 4.3.3. Valores dos testes de Kruskal-Wallis e Mann-Whitney U para os parâmetros acústicos nas amostras de voz plana e formante de cantor.
9. 4.3.4 Valores dos testes de Kruskal-Wallis e Mann-Whitney U para os parâmetros acústicos nas amostras de voz plana e formante de cantor.
10. 4.3.5 Valores dos testes de Kruskal-Wallis e Mann-Whitney U para os parâmetros acústicos nas amostras de voz plana e formante de cantor.
11. 4.4 Valores extraídos para caracterização do vibrato encontrado em amostras de voz de Teresa Salgueiro.

---

## Lista de Gráficos

1. 2.11.3. Frequências fundamentais relativas a fala e canto entre indivíduos do sexo masculino e feminino.
2. 4.3.2.1. Teste de Kruskal-Wallis para *Spectral Entropy* entre registos Modal e Falsete.
3. 4.3.2.2. Teste de Kruskal-Wallis para Autocorrelação média entre registos Modal e Falsete.
4. 4.3.2.3. Teste de Kruskal-Wallis para valores de NHR médio entre registos Modal e Falsete.
5. 4.3.3.1. Teste de Kruskal-Wallis para parâmetros *jitter* (local) para exemplos de voz plana e formante de cantor.
6. 4.3.3.2. Teste de Kruskal-Wallis para parâmetros *shimmer* (local) para exemplos de voz plana e formante de cantor.
7. 4.3.3.3. Teste de Kruskal-Wallis para parâmetros *HNR* médio para exemplos de voz plana e formante de cantor.
8. 4.3.4.1. Teste de Kruskal-Wallis para parâmetros *jitter* (*rap*) para exemplos de voz límpida e voz com sopro.
9. 4.3.4.2. Teste de Kruskal-Wallis para valores de *Spectral Flux* para exemplos de voz límpida e voz com sopro.
10. 4.3.4.3. Teste de Kruskal-Wallis para parâmetros *jitter* (*rap*) para exemplos de voz límpida e voz com sopro.
11. 4.3.5.1. Teste de Kruskal-Wallis para valores de Tempo Total de Transição para exemplos com portamento e sem portamento.
12. 4.3.5.2. Teste de Kruskal-Wallis para valores de Tempo Médio por Nota para exemplos com portamento e sem portamento.
13. 4.3.5.3. Teste de Kruskal-Wallis para valores de Declive da Transição para exemplos com portamento e sem portamento.

- 
14. 4.4.2 Teste de Kruskal-Wallis para valores *jitter (local)* para exemplos de voz plana, voz com formante de cantor e exemplo de Teresa Salgueiro.
  15. Teste de Kruskal-Wallis para valores *shimmer (local)* para exemplos de voz plana, voz com formante de cantor e exemplo de Teresa Salgueiro.
  16. Teste de Kruskal-Wallis para valores *HNR* médio para exemplos de voz plana, voz com formante de cantor e exemplo de Teresa Salgueiro.
  17. 4.4.3. Teste de Kruskal-Wallis para valores se *Spectral Entropy* nas vozes de Teresa Salgueiro e nas amostras em registo modal e em registo de falsete.
  18. Teste de Kruskal-Wallis para valores se Autocorrelação média nas vozes de Teresa Salgueiro e nas amostras em registo modal e em registo de falsete.

---

## Abreviaturas e Símbolos

- ESMAE – Escola Superior de Música, Artes e Espectáculos
- FCT – Fundação para a Ciência e Tecnologia
- FEUP – Faculdade de Engenharia da Universidade do Porto
- GRBAS – *Grade, Roughness, Breathiness, Asteny, Strain*
- HNR – *Harmonics-to-noise Ratio*
- NHR – *Noise-to-harmonic Ratio*
- ZCR – *Zero-Crossing Rate*



---

# Capítulo 1

---

# 1. Introdução

---

## 1.1. Motivação

Não é de forma leviana que a emissão da voz através do canto é considerada por muitos autores, a forma mais autêntica e expressiva de todas as manifestações artísticas. Basicamente existem duas formas de utilizar a voz num contexto profissional, a voz falada (locutores de rádio, apresentadores, jornalistas) e a voz cantada. A emissão falada é geralmente natural e inconsciente, não necessitando (salvo casos extremos) de treino ou ajustes técnicos. Por outro lado, a voz utilizada num contexto profissional, em particular a voz cantada, necessita de melhoramentos e da aquisição de conceitos de forma a executar uma performance correcta a nível estético e sem prejuízo para o aparelho fonatório. As características da voz cantada têm sido alvo de estudo pelos mais variados autores, em particular, o canto lírico, pela sua estética particular e pela técnica impregnada que necessita de muitos anos de aperfeiçoamento e estudo constante sendo, por isso, objecto de interesse de muitos autores.

Não é contudo consensual a existência de uma técnica e estética universal pois sendo a música uma arte e a arte uma construção cultural, é natural que residam algumas diferenças sobretudo na parte estética. Tal como na grande maioria dos

---

instrumentos com alguma preponderância na História da Música Ocidental, a voz como instrumento e a didática do canto, sofrem muito com a questão das “Escolas” o que faz com que consigamos distinguir por exemplo uma voz formada em Itália de uma voz formada na Alemanha, não só pelas suas características estéticas mas, muitas vezes, pela técnica impregnada e claro, pela própria fisionomia do cantor. O ensino do canto, como qualquer outro instrumento, é um ensino que se baseia na imitação do docente, porém apresenta uma pequena mas muito significativa diferença: ao contrário de qualquer outro instrumento que quando adquirido já vem construído sem necessitar de adaptação ao utilizador, a voz como instrumento é construída pelo próprio de forma pessoal e única, o que faz com que não haja um método padrão para o ensino. Contudo, a má construção e/ou má utilização do aparelho vocal pode levar a sérios problemas no futuro, sendo trabalho do professor, a orientação da construção da voz como instrumento e a sua utilização numa performance.

Pessoalmente, como Cantor, Estudante e mais recentemente Professor, tenho consciência que a procura de determinados conceitos e objectivos acaba por ser comum em todos os estudantes de canto e embora o tenham e consigam nas aulas, uma vez quebrado o ambiente entre professor e aluno com o final da mesma e/ou com a falta de estudo constante, os vícios e as dificuldades acabam por vir novamente ao de cima, sendo portanto este um processo progressivo, até serem eliminados ou controlados. A superação das dificuldades é portanto um processo dificultado pela ausência do professor no estudo fora do período da orientação deste, o que exige maturidade e sobretudo tempo para interiorização na memória física das noções e atributos técnicos do canto.

## 1.2. Enquadramento

Este trabalho enquadra-se no âmbito de um projeto financiado pela Fundação para a Ciência e Tecnologia (FCT), que tem como objetivo o desenvolvimento de meios tecnológicos de apoio em tempo real ao ensino do Canto e para uma monitorização preventiva da utilização da voz cantada.

O projeto reúne profissionais não só relacionados com o Canto como arte performativa, mas também de outras áreas distintas como Otorrinolaringologia e Engenharia

---

Electrotécnica, com o propósito de otimizar o ensino e o estudo do canto, assegurando a sua correta performance, prevenindo eventuais distúrbios vocais. Deste modo é previsto o desenvolvimento de um sistema de feedback visual em tempo real que avalia a qualidade de diferentes parâmetros da voz cantada de forma a ser agregado ao software SingingStudio anteriormente desenvolvido pela spin-off da Faculdade de Engenharia da Universidade do Porto, Seegnal Research, Lda.

Este sistema informático pretende identificar e caracterizar parâmetros de avaliação perceptiva da voz cantada, diminuindo a subjetividade inerente à avaliação do ato de cantar, através do estudo das características acústicas que melhor se correlacionam, desenvolvendo algoritmos eficientes para estimar essas relações. Estes algoritmos serão utilizados no software de apoio ao ensino de canto. Para o desenvolvimento destes, o projeto é desenvolvido em colaboração com um estudante de mestrado da Faculdade de Engenharia da Universidade do Porto (FEUP).

### 1.3. Objectivos

O objectivo deste trabalho é apresentar soluções inovadoras de apoio ao ensino do canto no que respeita aos seus atributos técnicos e estéticos, de modo a ajudar tanto o estudante como o docente, tomando como ponto de partida os conceitos artísticos já existentes, no que respeita ao domínio da voz como instrumento.

Uma vez que o material de apoio tecnológico ao canto é escasso e portador de algumas lacunas, procurar-se-á desenvolver um material de apoio constituído por uma base de dados. Esta correlaciona atributos perceptivos e objectivos de natureza técnica e/ou estética com parâmetros acústicos específicos de modo a promover um conhecimento aprofundado sobre o sistema de emissão de voz e evitar assim, também, perturbações futuras através da associação de exemplos de mau uso vocal (uso excessivo, stress vocal) a fenómenos acústicos. É também objectivo criar um modelo de avaliação de todos os parâmetros já referidos através de um bio-feedback em tempo-real.

Serão estudados determinados parâmetros, objectivos ou não, utilizados regularmente na aprendizagem da voz como instrumento tais como:

- 
- Reconhecimento e diferenciação dos diferentes tipos de voz consoante o timbre e tessitura;
  - Afinação;
  - Visualização e posterior estudo de aspectos musicais relacionados como a voz cantada tais como *legato*, timbre, coloratura, intensidade vocal, ataque e microdinâmicas;
  - Articulação do texto e posição do tracto vocal nas diferentes vogais;
  - Cobertura vocal e *Messa di Voce* (mistura de voz) directamente relacionada com a riqueza em harmónicos e brilho vocal;
  - Amplitude e controlo do Vibrato;
  - Eliminação de perturbações vocais tais como soprosidade, aspereza;
  - Formante do cantor.

A elaboração deste material tecnológico trará, pensamos nós, um maior aproveitamento no ensino do Canto, particularmente no estudo fora do horário de interação entre aluno e professor. Permitirá também detectar mais facilmente perturbações vocais capazes de provocar situações de disfonia no futuro. Além disso, poderá também ajudar ao cantor profissional, a manutenção do seu aparelho vocal.

## 1.4. Estrutura

Esta dissertação encontra-se dividida em cinco capítulos. No primeiro capítulo é feita uma introdução que contempla os objectivos, o enquadramento e a motivação que nos levou a realizar este trabalho. No segundo capítulo é feita uma revisão do estado da arte que contempla os atributos perceptivos mais importantes da voz cantada bem como uma descrição da voz e do fenómeno de fonação.

No terceiro capítulo apresentamos a forma como foi constituída a nossa base de dados e uma forma ainda primária de alguns dos resultados obtidos que vão de encontro às características descritas no capítulo anterior. Resultados estes que iremos tentar provar no capítulo 4 onde referimos os testes que foram efectuados e a sua análise estatística.

---

No capítulo final apresentamos uma conclusão que visa apurar os resultados que conseguimos alcançar com o trabalho realizado e a sua importância. É também apresentada uma descrição do trabalho futuro que envolve não só aspetos a melhorar no desenvolvimento deste tipo de trabalho mas também as aplicações que este pode vir a ter.

---

## Capítulo 2

---

## 2. Estado da Arte

---

### 2.1. Introdução

Neste capítulo faremos uma compilação da bibliografia consultada no decorrer da construção desta dissertação. Verificamos que, em alguns pontos a bibliografia é escassa ou pouco precisa e sem resultados conclusivos. Um dos objectivos desta dissertação é, no seguimento disto, clarificar alguns aspectos que foram alvo de pouca atenção por parte da comunidade científica, como por exemplo, o falsete.

É então feita uma descrição de alguns dos parâmetros mais importantes da voz cantada e falada, enumerando as suas principais diferenças. Estes parâmetros são muitas vezes dados como conhecimento empírico mas, na altura de os explicar de forma clara e precisa, a tarefa torna-se mais complicada do que o previsto. A revisão bibliográfica bem como algumas caracterizações e explicações dos parâmetros musicais e estilísticos foi feita em articulação com os restantes colegas envolvidos o projecto de investigação que, não estando ligados de forma profissional ao mundo da música e em particular do canto, nos ajudaram a clarificar determinados conceitos no sentido de os tornarmos mais legíveis para o público leigo.

---

## 2.2. Voz – O fenómeno de fonação

A produção da voz humana, também designada por fonação, consiste num aumento da pressão de ar a nível dos pulmões, originando assim pulsos de ar que passam pelas pregas vocais que vibram. As pregas vocais (ou cordas vocais) consistem em duas pregas musculares e membranosas situadas na zona da laringe e que constituem o elemento que vibra no fenómeno de produção sonora, ou seja, na fonação. A vibração das mesmas é consequência da sua adução, que impõe uma resistência à saída do ar e conseqüentemente, uma modulação dos fluxos de ar. A frequência típica de vibração das pregas vocais (i.e. as pregas vocais abrem e fecham) é de 200-220 vezes/segundo nos indivíduos do sexo feminino e 100-120 vezes/segundo nos indivíduos do sexo masculino, embora possa variar em ambos os casos, sobretudo com a natural variação no canto da frequência fundamental.

Sendo que as pregas vocais se situam na laringe, o som produzido através dos fenómenos acima referidos é designado por som laríngeo, sendo este a base da fala e do canto. O som laríngeo é composto pela sua frequência fundamental, a frequência mais baixa da onda sonora produzida que corresponde à vibração das pregas vocais, e pelos seus parciais harmónicos. A frequência fundamental emitida, está directamente relacionada com as características morfológicas das pregas vocais, nomeadamente o tamanho, elasticidade e grossura. Neste sentido, os diferentes valores de frequência fundamental entre homens, mulheres e crianças, e mesmo a variabilidade que existe entre a frequência fundamental dentro do mesmo sexo, é devida às características específicas das pregas vocais e da laringe de cada indivíduo (Sundberg 1991).

O som laríngeo produzido a nível das pregas vocais, é muito fraco para ser ouvido. Para que este se torne no som que habitualmente percebemos é então submetido à passagem pelas cavidades supra-glóticas nomeadamente a laringe, faringe, boca e cavidade nasal que constituem o tracto vocal, sendo este uma espécie de caixa de ressonância às frequências emitidas. As ressonâncias das cavidades supra-glóticas amplificam, não só a frequência fundamental mas também os seus parciais harmónicos, sendo que uns são mais amplificados que outros como resultado da interacção destes com as diversas ressonâncias.

---

### 2.3. O Tracto Vocal

O tracto vocal é constituído essencialmente pela cavidade oral, cavidade nasal, faringe e laringe. Podemos também definir dois tipos de órgãos que se encontram nestes componentes do tracto vocal, os órgãos articulatórios ativos e passivos. Os órgãos articulatórios activos são a língua, os lábios, o palato mole e a mandíbula (ou maxilar inferior) ao passo que os órgãos passivos são respectivamente o palato duro, os dentes e o maxilar inferior. Através de várias posições dos diferentes órgãos articulatórios, o tracto vocal assume então várias configurações que correspondem a diferentes “Filtros” ao som produzido a nível das pregas vocais, sendo portanto o som escutado diferente para cada um dos filtros.

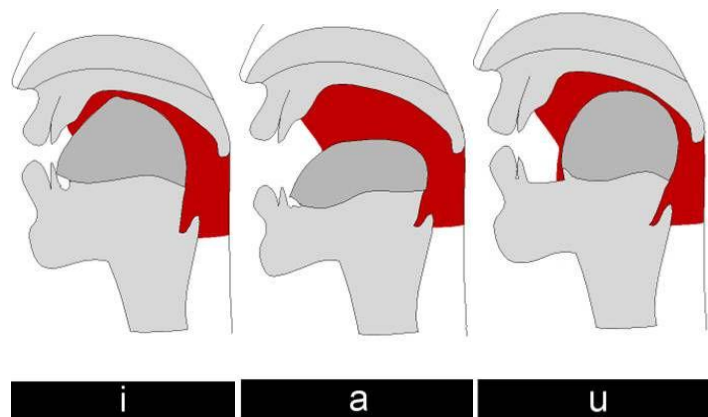


Figura 1 - Diferentes configurações do tracto vocal durante as emissões das vogais “i”, “a” e “u”.

Como já vimos, o tamanho das cordas vocais, tanto em comprimento como em espessura, desempenha um papel importante na distinção entre os vários tipos de voz, designadamente entre voz masculina e feminina. Contudo, não é apenas o tamanho das cordas vocais que determina o timbre de um cantor. Segundo o modelo Fonte-Filtro (Fant 1970), a fonação é dividida em três partes: fonte sonora, filtro acústico e radiação acústica. O ar projectado dos pulmões provoca a vibração das pregas vocais (Fonte) produzindo ondas acústicas sob a forma de som que por sua vez é filtrado pelo tracto vocal (Filtro) e projectado (Radiação Labial). Isto faz com que a fisionomia do cantor seja um dos aspectos preponderantes no Timbre do cantor, uma vez que determina as ressonâncias que actuam como filtro ao som produzido e consequentemente na

---

catalogação do tipo de voz do cantor, não no que respeita à diferenciação entre homens e mulheres mas sim, à diferenciação entre os vários tipos de vozes masculinas (ex. diferenças tímbricas entre tenor e barítono) e femininas (ex. diferenças tímbricas entre soprano e contralto). É também utilizado para estabelecer diferenças entre subclassificações do mesmo tipo de voz (ex. tenor lírico, tenor ligeiro, *spinto*, dramático etc.) e mesmo estabelecer diferenças entre essas mesmas subclassificações (ex. diferenças tímbricas entre dois tenores líricos ou entre dois sopranos ligeiros).

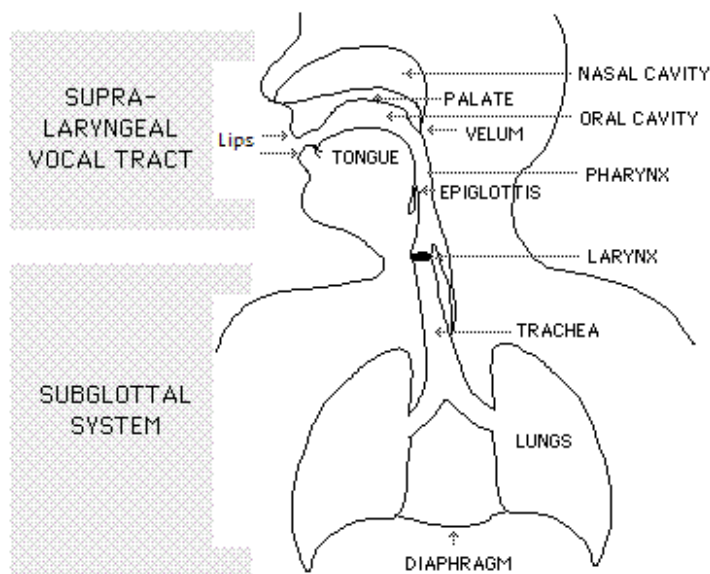


Figura 2 - O Tracto Vocal

## 2.4. Modelo Fonte-Filtro

O modelo Fonte-Filtro, proposto por Fant (1970) é um dos modelos mais comumente utilizados para descrever a produção da voz. Este modelo divide a fonação em três partes independentes sendo estas a fonte sonora, o filtro acústico e a radiação acústica. A justificação da separação entre fonte e filtro prende-se com o facto de a abertura da glote ser bastante menor do que a secção transversal das cavidades supraglóticas.

Como já foi referido, a produção da voz passa por um aumento da pressão de ar nos pulmões que passam pelas pregas vocais que vibram como resultado da interacção entre o ar e a adução das mesmas. O resultado desta interacção é o chamado som laríngeo ou

---

*voice source* que, segundo a definição de Sundberg (1987), é o som resultante da vibração das cordas vocais pelo ar vindo dos pulmões. Este fenómeno origina o som laríngeo e é então considerado a fonte. O som laríngeo produzido é a matéria-prima, que constitui a fonte primordial da voz cantada e da voz falada (Henrique, 2002). O som irá passar pelo tracto vocal onde é filtrado e modulado pelas suas cavidades e componentes traduzindo assim o tipo de sonoridade pretendida. Por fim, dá-se a projecção do som por intermédio da radiação através dos lábios, sendo estes a componente mais periférica do tracto vocal. O tracto nasal está desacoplado devido à elevação do palato e a posição do tracto vocal é determinada pela língua, lábios, mandíbula e posição horizontal da laringe. Cada vogal exige uma interacção específica entre a cavidade oral e a vibração das pregas vocais. Podemos então afirmar que o som produzido a nível das pregas vocais é diferente do som por nós percebido, uma vez que sofre diversas transformações desde que sai da fonte até à radiação labial.

Fant (1960) defende que o tracto vocal é o maior responsável pelo som que chega aos nossos ouvidos. Este modelo caracteriza os fenómenos acústicos no domínio das frequências.

## 2.5. Formantes

Tal como em outros instrumentos ressoadores, o som no tracto vocal é amplificado por ressonâncias adjacentes ao ressoador, neste caso o tracto vocal. Neste caso, é a forma do mesmo que possibilita a percepção de diferentes sons.

O tracto vocal apresenta então quatro ou cinco importantes ressonâncias que originam picos no espectro de frequências que correspondem ao modo normal dos tubos acústicos para as diferentes vogais. Estas ressonâncias F1, F2, F3, F4 e F5 são chamados de formantes e são responsáveis pela percepção de sons distintos, nomeadamente as vogais (Henrique 2002). Os três primeiros formantes têm menor dependência com o locutor e prestam-se, principalmente, para diferenciar as vogais <i>, <a> e <u>. Os formantes superiores (F4, F5, etc.), por outro lado, têm menor conteúdo linguístico e maior variação com o locutor. Acusticamente, os formantes amplificam selectivamente os harmónicos gerados pela vibração das pregas vocais, ou seja, pelo som laríngeo. Esta amplificação selectiva de harmónicos da frequência fundamental F0 é feita através de

modelações da forma do tracto vocal e são estas as responsáveis pela percepção das diferentes vogais. As transformações na forma do tracto vocal podem incluir modificações na posição da língua, lábios, dentes e palato mole. Como podemos ver na figura 3 para vogais diferentes com iguais níveis de F0, há um diferente posicionamento dos formantes ao longo do espectro, o que corresponde a diferenças na forma do tracto vocal.

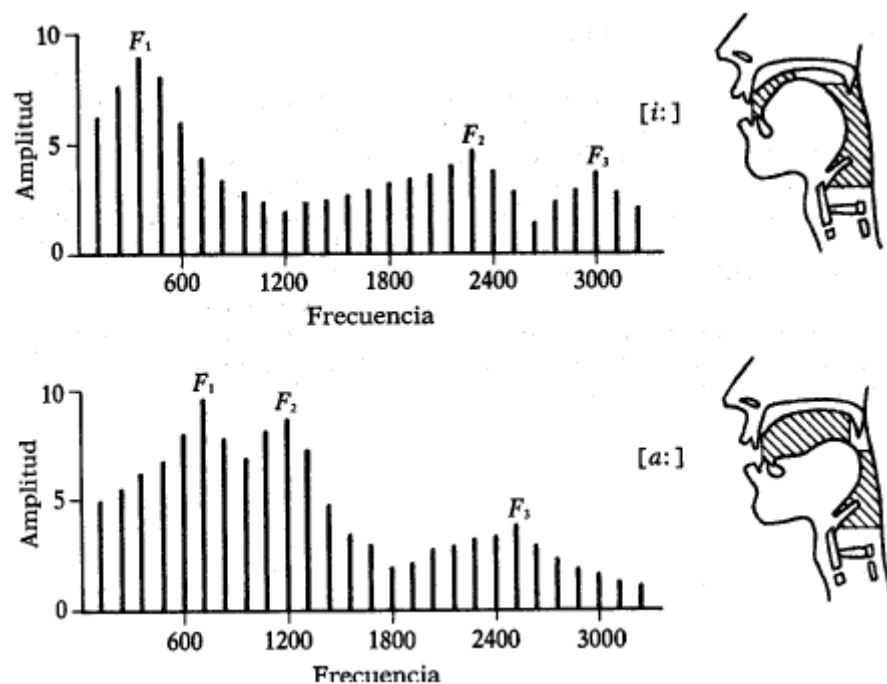


Figura 3 – Configuração do tracto vocal e distribuição de formantes para as vogais <i> e <a>.

Como podemos observar, há uma maior separação dos formantes na vogal <i> relativamente à vogal <a>, bem como uma forma diferente no tracto vocal para ambas as vogais. No caso do tracto vocal “neutro”, a nível spectral verifica-se uma separação igual entre todos os formantes.

Acontece que, na voz cantada, como os níveis de frequência fundamental são bastante superiores aos utilizados no contexto da voz falada, há uma necessidade de adaptação

---

dos diferentes formantes à frequência fundamental emitida, com o objectivo de a vogal cantada ser perceptível.

No caso do canto lírico feminino, Fant (1986) defende um modelo de ajustamento do primeiro formante à frequência fundamental emitida conhecido por Sintonia F0-F1 para a emissão da voz cantada no caso das cantoras. Este modelo pode também ser conhecido como “afinação dos formantes”. Sendo F0 a frequência fundamental, o que o cantor faz é ajustar os formantes até à frequência fundamental desejada. Esta mudança faz-se através de alterações à forma do tracto vocal, por outras palavras, alterando o Filtro no modelo Fonte-Filtro descrito pelo mesmo autor. Vejamos o seguinte exemplo: uma cantora emite uma nota em F0=500Hz, como podemos observar na figura 4, está entre o primeiro e o segundo formante.

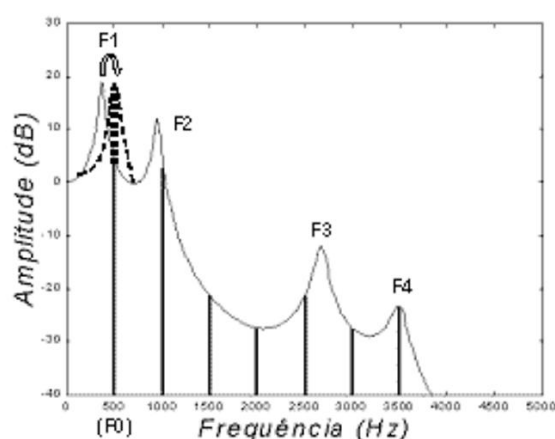


Figura 4 – Representação gráfica do fenómeno Sintonia F0-F1 também chamado de “afinação dos formantes”.

O que acontece, neste caso, é um ajustamento do primeiro formante, F1, à frequência desejada, F0, através do abaixamento da mandíbula. Com isso, há um reforço de ressonâncias em F0 pela sua sobreposição com F1 e a voz ganha audibilidade.

A frequência dos formantes pode ser alterada com a intervenção dos chamados órgãos articuladores, sendo estes os lábios, maxilar inferior, a língua, a laringe e o palato mole. Como o próprio nome indica, este conjunto de órgãos é responsável pelo processo de articulação, pelo qual se processam as modificações da forma do tracto vocal que se traduzem depois em fenómenos acústicos e perceptivos.

Sundberg (1974) classifica este fenómeno como afinação de formantes. Segundo este, a afinação dos formantes traz consequências ao nível da percepção das vogais. As notas

---

cuja frequência fundamental mais se distancia da frequência fundamental da fala, ou seja, as notas mais agudas, são executadas com uma grande abertura do maxilar inferior originando a subida do primeiro formante como foi mostrado anteriormente. De facto, quando as vozes de natureza aguda (tenores e sopranos) emitem notas de frequência fundamental muito elevada, as vogais aproximam-se todas da vogal <a>, o que é particularmente notório no caso das sopranos. A percepção da vogal é então tanto menor quanto maior for a frequência fundamental emitida. Alguns autores referem que os tenores podem evitar esta afinação do formante de modo a preservar uma sonoridade “masculina”.

## 2.6. Frequência Fundamental

A frequência fundamental é a frequência mais baixa de uma onda sonora periódica, sendo muitas vezes alvo da designação “tom natural” da voz ou ainda “o primeiro harmónico”, pois como o próprio nome indica, é o primeiro da série harmónica de um som. Apesar do som ser constituído por vários harmónicos (*overtones*), a frequência fundamental (F0) é a mais proeminente auditivamente, sendo portanto a responsável pela percepção da altura (em Hz) do som emitido. A frequência fundamental está directamente relacionada com a percepção da altura (*pitch*) do som.

A frequência fundamental da fala corresponde ao inverso do período fundamental que é o intervalo de tempo relativo a um ciclo de abertura das cordas vocais aquando do processo de fonação. Esta frequência, ou seja, o tom natural de um indivíduo na emissão da voz cantada, depende do comprimento das pregas vocais e da sua massa modal, características ligadas à componente fisiológica da laringe. Portanto, diferenças entre as frequências fundamentais da fala entre indivíduos do sexo masculino e feminino e crianças, resultam fundamentalmente de comprimentos diferentes das pregas vocais.

Quando se dá a fonação na voz cantada, a gama de frequências fundamentais é, como já foi referido, normalmente superior à voz falada. A gama de frequências na voz falada está contida no intervalo de 80 a 400 Hz enquanto que na voz cantada, podem ir até 1024 Hz no caso da voz de soprano (Vieira, 2005).

---

A variação da frequência fundamental na voz cantada é controlada pelo cantor, e está associada a fenómenos estritamente musicais, ao passo que na voz falada está sobretudo associada a fenómenos emocionais (Henrique 2002).

## 2.7. Voz Cantada vs. Voz falada: Principais diferenças

Existem diferenças significativas no que respeita à voz cantada, relativamente à voz falada. Embora sejam duas formas de utilizar o mesmo aparelho vocal, existem diferenças perceptivas inerentes à audição que são facilmente detectáveis, havendo inclusivamente explicações do ponto de vista acústico que as comprovam.

Na voz cantada, a fonação é mais sustentada e a separação entre os parciais harmónicos é maior. A duração dos fonemas das vogais é também maior na voz cantada bem como a intensidade (*loudness*).

Quanto à altura do som (*pitch*), este é normalmente superior na voz cantada, embora possam acontecer casos em que, a escrita musical induza o cantor a produzir uma frequência fundamental abaixo da que normalmente utiliza na fala.

A componente musical pode também significar uma diferença entre as duas formas de utilizar o aparelho vocal, no sentido em que, tanto a nível de *pitch*, de *loudness* ou mesmo de ressonância, há a introdução da percepção da musicalidade que, normalmente não se encontra no discurso falado de forma tão acentuada.

## 2.8. Formante de Cantor

Como já foi referido, as notas emitidas por um cantor, situam-se tipicamente em frequências muito mais altas do que as utilizadas como frequência fundamental da fala. Neste sentido, de modo a fazer a voz ouvir-se e sobrepor-se ao som de uma orquestra, sem recurso a amplificação adicional e sem prejuízo para o aparelho fonatório, uma das batalhas mais travadas na didáctica do canto é a aquisição e controlo do chamado formante do cantor ou formante extra, designação proposta pelo cientista

---

sueco Johan Sundberg (1974). O formante de cantor promove um reforço energético nas frequências de 2000, 3000 e 4000Hz, reforço este que é demonstrado no espectro acústico pela junção dos formantes F3, F4 e F5. Estes formantes (ressonâncias) quando aglutinados, fornecem então um reforço de frequências a que chamamos formante do cantor. A frequência central do formante de cantor varia conforme a vogal emitida e/ou a altura do som (*pitch*). O nível de formante de cantor pode ser influenciado de acordo com a variação da frequência fundamental (som laríngeo, a vogal que é emitida, o modo de fonação ou a intensidade vocal (Henrique, 2002). A voz apresenta um pico espectral na região dos 3000Hz para as vozes masculinas e femininas graves, e 3600Hz para vozes femininas agudas, sendo este valor não consensual (Gusmão, 2010). Este pico espectral eleva a amplitude relativa dos harmônicos aí situados, destacando a voz sobre o som da orquestra (Figura 5). Segundo Sundberg (1974), a produção deste pico, denominado formante do cantor, está relacionada com o abaixamento da laringe e o alargamento da cavidade faríngea embora haja outros mecanismos existentes como a elevação do palato mole, cuja sustentação, não se faz sem apoio do diafragma. Segundo estudos de Iwarsson (1998), através de um comportamento respiratório adequado, o fenómeno de fonação pode melhorar consideravelmente. Este reforço de ressonâncias é maior no caso dos tenores, uma vez que, sendo a frequência fundamental maior, a área espectral do formante de cantor é também superior (de 3000 a 3800 Hz) enquanto que nos baixos, os valores variam entre 2300 a 3000 Hz. No caso dos contratenores, devido à produção vocal através da técnica de falsete, os níveis de ressonância são menores. Estes níveis, geralmente não variam muito com a fonação das diferentes vogais, contudo verifica-se que aumentam ligeiramente na produção da vogal “A” e por vezes na vogal “O” (Ekholm, 1998).

A aquisição do formante de cantor permite um grau de afinação diferente do que apenas igualar a frequência emitida a uma frequência padrão. Muitas vezes, a não elevação do palato mole causa desafinação, o que faz que não seja uma questão meramente auditiva mas também de natureza técnica. A falta de apoio do diafragma é muitas vezes a causa principal, mas podem ocorrer outras como tensões no maxilar ou na garganta ou um excessivo levantamento da laringe no caso de um mau ataque da nota. Podemos então tirar a conclusão de que a afinação, não é apenas um parâmetro de qualidade mas também um parâmetro perceptivo.

---

Nos cantores, verifica-se que há uma maior variação da pressão subglótica, variação esta que está associada à intensidade e frequência fundamental (Sundberg 1995). O formante do cantor induz uma pressão subglotal acrescida, devido ao facto de as frequências emitidas no canto serem bastante superiores relativamente à frequência fundamental da fala. A pressão subglotal varia consoante a frequência fundamental, a dinâmica pretendida (dB) e as características vocais de cada indivíduo (o que acontece entre duas vozes similares). Tenores e contratenores apresentam os dois extremos opostos no que respeita aos níveis da pressão subglotal relativamente às outras vozes, sendo muito maior no caso dos tenores (cerca de 50 cm H<sub>2</sub>O) e bastante menor no caso dos contratenores (cerca de 15 cm H<sub>2</sub>O) (Henrique, 2002).

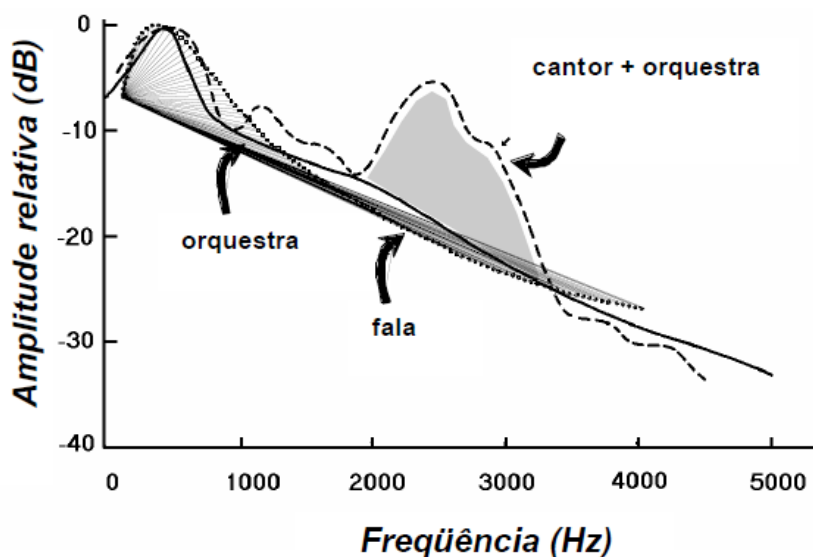


Figura 5 – Espectro de longa duração do som de uma orquestra com e sem cantor e de conversação normal.

## 2.9. Vozeamento

Outro aspecto preponderante em termos de diferenças entre a voz cantada e a voz falada é o facto de, na voz cantada, assistirmos a um maior número de sons vozeados relativamente à voz falada. Se na voz falada temos uma percentagem de 60% de sons

---

vozeados, esta pode ascender aos 95% no caso da voz falada (Loscos, 2007). Este processo deve-se à maior utilização dos órgãos articuladores e consequentemente das ressonâncias. Por outro lado, pode haver a modificação de algumas vogais, voluntaria ou involuntariamente, no decorrer da emissão da voz cantada.

## 2.10. Percepção da Voz Cantada

Muitos aspectos qualitativos da voz cantada são estudados. Porém, a definição da qualidade da voz é uma tarefa difícil, devido ao facto de a mesma estar associada à sensação e percepção auditiva por parte do professor de canto. A qualidade é um aspecto multidimensional e a medição da mesma, não se prende com aspectos como o *loudness* ou o *pitch*, sendo na maioria das vezes comparada com outro som de referência. Alguns investigadores propõem um determinado número de parâmetros para avaliação de vozes patológicas como é o caso dos parâmetros GIRBAS (Grade, Instability, Roughness, Breathiness, Asthenia, e Strain ou em português Disfonia, Instabilidade, Rouquidão, Soprosidade, Astenia e Tensão), utilizados para a medição da qualidade na emissão de voz falada (Ferreira 2008).

No caso da voz cantada, como já foi dito, a avaliação depende do professor de canto ou orientador que, além de avaliar a qualidade de voz do executante tendo como referência um som considerado como sendo “o ideal”, deve ter em conta o contexto quer espacial, quer estritamente musical onde a mesma voz é produzida. Além disto, as dificuldades técnicas ou musicais apresentadas pelo estudante de canto, são identificadas pelo professor de canto com base no tipo de sonoridade a que estão associadas, podendo estas ser mais ou menos detectáveis. Ora este tipo de avaliação baseado na percepção do som emitido pelo cantor, é subjectiva e pouco precisa.

Em todo o caso podemos dizer que a avaliação da qualidade vocal, quer no caso da voz cantada, quer no caso da voz falada, é resultante de uma interacção entre o professor de canto/ouvinte e um sinal acústico emitido pela voz do executante.

Os estudos efectuados com base na percepção, prendem-se muito com aspetos e estilos musicais enquanto que os que visam a classificação de diferentes tipos de voz e suas características acústicas, são escassos. Em todo o caso encontramos estudos que

---

correlacionam diferentes classificações da voz com o *pitch* (Tessitura), presença ou não do formante de cantor, e em alguns casos até com o vibrato, apesar de este não ser considerado uma característica exclusivamente perceptiva, uma vez que varia em função da altura do som.

Encontramos também estudos como os realizados por Sundberg (1991), que refere que os níveis de pressão subglotal e adução das pregas vocais, é inversamente proporcional à quantidade de ar expelido na produção da voz cantada, relacionando este *ratio* com diversos estilos musicais tais como Jazz, Pop ou o Canto Lírico.

São utilizadas também certas subclassificações (McCoy, 2004) entre os vários tipos de voz, tais como lírico, dramático, ligeiro, *spinto*, entre outros, tendo estas subclassificações fortes correlações com o *pitch*, embora a sua natureza seja também uma questão tímbrica.

Outras características perceptivas são apontadas, algumas delas em forma de pares antagónicos, como por exemplo claro/escuro ou suavidade/aspereza referindo inclusivamente que uma voz não é exclusivamente “clara” ou “escura”, misturando características referentes a ambas as classificações. Segundo McCoy (2004), uma voz clara está associada à sensação de brilho e poder, sendo normalmente rica em harmónicos na região das altas frequências (*high-pitched overtones*) ao passo que uma voz escura, está associada à sensação de calor e rica em harmónicos de baixo *pitch* (*low-pitched overtones*).

No que respeita à produção do som, são utilizados termos como sons “frontais” ou “posteriores”, definições estas que assentam na percepção da sua produção. Sons posteriores dão a sensação de serem produzidos nas regiões mais recuadas do tracto vocal, como a raiz da língua, ao passo que os sons frontais são produzidos nas regiões mais periféricas, como os lábios ou a ponta da língua. No mesmo estudo, McCoy (2004) refere ainda que estes dois tipos de sons dão origem a sonoridades mais escuras, no caso dos sons posteriores e mais claras no caso dos sons frontais. A existência de sons mais escuros ou mais claros tem também uma forte correlação com as vogais emitidas.

Podemos também encontrar definições que assentam na existência ou não de um timbre nasal da voz cantada. A existência da voz nasalada, deve-se ao facto de haver uma ressonância extra, o que acontece quando a comunicação entre a cavidade bucal e nasal não se encontra totalmente fechada. Sobre este tema, existem autores que defendem a sua utilização para uma correcta emissão da voz cantada, enquanto outros

---

rejeitam a sua existência, sendo considerada não só como uma falta de sentido estético como também um déficit de técnica.

A existência de uma certa quantidade de ruído ou ar na voz aquando da sua emissão, é também alvo de caracterização e definição. Podemos caracterizar uma voz como sendo mais ou menos “soprosa”, ou seja, pela existência de uma certa quantidade de ar na sua emissão, o que acontece quando a adução das pregas vocais é deficiente ou incompleta. Este fenómeno, tal com a rouquidão, está normalmente associado a uma patologia vocal, embora esteticamente possa ser introduzido numa performance vocal com sentido puramente interpretativo.

## 2.11. Parâmetros Qualitativos e Perceptivos

### 2.11.1. Introdução

Nesta secção procura-se fazer uma abordagem acerca dos vários parâmetros e atributos musicais/estilísticos de natureza perceptiva, e que muitas vezes são entendidos como subjectivos para o público menos habituado às questões da voz cantada. Normalmente estas definições são utilizadas no contexto de uma aula de canto, e acompanham os cantores e orientadores durante todo o percurso artístico. Porém, e como tratamos de características muito subjectivas, que dependem da capacidade interpretativa do professor de canto e da resposta do aluno, tornam-se difíceis de compreender para o público leigo.

### 2.11.2. Afinação

A afinação define-se como a capacidade de produzir um som igual, em termos de frequência a outro e está sempre associada a uma referência (ex: Lá 440 Hz). O conceito de afinação varia também com a capacidade de distinguir as ditas frequências. A afinação varia igualmente com questões de natureza cultural. No caso da tradição europeia, é importante referir o padrão utilizado (seja este uma escala ou um modo). Na

---

escala igualmente temperada, a escala mais comumente utilizada, todos os meios-tons são idênticos. A relação entre a frequência de uma nota e de outra meio-tom acima é de  $1:2^{1/12}$ . Portanto, se multiplicarmos uma determinada frequência, por exemplo Lá 440Hz por  $2^{1/12}$  obtemos 466,163Hz correspondente ao Lá# acima do Dó central.

### 2.11.3. Tessitura

Designa-se por Tessitura, a zona de frequências emitidas que é confortável para um cantor. Neste conjunto de frequências, a voz é produzida sem esforço e com todas as suas qualidades tímbricas presentes.

A tessitura pode ser representativa do tipo de voz, existindo um padrão de alcance de frequências graves e agudas para cada classificação. Como é sabido, o tamanho, grossura e elasticidade das cordas vocais, influenciam e definem as características da voz de determinado indivíduo. Neste sentido, os elementos do sexo feminino, por possuírem pregas vocais de menor tamanho, emitem níveis mais altos de frequência fundamental na fala (cerca de 220 Hz), ao passo que os elementos do sexo masculino, na fala, emitem valores mais baixos de frequência fundamental (cerca de 120 Hz) (Vieira2005). Estes valores, no canto, podem ir até 1047 Hz, no caso das mulheres (sopranos) e 523 Hz no caso dos homens (tenores) (Gráfico 1). Nas Tabelas 1 e 2 vemos como podem ser classificadas as vozes masculinas e femininas tendo em conta a sua tessitura.

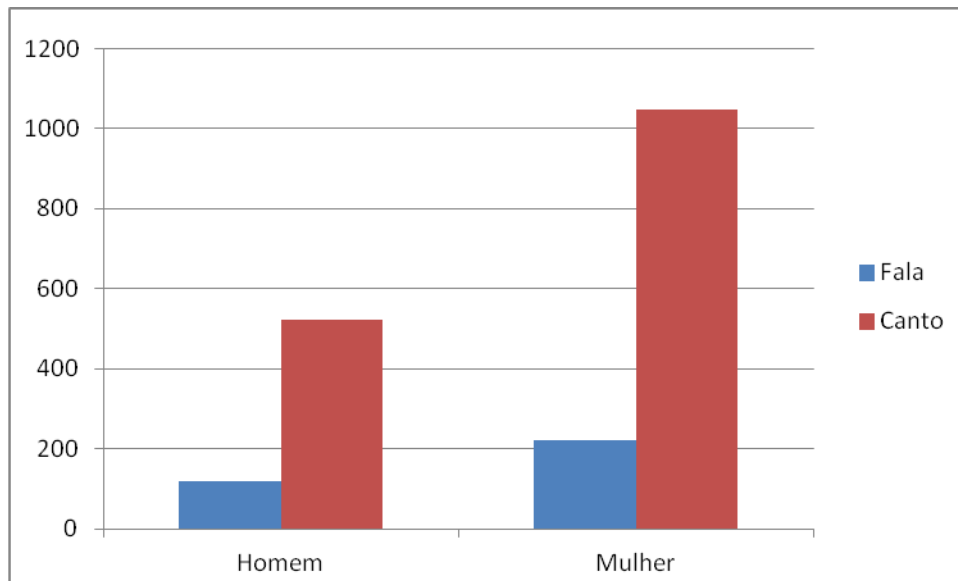


Gráfico 1 – Frequências fundamentais relativas a fala e canto entre indivíduos do sexo masculino e feminino.

Si	Dó	Ré	Mi	Fá	Sol	Lá	Si	Dó	Ré	Mi	Fá	Sol	Lá	Si	Dó	Ré	Mi	Fá	Sol	Lá	Si	Dó	Ré	Mi	
1	2	2	2	2	2	2	2	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	5	5	5
<b>Baixo Profundo – Ex. Ivan Rebroff</b>																									
		<b>Baixo – Ex. Kurt Moll</b>																							
				<b>Barítono – Ex. Dietrich Fischer-Dieskau</b>																					
								<b>Tenor Dramático/Wagneriano - Ex. James King</b>																	
								<b>Tenor Spinto – Ben Heppner</b>																	
								<b>Tenor Lírico – Ex. Fritz Wunderlich</b>																	
								<b>Tenor Ligeiro – Ex. Luigi Alva</b>																	

Tabela 1 – Tabela de catalogação de tipos de voz masculinas consoante a tessitura (*Fach*).

Mi	Fá	Sol	Lá	Si	Dó	Ré	Mi	Fá	Sol	Lá	Si	Dó	Ré	Mi	Fá	Sol	Lá	Si	Dó	Ré	Mi	Fá
3	3	3	3	3	4	4	4	4	4	4	4	5	5	5	5	5	5	5	6	6	6	6
<b>Contralto/Contratenor (Masc.) – Ex. Marian Anderson</b>																						
(Contralto); Andreas Scholl (Contratenor)																						
		<b>Mezzo-Soprano – Ex. Christa Ludwig</b>																				
		<b>Soprano Dramático – Ex. Birgit Nilson</b>																				
				<b>Soprano Spinto – Ex. Leontyne Price</b>																		
				<b>Soprano Lírico – Montserrat Caballé</b>																		
				<b>Soprano Ligeiro (Soubrette) – Kathleen Battle</b>																		
				<b>Soprano Coloratura – Ex. Editha Gruberova</b>																		

Tabela 2 – Tabela de catalogação de tipos de voz femininas consoante a tessitura (*Fach*).

Este padrão não obsta porém que haja variabilidade, ou seja, determinadas vozes podem atingir ou mesmo não atingir, frequências mais graves e/ou mais agudas das que estão explícitas no padrão. No que respeita ao repertório abrangido por determinadas vozes, há vozes que possuem características especiais, que as dotam da capacidade de executar determinado repertório que, em teoria, está destinado a outro tipo de voz. Por exemplo um tenor lírico com características especiais, pode aspirar a papéis (não todos) de tenor ligeiro ou mesmo de tenor *spinto* ao mesmo tempo que uma mezzo-soprano pode aspirar a papéis de soprano dramático.

#### 2.11.4. Timbre

O timbre, acusticamente falando, é uma característica sonora que nos permite distinguir sons da mesma frequência e intensidade igual emitidos por diferentes fontes sonoras (Luis Henrique 2002). Resulta da correlação de todas as propriedades do som

---

que não exercem influência na altura do som (*pitch*) e na sensação de intensidade (*loudness*). Estas características podem incluir envolvente temporal, distribuição de energia espectral ou grau de inarmonicidade dos parciais e frequência. Na emissão vocal, o tracto vocal do cantor comporta-se como um filtro às ondas sonoras produzidas a nível das pregas vocais, o que nos faz concluir que estas são a fonte do som produzido, ou seja, do som laríngeo. O timbre está então directamente relacionado com a fisionomia do cantor, ou seja, a forma do tracto vocal.

O termo “voz timbrada” é associado a vários aspectos qualitativos que podem coexistir ou não na mesma voz, podendo ser definidos através de pares com termos antagónicos. Assim sendo, definimos então os seguintes pares: claro/escuro; voz na frente/voz recuada; leve/pesada; limpidez/soprosidade; suavidade/aspereza.

#### 2.11.4.1. **Claro/Escuro**

O termo vem do italiano *chiaroscuro*, expressão utilizada para descrever a técnica de pintura de Leonardo da Vinci (1452 – 1519). Em termos acústicos, estas características variam consoante a proeminência de baixas ou altas frequências (*overtones*). Então, uma voz clara possui um reforço nas frequências agudas, enquanto que uma voz escura possui um reforço nas frequências graves. Como já foi dito, o facto de haver vozes que se encaixam mais facilmente numa das classificações, não impede que possuam características que pertençam a outra. Uma voz pode então ser classificada como clara ou escura para cada um dos géneros, como se exemplifica na Tabela 3 com cantores reais.

Classificação	Cantor Exemplo
Masculina Clara	Luigi Alva
Masculina Escura	Jonas Kaufmann
Feminina Clara	Lucia Popp
Feminina Escura	Jessey Norman

Tabela 3 – Exemplos de vozes masculinas e femininas para os parâmetros Claro e Escuro.

---

#### 2.11.4.2. **Voz na Frente/Voz Recuada**

Este par de termos resulta da forma como, tecnicamente é feita a projecção da voz. Uma voz na frente (sensação que os cantores descrevem como “voz na máscara” precisamente por explorar os seios nasais) possui mais nitidez do que uma voz mais recuada. Pode entender-se por nitidez como havendo uma maior percepção das características vocais e articulação do texto durante a emissão, tanto no aspecto tímbrico como dinâmico. Pelo contrário, uma voz recuada em demasia dá a sensação de um aperto na garganta, e de uma voz baça e sem perspectiva de projecção, resultante da fraca exploração dos seios nasais.

A voz na frente é por vezes confundida, pelos alunos, nos primeiros anos de aprendizagem com a voz nasalada. A voz nasalada possui menos projecção e consequentemente, menos capacidade de se ouvir por cima do som de uma orquestra. Ao contrário da voz na máscara, a voz nasalada encontra-se focada só num ponto.

#### 2.11.4.3. **Voz Leve/Voz Pesada (ou Repertório Leve/Pesado)**

Trata-se de uma grande batalha para os professores de canto e/ou pedagogos no ensino do canto. Em termos técnicos, uma voz pesada é uma voz pouco maleável e menos ágil, sendo portando pouco propícia a flutuações de dinâmica e movimentos que exijam destreza, como por exemplo a coloratura. Pelo contrário, uma voz leve é maleável e executa com facilidade os movimentos referidos. Vozes mais pesadas produzem geralmente mais som do que vozes mais leves. Isto faz com que o tipo de repertório realizado esteja intrinsecamente ligado ao tipo de voz de um determinado cantor (ex. vozes pesadas não são aptas para papeis de coloratura). Na nossa opinião, faz portanto mais sentido falar em repertório leve e pesado, do que em voz leve e voz pesada uma vez que a voz determina o tipo de repertório adequado ao cantor e não o contrário, querendo isto dizer que é incorrecto modificar ou adaptar a voz a diferentes tipos de repertório. É de salientar que nenhum dos parâmetros é melhor que o outro mas sim diferente, o que faz com que a escolha de repertório seja alvo de uma análise minuciosa consoante as características vocais do indivíduo.

---

#### 2.11.4.4. **Limpidez/Soprosidade**

A soprosidade está relacionada com a quantidade de ar na emissão vocal. Esta resulta de uma fenda glotal (i.e. um mau contacto das cordas vocais) e quando não associada a fenómenos patológicos, deve-se a uma deficiência de suporte respiratório. É raro encontrar num cantor lírico e quando existe é considerado uma lacuna técnica. Contudo pode ser encontrada noutros estilos como *Jazz* ou canto ligeiro e etnográfico.

#### 2.11.4.5. **Suavidade/Aspereza**

A aspereza é definida como a quantidade de “ruído” na emissão vocal. Pode ser causado por fenómenos fisiológicos e/ou patológicos e podem ir desde a rouquidão até, num caso mais extremo, à afonia. Alguns cantores induzem um certo nível de Aspereza consoante a interpretação do texto, por exemplo numa cena de Ópera que assim o exija.

#### 2.11.5. **Falsete/Voz de Cabeça**

O Falsete (que deriva do termo *falsetto*) é produzido através da vibração de uma fracção das pregas vocais. Normalmente o termo falsete é associado à voz masculina sendo o termo antagónico o registo modal ou de peito (*modal, chest* ou *heavy*). No caso da voz feminina o termo utilizado é voz de cabeça. A diferença entre estes registos reside nas diferentes formas de vibração das pregas vocais resultantes de diferentes formas de adução. Estas formas de vibração distintas dão portanto origem a diferentes tipos de som laríngeo. O falsete traduz-se no registo utilizado pelo cantor na fonação de níveis mais elevados de frequência fundamental de um modo não natural (e por isso falso, *Falsetto = tom falso*). O resultado é um registo mais leve e suave que contrasta com o registo mais encorpado e audível que é o registo de “peito”.

Difere da voz modal no que respeita à posição da laringe, sendo portanto impossível uma combinação entre estas duas formas vocais. De acordo com Van Den Berg (1980), o registo modal é caracterizado por uma tensão longitudinal acumulada nos músculos vocais, ao passo que o registo de falsete encontra-se uma forte tensão longitudinal nos ligamentos vocais. A regulação da tensão efectiva das pregas vocais é

---

feita pelos músculos tiroaritenóides que criam uma estrutura mais espessa e profunda dando origem ao registo modal (Henrique 2002). As pregas vocais no registo de falsete estão mais finas e esticadas sendo menor a área de contacto portanto a fase em que estão fechadas é menor relativamente ao registo modal. Assim sendo, as pregas vocais vibram menos no registo de falsete do que no registo modal pois estão mais tensas.

O Falsete apresenta um dispêndio superior de energia, uma vez que apenas parte das cordas vocais se juntam para proceder à fonação. A amplitude de fonação é portanto pequena e pouco passível de ser aumentada, sendo também de frequência fundamental elevada. Acusticamente a voz possui uma carência de harmónicos (sobretudo na região das altas-frequências) devido a uma falta de vigor na utilização dos formantes. Isto traduz-se numa diferença tímbrica acentuada entre os registos de falsete e modal (Castellengo 1985).

É de realçar que, na voz cantada, todos estes aspectos são potencializados em qualidades e direccionados para um determinado repertório. A voz de contratenor é muitas vezes entendida como uma técnica ou um registo de falsete embora outros autores como Sundberg e Hogset (2001) a possam considerar como um registo diferente do falsete.

A utilização do falsete acarreta predominantemente uma fenda glótica, o que se traduz numa componente significativa de ruído (soprosidade). No canto, esta soprosidade é suprimida através da técnica.

A passagem de um registo para outro é designada quebra de registo podendo ser voluntária ou involuntária.

## 2.11.6. Ataque

Podemos definir o ataque como o começo da emissão do som numa frase vocal. Um bom posicionamento dos vários componentes do tracto vocal é determinante para a qualidade do ataque. A formação do ataque envolve todos os componentes do tracto vocal, desde a garganta e a laringe até à posição da língua e lábios, sendo suportado por uma forte componente de apoio respiratório. Uma posição correcta do tracto vocal varia necessariamente com a morfologia de cada indivíduo, por exemplo, uma correcta

---

posição do maxilar não é necessariamente a mesma para indivíduos de morfologias diferentes.

Um aspecto ligado intrinsecamente com o ataque é a articulação do texto. A qualidade da articulação do texto prende-se com a forma correcta do tracto vocal e o suporte do apoio respiratório. A junção destes dois componentes dá uma facilidade na emissão das vogais e consoantes.

### 2.11.7. *Vibrato*

O *Vibrato* é definido como uma série de modulações periódicas de altura de som (variações na frequência fundamental de fonação), intensidade e espectrais que, podendo ser combinadas ou não, influenciam o som produzido e o próprio timbre. É uma consequência da projecção vocal e portanto uma voz com *vibrato* é uma voz timbrada (Garnier, 2004). O *vibrato* é quase sempre desenvolvido sem que os cantores pensem em adquiri-lo activamente (Björklund, 1961). A frequência do vibrato é geralmente considerada constante para cada cantor e é muito difícil ou mesmo impossível de ser alterada através de treino. Pode ser considerado tanto um atributo perceptual como um atributo de qualidade. O *vibrato* pode ser medido através da extensão (no caso das modulações de frequência, em semi-tons), através da taxa (nº de ciclos por segundo ou Hz, sendo o normal entre 5,5 e 7,5 Hz) e através da forma tendo este ultimo parâmetro sido alvo de pouca atenção por parte da comunidade científica. No entanto, alguns cientistas propõem a classificação de sinusoidal, triangular, trapezoidal e não-identificável. Quando a taxa (nº de ciclos por segundo) é muito elevada (superior a 7,5 ciclos/s), o *vibrato* passa a ter a designação de trémolo (Bunch, 1982). No entanto, alguns autores defendem uma distinção entre o vibrato e o trémolo, sendo este último uma modulação da amplitude de frequência de 5 a 8 ciclos/s. Segundo Sundberg (1987), vibratos com taxa inferior a 5,5 ciclos/s são considerados demasiado lentos, ao passo que os que possuem uma taxa superior a 7,5 ciclos/s são considerados “nervosos”. A taxa desejável para o *vibrato* anda à volta dos 6 ciclos/s (Hall 1991).

No canto, o *vibrato* é produzido pela modulação da tensão da musculatura da laringe responsável pela tensão e endurecimento das pregas vocais. Aumenta ao longo das notas sustentadas e muitas vezes, camufla eventuais desacertos de afinação. A amplitude de frequência está directamente relacionado com a altura do som (em dB).

---

O *vibrato* depende de vários factores, como sexo, idade, características fisiológicas do cantor e envolvimento emocional. Sendo também um parâmetro de qualidade, está naturalmente sujeito às condicionantes estéticas da obra a interpretar, ou seja, enquanto numa obra renascentista é utilizado apenas como ornamento, numa obra mais romântica é um atributo musical essencial no campo da expressividade. Trata-se de uma característica bastante importante e quase essencial nas vozes de ópera e *lied* da cultura musical ocidental (Prame, 1997).

Usualmente designa-se por Voz Branca, uma voz de uma criança antes do início da puberdade. Trata-se de um timbre puro e cristalino que faz da ausência de *vibrato*, a sua principal característica. Em determinados estilos musicais como a música renascentista, onde o *vibrato* é utilizado apenas como ornamento, utiliza-se a chamada “voz lisa”, que podemos entender como uma aproximação à anterior, feita por indivíduos de idade adulta.

#### 2.11.8. *Legato*

Em canto, o termo *legato* é caracterizado pela continuidade da linha vocal sem hiatos perceptivos na emissão da voz, quer na passagem para frequências mais graves ou mais agudas. A transição entre as notas é feita de forma contínua e progressiva sem interrupção da emissão do som ao inverso do *staccato*. Exige um controlo da técnica vocal com vista a manter as características do som fundamental, de modo a que o som não sofra alterações qualitativas à medida que a frase se desenrola, o que levanta alguns problemas sobretudo na articulação das consoantes do texto. Em termos artísticos, o *legato* coexiste com a articulação e prosódia do texto.

#### 2.11.9. *Staccato*

É uma oposição ao *legato* e consiste na realização de pequenas pausas entre as notas, ou seja, a emissão do som é interrompida fazendo com que as notas possuam uma duração mais curta do que se não tivessem a indicação de *staccato*. A palavra *staccato* em italiano significa destacado ou separado.

---

## 2.11.10. Microdinâmicas e Destreza Vocal

Além das terminologias básicas de *forte* e *piano* indicadas na partitura, estão atribuídas *nuances* de condução dinâmica, fraseado e intensidade inerentes a vários factores. São estes a própria estética da composição, quer seja pela prosódia do texto (presença de sílabas tónicas e construção frásica), pela construção musical, uma vez que, os compassos numa partitura possuem tempos fortes (que devem ser acentuados), fracos e meio-fortes, conforme características estilísticas e de textura musical de cada compositor. Assim sendo, conforme a peça a ser executada, diferentes microdinâmicas serão utilizadas. Muitas das vezes, a utilização das mesmas está directamente relacionada com a interpretação do executante (a forma como o intérprete sente o texto e o comunica), podendo ser, ou não, orientada pelo professor. Assim sendo, manifestações musicais como crescendos, diferentes formas de ataque e articulação, mudanças de timbre e intensidade vocal e por vezes a utilização de fenómenos de emissão mais extremos e tidos como indesejáveis como a voz com um certo grau de sopro ou aspereza, ou ainda o sussurro podem ser utilizados.

A realização das chamadas microdinâmicas requer acima de tudo, sensibilidade e alguma destreza vocal com vista a não prejudicar a qualidade do som emitido. É também necessário compreender a prosódia do texto, bem como o seu significado de modo a construir uma frase coerente não só em termos musicais mas também literários.

### 2.11.10.1. Coloratura

A coloratura exige um grande nível de agilidade vocal e virtuosismo. Define-se por coloratura, a realização de várias notas numa única sílaba, num tempo mais ou menos rápido e com saltos entre notas mais ou menos longos, consoante o indicado na partitura. A coloratura tem que ser bem articulada, com uma percepção clara das diferentes notas, geralmente feita em *legato* podendo também ser efectuada em *staccato*, quando mencionada na partitura. A dificuldade está na manutenção da estrutura do tracto vocal, que se faz à custa da sustentação pelo diafragma durante a execução (há sempre uma tendência para a subida da laringe) e nas dificuldades respiratórias que acarreta.

---

### 2.11.11. Portamento

A palavra em italiano significa “transporte” e o *portamento* é precisamente uma ligação entre duas notas feito em *slide* (deslize). Uma indicação deste tipo pode estar contida na partitura, porém, a sua utilização quando não incitada, é muitas vezes entendida como déficit de técnica e/ou com falta de sentido estético.

---

## Capítulo 3

---

## 3. Análise dos parâmetros Qualitativos e Perceptivos

---

### 3.1. Introdução

Neste capítulo procederemos à apresentação e interpretação dos resultados obtidos no desenrolar da nossa pesquisa. Foram seleccionados quatro dos parâmetros perceptivos acima referidos, cuja escolha foi efectuada criteriosamente tendo em conta os que mais evidenciam diferenças, do ponto de vista quer perceptivo, quer, como mais adiante veremos, acústico.

O principal objectivo deste capítulo e desta pesquisa será identificar os fenómenos acústicos associados aos parâmetros perceptivos analisados, e estabelecer uma diferença entre estes parâmetros, com base na existência de tais fenómenos ou na sua inexistência. No caso particular do vibrato, analisaremos as suas características e a forma como o mesmo se comporta em cada cantor, tendo em conta os vários exercícios que são realizados.

Esta parte da investigação, foi realizada em conjunto com os nossos colegas da Faculdade de Engenharia da Universidade do Porto e sempre com o supervisionamento do orientador do projecto de investigação.

---

## 3.2. Base de Dados

Para a realização da análise dos parâmetros perceptivos acima definidos, foi necessária uma base de dados que contemplasse os parâmetros descritos anteriormente assentando em exemplos de voz cantada. Foi construída uma base de dados constituída por seis Cantores pertencentes à classe de Canto da ESMAE.

As gravações foram efectuadas nos estúdios do curso de Produção e Tecnologias da Música da ESMAE e no estúdio particular de um profissional da área da produção musical e áudio.

As gravações efectuadas foram supervisionadas e aprovadas quanto à sua veracidade, por dois especialistas da área do Canto intervenientes no projecto científico.

Na construção da desta base de dados foram utilizados os seguintes aparelhos:

- Microfones: Neuman Studio Microphone TLM 103; Rode K2
- Pré-amplificadores: Onyx 800 R; Tl áudio 5051
- Software: Digital Performer 7.24; Cubase 5.5.3
- Hardware: Mac Pro 1,1 Dual-Core Intel Xeon; AsusP5Q Intel Quad 2 Core.

A título de curiosidade foi também incluída na nossa base de dados a voz da conhecida cantora Teresa Salgueiro. A análise não foi tão extensiva relativamente às outras vozes pois apenas foram analisados os parâmetros de voz plana e falsete e caracterização de vibrato não comparando com nenhum dos elementos das amostras.

## 3.3. Parâmetros escolhidos

De entre os vários parâmetros subjectivos incluídos na base de dados, por questões de logística e de tempo apenas cinco serão alvo de estudo nesta dissertação. Os parâmetros escolhidos nesta fase foram os que evidenciaram as características mais claras e definidas quer perceptíveis auditivamente para um indivíduo não especialista na área do canto, quer traduzidas em fenómenos acústicos representados visualmente.

---

Neste ponto apresentaremos os parâmetros escolhidos e os resultados obtidos na primeira análise a nível espectral e no software SingingStudio.

### 3.3.1. Portamento

O portamento existe quando a ligação entre duas notas é feita com efeito deslizante. Assim, preliminarmente foi estudada a frequência fundamental em trechos de canto com e sem portamento, tendo-se verificado que é claro quando existe ou não portamento na transição entre notas quer auditivamente quer, como mais adiante veremos, visualmente. Como se pode ver na Figura 6, uma transição de nota sem portamento é quase instantânea, enquanto que no caso de utilização de portamento na transição de nota, esta é feita mais lentamente passando por um curto período de tempo pelas notas intercalares, como se verifica na Figura 2.

Não queremos com isto dizer que não há transição entre notas nas passagens sem portamento. Acontece que, com portamento, a transição é mais demorada e passa de forma gradual por cada nota até chegar à nota final e nas passagens sem portamento as transições são de tal maneira curtas que não são reconhecidas.

O software utilizado para chegar a esta conclusão foi o SingingStudio, escolhido por ser o que melhor representa visualmente a transição entre as duas notas.

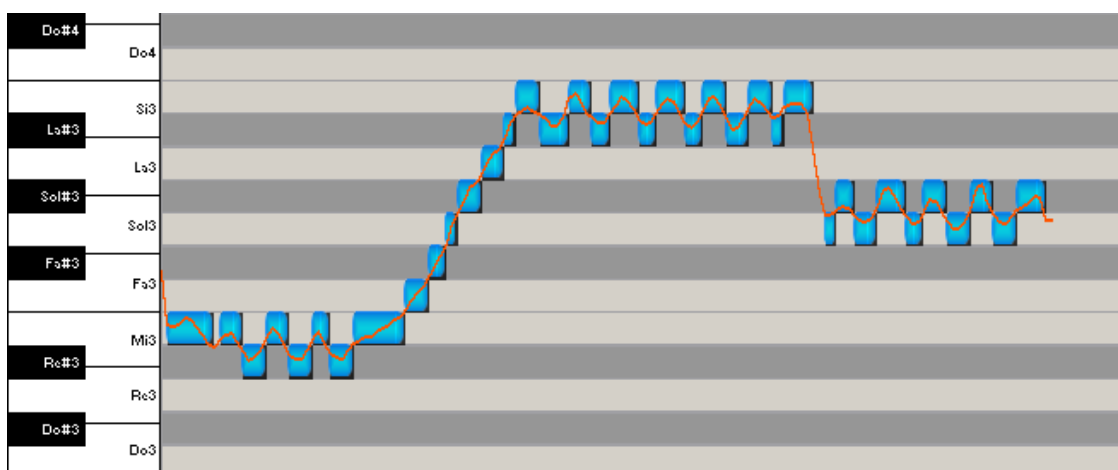


Figura 6 – Transição entre as notas Mi 3 e Si 3 feita através de portamento.

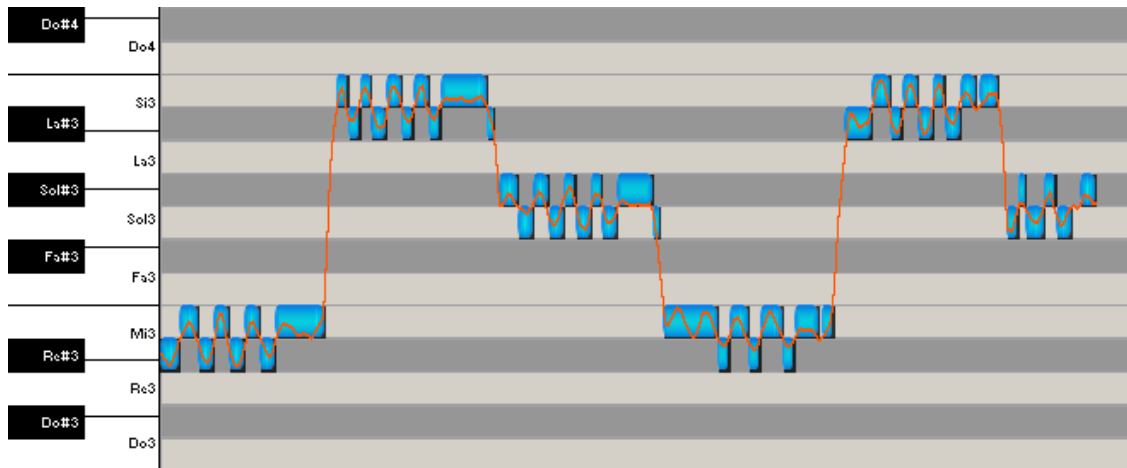


Figura 7 – Transição entre as notas Mi 3, Si 3 e Sol# 3 feitas sem portamento.

Uma outra condição acústica que se verificou foi a definição dos harmónicos na transição de nota. No caso do uso de portamento a perda de definição dos harmónicos não é tão acentuada como no caso da não utilização de portamento devido ao facto de, visto as transições serem muito pequenas, os harmónicos estarem mergulhados no ruído. Esta condição está ilustrada nas Figuras 8 e 9. Para chegar a esta conclusão foi utilizado o software MatLab®.

### **Uso de Portamento**

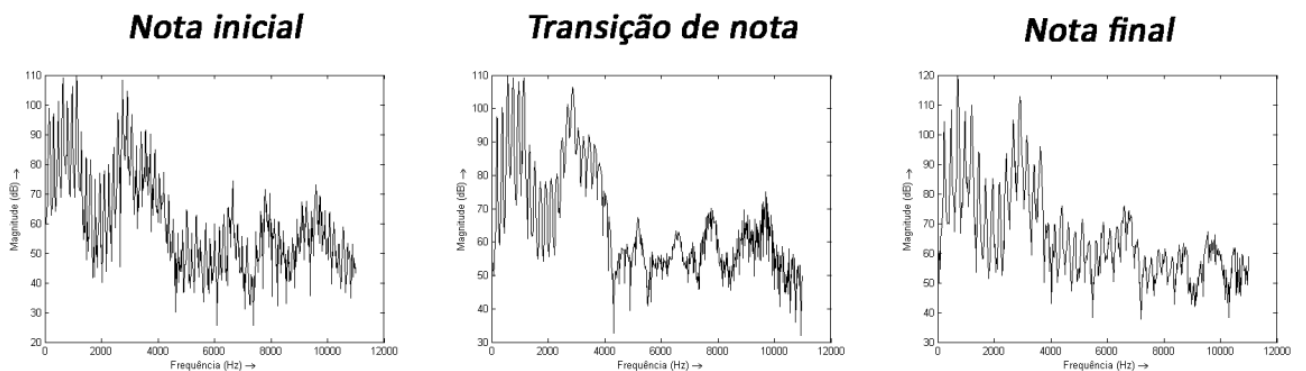


Figura 8 – Representação espectral da frequência fundamental e parciais harmónicos na transição com uso de portamento.



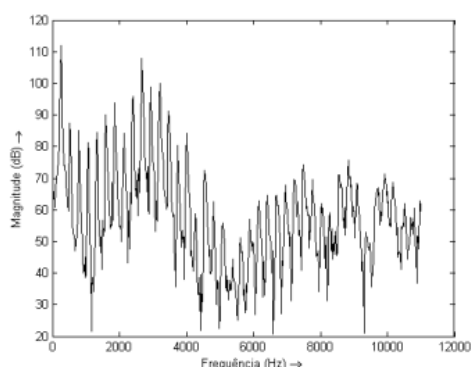
Figura 9 – Representação espectral da frequência fundamental e parciais harmônicos na transição sem uso de portamento.

### 3.3.2. Falsete/Voz de Cabeça (Falsete vs. Voz Modal)

O falsete está normalmente associado a produção de sons de mais alta frequência fundamental de uma forma não natural, sendo por isso uma técnica que necessita de muito treino. Está também associado à produção de sons não atingíveis no registo de voz modal embora tal não seja totalmente verdade.

Acusticamente está associado à perda de harmónicos nas altas frequências, e ao aumento da amplitude do harmónico fundamental em relação aos precedentes. O facto de os harmónicos existirem mas, ao invés de serem claros, estão mergulhados no ruído o que reduz o possível efeito da existência de formante de cantor. Estas características estão ilustradas na Figura 5 em voz masculina, e na Figura 6 em voz feminina.

**Voz modal  
(voz masculina)**



**Voz em falsete  
(Voz masculina)**

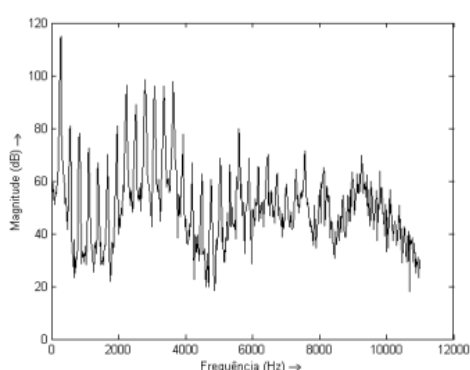
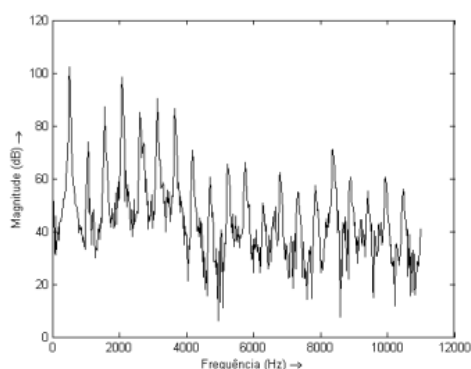


Figura 10 – Comparação entre parciais harmônicos nos registros de voz modal e registro de falsete para voz masculina.

**Voz modal  
(voz feminina)**



**Voz em falsete  
(Voz feminina)**

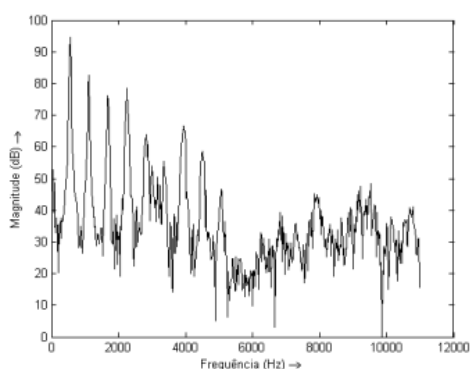


Figura 11 – Comparação entre parciais harmônicos nos registros de voz modal e registro de falsete para voz feminina.

Verifica-se claramente nas Figuras 10 e 11, a proeminência da região espectral correspondente ao formante de cantor na voz modal relativamente ao registro de falsete. Verifica-se também a dispersão dos harmônicos na região das altas frequências nomeadamente junto aos formantes 4 e 5.

---

### 3.3.3. Limpidez vs. Soprosidade

A soprosidade na voz resulta de uma fenda glótica provocando fuga de ar pelas pregas vocais e está normalmente associada a uma condição patológica. Contudo, no caso da voz cantada pode-se dever a deficiente suporte respiratório. Esteticamente pode coexistir ou não em alguns estilos de canto, sendo que no canto lírico é considerado como um déficit de técnica ao passo que em outros estilos como *Country*, *Jazz* ou *Folk* é de existência habitual.

Perceptivamente existe um ruído de fundo audível. Acusticamente denota-se a diminuição do destaque dos harmónicos em relação ao ruído no caso da existência de soprosidade como está demonstrado na Figura 7 para voz masculina e na Figura 8 para voz feminina. Os harmónicos encontram-se mergulhados no ruído e a proporção entre harmónicos e ruído é diminuída. De notar que o caso da voz feminina demonstrada na Figura 8 é um caso de grande nível de soprosidade.

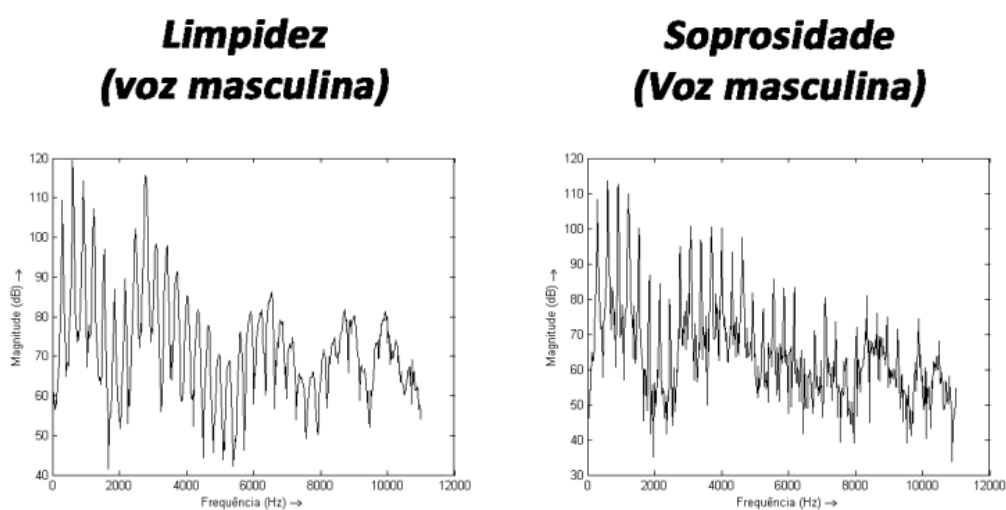
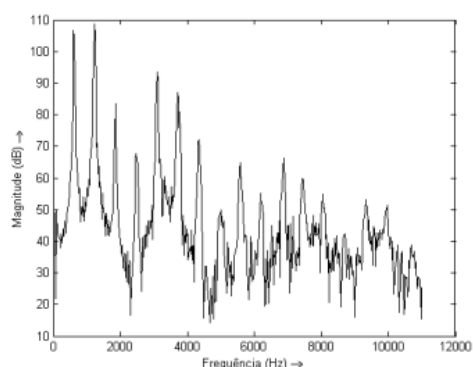


Figura 12 – Representação espectral de voz limpa e voz soprosa para voz masculina.

---

### **Limpidez (voz feminina)**



### **Soprosidade (Voz feminina)**

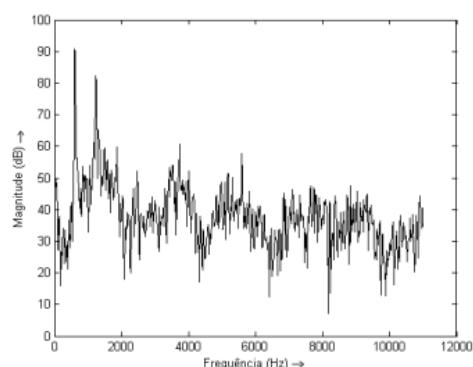


Figura 13 – Representação espectral de voz limpa e voz soprosa para voz feminina.

Denota-se claramente a perda de definição dos harmónicos nos dois casos de voz soprosa face ao exemplo de Limpidez. Esta afirmação ganha particular ênfase com o exemplo da voz feminina onde só são distinguidos praticamente o primeiro e segundo harmónico estando os restantes, completamente mergulhados no ruído.

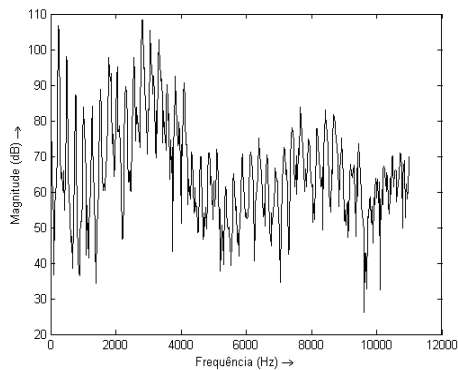
#### 3.3.4. Formante de Cantor vs. Voz Plana

O formante de cantor é uma característica que reforça a amplitude relativa dos harmónicos na região dos 3000 Hz tornando possível a voz de um cantor sobrepor-se ao som de uma orquestras. É um parâmetro de qualidade do desempenho do cantor em emitir sons de forma fácil, clara e confortável com a maior energia possível.

Acusticamente com a utilização de formante de cantor verifica-se um aumento da amplitude dos harmónicos na sua região bem como a utilização de outras ressonâncias em zonas de mais altas frequências, como está ilustrado nas Figuras 9 e 10. Verifica-se também o aumento da energia relativa dos harmónicos situados na zona do formante de cantor contrastando com a menor energia no caso da voz plana. Em alguns casos como o da figura 14, a energia do formante de cantor sobrepõe-se até ao valor da frequência fundamental  $F_0$ .

---

**Formante de Cantor  
(voz masculina)**



**Voz Plana  
(Voz masculina)**

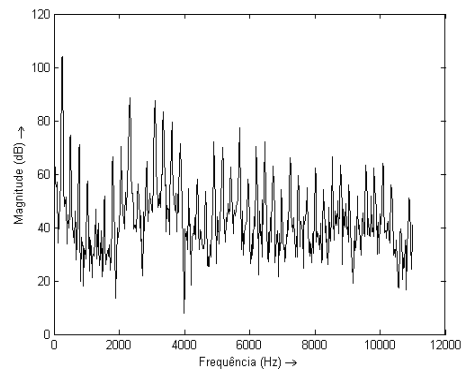
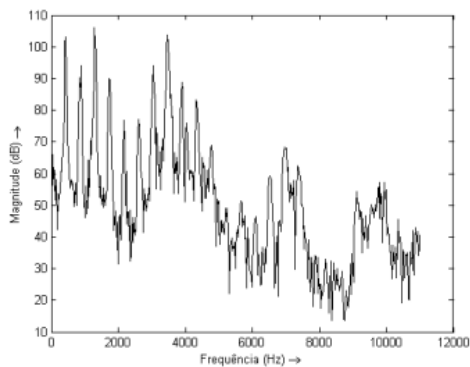


Figura 14 – Análise espectral dos parciais harmônicos numa voz masculina no registo voz plana e voz com formante de cantor.

**Formante de Cantor  
(voz feminina)**



**Voz Plana  
(Voz feminina)**

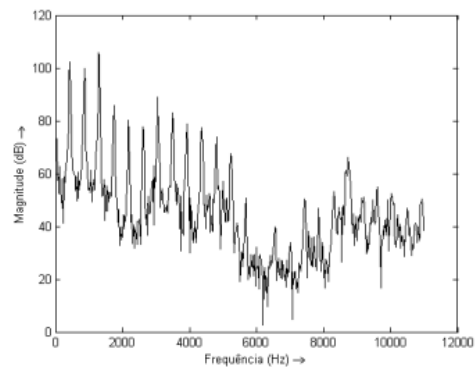


Figura 15 – Análise espectral dos parciais harmônicos numa voz feminina no registo voz plana e voz com formante de cantor.

Verificamos também a existência de vibrato com a proeminência do formante de cantor contrastando com a ausência de vibrato na amostra de voz plana. O vibrato é uma consequência da existência do formante de cantor sendo portanto uma característica indispensável quando o formante de cantor está presente.

### 3.3.5. Vibrato

O caso do *vibrato* é da nossa parte, alvo de um tipo de estudo diferente.

Foram estudados dois modos diferentes de utilização da voz cantada. No primeiro utilizamos as cinco vogais abertas (<a>, <e>, <i>, <o> e <u>) emitidas de forma prolongada e sustentada. No segundo modo de utilização do canto utilizamos um excerto da ária antiga *Caro mio ben* de Tommaso Giordani (1695 – 1762).

Foi utilizado o software SingingStudio e obtidos os valores de *pitch*, frequência de vibrato (ou taxa), extensão e pureza sinusoidal como mostra a figura 17. Para cada trecho da ária cantada, foram seccionados diversos *frames* contendo vibrato reconhecidos pelo SingingStudio (Figura 16). No capítulo seguinte damos ênfase à comparação entre os vários tipos de vibrato para cada cantor.



Figura 16 – Detecção de *frame* contendo vibrato no programa SingingStudio



Figura17 – Valores extraídos do vibrato detectado anteriormente.

Vamos também tentar perceber se existem características inerentes ao vibrato que se mantêm (ou apresentam uma variação pouco significativa) independentemente dos diferentes tipos de utilização da voz cantada (execução da ária ou vogal sustentada) ou se o mesmo vibrato sofre alguma alteração com a variação da gama de frequências emitida pelo cantor. Esperamos também chegar à conclusão de que o vibrato varia de acordo com o interlocutor através da comparação entre diferentes vibratos emitidos por diferentes cantores para o mesmo exercício.

---

## Capítulo 4

---

## 4. Análise Estatística – Resultados Obtidos

---

### 4.1. Introdução

Neste capítulo apresentaremos os resultados obtidos na análise estatística dos parâmetros da voz cantada escolhidos no capítulo anterior. Para a obtenção dos resultados aqui descritos, diversos parâmetros acústicos foram retirados e analisados estatisticamente.

Nesta parte da dissertação é então apresentada uma enumeração e descrição dos vários parâmetros extraídos e a forma como os mesmos se extraíram para dar origem aos resultados obtidos.

Para a análise dos binómios Falsete/Modal, Formante de Cantor/Voz plana, Limpidez/Soprosidade e Portamento (vs. Não Portamento) foram utilizados os testes de Kruskal-Wallis e Mann-Whitney U. Estes testes baseiam-se na comparação entre os parâmetros acústicos extraídos das amostras de voz evidenciando quais aqueles que melhor demonstram a diferença entre os elementos do binómio analisado (ex. diferenças entre registo de falsete e modal ou entre limpidez e soprosidade). No caso

---

do teste de Kruskal-Wallis, quanto maior for o valor de Chi-quadrado, mais as diferenças nos valores do parâmetro analisado (ex. *jitter*) são esclarecedoras relativamente à diferença entre os dois constituintes do binómio. No caso do teste de Mann-Whitney U, quanto menor for o valor de “U”, mais as diferenças são esclarecedoras.

## 4.2. Parâmetros acústicos extraídos

### 4.2.1. *Jitter e Shimmer*

*Jitter* e *Shimmer* são características acústicas de sinais de vozes que podem ser quantificados como variações da frequência fundamental ciclo glótico a ciclo glótico no caso do *Jitter* e amplitude da onda sonora no caso do *Shimmer*. Estes parâmetros são normalmente utilizados na definição e caracterização de vozes patológicas e são normalmente medidos em vogais sustentadas. Considera-se que podem existir diferenças significativas nos parâmetros *Jitter* e *Shimmer* para diferentes estilos de voz falada especialmente no parâmetro *Shimmer*.

#### a) *Jitter*

O parâmetro *Jitter* pode ser dividido em 4 subclassificações sendo estas: *absolute*, *relative*, *rap* e *ppq5*.

- *Jitter (absolute)*: consiste na variação da frequência fundamental ciclo-a-ciclo (falamos em ciclos glóticos) neste caso, a média da diferença absoluta entre dois períodos consecutivos.
- *Jitter (relative)*: é definido como a média da diferença absoluta entre dois períodos consecutivos dividida pelo período médio.

- 
- *Jitter (rap)*: é definido como *Relative Average Perturbation*, ou seja, a média da diferença absoluta entre um período e a média desse mesmo período e do seguinte e anterior, dividido pelo período médio.
  - *Jitter (ppq5)*: definido como o quociente de cinco pontos da perturbação do período (*Period Perturbation Quocient five-point*), a média da diferença absoluta entre um período e a média desse mesmo período e dos dois seguintes e dois anteriores (ou seja dos 4 períodos mais próximos).

#### **b) Shimmer**

À semelhança do parâmetro *jitter*, o *shimmer* também é dividido em 4 subclassificações: *Shimmer (dB)*, *relative*, *apq3* e *apq5*.

- *Shimmer (dB)*: consiste na variação da amplitude do sinal pico-a-pico. É expressa em decibéis (dB). A amplitude do sinal está directamente relacionada com a intensidade (*loudness*).
- *Shimmer (relative)*: é definido como a média da diferença absoluta entre amplitudes de períodos consecutivos dividida pela amplitude média. É expressa em percentagem.
- *Shimmer (apq3)*: trata-se do quociente de três pontos de perturbação de amplitude (*Amplitude Perturbation Quotient three-point*), a média da diferença absoluta entre a amplitude de um período e a média de amplitudes dos períodos seguinte e anterior, dividida pela amplitude média.
- *Shimmer (apq5)*: trata-se do quociente de cinco pontos de perturbação de amplitude (*Amplitude Perturbation Quotient five-point*), a média da diferença absoluta entre a amplitude de um período e a média de amplitudes dos 4 períodos mais próximos (2 seguintes e 2 anteriores), dividida pela amplitude média.

- 
- *Shimmer (apq11)*: trata-se do quociente de onze pontos de perturbação de amplitude (*Amplitude Perturbation Quotient eleven-point*), a média da diferença absoluta entre a amplitude de um período e a média de amplitudes dos 10 períodos mais próximos (5 seguintes e 5 anteriores), dividida pela amplitude média.

#### 4.2.2. Autocorrelação

A autocorrelação representa a distribuição da magnitude espectral do sinal vocal num domínio de tempo (a autocorrelação de um sinal é a transformação inversa de Fourier de um espectro de energia distribuída desse sinal). Em termos práticos a autocorrelação caracteriza a similaridade de um sinal com uma cópia de si mesmo, para um certo desalinhamento entre os dois.

#### 4.2.3. *Harmonics-to-noise Ratio* (HNR)

O parâmetro HNR (*Harmonics-to-noise Ratio* – Rácio entre harmónicos e ruído) é um dos parâmetros objetivos mais relevantes na análise acústica da voz. Trata-se de uma avaliação de base matemática da relação entre a componente periódica, representativa da vibração das pregas vocais e a componente aperiódica que decorre do ruído glótico. O valor do HNR é tanto maior quanto melhor for a qualidade e eficiência da fonação isto é, quanto melhor o fluxo de ar vindo dos pulmões for aproveitado para a vibração das pregas vocais traduzindo-se num ciclo vibratório mais estável e eficiente. Neste sentido, o contrário também é válido, ou seja, quanto menor for a qualidade do ciclo vibratório das pregas vocais em termos de estabilidade, mais baixa será a relação HNR resultado de um maior ruído glótico.

Assim sendo, um elevado valor de HNR corresponde a uma voz saudável auditivamente e visualmente ao passo que um baixo valor de HNR corresponde a uma voz com alto teor de ruído não-harmónico.

---

#### 4.2.4. *Noise-to-harmonics Ratio (NHR)*

Noise-to-harmonics Ratio (Rácio entre ruído e harmónicos) é definido como a relação entre a energia dos componentes não-harmónicos no intervalo de 1500-4500 Hz e a energia dos componentes harmónicos no intervalo de 70-4500 Hz. Trata-se da avaliação da presença de ruído num sinal vocal (assim como variações de amplitude e frequência, componentes sub-harmónicos e quebras de registo).

#### 4.2.5. *Short-time energy*

Este parâmetro fornece-nos a indicação da amplitude de um sinal de voz num determinado intervalo de tempo.

#### 4.2.6. *Spectral Flux*

O *Spectral Flux* é definido como o quadrado da diferença entre magnitudes normalizadas de distribuições espectrais sucessivas que correspondem a sinais de *frames* (excertos de uma mesma gravação ou amostra) sucessivos. Através do *Spectral Flux* conseguimos perceber as variações temporais na forma do espectro.

#### 4.2.7. *Spectral Centroid*

É definido como o centro de gravidade da energia de um espectro. Está normalmente associado à medição do brilho (*brightness*) de um som que está relacionada com a proeminência das altas frequências (*high-frequency*) do espectro. Valores altos no *Spectral Centroid* correspondem a um grande reforço do som nas altas frequências.

---

#### 4.2.8. *Spectral Entropy*

Este parâmetro é utilizado para detectar espaços vozeados e de silêncio num sinal de voz. Pode também ser utilizado na detecção de formantes e na distribuição dos seus picos espectrais. Nos segmentos vozeados denota-se um espectro mais organizado relativamente aos espaços não vozeados ou com ruído não-harmónico.

#### 4.2.9. *Spectral Rolloff*

O *Rolloff* é caracterizado como sendo a frequência abaixo da qual 85% da magnitude espectral está concentrada. Tal como no caso do *Spectral Centroid*, pode ser utilizado para a análise da forma do espectro e apresenta valores mais altos quanto maior for o reforço nas altas frequências.

#### 4.2.10. *Zero Crossing rate (ZCR)*

O parâmetro *Zero Crossing rate* é uma medição do número de vezes que o sinal de voz atravessa a o valor de 0 (zero) por unidade de tempo. Pode ser um parâmetro utilizado para fazer uma estimativa da frequência fundamental ou do ruído de um determinado sinal de voz. Sons periódicos têm normalmente valores baixos de ZCR ao passo que sons com elevado grau de ruído não-harmónico tendem a apresentar valores mais elevados.

---

## 4.3. Resultados Obtidos

### 4.3.1. Vibrato

Como já dissemos, foi analisada uma amostra de 6 cantores pertencentes à classe de canto da ESMAE. Na análise dos parâmetros acústicos relacionados com o vibrato, ou seja, Frequência do Vibrato, Extensão (pitch) e Pureza Sinusoidal, optamos por excluir o caso do Barítono de qualquer análise de parâmetros relacionados com o vibrato pelo facto de os dados obtidos serem considerados inválidos. Isto acontece pelo facto de não haver amostras suficientes no caso da vogal sustentada tendo sido obtidos valores escassos e muito díspares. Como convenção utilizamos também o valor de 0.1 semitons de extensão como o mínimo para a ocorrência de vibrato. Sendo que, no caso do Barítono, os valores apresentados são, além de poucos, possuidores de uma média de 0.0600 semitons, optamos também por excluí-los da nossa análise estatística.

#### 4.3.1.1. Frequência do vibrato

Na análise das duas formas de utilização da voz cantada emitidas pelos cantores através do programa de análise estatística SPSS® (IBM®), verificamos que a média da frequência do vibrato aumenta na execução da ária *Caro mio bem* relativamente à execução das vogais sustentadas. Isto acontece para todas as vozes analisadas à excepção do tenor e do barítono, este último pelo facto do teste ter sido considerado inválido. No caso específico do tenor, não encontramos uma variação significativa na média da frequência de vibrato comparando a execução da ária com a vogal sustentada denotando um especial caso de estabilidade.

Frequência de Vibrato (Hz)		Baixo	Barítono	Tenor	Contralto	Mezzo	Soprano
	<i>N</i>	17	2	11	6	7	18
<b>Vogal Sustentada</b>	Média (Hz)	5.2424	5.5650	5.4836	5.2433	6.5700	4.6111
	Desvio Padrão	0.9175	0.6010	0.1470	0.0599	0.7279	0.1407
	<i>N</i>	12	12	14	7	18	6
<b>Ária</b>	Média	5.0883	6.5512	5.4664	5.7400	7.7756	5.4667
	Desvio Padrão	0.5223	0.2353	0.3266	0.1638	0.4733	0.3080

Tabela 4 – Valores médios e desvio padrão das amostras recolhidas para o parâmetro Frequência do vibrato para as vozes de Baixo, Barítono, Tenor, Contralto, Mezzo-Soprano e Soprano. *N* – número de amostras.

Conseguimos também deprender que a maioria dos cantores apresenta uma variação da frequência de vibrato maior no caso da execução da ária relativamente à vogal sustentada. Esta afirmação é apoiada pelos resultados obtidos na análise das vozes de tenor, contralto e soprano. Esta variação poderá dever-se à introdução da sensação de musicalidade na emissão da voz cantada, assumindo o vibrato neste caso, uma função estética/estilística. Contrariamente ao esperado, as vozes de baixo e mezzo-soprano, apresentam uma variação maior da frequência de vibrato nas vogais sustentadas relativamente à ária. O barítono, como já vimos, apresenta um teste inválido. Podemos também realçar a estabilidade do vibrato nas amostras recolhidas para a voz de Contralto pois, tanto para o exercício de vogais sustentadas como para a ária cantada, é a que apresenta valores de desvio padrão mais baixos, traduzindo-se assim numa menor variação da frequência do vibrato comparativamente às outras vozes apesar de, Tenor e Soprano apresentarem também bons valores de desvio padrão.

Conseguimos também verificar que os valores obtidos na frequência de vibrato vão de encontro aos valores citados na literatura como sendo “normais” (neste caso de 5.5 a 7.5 Hz). Segundo a literatura, o vibrato aconselhável anda à volta dos 6 Hz Hall

(1991). As vozes que estão mais perto deste valor são as de Tenor e Mezzo para as vogais sustentadas e Soprano, Tenor e Contralto na execução da ária.

#### 4.3.1.2. Extensão do vibrato (*pitch*)

Relativamente à análise das médias obtidas para a extensão do vibrato, verifica-se a tendência para a extensão ser menor na vogal sustentada do que na execução da ária como mostram os gráficos resultantes da análise. Para a obtenção destas tabelas foi utilizado novamente o programa SPSS® (IBM®). Os resultados obtidos são assim de acordo com o esperado relativamente à comparação da extensão para as duas formas de utilização da voz cantada. Como já foi referido, consideramos como existência de vibrato, todas as modulações de *pitch* acima de 0.1 semitons, o que acontece em todos os casos à excepção do barítono.

Verifica-se também um maior desvio padrão na execução da ária, resultante de uma maior variação na extensão média relativamente ao mesmo cantor realizando o exercício de vogal sustentada.

Extensão (Semitons)		Baixo	Barítono	Tenor	Contralto	Mezzo	Soprano
	<i>N</i>	17	2	11	6	7	18
<b>Vogal Sustentada</b>	Média (Semitons)	0.1388	0.0600	0.4618	0.4583	0.2229	0.4672
	Desvio Padrão	0.0512	0.0283	0.0924	0.1074	0.0655	0.1060
	<i>N</i>	12	12	14	7	18	6
<b>Ária</b>	Média (Semitons)	0.2583	0.3392	0.5214	0.4600	0.3611	0.5667
	Desvio Padrão	0.1566	0.0624	0.1033	0.2146	0.1058	0.1155

Tabela 5 – Valores médios e desvio padrão das amostras recolhidas para o parâmetro Extensão do vibrato (*pitch*) para as vozes de Baixo, Barítono, Tenor, Contralto, Mezzo-Soprano e Soprano. *N* – número de amostras.

#### 4.3.1.3. Pureza Sinusoidal

Observando os resultados obtidos para a Pureza Sinusoidal, verificamos que este parâmetro apresenta valores médios superiores nos exercícios de vogal sustentada relativamente à performance da ária. Esta afirmação é apoiada pelos dados recolhidos em todas as vozes à excepção da Soprano e do Barítono que, como já dissemos, é considerado como um resultado inválido.

Observa-se também um desvio padrão bastante superior no caso dos exercícios de vogal sustentada relativamente à ária para todas as vozes excepto para a amostra recolhida na voz de Soprano. Os indivíduos que apresentam maiores valores de pureza sinusoidal são o Tenor para as vogais sustentadas e a Soprano para o trecho da ária. É de realçar que, embora numa situação de vibrato ideal, a pureza sinusoidal ronde os 100%, este valor foi impossível de atingir sendo o máximo atingido de 92.59% num excerto da ária cantada pelo Tenor.

Pureza Sinusoidal (%)		Baixo	Barítono	Tenor	Contralto	Mezzo	Soprano
	<i>N</i>	17	2	11	6	7	18
<b>Vogal Sustentada</b>	Média (%)	60.5066	49.2150	86.4255	84.8400	66.3114	85.4872
	Desvio Padrão	5.7855	9.9773	2.6630	3.5364	5.5427	6.1886
	<i>N</i>	12	12	14	7	18	6
<b>Ária</b>	Média (%)	57.0450	56.3950	78.9729	67.1100	51.2561	86.4517
	Desvio Padrão	11.9083	18.0596	9.3258	26.1176	22.9651	1.7407

Tabela 6 – Valores médios e desvio padrão das amostras recolhidas para o parâmetro Pureza Sinusoidal para as vozes de Baixo, Barítono, Tenor, Contralto, Mezzo-Soprano e Soprano. *N* – número de amostras.

---

### 4.3.2. Falsete

Para o estudo do binómio falsete/modal foram utilizados os testes de validação de hipótese de Kruskal-Wallis e Man-Whitney U. Dos parâmetros acústicos acima descritos, aqueles que melhor demonstram as diferenças entre os dois elementos do binómio são *spectral entropy*, Autocorrelação média e NHR médio. Podemos verificar esta afirmação através dos valores obtidos para U (4,6 e 6 respectivamente) e Chi-quadrado (51.083, 50.761 e 50.761 respectivamente) dos diferentes testes.

	Mann-Whitney U		Kruskal Wallis	
	U	Sig. Assint. (2 caudas)	Chi-quadrado	Significância Assintótica
"Jitter Local"	40.500	0.000	45.380	0.000
"Jitter Local (Absolute)"	140.500	0.000	31.456	0.000
"Jitter RAP"	69.000	0.000	41.215	0.000
"Jitter PPQ5"	56.500	0.000	43.051	0.000
"Jitter DDP"	69.500	0.000	41.087	0.000
"Shimmer Local"	13.000	0.000	49.646	0.000
"Shimmer Local (dB)"	132.000	0.000	32.552	0.000
"Shimmer APQ3"	37.500	0.000	45.829	0.000
"Shimmer APQ5"	73.000	0.000	40.570	0.000
"Shimmer APQ11"	117.000	0.000	34.497	0.000
"Shimmer DDA"	38.000	0.000	45.751	0.000
Autocorrelação média	6.000	0.000	50.761	0.000
NHR médio	6.000	0.000	50.761	0.000
HNR médio	18.000	0.000	48.852	0.000
"Short Time Energy"	79.000	0.000	39.712	0.000
"Spectral Flux"	36.000	0.000	46.057	0.000
"Spectral Centroid"	11.000	0.000	49.961	0.000
"Spectral Entropy"	4.000	0.000	51.083	0.000
"Spectral Roll-Off" a 95%	626.000	0.892	0.018	0.892
"Zero Crossing Rate"	15.000	0.000	49.326	0.000

Tabela 7 – Valores dos testes de Kruskal-Wallis e Mann-Whitney U para os parâmetros acústicos nas amostras de voz em falsete e em registro modal.

---

#### 4.3.2.1. *Spectra Entropy*

O parâmetro *Spectral Entropy* é dos três seleccionados, aquele que mais diferenças apresenta na análises dos registos de Falsete e Modal.

Registou-se uma diferença bastante grande e notória dos níveis de *spectral entropy* favorável à voz modal relativamente ao registo de falsete. Esta afirmação é suportada pelo gráfico que mostra claramente que os valores de *spectral entropy* no registo modal estão acima.

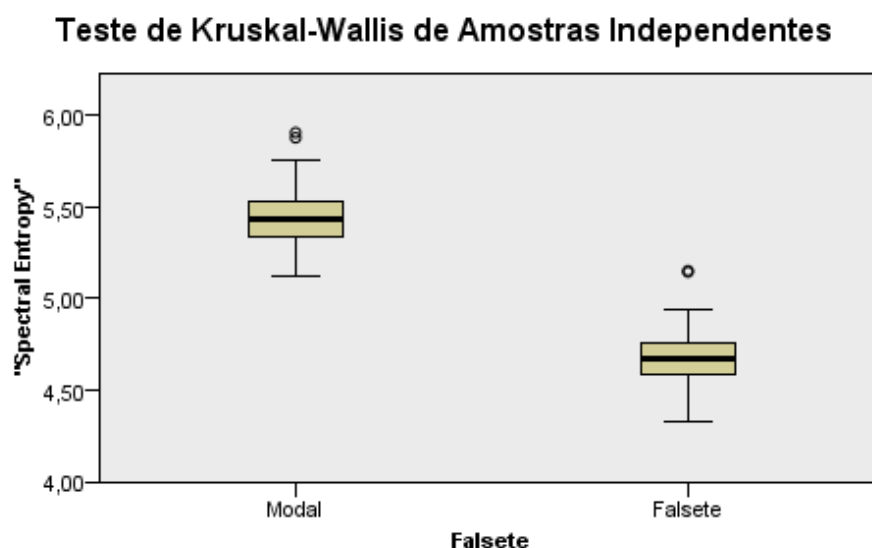


Gráfico 2 – Teste de Kruskal-Wallis para *Spectral Entropy* entre registos Modal e Falsete.

Sendo que este parâmetro detecta a existência de formantes e seus picos e os espaços vozeados num sinal sonoro, podemos então afirmar que, o facto de *spectral entropy* ser maior no caso do registo modal, quer dizer que há uma maior definição dos picos de formantes e um maior número de sons vozeados no registo modal. Contrariamente, o registo de falsete possui menos sons vozeados (portanto mais ruído). Estes resultados vão de encontro aos resultados expressos no Capítulo 3 desta dissertação ou seja, há uma maior definição dos formantes no registo modal

---

relativamente ao registo de falsete o que se evidencia numa maior definição dos harmónicos. Vemos claramente no Gráfico 2 que os harmónicos no registo de modal, à medida que se avança no espectro, continuam bastante definidos na região das altas frequências ao passo que, no registo de falsete, à medida que se avança no espectro, os harmónicos na região das altas frequências vão perdendo definição ficando mergulhados no ruído. Havendo menor definição harmónica a voz é menos audível o que vai de encontro ao que é referido na literatura.

#### 4.3.2.2. Autocorrelação média

Através do valor de autocorrelação média conseguimos saber a distribuição espectral do sinal vocal num domínio de tempo. Através dos valores de autocorrelação podemos saber se a distribuição espectral varia muito ao longo do tempo ou não. Os valores de autocorrelação vão de 0 a 1, sendo 0 o valor mais baixo correspondente a uma grande variação do sinal espectral (e, portanto, a uma menor similaridade entre o sinal e uma cópia deslocada), e 1 o valor mais alto, correspondente a uma total estabilidade do sinal espectral (e, portanto, a uma grande similaridade entre o sinal e uma cópia afetada de um determinado deslocamento).

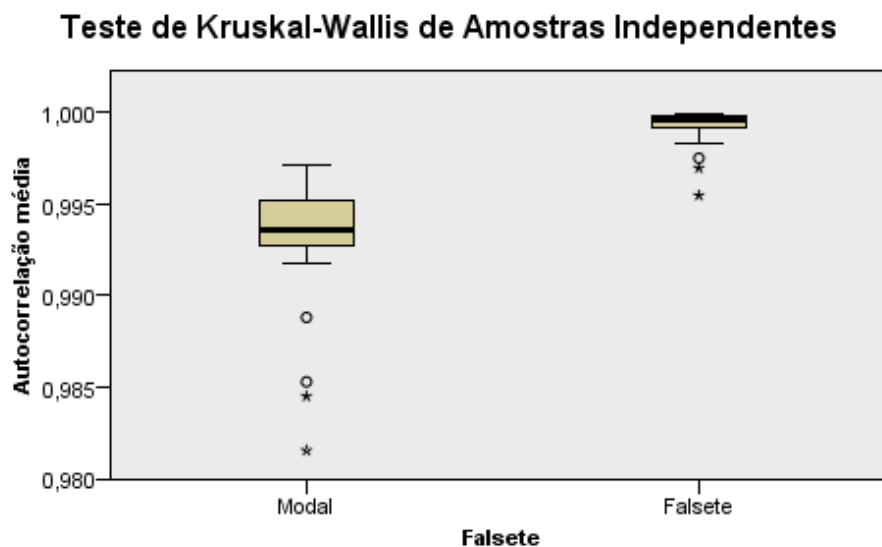


Gráfico 3 – Teste de Kruskal-Wallis para Autocorrelação média entre registos Modal e Falsete.

---

Pela análise do gráfico referente ao teste de Kruskal-Wallis para a comparação entre o registo de falsete e modal, verificamos que existe um maior valor de autocorrelação para o falsete sendo este bastante próximo de 1. Isto sugere-nos uma menor variação da distribuição espectral ao longo do tempo.

Através destes resultados podemos deduzir que, sendo o registo de falsete uma voz plana relativamente ao registo modal (onde há proeminência do formante de cantor), a ausência de vibrato e a pouca flutuação dinâmica no decorrer do exercício podem resultar numa distribuição espectral com pouca variabilidade.

#### 4.3.2.3. *NHR (Noise-to-harmonics Ratio)*

A medida NHR mede a relação entre a energia dos componentes não-harmónicos e a energia dos componentes harmónicos. É então de esperar um maior valor de NHR para o falsete relativamente ao registo modal, evidenciando o maior grau de ruído não-harmónico que a emissão do registo de falsete acarreta devido à permanente fenda-glótica associada à emissão deste registo.

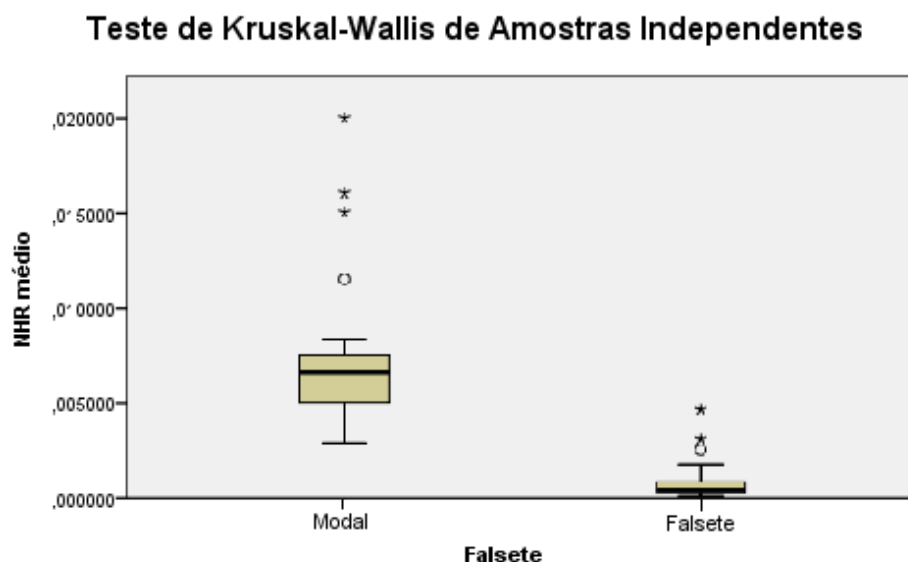


Gráfico 4 – Teste de Kruskal-Wallis para valores de NHR médio entre registos Modal e Falsete.

---

Analisando o gráfico vemos que tal não acontece, havendo inclusivamente um valor mais alto de NHR para o registo modal. Uma vez que também não obtivemos valores de HNR (que evidenciam a relação entre componentes harmónicos e ruído não-harmónico) conclusivos relativamente às diferenças entre o falsete e o registo modal podemos afirmar a falta de capacidade dos algoritmos utilizados para a detecção de componentes harmónicos nas vozes com vibrato. Podemos então deduzir que a existência de vibrato pode complicar a tarefa de reconhecimento dos componentes harmónicos da voz modal traduzindo-se em valores de relação harmónica diferentes dos esperados.

### 4.3.3. Formante de Cantor

Tal como na análise do falsete, para a análise do binómio formante de cantor/voz plana, foram utilizados os testes de Kruskal-Wallis e Mann-Whitney U. Neste caso, os parâmetros acústicos escolhidos com base nos resultados destes testes foram: *Jitter (local)*, *Shimmer (local)* e *HNR (Harmonics-to-noise Ratio)* médio. Podemos verificar esta afirmação através dos valores obtidos para U (0 nos três casos) e Chi-quadrado (10.125 nos três casos) dos diferentes testes.

	Mann-Whitney U		Kruskal Wallis	
	U	Sig. Assint. (2 caudas)	Chi-quadrado	Significância Assintótica
"Jitter Local"	0	,000 <sup>b</sup>	10.143	0.001
"Jitter Local (Absolute)"	0	,000 <sup>b</sup>	10.125	0.001
"Jitter RAP"	1	,001 <sup>b</sup>	9.406	0.002
"Jitter PPQ5"	0	,000 <sup>b</sup>	10.125	0.001
"Jitter DDP"	1	,001 <sup>b</sup>	9.389	0.002
"Shimmer Local"	0	,000 <sup>b</sup>	10.125	0.001
"Shimmer Local (dB)"	0	,000 <sup>b</sup>	10.125	0.001
"Shimmer APQ3"	0	,000 <sup>b</sup>	10.125	0.001
"Shimmer APQ5"	0	,000 <sup>b</sup>	10.125	0.001
"Shimmer APQ11"	0	,000 <sup>b</sup>	10.125	0.001
"Shimmer DDA"	0	,000 <sup>b</sup>	10.125	0.001
Autocorrelação média	0	,000 <sup>b</sup>	10.125	0.001
NHR médio	9	,036 <sup>b</sup>	4.5	0.034
HNR médio	0	,000 <sup>b</sup>	10.125	0.001
"Short Time Energy"	0	,000 <sup>b</sup>	10.125	0.001
"Spectral Flux"	0	,000 <sup>b</sup>	10.125	0.001
"Spectral Centroid"	6	,012 <sup>b</sup>	6.125	0.013
"Spectral Entropy"	1	,001 <sup>b</sup>	9.389	0.002
"Spectral Roll-Off" a 95%	17	,272 <sup>b</sup>	1.389	0.239
"Zero Crossing Rate"	15.000	0.000	49.326	0.000

Tabela 8 – Valores dos testes de Kruskal-Wallis e Mann-Whitney U para os parâmetros acústicos nas amostras de voz plana e formante de cantor.

---

#### 4.3.3.1. Jitter

O jitter é um parâmetro acústico que serve como medida de curto termo (ciclo glótico a ciclo glótico) de variações involuntárias na frequência fundamental. Se na avaliação de vozes patológicas ele serve para medir a estabilidade do aparelho fonatório, na voz cantada pode assumir outro papel. Vejamos o gráfico obtido no teste de Kruskal-Wallis:

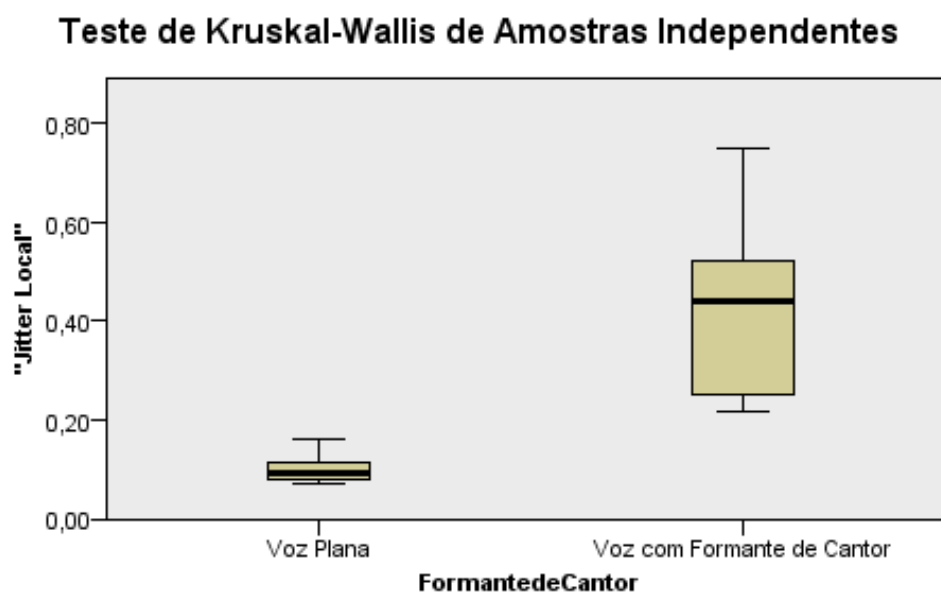


Gráfico 5 - Teste de Kruskal-Wallis para parâmetros *jitter* (local) para exemplos de voz plana e formante de cantor.

O gráfico obtido na análise do *jitter* para as amostras em questão mostra-nos que o formante de cantor apresenta valores bastante mais altos de jitter relativamente à voz plana sendo a diferença bastante significativa. Este valor deve-se à muito provavelmente às variações de frequência fundamental a que o vibrato está associado. De facto, o vibrato, como consequência da projecção vocal que o formante de cantor proporciona, está presente quando os cantores utilizam o formante de cantor e é constituído, como está acima descrito e provado, por variações da frequência fundamental (*pitch*), variações essas que em cima definimos como Extensão. Acontece que uma das características da voz plana é a não existência de vibrato portanto, as

---

variações na frequência fundamental são menores o que se traduz em menores valores de *jitter*.

As variações de frequência fundamental induzidas pelo vibrato são então responsáveis por esta distinção clara entre a voz plana e o formante de cantor.

#### 4.3.3.2. *Shimmer*

O *shimmer* é um parâmetro que mede perturbações na amplitude do sinal pico a pico. A amplitude está directamente relacionada com a intensidade do som, ou seja com o *loudness*. Teoricamente, será de esperar que obtenhamos valores maiores de shimmer para a amostra de formante de cantor relativamente à amostra de voz plana dado que por causa da existência de vibrato na primeira, decorre naturalmente uma maior variação de amplitude (é conhecido que uma modulação de frequência tipicamente ocasiona também uma modulação de amplitude). Pela análise do Gráfico 6, vemos que acontece o que teoricamente é esperado:

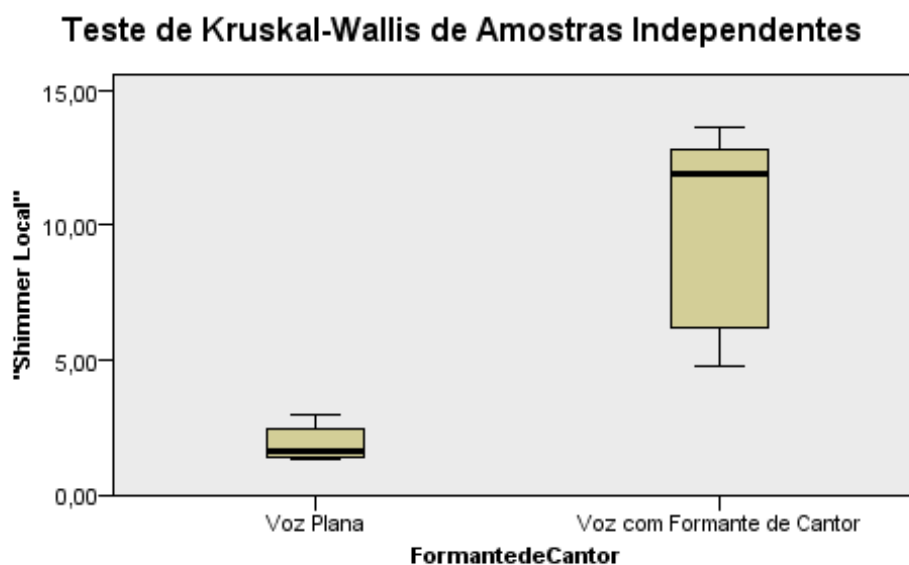


Gráfico 6 - Teste de Kruskal-Wallis para parâmetros *shimmer* (local) para exemplos de voz plana e formante de cantor.

---

Confirma-se, como se pode observar no gráfico, que não só a amplitude é maior como a variação da mesma é bastante maior no caso de formante de cantor relativamente à voz plana. Isto leva-nos também a suportar a afirmação de que a utilização do formante de cantor confere audibilidade à voz e isso traduz-se numa maior intensidade espectral (*loudness*).

O reforço de ressonâncias proporcionado pela utilização do formante de cantor fornece à voz a intensidade necessária para se destacar relativamente ao som de uma orquestra, intensidade essa que se traduz em um aumento da gama de valores do *shimmer*, quer em termos médios quer em termos de variabilidade desses mesmos valores.

#### 4.3.3.3. *HNR (Harmonics-to-noise Ratio)*

Esta medida fornece-nos a relação entre os componentes harmónicos do sinal de voz e o ruído não-harmónico. É portanto de esperar que, no exemplo de formante de cantor, haja uma maior relação HNR do que na voz plana devido à maior componente harmónica proporcionada pelo reforço de ressonâncias que é característico da utilização do formante de cantor. Vejamos então o que nos demonstra o Gráfico 7

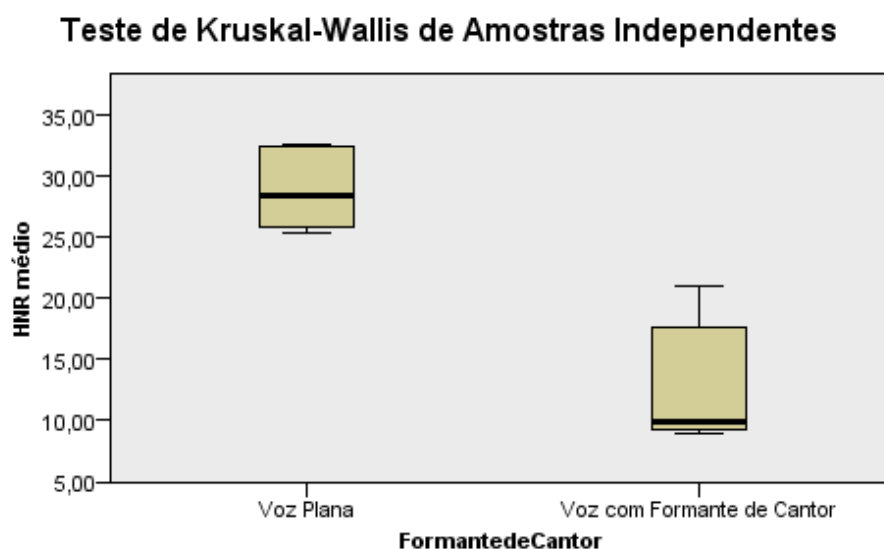


Gráfico 7 - Teste de Kruskal-Wallis para parâmetros *HNR* médio para exemplos de voz plana e formante de cantor.

---

Tal como acontece na medida *NHR* na análise do falsete, os resultados obtidos encontram-se no oposto dos resultados esperados. Mais uma vez devemos considerar seriamente a dificuldade do algoritmo utilizado em detectar a componente harmónica em vozes com vibrato.

#### 4.3.4. Limpidez/Soprosidade

Na análise do binómio Limpidez/Soprosidade foram novamente utilizados os testes de Kruskal-Wallis e Mann-Whitney U. Neste caso, os parâmetros acústicos escolhidos com base nos resultados destes testes foram: *Jitter (rap)*, *Zero crossing rate* e *Spectral Flux*. Podemos verificar esta afirmação através dos valores obtidos para U (147, 69 e 80 respectivamente) e Chi-quadrado (16.613, 28.07 e 26.272 respectivamente) dos diferentes testes.

	Mann-Whitney U		Kruskal Wallis	
	U	Sig. Assint. (2 caudas)	Chi-quadrado	Significância Assintótica
"Jitter Local"	227	0.005	7.961	0.005
"Jitter Local (Absolute)"	199	0.001	10.63	0.001
"Jitter RAP"	147	0	16.613	0
"Jitter PPQ5"	194	0.001	11.155	0.001
"Jitter DDP"	147.5	0	16.548	0
"Shimmer Local"	395.5	0.857	0.032	0.857
"Shimmer Local (dB)"	394.5	0.845	0.038	0.845
"Shimmer APQ3"	271.5	0.034	4.511	0.034
"Shimmer APQ5"	336	0.266	1.239	0.266
"Shimmer APQ11"	361	0.471	0.52	0.471
"Shimmer DDA"	271	0.033	4.544	0.033
Autocorrelação média	368	0.541	0.374	0.541
NHR médio	368	0.541	0.374	0.541
HNR médio	390	0.79	0.071	0.79
"Short Time Energy"	301	0.097	2.761	0.097
"Spectral Flux"	80	0	26.272	0
"Spectral Centroid"	368	0.541	0.374	0.541
"Spectral Entropy"	263	0.024	5.095	0.024
"Spectral Roll-Off" a 95%	164	0	14.508	0
"Zero Crossing Rate"	69	0	28.07	0

Tabela 9 – Valores dos testes de Kruskal-Wallis e Mann-Whitney U para os parâmetros acústicos nas amostras de voz plana e formante de cantor.

---

#### 4.3.4.1. Jitter (rap)

O jitter, como já vimos, mede variações na frequência fundamental nas amostras de voz. Sendo que, nos exemplos de voz límpida, o tipo de voz executada pelos cantores incluídos na amostra é uma voz timbrada e, portanto, com formante de cantor. Conseqüentemente, a inclusão desta medida como uma das três que mais evidenciam diferenças entre as amostras de voz límpida e voz com sopro, é consequência da existência de vibrato na voz límpida e não de outra característica que diferencie a voz límpida da voz com sopro.

Vemos também através do gráfico obtido (Gráfico 8) que, os valores não são de todo conclusivos para suportarmos qualquer afirmação descrita no capítulo 3 desta dissertação.

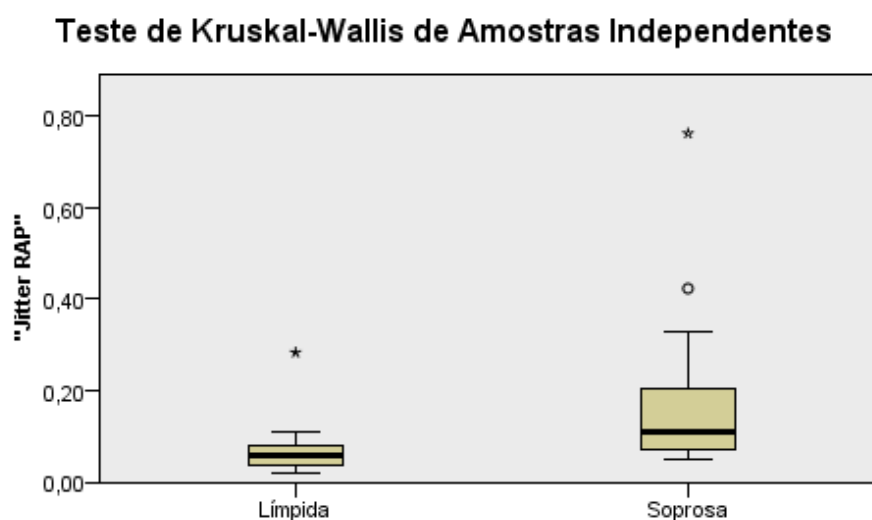


Gráfico 8 - Teste de Kruskal-Wallis para parâmetros *jitter (rap)* para exemplos de voz límpida e voz com sopro.

Apesar de, a nível espectral, haver uma clara distinção entre uma voz límpida e uma voz com sopro através da representação espectral de harmônicos e ruído (gráfico 8), o mesmo não podemos afirmar a nível estatístico devido a, mais uma vez,

---

uma incapacidade dos algoritmos de reconhecer parciais harmónicos numa voz com vibrato.

#### 4.3.4.2. *Spectral Flux*

Através do *Spectral Flux* conseguimos perceber as variações na forma do espectro. Na música, existe uma alta gama de frequências e uma grande variação de sons em termos de altura portanto é de esperar que na voz cantada aconteça o mesmo. Vejamos o que acontece no Gráfico 9.

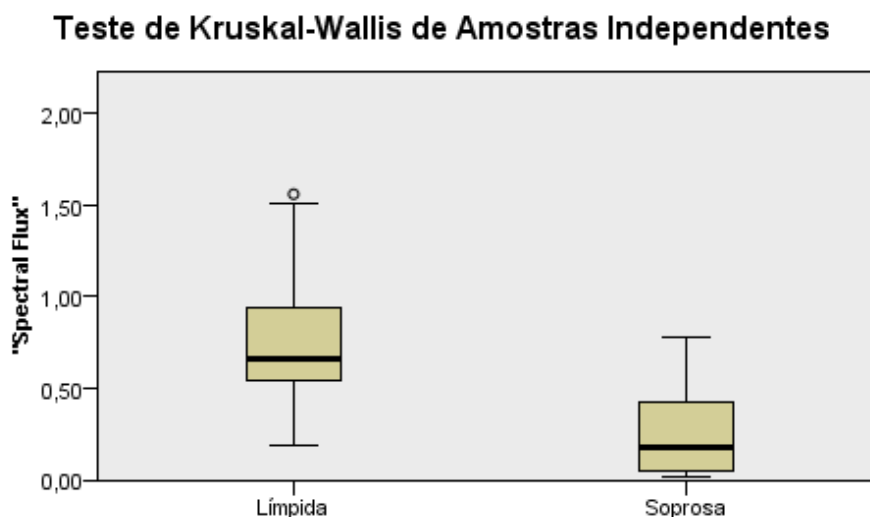


Gráfico 9 - Teste de Kruskal-Wallis para valores de *Spectral Flux* para exemplos de voz límpida e voz com sopro.

Devido ao facto de, tanto no caso da voz límpida como no caso da voz com algum nível de sopro, terem sido efectuados os mesmos exercícios, neste caso um vocalizo específico, não conseguimos ter uma diferenciação substancial no gráfico de Kruskal-Wallis. Contudo podemos ver que há uma ligeira diferença superior nos valores de *Spectral Flux* para a amostra de voz límpida. Podemos interpretar isto da seguinte forma: sendo a emissão da voz límpida neste caso, um vocalizo, isso pressupõe que

---

tenha alguma musicalidade ao contrário da voz com sopro que, por questões de natureza técnica, possui menos propensão a flutuações de dinâmica. Esta *décalage* entre os sons emitidos resulta então nestes valores de *Spectral Flux* embora, tal como no parâmetro *jitter (rap)* não seja totalmente válido na tarefa de chegar a uma conclusão através da análise estatística nem tão pouco vão de encontro aos resultados apontados no ponto 3.4.3. desta dissertação. Estamos então perante outra característica que, tal como a anterior, embora demonstre alguma diferença entre os dois tipos de emissão de voz cantada, é inconclusiva tendo em conta a meta que pretendemos esbater.

#### 4.3.4.3. Zero-cross Rating

Tal como foi descrito no ponto 4.2.10., o Zero-Cross Rating indica através dos seus valores se uma amostra possui alto teor de ruído não-harmónico ou se se trata de um som periódico e portanto tido como “saudável). Neste caso, são esperados baixos valores de ZCR para as amostras de voz límpida e por conseguinte, altos valores de ZCR para amostras de voz com sopro, isto é, que evidenciem a existência de ruído não-harmónico. Olhemos o Gráfico 10:

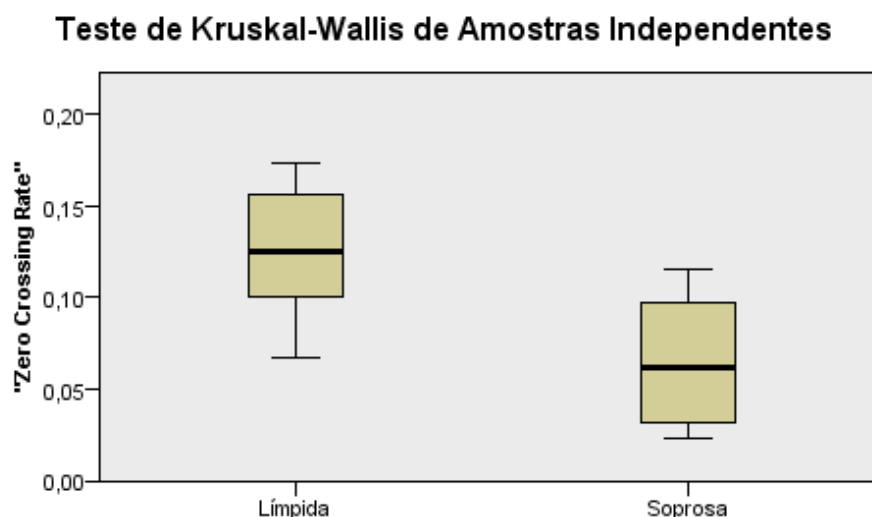


Gráfico 10 - Teste de Kruskal-Wallis para parâmetros *jitter (rap)* para exemplos de voz límpida e voz com sopro.

---

Estamos mais uma vez perante um caso de falha no algoritmo do programa IMB® SPSS®. Neste caso, a detecção dos valores de *Zero-crossing Rate* vão em sentido oposto ao que seria de esperar. Pela análise do gráfico observamos valores mais altos de ZCR para vozes límpidas, evidenciando que a amostra possui várias regiões de ruído não-harmónico. Novamente vemos aqui evidenciadas as dificuldades do algoritmo quando se depara com vozes com vibrato.

#### 4.3.4.4. O caso especial da análise do par Limpidez/Soprosidade

Como foi dito anteriormente, há uma limitação bastante acentuada por parte do algoritmo para a detecção de componentes harmónicos em vozes timbradas, isto é, com vibrato. Caso o algoritmo admitisse esta possibilidade, não temos dúvidas que os parâmetros acústicos que melhor evidenciariam a diferença entre limpidez e soprosidade seriam o HNR (*Harmonics-to-noise Ratio*) e o NHR (*Noise-to-harmonic Ratio*) devido ao facto de apresentarem numericamente a relação entre componentes harmónicos e ruído não harmónico.

#### 4.3.5. Portamento

Os parâmetros acústicos analisados no caso do Portamento são um pouco diferentes dos que foram utilizados nos casos anteriores. Visto que, o que vamos analisar são as notas intermédias entre a transição de uma nota para a outra, no caso da inexistência estas notas de transição são pouco perceptíveis e na maioria dos casos, não detectadas. Assim sendo, não faz sentido calcular valores de *jitter* ou *shimmer* ou HNR e NHR pois as amostras obtidas da detecção de notas intermédias nas transições com ausência de portamento vão ser muito pequenas.

A tabela de parâmetros acústicos com os respectivos valores de U e Chi-quadrado para os diferentes testes estatísticos é então um pouco diferente das outras.

	Mann-Whitney U		Kruskal Wallis	
	Mann-Whitney U	Sig. Assint. (2 caudas)	Chi- quadrado	Significância Assintótica
<b>"Sort Time Energy"</b>	1193	0.011	6.527	0.011
<b>"Spectral Flux"</b>	1055	0.001	10.998	0.001
<b>"Spectral Centroid"</b>	1446	0.247	1.343	0.247
<b>"Spectral Entropy"</b>	414	0	46.966	0
<b>"Spectral Roll-Off" a 95%</b>	1336	0.077	3.118	0.077
<b>"Zero Crossing Rate"</b>	1481	0.334	0.932	0.334
<b>Tempo da transição</b>	0	0	84.851	0
<b>Tempo médio por nota</b>	37	0	80.636	0
<b>Declive da transição</b>	175	0	67.475	0

Tabela 10 – Valores dos testes de Kruskal-Wallis e Mann-Whitney U para os parâmetros acústicos nas amostras de voz plana e formante de cantor.

Os parâmetros acústicos que vão ser analisados neste caso serão o Tempo de Transição, o Tempo médio por nota (entenda-se nota de transição) e o declive da transição pois são os que apresentam menores valores de U (0,37 e 175 respectivamente) e maiores valores de Chi-quadrado (0 nos três casos).

---

#### 4.3.5.1. Tempo da transição

A existência de portamento caracteriza-se pela passagem por todas as notas que separam a nota de partida da nota de destino (ex. na passagem com portamento de Mi para Sol passamos por Fá e Fá# até chegar finalmente a Sol). Os resultados esperados são portanto um maior tempo de transição na existência de portamento pois, partindo do princípio que a execução musical deve ser feita num determinado tempo, a obrigatoriedade de passagem pelas notas que separam a nota de partida da nota de destino vai fazer com que o tempo total de passagem por essas notas (entenda-se tempo de transição) seja obviamente maior neste caso do que no caso da inexistência de portamento.

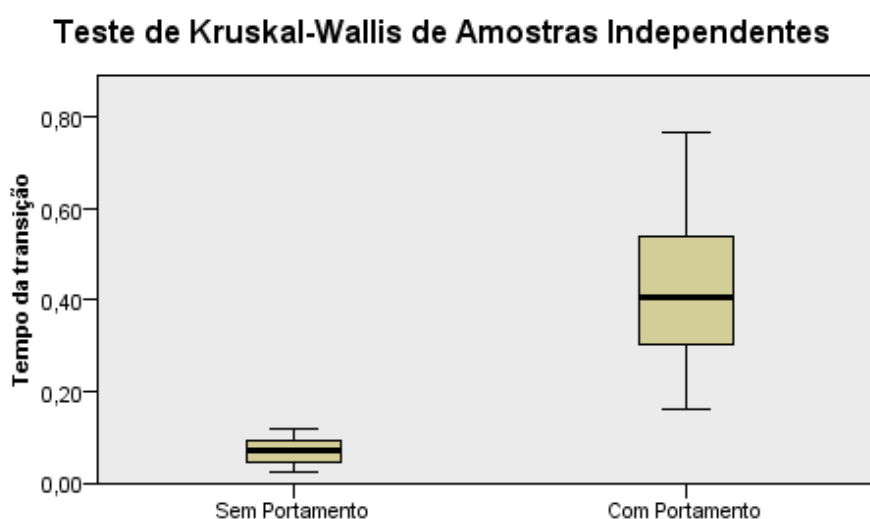


Gráfico 11 - Teste de Kruskal-Wallis para valores de Tempo Total de Transição para exemplos com portamento e sem portamento.

Observamos pela visualização do Gráfico 11 que os resultados obtidos são esclarecedores no que respeita à diferença entre os dois casos. De facto, com portamento, o tempo de transição, ou seja, o tempo total de passagem pelas notas intermédias é significativamente maior do que no caso de não-portamento. De realçar também que existe uma maior variabilidade nos tempos de transição nas passagens com

---

portamento relativamente às passagens sem portamento. Podemos então concluir que os resultados vão de encontro ao esperado.

#### 4.3.5.2. Tempo Médio por Nota

Na inexistência de portamento, como já foi referido neste capítulo e no ponto 3.4.1., as notas intermédias que separam a nota de partida da nota de destino, são tão curtas ou mesmo inexistentes que não chegam a ser reconhecidas pelo algoritmo e consequentemente não são contabilizadas. Assim sendo, é de esperar uma diferença muito substancial dos valores de Tempo Médio por Nota entre os dois tipos de amostra, sendo estes valores bastante superiores no caso de utilização de portamento.

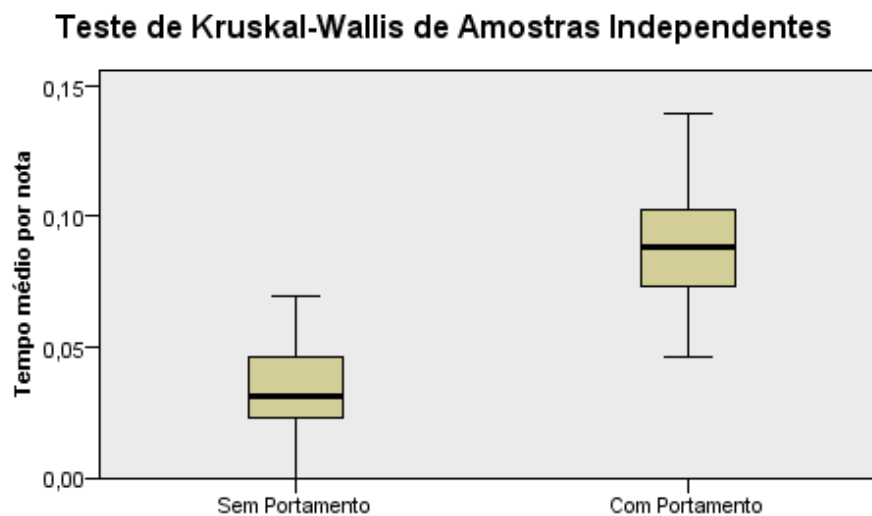


Gráfico 12 - Teste de Kruskal-Wallis para valores de Tempo Médio por Nota para exemplos com portamento e sem portamento.

Os resultados obtidos demonstrados pelo Gráfico 12 são mais uma vez esclarecedores, no sentido em que obtemos valores de Tempo Médio por Nota muito superiores nas amostras com portamento.

---

O facto de os valores de Tempo Médio por Nota nas amostras sem portamento não serem tão próximos de zero como seria de esperar poderá ser devido ao facto de, mesmo sem portamento, há sempre pequenas transições entre a nota de partida e a nota de destino que são normais tendo em conta a nossa condição humana.

#### 4.3.5.3. Declive da transição

Neste caso, o declive obtido pela análise dos dados reflecte a relação entre o número de notas dividido pelo tempo de transição entre a nota de partida e a nota de destino. É portanto o inverso do Tempo de Transição, quantificado anteriormente. O declive será então tanto maior quanto menor for o tempo de transição entre a primeira nota e a última. Os resultados esperados são então um maior declive na transição sem portamento uma vez que o tempo de transição é substancialmente menor. Olhemos o Gráfico 13:

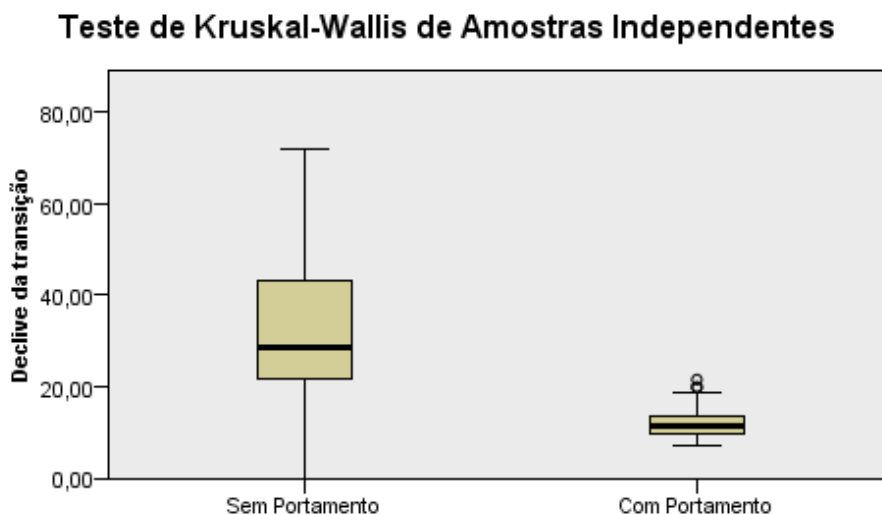


Gráfico 13 - Teste de Kruskal-Wallis para valores de Declive da Transição para exemplos com portamento e sem portamento.

---

Verificamos então que os resultados vão de encontro ao esperado e é observado um maior declive nas transições com portamento relativamente às transições sem portamento. A obtenção deste resultado reforça também a ideia que tínhamos do portamento expressa no ponto 3.3.1. e no ponto 4.3.5.1. onde caracterizamos o tempo de transição.

## 4.4. O Caso de Teresa Salgueiro

Tal como já foi referido, com a autorização da entidade gestora da sua carreira artística, utilizamos a voz da cantora Teresa Salgueiro na nossa análise estatística. Foi-nos enviado um trecho de uma canção da cantora o qual foi segmentado e dividido em amostras para vários parâmetros perceptivos. A análise das amostras e resultados obtidos servem, não para chegamos a uma conclusão, mas sim, a título de curiosidade, tentar perceber como esta voz tão peculiar se enquadra neste padrão definido por nós.

### 4.4.1. Vibrato

Na análise auditiva do trecho enviado, foi reconhecido que havia uma quase inexistência de vibrato na voz da cantora particularmente nas notas sustentadas. Onde há alguma ocorrência de vibrato é no final das frases musicais.

À semelhança do que fizemos com as outras amostras, foi elaborada uma tabela (Tabela 10) com as características do vibrato reconhecido pelo SinginsStudio.

		Frequência de Vibrato média	Desvio Padrão Frequência Vibrato	Pureza Sinusoidal
<b>N</b>	Válido	6	6	6
	Ausente	0	0	0
<b>Média</b>		6,2850	,0000	62,8800
<b>Mediana</b>		6,3850	,0000	62,1050
<b>Erro padrão</b>		0,49614	,00000	7,29526
25		5,7350	,0000	55,4950
<b>Percentis</b>	50	6,3850	,0000	62,1050
	75	6,5900	,0000	69,9875

Tabela 11 – Valores extraídos para caracterização do vibrato encontrado em amostras de voz de Teresa Salgueiro.

#### 4.4.2. Existência de Voz plana

O parâmetro perceptivo atribuído à voz cantada que mais se evidenciou foi a Voz Plana e por consequência, a não existência do formante de cantor. Analisando auditivamente vemos que há uma estabilidade vocal quer a nível de inexistência de vibrato quer a nível de flutuações de dinâmica. Perceptivamente a sensação que se obtém é de uma voz com um grau de limpidez muito acentuado. Foram extraídos os parâmetros de jitter, shimmer e HNR para os segmentos que consideramos como Voz Plana. Os resultados obtidos foram os seguintes:

---

### Teste de Kruskal-Wallis de Amostras Independentes

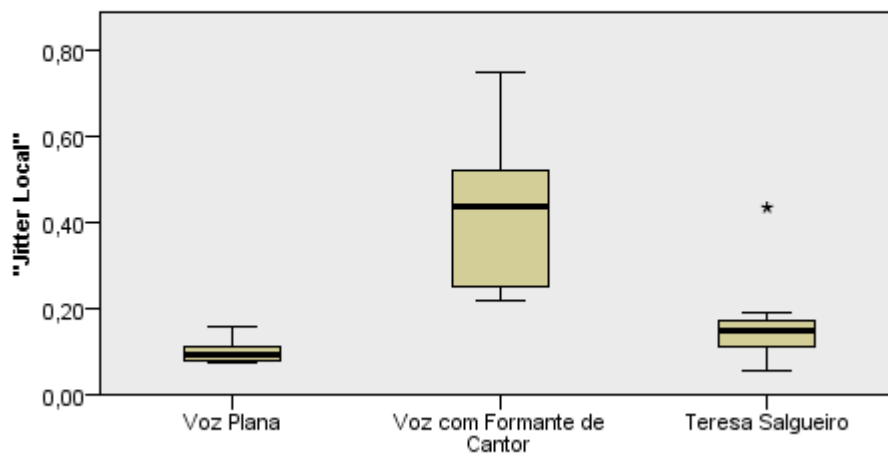


Gráfico 14 - Teste de Kruskal-Wallis para valores *jitter (local)* para exemplos de voz plana, voz com formante de cantor e exemplo de Teresa Salgueiro.

O Gráfico 14 obtido na sequência da análise do parâmetro jitter, vai de encontro ao que afirmamos anteriormente, ou seja, há muito poucas variações de frequência fundamental resultantes de uma voz sem vibrato o que reforça a afirmação de que, perceptivamente se trata de uma voz plana.

### Teste de Kruskal-Wallis de Amostras Independentes

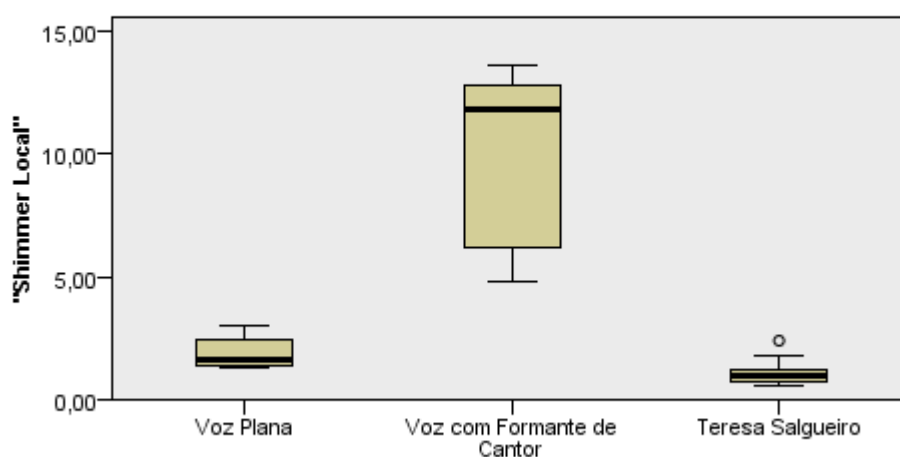


Gráfico 15 - Teste de Kruskal-Wallis para valores *shimmer (local)* para exemplos de voz plana, voz com formante de cantor e exemplo de Teresa Salgueiro.

---

Verificamos também que, as variações de amplitude (que traduzem variações de intensidade – *loudness*) demonstradas pelos excertos de voz da cantora no Gráfico 15, vão de encontro à análise perceptiva que tivemos, encontrando-se a voz da cantora, mais perto do conceito de voz plana do que da existência de formante de cantor.

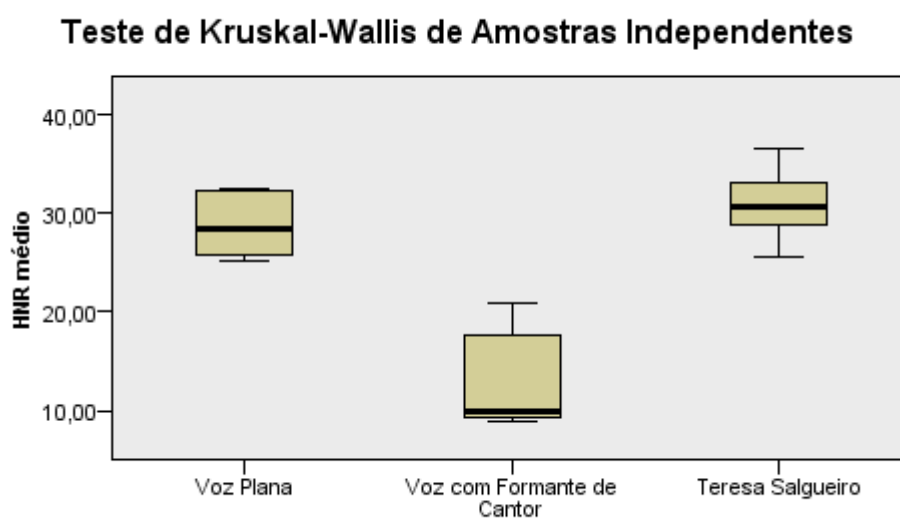


Gráfico 16 - Teste de Kruskal-Wallis para valores *HNR* médio para exemplos de voz plana, voz com formante de cantor e exemplo de Teresa Salgueiro.

A alta relação *HNR* demonstrada pelo Gráfico 16, sugere-nos um alto número de detecção de harmónicos relativamente à detecção de ruído não-harmónico. Embora este parâmetro acústico não seja de todo aquele que melhor evidencie as diferenças entre o formante de cantor e a voz plana, a inclusão do mesmo foi para provar que, para vozes com pouco ou mesmo nenhum vibrato, como é o caso da voz de Teresa Salgueiro, o algoritmo apresenta resultados mais satisfatórios estando até um pouco acima dos valores obtidos para as amostras de cantores emitindo voz plana.

#### 4.4.3. Possível existência de falsete

Através da análise perceptiva das gravações, encontramos segmentos onde nos parece que a emissão vocal foi feita em registo de falsete. Pelo menos é perceptível uma certa mudança de registo relativamente ao resto da execução vocal. Para tais segmentos

---

foram então extraídos os devidos parâmetros acústicos e foi verificada a existência de uma possível conclusão depois da comparação desses mesmos parâmetros acústicos com os obtidos anteriormente para caracterização do falsete. Os resultados obtidos foram os seguintes:

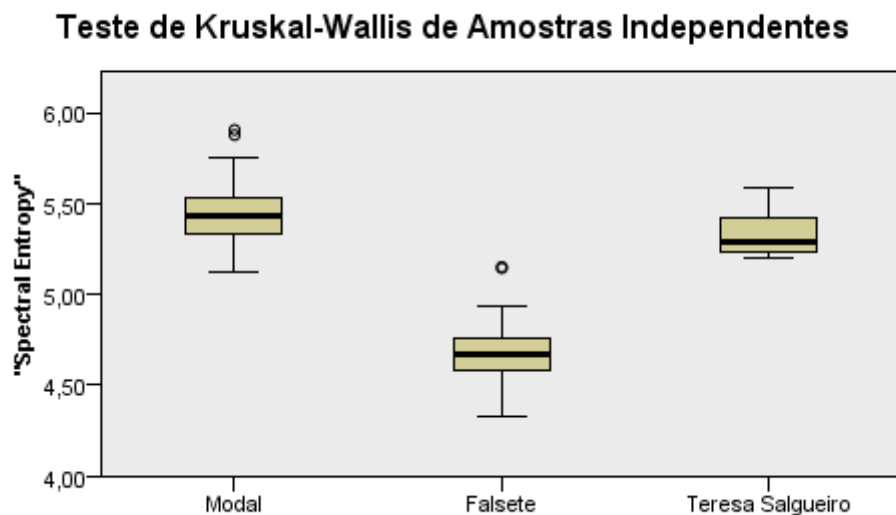


Gráfico 17 – Teste de Kruskal-Wallis para valores se *Spectral Entropy* nas vozes de Teresa Salgueiro e nas amostras em registo modal e em registo de falsete.

Pela análise do Gráfico 17 dos valores de *Spectral Entropy*, vemos que os valores deste estão um pouco entre os valores obtidos para as amostras de voz em registo modal e em registo de falsete contidas na nossa base de dados.

---

### Teste de Kruskal-Wallis de Amostras Independentes

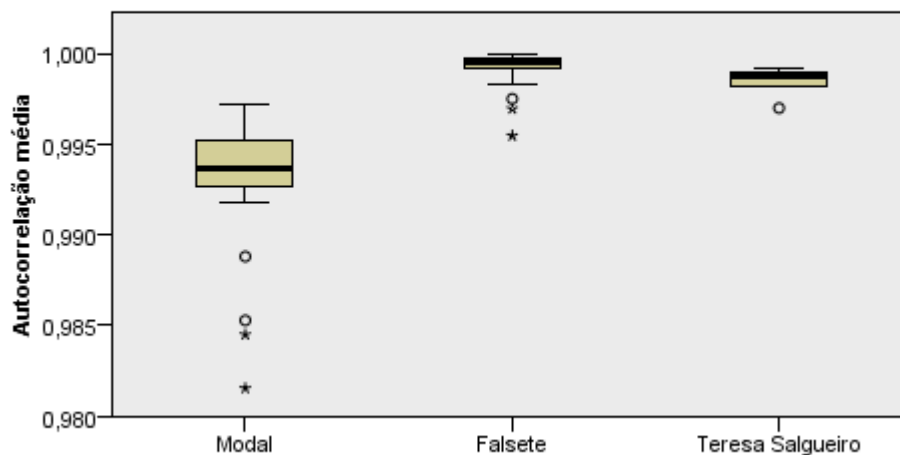


Gráfico 18 – Teste de Kruskal-Wallis para valores se Autocorrelação média nas vozes de Teresa Salgueiro e nas amostras em registo modal e em registo de falsete.

O Gráfico 18 obtido na sequência do teste de Kruskal-Wallis para os valores de Autocorrelação Média, estão de acordo com o esperado, havendo uma sintonia entre os valores obtidos para as amostras de voz em registo de falsete e os valores obtidos para a voz de Teresa Salgueiro.

Através destes resultados podemos dizer que, apesar de no caso do parâmetro *Spectral Entropy* os valores se encontrarem um pouco entre os obtidos para vozes em registo de falsete e modal, a voz de Teresa Salgueiro apresenta um registo muito aproximado do falsete.

---

## 4.5. Conclusões

Com a análise estatística levada a cabo por mim com a preciosa ajuda dos meus colegas da FEUP intervenientes no projecto conseguimos concluir que no geral, os resultados obtidos vão de encontro ao que era esperado. Confirmam também os resultados obtidos a nível spectral (e não só) no capítulo 3 desta dissertação através do SingingStudio, Praat® e MatLab®. Esta concordância entre os resultados esperados e os resultados obtidos tem no entanto um caso de insucesso: o binómio Limpidez/Soprosidade: Segundo as características que apurámos no ponto 3.3.3. os parâmetros acústicos que pensamos que melhor evidenciaríamos as diferenças entre os constituintes deste par antagónico seriam o HNR e NHR bem como o ZCR que foi de facto contabilizado. Acontece que, segundo os testes de Kruskal-Wallis e Mann-Whitney U, não aparecem como características que melhor evidenciam as diferenças entre as duas formas de emissão vocal devido ao facto de haver uma impreparação do algoritmo em reconhecer componentes harmónicos em vozes com vibrato. Este problema é inexistente no caso da voz da Teresa Salgueiro onde o algoritmo para detecção destes três parâmetros acústicos referenciados funciona aparentemente melhor. Em vozes com menos vibrato, o algoritmo apresenta menor dificuldade na detecção de harmónicos o que se comprova através dos resultados mais satisfatórios de HNR. Futuramente no desenrolar deste projecto, um aspecto a melhorar sem dúvida é a funcionalidade do algoritmo.

Houve algumas dificuldades nos cantores escolhidos por mim para efectuar vozes que se desviassem do seu padrão vocal normal como é o caso da voz com soprosidade. Uma das conclusões a que cheguei com a construção da base de dados foi o facto de os cantores treinados apresentarem alguma dificuldade em suprimir pormenores técnicos já adquiridos. Por outro lado, isto também evidencia a solidez da técnica encontrada neste tipo de cantores.

---

## Capítulo 5

---

## 5. Conclusão e Futuras Aplicações

---

O trabalho contido nesta dissertação reflecte dois objectivos primordiais. O primeiro, o de encontrar definições claras e o menos subjectivas possível para termos que são tudo menos objectivos. De facto a linguagem que é comum no universo da música e neste particular caso do canto é muitas vezes alvo de tentativas de descrição mas a verdade é que se tratam de definições de entendidos para entendidos sem ter a preocupação em que estes termos cheguem de forma clara aos ouvidos do senso comum. A colaboração com pessoas de outras áreas que não a música e o canto (neste caso de engenharia) foi preciosa na elaboração de tais definições pois sentiu-se a obvia necessidade de fazer aproximar este mundo tão subjectivo a indivíduos de um meio tão cientificamente objectivo. Penso também que foi uma experiencia enriquecedora para ambas as partes pois ficamos todos a saber mais não só sobre este assunto mas também sobre determinadas áreas que até então se mantinham inexploradas (no meu caso, tudo o que envolve análise estatística e processamento de sinal). O segundo objectivo principal desta dissertação foi o de obter um *feedback* visual de parâmetros e atributos musicais

---

e/ou estilísticos da voz cantada através da análise dos seus parâmetros acústicos. Esse objectivo foi atingido (embora no parâmetro de limpidez/soprosidade os resultados não terem sido os inicialmente esperados) de forma pioneira e com bastante empenho dos que nele trabalharam. Penso também que este projecto de investigação do qual esta dissertação faz parte, serve para aproximar estes dois mundos tão distintos e ver que, ao contrário do que possamos pensar, tudo está relacionado pela lógica. Podemos também dizer que com os resultados obtidos nesta dissertação e os que hão-de- vir, passaremos a ver as coisas com outros olhos. Foram encontradas explicações científicas e fornecidos dados visuais que suportam aquilo que para nós, músicos, já é inato e foi clarificado para pessoas interessadas no tema que, afinal, não é assim tão subjectivo como aquilo que os professores de canto defendem abraçando tudo uma lógica. Posso/podemos dizer que hoje, se deu um passo em frente na construção dessa lógica.

## 5.1. Futuras aplicações

No decorrer deste projecto foram experimentadas algumas dificuldades já acima referidas. Um dos pontos a melhorar no desenvolvimento futuro do projecto de investigação e/ou por alguém que queira continuar o que aqui foi feito é de facto alterar ou melhorar significativamente a resposta dos algoritmos para vozes com vibrato pois de facto, o vibrato é uma característica muito importante no canto lírico. Pensamos que, com o devido melhoramento, poderemos começar a explorar outros caminhos no estudo de parâmetros ainda mais subjectivos como a questão da musicalidade (em particular as microdinâmicas).

O objectivo final do projecto de investigação visa a criação de ferramentas de apoio ao ensino ou estudo do canto. Hoje foi passo em frente para esse objectivo que pretendemos aprofundar mas, no futuro, iremos precisar de mais amostras de mais cantores para a realização de algoritmos que detectem a presença ou não de alguns (ou mesmo de todos) os parâmetros acima descritos, de forma fácil e rápida.



---

# Bibliografia

---

1. Di Nicola, V, *et al* (2006), Acoustic analysis of voice in patients treated by reconstructive subtotal laryngectomy. Evaluation and critical review
2. Ekholm, E. (1998) *Relating Objective Measurements to Expert Evaluation Voice Quality in Western Classical Singing" Critical Perceptual Parameters*. Journal of Voice, Vol.12, nº2.
3. Farrús, M. *et al*. *Jitter and Shimmer Measurements for Speaker Recognition*, p. 1-3.
4. Ferreira, A. *et al*. (2008), *A medida HNR: sua relevância na análise acústica da voz e sua estimação precisa*, p. 1-4.
5. Gusmão, C. (2010) de S.; Campos, P. H.; Maia, M. E. O. *O formante do cantor e os ajustes laríngeos para utilizá-lo. Per Musi*, Belo Horizonte, n.21, p.43-50.
6. Guzman, M. N. (2010), *Acustica del tracto vocal*, [www.vozproesional.cl](http://www.vozproesional.cl)
7. Hall, D. (1991), *Musical Acoustics* , Pacific Grove
8. Henrique, L (2002), *Acústica Musical*, Cap. 16 AEROFONES: A VOZ Fundação Calouste Gulbenkian, p. 606-702
9. Knobel, M. *Física da Fala e da Audição*. Instituto de Física Gleb Wataghin/UNICAMP. <http://www.ifi.unicamp.br/~knobel/f105/>
10. Kofi, A. B. (2002/2005), *Audio Segmentation for Meetings Speech Processing*, Electrical Engineering and Computer Sciences, University of California, Berkeley

- 
11. Loscos, A. (2007), *Spectral Processing of the Singing Voice*, Universitat Pompeu Fabra
  12. McCoy, S. (2004) *Your voice: an inside view. Chapter1: Listening to singers.*
  13. Nam, U. (2001) *Special Area Exam Part II, p.1-5*
  14. Peeters, G. (2004), *A large set of audio features for sound description (Similarity and Classification) in the CUIDADO project*, p. 1, 11, 12, 17
  15. Prame, E. *Vibrato Extent and Intonation in professional western lyric singing*, p. 616-619
  16. Subramanian, H (2004), *Audio Signal Classification*, M.Tech. Credit Seminar Report, Electronic Systems Group, EE. Dept, IIT Bombay, p.1-4
  17. Sundberg, J. (1974), *Articulatoru interpretation of the singers formant*, p. 838-844
  18. Sundberg, J (1990), *What's so special about singers?*, Journal of Voice, p. 107-119
  19. Sundberg, J. (1991), *How constant is subglottal pressure in singing?*, STL-QPSR, Volume 32
  20. Sundberg, J. (1991), *The science of musical sounds*, Academic Press
  21. Sundberg, J.; Högset, C. (2001) *Voice source differences between falsetto and modal registers in counter tenors, tenors and baritones*. Logopedics Phoniatrics Vocology.

- 
22. Teixeira, J. P. *et al.* (2011), *Análise acústica vocal - determinação do Jitter e Shimmer para diagnóstico de patologias da fala*, p.1, 13-16
23. Titze, I. R. (1994), *Acoustics of the Tenor High Voice*, p. 1133-1142
24. Toh, A. M. *et. Al.* (2010), *Spectral Entropy as speech features for speech recognition*, p. 1-2
25. Vieira, M. N. (2005). Uma Introdução à Acústica da Voz cantada. *I Seminário Música Ciência Tecnologia: Acústica Musical*. Departamento de Física/ICEx/UFMG. p. 70-79