

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO



Relação entre características objetivas da voz cantada e seus atributos artísticos e estéticos.

Vítor Filipe Maia Baptista Fonseca Almeida

Mestrado Integrado em Engenharia Electrotécnica e de Computadores

Orientador: Aníbal Ferreira (Professor Doutor)

Janeiro de 2012

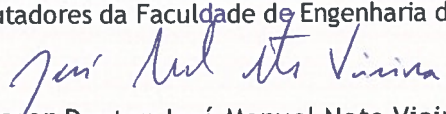
A Dissertação intitulada

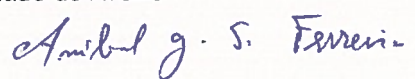
“Relação Entre Características Objectivas da Voz Cantada e Seus Atributos
Artísticos e Estéticos”

foi aprovada em provas realizadas em 15-02-2012

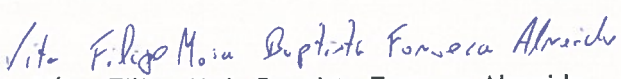
o júri


Presidente Professor Doutor Carlos Alberto Barbosa da Cunha Mendonça Guedes
Professor Associado Convidado do Departamento de Engenharia Electrotécnica e de
Computadores da Faculdade de Engenharia da Universidade do Porto


Professor Doutor José Manuel Neto Vieira
Professor Auxiliar Departamento de Electrónica, Telecomunicações e Informática da
Universidade de Aveiro


Professor Doutor Aníbal João de Sousa Ferreira
Professor Associado do Departamento de Engenharia Electrotécnica e de
Computadores da Faculdade de Engenharia da Universidade do Porto

O autor declara que a presente dissertação (ou relatório de projeto) é da sua exclusiva autoria e foi escrita sem qualquer apoio externo não explicitamente autorizado. Os resultados, ideias, parágrafos, ou outros extratos tomados de ou inspirados em trabalhos de outros autores, e demais referências bibliográficas usadas, são corretamente citados.


Autor - Vitor Filipe Maia Baptista Fonseca Almeida

Faculdade de Engenharia da Universidade do Porto

Resumo

A classificação da qualidade da voz, de um cantor, está geralmente associada à percepção auditiva. Existem vários parâmetros de avaliação e classificação de voz cantada, mas com várias definições pouco coerentes, sendo normalmente de carácter subjetivo e dependendo dos gostos de quem avalia. Estes parâmetros estão normalmente associados ao resultado final; sem haver a preocupação de compreender os fenómenos que originam um determinado tipo de voz, ou uma característica que a faça distinguir e sobressair do ponto de vista estético ou de qualidade global.

Esta dissertação enquadra-se num projeto financiado pela Fundação para a Ciência e Tecnologia com vista ao desenvolvimento de ferramentas de apoio ao ensino e aprendizagem de canto.

Neste projeto de dissertação foram estudados os parâmetros perceptivos e de qualidade associados ao canto. Procurou-se, assim, clarificar quais os fenómenos responsáveis por essas características perceptivas da voz, e as suas origens, para que estes parâmetros possam ser caracterizados em termos acústicos e perceber de que forma se podem desenvolver ferramentas de análise e de *feedback* visual.

Dos estudos efetuados sobre os parâmetros perceptivos, foi possível identificar a importância que as ressonâncias do trato vocal têm para a caracterização da qualidade da voz. Assim, foram estudados métodos de estimação de formantes para voz cantada. O objetivo é o desenvolvimento de um sistema robusto para estimação de formantes, utilizando tanto informação de todo o sinal como do ruído da voz, e de representação em tempo real sobre um espectrograma.

Durante o decorrer do projeto, foi construída uma base de dados de canto com cantores treinados que serviu de elemento de teste e estudo e servirá, também, como material de apoio para as várias realizações do projeto financiado pela Fundação para a Ciência e Tecnologia, no qual esta dissertação se enquadra.

Abstract

Classification of a singer's voice is usually associated with auditory perception. There are several parameters for evaluation and classification of singing voice, but several inconsistent definitions, being usually subjective in nature and dependent on aesthetical preferences of who evaluates. These parameters are usually associated with the final result, with no concern to understand the phenomena that cause a particular type of voice or a characteristic that makes it stand out and distinguish from aesthetic point of view or overall quality.

This dissertation is part of a project supported by “Fundação para a Ciência e Tecnologia”, to develop support tools for teaching and learning of singing.

In this dissertation perceptual and quality parameters associated with the singing voice were studied in order to clarify which phenomena are responsible for these perceptual characteristics of voice, understanding their origins, so that these parameters can be categorized in terms of acoustic and see how we can develop tools to help and support the learning/teaching of singing.

With the studies carried out in the perceptual parameters, it was possible to identify the importance of the vocal tract resonances to characterize the voice quality. Thus, we studied methods for formants estimation in the singing voice. The goal is the development of a robust formants estimation method for singing voice, using the entire signal and also the noise of the voice, so that they can be represented in real time on a spectrogram.

During the course of the project, a database of singing voices was recorded by trained singers. This database served as element of study and test for this project. It will also serve to support the various achievements of the project financed by the “Fundação para a Ciência e Tecnologia”, where this thesis fits.

Agradecimentos

Quero agradecer ao Professor Doutor Aníbal Ferreira pela sua orientação e apoio, bem como pela oportunidade de trabalhar neste projeto. Uma palavra também ao Doutor Ricardo Sousa pela sua disponibilidade e apoio durante a dissertação; bem como a todas as pessoas com quem partilhei o espaço de trabalho, pelo ambiente de entreaajuda criado.

Aos meus pais, por todo o apoio e disponibilidade, principalmente nas adversidade do caminho, onde estiveram sempre ao meu lado. À minha irmã, avó e restante amigos, pelo apoio durante toda esta caminhada.

Vítor Almeida

“A ciência de hoje é a tecnologia de amanhã.”

Edward Teller

Conteúdo

1	Introdução	1
1.1	Motivação	1
1.2	Enquadramento	2
1.3	Estrutura	2
2	Caracterização do estado da arte	3
2.1	Introdução	3
2.2	Voz	3
2.3	Voz Cantada	5
2.3.1	Frequência Fundamental	6
2.3.2	Vibrato	6
2.3.3	Formantes e formante de cantor	7
2.3.4	Vozeamento	9
2.4	Percepção	9
2.5	Parâmetros qualitativos e perceptivos	12
2.5.1	Afinação	12
2.5.2	Tessitura	12
2.5.3	Timbre	12
2.5.4	Falsete	15
2.5.5	Ataque	15
2.5.6	Vibrato	15
2.5.7	Legato	15
2.5.8	Staccato	16
2.5.9	Micro-dinâmicas e destreza vocal	16
2.6	Aplicações	17
2.6.1	Estimação e mapeamento de formantes	17
2.6.2	Deteção e caracterização de vibrato	19
2.7	Conclusões	19
3	Estimação e mapeamento de formantes	21
3.1	Introdução	21
3.2	Modelo fonte-filtro	21
3.3	Estimação de trato vocal	22
3.3.1	Linear Prediction Coding	22
3.3.2	Cepstrum	23
3.4	Estimação da frequência dos formantes	24
3.5	Mapeamento temporal dos formantes	30
3.6	Conclusões	33

4	Teste e resultados	35
4.1	Introdução	35
4.2	Base de dados	35
4.2.1	Voz sintetizada	35
4.2.2	Voz natural	37
4.3	Testes	38
4.3.1	Testes com vozes sintéticas	39
4.3.2	Testes com vozes naturais	42
4.4	Conclusões	51
5	Conclusão	53
5.1	Trabalho Futuro	54
A	Anexo A - Descrições estatísticas dos testes	55
A.1	Descrições estatísticas dos testes com vozes naturais	55
A.1.1	<i>Praat</i>	55
A.1.2	Métodos implementados	58
A.2	Descrições estatísticas dos testes com vozes sintéticas	66
A.2.1	<i>Praat</i>	66
A.2.2	Métodos implementados	68
A.2.3	Gráficos	74
A.2.4	Relação de proximidade aos harmónicos	78
B	Anexo B - Equipamentos utilizados nas gravações	81
B.1	Microfone	81
B.2	Pré-amplificador	82
	Referências	83

Lista de Figuras

2.1	Trato Vocal	4
2.2	Espectrogramas da mesma frase na forma cantada e falada	5
2.3	Exemplos de vibrato (<i>adaptada</i>)	7
2.4	Afinação de formantes	8
2.5	Espectro do formante de cantor e sua comparação com a orquestra (<i>adaptada</i>) . .	9
2.6	Jitter, Shimmer, SNR e D2 para seis estilos de música	10
2.7	Catálogo de tipos de voz masculinas consoante a tessitura (<i>Fach</i>)	13
2.8	Catálogo de tipos de voz femininas consoante a tessitura (<i>Fach</i>)	13
2.9	Erros típicos de estimação de formantes (<i>adaptada</i>)	18
2.10	Espectrograma com o <i>tracking</i> dos formantes encontrados (Vogal 'o' de um Contralto), de notar que o padrão de cores está inverso ao habitual.	19
3.1	Modelo fonte-filtro (<i>adaptada</i>)	22
3.2	Exemplo de envolvente obtida com modelo baseado em LPC (preto) do espectro de um sinal (azul) para modelação do trato vocal	23
3.3	Primeiro nível de processamento.	25
3.4	Processo de estimação do resíduo	26
3.5	As envolventes espectrais calculadas	27
3.6	Processo de estimação de candidatos a formantes pela segunda derivada da envolvente espectral	28
3.7	Segundo nível de processamento.	29
3.8	Dois possíveis erros de <i>tracking</i> (Vogal 'o' de um Contralto)	30
3.9	Exemplificação da alocação de formantes no <i>stream</i>	31
3.10	Processamento do <i>tracking</i> dos formantes	31
3.11	Pormenor da possibilidade de erro por proximidade (3º e 4º formantes da vogal 'a' de um Baixo).	32
4.1	Diagrama de blocos do sintetizador utilizado	36
4.2	Diferença dos resultados dos dois sintetizadores abordados	37
4.3	Configuração de estimação de formantes do <i>Praat</i>	38
4.4	Gráfico representativo do erro médio de estimação dos formantes da vogal sintética 'i'	40
4.5	Gráfico representativo do erro médio de estimação dos formantes da vogal sintética 'a'	41
4.6	Estimação de formantes da vogal 'a' de um Baixo	42
4.7	Estimação de formantes da vogal 'i' de um Baixo	43
4.8	Estimação de formantes da vogal 'a' de um Tenor	44
4.9	Estimação de formantes da vogal 'i' de um Tenor	45

4.10	Estimação de formantes da vogal 'a' de uma Contralto	46
4.11	Estimação de formantes da vogal 'i' de uma Contralto	47
4.12	Estimação de formantes da vogal 'a' de uma Soprano	48
4.13	Estimação de formantes da vogal 'i' de uma Soprano	49
A.1	Estatísticas relativos do 1º formante da vogal sintética 'i'	74
A.2	Estatísticas relativos do 2º formante da vogal sintética 'i'	74
A.3	Estatísticas relativos do 3º formante da vogal sintética 'i'	75
A.4	Estatísticas relativos do 4º formante da vogal sintética 'i'	75
A.5	Estatísticas relativos do 5º formante da vogal sintética 'i'	75
A.6	Estatísticas relativos do 1º formante da vogal sintética 'i'	76
A.7	Estatísticas relativos do 2º formante da vogal sintética 'i'	76
A.8	Estatísticas relativos do 3º formante da vogal sintética 'i'	76
A.9	Estatísticas relativos do 4º formante da vogal sintética 'i'	77
A.10	Estatísticas relativos do 5º formante da vogal sintética 'i'	77
B.1	Especificações do microfone utilizado nas gravações	81
B.2	Especificações do pré-amplificador utilizado nas gravações	82

Lista de Tabelas

2.1	Exemplos de cantores Claro/Escuro	14
2.2	Exemplos de cantores Voz leve/Voz pesada	14
4.1	Frequências dos formantes das vogais sintetizadas	36
4.2	Erros relativos das frequências estimadas pelos vários métodos para a vogal 'i'	39
4.3	Erros relativos em relação aos harmônicos teóricos, vogal 'i' com F0 = 880 Hz	40
4.4	Erros relativos das frequências estimadas pelos vários métodos para a vogal 'a'	40
4.5	Erros relativos em relação aos harmônicos teóricos, vogal 'a' com F0 = 880 Hz	41
4.6	Médias dos formantes, vogal 'a' de um Baixo	43
4.7	Médias dos formantes, vogal 'i' de um Baixo	44
4.8	Médias dos formantes, vogal 'a' de um Tenor	45
4.9	Médias dos formantes, vogal 'i' de um Tenor	46
4.10	Médias dos formantes, vogal 'a' de uma Contralto	47
4.11	Médias dos formantes, vogal 'i' de uma Contralto	48
4.12	Médias dos formantes, vogal 'a' de uma Soprano	49
4.13	Médias dos formantes, vogal 'i' de um Soprano	50
A.1	Descrições estatísticas dos resultados do <i>Praat</i> da vogal 'a' de um Baixo	55
A.2	Descrições estatísticas dos resultados do <i>Praat</i> da vogal 'i' de um Baixo	55
A.3	Descrições estatísticas dos resultados do <i>Praat</i> da vogal 'a' de um Tenor	56
A.4	Descrições estatísticas dos resultados do <i>Praat</i> da vogal 'i' de um Tenor	56
A.5	Descrições estatísticas dos resultados do <i>Praat</i> da vogal 'a' de uma Contralto	56
A.6	Descrições estatísticas dos resultados do <i>Praat</i> da vogal 'i' de uma Contralto	57
A.7	Descrições estatísticas dos resultados do <i>Praat</i> da vogal 'a' de uma Soprano	57
A.8	Descrições estatísticas dos resultados do <i>Praat</i> da vogal 'i' de uma Soprano	57
A.9	Descrições estatísticas dos resultados dos métodos implementados da vogal 'a' de um Baixo	58
A.10	Descrições estatísticas dos resultados dos métodos implementados da vogal 'i' de um Baixo	59
A.11	Descrições estatísticas dos resultados dos métodos implementados da vogal 'a' de um Tenor	60
A.12	Descrições estatísticas dos resultados dos métodos implementados da vogal 'i' de um Tenor	61
A.13	Descrições estatísticas dos resultados dos métodos implementados da vogal 'a' de uma Contralto	62
A.14	Descrições estatísticas dos resultados dos métodos implementados da vogal 'i' de uma Contralto	63

A.15	Descrições estatísticas dos resultados dos métodos implementados da vogal 'a' de uma Soprano	64
A.16	Descrições estatísticas dos resultados dos métodos implementados da vogal 'i' de uma Soprano	65
A.17	Descrições estatísticas dos resultados do <i>Praat</i> da vogal sintética 'i' com frequência fundamental de 220 Hz	66
A.18	Descrições estatísticas dos resultados do <i>Praat</i> da vogal sintética 'i' com frequência fundamental de 440 Hz	66
A.19	Descrições estatísticas dos resultados do <i>Praat</i> da vogal sintética 'i' com frequência fundamental de 880 Hz	66
A.20	Descrições estatísticas dos resultados do <i>Praat</i> da vogal sintética 'a' com frequência fundamental de 220 Hz	67
A.21	Descrições estatísticas dos resultados do <i>Praat</i> da vogal sintética 'a' com frequência fundamental de 440 Hz	67
A.22	Descrições estatísticas dos resultados do <i>Praat</i> da vogal sintética 'a' com frequência fundamental de 880 Hz	67
A.23	Descrições estatísticas dos resultados dos métodos implementados da vogal sintética 'i' com frequência fundamental de 220 Hz	68
A.24	Descrições estatísticas dos resultados dos métodos implementados da vogal sintética 'a' com frequência fundamental de 220 Hz	69
A.25	Descrições estatísticas dos resultados dos métodos implementados da vogal sintética 'i' com frequência fundamental de 440 Hz	70
A.26	Descrições estatísticas dos resultados dos métodos implementados da vogal sintética 'a' com frequência fundamental de 440 Hz	71
A.27	Descrições estatísticas dos resultados dos métodos implementados da vogal sintética 'i' com frequência fundamental de 880 Hz	72
A.28	Descrições estatísticas dos resultados dos métodos implementados da vogal sintética 'a' com frequência fundamental de 880 Hz	73
A.29	Relações dos resultados estimados com os harmônicos teóricos da vogal sintética 'i' com frequência fundamental de 220 Hz	78
A.30	Relações dos resultados estimados com os harmônicos teóricos da vogal sintética 'a' com frequência fundamental de 220 Hz	78
A.31	Relações dos resultados estimados com os harmônicos teóricos da vogal sintética 'i' com frequência fundamental de 440 Hz	78
A.32	Relações dos resultados estimados com os harmônicos teóricos da vogal sintética 'a' com frequência fundamental de 440 Hz	79
A.33	Relações dos resultados estimados com os harmônicos teóricos da vogal sintética 'i' com frequência fundamental de 880 Hz	79
A.34	Relações dos resultados estimados com os harmônicos teóricos da vogal sintética 'a' com frequência fundamental de 880 Hz	79

Abreviaturas e Símbolos

FEUP	Faculdade de Engenharia da Universidade do Porto
FCT	Fundação para a Ciência e Tecnologia
ESMAE	Escola Superior de Música e Artes do Espetáculo
GIRBAS	<i>Grade, instability, roughness, breathiness, asthenia, and strain</i>
LPC	<i>Linear Predictive Coding</i>
HMM	<i>Hidden Markov Models</i>
ODFT	<i>Odd-Discrete Fourier Transform</i>
F0	Frequência fundamental
F _{<i>n</i>}	Formante <i>n</i>
H _{<i>n</i>}	Harmónico <i>n</i>

Capítulo 1

Introdução

1.1 Motivação

De um modo geral, a voz humana, como importante órgão de comunicação, tem sido tema para estudos dentro de vários âmbitos, desde o foro médico com estudos da qualidade e da correta utilização da voz, ao foro forense como estudos para a criação de soluções de identificação de orador.

A voz seja ela falada ou cantada, é o resultado da articulação dos órgãos que compõem o aparelho fonatório humano, sendo assim um fenómeno sujeito a variações recorrentes de vários fatores, como condições físicas, psicológicas, médicas ou de postura.

Devido a estas características complexas do sistema de voz, surge a necessidade de estudar e relacionar entre várias características, de forma a melhor poder compreender o seu mecanismo; podem, assim, ser criadas soluções de reconhecimento de orador, de estudo das condições fisiológicas e psicológicas do orador, de medição da qualidade da voz e, no caso da voz cantada, de mensuração e avaliação tanto do ponto de vista de saúde, como do ponto de vista artístico e estético.

Do ponto de vista de estudos da voz cantada e a relação entre as características acústicas da mesma e os princípios de avaliação perceptiva dessa mesma voz do ponto de vista artístico e estético, poucos são os estudos existentes. Salienta-se a ambiguidade nas relações já estabelecidas entre as características acústicas e esses atributos de avaliação perceptiva, faltando estabelecer um certo consenso e normalização destas definições.

É neste contexto que se enquadra este projeto de dissertação, com o propósito do desenvolvimento de novas ferramentas de correspondência entre as características objetivas e os atributos perceptivos de avaliação e caracterização; bem como ferramentas de *feedback* visual que forneçam ao cantor e ao seu formador informação em tempo real das características acústicas da voz, relacionando-as com possíveis atributos perceptivos utilizados na avaliação do tipo e qualidade da voz de um cantor.

1.2 Enquadramento

O projeto de dissertação “Relação entre características objetivas da voz cantada e seu atributos artísticos e estéticos” enquadra-se num projeto financiado pela Fundação para a Ciência e Tecnologia (FCT) que visa o desenvolvimento de tecnologias interativas de apoio ao ensino e aprendizagem de canto, bem como para a monitorização preventiva da voz, seja na forma cantada ou falada.

Este projeto FCT é multidisciplinar, englobando especialistas nas áreas de ensino de canto, engenharia e otorrinolaringologia.

Tem como objetivo desenvolver soluções que auxiliem e otimizem o ensino e treino de canto, bem como a sua execução correta e segura prevenindo eventuais distúrbios vocais. Assim é esperado o desenvolvimento de um sistema de *feedback* visual em tempo real da qualidade de vários parâmetros da voz cantada de forma a ser aglutinado ao *software* *SingingStudio*® anteriormente desenvolvido pela *spin-off* da Faculdade de Engenharia da Universidade do Porto *Seegnal*.

A tarefa de relacionamento entre características subjetivas e parâmetros acústicos da voz cantada tem como objetivo identificar e caracterizar parâmetros de avaliação perceptiva da voz cantada, investigando que características acústicas melhor se correlacionam, e desenvolver algoritmos eficientes para estimar essas relações. Estes algoritmos serão então utilizados no desenvolvimento de uma plataforma interativa de apoio ao ensino de canto para implementação de uma representação visual e em tempo real dos parâmetros perceptivos associados. Para este efeito, este projeto é desenvolvido em articulação com um cantor da Escola Superior de Música e Artes do Espetáculo (ESMAE) do Instituto Politécnico do Porto, na vertente da definição dos conceitos perceptivos de caracterização da voz cantada.

1.3 Estrutura

Esta dissertação encontra-se dividida em cinco capítulos. Neste primeiro é feita uma introdução ao tema do projeto bem como ao seu enquadramento.

No segundo capítulo é feita uma descrição de conceitos fundamentais de características acústicas e perceptivas da voz, em especial na forma cantada; bem como a revisão do estado de arte no âmbito do tema desta dissertação.

No terceiro capítulo é descrito o sistema de estimação e mapeamento de formantes que foi desenvolvido, sendo descritos os métodos e as ferramentas utilizadas bem como as dificuldades que foram encontradas durante o desenvolvimento.

No quarto capítulo são descritos todos os testes efetuados, bem como a elaboração da base de dados de canto utilizada nesta dissertação.

Para terminar, no quinto capítulo, constam as conclusões finais do trabalho realizado, tendo em conta os resultados conseguidos; e a descrição de trabalho futuro a realizar de forma a melhorar o desempenho do sistema desenvolvido bem como na construção e implementação de novos sistemas relacionados com o tema.

Capítulo 2

Caracterização do estado da arte

2.1 Introdução

Neste capítulo é feita uma breve descrição da voz distinguindo a voz falada da voz cantada e enumeradas as características objetivas relevantes para o desenvolvimento desta dissertação.

É feito um estudo de parâmetros de avaliação perceptiva da voz cantada que foi levado a cabo por mim, bem como um resumo do levantamento efetuado pelo cantor da ESMAE que trabalhou em articulação comigo nesta vertente do projeto.

É também feito um levantamento de métodos de análise de características acústicas de sinal relevantes para o desenvolvimento desta dissertação, mais concretamente metodologias de estimação e mapeamento de formantes.

2.2 Voz

A produção de voz começa com o aumento da pressão do ar nos pulmões originando assim um fluxo de ar que irá passar pelas pregas vocais. Se as pregas vocais estiverem aduzidas haverá então resistência à saída do ar, causando assim a vibração das pregas vocais a qual gera o chamado som laríngeo que é fonte sonora da voz, sendo a base da fala e do canto [1]. Este som é composto pela frequência fundamental que é a frequência de vibração das pregas vocais e pelos seus parciais harmónicos [1] [2]. A frequência fundamental depende das características morfológicas das pregas vocais e da laringe. Assim, as diferenças na frequência fundamental entre crianças, adultos do sexo feminino e adultos do sexo masculinos são devidos aos diferentes tamanhos da laringe e das pregas vocais [1].

O som laríngeo passa então pelo conjunto de cavidades supra-glóticas constituídos pela laringe, faringe, boca e cavidade nasal normalmente designado de trato vocal [1], sendo este não mais do que uma caixa de ressonância na qual a transmissão dos parciais do som laríngeo vai ser dependente das frequências de ressonância sendo estes parciais modelados de forma diferente ao longo do espectro, uns mais atenuados que outros [2].

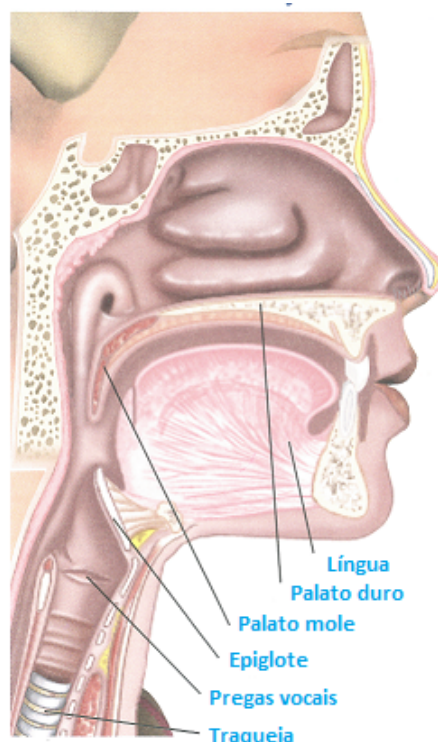


Figura 2.1: Trato Vocal [3]

O trato vocal, representado na Figura 2.1 é responsável por quatro ou cinco importantes ressonâncias chamadas de formantes. A presença destes formantes impõe picos ao espectro do som laríngeo modificando a sua envolvente espectral. São estas perturbações que permitem a produção de sons distintos [2] como a produção das várias vogais.

As frequências destas ressonâncias podem ser alteradas por meio da modificação do trato vocal, sendo dependentes da forma do trato vocal [2]. Assim, uma constrição ou uma expansão de um ponto do trato vocal afeta a localização das frequências dos formantes, sendo que o trato vocal tem a característica de poder ser modificado com relativa rapidez [1], facilitando assim modificações acústicas ao som laríngeo, de forma à produção linguística.

Segundo Sundberg [2] existem três importantes formas de modelação do trato vocal de forma a deslocar a frequência de três das importantes ressonâncias, que originam os três formantes de frequência mais baixa. Assim, a abertura da mandíbula é decisiva na localização da frequência do formante de frequência mais baixa, primeiro formante; quanto mais aberta a mandíbula, mais alta é a frequência do primeiro formante. Quanto à frequência do segundo formante, esta é sensível à variação da forma do tronco da língua, enquanto que a posição da extremidade da língua é a maior responsável pela variação de frequência do terceiro formante. Com estes três importantes formantes, a envolvente espectral pode ser modelada de forma a originar várias configurações associadas às várias vogais.

2.3 Voz Cantada

Falar e cantar são dois modos de uso do mesmo sistema [4]; sendo que os dois têm muitas propriedades em comum por serem produtos do mesmo mecanismo, mas há importantes diferenças entre os dois, desde diferenças perceptivas às diferentes características acústicas.

Uma representação ilustrativa desta diferença entre a voz falada e cantada está na Figura 2.2 onde se pode verificar que a forma cantada é normalmente mais sustentada, tendo também a energia mais distribuída pelo espectro ou a maior separação entre os harmônicos.

Pelos resultados de estudos que mostram que um humano distingue a voz cantada da voz falada com 70% e 95% de exatidão para amostras de trezentos milissegundos e um segundo respectivamente [5] podemos concluir que os dois modos são bem distintos.

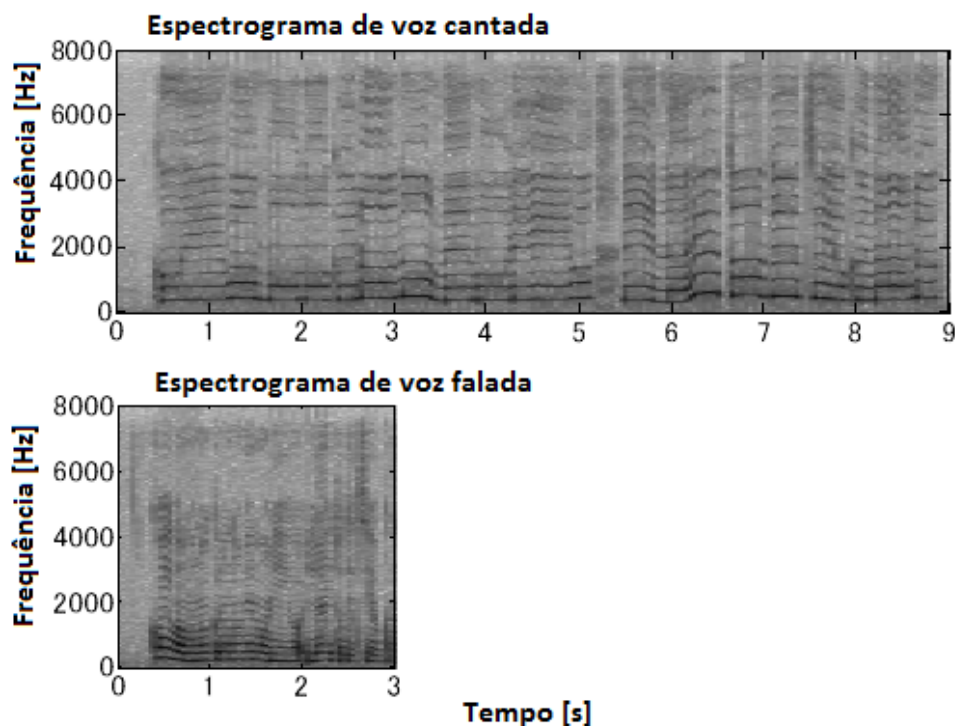


Figura 2.2: Espectrogramas da mesma frase na forma cantada e falada [5] (*adaptada*)

Segundo estudos de Johan Sundberg [6], os cantores utilizam a voz de uma forma bem diferente dos não cantores, mostrando uma grande independência de parâmetros fonatórios na produção de voz [1].

Existem várias diferenças entre a voz falada e a cantada, por exemplo, as durações dos fonemas das vogais no caso da voz cantada são superiores às da voz falada; sendo o *loudness* (intensidade), na voz cantada normalmente superior ao da voz falada.

Quanto ao *pitch* (frequência fundamental), geralmente é superior na forma cantada em relação à forma falada, sendo que na forma cantada poderá existir uma variação de forma a introduzir uma certa percepção de musicalidade à sua produção.

2.3.1 Frequência Fundamental

Como já referido, pela frequência fundamental podemos distinguir a voz cantada da voz falada.

Um cantor treinado quando canta utiliza uma gama de frequências fundamentais superior às utilizadas na voz falada [7], estando compreendida entre os 80 e os 400 Hz na voz falada, enquanto na voz cantada a gama é mais ampla podendo chegar ao 1400 Hz na voz de um soprano [8]; no geral, a frequência fundamental é superior na voz cantada em relação à da voz falada. Para além desta diferença, na voz cantada a energia da frequência fundamental é também ela superior à da voz falada.

Na voz falada a variação da frequência fundamental está associada a estados emocionais [7] enquanto a variação no caso da voz cantada, é controlada pelo cantor [9].

A nível percetivo, e numa primeira instância, correlaciona-se diretamente a frequência fundamental com o *Pitch* ou “altura” do som.

2.3.2 Vibrato

Uma outra importante característica acústica da voz cantada é o vibrato, que do ponto de vista percetivo é uma modificação periódica ou quase periódica da frequência fundamental de uma determinada nota [10]. Não existe na voz falada, sendo então considerada uma especificação musical [11] que não se encontra presente somente na voz, mas em instrumentos musicais. De facto é mais evidente percetivamente a sua presença nestes instrumentos musicais do que na voz [10].

Do ponto de vista acústico consiste numa variação de frequência de um conjunto de parciais que compõem a voz, como ilustrado na Figura 2.3, normalmente acompanhada também de uma variação de amplitude [11]. Esta condição acústica está relacionada com características percetivas tais como pitch, timbre e sonoridade; sendo que destas o pitch é o mais estudado [12].

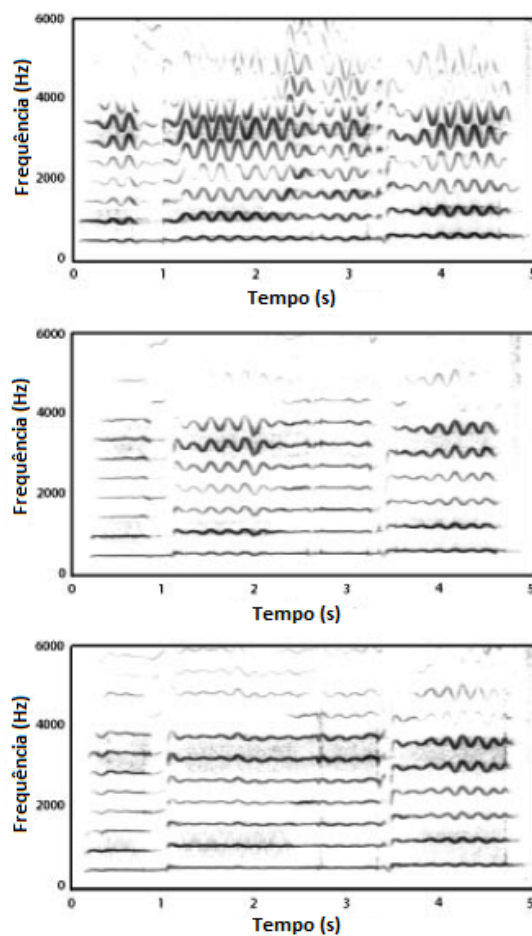
Embora seja uma das características mais abordadas no estudo do canto, muitos aspetos permanecem ainda pouco claros; como a razão das suas características acústicas mais relevantes ou a relação com aspetos psicológicos [12].

O vibrato pode ser caracterizado pela sua frequência, a sua regularidade, extensão e forma de onda [14]. A frequência do vibrato é considerada constante para cada cantor, sendo difícil de ser alterado até pelo treino [1] e normalmente encontra-se compreendida entre os 5 e 7 Hz [1] [10] [14].

O seu estudo tem sido levado a cabo seguindo abordagens diferentes, dependendo da relação ou característica a estudar os investigadores têm escolhido a abordagem que acham mais pertinente para o caso.

Estudos que contemplam só a variação da amplitude no vibrato foram levados a cabo recorrendo a uma abordagem de modelo sinusoidal [12] utilizando modelos de sinais puros sem dar grande importância à relação entre a variação da amplitude e a variação da frequência, ignorando processos importantes presentes na produção de voz [12].

Devido à complexidade na produção de voz, e sem descurar os processos que nas abordagens de modelo sinusoidal são ignorados, entres os quais a relação entre a cavidade sub-glótica

Figura 2.3: Exemplos de vibrato (*adaptada*) [13]

e a supra-glótica, têm sido utilizadas abordagens de sistemas interativos fonte-filtro no estudo do vibrato [15].

Quanto a uma possível correlação com características perceptivas para além da sentida variação da nota, as definições perceptivas sobre o vibrato baseiam-se na sua qualidade, considerando o vibrato como uma característica perceptiva.

Segundo Johan Sundberg [6], vibratos abaixo do 5.5 Hz são vibratos demasiado "lentos" e acima dos 7.5 Hz são demasiados "nervosos" [1].

2.3.3 Formantes e formante de cantor

Um possível problema na voz cantada prende-se com o fato de a frequência fundamental poder ser superior à frequência do primeiro ou até dos dois primeiros formantes de uma vogal, o que a acontecer torna o som fraco, sem que o cantor tire partido das ressonâncias do trato vocal de forma a amplificar o som da sua voz. Assim, surge a necessidade de elevar a frequência dos primeiros formantes, arrastando o primeiro formante para a frequência fundamental ilustrada na Figura 2.4, utilizando assim todo o potencial das ressonâncias do trato vocal de forma a amplificar a voz, diminuindo o esforço vocal [2].

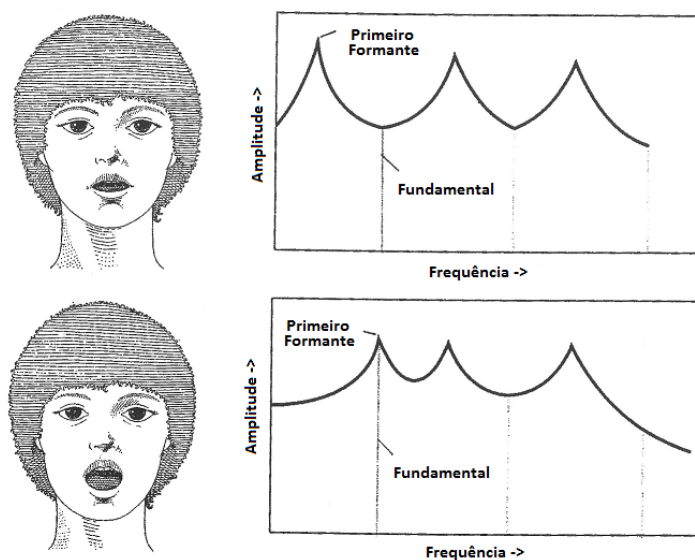


Figura 2.4: Afinação de formantes [2] (*adaptada*)

Devido à importância dos formantes na definição das vogais, ao ser efetuado este deslocamento dos formantes é de esperar que se perca a capacidade de identificação das mesmas, mas isso não acontece assim; principalmente por estarmos largamente habituados a ouvir vogais produzidas com diversas frequências fundamentais e com localizações diferentes dos formantes que as originam [2].

Do ponto de vista acústico, uma das principais características que destaca a voz cantada da voz falada é a existência de um formante extra de elevada energia situado entre os 2000 e os 3500 Hz, vulgarmente designado de formante dos cantores [1] que facilita a percepção da voz do cantor quando o canto é acompanhado por orquestra [6], estando o seu efeito ilustrado na Figura 2.5.

De acordo com estudos anteriores, o formante de cantor é um fenómeno de ressonância de um conjunto formado pelos formantes três, quatro e cinco, onde a separação entre estes três formantes é menor na voz cantada do que na voz falada.

O nível e a frequência central deste formante extra é relacionado com a frequência fundamental, a vogal emitida, a intensidade, entre outros fatores acústicos [1] [7] [6].

O nível do formante do cantor aumenta de entre 16 e 19 dB, dependendo do cantor ou da vogal por cada aumento do nível de som global em 10 dB [6].

Quanto à frequência central deste formante, segundo Johan Sundberg [6], relaciona-se com a classificação da voz; sendo menor para cantores cuja voz é classificada como baixos e maior para tenores; no mesmo estudo, também, Johan Sundberg mostrou que para a maioria dos sopranos observam-se dois picos no formante do cantor, o que sugere que os formantes que o constituem não se encontram muito próximos [6].

Sendo um fenómeno de ressonância, é normalmente associado a características perceptivas também relacionadas com a ressonância [16].

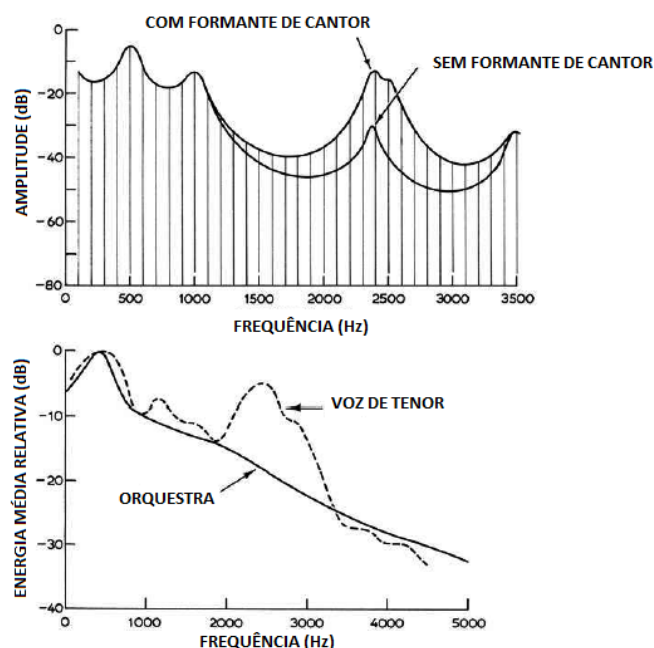


Figura 2.5: Espectro do formante de cantor e sua comparação com a orquestra [2] (adaptada)

2.3.4 Vozeamento

Outra característica da voz cantada em relação à voz falada é o rácio de som vozeado e não vozeado que constitui esses dois modos. Enquanto, na voz falada teremos um rácio de 60% de sons vozeados, no caso da voz cantada esse rácio pode subir até aos 95% [7].

Para além deste aumento do rácio de sons vozeados é também possível que algumas vogais, sejam modificadas propositadamente ou involuntariamente no caso da voz cantada [9].

2.4 Percepção

A qualidade da voz, seja do ponto de vista fisiológico clínico ou artístico, tem sido um dos temas abordados por investigadores, mas definir a qualidade da voz é problemático, pois a qualidade de um som está normalmente associada à sensação auditiva [17] que é claramente multidimensional. Neste sentido aparece a necessidade de estudar a correlação entre características perceptivas e características acústicas da voz, como uma alternativa, tanto para a medição da qualidade da voz falada como da voz cantada.

Existem estudos com resultados contraditórios em relação à correlação entre os parâmetros de avaliação perceptiva e as características acústicas; isto provavelmente acontece por existirem várias normalizações para a avaliação de voz, não havendo em certos casos consenso nas suas definições.

Estudos sobre a correlação dos parâmetros GIRBAS (*Grade, instability, roughness, breathiness, asthenia, and strain*) para avaliação de vozes patológicas e características acústicas da voz chegam à conclusão que esta correlação, a existir, não é obtida linearmente por nenhum dos conhecidos parâmetros acústicos; havendo vários motivos que dificultam este relacionamento [18].

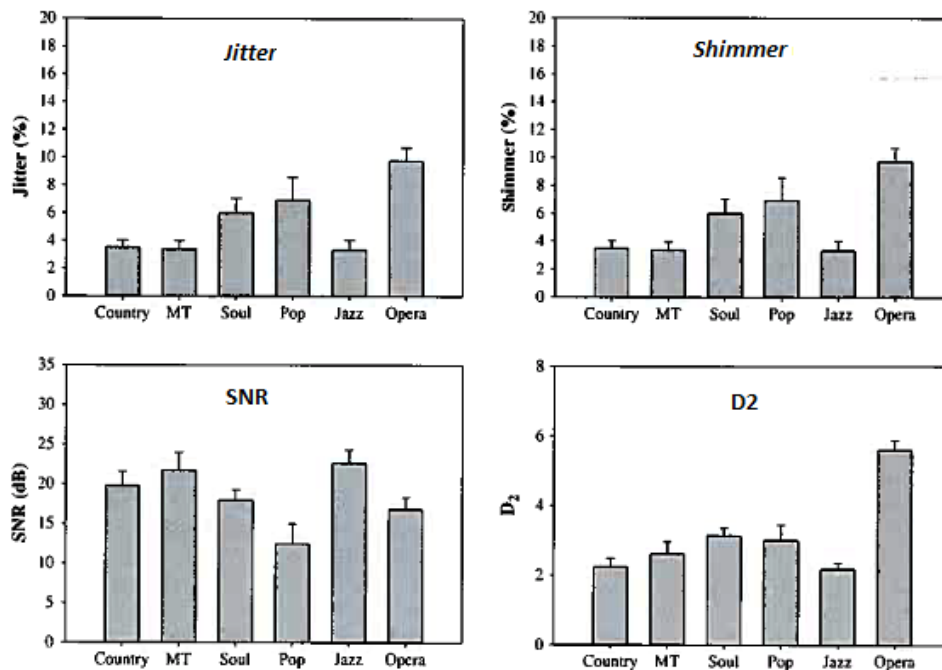


Figura 2.6: Jitter, Shimmer, SNR e D2 para seis estilos de música [19]

No caso da percepção na voz cantada, para além de em algumas das definições de parâmetros perceptivos já normalizados não haver correlação direta entre esses parâmetros e as características acústicas mais conhecidas, temos também o problema da avaliação perceptiva de uma voz depender fortemente do avaliador que a escuta e que tem de julgar se o som escutado é similar ou não com um outro som de referência [17].

Do ponto de vista da voz cantada, como para as vozes patológicas, a avaliação perceptiva é subjetiva e pouco precisa. Pode-se ter a avaliação de uma voz baseada em parâmetros já pré-definidos ou uma descrição pessoal da voz.

A nível perceptivo, existem vários estudos referentes a estilos musicais, enquanto que os de classificação do tipo de voz, como é o objetivo deste projeto, não abundam. Apesar desta discrepância de objetivos, algumas características destes estudos são interessantes do ponto de vista da classificação do tipo de voz, visto que alguns estilos de canto são diretamente relacionados a certos tipos de voz e as características acústicas estudadas são próximas das que vão ser estudadas neste projeto.

No estudo de Caitlin J. Butte intitulado *Perturbation and Nonlinear Dynamic Analysis of Different Singing Styles*, utilizando a análise de perturbações e de dinâmica não linear, a autora estuda a correlação entre características acústicas e diversos estilos musicais [19]. Pode-se verificar que em estilos musicais diferentes verificam-se características acústicas também distintas como mostra a Figura 2.6.

A nível da avaliação e descrição de uma voz cantada, a terminologia para descrever esse som é muito variada e, deste exercício de avaliação podem aparecer termos como lírica, dramática, soprosa, clara, pesada, nasal, trémula, focada, entre outros [20].

Do ponto de vista da correlação das classificações do tipo de vozes com características acústicas, normalmente encontramos associações com o *pitch*, frequência fundamental e com o formante do cantor, no caso de características perceptivas de ressonância [16]. Para algumas definições também se encontra alguma correlação com o vibrato, mas como já mencionado, normalmente o vibrato não é associado por si só a uma característica perceptiva, visto este derivar do *pitch* e da sua variação.

A nível de avaliação de vozes estão estabelecidos parâmetros de avaliação quanto à extensão vocal, que se refere ao grupo de notas que um cantor consegue produzir, mesmo com baixa qualidade tímbrica; estas avaliações são já bastante estudadas no que toca à frequência fundamental do som.

Algumas definições foram tomadas para certos termos de avaliação vocal. Por exemplo a caracterização de uma voz "brilhante" ou "escura" demonstra a complexidade da voz, pois são muito poucas as vozes que são exclusivamente "claras" ou "escuras", contendo normalmente elementos das duas classificações [20]. Uma voz brilhante está relacionada com o brilho e o poder, a energia de execução, normalmente tendo um *pitch* elevado, enquanto uma escura tem um baixo *pitch* estando associada a calor e plenitude. Sendo que este parâmetro de classificação está associado à distribuição espectral de energia.

Também existem descrições de sons consoante a percepção da sua origem; os "posteriores", que dão a sensação de serem produzidos na laringe ou na raiz da língua, e que estão normalmente associados a tons "escuras"; e os "frontais" associados a sons mais brilhantes, e que dão a sensação de serem produzidos na caixa de ressonância bocal [20].

Podemos também definir as vozes consoante a quantidade de ar que é expelido na sua produção, temos então o caso de uma voz ser mais soprosa, sendo expelido maior quantidade de ar aquando esta é produzida sendo acompanhada de um género de ruído, ou o caso de uma voz mais clara, livre, quando necessita de menor quantidade de ar a ser expelido para ser formada [20]. Existe uma descrição do tipo de voz que também coloca num extremo as vozes claras e limpas, com baixo ruído, e num outro o tipo de voz "rouca", que se diferencia do tipo de voz soprosa no tipo de ruído por que é acompanhada, sendo neste caso um ruído não branco. Normalmente, este tipo está associada a patologia vocal, mas alguns cantores incluem deliberadamente este efeito [20].

Uma das características perceptivas mais estudadas em relação à voz cantada, em particular em performances clássicas, é a característica de *Ring*. Acusticamente, esta é uma amplificação nas altas frequências [20], estando normalmente associada ao formante de cantor, que desta forma consegue que a sua voz sobressaia em relação a uma orquestra. Esta característica está ilustrada na Figura 2.5.

Outra descrição que se pode fazer de uma voz é a existência ou não de timbre nasal. Esta característica acontece quando a comunicação entre a cavidade bocal e a cavidade nasal não se encontra totalmente fechada, ou até se encontra totalmente aberta, fazendo com que haja uma ressonância extra. Alguns especialistas em música dizem que ressonância nasal não deveria existir em voz cantada, enquanto outros referem que esta é importante para a correta utilização da voz

cantada [21]. Perceptivamente, a existência de mais ou menos ressonância nasal acaba por ser uma questão subjetiva, estética [21] ou cultural.

De notar também que, embora as definições sejam feitas em termos de extremos, poucas são as vozes que realmente se encontram nos extremos destas definições, a maioria encontra-se em patamares intermédios das várias definições.

Ao longo do tempo, tempo este numa perspetiva longa, é perceptível a evolução das vozes, variando as suas características, consoante o treino o desgaste ou o envelhecimento do sistema fonatório dos cantores; como por exemplo, para quem acompanhou a carreira de Frank Sinatra notou um aumento de rouquidão na sua voz [20].

2.5 Parâmetros qualitativos e perceptivos

Como já referido, no decorrer deste projeto tive o auxílio de um cantor da ESMAE, João Ferreira, para a caracterização da voz em termos de parâmetros qualitativos e perceptivos da voz cantada, sendo importante o ponto de vista de um cantor profissional e de grande relevância e acréscimo o estudo de várias definições de parâmetros perceptivos da voz levado a cabo por ele. Com autorização do João Ferreira é são aqui descritos os resultados do seu estudo.

2.5.1 Afinação

Afinação é a capacidade de produção de um som, do ponto de vista de frequência fundamental, igual a outro. É de realçar que no conceito de afinação é importante a referência à escala utilizada; sendo a mais comumente usada a escala igualmente temperada, sendo que nesta cada oitava é igualmente dividida em doze semitons. A relação entre a frequência de uma nota e um semitom acima é de $\sqrt[12]{2}$ [22].

2.5.2 Tessitura

Designa-se tessitura ao conjunto de notas que um cantor consegue produzir sem esforço mantendo todas as suas qualidades tímbricas. Pode ser representativa do tipo de voz consoante os limites graves e agudos para cada tipo de voz. Como pode ser verificado na Figura 2.7 para o caso de voz masculina e na Figura 2.8 no caso de voz feminina, encontram-se representados nestas figuras os limites em termos de notas e das respetivas frequências para os vários tipos de voz, sendo contudo que esta representação não é absolutamente rígida. Por exemplo, um Tenor Lírico com características especiais pode aspirar a alguns papéis de Tenor Ligeiro, ou até mesmo a Tenor Spinto [22].

2.5.3 Timbre

O timbre é a característica que permite distinguir sons da mesma frequência emitidos por fontes diferentes. Na voz cantada, o timbre está diretamente relacionado com a fisionomia do cantor, tanto interna como externa [22].

Si 1	Dó 2	Ré 2	Mi 2	Fá 2	Sol 2	Lá 2	Si 2	Dó 3	Ré 3	Mi 3	Fá 3	Sol 3	Lá 3	Si 3	Dó 4	Ré 4	Mi 4	Fá 4	Sol 4	Lá 4	Si 4	Dó 5	Ré 5	Mi 5
61,74	65,41	73,42	82,41	87,31	98,00	110,00	123,47	130,82	146,83	164,82	174,62	196,00	220,00	246,94	261,63	293,66	329,63	349,23	392,00	440,00	493,88	523,26	587,32	659,26
Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz
Baixo Profundo – Ex. Ivan Rebroff																								
		Baixo – Ex. Kurt Moll																						
				Baritono – Ex. Dietrich Fischer-Dieskau																				
							Tenor Dramático/Wagneriano – Ex. James King																	
							Tenor Spinto – Ben Heppner																	
							Tenor Lírico – Ex. Fritz Wunderlich																	
							Tenor Ligeiro – Ex. Luigi Alva																	

Figura 2.7: Catalogação de tipos de voz masculinas consoante a tessitura (*Fach*) [22]

Mi 3	Fá 3	Sol 3	Lá 3	Si 3	Dó 4	Ré 4	Mi 4	Fá 4	Sol 4	Lá 4	Si 4	Dó 5	Ré 5	Mi 5	Fá 5	Sol 5	Lá 5	Si 5	Dó 6	Ré 6	Mi 6	Fá 6
164,82	174,62	196,00	220,00	246,94	261,63	293,66	329,63	349,23	392,00	440,00	493,88	523,26	587,32	659,26	698,46	784,00	880,00	987,76	1046,52	1174,64	1318,52	1396,92
Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz	Hz
Contralto/Contratenor (Masc.) – Ex. Marian Anderson (Contralto); Andreas Scholl (Contratenor)																						
		Mezzo-Soprano – Ex. Christa Ludwig																				
		Soprano Dramático – Ex. Birgit Nilson																				
					Soprano Spinto – Ex. Leontyne Price																	
					Soprano Lírico – Montserrat Caballé																	
					Soprano Ligeiro (Soubrette) – Kathleen Battle																	
					Soprano Coloratura – Ex. Editha Gruberova																	

Figura 2.8: Catalogação de tipos de voz femininas consoante a tessitura (*Fach*) [22]

Segundo Fant [23] e o seu modelo fonte-filtro, a fonação é dividida em três partes: fonte sonora, o filtro e a radiação, a vibração das pregas vocais (fonte) produzindo o som laríngeo que é filtrado pelo trato vocal (filtro) e projetado (radiação). Isto faz com que a fisionomia do cantor seja um ponto preponderante do seu timbre [22].

O timbre é também outra característica de catalogação de voz entre os vários tipos de voz masculina e feminina, por exemplo, diferenças tímbricas entre Soprano e Contralto; servindo também para diferenciar entre subclasses do mesmo tipo de voz, por exemplo, Tenor Lírico, Ligeiro, Spinto ou Dramático [22].

O termo "voz timbrada" é associado a vários aspetos que podem existir ou não na mesma voz podendo ser definidos por pares de termos antagónicos. Temos, então, os pares clara/escuro, voz na frente/voz recuada, leve/pesada, limpidez/soprosidade e limpidez/aspereza. O facto de haver vozes que encaixam melhor numa classificação não impede que possuam características de outra, por exemplo uma voz pode ser recuada e ao mesmo tempo ter limpidez [22].

2.5.3.1 Claro/Escuro

O termo vem do italiano *chiaroscuro*, expressão utilizada para descrever a técnica de pintura de Leonardo da Vinci. Em termos de características acústicas uma voz "clara" possui um reforço nas frequências agudas enquanto que uma voz "escuro" possui um reforço nas graves. Assim uma voz "clara" possui brilho e energia sendo que a "escuro" transmite uma sensação de calor,

sendo redonda e cheia. Na Tabela 2.1 são representados exemplos de cantores classificados desta forma [22].

Tabela 2.1: Exemplos de cantores Claro/Escuro [22]

Classificação	Cantor Exemplo
Masculina clara	Luigi Alva
Masculina escura	Jonas Kaufmann
Feminina clara	Lucia Popp
Feminina escura	Jessye Norman

2.5.3.2 Voz na frente/Voz recuada

Uma voz "na frente", sensação que os cantores descrevem como "voz de máscara", possui mais brilho do que uma voz mais recuada. Estes parâmetro possui uma grande relação com o anterior, pois uma voz mais frontal dá origem a um timbre mais claro [22].

Uma voz demasiado "recuada" dá a sensação de a faringe estar estrangulada explorando pouco os seios nasais, dando a sensação de a voz estar recuada e difusa [22].

Esta técnica é muitas vezes confundida com a voz nasalada que tem menos projeção em comparação com a voz na frente ou focada na máscara que é colocada num só ponto tendo assim maior capacidade de ser ouvida por cima de uma orquestra [22].

2.5.3.3 Voz leve/Voz pesada

Uma voz "pesada" é pouco maleável, sendo pouco propícia a flutuações de dinâmica. É resultado da não elevação do palato mole tornando-a menos rica em harmónicos, fazendo com que o dispêndio de energia seja grande, causando cansaço e dificultando a performance. Uma voz "leve", pelo contrário, é mais brilhante e ressoante devido à riqueza em harmónicos [22].

Vozes mais "pesadas", geralmente, produzem mais som do que vozes mais "leves". Isto faz com que o tipo de repertório esteja intrinsecamente ligado ao tipo de voz de um determinado cantor. Faz mais sentido, portanto, falar em repertório "pesado" e "leve" ao invés de uma voz "leve" ou "pesada", visto se determinar o tipo de repertório adequado ao cantor e não o contrário. Na Tabela 2.2 são representados exemplos de cantores classificados desta forma [22].

Tabela 2.2: Exemplos de cantores Voz leve/Voz pesada [22]

Classificação	Cantor Exemplo
Voz pesada Masculina	James King
Voz leve Masculina	Luigi Alva
Voz pesada Feminina	Birgit Nilsson
Voz leve Feminina	Cecilia Bartoli

2.5.3.4 Limpidez/Soprosidade/Aspereza

A soprosidade na voz resulta de uma fenda glotal que quando não é associada a fenómenos patológicos. Deve-se, por exemplo, a deficiência de suporte respiratório. Quando encontrada num cantor lírico é considerado um défice de técnica. Contudo pode ser encontrada noutros estilos como o *jazz* ou o *Folk* [22].

A aspereza é definida pela quantidade de ruído na voz, podendo ser causada por fenómenos fisiológicos ou patológicos; pode também ser introduzida propositadamente, consoante a interpretação do tema [22].

2.5.4 Falsete

O falsete caracteriza-se pela produção não natural de tons de elevada frequência fundamental através da vibração parcial das pregas vocais. O resultado é um registo mais leve e suave contrastando com o registo de "peito" [22].

Apresenta um dispêndio superior de energia, visto que só parte das pregas vocais estão ativas na produção sonora, sendo a amplitude sonora baixa [22].

Acusticamente, a voz de falsete é pouco timbrada tendo carência de harmónicos, principalmente graves devido à falta de vigor na utilização das ressonâncias do trato vocal [22].

Contém alguma soprosidade, devido à existência de uma fenda glótica, pela utilização parcial das pregas vocais [22].

2.5.5 Ataque

Um bom ataque é determinante para uma correta emissão vocal. Pode ser definido como o posicionamento de todos os constituintes do trato vocal no momento em que se inicia a produção da nota desejada. A qualidade do ataque relaciona-se diretamente com a qualidade da articulação do texto, mais precisamente das consoantes [22].

2.5.6 Vibrato

Como já referido, o vibrato é uma quase periódica variação da frequência fundamental, podendo este ser combinado com variações de intensidade, enriquecendo o som produzido e o próprio timbre. Sendo também um parâmetro de qualidade está naturalmente sujeito às condicionantes estéticas da obra a interpretar. Enquanto que numa obra renascentista é utilizado apenas como ornamento, numa obra romântica ou contemporânea é um atributo essencial em termos de expressividade [22].

2.5.7 Legato

Legato é caracterizado pela continuidade da linha vocal sem percepção de hiatos, quer na mudança de notas mais graves para mais agudas quer no contrário. A transição entre notas é feita de

forma contínua e progressiva, exigindo muita técnica de forma a manter as características do som fundamental [22].

2.5.8 Staccato

Staccato deriva do italiano *staccato* que significa destacado, separado. É, em oposição ao *Legato*, a introdução de pausas entre notas [22].

2.5.9 Micro-dinâmicas e destreza vocal

Além das indicações na partitura, estão atribuídas dinâmicas inerentes à própria estética da composição, quer pela prosódia do texto, pela construção musical ou pelas características estilísticas do próprio compositor. Sendo que, conforme a peça, diferentes micro-dinâmicas podem ser utilizadas, sendo muitas vezes relacionadas com a interpretação do executante [22].

Assim, manifestações musicais como crescendos, diferentes formas de ataque e de articulação, mudanças de timbre e de intensidade vocal e, até mesmo características que no geral são consideradas indesejadas no canto como a soproidade ou a aspereza, podem ser empregues na interpretação de peças que assim o exijam [22].

A realização das micro-dinâmicas requer sensibilidade e destreza vocal, de forma a não ser prejudicado o som emitido, bem como a compreensão do texto e do seu significado, de forma à boa execução da peça [22].

2.5.9.1 Coloratura

A coloratura é 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. Pode ser efetuado tanto em *Legato* como em *Staccato*. Sendo a dificuldade a manutenção da estrutura do trato vocal que é feita à custa da sustentação pelo diafragma durante a execução [22].

2.5.9.2 Melisma

Trata-se da realização de uma sílaba em várias notas sucessivas. É um termo muito utilizado na música renascentista, nas passagens do canto Gregoriano [22].

2.5.9.3 Portamento

Portamento em italiano significa transporte. É a ligação entre duas notas com efeito deslizante. A indicação da sua utilização pode estar na partitura da peça, sendo que quando é feita sem indicação é considerada, muitas vezes, défice de técnica [22].

2.6 Aplicações

2.6.1 Estimação e mapeamento de formantes

São muitas as utilidades que se podem dar à estimação da frequência e das trajetórias de formantes, tais como o desenvolvimento de codificadores, de aplicações de reconhecimento de discurso e/ou orador [24], aplicações de modificação de características da voz, aplicações de síntese de voz (*text-to-speech*) [25] ou aplicações de *feedback* visual [26].

Embora a frequência aproximada dos formantes no espectro geralmente seja de fácil detecção ao olho humano, o desenvolvimento de um sistema automático para este efeito está longe de ser fácil [25]; assim, vários métodos de detecção de formantes têm sido propostos [24], sendo alguns baseados em detecção de picos espectrais e outros em extração de raízes [27].

Os métodos baseados em extração das raízes requerem grande esforço computacional ao tentarem encontrar as raízes do polinómio obtido, por exemplo, por predição linear (LPC) [27] sendo que para aplicações em tempo real não são a melhor opção, devido à complexidade computacional requerida. A maior vantagem desta abordagem é a sua precisão [28], embora esteja condicionada pelas limitações do método de obtenção dos polinómios utilizados para o efeito.

Os métodos baseados em detecção de picos espectrais têm sido os mais utilizados para estimação de formantes [27]; embora haja algumas abordagens feitas na detecção de picos diretamente do espectro do sinal, tradicionalmente emprega-se a detecção de picos a uma suavização da envolvente espectral, obtida através de, por exemplo, predição linear (LPC) muito utilizada na análise de voz falada, ou o *cepstrum* real do sinal, utilizado em aplicações onde os autores abordam a ideia de separação da componente harmónica do sinal da sua envolvente espectral. O princípio baseia-se no facto de os coeficientes (*quefrecies*) mais baixos do *cepstrum* serem mais afetados pela envolvente espectral do que os mais altos, sendo também menos afetados pela componente harmónica [24].

Para a detecção dos máximos da envolvente espectral têm sido aplicados vários métodos, desde a simples detecção direta dos picos no logaritmo do valor absoluto da envolvente espectral ou a detecção dos mínimos da segunda derivada do mesmo logaritmo ou, ainda, a detecção de picos na primeira derivada da fase da parte complexa do espectro [25] [28]. A detecção de picos no logaritmo do valor absoluto da envolvente espectral falha na detecção de formantes que, estando próximos de outros mais fortes, não causam a existência de um pico bem definido na envolvente espectral, ao contrário da abordagem de detecção dos mínimos da segunda derivada do mesmo logaritmo que obtém melhores resultados nessa situação, definindo bem dois picos que estejam muito próximos [26].

Para voz falada com baixa frequência fundamental, no geral o método de detecção de formantes pelos máximos da envolvente espectral referidos tem bons resultados na detecção dos primeiros dois ou três formantes, porque os harmónicos estão próximos o suficiente para o seu efeito não ser um fator limitador [24]. Com o aumento da frequência fundamental, como acontece no caso da voz cantada, os métodos por detecção de picos, devido ao aumento da distância dos harmónicos,

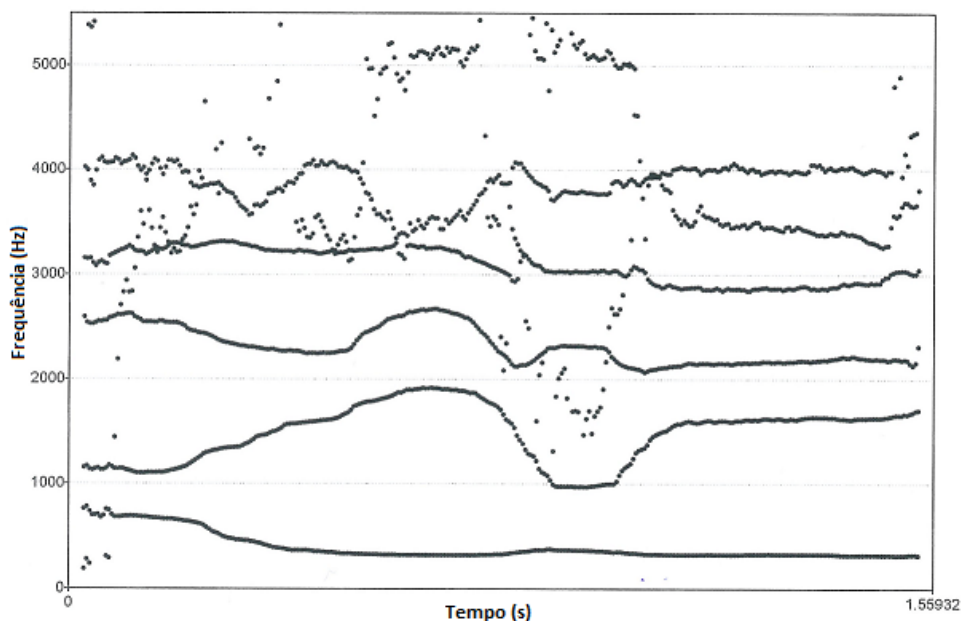


Figura 2.9: Erros típicos de estimação de formantes [29] (*adaptada*)

vão perdendo eficiência detetando as localizações dos harmônicos e confundindo aqueles com estas [24].

O mapeamento dos formantes é também importante em áreas como o estudo de variações das características do trato vocal, como por exemplo, as transições vogal-consoante [30] ou em aplicações de *feedback* visual, entre outras; sendo um último nível de validação de candidatos a formantes detetados pelos métodos já descritos, que geralmente contêm alguns erros, como é mostrado na Figura 2.9. Estes podem contudo ser minimizados com um algoritmo robusto de mapeamento temporal [29].

Assim, vários métodos podem ser utilizados para encontrar a trajetória dos formantes, desde aqueles que usam algum tipo de interpolação e/ou suavização da localização dos candidatos encontrados [29], que para aplicações em tempo real não são apropriados, tendo também em conta que ao efetuar estas operações poderemos perder informação relevante para o estudo em causa. Outros métodos como a simples escolha dos candidatos mais próximos aos encontrados na janela anterior [26] ou à média de um número determinado de janelas anteriores são mais apropriados a aplicações em tempo real, mas têm o problema da possível falha de deteção de um formante, o que compromete este, podendo ser atribuídos candidatos a formantes aos quais não pertencem. Uma alternativa de forma a tentar evitar estes erros é a utilização de mapeamento baseado em regras como a limitação de máximo deslocamento de um formante de uma janela para outra e utilizando informação extra dos candidatos e dos formantes, como a amplitude ou a banda para regras mais elaboradas [29]. Estes métodos têm a dificuldade de mapeamento no caso de os formantes estarem muito próximos e podem causar alguns erros ao não considerar erradamente alguns candidatos que fujam às regras estabelecidas, para além de serem de difícil implementação. Outros métodos implementados para o mapeamento de formantes são os baseados em modelos escondidos de Markov

(HMM), onde o resultado é definido pelas combinações de probabilidades de conjuntos sucessivos de janelas [29].

Geralmente este mapeamento é feito sobreposto ao espectrograma, de forma a relacionar estas duas visualizações como é mostrado na Figura 2.10.

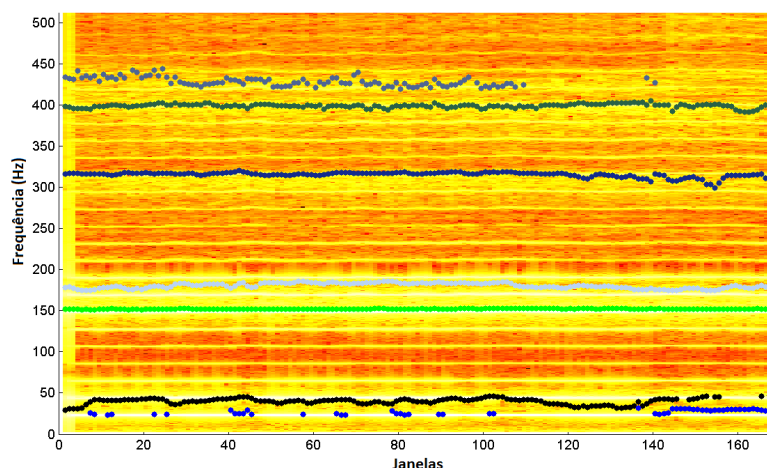


Figura 2.10: Espectrograma com o *tracking* dos formantes encontrados (Vogal 'o' de um Contralto), de notar que o padrão de cores está inverso ao habitual.

2.6.2 Detecção e caracterização de vibrato

Como já referido o vibrato é uma importante característica da voz cantada, sendo caracterizado por uma variação de frequência fundamental de forma quase periódica.

Embora as suas características sejam bastante estudadas, a sua deteção e um *feedback* visual é um tema muito pouco abordado, tendo sido desenvolvido por José Ventura na sua dissertação intitulada "*Biofeedback da voz cantada*" [31] também no âmbito do mesmo projeto FCT no qual esta dissertação se enquadra, um algoritmo de estimação de vibrato que o deteta bem como também estima a sua frequência (em Hertz) e a extensão (em semi-tons) de forma a ser visualmente mostrado numa aplicação interativa.

2.7 Conclusões

Do levantamento de atributos artísticos e estéticos feito tanto por mim como pelo cantor da ESMAE, João Ferreira com o qual trabalhei, denota-se a maior parte dos atributos de avaliação e catalogação de voz cantada, tanto do ponto de vista técnico como estético, prendem-se com o efeito do trato vocal e suas ressonâncias sobre o sinal produzido pelas pregas vocais.

Temos, então, a catalogação dos tipos de voz associados a timbres característicos e não só à capacidade de produção de uma determinada nota.

Foi visto o efeito que modificações da dinâmica do trato vocal têm tanto na interpretação de peças como na correta utilização da voz, bem como o efeito que a distribuição de energia

pelo espectro tem do ponto de vista da caracterização estética da voz cantada e a importância da elevação dos harmónicos de mais alta frequência e reposicionamento de formantes, para uma maior perceptibilidade. Todas estas modelações do sinal glótico são da responsabilidade do trato vocal.

A caracterização do vibrato, não será tratada neste projeto, visto esta tarefa já ter sido realizada, por José Ventura, no âmbito do mesmo projeto FCT, no qual esta dissertação se enquadra.

Capítulo 3

Estimação e mapeamento de formantes

3.1 Introdução

Neste capítulo, depois de uma breve descrição dos métodos utilizados, é descrito o sistema de estimação da frequência dos formantes e o seu mapeamento desenvolvido no decorrer desta dissertação.

Os algoritmos desenvolvidos foram desenvolvidos em *MatLab* por esta ser uma ferramenta boa no processamento e análise de sinal, e por ser fácil a visualização de grafismos dos resultados do funcionamento do algoritmo. Foram desenvolvidos de forma a uma fácil aplicação em ambiente C/C++.

3.2 Modelo fonte-filtro

A mais comum abordagem de um modelo da produção de voz tem sido o modelo fonte-filtro [7] ilustrado na Figura 3.1. Basicamente, este modelo inclui uma fonte de sinal que é modelada por um filtro para a sua forma final [23].

Assim, a fonte é o som laríngeo (fonte glótica) produzido pela vibração das pregas vocais no caso de voz vozeada e ruído resultante nomeadamente de turbulência, no caso de voz não vozeada [7] [33]. O filtro é o trato vocal que vai modelar a fonte quando o sinal desta passa pelas várias cavidades e formas que o constituem, sendo assim definida a qualidade e o tipo de sonoridade a ser percebida [34]. Segundo Fant, o trato vocal é responsável pela maior parte da informação de um sinal de voz [23].

O trato vocal é modelado por um filtro caracterizado por uma função de transferência $H(z)$, assumindo-se tipicamente na literatura que se trata de um filtro *all-pole*.

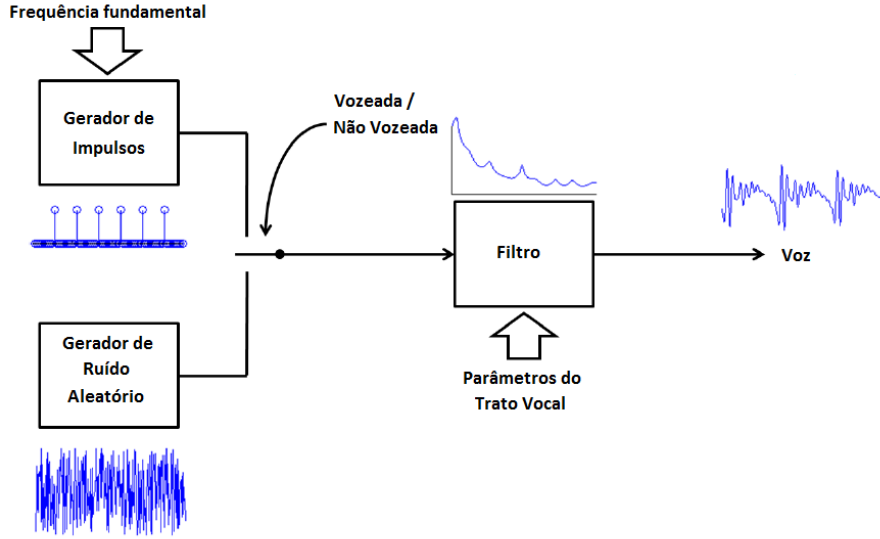


Figura 3.1: Modelo fonte-filtro [32] (*adaptada*)

3.3 Estimação de trato vocal

Como referido, usualmente o trato vocal é modelado por um filtro *all-pole*. Assim, surge a necessidade de estimação dos seus coeficientes de forma a poder ser analisado. Neste projeto foram utilizados dois métodos reconhecidos de análise de sinal, um baseado em predição linear e um outro baseado no *cepstrum*.

3.3.1 Linear Prediction Coding

Em análise de sinal de voz é muito utilizada a predição linear onde se assume que o sinal de voz é a combinação linear dos seu valores passados e do valor atual [35] [33]. Muitos sinais, como a voz, são parte preditivos, parte aleatórios; estes sinais podem ser modelados pela saída de um sinal não relacionado aplicado a um determinado filtro [35]. Assim, segundo o modelo fonte-filtro, o trato vocal pode ser então modelado por predição linear.

A expressão 3.1 representa a saída atual. O valor da saída de índice m é obtido usando uma combinação linear dos P valores passados onde m é o índice do tempo discreto, $\hat{x}(m)$ a predição de $x(m)$ e a_k os coeficientes de predição [35].

$$\hat{x}(m) = \sum_{k=1}^P a_k x(m-k) \quad (3.1)$$

O erro de predição é dado pela expressão 3.2, que é a diferença entre o sinal $x(m)$ e o sinal estimado $\hat{x}(m)$.

$$e(m) = x(m) - \sum_{k=1}^P a_k x(m-k) \quad (3.2)$$

A ordem P de uma predição linear para modelação do trato vocal é importante para as características do modelo do trato vocal a analisar, se for muito baixa, ressonâncias importantes podem ser descartadas, se for muito alta será modelada a fonte do sinal e não as características do trato vocal [32].

Na Figura 3.2 está demonstrada a envolvente espectral obtida por um método baseado em predição linear.

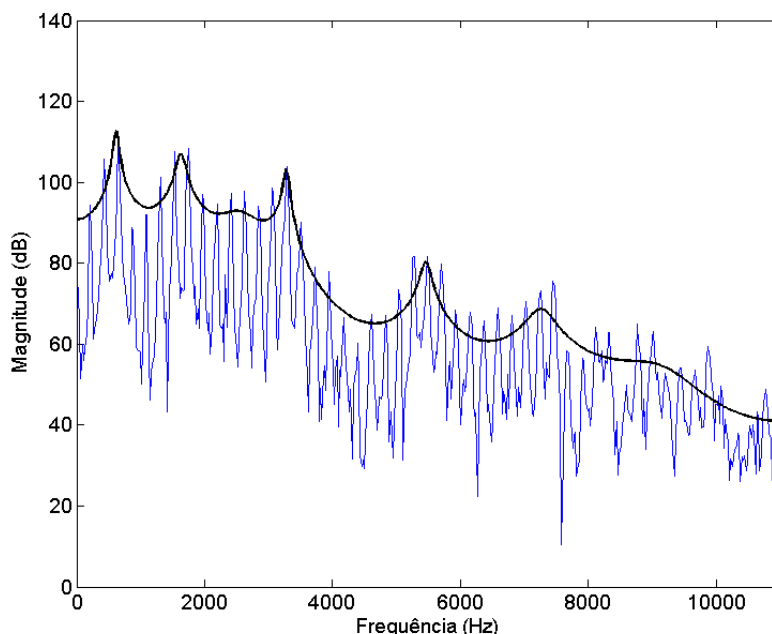


Figura 3.2: Exemplo de envolvente obtida com modelo baseado em LPC (preto) do espectro de um sinal (azul) para modelação do trato vocal

3.3.2 Cepstrum

O termo *cepstrum* deriva da inversão da primeira sílaba do termo *spectrum* dado que é obtido pela transformada inversa de *Fourier* do espectro logaritmo do sinal. Foi desenvolvido por Bogert *et al* e definido o termo *quefrency* para representar a variável independente n do valor do *cepstrum* $c(n)$ [36].

O *cepstrum* é uma transformação homomórfica que permite a separação entre a fonte e o filtro de um sinal, convertendo uma convolução numa soma [36], expressões 3.3 e 3.4.

$$x(n) = s(n) * h(n) \quad (3.3)$$

$$\hat{x}(n) = \hat{s}(n) + \hat{h}(n) \quad (3.4)$$

Nestas equações $h(n)$ representa a característica do filtro, $s(n)$ representa o sinal de entrada.

Pode-se então recuperar o sinal $s(n)$ e $h(n)$ do sinal $\hat{x}(n)$, pois assume-se que é possível encontrar um valor N de forma a que $\hat{h}(n) \approx 0$ para $n \geq N$ e $\hat{s}(n) \approx 0$ para $n < N$ [36].

O *cepstrum* complexo é definido pela expressão 3.5, sendo o real definido pela expressão 3.6.

$$\hat{x}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln X(\omega) e^{j\omega n} d\omega \quad (3.5)$$

$$c(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln |X(\omega)| e^{j\omega n} d\omega \quad (3.6)$$

Se o sinal a analisar tem informação fase mínima o *cepstrum* pode então ser determinado só pelo *cepstrum* real [36]. Isto acontece no caso da análise de voz, onde a informação cuja natureza é de fase mínima não tem tanta importância como a de magnitude. À utilização do *cepstrum* real acresce-se a vantagem do menor peso computacional, em comparação ao *cepstrum* complexo, tornando mais fácil a análise [37].

Como já referido, o princípio de utilização do *cepstrum* prende-se com a separação entre a fonte de sinal e do filtro. No caso da análise de voz traduz-se na separação do trato vocal e do sinal glotal. Assim, os primeiros N valores do *cepstrum* são representativos do trato vocal, especificando a envolvente espectral, enquanto que os restantes são representativos da fonte glótica [37].

3.4 Estimação da frequência dos formantes

Como já mencionado, para deteção e estimação de formantes têm sido utilizados maioritariamente duas abordagens, a deteção de picos espectrais e a extração de raízes do filtro que modela o trato vocal.

No caso deste projeto foi escolhida a abordagem baseada em deteção de picos espectrais, que sendo menos pesada computacionalmente é mais adaptada a análise em tempo real o que é uma das características pretendidas. Foi também decidido que o número de formantes a ser estimado seria de oito, de forma a poder serem localizados possíveis formantes nas regiões de mais alta frequência, tendo sido desenvolvido este algoritmo baseado nesse princípio.

O sinal é analisado em janelas (obtidas pela raiz quadrada da janela de Hanning) de 1024 amostras às quais é aplicada a transformada ODFT sendo calculadas a sua densidade espectral de potência, energia e fase. Do ponto de vista do avanço temporal de análise, a janela a ser processada terá uma sobreposição de 75% da anterior, tendo a análise um deslocamento de 256 amostras por ciclo de processamento.

A ideia geral do sistema desenvolvido é a utilização de dois métodos para extração das características do trato vocal e estimados os seus formantes. Estes dois métodos, um baseado em LPC e outro em *cepstrum*, são aplicados tanto sobre o espectro do sinal como sobre o espectro da estimação de ruído presente na voz, de forma a tirar partido do fato do ruído da voz também ser modelado pelo trato vocal. Esta estimação do ruído é baseada no princípio da subtração dos

harmónicos de forma a extrair o som laríngeo (fonte glótica) do sinal, deixando só a componente não vozeada do sinal, o ruído da voz.

Desta forma para a sua estimação é necessário obtermos a localização dos harmónicos no espectro sendo então necessário o recurso a um algoritmo de estimação de componente harmónica. Assim é usado o algoritmo *Searchtonal*, que é um detetor de *pitch* (frequência fundamental) e da respetiva componente harmónica; que foi desenvolvido pelo professor Doutor Aníbal Ferreira.

Este detetor de *pitch* implementa análise *cepstral* e um banco de regras heurísticas de seleção dos candidatos à frequência fundamental e componentes harmónicas respetivas. Tem vindo a ser melhorado ao longo do tempo de forma a estimar de uma forma robusta e precisa a frequência fundamental e os seus respetivos harmónicos [31].

Com as localizações das componentes harmónicas no sinal, obtidas pelo detetor de *pitch*; bem como as suas características de fase e energia obtidas a partir da sua ODFT; vai ser sintetizada a estrutura harmónica para ser subtraída à ODFT do sinal. De seguida é calculada a energia deste resultado a qual passará por um sistema suavização de três níveis de forma a ser reduzido o efeito do processamento efetuado.

É possível ver estes passos na Figura 3.4 e o diagrama de blocos deste primeiro nível de processamento na Figura 3.3.

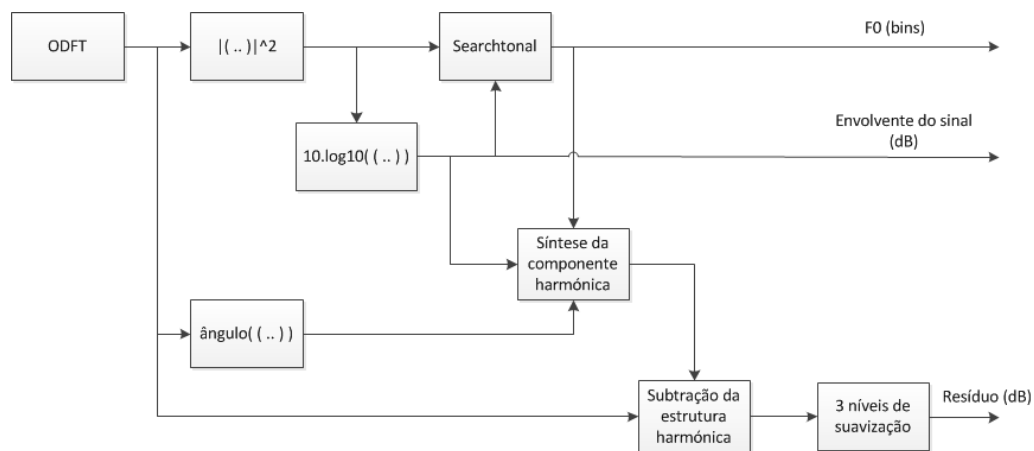
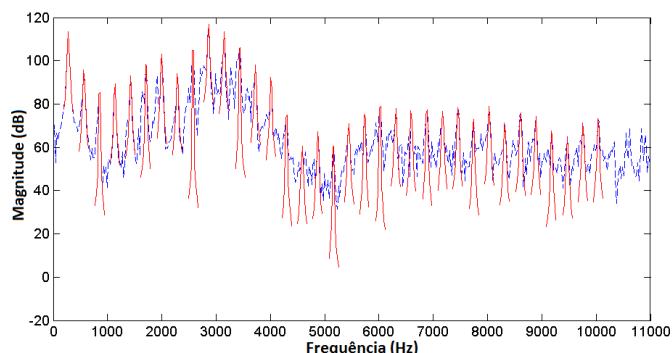


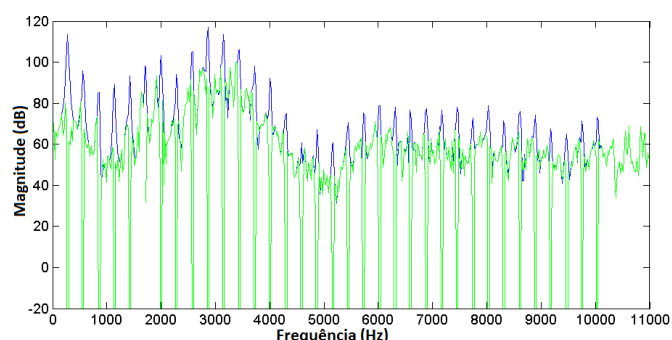
Figura 3.3: Primeiro nível de processamento.

São aplicados tanto sobre o espectro de energia do sinal, como do resíduo, os métodos já referidos, um baseado em LPC e outro em *cepstrum*, de forma a serem obtidas as respetivas envolventes espectrais para, como já mencionado, serem detetados os possíveis formantes por análise espectral. Estas envolventes estão exemplificadas na Figura 3.5.

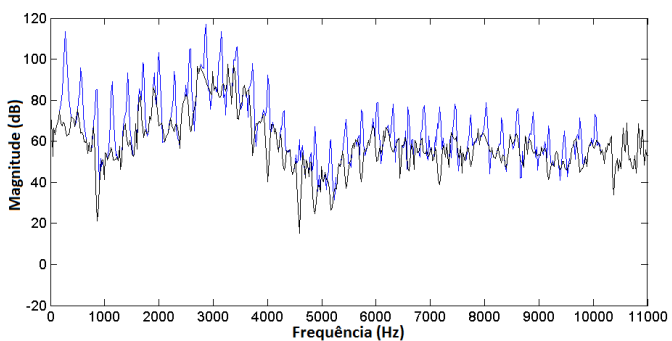
No caso da envolvente obtida por LPC, tanto do sinal como do resíduo, a ordem do polinómio LPC foi definida como sendo de 16, duas vezes o número de formantes que se pretende estimar. No caso da envolvente do sinal obtida por *cepstrum* o número de coeficientes *quefrequency* a utilizar vai depender da frequência fundamental de forma a só ser utilizados os coeficientes que representam o trato vocal. E no caso da envolvente do resíduo pelo *cepstrum*, e visto este teoricamente não



(a) O espectro dos harmónicos sintetizados a vermelho e do sinal original a azul



(b) Representam-se a azul o sinal original e a verde o resultante da extração dos harmónicos (i.e. o resíduo)

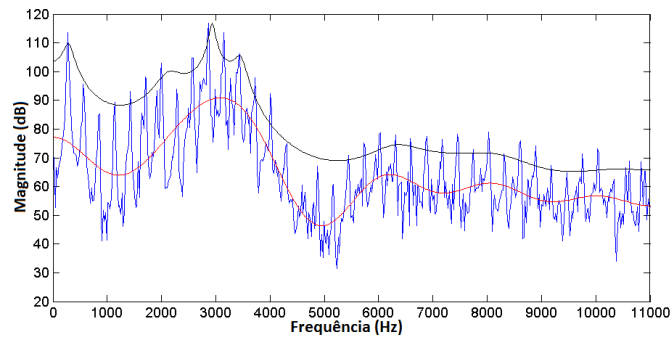


(c) Representam-se a azul o sinal original e a preto o resultante da suavização do resíduo

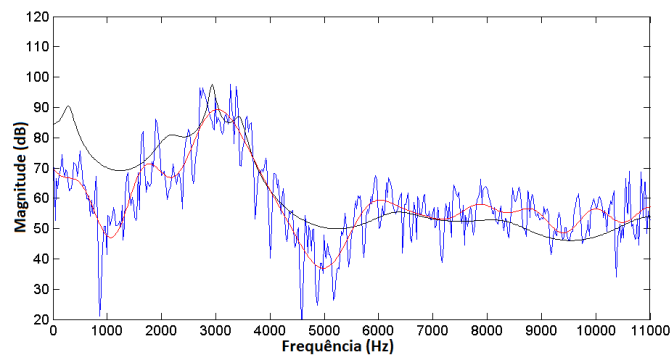
Figura 3.4: Processo de estimação do resíduo

conter componente harmónica, o número de coeficientes *quefrequency* foi definido como sendo de 35.

Numa primeira abordagem foi utilizado um detetor de picos para a localização dos picos espectrais que representam formantes, mas, o que se verificou é que com uma localização só de picos espectrais não se obtém a deteção de muitos formantes, isto por estes não terem a capacidade de se evidenciarem, tanto por não terem muita energia, ou por serem desvalorizados pela presença de



(a) O espectro do sinal a azul, as suas envolventes LPC a preto e a *cepstral* a vermelho



(b) O espectro do resíduo a azul, as suas envolventes LPC a preto e a *cepstral* a vermelho

Figura 3.5: As envolventes espectrais calculadas

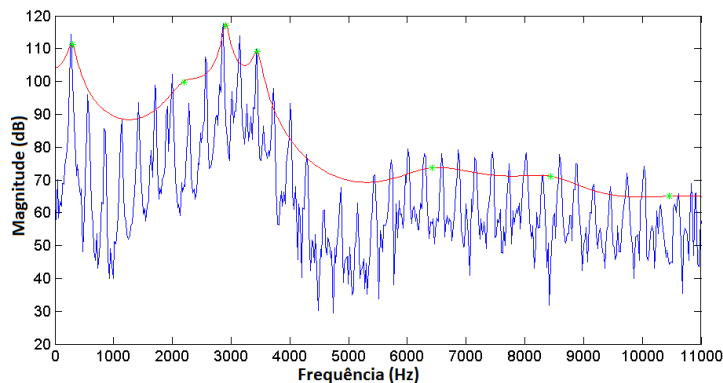
outros mais poderosos na sua vizinhança.

Assim, em vez da deteção de picos espectrais, é utilizada a localização dos mínimos da segunda derivada da envolvente para localizar os possíveis formantes. Na Figura 3.6 está ilustrado este processo.

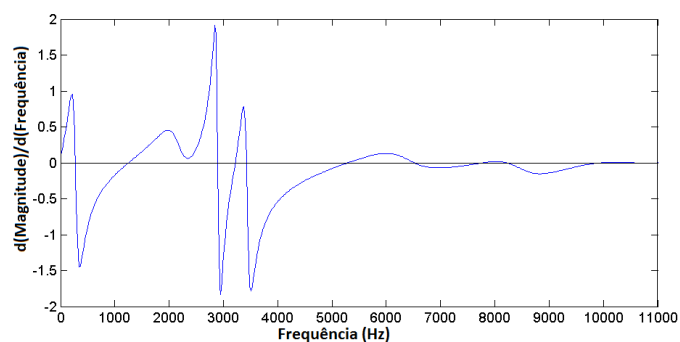
Obtidos os candidatos, estes serão validados por métodos heurísticos de forma a reduzir a possibilidade de estimação de falsos positivos. No caso da envolvente *cepstral* do resíduo este processo tem maior importância, pois, a aplicação deste método sobre a estimação da componente não vozeada da voz funciona mais como um último nível de suavização do que o princípio em que geralmente é utilizado, para separação da fonte do filtro; dando normalmente e dependendo do grau que é aplicado, um número de candidatos superior ao número de reais formantes existente.

Obtidos e validados os candidatos dos quatro métodos utilizados, a seleção de quais os candidatos da janela que está a ser analisada vão ser escolhidos como possíveis formantes, depende da frequência fundamental, obtida pelo detetor de *pitch*.

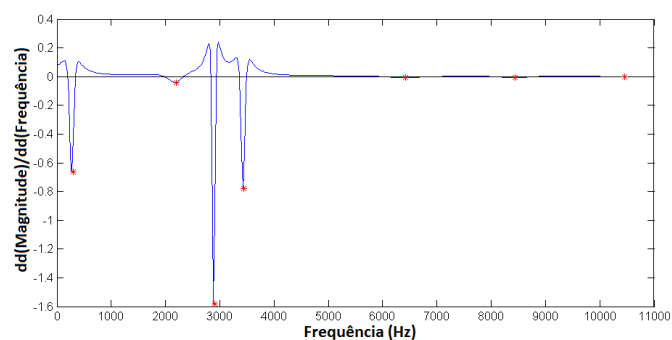
Este segundo nível de processamento está ilustrado no diagrama de blocos na Figura 3.7.



(a) O espectro do sinal a azul, a envolvente LPC do mesmo a vermelho e a verde as marcações dos candidatos a formantes encontrados pelo processamento da segunda derivada



(b) A primeira derivada da envolvente representada na Figura 3.6(a)



(c) A segunda derivada da envolvente representada na Figura 3.6(a) e a marcação a vermelho dos candidatos encontrados

Figura 3.6: Processo de estimação de candidatos a formantes pela segunda derivada da envolvente espectral

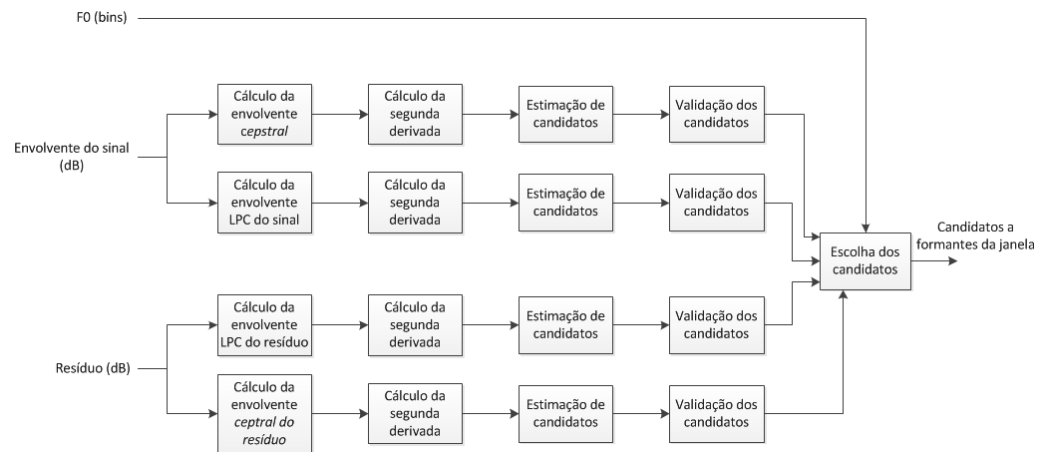


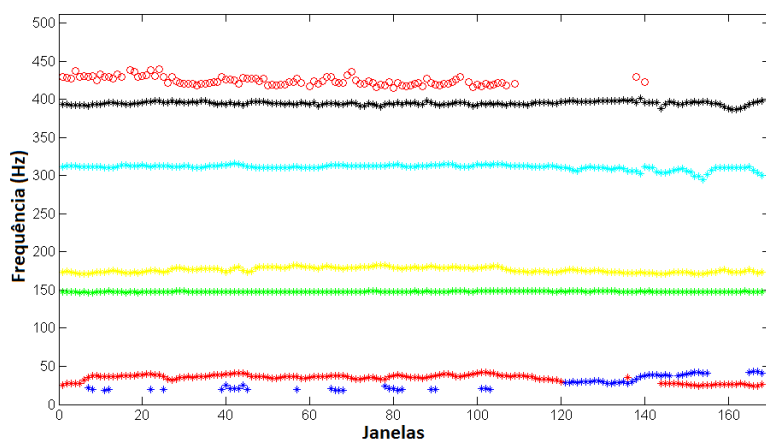
Figura 3.7: Segundo nível de processamento.

3.5 Mapeamento temporal dos formantes

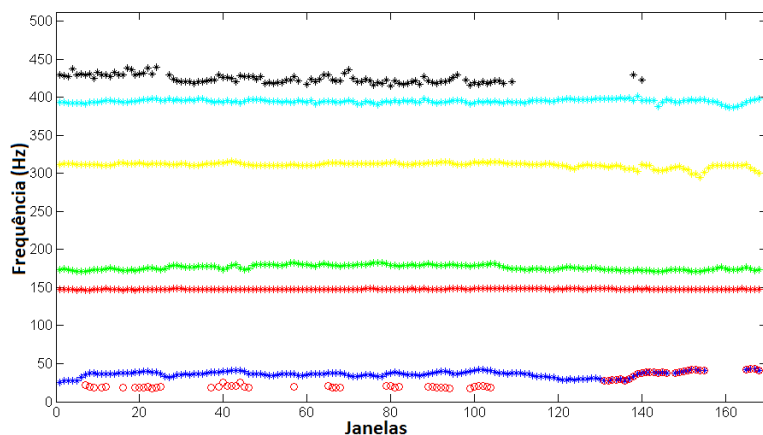
Com a estimação de possíveis formantes a nível da janela já efetuada, queremos verificar o seu comportamento do ponto de vista temporal, como os deslocamentos e a interação entre eles.

Assim foi pensado um algoritmo de catalogação dos formantes do ponto de vista da distribuição espectral e temporal com o intuito de poder serem representadas as trajetórias ao longo do espectro e do tempo.

Como já mencionei, muitos problemas dificultam este *tracking*; sendo a possível proximidade entre formantes um fator importante para a existência de erros nesse processo. Estão ilustrados na Figura 3.8 dois possíveis erros devido à proximidade entre formantes; na Figura 3.8(a) é um caso de cruzamento dos dois primeiros *streams*, enquanto que na Figura 3.8(b) é um caso de sobreposição, isto por escolha dos mesmos candidatos descartando os outros.



(a) Possível erro de *tracking* devido à proximidade entre formantes



(b) Possível erro de *tracking* devido à proximidade entre formantes

Figura 3.8: Dois possíveis erros de *tracking* (Vogal 'o' de um Contralto)

Assim, o método desenvolvido é implementado à medida que o processamento do sinal vai sendo efetuado, e os resultados da estimação de cada uma das janelas já processadas for obtido,

estando pensado para funcionar em tempo real.

O princípio do algoritmo consiste na colocação dos possíveis formantes num *stream* consoante a sua localização espectral; esta colocação é regida pela regra da proximidade, ou seja, os resultados mais recentes serão comparados só com os últimos valores que foram colocados em cada *stream* de forma a ser encontrada a melhor relação de proximidade para a alocação. A ideia está ilustrada na Figura 3.9.

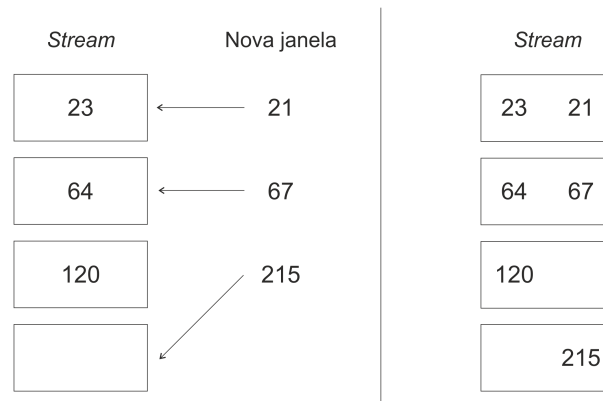


Figura 3.9: Exemplificação da alocação de formantes no *stream*

Esta relação de proximidade é calculada por uma matriz das diferenças entre os formantes da janela a ser analisada com os últimos valores alocados em cada *stream*; sendo verificada a proximidade dos novos formantes aos já existentes pela ordem da diferença mais pequena para a maior. Essa informação de proximidade, bem como a informação de formantes que não estão próximos a nenhum formante já existente, será então utilizada para a alocação destes novos formantes nos respetivos *streams* e o preenchimento de algum que se encontra vazio com novos formantes. O diagrama de blocos deste algoritmo está representado na Figura 3.10.

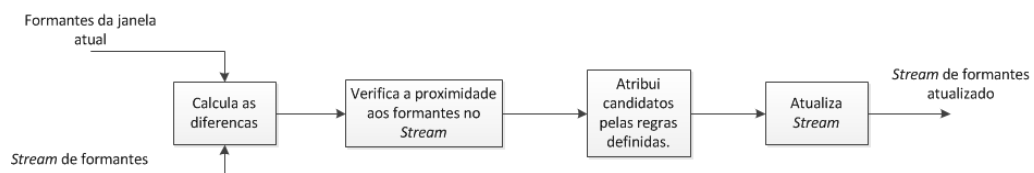


Figura 3.10: Processamento do *tracking* dos formantes

Está ilustrado na Figura 3.11 um possível erro de alocação de formantes; a vermelho estão representados os candidatos que foram descartados e que deveriam ter sido alocados no *stream* azul enquanto que os superiores deveriam ter sido alocados no *stream* verde.

O *stream* tem tamanho finito, dependendo do número de formantes que se quer estimar e seguir, sendo assim, só são colocados no *stream* os n formantes com maior amplitude por serem os mais relevantes. É também utilizada uma regra de continuidade, ou seja, de uma janela para

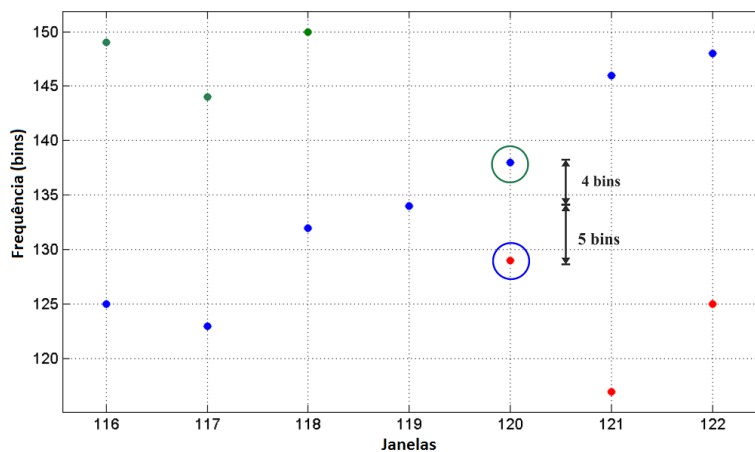


Figura 3.11: Pormenor da possibilidade de erro por proximidade (3^o e 4^o formantes da vogal 'a' de um Baixo).

outra, se houver um grande deslocamento do suposto formantes este não será alocado no *stream* mais próximo havendo uma falha nesse *stream*.

3.6 Conclusões

Neste capítulo foram descritos os dois métodos de caracterização do trato vocal utilizados nos desenvolvimentos feitos, o LPC e o *cepstrum* e explicado o seu enquadramento neste projeto.

Seguidamente foi explorado o algoritmo desenvolvido, sendo descrito e explicado o seu princípio de funcionamento, bem como as ideias gerais das abordagens, como a metodologia de estimação do ruído da voz, para a partir dele se extrair as características do trato vocal; e o método empregue para diferenciar e seguir, do ponto de vista temporal, os formantes bem como são descritas as dificuldades inerentes às abordagens tomadas, principalmente, o caso da possibilidade da existência de formantes muito próximos uns dos outros, sendo muito comum na voz cantada.

No seguinte capítulo, são explorados resultados obtidos pelo algoritmo desenvolvido.

Capítulo 4

Teste e resultados

4.1 Introdução

Neste capítulo são descritas as base de dados de voz, tanto sintéticas como naturais que foram criadas.

São também descritos os testes efetuados e apresentados exemplos demonstrativo dos resultados obtidos e a descrição das conclusões a que cheguei e as dificuldades que enfrentei.

4.2 Base de dados

Para o desenvolvimento deste projeto foi necessário acesso a exemplos de voz, de preferencialmente cantada. Assim, e como a dissertação se engloba no já mencionado projeto FCT para o qual também é necessário o acesso a base de dados de voz, foi criada tanto uma base de voz natural cantada como também uma base de dados de voz sintetizada.

Por questões de logística e de disponibilidade de cantores treinados foi mais demorada do que o esperado a criação da base de dados de canto, tendo o desenvolvimento sido levado a cabo nos primeiros tempos só com a utilização de vozes sintéticas.

4.2.1 Voz sintetizada

Uma das primeiras tarefas a ser feita foi a criação de uma base de dados de voz sintética, a qual foi criada de forma a representarem oito vogais cardinais, com frequências fundamentais de 220 Hz, 440 Hz e 880 Hz.

Para as frequências das ressonâncias das vogais representadas foram escolhidos os valores representados na Tabela 4.1. Os valores das primeiras quatro ressonâncias foram retirados do artigo *A two-formant model and the cardinal vowels* de Fant e Bladon [38], enquanto que o valor da quinta ressonância foi adicionada acrescentando 1 kHz à quarta ressonância.

Foi primeiro utilizado o sintetizador *Madde* [39], mas este, sintetiza as vozes dando mais importância à componente harmónica do sinal e deixando em segundo plano a componente não vozeada, o ruído da voz, não sendo este modelado pelo trato vocal.

Tabela 4.1: Frequências dos formantes das vogais sintetizadas

	Vogal	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)	F5 (Hz)
1	i	300	2300	3070	3590	4590
2	e	470	2180	2720	3790	4790
3	ε	680	1890	2580	3940	4940
4	a	770	1400	2460	3710	4710
5	α	660	1170	2770	3650	4650
6	ɔ	570	840	2640	3310	4310
7	o	370	730	2670	3240	4240
8	u	290	700	2550	3280	4280

Assim, e porque nos interessa estudar a utilização da componente não vozeada da voz para estimação dos formantes, foi então utilizado, com autorização do autor, um sintetizador desenvolvido em *MatLab* por Ricardo Sousa para a sua tese de doutoramento intitulada de "Metodologias de Avaliação Percetiva e Acústica do Sinal de Voz em Aplicações de Ensino do Canto e Diagnóstico/Reabilitação da Fala" [40]. Este sintetizador tem a preocupação de, para além da componente harmónica, modelar segundo o trato vocal, a componente não vozeada.

Este sintetizador foi criado baseado na teoria fonte-filtro, e que permite a reprodução de eventos específicos no sinal, como o ruído de origem natural existente no sinal de voz, a frequência fundamental, o *Jitter*, o *Shimmer*, a resposta impulsional do filtro representativo do trato vocal e ainda vários parâmetros do impulso glótico [40]. O diagrama de blocos está representado na Figura 4.1.

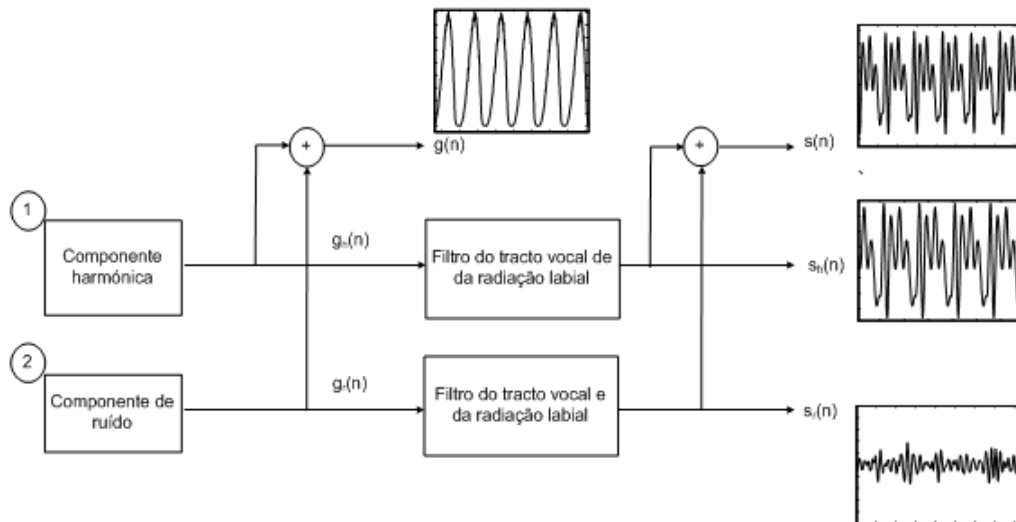
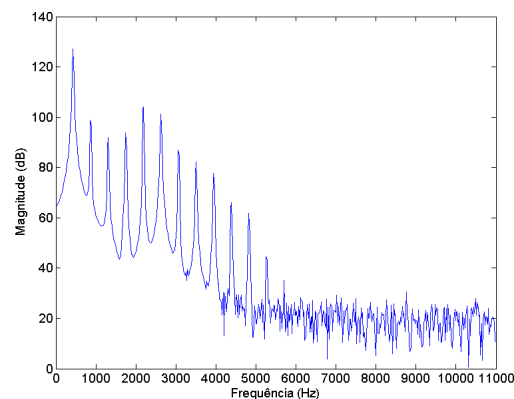


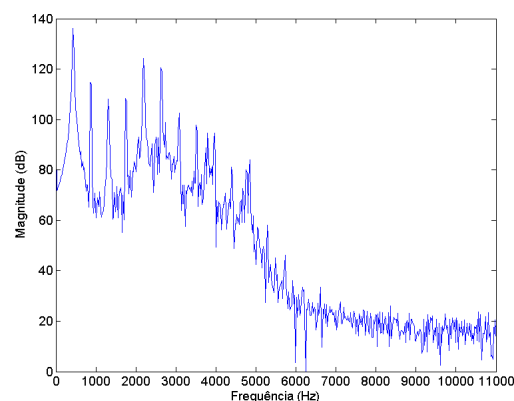
Figura 4.1: Diagrama de blocos do sintetizador utilizado [40]

As diferenças entre a mesma vogal sintetizada pelos dois sintetizadores podem ser vistas na Figura 4.2. É clara a diferença de abordagem em relação à componente de ruído da voz.

Ficou então decidido a utilização das vozes sintetizadas pelo sintetizador desenvolvido pelo Ricardo Sousa, por este modelar também a componente não vozeada da voz.



(a) O espectro da vogal sintética "e" obtida pelo sintetizador *Madde*



(b) O espectro da vogal sintética "e" obtida pelo sintetizador desenvolvido por Ricardo Sousa

Figura 4.2: Diferença dos resultados dos dois sintetizadores abordados

4.2.2 Voz natural

Como já mencionado houve algumas dificuldades de logística e de disponibilidade de cantores para a realização das gravações da base de dados de canto. Ultrapassadas estas dificuldades foram feitas as gravações com o intuito de criar não só uma base de dados para este projeto de dissertação, mas também para o projeto FCT ao qual esta dissertação reporta.

Assim, as gravações tiveram lugar num estúdio da ESMAE (Escola Superior de Música e Artes do Espetáculo). Embora não tenham sido realizadas em câmara anecoica, teve-se o devido cuidado de utilizar uma sala com bom isolamento acústico, equipamento de qualidade profissional e de a distância ao microfone ser sempre a mesma.

O microfone utilizado foi um microfone de estúdio *Neumann*, referencia *TLM 103*, as suas características encontram-se em anexo. O pré-amplificador foi o *Onyx800R* da *Mackie*, estando as suas especificações também em anexo. O *software* utilizado foi o *Digital Performer 5* da *MOTU* a correr numa máquina *Apple, Mac Pro*.

As gravações foram feitas de forma a cobrir as necessidades do projeto FCT, assim, foram gravadas recorrendo a cantores treinados e de forma a serem obtidos registos de Baixo, Baritono e Tenor de vozes masculinas, e registos de Contralto, Mezzo e Soprano de vozes femininas.

Os registos destes seis tipo de voz foram as cinco vogais; 'a', 'e', 'i', 'o' e 'u', na forma sustentada; uma frase de um tema, sendo a mesma frase e o mesmo tema para todos.

Foram feitos registos extra de Tenor e de Soprano, um exercício de demonstração de coloratura, de *legato*, portamento e *staccato*; demonstração da existência e do efeito de formante de cantor; de exemplificação da utilização de voz na frente ou de voz recuada e do efeito da existência de ar na voz, soproidade.

4.3 Testes

Os testes efetuados foram utilizando tanto a base de dados sintéticas como a de canto natural, e, mesmo tendo sido utilizadas todas as vozes pertencentes a estas duas bases de dados, os resultados aqui publicados estão centrados em vogais extremas no mapa de vogais cardinais; a vogal 'i' e a vogal 'a', no caso das sintéticas esta é a vogal 'a' aberta posterior.

Assim, para os testes das vozes sintéticas, e por ser sabido o valor teórico dos formantes do trato vocal foram efetuados testes e precisão e comparados com os resultados obtidos por uma importante ferramenta de análise de voz, o *Praat*.

Quanto aos testes com canto natural, e por não se conhecer a localização dos seus formantes, foram feitos testes e comparados com os resultados do *Praat*.

Para a extração das estimações dos formantes do *Praat*, este foi então configurado de forma a se aproximar do algoritmo desenvolvido. Assim, como mostra na Figura 4.3, foi configurado para deteção de oito formantes e com o tamanho da janela igual ao utilizado no algoritmo desenvolvido, sendo que, o *Praat* também utiliza sobreposição de 75% das janelas de análise para deteção de formantes.

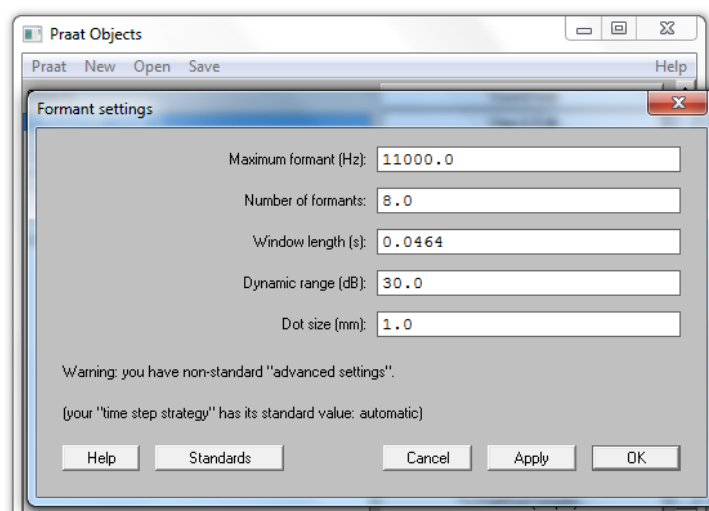


Figura 4.3: Configuração de estimação de formantes do *Praat*

4.3.1 Testes com vozes sintéticas

Os testes realizados com as vozes sintéticas, sendo conhecidos os formantes das vozes, foram feitos de forma a tentar verificar a boa estimação pelos quatro métodos aplicados neste projeto, de forma a poder ser percebido qual o efeito que o aumento da frequência fundamental traz à capacidade de os métodos localizarem os formantes.

É de conhecimento, que com o aumento da frequência fundamental e respetivo aumento da separação entre os harmónicos, os métodos de estimação de formantes têm tendência a serem influenciados pelos harmónicos estimando estes como possíveis formantes, por isso foi verificada também a relação dos valores obtidos com os valores teóricos dos harmónicos para se perceber a influência destes sobre os métodos.

Como referência para comparação, foi utilizado a estimação de formantes pelo *Praat*.

Os resultados obtidos para os erros relativos dos formantes individualizados estão representados nas Tabelas 4.2 e 4.4; e nas Figuras 4.4 e 4.5, estão representados os gráficos das médias dos erros.

Os resultados dos testes para as vozes com frequência fundamental mais elevada, de 880 Hz, estão representados nas Tabelas 4.3 e 4.5, os restantes resultados estão em anexo.

As tabelas com a descrição estatística dos valores obtidos encontram-se em anexo.

Tabela 4.2: Erros relativos das frequências estimadas pelos vários métodos para a vogal 'i'

	F0 (Hz)	F1	F2	F3	F4	F5
		Erro rel. (%)	Erro rel. (%)	Erro rel. (%)	Erro rel. (%)	Erro rel. (%)
Praat	220	18.36	59.60	57.01	50.77	75.46
	440	74.94	5.50	1.62	3.74	19.00
	880	191.21	24.09	15.19	3.40	5.74
LPC do sinal	220	14.25	1.94	0.56	2.35	0.97
	440	57.91	3.57	1.00	1.50	0.11
	880	187.11	23.23	13.03	1.63	3.83
Cepstrum do sinal	220	6.72	12.92	3.55	-	56.20
	440	76.24	11.27	4.78	-	1.57
	880	132.17	11.51	6.08	-	1.78
LPC do ruído	220	13.70	1.29	0.76	2.83	0.29
	440	57.91	3.57	1.00	1.41	2.55
	880	179.93	22.48	13.03	1.63	3.36
Cesptrum do ruído	220	63.54	7.48	-	6.65	2.44
	440	88.93	7.39	8.68	25.18	67.90
	880	14.38	1.45	-	6.51	1.23

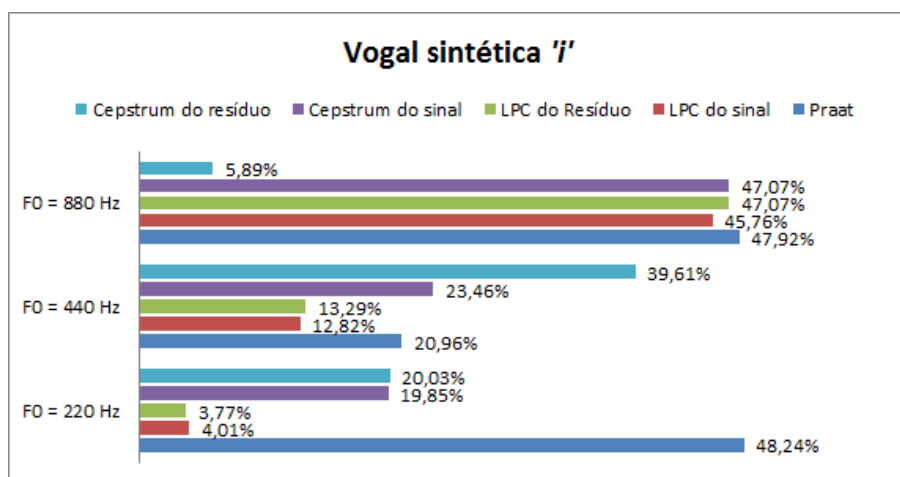


Figura 4.4: Gráfico representativo do erro médio de estimação dos formantes da vogal sintética 'i'

Tabela 4.3: Erros relativos em relação aos harmônicos teóricos, vogal 'i' com F0 = 880 Hz

	H1 (880 Hz)	H2 (1760 Hz)	H3 (2640 Hz)	H4 (3520 Hz)	H5 (4400 Hz)
	Erro rel. (%)	Erro rel. (%)	Erro rel. (%)	Erro rel. (%)	Erro rel. (%)
Praat	0.72	0.80	1.33	1.47	1.66
LPC do sinal	2.12	0.33	1.14	0.33	0.33
Cepstrum do sinal	20.85	15.65	23.36	32.72	6.17
LPC do resíduo	4.57	1.31	1.14	0.33	0.81
Cepstrum do resíduo	70.81	32.58	-	4.65	3.04

Tabela 4.4: Erros relativos das frequências estimadas pelos vários métodos para a vogal 'a'

		F1	F2	F3	F4	F5
	F0 (Hz)	Erro rel. (%)	Erro rel. (%)	Erro rel. (%)	Erro rel. (%)	Erro rel. (%)
Praat	220	2.18	3.46	7.27	9.12	8.97
	440	20.72	0.89	27.38	14.78	12.75
	880	31.31	47.19	6.03	4.85	5.43
LPC do sinal	220	1.14	4.33	0.60	0.64	0.45
	440	23.98	1.22	4.71	3.46	0.04
	880	33.77	49.08	4.38	3.07	3.22
Cepstrum do sinal	220	19.88	-	5.87	14.63	0.70
	440	33.09	-	6.62	-	0.42
	880	41.53	-	11.19	16.54	1.13
LPC do ruído	220	1.14	4.14	1.22	0.67	1.59
	440	24.29	1.84	4.28	2.38	2.09
	880	33.77	49.14	3.97	2.66	3.58
Cepstrum do ruído	220	10.77	38.81	2.02	1.10	1.15
	440	27.84	36.91	2.10	0.61	1.41
	880	40.64	46.60	0.95	0.34	1.37

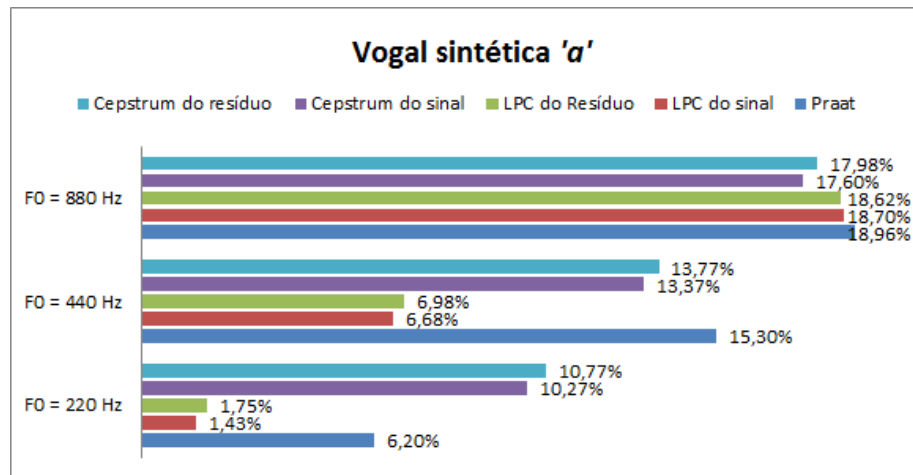


Figura 4.5: Gráfico representativo do erro médio de estimação dos formantes da vogal sintética 'a'

Tabela 4.5: Erros relativos em relação aos harmónicos teóricos, vogal 'a' com F0 = 880 Hz

	H1 (880 Hz)	H2 (1760 Hz)	H3 (2640 Hz)	H4 (3520 Hz)	H5 (4400 Hz)
	Erro rel. (%)	Erro rel. (%)	Erro rel. (%)	Erro rel. (%)	Erro rel. (%)
Praat	1.51	2.15	1.40	1.34	0.06
LPC do sinal	0.33	0.90	0.33	0.51	2.28
Cepstrum do sinal	6.15	-	6.82	13.46	6.87
LPC do resíduo	0.33	0.85	0.76	0.93	1.90
Cepstrum do resíduo	5.48	2.54	3.92	3.34	4.24

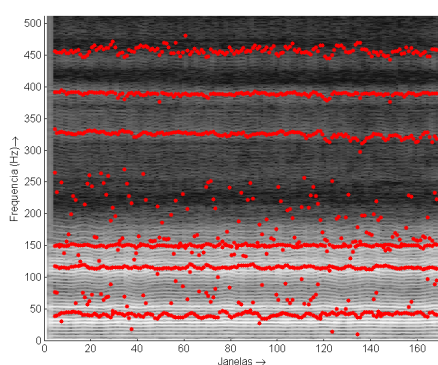
4.3.2 Testes com vozes naturais

Os testes realizados com vozes naturais, baseou-se mais em visualização dos resultados obtidos observando o espectro dos sinais e a sobreposição no espectrograma de forma a tentar perceber o comportamento dos métodos utilizados. Foram utilizadas vozes de um Baixo e um Tenor do género masculino, e de uma Contralto e de uma Soprano do género feminino; como já referido, as vogais 'i' e 'a'

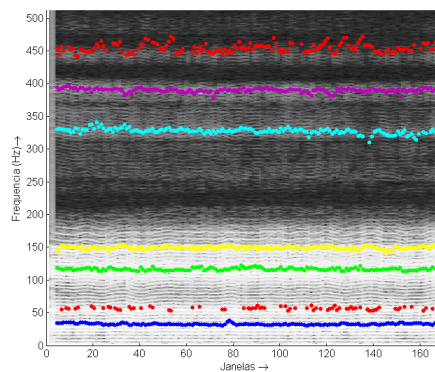
Foi também utilizado o *Praat* para a obtenção dos formantes para as mesmas vozes, de forma a serem comparados estes resultados com os métodos aplicados no algoritmo desenvolvido.

Estão ilustrados nas tabelas e figuras seguintes, os resultados obtidos sobrepostos ao espectrograma, tanto dos métodos utilizados no projeto como os resultados do *Praat*. De notar que o *Praat* não ordena e distingue os formantes detetados, sendo a representação monocromática e sem agrupar os formantes pela sua ordem, ao contrário do que se pretende neste projeto.

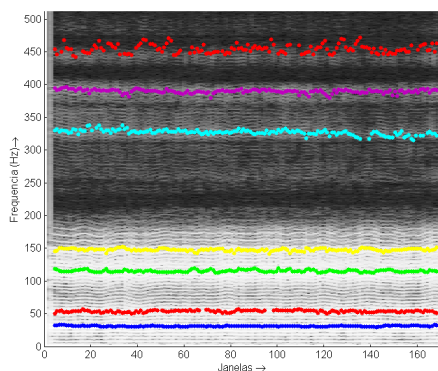
As tabelas completas com a descrição estatística dos valores obtidos detalhe encontra-se em anexo.



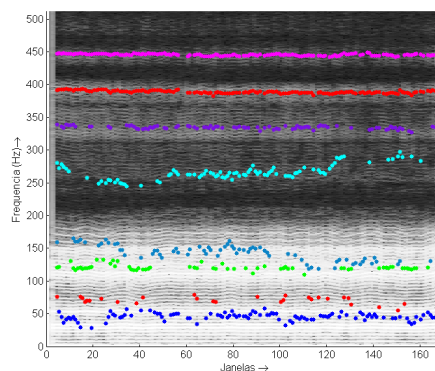
(a) Espectrograma e estimação de formantes do *Praat* da vogal 'a' de um Baixo



(b) Espectrograma e estimação de formantes do LPC do resíduo da vogal 'a' de um Baixo



(c) Espectrograma e estimação de formantes do LPC do sinal da vogal 'a' de um Baixo

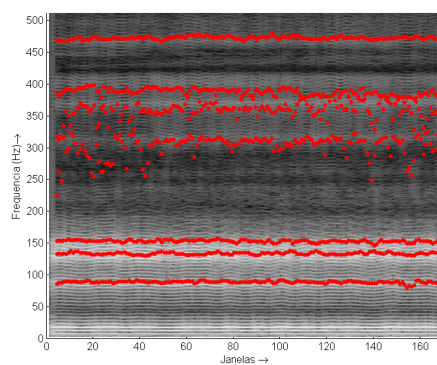
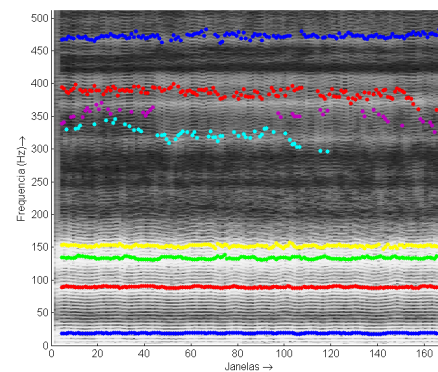


(d) Espectrograma e estimação de formantes do *Cepstrum* do resíduo da vogal 'a' de um Baixo

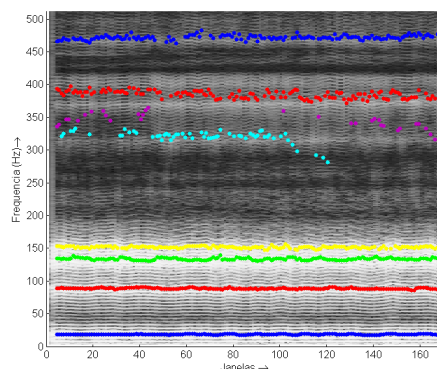
Figura 4.6: Estimação de formantes da vogal 'a' de um Baixo

Tabela 4.6: Médias dos formantes, vogal 'a' de um Baixo

		F1	F2	F3	F4	F5	F6	F7
<i>Praat</i>	Média	775.72	1947.74	2779.25	3267.26	5939.76	7716.68	9139.50
	D. P.	103.50	564.62	337.94	270.34	1178.10	663.80	727.38
LPC do sinal	Média	578.82	1071.58	2404.88	3086.98	6957.68	8294.80	9748.72
	D. P.	15.64	32.97	36.09	42.95	86.13	58.80	140.52
LPC do resíduo	Média	613.89	1135.24	2434.15	3118.70	6964.43	8297.37	9743.22
	D. P.	29.03	44.00	42.24	47.07	88.73	58.45	148.41
<i>Cepstrum</i> do sinal	Média	2611.45	7009.06	-	-	-	-	-
	D. P.	30.97	45.68	-	-	-	-	-
<i>Cepstrum</i> do resíduo	Média	863.77	2794.28	5579.18	5951.33	7109.60	8273.86	9504.78
	D. P.	123.46	253.85	274.92	157.12	48.98	44.78	31.86

(a) Espectrograma e estimação de formantes do *Praat* da vogal 'i' de um Baixo

(b) Espectrograma e estimação de formantes do LPC do resíduo da vogal 'i' de um Baixo



(c) Espectrograma e estimação de formantes do LPC do sinal da vogal 'i' de um Baixo

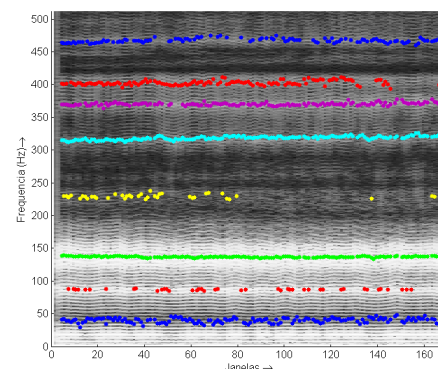
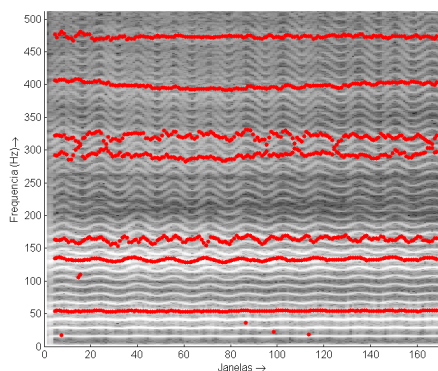
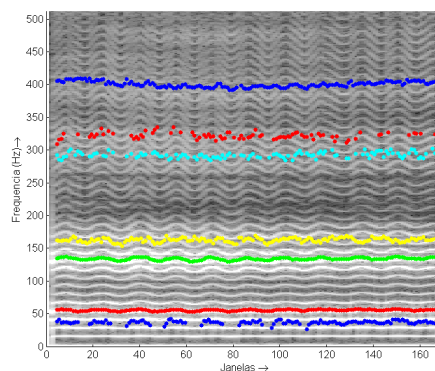
(d) Espectrograma e estimação de formantes do *Cepstrum* do resíduo da vogal 'i' de um Baixo

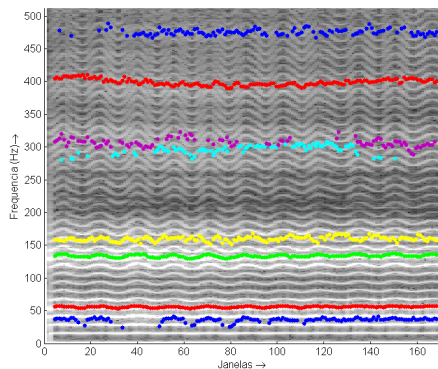
Figura 4.7: Estimação de formantes da vogal 'i' de um Baixo

Tabela 4.7: Médias dos formantes, vogal 'i' de um Baixo

		F1	F2	F3	F4	F5	F6	F7
<i>Praat</i>	Média	1819.80	2787.71	3206.31	6421.47	7632.48	8259.75	10089.81
	D. P.	36.94	37.05	39.05	383.34	281.55	113.79	51.58
LPC do sinal	Média	303.79	1824.91	2783.59	3170.43	7538.67	8211.76	10071.34
	D. P.	13.58	21.75	37.40	34.24	149.33	109.74	67.71
LPC do resíduo	Média	311.01	1837.03	2782.17	3178.22	6959.46	8239.85	10084.96
	D. P.	13.50	20.32	33.97	35.16	262.84	141.68	73.04
<i>Cepstrum</i> do sinal	Média	-	-	-	-	-	-	-
	D. P.	-	-	-	-	-	-	-
<i>Cepstrum</i> do resíduo	Média	774.06	2850.73	4852.90	6773.27	7892.99	8580.39	9973.99
	D. P.	81.70	21.73	57.83	60.85	46.63	77.73	65.65

(a) Espectrograma e estimação de formantes do *Praat* da vogal 'a' de um Tenor

(b) Espectrograma e estimação de formantes do LPC do resíduo da vogal 'a' de um Tenor



(c) Espectrograma e estimação de formantes do LPC do sinal da vogal 'a' de um Tenor

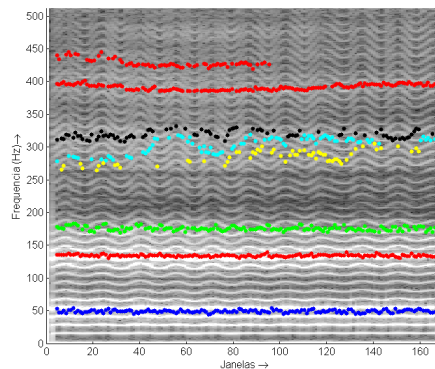
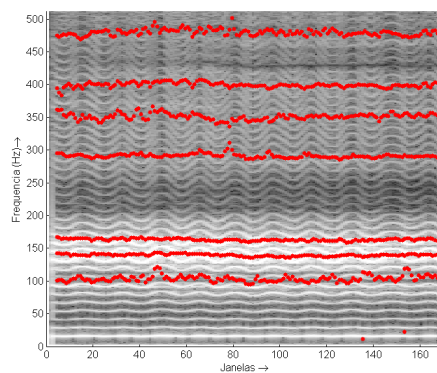
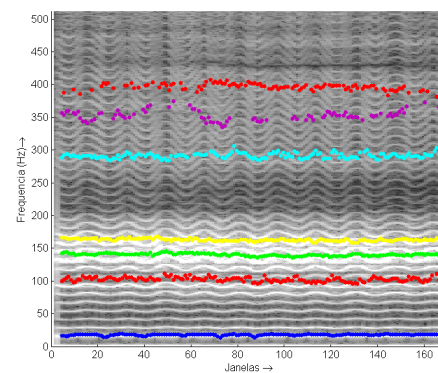
(d) Espectrograma e estimação de formantes do *Cepstrum* do resíduo da vogal 'a' de um Tenor

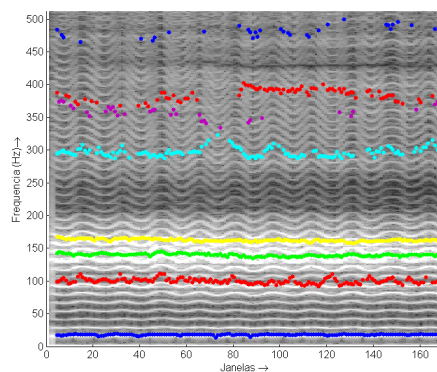
Figura 4.8: Estimação de formantes da vogal 'a' de um Tenor

Tabela 4.8: Médias dos formantes, vogal 'a' de um Tenor

		F1	F2	F3	F4	F5	F6	F7
<i>Praat</i>	Média	1064.14	2727.13	3413.21	6104.04	6798.54	8433.97	10043.30
	D. P.	105.97	269.66	143.22	527.05	139.90	328.43	292.04
LPC do sinal	Média	689.53	1114.70	2782.68	3352.35	6386.86	8496.46	10155.06
	D. P.	67.11	22.17	41.19	75.04	158.79	102.22	94.65
LPC do resíduo	Média	702.94	1105.67	2787.84	3411.12	6251.92	6759.21	8525.46
	D. P.	54.33	23.75	40.17	69.08	146.99	184.38	93.49
<i>Cepstrum</i> do sinal	Média	961.60	3164.87	6107.76	-	-	-	-
	D. P.	110.69	36.61	53.70	-	-	-	-
<i>Cepstrum</i> do resíduo	Média	970.93	2793.13	3693.92	6063.04	6740.88	8332.66	9125.51
	D. P.	45.76	33.85	71.23	200.58	153.98	92.18	127.73

(a) Espectrograma e estimação de formantes do *Praat* da vogal 'i' de um Tenor

(b) Espectrograma e estimação de formantes do LPC do resíduo da vogal 'i' de um Tenor



(c) Espectrograma e estimação de formantes do LPC do sinal da vogal 'i' de um Tenor

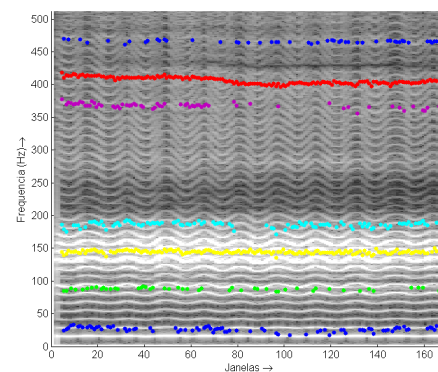
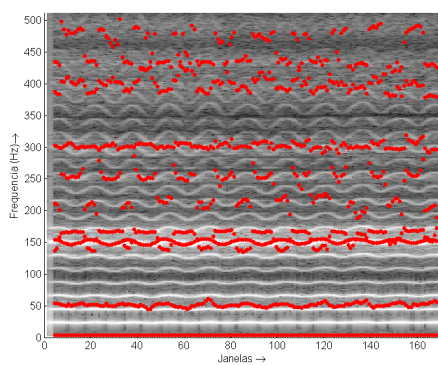
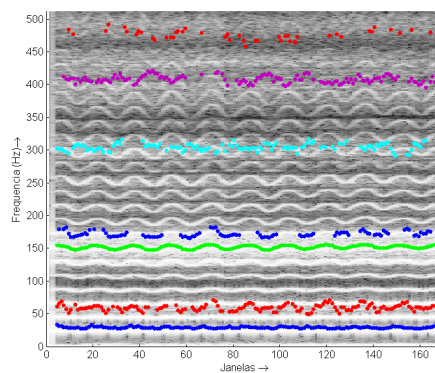
(d) Espectrograma e estimação de formantes do *Cepstrum* do resíduo da vogal 'i' de um Tenor

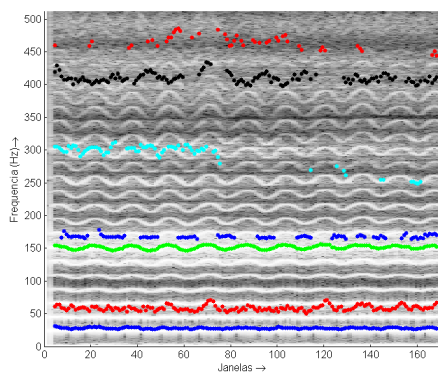
Figura 4.9: Estimação de formantes da vogal 'i' de um Tenor

Tabela 4.9: Médias dos formantes, vogal 'i' de um Tenor

		F1	F2	F3	F4	F5	F6	F7
<i>Praat</i>	Média	2122.70	2913.87	3420.26	6151.54	7459.70	8502.70	10219.88
	D. P.	228.57	66.69	70.69	309.20	178.90	130.87	216.37
LPC do sinal	Média	294.50	2076.09	2913.17	3403.79	6301.17	7715.29	10240.15
	D. P.	17.89	87.78	44.16	41.59	145.52	261.07	153.73
LPC do resíduo	Média	290.38	2128.22	2923.74	3413.59	6182.27	7569.78	8477.67
	D. P.	22.53	72.11	41.78	41.06	87.40	134.02	103.91
<i>Cepstrum</i> do sinal	Média	483.42	3070.87	6463.00	-	-	-	-
	D. P.	137.34	33.33	118.61	-	-	-	-
<i>Cepstrum</i> do resíduo	Média	522.18	1810.55	3012.82	3916.81	8649.00	9945.81	-
	D. P.	73.34	43.92	50.07	88.01	106.67	40.32	-

(a) Espectrograma e estimação de formantes do *Praat* da vogal 'a' de uma Contralto

(b) Espectrograma e estimação de formantes do LPC do resíduo da vogal 'a' de uma Contralto



(c) Espectrograma e estimação de formantes do LPC do sinal da vogal 'a' de uma Contralto

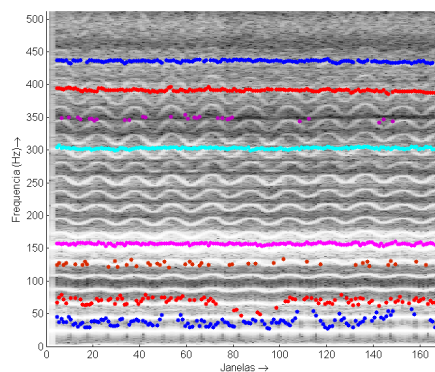
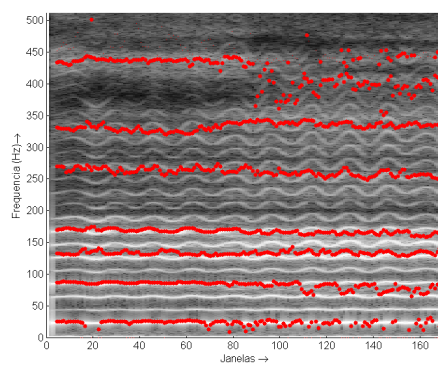
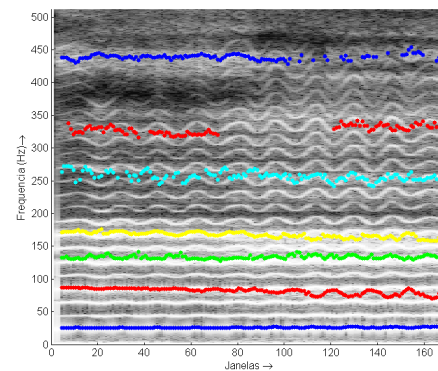
(d) Espectrograma e estimação de formantes do *Cepstrum* do resíduo da vogal 'a' de uma Contralto

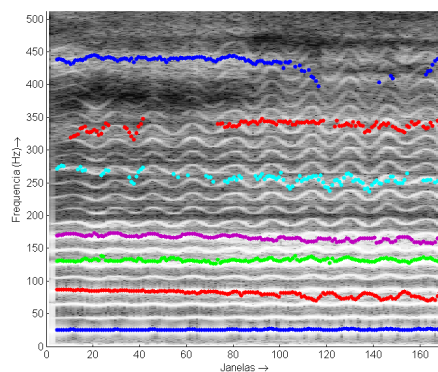
Figura 4.10: Estimação de formantes da vogal 'a' de uma Contralto

Tabela 4.10: Médias dos formantes, vogal 'a' de uma Contralto

		F1	F2	F3	F4	F5	F6	F7
<i>Praat</i>	Média	1019.86	3083.53	3399.20	4984.18	6421.92	8512.76	9704.46
	D. P.	61.48	104.82	130.93	522.87	99.37	249.36	531.16
LPC do sinal	Média	510.22	1179.04	3170.41	3517.45	5436.24	6405.63	8846.18
	D. P.	21.37	85.86	57.55	42.12	103.17	94.92	188.31
LPC do resíduo	Média	530.98	1193	31174.92	3611.65	6454.58	8691.07	9092.75
	D. P.	26.51	103.58	52.67	93.01	114.09	130.66	103.74
<i>Cepstrum</i> do sinal	Média	921.16	1602.37	3232.30	6465.12	8115.48	9034.69	-
	D. P.	110.05	79.43	34.17	65.68	93.32	91.45	-
<i>Cepstrum</i> do resíduo	Média	690.59	1274.46	2638.84	3283.75	6428.37	8325.13	9292.73
	D. P.	99.83	265.49	55.71	29.79	27.54	39.54	32.47

(a) Espectrograma e estimação de formantes do *Praat* da vogal 'i' de uma Contralto

(b) Espectrograma e estimação de formantes do LPC do resíduo da vogal 'i' de uma Contralto



(c) Espectrograma e estimação de formantes do LPC do sinal da vogal 'i' de uma Contralto

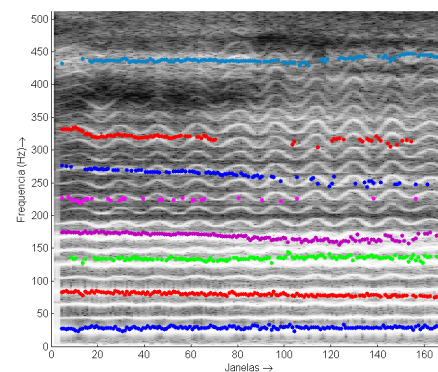
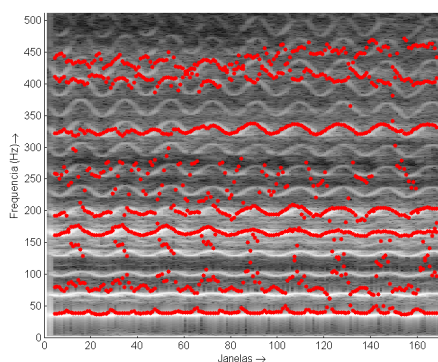
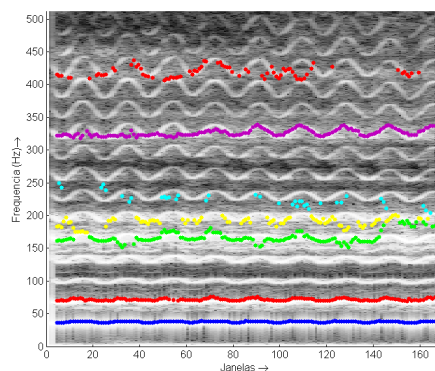
(d) Espectrograma e estimação de formantes do *Cepstrum* do resíduo da vogal 'i' de uma Contralto

Figura 4.11: Estimação de formantes da vogal 'i' de uma Contralto

Tabela 4.11: Médias dos formantes, vogal 'i' de uma Contralto

		F1	F2	F3	F4	F5	F6	F7
<i>Praat</i>	Média	626.59	1924.90	2914.38	3897.34	5848.27	7356.52	9037.36
	D. P.	455.41	429.07	283.14	790.59	642.78	643.55	512.57
LPC do sinal	Média	462.00	1658.31	2750.19	3489.80	5497.94	7014.30	9306.95
	D. P.	13.52	99.61	48.83	89.85	199.85	163.68	125.59
LPC do resíduo	Média	463.80	1676.37	2779.74	3509.14	54.37	6991.64	9226.35
	D. P.	14.71	93.28	57.41	92.51	131.87	139.05	287.92
<i>Cepstrum</i> do sinal	Média	500.95	1614.79	3250.35	7183.01	-	-	-
	D. P.	133.90	170.10	76.71	127.67	-	-	-
<i>Cepstrum</i> do resíduo	Média	525.69	1634.09	2806.39	3526.23	4747.70	5564.43	6804.23
	D. P.	40.96	54.50	58.15	111.39	38.44	162.00	127.48

(a) Espectrograma e estimação de formantes do *Praat* da vogal 'a' de uma Soprano

(b) Espectrograma e estimação de formantes do LPC do resíduo da vogal 'a' de uma Soprano

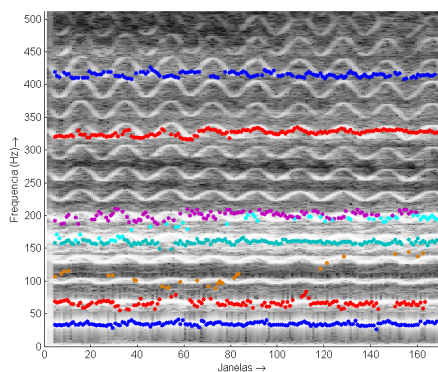
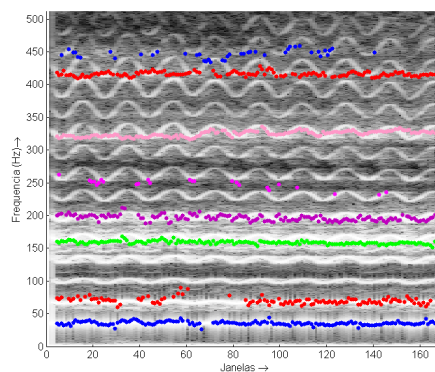
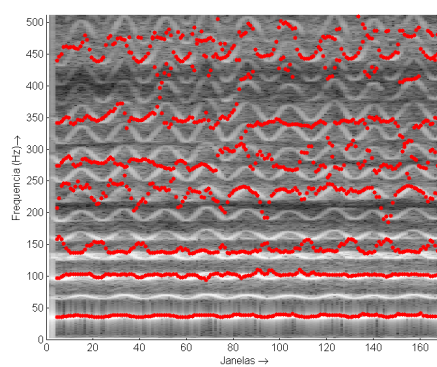
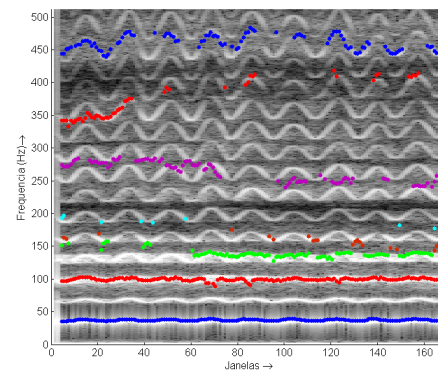
(c) Espectrograma e estimação de formantes do *Cepstrum* do sinal da vogal 'a' de uma Soprano(d) Espectrograma e estimação de formantes do *Cepstrum* do resíduo da vogal 'a' de uma Contralto

Figura 4.12: Estimação de formantes da vogal 'a' de uma Soprano

Tabela 4.12: Médias dos formantes, vogal 'a' de uma Soprano

		F1	F2	F3	F4	F5	F6	F7
<i>Praat</i>	Média	779.80	1744.20	3321.56	3953.84	5180.66	7027.30	8740.06
	D. P.	67.36	318.91	257.55	268.53	718.79	334.99	239.74
LPC do sinal	Média	717.17	1441.95	3521.52	4348.01	5049.54	6947.70	8959.25
	D. P.	18.42	24.44	115.30	186.38	45.68	113.92	214.93
LPC do resíduo	Média	719.11	1460.08	3435.31	3976.94	4542.63	6949.26	8677.88
	D. P.	18.60	35.02	118.05	134.40	362.94	113.61	109.44
<i>Cepstrum</i> do sinal	Média	650.30	1422.58	3348.07	4202.03	5271.65	6951.12	8839.99
	D. P.	50.51	247.50	69.29	117.92	103.45	88.83	75.85
<i>Cepstrum</i> do resíduo	Média	676.14	1426.77	3329.78	4142.38	5285.11	6910.72	8863.62
	D. P.	59.75	103.60	52.18	89.87	121.98	93.23	67.64

(a) Espectrograma e estimação de formantes do *Praat* da vogal 'i' de uma Soprano

(b) Espectrograma e estimação de formantes do LPC do resíduo da vogal 'i' de uma Soprano

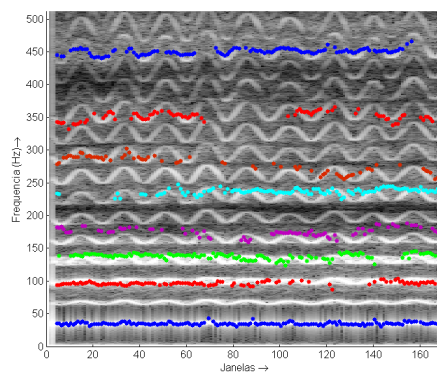
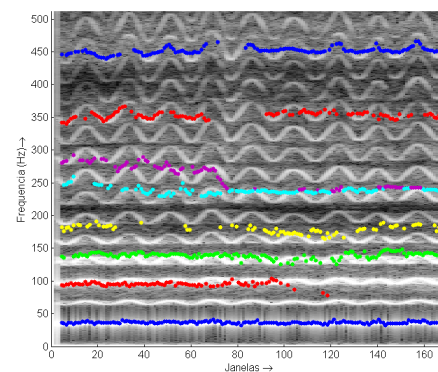
(c) Espectrograma e estimação de formantes do *Ceps-* (d) Espectrograma e estimação de formantes do *Cepstrum* do resíduo da vogal 'i' de uma Soprano

Figura 4.13: Estimação de formantes da vogal 'i' de uma Soprano

Tabela 4.13: Médias dos formantes, vogal 'i' de um Soprano

		F1	F2	F3	F4	F5	F6	F7
<i>Praat</i>	Média	718.15	2108.72	2992.28	4879.92	5952.16	7317.65	9460.03
	D. P.	29.81	50.93	145.48	286.14	253.66	261.76	391.61
LPC do sinal	Média	711.63	2059.88	2903.54	4021.33	5618.31	7304.52	9883.48
	D. P.	23.14	47.04	143.48	104.57	319.05	162.64	252.16
LPC do resíduo	Média	710.86	2077.82	2908.99	4026.71	5844.68	7376.44	9769.11
	D. P.	22.36	42.98	108.20	160.32	135.08	205.64	193.23
<i>Cepstrum</i> do sinal	Média	660.05	1997.91	2867.83	3678.19	5025.33	5889.03	7480.99
	D. P.	42.64	53.80	84.59	141.58	91.61	255.50	143.62
<i>Cepstrum</i> do resíduo	Média	695.94	1967.28	2897.53	3739.60	5711.51	5978.26	7520.27
	D. P.	34.76	74.95	94.19	134.68	135.16	101.03	103.08

4.4 Conclusões

Dos teste efetuados com as vozes sintéticas, e utilizando as vogais 'i' e 'a', denota-se bastante dificuldade em estimar corretamente o primeiro formante; sendo que para a vogal 'i' com frequência fundamental de 440 Hz e 880 Hz, e para a vogal 'a' com 880 Hz, este primeiro formante é mais baixo que o seu harmónico fundamental, havendo assim um acréscimo de dificuldade. Neste caso, o *Cepstrum* do resíduo é o que mais próximo se encontra da estimação desse formante, por menos depender da componente harmónica.

Pôde-se também verificar, que por exemplo no caso da vogal 'a' que tem os primeiros três formantes mais baixos do que a vogal 'i', que os métodos testados encontram mais dificuldade em deteta-los com precisão.

Foi também realizado o teste de proximidade das estimações em relação aos harmónicos teóricos do sinal, para se verificar a tendência dos métodos serem afetados pelos harmónicos, especialmente em vozes com frequência fundamental elevada. Assim, pôde-se verificar, que no caso do *Praat* e para voz com frequência fundamental de 880 Hz, existe tendência da detecção ser afetada pelos harmónicos, bem como com o LPC do sinal e até mesmo o do resíduo que foram implementados; sendo o *Cepstrum* do resíduo o menos influenciável pelos harmónicos.

Embora se possam tirar algumas conclusões com os testes sobre voz sintetizada, estas na minha opinião não são muito esclarecedoras, visto as vozes sintéticas não terem características importantes que existem na voz natural, não tendo componente harmónica nas alta frequência, sendo muito estável e limpo; ao contrario do sinal de voz natural de canto, que pode ter componente harmónica bem definidas nas altas frequências, podendo também ter mais ruído.

Dos teste com vozes naturais de canto, pode-se verificar que para frequências fundamentais baixas, o *Cepstrum* do sinal não tem bons resultados, devido à dependência que têm da frequência fundamental para estimação do trato vocal; obtém contudo resultados interessantes para a voz de uma Soprano.

O método do *Cepstrum* do resíduo, embora tenha resultados interessantes para todas as vozes testadas, este para frequências fundamentais baixas é pouco regular e preciso, melhorando os resultados à medida que a frequência fundamental seja elevada.

Os métodos de LPC do sinal e do resíduo, têm comportamento bastante paralelo, similar, dando bons resultados na generalidade das vozes, inclusive para a voz de Soprano. Têm, em comparação com o *Praat*, uma maior definição nos formantes de baixa frequência.

De notar também que, o *Praat*, não faz qualquer tipo de alocação e diferenciação dos formantes a nível temporal, de janela para janela, estando até representados como formantes picos esporádicos sem qualquer tipo de continuidade. A obtenção da listas dos formantes do *Praat* serviu para eu verificar isso.

As principais dificuldades, dando origem a resultados pouco estáveis e precisos, prendem-se com a existência de vibrato; e a proximidade entre formantes, especialmente pelas vozes estudadas serem de canto, havendo assim a formação do formantes de cantor.

Capítulo 5

Conclusão

O trabalho realizado no decorrer desta dissertação foi debruçado sobre duas perspectivas. por outro lado, do ponto de vista da caracterização de parâmetros perceptivos utilizados na avaliação da voz cantada, tanto a nível de qualidade, como a nível estético. Por outro lado, do estudo e desenvolvimento de um método de estimação de formantes da voz cantada.

Verificou-se que os parâmetros de qualidade e percepção usualmente utilizados no canto são subjetivos, não havendo uma definição bem definida dos fatos que têm importância para que essa determinada característica perceptiva esteja presente na voz.

Foi identificado, que o trato vocal tem grande importância no canto, e nomeadamente devido à influência dos seus formantes. Assim estabeleceu-se como ponto de partida para a objetivação dos referidos parâmetros perceptivos, o estudo dos formantes na voz cantada.

Começou-se tendo como objeto de estudo vozes sintéticas, mas foram encontradas dificuldades relacionadas com as diferenças entre as características morfológicas das vozes sintéticas e as vozes de canto naturais, existindo fatores nas vozes naturais que os sintetizadores ignoram, como o ruído da voz e as componentes nas altas frequências.

Outra dificuldade com que me deparei, é a proximidade entre dois ou mais formantes dificultando tanto a sua deteção como o seu seguimento e catalogação em tempo real, isto acontece muito na voz cantada com a existência de aglomerações de formantes, como o caso do formante de cantor.

Do ponto de vista de vozes com frequência fundamental alta, restringi-me só à análise de voz sintética, devido à base de dados de canto natural criada ter como voz com a fundamental mais alta, a duma Soprano. Neste caso em que a voz tem uma frequência fundamental de aproximadamente 660 Hz, do ponto de vista de análise de formantes, o simples método baseado em LPC do sinal obtém bons resultados.

Outras dificuldades, do ponto de vista da deteção e estimação dos formantes prende-se com a existência de vibrato, com uma frequência assinalável, causando instabilidade da componente harmónica de uma janela de análise para outra, dificultando tanto a estimação por métodos baseados no sinal, ou mesmo dificuldade de estimação do resíduo dificultado assim o estudo dos métodos baseados no resíduo.

Penso que foi conclusivo, que para vozes com frequência fundamental elevada a utilização do ruído da voz é uma boa abordagem para a estimação dos formantes do trato vocal, por si só, ou como é o objetivo deste projeto, com a distribuição de esforço por outros métodos. Não cheguei porém a um algoritmo final e funcional pois não consegui, ainda, identificar quais as características diferenciadoras de que método utilizar e quando o utilizar.

5.1 Trabalho Futuro

Penso que um dos primeiros passos a ser feito é a migração dos algoritmos desenvolvidos em *MatLab* para ambiente *C/C++* e seguir os desenvolvimentos a partir dessa plataforma, isto devido a certos problemas encontrados na plataforma *MatLab* no decorrer do desenvolvimento.

O desenvolvimento do algoritmo de estimação de formantes ainda não cumpre os objetivos pretendidos, faltando determinar limites de utilização dos métodos, dependendo da frequência e de outros possíveis fatores, para cada um dos métodos envolvidos. Para isso será necessário levar a cabo testes com vozes, de preferência naturais, com elevada frequência fundamental de forma a ser estudada a resposta dos métodos e serem estabelecidas essas regras.

O algoritmo de catalogação e seguimento de formantes em tempo real, pode também ainda ser melhorado, possivelmente com a utilização de métodos estatísticos, como por exemplo modelos de *Markov* (HMM), de forma a arranjar um método robusto e resistente a erros devido a proximidade de formantes, erros esses que por vezes ainda acontecem.

Partindo da estimação dos formantes, a ideia é também identificar a existência do formante de cantor, bem como as suas características morfológicas, que estão largamente relacionadas a certos parâmetros percetivos; podendo assim ser desenvolvida uma representação visual e em tempo real dessas características, relacionando as com os parâmetros percetivos relevantes.

Foram identificadas outras características acústicas relacionadas a parâmetros percetivos, como é o caso da distribuição harmónica no espectro; pode assim ser desenvolvido um método de representação visual desta característica de forma a ser relacionada com os respetivos parâmetros percetivos, visto essa informação da componente harmónica já ser obtida pelo algoritmo de deteção de *pitch*, o *Searchtonal*.

Anexo A

Anexo A - Descrições estatísticas dos testes

A.1 Descrições estatísticas dos testes com vozes naturais

A.1.1 *Praat*

Tabela A.1: Descrições estatísticas dos resultados do *Praat* da vogal 'a' de um Baixo

	N	Gama	Média		Desvio P.	Variância
	Estatística	Estatística	Estatística	Erro	Estatística	Estatística
F1	165	791,30	775,72	8,06	103,50	10712,47
F2	165	1773,72	1947,74	43,96	564,62	318792,38
F3	165	845,44	2779,25	26,31	337,94	114200,92
F4	165	2002,69	3267,26	21,05	270,34	73084,61
F5	165	3925,29	5939,76	91,71	1178,10	1387911,96
F6	165	1653,98	7716,68	51,68	663,80	440631,06
F7	165	2152,53	9139,50	56,63	727,38	529076,80
F8	69	529,58	9757,92	15,08	125,27	15692,89

Tabela A.2: Descrições estatísticas dos resultados do *Praat* da vogal 'i' de um Baixo

	N	Gama	Média		Desvio P.	Variância
	Estatística	Estatística	Estatística	Erro	Estatística	Estatística
F1	165	262,71	1819,80	2,88	36,94	1364,80
F2	165	158,30	2787,71	2,88	37,05	1372,57
F3	165	241,11	3206,31	3,04	39,05	1524,95
F4	165	2173,19	6421,47	29,84	383,34	146951,68
F5	165	1491,48	7632,48	21,92	281,55	79268,51
F6	165	491,58	8259,75	8,86	113,79	12948,68
F7	165	305,68	10089,81	4,02	51,58	2660,78
F8	0	-	-	-	-	-

Tabela A.3: Descrições estatísticas dos resultados do *Praat* da vogal 'a' de um Tenor

	N	Gama	Média		Desvio P.	Variância
	Estatística	Estatística	Estatística	Erro	Estatística	Estatística
F1	165	819,73	1064,14	8,25	105,97	11229,52
F2	165	1771,00	2727,13	20,99	269,66	72717,91
F3	165	842,22	3413,21	11,15	143,22	20511,24
F4	165	3135,84	6104,04	41,03	527,05	277779,98
F5	165	863,07	6798,54	10,89	139,90	19571,70
F6	165	2125,19	8433,97	25,57	328,43	107864,36
F7	165	1829,00	10043,30	22,74	292,04	85289,53
F8	6	194,94	10169,69	34,30	84,01	7057,15

Tabela A.4: Descrições estatísticas dos resultados do *Praat* da vogal 'i' de um Tenor

	N	Gama	Média		Desvio P.	Variância
	Estatística	Estatística	Estatística	Erro	Estatística	Estatística
F1	165	2371,31	2122,70	17,79	228,57	52243,51
F2	165	630,90	2913,87	5,19	66,63	4439,60
F3	165	651,30	3420,26	5,50	70,69	4996,59
F4	165	3181,54	6151,54	24,07	309,20	95602,10
F5	165	1645,65	7459,70	13,93	178,90	32005,06
F6	165	1158,19	8502,70	10,19	130,87	17127,66
F7	165	2213,42	10219,88	16,84	216,37	46814,18
F8	2	109,21	10174,64	54,60	77,22	5963,30

Tabela A.5: Descrições estatísticas dos resultados do *Praat* da vogal 'a' de uma Contralto

	N	Gama	Média		Desvio P.	Variância
	Estatística	Estatística	Estatística	Erro	Estatística	Estatística
F1	165	350,63	1019,86	4,79	61,48	3780,01
F2	165	430,28	3083,53	8,16	104,82	10988,02
F3	165	435,12	3399,20	10,19	130,93	17142,78
F4	165	2113,68	4984,18	40,71	522,87	273391,94
F5	165	609,12	6421,92	7,74	99,37	9873,59
F6	165	1165,25	8512,76	19,41	249,36	62180,00
F7	163	1889,76	9704,46	41,60	531,16	282136,13
F8	0	-	-	-	-	-

Tabela A.6: Descrições estatísticas dos resultados do *Praat* da vogal 'i' de uma Contralto

	N	Gama	Média		Desvio P.	Variância
	Estatística	Estatística	Estatística	Erro	Estatística	Estatística
F1	165	1641,34	626,59	35,45	455,41	207398,85
F2	165	1501,83	1924,90	33,40	429,07	184105,00
F3	165	982,68	2914,38	22,04	283,14	80169,22
F4	165	2429,55	3897,34	61,55	790,59	625038,85
F5	165	2036,99	5848,27	50,04	642,78	413170,14
F6	165	2654,23	7356,52	50,10	643,55	414155,31
F7	160	3224,03	9037,36	40,52	512,57	262726,35
F8	124	1536,64	9613,92	28,20	314,05	98629,54

Tabela A.7: Descrições estatísticas dos resultados do *Praat* da vogal 'a' de uma Soprano

	N	Gama	Média		Desvio P.	Variância
	Estatística	Estatística	Estatística	Erro	Estatística	Estatística
F1	165	527,22	779,80	5,24	67,36	4536,72
F2	165	1854,82	1744,20	24,83	318,91	101706,44
F3	165	1351,35	3321,56	20,05	257,55	66331,87
F4	165	925,41	3953,84	20,90	268,53	72106,57
F5	165	3075,73	5180,66	55,96	718,79	516656,72
F6	165	2008,97	7027,30	26,08	334,99	112219,67
F7	165	1684,26	8740,06	18,66	239,74	57475,28
F8	153	1123,01	9434,63	21,85	270,23	73026,51

Tabela A.8: Descrições estatísticas dos resultados do *Praat* da vogal 'i' de uma Soprano

	N	Gama	Média		Desvio P.	Variância
	Estatística	Estatística	Estatística	Erro	Estatística	Estatística
F1	165	134,23	718,15	2,32	29,81	888,80
F2	165	343,44	2108,72	3,96	50,93	2593,44
F3	165	578,21	2992,28	11,33	145,48	21164,20
F4	165	1595,72	4879,92	22,28	286,14	81874,88
F5	165	1277,82	5952,16	19,75	253,66	64341,50
F6	165	1901,53	7317,65	20,38	261,76	68519,57
F7	165	1790,20	9460,03	30,49	391,61	153356,44
F8	78	1332,58	10289,10	26,32	232,48	54047,08

A.1.2 Métodos implementados

Tabela A.9: Descrições estatísticas dos resultados dos métodos implementados da vogal 'a' de um Baixo

		N	Gama	Média	Desvio P.	Variância	
		Estatística	Estatística	Estatística	Erro	Estatística	
	F0	167	8,38	110,88	0,14	1,77	0,14
LPC Sinal	F1	167	86,13	578,82	1,21	15,64	11,36
	F2	161	150,73	1071,58	2,60	32,97	50,47
	F3	167	150,73	2404,88	2,79	36,09	60,47
	F4	167	258,40	3086,98	3,32	42,95	85,67
	F5	158	495,26	6957,68	6,85	86,13	344,52
	F6	167	344,53	8294,80	4,55	58,80	160,55
	F7	74	559,86	9748,72	16,33	140,52	916,96
Cepstrum do sinal	F1	167	172,27	2611,45	2,40	30,97	44,54
	F2	2	64,60	7009,06	32,30	45,68	96,90
	F3	0	-	-	-	-	-
	F4	0	-	-	-	-	-
	F5	0	-	-	-	-	-
	F6	0	-	-	-	-	-
	F7	0	-	-	-	-	-
LPC do resíduo	F1	167	172,27	613,89	2,25	29,03	39,14
	F2	68	172,27	1135,24	5,34	44,00	89,89
	F3	167	193,80	2434,15	3,27	42,24	82,85
	F4	167	279,93	3118,70	3,64	47,07	102,91
	F5	159	602,93	6964,43	7,04	88,73	365,64
	F6	167	366,06	8297,37	4,52	58,45	158,63
	F7	78	602,93	9743,22	16,80	148,41	1022,81
Cepstrum do resíduo	F1	97	559,86	863,77	12,54	123,46	707,85
	F2	77	968,99	2794,28	28,93	253,85	2992,49
	F3	62	1098,19	5579,18	34,92	274,92	3510,06
	F4	29	646,00	5951,33	29,18	157,12	1146,51
	F5	65	236,87	7109,60	6,08	48,98	111,41
	F6	156	215,33	8273,86	3,59	44,78	93,11
	F7	152	150,73	9504,78	2,58	31,86	47,13

Tabela A.10: Descrições estatísticas dos resultados dos métodos implementados da vogal 'i' de um Baixo

		N	Gama	Média		Desvio P.	Variância
		Estatística	Estatística	Estatística	Erro	Estatística	Estatística
LPC Sinal	F0	167	5,89	109,38	0,10	1,31	0,08
	F1	167	43,07	303,79	1,05	13,58	8,57
	F2	167	107,67	1824,91	1,68	21,75	21,98
	F3	167	193,80	2783,59	2,89	37,40	64,97
	F4	162	193,80	3170,43	2,69	34,24	54,45
	F5	21	559,86	7538,67	32,59	149,33	1035,54
	F6	133	581,40	8211,76	9,52	109,74	559,27
	F7	139	366,06	10071,34	5,74	67,71	212,91
Cepstrum do sinal	F1	167	150,73	2771,34	2,29	29,57	40,61
	F2	1	0,00	6309,23	-	-	-
	F3	0	-	-	-	-	-
	F4	0	-	-	-	-	-
	F5	0	-	-	-	-	-
	F6	0	-	-	-	-	-
	F7	0	-	-	-	-	-
LPC do resíduo	F1	167	43,07	311,01	1,04	13,50	8,46
	F2	167	64,60	1837,03	1,57	20,32	19,17
	F3	167	172,27	2782,17	2,63	33,97	53,58
	F4	166	215,33	3178,22	2,73	35,16	57,42
	F5	61	1119,73	6959,46	33,65	262,84	3208,32
	F6	114	839,79	8239,85	13,27	141,68	932,15
	F7	145	430,66	10084,96	6,07	73,04	247,75
Cepstrum do resíduo	F1	152	430,66	774,06	6,63	81,70	309,96
	F2	160	129,20	2850,73	1,72	21,73	21,94
	F3	38	215,33	4852,90	9,38	57,83	155,30
	F4	160	301,46	6773,27	4,81	60,85	171,94
	F5	151	301,46	7892,99	3,79	46,63	100,97
	F6	127	387,60	8580,39	6,90	77,73	280,56
	F7	115	323,00	9973,99	6,12	65,65	200,18

Tabela A.11: Descrições estatísticas dos resultados dos métodos implementados da vogal 'a' de um Tenor

		N	Gama	Média		Desvio P.	Variância
		Estatística	Estatística	Estatística	Erro	Estatística	Estatística
LPC Sinal	F0	167	21,23	278,61	0,42	5,45	1,38
	F1	137	366,06	689,53	5,73	67,11	209,15
	F2	167	86,13	1114,70	1,72	22,17	22,82
	F3	167	172,27	2782,68	3,19	41,19	78,78
	F4	167	366,06	3352,35	5,81	75,04	261,53
	F5	119	839,79	6386,86	14,56	158,79	1170,90
	F6	167	452,20	8496,46	7,91	102,22	485,25
	F7	90	452,20	10155,06	9,98	94,65	416,05
<i>Cepstrum</i> do sinal	F1	134	495,26	961,60	9,56	110,69	568,96
	F2	167	193,80	3164,87	2,83	36,61	62,25
	F3	160	236,87	6107,76	4,25	53,70	133,90
	F4	1	0,00	8096,48	-	-	-
	F5	0	-	-	-	-	-
	F6	0	-	-	-	-	-
	F7	0	-	-	-	-	-
<i>LPC</i> do resíduo	F1	135	344,53	702,94	4,68	54,33	137,06
	F2	167	86,13	1105,67	1,84	23,75	26,18
	F3	167	172,27	2787,84	3,11	40,17	74,92
	F4	165	344,53	3411,12	5,38	69,08	221,63
	F5	124	689,06	6251,92	13,20	146,99	1003,44
	F6	107	882,86	6759,21	17,82	184,38	1578,80
	F7	166	387,60	8525,46	7,26	93,49	405,87
<i>Cepstrum</i> do resíduo	F1	167	215,33	970,93	3,54	45,76	97,24
	F2	167	193,80	2793,13	2,62	33,85	53,21
	F3	154	323,00	3693,92	5,74	71,23	235,62
	F4	67	882,86	6063,04	24,50	200,58	1868,31
	F5	87	689,06	6740,88	16,51	153,98	1101,15
	F6	157	344,53	8332,66	7,36	92,18	394,62
	F7	66	559,86	9125,51	15,72	127,73	757,65

Tabela A.12: Descrições estatísticas dos resultados dos métodos implementados da vogal 'i' de um Tenor

		N	Gama	Média		Desvio P.	Variação
		Estatística	Estatística	Estatística	Erro	Estatística	Estatística
LPC Sinal	F0	167	23,86	280,50	0,47	6,08	1,71
	F1	167	129,20	294,50	1,38	17,89	14,86
	F2	162	430,66	2076,09	6,90	87,78	357,81
	F3	167	236,87	2913,17	3,42	44,16	90,57
	F4	167	215,33	3403,79	3,22	41,59	80,31
	F5	131	775,20	6301,17	12,71	145,52	983,35
	F6	37	1033,59	7715,29	42,92	261,07	3165,22
	F7	29	667,53	10240,15	28,55	153,73	1097,56
Cepstrum do sinal	F1	20	516,80	483,42	30,71	137,34	876,00
	F2	167	193,80	3070,87	2,58	33,33	51,59
	F3	92	559,86	6463,00	12,37	118,61	653,37
	F4	0	-	-	-	-	-
	F5	0	-	-	-	-	-
	F6	0	-	-	-	-	-
	F7	0	-	-	-	-	-
LPC do resíduo	F1	167	129,20	290,38	1,74	22,53	23,57
	F2	163	323,00	2128,22	5,65	72,11	241,45
	F3	167	215,33	2923,74	3,23	41,78	81,05
	F4	167	215,33	3413,59	3,18	41,06	78,30
	F5	154	473,73	6182,27	7,04	87,40	354,71
	F6	63	732,13	7569,78	16,88	134,02	834,12
	F7	131	624,46	8477,67	9,08	103,91	501,46
Cepstrum do resíduo	F1	144	409,13	522,18	6,11	73,34	249,82
	F2	49	172,27	1810,55	6,27	43,92	89,58
	F3	165	323,00	3012,82	3,90	50,07	116,44
	F4	106	473,73	3916,81	8,55	88,01	359,68
	F5	167	430,66	8649,00	8,25	106,67	528,41
	F6	51	193,80	9945,81	5,65	40,32	75,49
	F7	0	0,00	0,00	0,00	0,00	0,00

Tabela A.13: Descrições estatísticas dos resultados dos métodos implementados da vogal 'a' de uma Contralto

		N	Gama	Média		Desvio P.	Variância
		Estatística	Estatística	Estatística	Erro	Estatística	Estatística
LPC Sinal	F0	167	28,09	453,37	0,59	7,63	2,70
	F1	167	86,13	510,22	1,65	21,37	21,20
	F2	167	452,20	1179,04	6,64	85,86	342,32
	F3	167	215,33	3170,41	4,45	57,55	153,83
	F4	80	193,80	3517,45	4,71	42,12	82,37
	F5	24	366,06	5436,24	21,06	103,17	494,29
	F6	130	473,73	6405,63	8,32	94,92	418,38
	F7	38	710,60	8846,18	30,55	188,31	1646,83
<i>Cepstrum</i> do sinal	F1	113	387,60	921,16	10,35	110,05	562,45
	F2	29	344,53	1602,37	14,75	79,43	293,03
	F3	167	150,73	3232,30	2,64	34,17	54,23
	F4	167	279,93	6465,12	5,08	65,68	200,34
	F5	119	452,20	8115,48	8,55	93,32	404,45
	F6	114	430,66	9034,69	8,56	91,45	388,35
	F7	0	-	-	-	-	-
<i>LPC</i> do resíduo	F1	167	129,20	530,98	2,05	26,51	32,63
	F2	147	473,73	1193,55	8,54	103,58	498,21
	F3	167	172,27	3174,92	4,08	52,67	128,84
	F4	98	366,06	3611,65	9,39	93,01	401,71
	F5	104	667,53	6454,58	11,19	114,09	604,50
	F6	111	646,00	8691,07	12,40	130,66	792,88
	F7	15	301,46	9092,75	26,79	103,74	499,78
<i>Cepstrum</i> do resíduo	F1	127	452,20	690,59	8,86	99,83	462,85
	F2	70	882,86	1274,46	31,73	265,49	3273,23
	F3	42	258,40	2638,84	8,60	55,71	144,12
	F4	167	150,73	3283,75	2,31	29,79	41,22
	F5	167	172,27	6428,37	2,13	27,54	35,23
	F6	165	236,87	8325,13	3,08	39,54	72,60
	F7	159	150,73	9292,73	2,58	32,47	48,97

Tabela A.14: Descrições estatísticas dos resultados dos métodos implementados da vogal 'i' de uma Contralto

		N	Gama	Média		Desvio P.	Variância
		Estatística	Estatística	Estatística	Erro	Estatística	Estatística
LPC Sinal	F0	167	31,27	445,84	0,65	8,37	3,25
	F1	167	43,07	462,00	1,05	13,52	8,49
	F2	167	366,06	1658,31	7,71	99,61	460,76
	F3	167	258,40	2750,19	3,78	48,83	110,75
	F4	167	344,53	3489,80	6,95	89,85	374,92
	F5	71	882,86	5497,94	23,72	199,85	1854,79
	F6	78	646,00	7014,30	18,53	163,68	1244,14
	F7	131	646,00	9306,95	10,97	125,59	732,46
<i>Cepstrum</i> do sinal	F1	26	516,80	500,23	26,26	133,90	832,57
	F2	108	689,06	1614,79	16,37	170,10	1343,71
	F3	167	279,93	3250,35	5,94	76,71	273,25
	F4	166	581,40	7183,01	9,91	127,67	756,99
	F5	1	0,00	7579,69	-	-	-
	F6	0	-	-	-	-	-
	F7	0	-	-	-	-	-
<i>LPC</i> do resíduo	F1	167	43,07	463,80	1,14	14,71	10,05
	F2	167	366,06	1676,37	7,22	93,28	404,11
	F3	165	279,93	2779,74	4,47	57,41	153,05
	F4	167	387,60	3509,14	7,16	92,51	397,43
	F5	132	646,00	5437,30	11,48	131,87	807,58
	F6	110	559,86	6991,64	13,26	139,05	897,97
	F7	121	1141,26	9226,35	26,17	287,92	3849,80
<i>Cepstrum</i> do resíduo	F1	167	215,33	525,69	3,17	40,96	77,89
	F2	159	215,33	1634,09	4,32	54,50	137,92
	F3	143	344,53	2806,39	4,86	58,15	157,03
	F4	157	452,20	3526,23	8,89	111,39	576,26
	F5	29	172,27	4747,70	7,14	38,44	68,63
	F6	85	667,53	5564,43	17,57	162,00	1218,83
	F7	82	667,53	6804,23	14,08	127,48	754,72

Tabela A.15: Descrições estatísticas dos resultados dos métodos implementados da vogal 'a' de uma Soprano

		N	Gama	Média		Desvio P.	Variância
		Estatística	Estatística	Estatística	Erro	Estatística	Estatística
LPC Sinal	F0	167	50,78	701,94	1,15	14,91	10,32
	F1	167	43,07	717,17	1,43	18,42	15,75
	F2	167	86,13	1441,95	1,89	24,44	27,73
	F3	167	387,60	3521,52	8,92	115,30	617,39
	F4	89	818,26	4348,01	19,76	186,38	1613,14
	F5	2	64,60	5049,54	32,30	45,68	96,90
	F6	166	430,66	6947,70	8,84	113,92	602,63
	F7	75	882,86	8959,25	24,82	214,93	2145,37
<i>Cepstrum</i> do sinal	F1	165	323,00	650,30	3,93	50,51	118,49
	F2	93	1076,66	1422,58	25,66	247,50	2844,63
	F3	157	495,26	3348,07	5,53	69,29	222,95
	F4	120	495,26	4202,03	10,76	117,92	645,74
	F5	27	409,13	5271,65	19,91	103,45	496,98
	F6	152	409,13	6951,12	7,20	88,83	366,42
	F7	142	387,60	8839,99	6,36	75,85	267,16
<i>LPC</i> do resíduo	F1	167	64,60	719,11	1,44	18,60	16,07
	F2	165	150,73	1460,08	2,73	35,02	56,96
	F3	155	646,00	3435,31	9,48	118,05	647,19
	F4	90	559,86	3976,94	14,17	134,40	838,90
	F5	49	1335,06	4542,63	51,85	362,94	6117,19
	F6	166	452,20	6949,26	8,82	113,61	599,44
	F7	60	538,33	8677,88	14,13	109,44	556,21
<i>Cepstrum</i> do resíduo	F1	160	387,60	676,14	4,72	59,75	165,82
	F2	108	624,46	1426,77	9,97	103,60	498,43
	F3	167	366,06	3329,78	4,04	52,18	126,44
	F4	148	495,26	4142,38	7,39	89,87	375,08
	F5	25	559,86	5285,11	24,40	121,98	691,00
	F6	165	430,66	6910,72	7,26	93,23	403,65
	F7	147	344,53	8863,62	5,58	67,64	212,44

Tabela A.16: Descrições estatísticas dos resultados dos métodos implementados da vogal 'i' de uma Soprano

		N	Gama	Média		Desvio P.	Variação
		Estatística	Estatística	Estatística	Erro	Estatística	Estatística
LPC Sinal	F0	166	61,09	697,30	1,28	16,46	12,58
	F1	167	86,13	711,63	1,79	23,14	24,86
	F2	162	301,46	2059,88	3,70	47,04	102,76
	F3	100	753,66	2903,54	14,35	143,48	956,04
	F4	4	236,87	4021,33	52,29	104,57	507,82
	F5	81	968,99	5618,31	35,45	319,05	4727,18
	F6	95	710,60	7304,52	16,69	162,64	1228,39
	F7	83	968,99	9883,48	27,68	252,16	2952,93
<i>Cepstrum</i> do sinal	F1	164	279,93	660,05	3,33	42,64	84,44
	F2	115	366,06	1997,91	5,02	53,80	134,41
	F3	132	430,66	2867,83	7,36	84,59	332,31
	F4	81	624,46	3678,19	15,73	141,58	930,83
	F5	117	495,26	5025,33	8,47	91,61	389,72
	F6	72	861,33	5889,03	30,11	255,50	3031,63
	F7	84	581,40	7480,99	15,67	143,62	957,95
<i>LPC</i> do resíduo	F1	166	86,13	710,86	1,74	22,36	23,23
	F2	166	301,46	2077,82	3,34	42,98	85,81
	F3	118	581,40	2908,99	9,96	108,20	543,66
	F4	8	495,26	4026,71	56,68	160,32	1193,55
	F5	68	559,86	5844,68	16,38	135,08	847,39
	F6	41	839,79	7376,44	32,12	205,64	1963,88
	F7	99	818,26	9769,11	19,42	193,23	1734,01
<i>Cepstrum</i> do resíduo	F1	166	193,80	695,94	2,70	34,76	56,13
	F2	86	559,86	1967,28	8,08	74,95	260,89
	F3	148	473,73	2897,53	7,74	94,19	411,98
	F4	57	581,40	3739,60	17,84	134,68	842,36
	F5	62	646,00	5711,51	17,17	135,16	848,40
	F6	27	409,13	5978,26	19,44	101,03	473,98
	F7	108	581,40	7520,27	9,92	103,08	493,40

A.2 Descrições estatísticas dos testes com vozes sintéticas

A.2.1 Praat

Tabela A.17: Descrições estatísticas dos resultados do *Praat* da vogal sintética 'i' com frequência fundamental de 220 Hz

	N	Gama	Média		Desvio P.	Variância
	Estatística	Estatística	Estatística	Erro	Estatística	Estatística
F1	168	323,00	355,0833	3,27858	42,49529	1805,849
F2	168	1732,61	929,1932	62,03089	804,01221	646435,632
F3	168	2321,71	1319,8688	87,96920	1140,21115	1300081,466
F4	168	3084,92	1767,2969	117,78983	1526,73071	2330906,661
F5	168	3589,45	2044,3286	136,20462	1765,41370	3116685,517

Tabela A.18: Descrições estatísticas dos resultados do *Praat* da vogal sintética 'i' com frequência fundamental de 440 Hz

	N	Gama	Média		Desvio P.	Variância
	Estatística	Estatística	Estatística	Erro	Estatística	Estatística
F1	168	501,92	524,8324	5,05584	65,53118	4294,336
F2	168	2133,76	2173,3971	21,83274	282,98462	80080,295
F3	168	2960,23	3020,2810	30,65341	397,31364	157858,131
F4	168	3375,13	3455,6324	34,95574	453,07816	205279,816
F5	168	3851,95	3718,0293	37,91553	491,44138	241514,626

Tabela A.19: Descrições estatísticas dos resultados do *Praat* da vogal sintética 'i' com frequência fundamental de 880 Hz

	N	Gama	Média		Desvio P.	Variância
	Estatística	Estatística	Estatística	Erro	Estatística	Estatística
F1	168	851,46	873,6268	8,68800	112,60929	12680,852
F2	168	1702,05	1745,9404	17,37267	225,17548	50703,997
F3	168	2530,17	2604,8221	25,89595	335,64981	112660,796
F4	168	3364,49	3468,1111	34,47238	446,81311	199641,952
F5	168	4195,74	4326,7431	42,96551	556,89661	310133,838

Tabela A.20: Descrições estatísticas dos resultados do *Praat* da vogal sintética 'a' com frequência fundamental de 220 Hz

	N	Gama	Média		Desvio P.	Variância
	Estatística	Estatística	Estatística	Erro	Estatística	Estatística
F1	168	624,30	645,6236	7,38431	95,71165	9160,719
F2	168	1106,45	1129,5053	11,96135	155,03688	24036,434
F3	168	2703,34	2568,5369	37,82228	490,23280	240328,203
F4	168	3532,01	3317,2387	44,48303	576,56598	332428,332
F5	168	4462,74	4232,9157	55,13897	714,68273	510771,411

Tabela A.21: Descrições estatísticas dos resultados do *Praat* da vogal sintética 'a' com frequência fundamental de 440 Hz

	N	Gama	Média		Desvio P.	Variância
	Estatística	Estatística	Estatística	Erro	Estatística	Estatística
F1	168	808,68	796,7400	11,54894	149,69141	22407,520
F2	168	1241,29	1159,5973	17,43855	226,02945	51089,311
F3	168	2590,26	2011,5240	56,58612	733,43988	537934,064
F4	168	3448,53	3110,6282	45,41554	588,65262	346511,907
F5	168	4461,54	4057,0331	56,25184	729,10723	531597,350

Tabela A.22: Descrições estatísticas dos resultados do *Praat* da vogal sintética 'a' com frequência fundamental de 880 Hz

	N	Gama	Média		Desvio P.	Variância
	Estatística	Estatística	Estatística	Erro	Estatística	Estatística
F1	168	840,81	866,6777	8,61531	111,66720	12469,564
F2	168	1674,77	1722,1209	17,12402	221,95263	49262,971
F3	168	2531,39	2603,0208	25,87743	335,40984	112499,758
F4	168	3379,16	3472,9903	34,52386	447,48043	200238,734
F5	168	4311,67	4397,4639	43,73263	566,83970	321307,245

A.2.2 Métodos implementados

Tabela A.23: Descrições estatísticas dos resultados dos métodos implementados da vogal sintética 'i' com frequência fundamental de 220 Hz

		N	Gama	Média		Desvio P.	Variância
		Estatística	Estatística	Estatística	Erro	Estatística	Estatística
LPC do sinal	F1	168	21,53	342,7368	,46054	5,96926	1,655
	F2	168	43,07	2255,4749	,79242	10,27089	4,899
	F3	168	21,53	3087,3230	,80669	10,45588	5,077
	F4	168	107,67	3505,5542	1,58036	20,48378	19,485
	F5	168	215,33	4634,5093	2,13345	27,65261	35,511
LPC do resíduo	F1	168	21,53	258,9111	,25403	3,29267	,503
	F2	168	43,07	2329,6875	,83220	10,78654	5,403
	F3	168	43,07	3093,2190	,83591	10,83469	5,452
	F4	166	193,80	3488,5086	2,61193	33,65239	52,592
	F5	168	193,80	4603,2349	2,80303	36,33140	61,299
Cepstrum do sinal	F1	53	409,13	320,1540	13,11326	95,46594	423,242
	F2	108	279,93	440,4338	5,10997	53,10434	130,964
	F3	153	150,73	2002,8694	2,05581	25,42892	30,029
	F4	168	172,27	3179,0955	2,98975	38,75157	69,738
	F5	26	301,46	7169,7284	12,38318	63,14210	185,152
Cepstrum do resíduo	F1	148	301,46	490,6078	4,34811	52,89707	129,944
	F2	28	301,46	1446,5698	15,58670	82,47704	315,906
	F3	149	323,00	2472,1273	4,54716	55,50513	143,073
	F4	168	193,80	3351,3611	2,81904	36,53892	62,002
	F5	159	279,93	4477,8228	3,82720	48,25921	108,156

Tabela A.24: Descrições estatísticas dos resultados dos métodos implementados da vogal sintética 'a' com frequência fundamental de 220 Hz

		N	Gama	Média	Desvio P.	Variância
		Estatística	Estatística	Estatística	Erro	Estatística
LPC do sinal	F1	168	,00	667,5293	,00000	,00000
	F2	168	21,53	1119,3420	,22067	2,86021
	F3	168	279,93	2753,4302	3,81171	49,40538
	F4	159	409,13	3626,6519	5,12446	64,61704
	F5	79	193,80	4671,0696	4,93418	43,85592
LPC do resíduo	F1	168	,00	667,5293	,00000	,00000
	F2	168	21,53	1121,5210	,46054	5,96926
	F3	168	193,80	2803,8025	2,68345	34,78150
	F4	149	344,53	3674,5183	4,68560	57,19509
	F5	154	581,40	4724,0212	6,90724	85,71656
Cepstrum do sinal	F1	168	107,67	791,2170	1,56393	20,27088
	F2	32	538,33	2499,8703	27,96174	158,17551
	F3	101	689,06	2932,5664	13,09187	131,57167
	F4	61	538,33	3115,9604	11,72251	91,55573
	F5	159	409,13	4682,4560	5,47723	69,06527
Cepstrum do resíduo	F1	168	193,80	731,1035	2,73201	35,41086
	F2	131	236,87	1624,0309	4,13475	47,32440
	F3	68	107,67	2714,1336	3,64950	30,09452
	F4	167	258,40	3609,7127	3,56704	46,09636
	F5	168	172,27	4596,4417	2,41094	31,24940

Tabela A.25: Descrições estatísticas dos resultados dos métodos implementados da vogal sintética 'i' com frequência fundamental de 440 Hz

		N	Gama	Média		Desvio P.	Variância
		Estatística	Estatística	Estatística	Erro	Estatística	Estatística
LPC do sinal	F1	168	,00	473,7305	,00000	,00000	,000
	F2	168	,00	2217,9199	,00000	,00000	,000
	F3	168	,00	3100,7813	,00000	,00000	,000
	F4	168	21,53	3536,0596	,68372	8,86205	3,647
	F5	168	86,13	4585,1624	1,19413	15,47764	11,125
LPC do resíduo	F1	168	,00	473,7305	,00000	,00000	,000
	F2	168	,00	2217,9199	,00000	,00000	,000
	F3	168	,00	3100,7813	,00000	,00000	,000
	F4	168	21,53	3539,3921	,80406	10,42184	5,044
	F5	168	64,60	4707,0557	1,26272	16,36668	12,440
<i>Cepstrum</i> do sinal	F1	168	215,33	528,7170	3,42917	44,44709	91,744
	F2	141	64,60	2040,7673	1,39408	16,55374	12,726
	F3	2	21,53	2293,2861	10,76660	15,22627	10,767
	F4	168	129,20	3216,7786	1,54189	19,98514	18,548
	F5	147	64,60	4662,0117	1,25367	15,20000	10,729
<i>Cepstrum</i> do resíduo	F1	168	172,27	566,7847	2,07548	26,90130	33,608
	F2	162	430,66	2469,9382	5,77777	73,53901	251,146
	F3	168	129,20	3336,4929	2,02663	26,26813	32,044
	F4	167	323,00	4493,9924	4,07136	52,61360	128,555
	F5	53	258,40	7706,4490	8,00966	58,31123	157,905

Tabela A.26: Descrições estatísticas dos resultados dos métodos implementados da vogal sintética 'a' com frequência fundamental de 440 Hz

		N	Gama	Média		Desvio P.	Variância
		Estatística	Estatística	Estatística	Erro	Estatística	Estatística
LPC do sinal	F1	168	,00	818,2617	,00000	,00000	,000
	F2	168	,00	1184,3262	,00000	,00000	,000
	F3	168	64,60	2639,4836	,88129	11,42278	6,059
	F4	168	129,20	3523,6267	1,63045	21,13307	20,740
	F5	167	279,93	4647,9483	3,59167	46,41457	100,046
LPC do resíduo	F1	168	21,53	820,3125	,48913	6,33983	1,867
	F2	168	86,13	1191,5039	1,73115	22,43826	23,381
	F3	168	129,20	2651,4038	1,31806	17,08406	13,554
	F4	168	193,80	3563,1042	2,81611	36,50091	61,873
	F5	168	236,87	4552,8625	3,05111	39,54690	72,630
Cepstrum do sinal	F1	168	64,60	878,3752	1,08863	14,11027	9,246
	F2	165	323,00	2953,3114	4,86843	62,53615	181,616
	F3	2	172,27	3660,6445	86,13281	121,81019	689,063
	F4	167	344,53	4669,3526	4,24222	54,82152	139,570
	F5	8	193,80	7649,6704	23,64104	66,86695	207,642
Cepstrum do resíduo	F1	168	236,87	843,7683	3,02391	39,19429	71,341
	F2	162	236,87	1601,8311	3,36914	42,88211	85,397
	F3	168	150,73	2711,7737	2,02566	26,25555	32,014
	F4	167	215,33	3627,7645	3,48860	45,08266	94,387
	F5	168	172,27	4584,3933	1,65274	21,42201	21,311

Tabela A.27: Descrições estatísticas dos resultados dos métodos implementados da vogal sintética 'i' com frequência fundamental de 880 Hz

		N	Gama	Média		Desvio P.	Variância
		Estatística	Estatística	Estatística	Erro	Estatística	Estatística
LPC do sinal	F1	168	,00	861,3281	,00000	,00000	,000
	F2	168	,00	1765,7227	,00000	,00000	,000
	F3	168	,00	2670,1172	,00000	,00000	,000
	F4	168	,00	3531,4453	,00000	,00000	,000
	F5	168	,00	4414,3066	,00000	,00000	,000
LPC do resíduo	F1	168	,00	839,7949	,00000	,00000	,000
	F2	168	21,53	1783,0261	,66201	8,58064	3,419
	F3	168	,00	2670,1172	,00000	,00000	,000
	F4	168	,00	3531,4453	,00000	,00000	,000
	F5	168	,00	4435,8398	,00000	,00000	,000
<i>Cepstrum</i> do sinal	F1	168	86,13	696,4966	1,32343	17,15357	13,665
	F2	153	86,13	2035,3803	1,11335	13,77137	8,807
	F3	168	64,60	3256,6406	1,28863	16,70252	12,956
	F4	168	86,13	4671,6797	1,35820	17,60433	14,392
	F5	33	172,27	7135,9730	7,89288	45,34113	95,472
<i>Cepstrum</i> do resíduo	F1	168	301,46	256,8604	3,89327	50,46252	118,258
	F2	70	430,66	2100,1025	10,18912	85,24833	337,492
	F3	77	516,80	2333,4162	11,86862	104,14670	503,712
	F4	163	495,26	3356,4055	5,57860	71,22280	235,575
	F5	153	279,93	4533,6541	3,80977	47,12428	103,129

Tabela A.28: Descrições estatísticas dos resultados dos métodos implementados da vogal sintética 'a' com frequência fundamental de 880 Hz

		N	Gama	Média		Desvio P.	Variância
		Estatística	Estatística	Estatística	Erro	Estatística	Estatística
LPC do sinal	F1	168	,00	882,8613	,00000	,00000	,000
	F2	168	,00	1744,1895	,00000	,00000	,000
	F3	168	,00	2648,5840	,00000	,00000	,000
	F4	168	21,53	3537,9822	,76616	9,93056	4,580
	F5	168	64,60	4500,4395	1,13538	14,71627	10,057
LPC do resíduo	F1	168	,00	882,8613	,00000	,00000	,000
	F2	168	43,07	1744,9585	,65281	8,46136	3,325
	F3	168	21,53	2659,9915	,83167	10,77964	5,396
	F4	168	43,07	3552,8503	,56028	7,26205	2,449
	F5	168	64,60	4483,5205	1,16616	15,11516	10,610
Cepstrum do sinal	F1	168	64,60	934,1309	,90557	11,73747	6,398
	F2	104	366,06	2459,9614	7,77625	79,30247	292,055
	F3	167	301,46	3046,1101	4,48289	57,93166	155,856
	F4	168	193,80	4702,4414	3,05744	39,62894	72,932
	F5	11	323,00	7483,7669	36,30740	120,41803	673,402
Cepstrum do resíduo	F1	167	323,00	928,2487	4,02390	52,00028	125,575
	F2	70	452,20	1715,2734	11,88653	99,44988	459,304
	F3	92	172,27	2743,6109	3,63295	34,84608	56,390
	F4	168	236,87	3637,5732	3,35517	43,48801	87,827
	F5	168	193,80	4586,4441	2,26711	29,38505	40,100

A.2.3 Gráficos

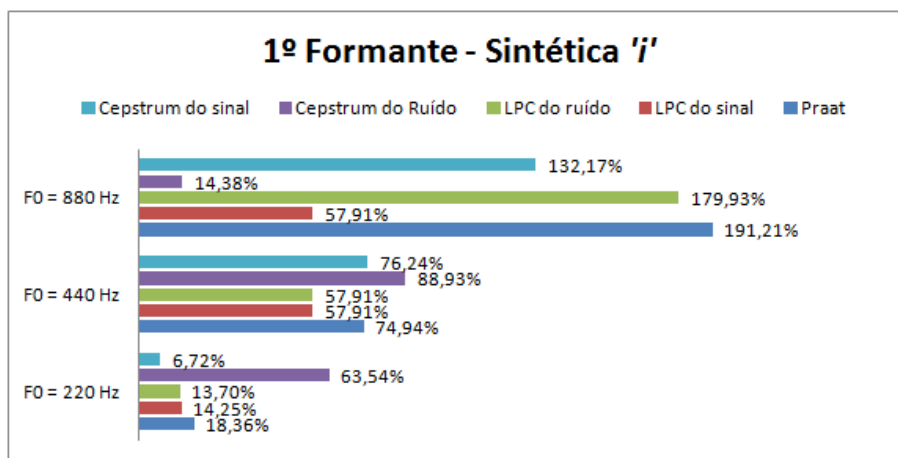


Figura A.1: Estatísticas relativos do 1º formante da vogal sintética 'i'

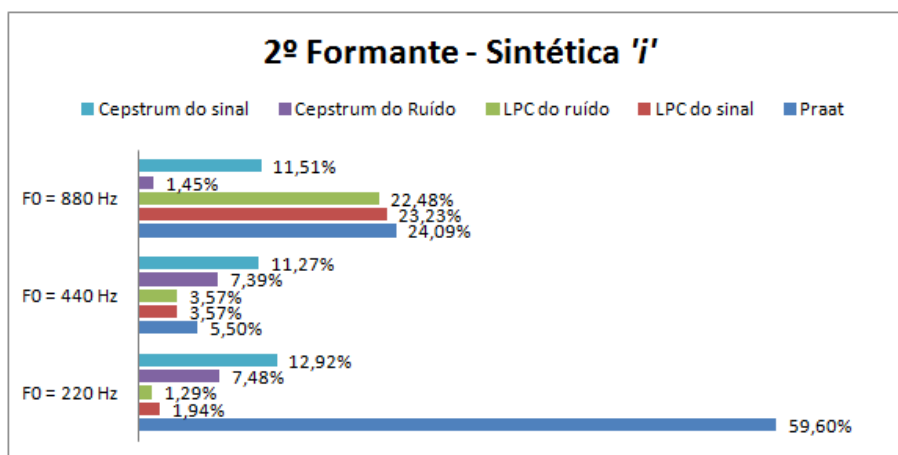


Figura A.2: Estatísticas relativos do 2º formante da vogal sintética 'i'

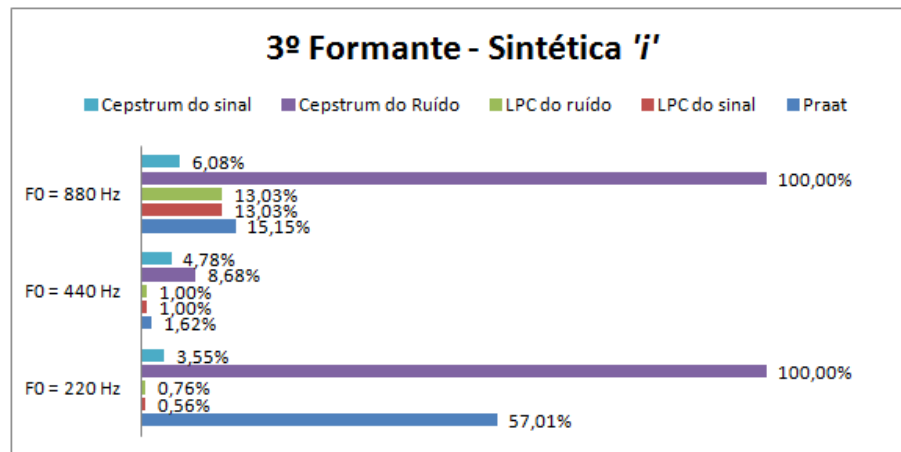


Figura A.3: Estatísticas relativos do 3º formante da vogal sintética 'i'

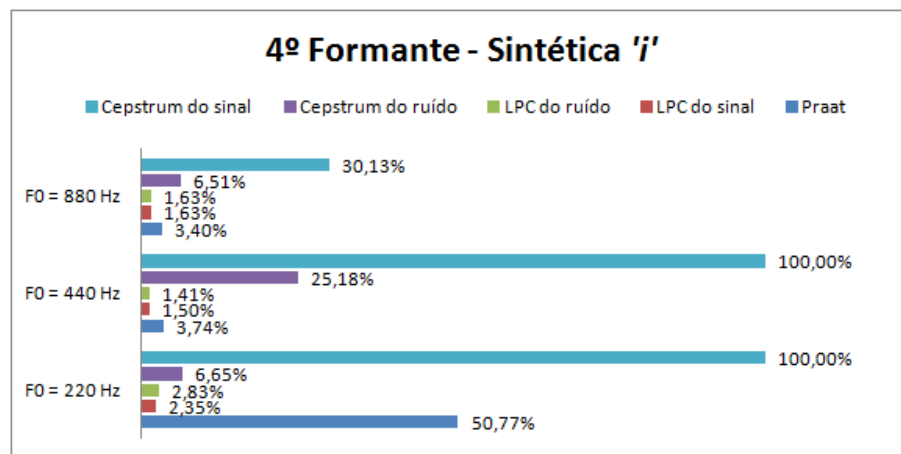


Figura A.4: Estatísticas relativos do 4º formante da vogal sintética 'i'

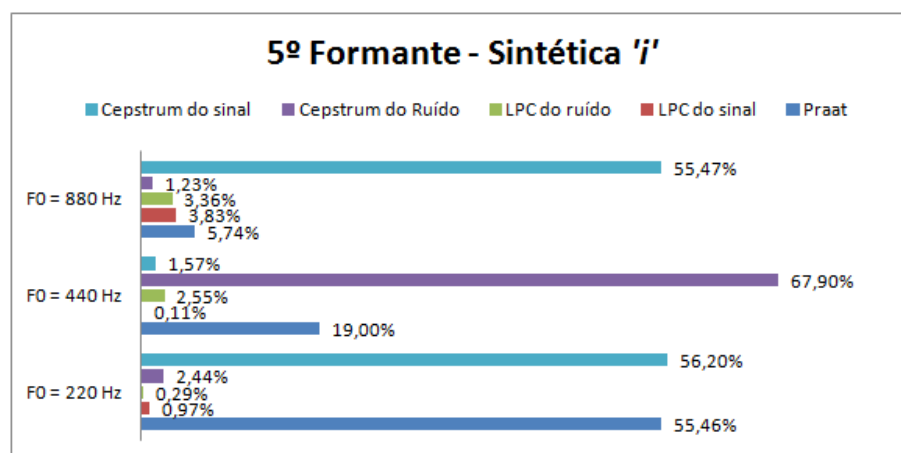


Figura A.5: Estatísticas relativos do 5º formante da vogal sintética 'i'

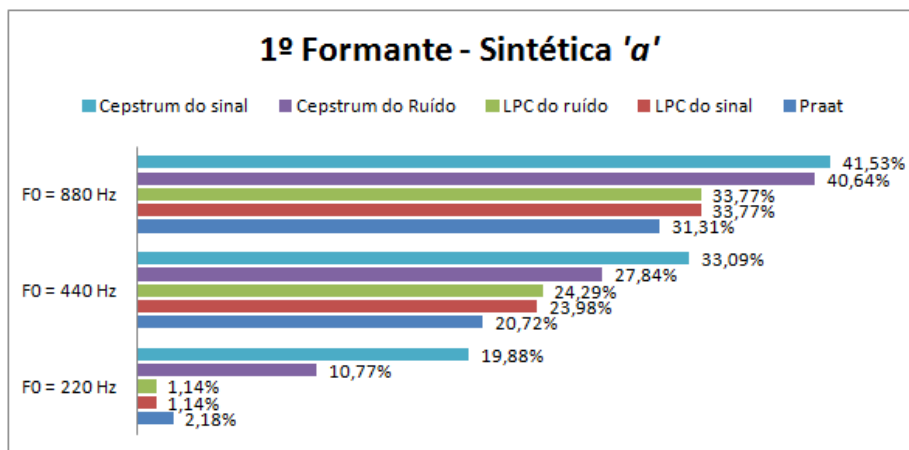


Figura A.6: Estatísticas relativos do 1º formante da vogal sintética 'i'

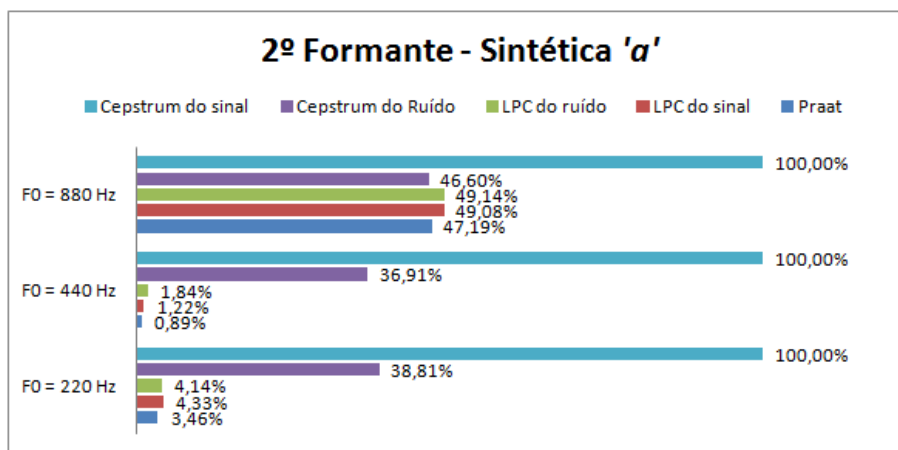


Figura A.7: Estatísticas relativos do 2º formante da vogal sintética 'i'

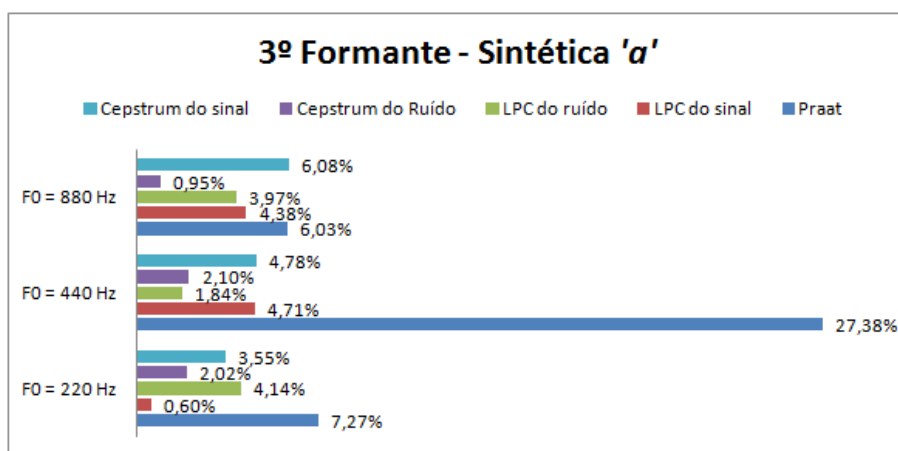


Figura A.8: Estatísticas relativos do 3º formante da vogal sintética 'i'

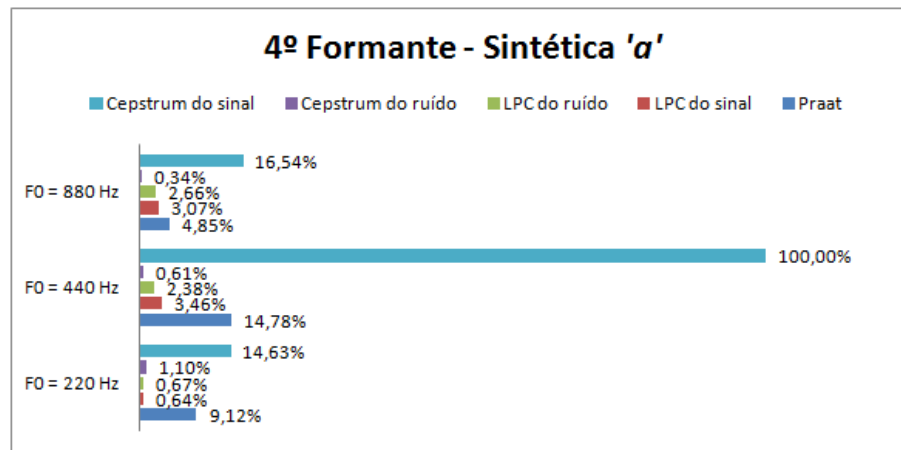


Figura A.9: Estatísticas relativos do 4º formante da vogal sintética 'i'

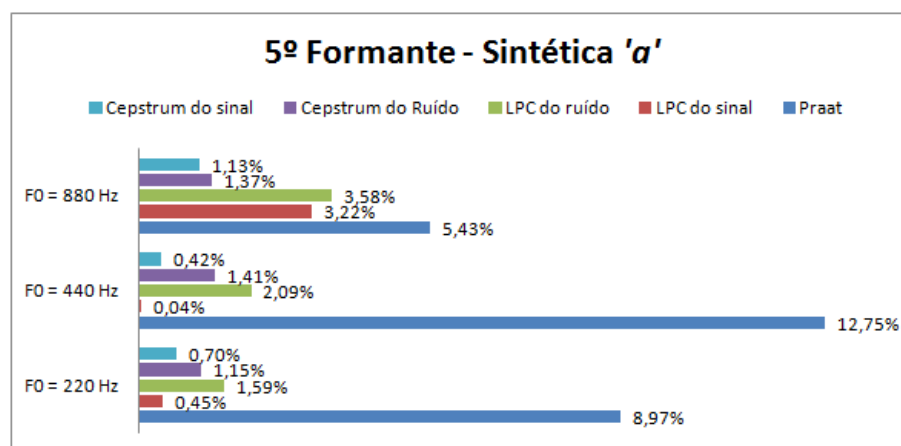


Figura A.10: Estatísticas relativos do 5º formante da vogal sintética 'i'

A.2.4 Relação de proximidade aos harmônicos

Tabela A.29: Relações dos resultados estimados com os harmônicos teóricos da vogal sintética 'i' com frequência fundamental de 220 Hz

	Vogal sintética 'i' com F0 = 220 Hz				
	H1	H2	H3	H4	H5
<i>Praat</i>	59,65%	47,20%	50,00%	49,79%	53,54%
LPC do sinal	60,21%	27,74%	15,63%	0,73%	4,99%
LPC do resíduo	70,58%	32,37%	17,17%	0,89%	4,62%
<i>Cepstrum</i> do sinal	63,62%	74,98%	24,13%	9,68%	62,95%
<i>Cepstrum</i> do resíduo	44,25%	17,81%	6,36%	4,79%	1,77%

Tabela A.30: Relações dos resultados estimados com os harmônicos teóricos da vogal sintética 'a' com frequência fundamental de 220 Hz

	Vogal sintética 'a' com F0 = 220 Hz				
	H1	H2	H3	H4	H5
<i>Praat</i>	26,63%	35,82%	2,71%	5,76%	3,80%
LPC do sinal	24,14%	36,40%	4,30%	3,03%	6,16%
LPC do resíduo	24,14%	36,28%	6,20%	4,39%	0,07364
<i>Cepstrum</i> do sinal	10,09%	42,04%	11,08%	11,48%	6,42%
<i>Cepstrum</i> do resíduo	16,92%	7,73%	2,81%	2,55%	4,46%

Tabela A.31: Relações dos resultados estimados com os harmônicos teóricos da vogal sintética 'i' com frequência fundamental de 440 Hz

	Vogal sintética 'i' com F0 = 440 Hz				
	H1	H2	H3	H4	H5
<i>Praat</i>	40,36%	23,49%	14,40%	1,83%	15,50%
LPC do sinal	60,21%	27,74%	15,63%	0,73%	4,99%
LPC do resíduo	46,17%	26,02%	17,45%	0,55%	6,98%
<i>Cepstrum</i> do sinal	39,92%	15,95%	13,13%	8,61%	5,95%
<i>Cepstrum</i> do resíduo	35,59%	40,34%	26,38%	27,67%	75,15%

Tabela A.32: Relações dos resultados estimados com os harmónicos teóricos da vogal sintética 'a' com frequência fundamental de 440 Hz

	Vogal sintética 'a' com F0 = 440 Hz				
	H1	H2	H3	H4	H5
<i>Praat</i>	9,46%	34,11%	23,81%	11,63%	7,79%
LPC do sinal	7,02%	32,71%	0,02%	0,10%	5,64%
LPC do resíduo	6,78%	32,30%	0,43%	1,22%	3,47%
<i>Cepstrum</i> do sinal	0,18%	67,80%	38,66%	32,65%	73,86%
<i>Cepstrum</i> do resíduo	4,12%	8,99%	2,72%	3,06%	4,19%

Tabela A.33: Relações dos resultados estimados com os harmónicos teóricos da vogal sintética 'i' com frequência fundamental de 880 Hz

	Vogal sintética 'i' com F0 = 880 Hz				
	H1	H2	H3	H4	H5
<i>Praat</i>	2,58%	1,76%	0,49%	0,24%	0,33%
LPC do sinal	60,21%	27,74%	15,63%	0,73%	4,99%
LPC do resíduo	4,57%	1,46%	1,14%	0,33%	0,81%
<i>Cepstrum</i> do sinal	65,48%	46,94%	33,31%	0,59%	0,93%
<i>Cepstrum</i> do resíduo	68,85%	27,44%	25,23%	100,00%	2,50%

Tabela A.34: Relações dos resultados estimados com os harmónicos teóricos da vogal sintética 'a' com frequência fundamental de 880 Hz

	Vogal sintética 'a' com F0 = 880 Hz				
	H1	H2	H3	H4	H5
<i>Praat</i>	0,21%	0,47%	0,59%	0,59%	2,56%
LPC do sinal	0,33%	0,90%	0,33%	0,51%	2,28%
LPC do resíduo	0,33%	0,61%	0,48%	0,59%	2,77%
<i>Cepstrum</i> do sinal	1,18%	100,00%	3,50%	1,99%	2,95%
<i>Cepstrum</i> do resíduo	5,66%	5,76%	100,00%	2,85%	4,68%

Anexo B

Anexo B - Equipamentos utilizados nas gravações

B.1 Microfone

Acoustical operating principle	Pressure gradient transducer
Directional pattern	Cardioid
Frequency range	20 Hz ... 20 kHz
Sensitivity at 1 kHz into 1 kohm	23 mV/Pa
Rated impedance	50 ohms
Rated load impedance	1 kohms
Equivalent noise level, CCIR ¹⁾	17.5 dB
Equivalent noise level, A-weighted ¹⁾	7 dB-A
Signal-to-noise ratio, CCIR ¹⁾ (rel. 94 dB SPL)	76.5 dB
Signal-to-noise ratio, A-weighted ¹⁾ (rel. 94 dB SPL)	87 dB
Maximum SPL for THD 0.5% ²⁾	138 dB
Maximum output voltage	13 dBu
Dynamic range of the microphone amplifier (A-weighted)	131 dB
Supply voltage (P48, IEC 61938)	48 V $\pm$ 4 V
Current consumption (P48, IEC 61938)	3 mA
Matching connector	XLR3F
Weight	approx. 450 g
Diameter	60 mm
Length	132 mm

1) according to IEC 60268-1; CCIR-weighting according to CCIR 468-3, quasi peak; A-weighting according to IEC 61672-1, RMS
2) measured as equivalent el. input signal

Figura B.1: Especificações do microfone utilizado nas gravações [41]

B.2 Pré-amplificador

Onyx 800R Specifications

Frequency Response

Mic Input to Line Output (Gain @ unity):
+0, -0.1 dB, 20 Hz to 30 kHz
+0, -3 dB, 10 Hz to 170 kHz

Mic Input to Digital Output (AES, 192 kHz sample rate):
+0, -0.2 dB, 20 Hz to 85 kHz
+0, -1 dB, 10 Hz to 90 kHz

Maximum Input Levels

Mic Input: +22 dBu, Gain @ unity
Inst Input: +21 dBu, Gain @ -20 dB
Line Input: > +22 dBu, Gain @ -20 dB

Input Impedance

Ch 1 and 2 Mic Input:
300Ω, 500Ω, 1.3 kΩ, or 2.4 kΩ balanced
Ch 3 through 8 Mic Input:
2.4 kΩ balanced
Inst Input: 1 MΩ
Line: 20 kΩ balanced, 10 kΩ unbalanced

Noise

Signal-to-Noise (A-weighted):
>103 dB (ref. +4 dBu, Mic In to Line Out, Gain @ unity)
Equivalent Input Noise (E.I.N.), 20 Hz to 20 kHz Bandwidth,
150Ω source impedance:
-129 dBu @ +60 dB gain
Residual Output Noise:
Line Out: < -102 dBu (Channel Gain at unity)
Digital Out (AES, 48 kHz): < -113 dB FS

Common Mode Rejection Ratio (CMRR)

Mic In: >60 dB @ 1 kHz, Gain @ maximum

Crosstalk

Mic Input to Line Output:
< -100 dB @ 1 kHz, +10 dBu signal on adjacent
input, 150Ω source impedance

Input Gain Control Range

Mic In: 0 dB to +60 dB
Line In: -20 dB to +40 dB

Phantom Power

+48 VDC

Equalization

High-Pass Filter: 75 Hz @ 18 dB/octave

Rated Output

Line: +4 dBu
Maximum Rated Output:
+24 dBu @ Balanced Line-Level Outputs

Distortion (THD & IMD)

Mic Input to Line Output (@ +4 dBu output):
THD+N: < 0.0007%, 20 Hz to 20 kHz BW, 1 kHz input
@ +4 dBu, preamp at unity gain
THD+N: < 0.002%, 10 Hz to 30 kHz BW, 1kHz input
@ 10mV rms (-37.8 dBu)
SMPTE IMD: < 0.005% (4:1 IMD)

Mic Input to Digital Output (AES, 48 kHz sample rate):
THD+N: < 0.004%, 10mV rms input,
gain at -1 dB FS output

Output Impedance

Line: 100 Ω balanced

Signal Level LEDs

-20 dBu, 0 dBu (normal operating level), 0L = 22 dBu

Sample Frequency Selections

32 kHz, 44.1 kHz, 48 kHz, 88.2 kHz, 96 kHz,
176.4 kHz, 192 kHz, External

Bit Depth Selections

24-bit (full range from converters), 16-bit (dithered)

Analog Input Connectors

Eight balanced XLR mic inputs
Two 1/4" TS high-impedance instrument inputs
One DB25 connector with eight balanced line-level inputs

Analog Output Connectors

One DB25 connector with eight balanced line-level
outputs

Digital Input Connectors

One BNC connector for external word clock input

Digital Output Connectors

Two Toslink Optical Connectors
Both transmit channels 1-8 at 44.1/48 kHz operation
One transmits channels 1-4 and one transmits
channels 5-8 at 88.2/96 kHz operation using
S/MUX standard
One transmits channels 1-2 and one transmits
channel 3-4 at 176.4/192 kHz operation using
S/MUX standard
One DB25 Connector
Transmits AES/EBU or S/PDIF formatted digital
audio with single-wire/dual-wire options available at
all sample rates

Figura B.2: Especificações do pré-amplificar utilizado nas gravações [42]

Referências

- [1] Luís L. Henrique. *Acústica Musical*. Fundação Calouste Gulbenkian, fundação calouste gulbenkian edição, 2002.
- [2] Johan Sundberg. The acoustics of the singing voice. *Scientific American*, páginas 104–114, March 1977.
- [3] The voice tablet. Blue Tree Publishing.
- [4] Evan Bradley. An investigation of the acoustic vowel space of singing. Em *Proceedings of the 11 th International Conference on Music Perception and Cognition, Seattle, W*. Department of Linguistics and Cognitive Science, University of Delaware, 2010.
- [5] Yasunori Ohishi, Masataka Goto, Katunobu Itou, e Kazuya Takeda. Discrimination between singing and speaking voices. Em *INTERSPEECH 2005*, páginas 1141–1144, 2005.
- [6] Johan Sundberg. Level and center frequency of the singer’s formant. *Quarterly Progress and Status Report*, vol. 40, páginas 087–094, 1999.
- [7] Alex Loscos. *Spectral Processing of the Singing Voice*. Tese de doutoramento, Universitat Pompeu Fabra, 2007.
- [8] Martín Rocamora e Perfecto Herrera. Comparing audio descriptors for singing voice detection in music audio files. Facultad de Ingeniería de la Universidad de la República and Universitat Pompeu Fabra.
- [9] Preeti Rao. Musical information extraction from the singing voice. Em *National Conference on Signal and Image Processing Applications, IET Mumbai*, 2007.
- [10] Ixone Arroabarren, Miroslav Zivanovic, José Bretos, Amaya Ezcurra, e Afonso Carlosena. Measurement of vibrato in lyric singers. *IEEE Transactions on Instrumentation and Measurement*, VOL. 51, NO.4, páginas 660–665, August 2002.
- [11] Ixone Arroabarren, Xavier Rodet, e Afonso Carlosena. On the measurement of the instantaneous frequency and amplitude of partials in vocal vibrato. *IEEE Transactions on Audio, Speech, and Language Processing*, VOL. 14, NO.4, páginas 1413–1421, July 2006.
- [12] Ixone Arroabarren e Afonso Carlosena. Voice production mechanisms of vocal vibrato in male singers. *IEEE Transactions on Audio, Speech, and Language Processing*, VOL. 15, NO.1, páginas 320–332, January 2007.
- [13] Helen F. Mitchell e Dianna T. Kenny. Em *Proceedings of the third Conference on Interdisciplinary Musicology, tallinn, Estonia*, páginas 15–19, 2007.

- [14] Dirk Mürbe, Thomas Zahnert, Eberhard Kuhlisch, e Johan Sundberg. Effects of professional singing education on vocal vibrato - a longitudinal study. *Journal of Voice*, Vol.21, No. 6, páginas 683–688, 2007.
- [15] Ixone Arroabarren e Afonso Carlosena. Vibrato in singing voice: The link between source-filter and sinusoidal models. *Eurasip Journal on Applied Signal Processing*, vol. 2004, páginas 1007–1020, 2004.
- [16] T. J. Millhouse e F. Clermont. Perceptual characterisation of the singer's formant region: A preliminary study.
- [17] Jody Kreiman, Diana Vanlancker-Sidtis, e Bruce Gerratt. Defining and measuring voice quality. Em *Sound to Sense*, MIT, June 2004.
- [18] Jan W. M. A. F. Martens, Huib Versnel, e Philippe H. Dejonchere. The effect of visible speech in the perceptual rating of pathological voices.
- [19] Caitlin J. Butte, Yu Zhang, Huangqiang Song, e Jack J. Jiang. Perturbation and nonlinear dynamic analysis of different singing styles. *Journal of Voice*, 2008.
- [20] S. McCoy. Chapter 1: Listening to singers. in *Your Voice: An Inside View*.
- [21] Barbara Patricia Burke. *Electronic Detection of nasality in the Singing Voice Using WaveShape Analysis*. Tese de doutoramento, The Florida State University, 1982.
- [22] João Filipe Terleira de Sá Ferreira. Tecnologia de apoio em tempo-real ao canto; abordagem acerca de parâmetros qualitativos e percetivos. Relatório técnico, Escola Superior de Música e Artes do Espetáculo, December 2011.
- [23] G. Fant. The source filter concept in voice production. *Quarterly Progress and Status Report*, Vol.22, No.1, páginas 021–037.
- [24] C. Zarras, K. Pastiadis, G. Papadelis, e G. Papanikolaou. Cepstrum-based estimation of resonance frequencies (formants) in high-pitch singing signals. Em *Proceedings DAGA, Berlin*, 2010.
- [25] Tamás Bőhm e Géza Németh. Algorithm for formant tracking, modification and synthesis. *Híradástechnika*, páginas 11–16, 2006.
- [26] Adam P. Kestian e Tamara Smyth. Real-time estimation of the vocal tract shape for musical control. 2010.
- [27] Chanwoo Kim, Kwang-deok Seo, e Wonyong Sung. A robust formant extraction algorithm combining spectral peak picking and root polishing. *EURASIP Journal on Applied Signal Processing*, páginas 1–16, 2006.
- [28] Géza Németh, Géza Kiss, Tomás Bőhm, e József Kiss. An algorithm for formant analysis, tracking and modification.
- [29] François Thibault. Formant trajectory detection using hidden markov models. Relatório técnico, Sound Processing and Control Lab, McGill University, Montreal, Canada, December 2003.

- [30] B. Yegnanarayana e Raymond N. J. Veldhuis. Extraction of vocal-tract system characteristics from speech signals. *IEEE Transactions on speech and audio processing*, Vol.6, No.4, páginas 313–327, July 1998.
- [31] José Ventura. Biofeedback da voz cantada. Tese de mestrado, Faculdade de Engenharia da Universidade do Porto, 2011.
- [32] Alan Ó Cinnéide. Linear prediction - the technique, its solution and application to speech. Relatório técnico, Dublin Institute of Technology, August 2008.
- [33] Diana Rocha Mendes. Reconhecimento de orador em dois segundos. Tese de mestrado, Faculdade de Engenharia da Universidade do Porto, 2011.
- [34] Katharine Murphy. *Digital signal processing techniques for application in the analysis of pathological voice and normaphonic singing voice*. Tese de doutoramento, Universidad Politécnica de Madrid, 2008.
- [35] Saeed V. Vaseghi. *Advanced Digital signal Processing and Noise Reduction*. John Wiley & Sons Ltd, second edition edição. Chapter 8: Linear Prediction Models.
- [36] Xuedong Huang, Alex Acero, e Hsiao-Wuen Hon. *Spoken Language Processing*. Prentice-Hall PTR, 2001.
- [37] G. Ravindran, S. Shenbagadevi, e V. Salai Selvam. Cepstral and linear prediction techniques for improving intelligibility and audibility of impaired speech. *J. Biomedical Science and Engineering*, páginas 85–94, 2010.
- [38] G Fant e A. Bladon. A two-formant model and the cardinal vowels. *Quarterly Progress and Status Report, KTH*, páginas 01–08, 1978.
- [39] Madde, an additive, real-time, singing synthesiser. <http://www.speech.kth.se/smptool/>.
- [40] Ricardo Sousa. *Metodologias de Avaliação Perceptiva e Acústica do Sinal de Voz em Aplicações de Ensino do Canto e Diagnóstico/Reabilitação da Fala*. Tese de doutoramento, Faculdade de Engenharia da Universidade do Porto, 2011.
- [41] Microfone Neumann TLM 103. <http://www.neumann.com/>.
- [42] Pré-amplificador Mackie ONYX - 800R. <http://www.mackie.com/products/800r/>.