

**Faculdade de Engenharia da Universidade do Porto**



**FEUP**

## **Biofeedback da Voz Cantada**

José Alexandre Pinto dos Santos Ventura

Dissertação de Projecto realizada no âmbito do  
Mestrado Integrado em Engenharia Electrotécnica e de Computadores  
Major Telecomunicações

Orientador: Prof. Dr. Aníbal Ferreira

27 de Junho de 2011

© José Ventura, 2011

# Resumo

O *feedback* da voz cantada é utilizado como meio de ajuda à aprendizagem do canto e envolve uma interacção entre o utilizador e o programa que representa visualmente as características da voz. A qualidade e dinâmica da interacção está directamente ligada não só com as características da voz, mas também com o método de representação das mesmas. Um programa que aplica este conceito é o *SingingStudio*. Este *software* é a base de todo o desenvolvimento efectuado na dissertação.

A representação visual da voz é também aplicada por outros programas que executam uma abordagem similar ao *SingingStudio*. Deste modo, foi necessário elaborar uma comparação entre os vários programas com o objectivo de avaliar os aspectos que os distinguem. Os parâmetros da voz cantada e suas características foram também alvo de estudo e a partir da quantidade de informação recolhida, foi possível identificar os parâmetros de voz associados ao vibrato que poderiam melhorar o *feedback* visual já produzido pelo *SingingStudio*. O estudo do vibrato e a elaboração de um algoritmo que permite extrair as suas características são os principais contextos de desenvolvimento desta dissertação. Este algoritmo pode ser integrado no *SingingStudio* para que este *software* proporcione melhor *biofeedback* da voz cantada.

Foram efectuados testes de comparação entre três dos melhores métodos de estimação da frequência fundamental da voz, incluindo o método usado pela plataforma a utilizar no projecto, de forma a avaliar a sua precisão, robustez e também identificar áreas de possível melhoramento. Os resultados obtidos desta comparação atestam a elevada precisão do algoritmo usado pelo *SingingStudio* e revelam os melhoramentos ainda possíveis de serem executados.

O algoritmo desenvolvido para detectar, estimar a frequência e a extensão do vibrato foi também testado em diversas condições, por forma a obter uma medida consolidada do seu desempenho. Os resultados obtidos demonstram a precisão e robustez do algoritmo de estimação do vibrato, tornando possível a sua integração no *SingingStudio*.

A dissertação “biofeedback da voz cantada” enquadra-se num projecto financiado pela Fundação para a Ciência e Tecnologia que visa o desenvolvimento de tecnologias interactivas de apoio ao ensino e aprendizagem de canto.

# Abstract

The feedback of the singing voice is a means to aid the learning of carrying a tune which involves an interaction between the user and the program that visually represents the features of the voice. The quality and dynamic of this interaction is directly proportional, not only with the features of the voice, but also with the method used to represent them. A program that applies this concept is SingingStudio. This software is used as a starting point for the developments made on the course of this master thesis.

The visual representation of the voice is also used by other programs, similar in function with the SingingStudio. Thus, a comparison between these programs is necessary in order to assess the aspects that distinguish them. The features of the singing voice were also studied in order to identify the features that describe vibrato and to improve the visual feedback already provided by SingingStudio. The study of vibrato and the development of a suitable algorithm for feature extraction are the main contexts of research and development of this master thesis. This algorithm would then be integrated in SingingStudio, improving its efficiency on biofeedback of the singing voice.

Comparison tests were conducted between the three best estimation methods of fundamental voice frequency, including the method used by the SingingStudio software. It is then possible to evaluate its precision, robustness and to identify areas for possible improvements. The results obtained from this comparison show the high accuracy of the algorithm used by SingingStudio and reveal some improvements that may be required.

The developed algorithm that detects and estimates both the frequency and the extension of the vibrato was also tested in a significant variety of contexts in order to get a consolidated assessment of its performance. The final results show high accuracy and robustness of the algorithm that estimates the parameters of the vibrato, proving that is suitable to be integrated into SinginStudio environment.

The “Biofeedback da voz cantada” master thesis is incorporated in a project supported by “Fundação para a Ciência e Tecnologia” and which aims to develop interactive technologies that help in both the teaching and learning of singing.



# Agradecimentos

O autor deste documento gostaria de agradecer todos os apoios e incentivos dirigidos ao longo da dissertação, e de uma forma especial à minha mãe, irmãos e namorada, pelo carinho e ajuda que me deram.

Tenho a agradecer ao meu orientador, Prof. Aníbal Ferreira, por toda a motivação e conhecimentos partilhados que tanto me ajudaram no decorrer da dissertação. Não posso deixar de agradecer ao Eng. Ricardo Sousa pelas inúmeras horas de confronto de ideias e enaltecer o seu companheirismo sempre que foi preciso.

Por fim, gostaria de agradecer ao meu Pai, por me ter proporcionado as condições para nunca deixar de sonhar e almejar sempre ir mais longe.



# Índice

Resumo .....	iii
Abstract.....	v
Agradecimentos .....	vii
Índice.....	ix
Lista de figuras .....	xiii
Lista de tabelas .....	xix
Abreviaturas e Símbolos .....	xxi
<b>Capítulo 1 .....</b>	<b>1</b>
Introdução.....	1
1.1 - Motivação .....	1
1.2 - Objectivos.....	1
1.3 - Planificação de trabalhos .....	2
1.4 - Estrutura da Dissertação.....	2
<b>Capítulo 2 .....</b>	<b>5</b>
Estado da Arte.....	5
2.1 - A Voz Humana .....	5
2.1.1 - Voz falada .....	8
2.1.2 - Voz cantada .....	9
2.2 - Análise da Voz cantada .....	10
2.3 - Características e Parâmetros da Voz Cantada .....	10
2.3.1 - <i>Loudness</i> .....	12
2.3.2 - <i>Pitch</i> .....	12
2.3.3 - Vibrato .....	12
2.4 - Estudo de plataformas existentes .....	13
<b>Capítulo 3 .....</b>	<b>15</b>
Estimação da Frequência Fundamental .....	15
3.1 - Definição de Frequência Fundamental.....	15
3.2 - Análise no domínio dos Tempos .....	17
3.2.1 - Eventos temporais.....	17
3.2.2 - <i>Zero-Crossing Rate</i> .....	18
3.2.3 - <i>Peak Detection</i> .....	19

3.2.4 -	<i>Slope Event Rate</i> .....	20
3.2.5 -	Autocorrelação.....	20
3.2.6 -	Algoritmo <i>Boersma</i> .....	21
3.2.7 -	<i>Yin Estimator</i> .....	22
3.3 -	Análise no domínio das Frequências.....	23
3.3.1 -	<i>Component Frequency Ratios</i> .....	23
3.3.2 -	Métodos baseados em filtros .....	24
3.3.3 -	Análise Cepstral .....	24
3.3.4 -	Métodos multi-resolucionais.....	25
3.3.5 -	Algoritmo <i>SearchTonal</i> .....	25
<b>Capítulo 4</b> .....		<b>29</b>
Estimador de Vibrato.....		29
4.1 -	Definição de Vibrato .....	29
4.2 -	Estrutura do algoritmo.....	30
4.3 -	Deteccção e duração do Vibrato .....	31
4.4 -	Estimação da Extensão .....	40
4.5 -	Estimação da frequência .....	42
4.6 -	<i>Biofeedback</i> do Vibrato.....	43
<b>Capítulo 5</b> .....		<b>45</b>
Planificação de testes .....		45
5.1 -	Ferramentas utilizadas .....	45
5.2 -	Estimação da Frequência Fundamental.....	46
5.3 -	Estimador de Vibrato.....	49
5.3.1 -	Estimação da Frequência .....	50
5.3.2 -	Estimação da Extensão .....	51
5.3.3 -	Deteccção e estimação da duração .....	52
<b>Capítulo 6</b> .....		<b>53</b>
Resultados .....		53
6.1 -	Estimação da Frequência Fundamental.....	53
6.1.1 -	Teste com Vozes Sintetizadas .....	53
6.1.2 -	Teste com Vozes sintetizadas na presença de ruído .....	58
6.1.3 -	Teste com segmentos de Vozes Naturais.....	63
6.1.4 -	Teste com amostras de longa duração .....	66
6.2 -	Estimador de Vibrato.....	69
6.2.1 -	Teste com Ondas Sinusoidais.....	69
6.2.2 -	Teste com Vozes Sintetizadas.....	70
6.2.3 -	Teste com segmentos de Vozes Naturais.....	71
6.2.4 -	Teste com amostras de longa duração .....	75
6.3 -	Discussão de Resultados .....	78
6.3.1 -	Estimação da Frequência Fundamental .....	78
6.3.2 -	Caracterização do Vibrato .....	78
<b>Capítulo 7</b> .....		<b>79</b>
Conclusão .....		79
7.1 -	Estimação da Frequência Fundamental.....	79
7.2 -	Estimador de Vibrato.....	80
7.3 -	Trabalho Futuro.....	81
<b>Anexo A</b> .....		<b>83</b>
<b>Anexo B</b> .....		<b>97</b>
<b>Anexo C</b> .....		<b>107</b>

<b>Anexo D.....</b>	<b>113</b>
<b>Anexo E.....</b>	<b>115</b>
<b>Anexo F.....</b>	<b>117</b>
<b>Anexo G.....</b>	<b>141</b>
<b>Referências .....</b>	<b>157</b>



## Lista de figuras

<b>Figura 1</b> - Planificação de trabalhos através de um diagrama de <i>Gantt</i> . ....	2
<b>Figura 2</b> - As pregas vocais em acção [6]. ....	6
<b>Figura 3</b> - Representação dos formantes da vogal [a], produzida por um adulto masculino. ....	7
<b>Figura 4</b> - Representação da tonalidade da vogal [a], produzida por um adulto masculino. ....	8
<b>Figura 5</b> - Ilustração de um sistema fonte-filtro. ....	10
<b>Figura 6</b> - Representação temporal de uma nota musical com F0 440Hz. ....	16
<b>Figura 7</b> - Espectrograma de uma voz sintetizada com F0 de valor 440Hz. ....	16
<b>Figura 8</b> - Análise de frequências de uma voz sintetizada com F0 de valor 440Hz. ....	17
<b>Figura 9</b> - A influência de harmónicos às altas frequências no ZCR. ....	18
<b>Figura 10</b> - Detecção de máximos e mínimos de uma forma de onda. ....	19
<b>Figura 11</b> - Forma de onda sinusoidal e correspondente autocorrelação. ....	20
<b>Figura 12</b> - Algoritmo de implementação do <i>Cepstrum</i> . ....	25
<b>Figura 13</b> - Diagrama de Blocos de alto nível do algoritmo <i>SearchTonal</i> . ....	26
<b>Figura 14</b> - Fluxograma do método de selecção de F0 por parte do algoritmo <i>SearchTonal</i> . .	27
<b>Figura 15</b> - Voz sintetizada com vibrato. ....	29
<b>Figura 16</b> - Implementação do estimador de Vibrato. ....	30
<b>Figura 17</b> - Diagrama de alto nível do estimador de Vibrato. ....	31
<b>Figura 18</b> - Fluxograma do algoritmo de detecção do Vibrato. ....	32
<b>Figura 19</b> - Evolução da janela de sinal F0. ....	33
<b>Figura 20</b> - Sobreposição das janelas de sinal F0. ....	34
<b>Figura 21</b> - Algoritmo de tratamento da janela de sinal F0. ....	34
<b>Figura 22</b> - Tratamento da janela de sinal F0. ....	35

<b>Figura 23</b> - Visualização da recta utilizada na interpolação linear do cálculo do Valor de F0 desconhecido.....	36
<b>Figura 24</b> - Cálculo do espectro logarítmico de potências do sinal. ....	36
<b>Figura 25</b> - Comparação do espectro logarítmico de potências do sinal com e sem componente DC.....	37
<b>Figura 26</b> - Comparação do espectro logarítmico de potências do sinal com a sua envolvente na presença de vibrato. ....	38
<b>Figura 27</b> - Comparação do espectro de potências logarítmico do sinal com a sua envolvente na ausência de vibrato. ....	39
<b>Figura 28</b> - Janela de sinal F0 com dois valores distintos.....	40
<b>Figura 29</b> - Fluxograma do algoritmo de extracção do valor da extensão do vibrato.....	41
<b>Figura 30</b> - Região da janela onde é efectuada a procura de máximos e mínimos locais. ....	42
<b>Figura 31</b> - Algoritmo de estimação do valor da frequência do vibrato. ....	43
<b>Figura 32</b> - Detecção de vibrato no <i>SingingStudio</i> . ....	44
<b>Figura 33</b> - Representação dos parâmetros do vibrato no <i>SingingStudio</i> .....	44
<b>Figura 34</b> - <i>Software</i> Madde utilizado para sintetizar voz cantada. ....	46
<b>Figura 35</b> - Gráfico de comparação dos algoritmos de estimação de F0, através do erro relativo dos mesmos. ....	57
<b>Figura 36</b> - Gráfico do desempenho de <i>SearchTonal</i> na presença de ruído.....	61
<b>Figura 37</b> - Gráfico do desempenho do estimador Yin na presença de ruído. ....	62
<b>Figura 38</b> - Gráfico dos desempenhos dos algoritmos estimadores de F0, através da voz natural JMA-cresc-C4-a. ....	63
<b>Figura 39</b> - Gráfico dos desempenhos dos algoritmos estimadores de F0, através da voz natural RE-cresc-A4-a. ....	64
<b>Figura 40</b> - Gráfico dos desempenhos dos algoritmos estimadores de F0, através da voz natural RE-cresc-D5-u. ....	65
<b>Figura 41</b> - Gráfico dos desempenhos dos algoritmos estimadores de F0, através da voz natural CT2-ppc-c-m1. ....	66
<b>Figura 42</b> - Gráfico dos desempenhos dos algoritmos estimadores de F0, através da voz natural RE-cresc-A4-a. ....	67
<b>Figura 43</b> - Gráfico dos desempenhos dos algoritmos estimadores de F0, através da voz natural soprano44. ....	68
<b>Figura 44</b> - Desempenho do estimador de vibrato, através da voz JMA-cresc-G3-e. ....	72
<b>Figura 45</b> - Desempenho do estimador de vibrato, através da voz RE-cresc-D5-u. ....	73
<b>Figura 46</b> - Desempenho do estimador de vibrato, através da voz JMA-cresc-C4-u. ....	74

<b>Figura 47</b> - Desempenho do estimador de vibrato, através da voz LP-cmem-6-a. ....	75
<b>Figura 48</b> - Desempenho do estimador de vibrato, através da voz JN-cmem-6-a. ....	76
<b>Figura 49</b> - Desempenho do estimador de vibrato, através da voz CT2-ppc-c-m1.....	77
<b>Figura A1</b> - comparação dos resultados obtidos pelos métodos <i>SearchTonal</i> e <i>Yin</i> , através de vozes sintetizadas com ruído.....	95
<b>Figura B1</b> - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz JMA-cresc-C4-e. ....	97
<b>Figura B2</b> - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz JMA-cresc-C4-u. ....	98
<b>Figura B3</b> - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz JMA-cresc-G3-a. ....	99
<b>Figura B4</b> - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz JMA-cresc-G3-e. ....	100
<b>Figura B5</b> - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz JMA-cresc-G3-u. ....	101
<b>Figura B6</b> - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz RE-cresc-A4-a. ....	102
<b>Figura B7</b> - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz RE-cresc-A4-e. ....	103
<b>Figura B8</b> - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz RE-cresc-A4-u. ....	104
<b>Figura B9</b> - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz RE-cresc-D5-a. ....	105
<b>Figura B10</b> - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz RE-cresc-D5-e. ....	106
<b>Figura C1</b> - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz CT2-ppc-c-m2. ....	107
<b>Figura C2</b> - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz JN-mem-6-a. ....	108
<b>Figura C3</b> - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz LP-mem-6-a. ....	109
<b>Figura C4</b> - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz suz32. ....	110
<b>Figura C5</b> - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz suz32solo. ....	111
<b>Figura E1</b> - Visualização do espectro de potências, através da ODFT do sinal Vibrato_concatenado_ST05-10.....	115
<b>Figura E2</b> - Visualização da estimação dos parâmetros do vibrato, através do sinal Vibrato_concatenado_ST05-10.....	116

<b>Figura F1</b> - Visualização da estimação dos parâmetros do vibrato, através do sinal JMA-cresc-C4-a. ....	117
<b>Figura F2</b> - Visualização do espectro de potências através da ODFT do sinal JMA-cresc-C4-a. ....	118
<b>Figura F3</b> - Visualização da estimação dos parâmetros do vibrato, através do sinal JMA-cresc-C4-e. ....	119
<b>Figura F4</b> - Visualização do espectro de potências através da ODFT do sinal JMA-cresc-C4-e. ....	120
<b>Figura F5</b> - Visualização da estimação dos parâmetros do vibrato, através do sinal JMA-cresc-C4-u. ....	121
<b>Figura F6</b> - Visualização do espectro de potências através da ODFT do sinal JMA-cresc-C4-u. ....	122
<b>Figura F7</b> - Visualização da estimação dos parâmetros do vibrato, através do sinal JMA-cresc-G3-a. ....	123
<b>Figura F8</b> - Visualização do espectro de potências através da ODFT do sinal JMA-cresc-G3-a. ....	124
<b>Figura F9</b> - Visualização da estimação dos parâmetros do vibrato, através do sinal JMA-cresc-G3-e. ....	125
<b>Figura F10</b> - Visualização do espectro de potências através da ODFT do sinal JMA-cresc-G3-a. ....	126
<b>Figura F11</b> - Visualização da estimação dos parâmetros do vibrato, através do sinal JMA-cresc-G3-u. ....	127
<b>Figura F12</b> - Visualização do espectro de potências através da ODFT do sinal JMA-cresc-G3-u. ....	128
<b>Figura F13</b> - Visualização da estimação dos parâmetros do vibrato, através do sinal RE-cresc-A4-a. ....	129
<b>Figura F14</b> - Visualização do espectro de potências através da ODFT do sinal RE-cresc-A4-a. ....	130
<b>Figura F15</b> - Visualização da estimação dos parâmetros do vibrato, através do sinal RE-cresc-A4-e. ....	131
<b>Figura F16</b> - Visualização do espectro de potências através da ODFT do sinal RE-cresc-A4-e. ....	132
<b>Figura F17</b> - Visualização da estimação dos parâmetros do vibrato, através do sinal RE-cresc-A4-u. ....	133
<b>Figura F18</b> - Visualização do espectro de potências através da ODFT do sinal RE-cresc-A4-u. ....	134
<b>Figura F19</b> - Visualização da estimação dos parâmetros do vibrato, através do sinal RE-cresc-D5-a. ....	135
<b>Figura F20</b> - Visualização do espectro de potências através da ODFT do sinal RE-cresc-D5-a. ....	136

<b>Figura F21</b> - Visualização da estimação dos parâmetros do vibrato, através do sinal RE-cresc-D5-e. ....	137
<b>Figura F22</b> - Visualização do espectro de potências através da ODFT do sinal RE-cresc-D5-e. ....	138
<b>Figura F23</b> - Visualização da estimação dos parâmetros do vibrato, através do sinal RE-cresc-D5-u. ....	139
<b>Figura F24</b> - Visualização do espectro de potências através da ODFT do sinal RE-cresc-D5-u. ....	140
<b>Figura G1</b> - Visualização da estimação dos parâmetros do vibrato, através do sinal CT2-ppc-c-m1. ....	141
<b>Figura G2</b> - Visualização do espectro de potências através da ODFT do sinal CT2-ppc-c-m1. ....	142
<b>Figura G3</b> - Visualização da estimação dos parâmetros do vibrato, através do sinal CT2-ppc-c-m2. ....	143
<b>Figura G4</b> - Visualização do espectro de potências através da ODFT do sinal CT2-ppc-c-m2. ....	144
<b>Figura G5</b> - Visualização da estimação dos parâmetros do vibrato, através do sinal JF-mem-6-a. ....	145
<b>Figura G6</b> - Visualização do espectro de potências através da ODFT do sinal JF-mem-6-a. ...	146
<b>Figura G7</b> - Visualização da estimação dos parâmetros do vibrato, através do sinal JN-mem-6-a. ....	147
<b>Figura G8</b> - Visualização do espectro de potências através da ODFT do sinal JN-mem-6-a. ...	148
<b>Figura G9</b> - Visualização da estimação dos parâmetros do vibrato, através do sinal LP-mem-6-a. ....	149
<b>Figura G10</b> - Visualização do espectro de potências através da ODFT do sinal LP-mem-6-a. ...	150
<b>Figura G11</b> - Visualização da estimação dos parâmetros do vibrato, através do soprano44. ...	151
<b>Figura G12</b> - Visualização do espectro de potências através da ODFT do sinal soprano44. ...	152
<b>Figura G13</b> - Visualização da estimação dos parâmetros do vibrato, através do suz32. ....	153
<b>Figura G14</b> - Visualização do espectro de potências através da ODFT do sinal suz32. ....	154
<b>Figura G15</b> - Visualização da estimação dos parâmetros do vibrato, através do suz32solo. ...	155
<b>Figura G16</b> - Visualização do espectro de potências através da ODFT do sinal suz32solo. ...	156



# Lista de tabelas

<b>Tabela 1</b> - Breve descrição dos programas existentes. ....	13
<b>Tabela 2</b> - Passos executados e erro do algoritmo ao longo da sua implementação [32]. ....	23
<b>Tabela 3</b> - Descrição dos parâmetros utilizados no primeiro teste com vozes sintetizadas dos algoritmos estimadores de F0. ....	47
<b>Tabela 4</b> - Descrição dos parâmetros utilizados no segundo teste com vozes sintetizadas. ...	48
<b>Tabela 5</b> - Descrição dos parâmetros utilizados no terceiro teste com vozes sintetizadas. ...	48
<b>Tabela 6</b> - Descrição dos parâmetros utilizados no quarto teste com vozes sintetizadas. ....	49
<b>Tabela 7</b> - Descrição dos parâmetros utilizados nos primeiros testes associados à avaliação da estimação da frequência do vibrato. ....	51
<b>Tabela 8</b> - Descrição dos parâmetros utilizados nos primeiros testes associados à avaliação da estimação da extensão do vibrato. ....	52
<b>Tabela 9</b> - Resultados do algoritmo SearchTonal obtidos com vozes sintetizadas. ....	54
<b>Tabela 10</b> - Resultados do algoritmo Yin obtidos com vozes sintetizadas. ....	55
<b>Tabela 11</b> - Resultados do algoritmo Boersma obtidos com vozes sintetizadas. ....	56
<b>Tabela 12</b> - Resultados da adição de ruído às vozes sintetizadas. ....	58
<b>Tabela 13</b> - Erro relativo do resultado do algoritmo SearchTonal obtido através da adição de ruído às vozes sintetizadas. ....	59
<b>Tabela 14</b> - Erro relativo do resultado do estimador Yin obtido através da adição de ruído às vozes sintetizadas. ....	60
<b>Tabela 15</b> - Resultados obtidos pelo estimador de vibrato nos testes com ondas sinusoidais. ....	69
<b>Tabela 16</b> - Resultados obtidos pelo estimador de vibrato nos testes com vozes sintetizadas com uma extensão de Valor 0,5 ST. ....	70
<b>Tabela 17</b> - Resultados obtidos pelo estimador de vibrato nos testes com vozes sintetizadas com uma extensão de Valor 1,0 ST. ....	70

<b>Tabela 18</b> - Resultados obtidos pelo estimador de vibrato nos testes com vozes sintetizadas com uma frequência de Valor 5,0Hz. ....	71
<b>Tabela A1</b> - Resultados obtidos pelo algoritmo <i>SearchTonal</i> através de vozes sintetizadas com um SNR=30db. ....	83
<b>Tabela A2</b> - Resultados obtidos pelo algoritmo <i>SearchTonal</i> através de vozes sintetizadas com um SNR=25db. ....	84
<b>Tabela A3</b> - Resultados obtidos pelo algoritmo <i>SearchTonal</i> através de vozes sintetizadas com um SNR=20db. ....	85
<b>Tabela A4</b> - Resultados obtidos pelo algoritmo <i>SearchTonal</i> através de vozes sintetizadas com um SNR=15db. ....	86
<b>Tabela A5</b> - Resultados obtidos pelo algoritmo <i>SearchTonal</i> através de vozes sintetizadas com um SNR=10db. ....	87
<b>Tabela A6</b> - Resultados obtidos pelo algoritmo <i>SearchTonal</i> através de vozes sintetizadas com um SNR=5db. ....	88
<b>Tabela A7</b> - Resultados obtidos pelo algoritmo <i>Yin</i> através de vozes sintetizadas com um SNR=30db. ....	89
<b>Tabela A8</b> - Resultados obtidos pelo algoritmo <i>Yin</i> através de vozes sintetizadas com um SNR=25db. ....	90
<b>Tabela A9</b> - Resultados obtidos pelo algoritmo <i>Yin</i> através de vozes sintetizadas com um SNR=20db. ....	91
<b>Tabela A10</b> - Resultados obtidos pelo algoritmo <i>Yin</i> através de vozes sintetizadas com um SNR=15db. ....	92
<b>Tabela A11</b> - Resultados obtidos pelo algoritmo <i>Yin</i> através de vozes sintetizadas com um SNR=10db. ....	93
<b>Tabela A12</b> - Resultados obtidos pelo algoritmo <i>Yin</i> através de vozes sintetizadas com um SNR=5db. ....	94
<b>Tabela D1</b> - Resultado da estimação de vibrato através de uma voz sintetizada com ST=0.5, com zona de ataque e duração 3s. ....	113
<b>Tabela D2</b> - Resultado da estimação de vibrato através de uma voz sintetizada com ST=1.0, com zona de ataque e duração 3s. ....	113

# Abreviaturas e Símbolos

Lista de abreviaturas (ordenadas por ordem alfabética)

ABS	Valor absoluto de um número
CSA	Coeficiente de Sobre-Amostragem
DFT	<i>Discrete Fourier Transform</i>
EKG	Electroglotografia
F0	Frequência fundamental
FFT	<i>Fast Fourier Transform</i>
HNR	<i>Harmonic to Noise Ratio</i>
LPC	<i>Linear Predictive Coding</i>
ODFT	<i>Odd Discrete Fourier Transform</i>
SNR	<i>Signal to Noise Ratio</i>



# Capítulo 1

## Introdução

### 1.1 - Motivação

A voz, ao longo dos séculos, desempenhou um papel preponderante na evolução da vida humana, permitindo-nos comunicar e organizar como sociedades até aos dias de hoje. Actualmente é comum dividirmos a voz em duas classes: a voz falada e a voz cantada. Ao analisarmos a longa evolução do Ser Humano, constatamos que em todas as civilizações modernas que existiram, a voz cantada era mais que um elemento de entretenimento. Esta sempre esteve ligada a aspectos muito importantes como a religião, grupos sociais ou mesmo a nível individual, onde por vezes o canto era visto como o alimento da alma ou uma fonte de riqueza, como demonstra a origem grega da palavra musical, que significa homem culto.

A voz cantada continua a ter um papel muito importante nas sociedades actuais, sendo alvo de vários estudos com o intuito de determinar o seu funcionamento e características inerentes. Estes estudos são muito importantes em várias áreas, desde a medicina à música. Porém, estes estudos focam-se maioritariamente na voz falada, como por exemplo o reconhecimento de orador, e o estudo da voz cantada permanece num segundo plano. O processamento em tempo real da voz cantada é por isso ainda muito limitado a vários níveis e existe nesta área uma possibilidade de crescimento tecnológico elevado. É neste contexto que esta dissertação se enquadra, no propósito de fornecer novas ferramentas de análise da voz cantada de forma objectiva, através de um *feedback* de parâmetros da mesma.

### 1.2 - Objectivos

O principal objectivo da dissertação é o reforço do *feedback* objectivo da voz cantada em contextos de ensino/aprendizagem, tendo como base de desenvolvimento tecnológico o *software* comercial *SingingStudio* da *SEEgnal* [1]. Devido à complexidade envolvida na análise da voz cantada, antes de serem delineados objectivos mais específicos foi efectuado um estudo sobre os parâmetros mais relevantes na voz cantada e a forma como estes podem ser analisados. Este estudo, contemplado no estado da arte, é muito importante pois não só serviu como base para efectuar comparações entre o *SingingStudio* e outras plataformas de

### 1.3 - Planificação de trabalhos

análise da voz cantada, como também permitiu uma análise mais criteriosa de parâmetros a serem analisados pelo *SingingStudio*.

Após a conclusão da etapa descrita no parágrafo anterior e com base na informação recolhida, foram identificados dois objectivos principais: a necessidade de comparação de novos algoritmos de estimação da frequência fundamental com o usado no *SingingStudio*; melhorar a objectividade do *biofeedback* da voz cantada através da análise de parâmetros que não está disponível na maior parte do *software* estudado. Tendo em conta a importância e exequibilidade da implementação no decorrer da dissertação, determinou-se que os parâmetros associados ao vibrato seriam estudados. A metodologia e planificação de todo o trabalho de análise dos parâmetros e selecção dos mesmos são demonstrados no capítulo seguinte.

### 1.3 - Planificação de trabalhos

De forma a organizar as tarefas a realizar ao longo da dissertação, foi elaborado uma planificação de trabalhos segundo um diagrama de *Gantt*, que pode ser visualizado na figura abaixo ou na página da dissertação.

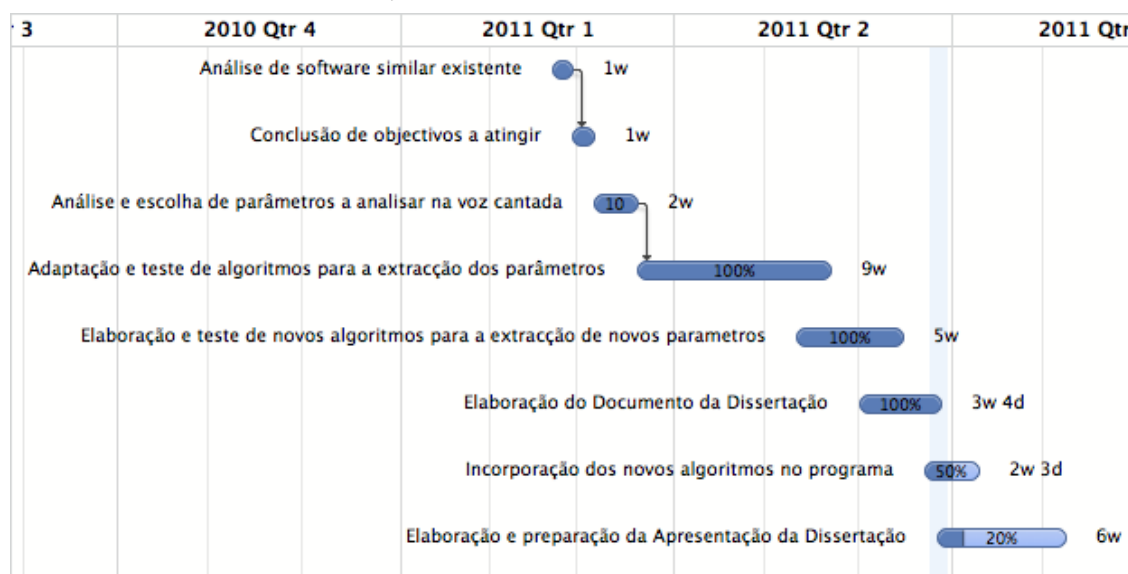


Figura 1 - Planificação de trabalhos através de um diagrama de *Gantt*.

O diagrama pode também ser consultado na página da dissertação [2][3], onde esteve sempre disponível ao longo da mesma, sendo um método de verificação dos avanços efectuados em cada tarefa ao longo do tempo.

### 1.4 - Estrutura da Dissertação

A dissertação encontra-se dividida em sete capítulos. No primeiro capítulo é abordado o tema da dissertação, a sua motivação e objectivos pretendidos. No segundo capítulo é efectuado o estado da arte, onde é reunida informação relevante para elaboração desta dissertação. No terceiro capítulo é abordado o tema da estimação da frequência fundamental, aspecto muito importante da voz cantada. Diferentes métodos serão estudados e os melhores algoritmos de estimação serão alvo de análise. No quarto capítulo será feito

um estudo sobre o vibrato, uma característica da voz cantada que, com base no estado da arte, se revelou bastante importante mas ao mesmo tempo uma lacuna nos *softwares* de análise da voz cantada. No quinto capítulo será delineado o conjunto de testes a efectuar, a sua metodologia e objectivos subjacentes. No sexto capítulo estão descritos os resultados dos testes efectuados e elaborar um resumo dos mesmos. O sétimo e último capítulo apresentarão as conclusões e o trabalho futuro.

## 1.4 - Estrutura da Dissertação

# Capítulo 2

## Estado da Arte

Neste capítulo são abordados os conceitos fundamentais relacionados com o trabalho desenvolvido, bem como uma breve análise de soluções já existentes relacionadas com o *feedback* da voz cantada. Devido à complexidade e abrangência de conceitos envolvidos no *biofeedback* da voz cantada, torna-se óbvia a necessidade de segmentar o seu estudo numa lista de objectivos simples, concretos e de importância relevante. Podemos facilmente delinear as quatro áreas chave que compõem a abordagem efectuada ao tema desta dissertação nos seguintes tópicos:

- A voz humana, suas características e funcionamento;
- Análise da voz cantada, a sua evolução ao longo do tempo e as características mais importantes do ponto de vista do cantor;
- Estudo de soluções existentes no mercado e a sua abordagem na resolução de problemas similares;
- Integração no *SingingStudio* de novos métodos de análise da voz cantada tendo em conta o desenvolvimento efectuada nos dois tópicos anteriores.

Nesta fase inicial do estudo, abordada na secção seguinte, é efectuada uma pequena abordagem sobre a origem da voz e o seu funcionamento, tal como foi referido no primeiro dos quatro pontos acima descritos. Esta introdução é importante pois só com a percepção do funcionamento da voz humana é possível uma maior eficácia no seu estudo e na formulação de novas abordagens à sua análise.

### 2.1 - A Voz Humana

A voz humana é principalmente produzida pela vibração das pregas vocais em resultado o ar expelido pelos pulmões, através da acção do diafragma, e que é modificado, particularmente quanto à coloração espectral, pelo tracto vocal, incluindo a acção dos lábios, dentes e língua [4].

O Ser Humano é o produto da sua própria evolução e a voz também é fruto de uma evolução entre a interacção do sistema nervoso, respiratório e digestivo, e de músculos, ligamentos e ossos que, todos eles em conjunto, actuam para que se possa obter uma emissão

## 2.1 - A Voz Humana

rica, profusa e eficiente. Uma das conclusões extraídas desta linha de pensamento prende-se com o propósito das pregas vocais, um par de músculos denominados tíreo-aritnóideo, muito provavelmente não se formaram para o uso da voz. Inicialmente, este par de músculos desenvolveu-se para as funções primárias de sobrevivência, respiração e alimentação. A voz como uma função apareceu depois, fruto da evolução humana e das suas actividades sempre mais complexas na qual a laringe, onde se situam as pregas, se especializou.

A produção da voz acontece quando o ar expirado a partir dos pulmões atravessa as pregas vocais, e através do comando do sistema nervoso ligado aos músculos do sistema fonatório, por meio de pequenos e múltiplos ajustes musculares, exerce pressões de diferentes graus na região abaixo das pregas vocais, fazendo-as vibrar. Contudo, esta forma de onda sonora produzida que se denomina voz, é caracterizada segundo várias propriedades, abordadas na secção seguinte. O ar expirado que faz as pregas vocais vibrarem, é modificado e os sons são articulados em vogais e consoantes. A fase seguinte, é a da modificação espectral que ocorre no trato vocal e nasal e de decorre a onda sonora, referida anteriormente [5].

Na análise da interação do ar com as pregas vocais, é constatado que quando respiramos, as pregas afastam-se deixando o ar circular nos pulmões, ao contrário da fala ou do canto onde as pregas vocais entram em contacto e restringem o fluxo de ar. A fala é produzida durante a expiração, uma vez que o fluxo do ar ao passar pela glote provoca a vibração das pregas vocais pelo efeito da pressão subglótica e o efeito de Bernoulli, produzindo assim ondas sonoras. A frequência das vibrações produzidas, bem como a altura do som, é condicionada pela tensão das pregas.

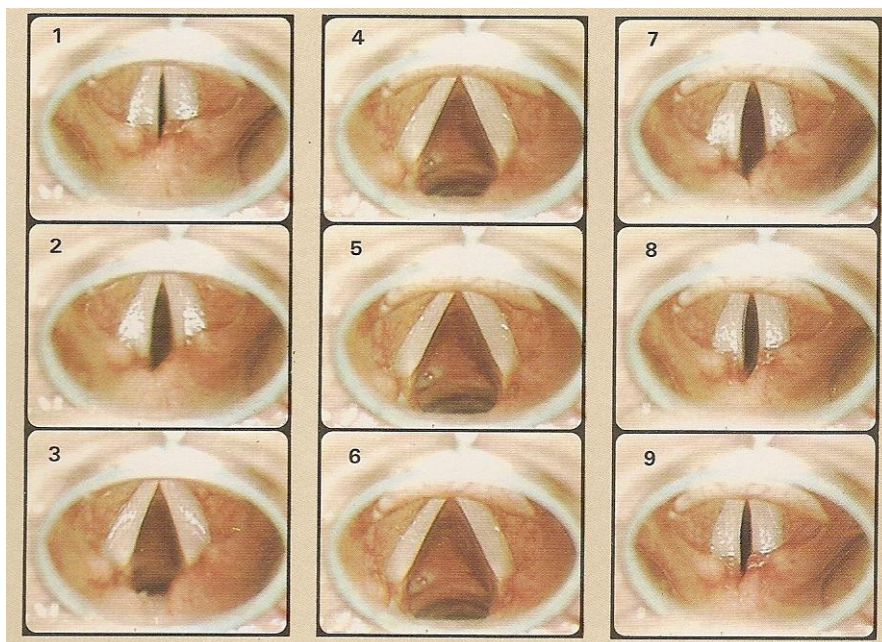


Figura 2 - As pregas vocais em acção [6].

Através das fotografias presentes na figura acima, é possível observar o funcionamento das pregas vocais, e da glote, onde cada uma delas foi tirada com um intervalo de um segundo. A glote é facilmente identificada pela abertura em fenda entre as pregas. Nas fotografias número um a número seis, a fala é interrompida, havendo um afastamento

gradual das pregas. Nas fotografias número sete a número nove, a fala é retomada [6], e observa-se o aproximar das pregas vocais referido no parágrafo anterior.

Os sons sonoros produzidos podem ser classificados em duas classes: sons vozeados e não vozeados. Nos sons vozeados, verifica-se a vibração das pregas vocais e nos não vozeados, ou sons surdos, as pregas vocais não vibram, permanecendo apenas abertas. Esta noção tem interesse nas perícias linguísticas e fonéticas, visto que os sons vozeados contêm um acervo de vogais predominante, enquanto os sons não vozeados têm como característica principal a predominância de consoantes. Esta definição é muito importante pois é partir dela que podemos elaborar um processo de análise/síntese de um sinal de voz. Exemplos de sons vozeados são os sons produzidos na pronúncia das vogais [a], [e], [i], [o], [u]. Por outro lado, a pronúncia de letras como o [f] e o [s] no meio de palavras são exemplos de sons não vozeados [7].

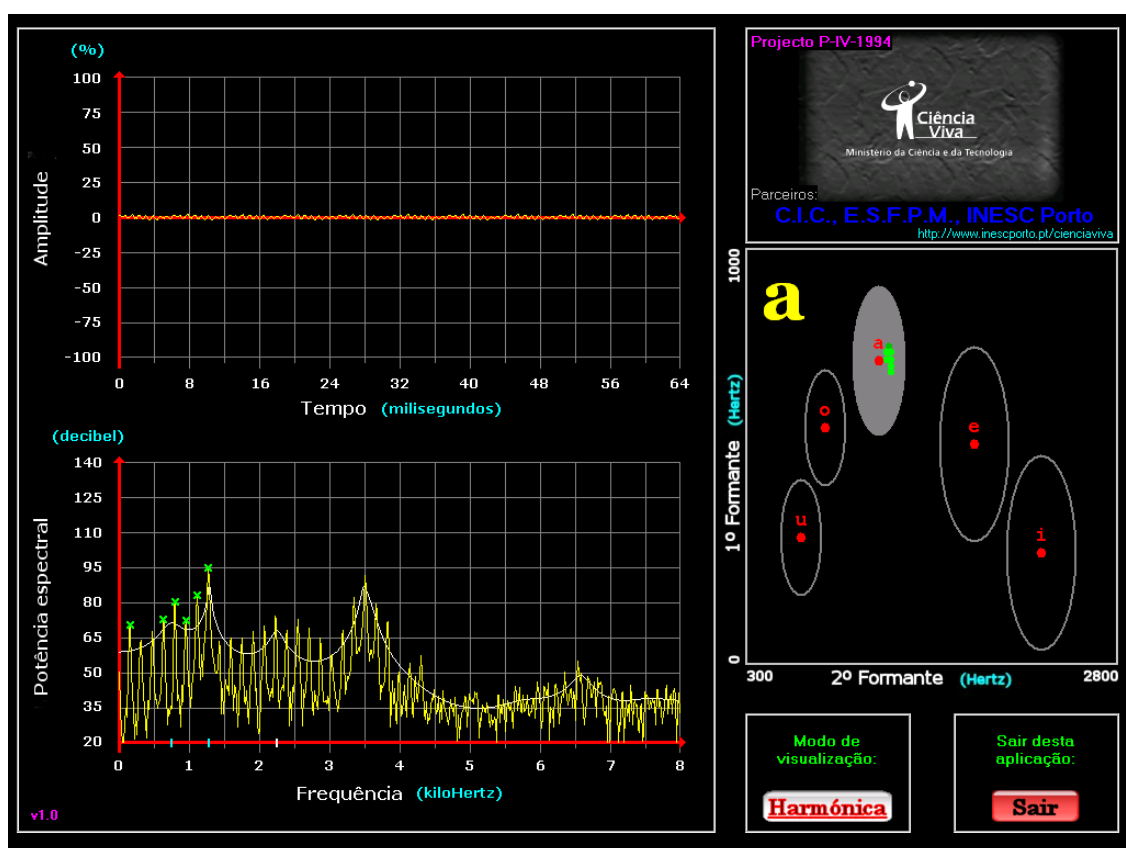


Figura 3 - Representação dos formantes da vogal [a], produzida por um adulto masculino.

A figura 3 foi obtida com o auxílio do programa de nome analisador de sinais de voz [8], e podemos observar a relação do primeiro e segundo formante na caracterização das vogais, bem como a potência espectral e amplitude do sinal analisado. O sinal analisado foi produzido por um adulto do sexo masculino. Na figura 4, também obtida através do mesmo programa, podemos observar a frequência fundamental e a sua relação com o género da pessoa que produziu a voz. O sinal avaliado foi semelhante ao ilustrado na figura 3.

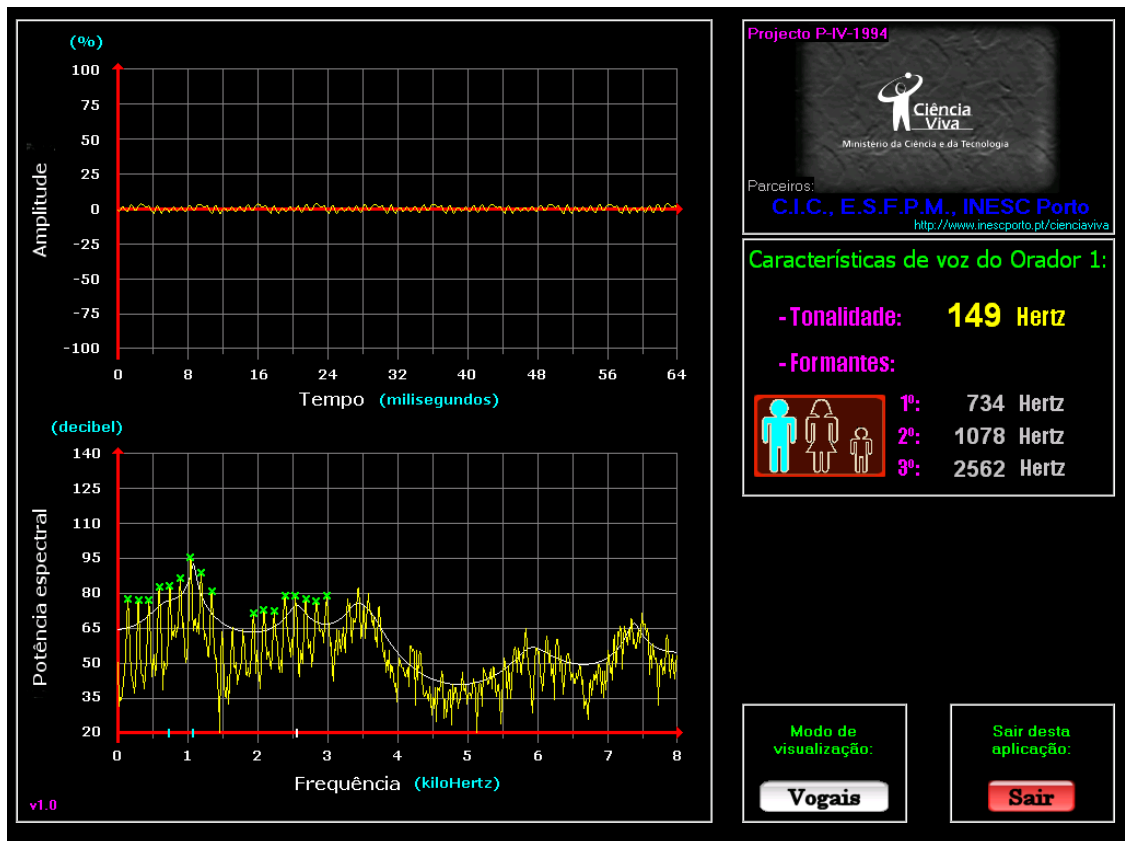


Figura 4 - Representação da tonalidade da vogal [a], produzida por um adulto masculino.

A vibração das pregas vocais é um dos principais fenómenos que contribuem para a produção de diferentes sons e acontece com uma certa frequência, designada de *pitch* e que varia com a voz e de pessoa para pessoa. As mulheres e crianças normalmente têm uma frequência maior, isto é, uma vibração mais rápida que os homens adultos, que produzem uma vibração mais lenta. A altura da voz, ou volume, é tipicamente determinada pela quantidade de ar expelido pelos pulmões.

### 2.1.1 -Voz falada

A fala envolve quase sempre a voz cuja diversidade psico-acústica pode variar quanto à intensidade, altura, ressonância, inflexão entre outras características, em que os fonemas são utilizados como unidades individuais da fala. As classes de fonemas do português são: oclusivas, fricativas, líquidas, nasais, semivogais e vogais.

Na voz falada, a respiração é natural e o ciclo respiratório completo varia de acordo com as emoções e o comprimento das frases. A inspiração é relativamente lenta e nasal nas pausas longas, sendo mais rápida e bucal durante a fala. O volume de ar utilizado é médio, e apenas nas situações de forte intensidade existe uma maior mobilização da caixa torácica.

O ciclo respiratório durante a voz falada é um dos parâmetros que mais se altera com as emoções, sendo que a mudança observada com mais frequência está relacionada com o ritmo inspiração/expiração. No uso da voz falada, é essencial a presença de uma boa articulação das palavras já que a mensagem que será transmitida depende delas.

### 2.1.2 -Voz cantada

O canto envolve activação e coordenação dos sistemas respiratório, fonatório, articulatório, ressonante e auditivo.

Ao contrário da fala, em que a entrada e saída de ar variam de acordo com a emoção e a mensagem transmitida, no canto a respiração é efectuada de acordo com as frases musicais e pausas. Quer na voz cantada quer na voz falada, o ruído produzido na inspiração é evitado, sendo que de um ponto de vista técnico ou estético, esta deva ser curta e oral. Para tal é necessária uma acção nas costelas inferiores, favorecendo assim a expansão da caixa torácica e que assegura o ar pleno, caracterizado como a soma do ar residual e do ar inspirado [8].

Do ponto de vista técnico, podemos classificar o tipo respiratório utilizado durante a fonação, quer este seja no canto ou na fala, como sendo costodiafragmático-abdominal. Este tipo possibilita uma respiração ampla e profunda, evitando o aumento de tensão muscular da região cervical. Durante a expiração, tendo em vista um melhor controlo da saída de ar, utiliza-se a musculatura abdominal. Este controlo da respiração varia conforme a intensidade do som, a altura, o timbre, a extensão e a duração da frase musical. Convém salientar que, no caso da voz cantada, um bom suporte respiratório é essencial para a projecção vocal adequada [8][9].

Tanto na voz cantada como na voz falada, a ressonância vocal consiste na modificação do som produzido pelas pregas vocais, modificação esta obtida através das cavidades de ressonância, onde ocorrem efeitos de atenuação e amplificação do sinal sonoro em regiões específicas de frequência. No canto, esta ressonância é de uma forma geral elevada, existindo uma maior concentração de energia sonora na parte superior do trato vocal. O objectivo de utilizar uma ressonância equilibrada tem a ver com facto de esta aliviar a sobrecarga muscular da laringe. Muitos cantores utilizam um certo grau de nasalidade como forma de dissipar a energia sonora sem sobrecarregar a laringe [9].

Uma das características presentes em cantores, não observadas em pessoas que não treinam de forma rigorosa a voz, é a estabilidade desta. Os melhores meios que um cantor tem para melhorar ainda mais esse parâmetro, aumentando assim a imunidade a factores externos, são: um bom controlo auditivo, técnica apropriada e propriocepção, que ajuda a coordenação muscular. No entanto, a qualidade da fala pode não só ser influenciada pelas emoções, tal como já foi mencionado, mas também pela natureza do discurso, sendo muito sensível ao interlocutor uma vez que tende a ser espontânea. No canto, o que acontece é o oposto, pois as peças musicais são geralmente conhecidas pelo ouvinte, o que torna a carga emocional e a retórica do texto mais previsível.

A intensidade da voz é relativamente constante na voz falada, o que não ocorre no canto. As variações observadas na voz cantada mostram que a intensidade pode variar de forma controlada de 45 a 110 decibéis. Para obter estes níveis, o cantor tem que controlar a quantidade de ar expirado, bem como a tensão das pregas vocais, que agem com resistência ao fluxo aéreo transglótico, e à amplificação do som ocorrida nas caixas de ressonância, o que proporciona ao cantor uma fonação com menos esforço [10].

## 2.2 - Análise da Voz cantada

No estudo da voz cantada existe um cruzamento de duas áreas de domínios distintos: psicofisiológico e tecnológico. Embora este cruzamento já exista há mais de 500 anos, os avanços tecnológicos efectuados nos últimos 50 permitiram melhorias significativas. A título de exemplo, podemos actualmente utilizar uma electroglotografia (EGG) que consiste em colocar dois eléctrodos no pescoço junto à laringe para retirar os parâmetros desejados directamente e de forma bastante precisa [11]. No entanto, devido à sua natureza pouco prática, é muito mais simples efectuar uma gravação e utilizar métodos computacionais para extrair os mesmos parâmetros. Os métodos computacionais têm evoluído ao longo dos anos, devido a avanços bastante importantes dos algoritmos computacionais tais como a FFT [12] e, posteriormente, algoritmos computacionais LPC [13]. De forma sucinta podemos dizer que este algoritmo consiste em extrair os formantes da voz, subtraí-los ao sinal original e posteriormente analisar a informação restante.

O modelo de análise da voz mais comum consiste na utilização de modelos matemáticos de forma a implementar um sistema fonte-filtro, que pode ser visualizado na imagem abaixo, em que a saída seria a voz e a fonte os impulsos glóticos que contêm os parâmetros pretendidos.

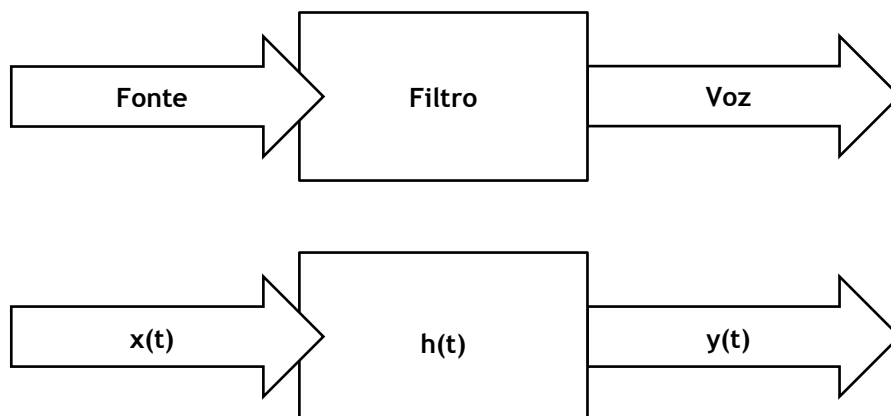


Figura 5 - Ilustração de um sistema fonte-filtro.

Numa fase posterior, usaríamos então um processo de *inverse filtering* para obter os parâmetros do sinal da fonte pretendida.

Os parâmetros mais comuns associados a estes processos são o *jitter*, o *shimmer* e o HNR [14]. O *jitter* e o *shimmer* são parâmetros associados a processos de detecção de marcas de *pitch*, enquanto que o HNR é um parâmetro associado à separação das componentes harmónica e de ruído.

## 2.3 - Características e Parâmetros da Voz Cantada

Ao focarmos a nossa análise na voz cantada, reparamos que existem similaridades com instrumentos de tubos e cordas. Quando comparamos o funcionamento de um instrumento de cordas, como o violino, com a nossa voz, verificamos que existem bastantes similaridades. É

fácil perceber que quando pretendemos modificar o *pitch*, ou a frequência fundamental, temos que encurtar o comprimento da corda, colocando o dedo na posição correspondente à nota pretendida. Da mesma forma, a vibração das cordas vocais é alterada segundo a sua tensão, as suas características físicas e a pressão exercida pelos pulmões. Se fizermos outra comparação, desta vez com um instrumento de sopro, como uma flauta, verificamos que a mesma mudança de F0 é efectuada tapando ou destapando os furos e, tal como na primeira comparação, também somos capazes de efectuar mudanças no trato vocal quando baixamos a laringe, o que origina um alongamento [15] e conseqüente mudança de F0.

Deste modo, define-se o *pitch* como a percepção do ouvido humano da frequência fundamental F0 ouvida, sendo este um dos aspectos a analisar na voz cantada.

Um trabalho bastante importante é a da formulação de um modelo matemático que descreva o funcionamento das pregas vocais tendo em conta todos os parâmetros descritos nos dois parágrafos anteriores. Para tal, foi muito importante o trabalho realizado por Ingo Titze [16] o qual estabelece que a frequência fundamental é determinada pela raiz quadrada da relação rigidez/massa do oscilador, sendo que a rigidez (*stiffness*) pode ser obtida matematicamente através de uma relação entre o comprimento, espessura, tensão e profundidade da prega vocal. Esta formulação poderá ser utilizada para efectuar uma melhor comparação entre a mudança nas cordas vocais e o resultado obtido na voz.

Para além do *pitch*, podemos definir o *loudness* como outro aspecto a analisar na voz cantada, sendo o *loudness* a percepção da amplitude da voz. Existem vários métodos para os cantores modificarem este parâmetro mas todos envolvem mudanças de pressão exercidas pelos pulmões, mudanças nos músculos do peito, pescoço e cabeça e melhor reprodução dos formantes de modo a obter uma projecção mais eficaz.

As mudanças de amplitude e frequência dos formantes conferem-lhes a sua coloração espectral e definem-nos. Como indicador de boa qualidade de canto, a frequência e amplitude devem manter-se estáveis ao longo de um período de tempo, sendo que o seu espectro é bem definido. Estas são características a analisar no domínio da voz cantada. Existe ainda uma forma peculiar de cantar usada por muito cantores denominada vibrato, que consiste em aumentar e diminuir sucessivamente a frequência F0. O vibrato pode ser comparado ao *shimmer* e ao *jitter*, parâmetros referidos anteriormente usados com frequência em análises da voz [17].

Para obter a melhor modelização, de forma a obter os aspectos da voz a analisar, foram estudados vários algoritmos, baseados no trabalho desenvolvido por Katharine Murphy [11] e Alex Loscos [18]. Em ambos os trabalhos, mas especialmente no segundo, são abordados algoritmos e métodos de extracção dos aspectos a analisar na voz cantada.

Após uma análise das referências anteriores, concluiu-se que seria possível a análise da voz cantada segundo os seguintes parâmetros:

- *Loudness*;
- *Pitch*;
- Vibrato e tremolo;
- *Rough* e *growl*;
- Regularidade dos impulsos;
- Soprosidade (*Breathiness*);

Dos parâmetros anteriores, nem todos tem a mesma importância quando pretendemos analisar a voz cantada. Para além disso, alguns deles podem ser bastante complexos de obter

## 2.3 - Características e Parâmetros da Voz Cantada

e o seu resultado apresentar uma percentagem de erro elevada, o que faz com que uma selecção dos parâmetros desejados seja executada no âmbito desta dissertação.

O método de extracção destes parâmetros também terá que ser analisado numa fase posterior, recorrendo aos modelos matemáticos apresentados nas bibliografias anteriormente mencionadas. Como em alguns casos existe mais de uma forma computacional da sua implementação, estas terão que ser discutidas e avaliadas antes de a sua escolha ser feita. Tal está descrito no plano de trabalhos e também nas metodologias a adoptar deste relatório.

### 2.3.1 - Loudness

O *loudness* é uma sensação subjectiva de intensidade e descreve a sensação auditiva provocada pelo nível de intensidade à medida que as frequências variam. O *loudness* relaciona a intensidade de um som com a sua frequência. O *loudness* também difere consoante o meio ambiente. Isto deve-se ao facto de este ser um método de avaliação subjectiva, facilmente influenciada por outros sons externos. Podemos resumir que apesar de se relacionarem, a intensidade da voz não tem o mesmo significado que o *loudness*.

### 2.3.2 - Pitch

O *pitch* caracteriza-se pela sensação auditiva que temos sobre a altura da voz, podendo ser classificado em grave, médio ou agudo. Tal como no caso do *loudness*, o *pitch* tem uma relação muito próxima com a definição da frequência fundamental, mas não podemos definir o *pitch* como sendo a frequência fundamental.

Podemos relacionar de forma matemática a frequência fundamental de um som, com a nota equivalente na escala igualmente temperada através da expressão seguinte:

$$f = 2^{n/12} \times 440 \text{ Hz} \quad (2.1)$$

Em que o valor de  $n$  é um número inteiro e representa o valor de salto entre notas, a partir da referência utilizada, a nota A4, de valor  $n = 0$  e  $f = 440 \text{ Hz}$ . A distância de um semitom na escala igualmente temperada é de 100 cents. Podemos então concluir que 1200cents representam uma oitava, com uma relação de frequência de 2:1 [19].

É também usual representar uma nota através do seu valor MIDI, usando como referência a mesma nota A4, em que a forma matemática que traduz a sua relação com a frequência é a seguinte:

$$P = 69 + 12 \times \log_2 \frac{f}{440} \quad (2.2)$$

### 2.3.3 - Vibrato

O vibrato caracteriza-se por uma variação regular na frequência fundamental de uma nota musical. Podemos avaliar a qualidade do vibrato através de três parâmetros: frequência do vibrato, a extensão do vibrato e a regularidade do vibrato. Destes três, apenas a regularidade não tem uma medida concreta mas graficamente deve ter um aspecto sinusoidal. No bel canto, a frequência aceitável do vibrato varia de  $f = 5,5 \text{ Hz}$  a  $f = 7,5 \text{ Hz}$  e a extensão varia de  $\pm 6\%$  a  $\pm 12\%$ , ou seja, de aproximadamente  $\pm 1$  a  $\pm 2$  semitons (ST) [8].

Para se produzir um som com vibrato, temos que modular a tensão da musculatura laríngea, responsável pela tensão das pregas. É também possível de produzir um som com vibrato através da variação da pressão sub-glótica, traduzida num maior deslocamento lateral das pregas vocais e, conseqüentemente, da elevação da tensão média das pregas. Contudo, este ultimo meio não parece ser a técnica natural utilizada pelos cantores experientes [5][8]

## 2.4 - Estudo de plataformas existentes

Apesar da análise da voz cantada ser um ponto fundamental desta dissertação, este apenas representa metade do desafio proposto. O restante desafio centra-se no *feedback* de forma simples e eficaz de modo a melhorar a performance do utilizador. Como ponto de partida foram analisados vários programas (ver a Tabela 1) que efectuam um *feedback* da voz cantada de modo a ser efectuada uma avaliação de aspectos positivos e negativos. Será esta avaliação, efectuada posteriormente, que indicará o caminho a seguir na formatação do ambiente gráfico para que a interacção com o utilizador seja o mais eficaz possível.

Outras funcionalidades para além do *feedback* em tempo real poderão ser implementadas de forma a tornar o programa mais poderoso e aumentar a sua abrangência. Contudo, tal está limitado pelo tempo e pela sua complexidade de implementação.

Tabela 1 - Breve descrição dos programas existentes.

<i>Nome</i>	<i>Descrição de funcionalidades</i>
<i>Sing and See</i> [20], [21]	Informação em tempo real do <i>pitch</i> e do timbre vocal através de gráficos; Espectrograma do sinal e representação gráfica da frequência F0 e respectivos harmónicos; Comparação com a pauta musical e representação da amplitude da voz em semitons.
<i>Music MasterWorks</i> [22]	Representação do <i>pitch</i> em tempo real; Comparação com a pauta musical e possibilidade de comparação com uma melodia MIDI; Representação da amplitude da voz.
<i>SingingStudio</i> [1]	Informação em tempo real do <i>pitch</i> e do timbre vocal através de gráficos; Comparação com a escala musical com possibilidade de comparação com uma melodia MIDI de forma a possibilitar uma pontuação; Representação gráfica da frequência F0 e respectivos harmónicos; Visualização do espectrograma numa futura versão já prevista;

## 2.4 - Estudo de plataformas existentes

<i>SingStar</i> [23]	<p>Não é um software de computador de ajuda ao canto, mas sim um jogo para uma consola;</p> <p>Sistema de pontuação baseado no <i>pitch</i> da voz, através da comparação com um valor pretendido;</p> <p>Informação em tempo real e comparação com a nota musical pretendida;</p>
<i>Carry-a-tune</i> <i>SingingCoach</i> [24]	<p>Um dos primeiros softwares disponíveis no mercado mas que neste momento não é oficialmente vendido nem actualizado; contudo ainda é possível comprar-se online [25];</p> <p>Visualização em tempo real do <i>pitch</i> e amplitude e comparação com a pauta musical.</p>
<i>Xitona</i> <i>SingingTutor</i> [26]	<p>Visualização em tempo real do <i>pitch</i>;</p> <p>Possibilidade de comparação com notas MIDI introduzidas manualmente e com a pauta musical.</p>
<i>Singing Superstar</i> [27]	<p>Funcionamento semelhante ao jogo SingStar, mas para computador;</p> <p>Possibilidade de leitura de músicas de karaoke, em que o <i>pitch</i> é avaliado em tempo real e comparado com o pretendido;</p> <p>Avaliação da performance através da comparação do <i>pitch</i>.</p>
<i>Music Tutor</i> [28]	<p>Software descontinuado e oficialmente sem suporte, contudo é possível obter-se online [29];</p> <p>Visualização em tempo-real do <i>pitch</i> e comparação com a pauta musical e nota MIDI pretendida para uma melhor avaliação.</p>

Após o levantamento de funcionalidades descrito na tabela anterior, podemos constatar que os programas mais simples apresentam muito pouca informação ao utilizador. Contudo, a informação apresentada, o *pitch*, é um dos parâmetros mais importantes e também dos mais simples de obter.

## Capítulo 3

# Estimação da Frequência Fundamental

Uma das conclusões que podemos retirar do capítulo anterior é a importância da análise do *pitch* no âmbito da voz cantada. Contudo, o *pitch* é uma medida psicoacústica e como tal, não é directamente mensurável. De forma sucinta, podemos dizer que o *pitch* é a percepção por parte do nosso cérebro da frequência fundamental, reforçada pelos restantes harmónicos existentes. A análise efectuada por cada individuo será por isso diferente e de difícil quantificação. Porém, a inerente relação entre o *pitch* e a frequência fundamental mostra que existe uma forma objectiva de efectuar uma caracterização e eleva a importância da estimação de F0 de forma robusta e precisa por parte do *software* de análise da voz cantada.

### 3.1 - Definição de Frequência Fundamental

A frequência fundamental define-se como o valor de frequência mais baixa numa estrutura harmónica ao qual se relacionam os harmónicos de uma onda periódica. A estimação de F0 tem sido estudada ao longo dos últimos anos e apesar dos inúmeros avanços, apresenta ainda algumas limitações, sendo o principal a robustez do algoritmo quando usado em vários contextos. A maioria dos estimadores usados funciona bastante bem para o contexto planeado, ou seja, num único contexto. Quando são utilizados em diferentes condições, contendo o sinal mais ruído ou até diferentes fontes, é difícil de encontrar um algoritmo que mantenha a precisão. Os diferentes algoritmos vão ser abordados nas secções seguintes, contudo existem duas abordagens possíveis de estimar a frequência fundamental: método temporal e método espectral.

A melhor forma de ilustrar o significado da frequência fundamental é através da análise da forma de onda de um sinal de características conhecidas, quer no domínio dos tempos como no das frequências, seguido de uma comparação de ambos. A melhor representação de um sinal no domínio das frequências é obtida com recurso a um espectrograma, onde o sinal é representado segundo a sua intensidade nas diversas frequências ao longo tempo. Para tornar mais fácil essa análise, também é recorrente executar-se uma visualização do sinal mas em que apenas são representadas a amplitude das frequências do sinal. Estas três representações estão ilustradas nas três figuras seguintes, em que o sinal a analisar é de origem conhecida e a sua frequência fundamental é de 440 Hz.

### 3.1 - Definição de Frequência Fundamental

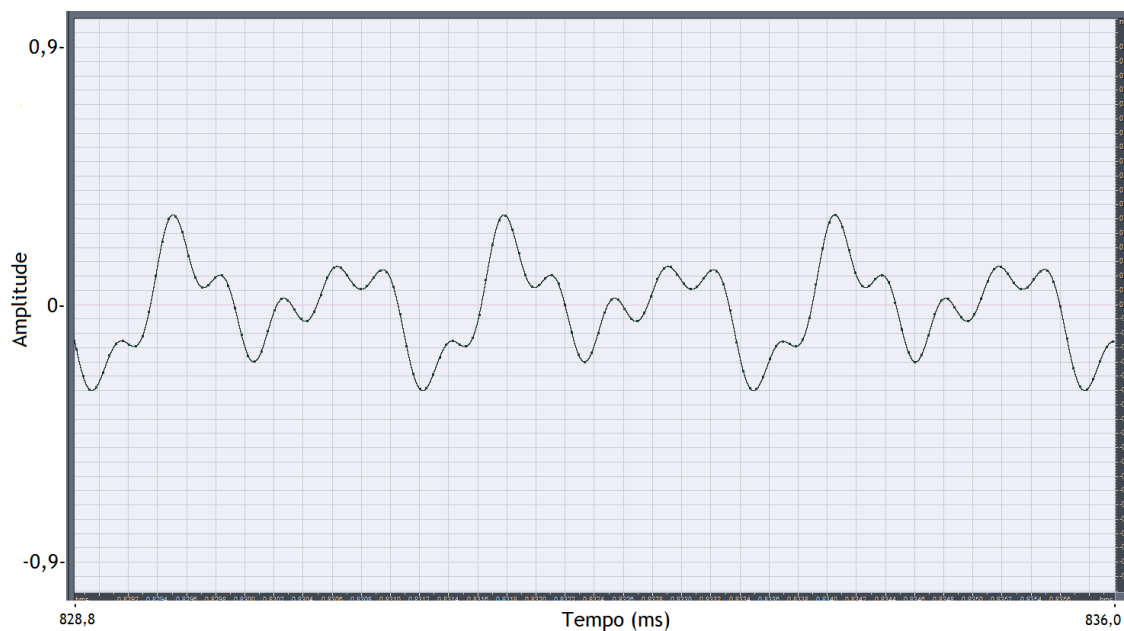


Figura 6 - Representação temporal de uma nota musical com F0 440Hz.

A figura 6 representa uma pequena secção de um sinal de voz cantada com frequência fundamental 440Hz e que foi sintetizado através da ferramenta *Madde* [30]. A forma de onda representada na figura acima e seguintes foi obtida através do *software* de análise *Adobe Audition* [31]. A periodicidade da forma de onda é notória, o que tornaria acessível a extracção da frequência fundamental por meios temporais.

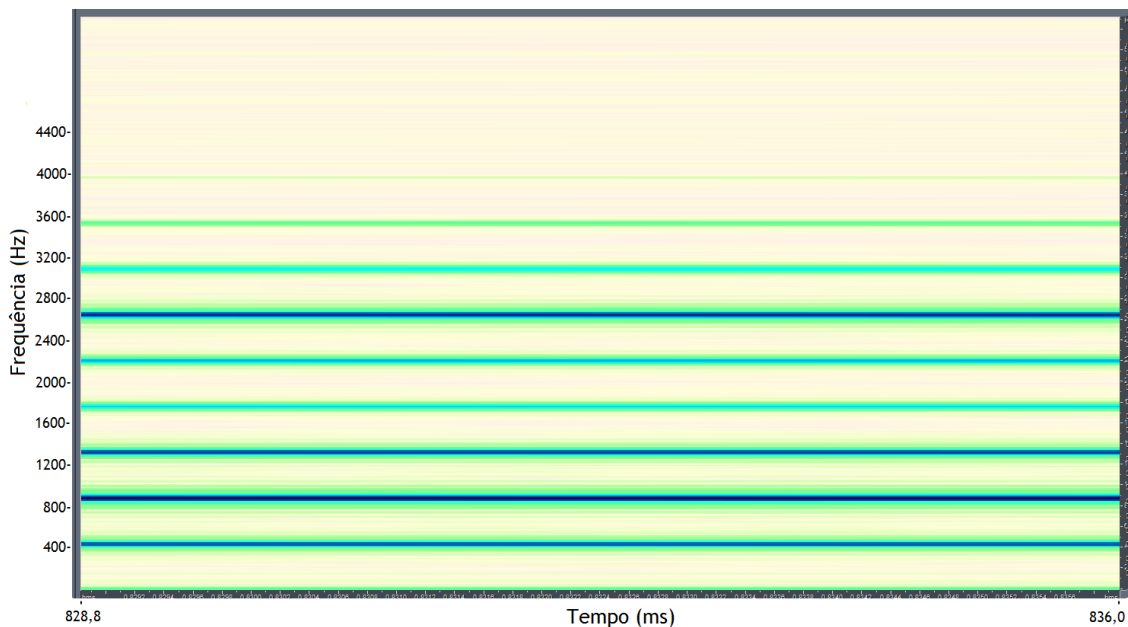
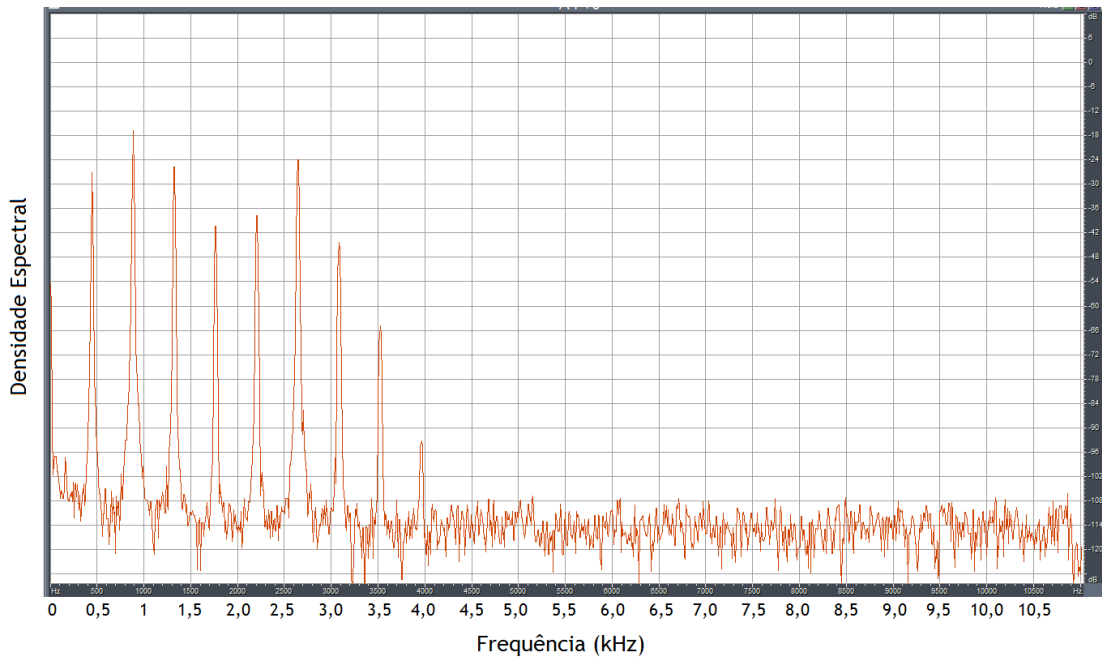


Figura 7 - Espectrograma de uma voz sintetizada com F0 de valor 440Hz.

Através do espectrograma representado na figura 7 identificamos claramente a frequência fundamental do sinal, visto esta estar ser identificada pela primeira linha a partir da referência. A densidade espectral está codificada através da cor segundo o eixo dos  $z$ , sendo tanto mais escura quanto mais intensa for aquela.

Se o mesmo espectrograma fosse visualizado de uma outra forma, tendo em conta os eixos da amplitude de frequência e o valor da frequência, obteríamos o gráfico da figura 8.



**Figura 8** - Análise de frequências de uma voz sintetizada com F0 de valor 440Hz.

Nesta visualização, o valor da frequência fundamental é representado pelo primeiro máximo local, excluindo valores na origem.

Convém notar que visto o sinal não ser de origem natural, isto é, não foi obtido com recurso a um orador, explica a ausência de perturbações no sinal. Contudo, um dos propósitos desta análise foi o de excluir qualquer subjectividade que pudesse ser conferida ao sinal, tornando o sinal o mais ideal possível.

O objectivo destas representações é de apenas ilustrar o sinal a ser analisado pelas duas abordagens de estimação de F0, que são o método temporal e o método espectral. Ambos os métodos serão explicados nas secções seguintes.

## 3.2 - Análise no domínio dos Tempos

O método temporal é um método mais simples que o espectral pois limita-se a tentar avaliar a frequência fundamental através do sinal original ao longo do tempo. Existem várias técnicas para o fazer, com as suas vantagens e desvantagens, mas como veremos mais adiante, os métodos mais robustos não utilizam uma mas várias técnicas sendo imperativo um aprofundamento do funcionamento de todas as técnicas envolvidas. O seu estudo encontra-se nas secções seguintes.

### 3.2.1 - Eventos temporais

Existem três meios eficazes de detectar os eventos temporais pretendidos:

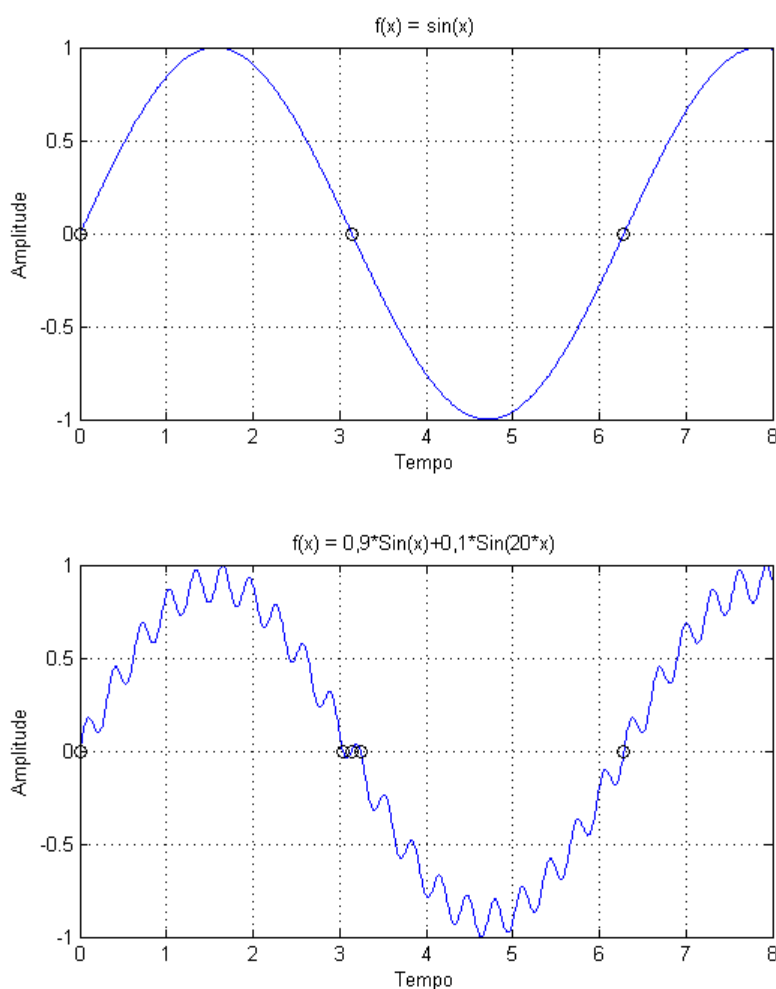
- Passagens por zero, conhecidos como *Zero-Crossing Rate*;
- Máximos e mínimos locais, ou *Peak detection*;
- Declive da onda periódica ou *Slope Event Rate*.

## 3.2 - Análise no domínio dos Tempos

Antes de os descrevermos com mais pormenor, é importante realçar que todos eles são métodos simples e com uma carga computacional baixa. São muito úteis quando sabemos a natureza do sinal e como meio de tornar mais robusto um algoritmo em determinados contextos. Contudo, não é de esperar que resulte um estimador de F0 robusto e preciso na análise da voz cantada recorrendo simplesmente ao uso destes métodos. Tendo em conta a riqueza em harmónicos da voz cantada, é de esperar que nestes casos a precisão destes métodos seja bastante limitada, como será ilustrado a seguir.

### 3.2.2 - Zero-Crossing Rate

A abordagem seguida pelos ZCR, tal como o seu nome indica, é a de estimar o número de vezes que a onda passa por zero, por unidade de tempo. É uma abordagem simplista e que se prendia com a relação da periodicidade da onda e a frequência fundamental do sinal [32]. É fácil perceber que se a potência espectral de um sinal se encontra em torno de F0, a onda passará por zero duas vezes num período. Contudo, em casos em que existam componentes elevadas nas altas frequências, o sinal vai cruzar zero mais que duas vezes num período, não sendo possível estimar com precisão o valor de F0. A figura seguinte ilustra as duas situações descritas.



**Figura 9** - A influência de harmónicos às altas frequências no ZCR.

Uma forma de ultrapassar esta limitação seria a de filtrar o sinal, contudo estaríamos a retirar harmónicos que poderiam ser importantes e como tal não é um processo recomendado.

### 3.2.3 - *Peak Detection*

Uma segunda forma da detecção de eventos seria através dos máximos ou mínimos locais. Devido à natureza sinusoidal e periódica da forma de onda, sabemos que num período existirão um máximo e mínimo. A relação entre o número de máximos ou mínimos num intervalo de tempo ou o intervalo de tempo entre a ocorrência de máximos ou mínimos são formas de obter a frequência fundamental.

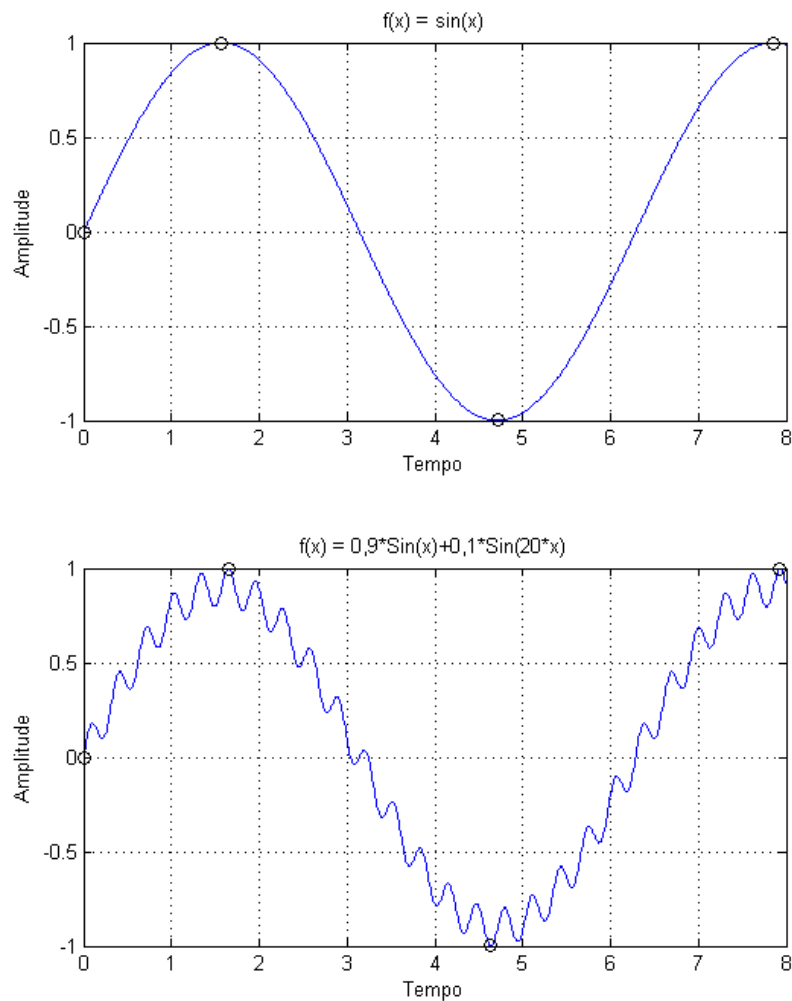


Figura 10 - Detecção de máximos e mínimos de uma forma de onda.

Os resultados do método *peak detection* podem ser comparados com o método ZCR através das figuras 8 e 9. Apesar de neste último caso os máximos e mínimos locais serem de fácil análise, convém lembrar que este sinal é sintetizado a partir de uma onda sinusoidal. Trata-se, por isso, de uma situação ideal que visa apenas a demonstração dos métodos e não a sua precisão como detecção de F0 de uma voz natural.

### 3.2.4 - Slope Event Rate

Sendo a forma do sinal de ordem periódica, a mudança de sinal associada ao declive será, também, periódica. A partir desta mudança de sinal, podemos caracterizar a frequência fundamental, partindo de uma abordagem em tudo semelhante à do método ZCR. Tal como os outros dois processos descritos anteriormente, a robustez não é uma mais-valia, mas poderá revelar-se mais útil em diferentes contextos.

### 3.2.5 - Autocorrelação

O objectivo da correlação entre duas formas de onda é o de obter uma medida da sua semelhança. A autocorrelação, tal como o nome implica, é a correlação entre uma forma de onda e ela própria mas afectada de um deslocamento. A autocorrelação de uma forma de onda periódica tem a particularidade de também ela ser periódica. A título de exemplo, representa-se na figura 11 a autocorrelação de dois períodos de uma onda sinusoidal.

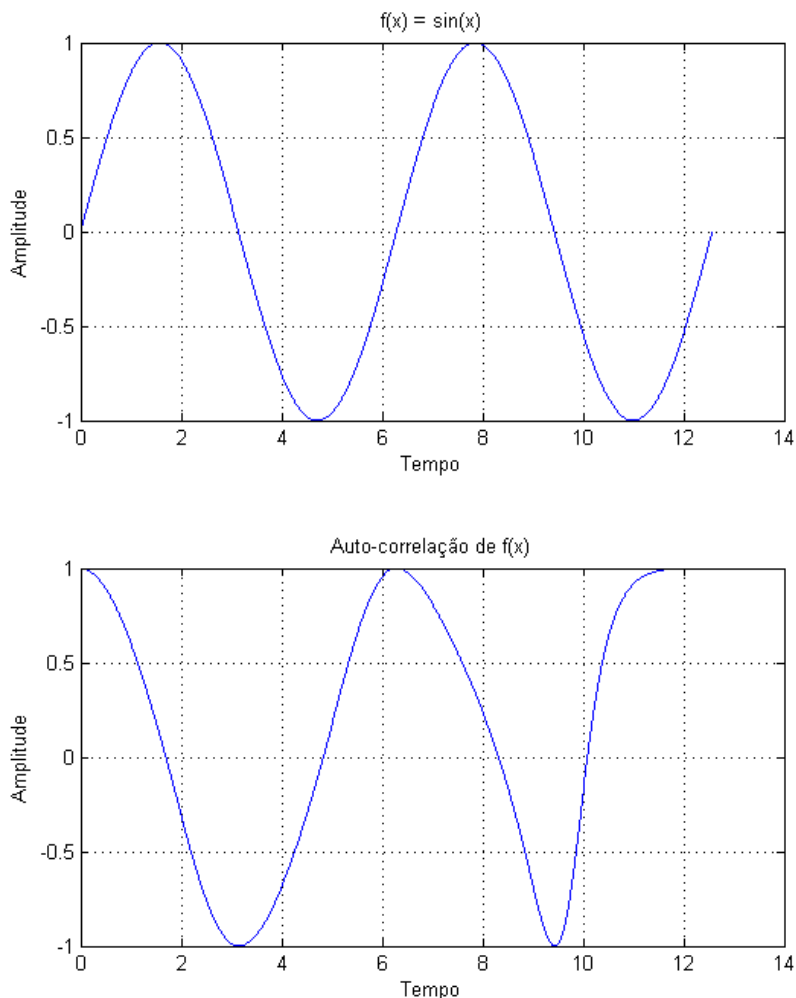


Figura 11 - Forma de onda sinusoidal e correspondente autocorrelação.

A representação anterior apenas tem como único objectivo a demonstração da periodicidade da função de autocorrelação, quando o próprio sinal também periódico.

A função de autocorrelação de um sinal estacionário pode ser calculada através da expressão matemática seguinte:

$$R_x(\tau) = \lim_{t_0 \rightarrow \infty} \frac{1}{t_0} \int_{-t_0/2}^{t_0/2} x(t + \tau)x(t)dt \quad (3.1)$$

Se o sinal for periódico, com período  $T$ , a sua função de autocorrelação também será periódica com o mesmo período.

Uma das particularidades inerentes a este método é o facto de existir um máximo global quando  $\tau = 0$ . Se existir um novo máximo global num instante  $T_0$ , significa que o sinal original é periódico e que os restantes máximos da função autocorrelação estão situados em  $nT_0$ , sendo  $n$  um número inteiro. A partir do valor de  $T_0$ , podemos obter o valor da frequência fundamental visto esta ser o seu inverso. As mesmas propriedades aplicam-se a um sinal discreto, sendo a expressão matemática a seguinte, se considerarmos um sinal  $x[n]$  com duração infinita.

$$R_x(\nu) = \sum_{n=-\infty}^{\infty} x[n]x[n + \nu] \quad (3.2)$$

Se o sinal  $x[n]$  estiver contido num intervalo de tempo finito de duração  $N$ , a expressão da função de autocorrelação será:

$$R_{x'}(\nu) = \sum_{n=0}^{N-1-\nu} x'[n]x'[n + \nu] \quad (3.3)$$

Um dos principais problemas surge quando a função de autocorrelação é usada num sinal harmónico complexo com pseudoperiodicidade, o que muitas vezes acontece na análise da voz cantada. Trata-se por isso, de um método eficaz mas por si só tem uma aplicação limitada. Na próxima secção vamos abordar um algoritmo conceituado, que se baseia na autocorrelação e em mais métodos de forma a tornar o método mais robusto e preciso.

### 3.2.6 - Algoritmo *Boersma*

O algoritmo *Boersma* [33] é um método robusto de análise de F0, e a sua escolha para este estudo deve-se ao facto de este ser o método usado para a estimação da frequência fundamental numa ferramenta que é referência nesta área: *Praat*. Tal como foi mencionado anteriormente, este algoritmo simples é baseado no conceito da autocorrelação mas assenta num conceito mais preciso. Voltando um pouco atrás, quando referimos que a existência de um máximo local num instante  $T_0$  significa que o sinal original é periódico, não definimos o seu significado quando tal não acontece, ou seja, quando não existe um máximo local noutro instante que não em  $\tau = 0$ . Efectivamente a função autocorrelação exibe sempre um máximo global em  $\tau = 0$ . Se ocorrer um máximo local quando  $\tau_{max}$  e se o valor da função nesse ponto  $R_x(\tau_{max})$  é suficientemente elevado, podemos afirmar que o sinal original têm uma

## 3.2 - Análise no domínio dos Tempos

componente periódica, sendo a sua componente harmónica  $R_0$  um número entre 0 e 1, igual ao máximo local  $r'_x(\tau_{max})$  da função de autocorrelação normalizada:

$$r'_x(\tau) \equiv \frac{r_x(\tau)}{r_x(0)} \quad (3.4)$$

Podemos criar um sinal  $x(t)$ , composto por um sinal periódico  $H(t)$  ao qual se adicionou ruído  $N(t)$ , obtendo o sinal  $x(t) = H(t) + N(t)$ . Se estes sinais não forem correlacionados, podemos efectuar a autocorrelação entre cada um deles visto que nesta situação a autocorrelação de sinal é equivalente à soma da autocorrelação das suas partes. Se o ruído for branco, isto é, não se correlaciona com ele próprio, excepto na origem, a função de autocorrelação irá conter um máximo local  $\tau_{max} = T_0$ , com valor  $r_x(\tau_{max}) = r_H(T_0)$ . Como a função de autocorrelação no ponto inicial representa a potência do sinal, a função de autocorrelação normalizada no ponto  $\tau_{max}$  representará a potência relativa da componente periódica do sinal e o seu complemento representará a potência relativa da componente associada ao ruído, descritas nas equações seguintes:

$$\frac{r_H(\tau_{max})}{r_x(0)} = r'_x(\tau_{max}) \quad (3.5)$$

$$\frac{r_N(0)}{r_x(0)} = 1 - \frac{r_H(\tau_{max})}{r_x(0)} = 1 - r'_x(\tau_{max}) \quad (3.6)$$

A partir destas duas expressões é possível definir a relação logarítmica *harmonics-to-noise ratio*:

$$HNR(dB) = 10 \cdot \log_{10} \frac{r'_x(\tau_{max})}{1 - r'_x(\tau_{max})} \quad (3.7)$$

O algoritmo *Boersma* assenta neste princípio base para determinar a frequência fundamental de um sinal. A sua ideia é semelhante à abordagem efectuada pelos métodos espectrais. Em termos teóricos, é um método bastante robusto e preciso, visto o HNR ser infinito para um sinal perfeitamente periódico. No mundo real, estes resultados podem não ser os desejados, e por ser uma referência, será usado nos testes de comparação no respectivo capítulo.

De realçar que o seu algoritmo computacional é um pouco mais complexo, envolvendo um método de *windowing* de forma a melhorar os seus resultados. Contudo o objectivo pretendido neste capítulo é o de analisar as filosofias e abordagens no qual assentam os algoritmos em questão.

### 3.2.7 - Yin Estimator

O algoritmo *Yin* [34] também tem como base de análise de F0 a autocorrelação. Contudo, implementa outros métodos de forma a melhorar o seu desempenho, relativamente ao algoritmo *Boersma*. Existem dois passos muito importantes segundo o autor que melhoram

substancialmente o seu desempenho quando comparado com outros métodos também baseados na autocorrelação:

- Uso de uma função *cumulative mean normalized difference* em vez de uma simples função de diferenças;
- Execução de uma interpolação parabólica de modo a aumentar a precisão.

A melhor forma de avaliarmos o impacto dos melhoramentos do algoritmo é recorrendo à tabela seguinte, que representa os passos efectuados pelo algoritmo e o erro associado à medida que os passos são aplicados.

Tabela 2 - Passos executados e erro do algoritmo ao longo da sua implementação [32].

<i>Implementação</i>	<i>Erro (%)</i>
Autocorrelação	10,0
Função das diferenças	1,95
<i>cumulative mean normalized difference</i>	1,69
<i>Absolute threshold</i>	0,78
Interpolação parabólica	0,77
Estimação do melhor valor	0,5

### 3.3 - Análise no domínio das Frequências

O domínio das frequências contém muita informação relevante na análise da frequência fundamental de um sinal, neste caso concreto, a voz cantada. O sinal extraído de uma voz cantada natural é composto por uma série de parciais, harmonicamente relacionados entre si. Estes parciais permitem uma identificação e consequente extracção do valor de F0 do sinal analisado. Existem, tal como no domínio dos tempos, vários métodos de implementação. Nas secções seguintes são abordados alguns dos métodos mais utilizados e um algoritmo bastante recente: *SearchTonal*.

#### 3.3.1 -Component Frequency Ratios

Um dos métodos pioneiros na análise da voz no domínio das frequências data de 1979, quando Martin Piszczalski trabalhava para implementar um sistema automático de transcrição de música. Um dos componentes fundamentais seria a da detecção e extracção do valor de F0 [35][36]. O seu método de estimação de F0 consiste em executar uma transformada de forma ao sinal original poder ser analisado no domínio das frequências. De seguida, são estimados os máximos locais, através de um método de detecção de picos do sinal. Para cada um destes parciais encontrados, o algoritmo estima o menor número harmónico possível que corresponderia a uma série harmónica que contivesse esses parciais. Tendo encontrado valores para vários pares de harmónicos, os seus resultados seriam pesados e avaliados de forma a encontrar a frequência fundamental do sinal. O peso de cada um dos parciais está relacionado com a amplitude do sinal: pares com maior amplitude tem um peso maior que pares com menor amplitude.

Uma das vantagens óbvias desta abordagem é a de que a própria frequência fundamental poderia até não estar contida no sinal, não deixando mesmo assim de ser detectada,

### 3.3 - Análise no domínio das Frequências

bastando para isso a existência no sinal de pares de harmónicos suficientes. Existem versões melhoradas e com melhores resultados, mas a ideia inerente à abordagem permanece a mesma.

#### 3.3.2 - Métodos baseados em filtros

O uso de filtros na estimação de F0 é implementada recorrendo à comparação do resultado da implementação de filtros com diferentes frequências centrais. A ideia base é a da que quando existe um pico espectral do sinal na zona central do filtro, o resultado é visível na sua saída. Com base no conhecimento da frequência central do filtro, é fácil extrapolar o valor de F0.

As duas implementações mais robustas com base neste método recorrem a duas utilizações diferentes dos filtros. Uma das abordagens seria a implementação de filtros passa-banda com uma certa frequência central, espaçados entre si com o valor da frequência central do primeiro filtro. Do ponto de vista computacional, este método é bastante intenso e apresenta também um problema sério que se revela quando o sinal original apenas apresenta um parcial harmónico [37].

Uma outra abordagem seria a da implementação de um filtro passa-banda deslizante ao longo da frequência. A ideia base é em muito semelhante ao funcionamento de um estroboscópio, presente em alguns leitores de discos “vinil” ou mais semelhante ainda, ao estroboscópio presente na ferramenta usada para afinar pianos [38]. Seria um método menos intenso a nível de computação mas o facto de o sistema necessitar de distinguir o que seriam perturbações dos pares de harmónicos, especialmente se se tratarem de perturbações também harmónicas como instrumentos de fundo, é ainda um problema.

#### 3.3.3 - Análise Cepstral

A análise cepstral ou *Cepstrum*, tem origem em 1963 por Bogert et. al. [39] como forma de melhorar um sistema de detecção de sismos baseado no método da autocorrelação. O seu nome deriva do facto de este ser, em síntese, um espectro de um espectro e daí a inversão das primeiras quatro letras da palavra *Spectrum* para *Cepstrum*. A sua definição exacta é: a transformada inversa de Fourier do logaritmo do espectro.

A expressão matemática do cepstro de potências é

$$C_{AA}(\tau) = \mathcal{F}^{-1}\{\log S_{AA}(f)\}, \quad (3.8)$$

onde  $S_{AA}(f)$  é um auto espectro médio de potências, bilateral

$$S_{AA}(f) = \overline{|\mathcal{F}\{a(t)\}|^2} \quad (3.9)$$

Em termos de implementação como algoritmo, o *cepstrum* pode ser representado como ilustra a figura 12.

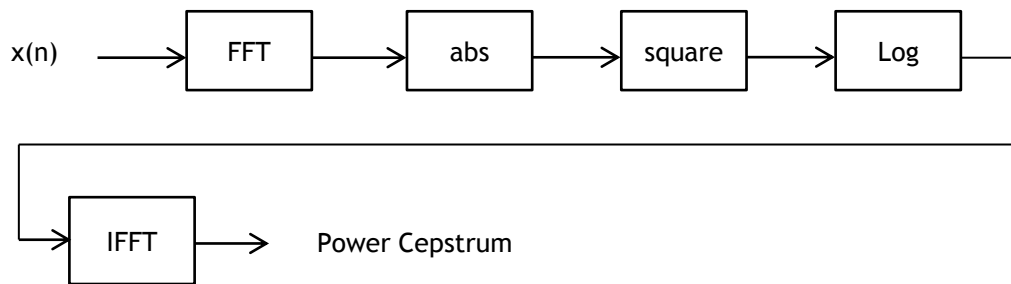


Figura 12 - Algoritmo de implementação do *Cepstrum*.

Devido à sua natureza, o cepstro tem vantagens importantes na análise da voz pois permite detectar estruturas periódicas no espectro logarítmico, como por exemplo, famílias de harmónicos e ou bandas laterais com espaçamento uniforme. É, por isso, um método muito usado na análise da voz, principalmente na identificação de regiões do sinal vozeadas ou não vozeadas, ou como método de reforço de estimação de  $F_0$ . Não é, contudo, usado como único método na extracção da frequência fundamental pois o cepstro assume que o sinal contém parciais harmónicos distanciados igualmente entre si, o que por vezes não acontece e nestes casos os resultados não serão precisos.

### 3.3.4 -Métodos multi-resolucionais

Os métodos multi-resolucionais representam uma abordagem ao problema que se prende com a diversidade de sinais que se pretende analisar. A ideia central prende-se com o facto de que se um dado método usado para determinar  $F_0$  não funciona como o esperado, poderemos executar o mesmo cálculo aplicando resoluções diferentes, confirmando assim os valores de  $F_0$  determinados [40].

Para implementar este método, bastaria recalcular o espectro usando uma janela com diferentes valores. Contudo, esta medida tem um custo muito elevado a nível computacional o que a torna pouco utilizável.

### 3.3.5 -Algoritmo *SearchTonal*

O método *SearchTonal* resulta da evolução de um algoritmo desenvolvido inicialmente por Aníbal Ferreira durante um outro projecto. Este algoritmo foi desde então melhorado de forma a estimar a frequência fundamental de um sinal de forma robusta e precisa. Podemos afirmar com certeza que este algoritmo foi realmente dimensionado de forma a obter uma precisão impar na estimação da frequência fundamental, não encontrada noutro método utilizado nesta função.

O seu funcionamento é descrito nas figuras 13 e 14, onde são representados diagramas de blocos das operações efectuadas durante o processamento do sinal.

### 3.3 - Análise no domínio das Frequências

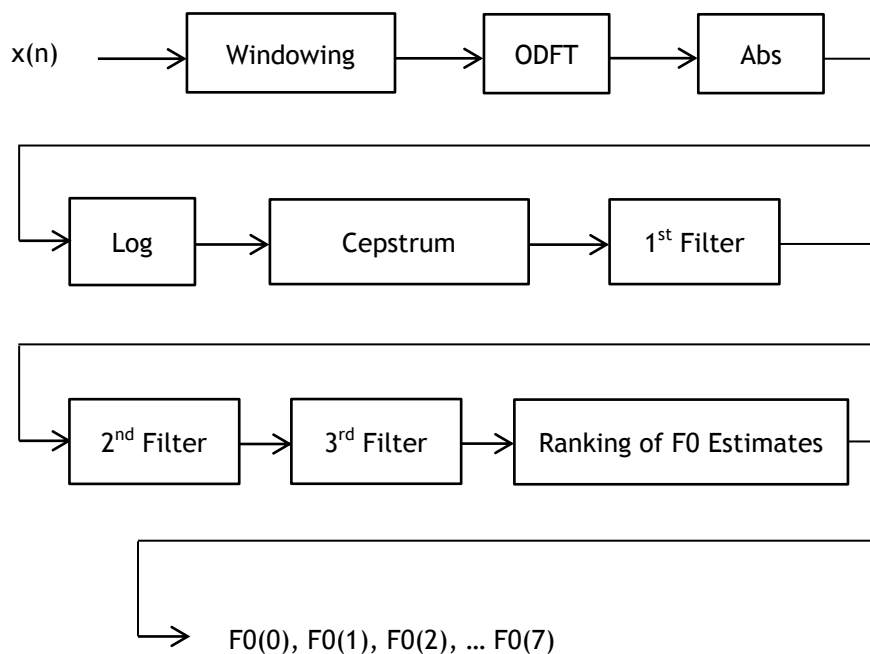


Figura 13 - Diagrama de Blocos de alto nível do algoritmo *SearchTonal*.

Este algoritmo, contudo, não se baseia apenas no exposto no diagrama de blocos apresentado na figura 13. Após ter estimado os oito valores mais prováveis de serem identificados como o valor de  $F0$ , através da análise cepstral, o algoritmo implementa um método de selecção do valor mais provável de  $F0$ .

A figura 14 apresentada na página seguinte ilustra o processo de escolha do valor da frequência fundamental.

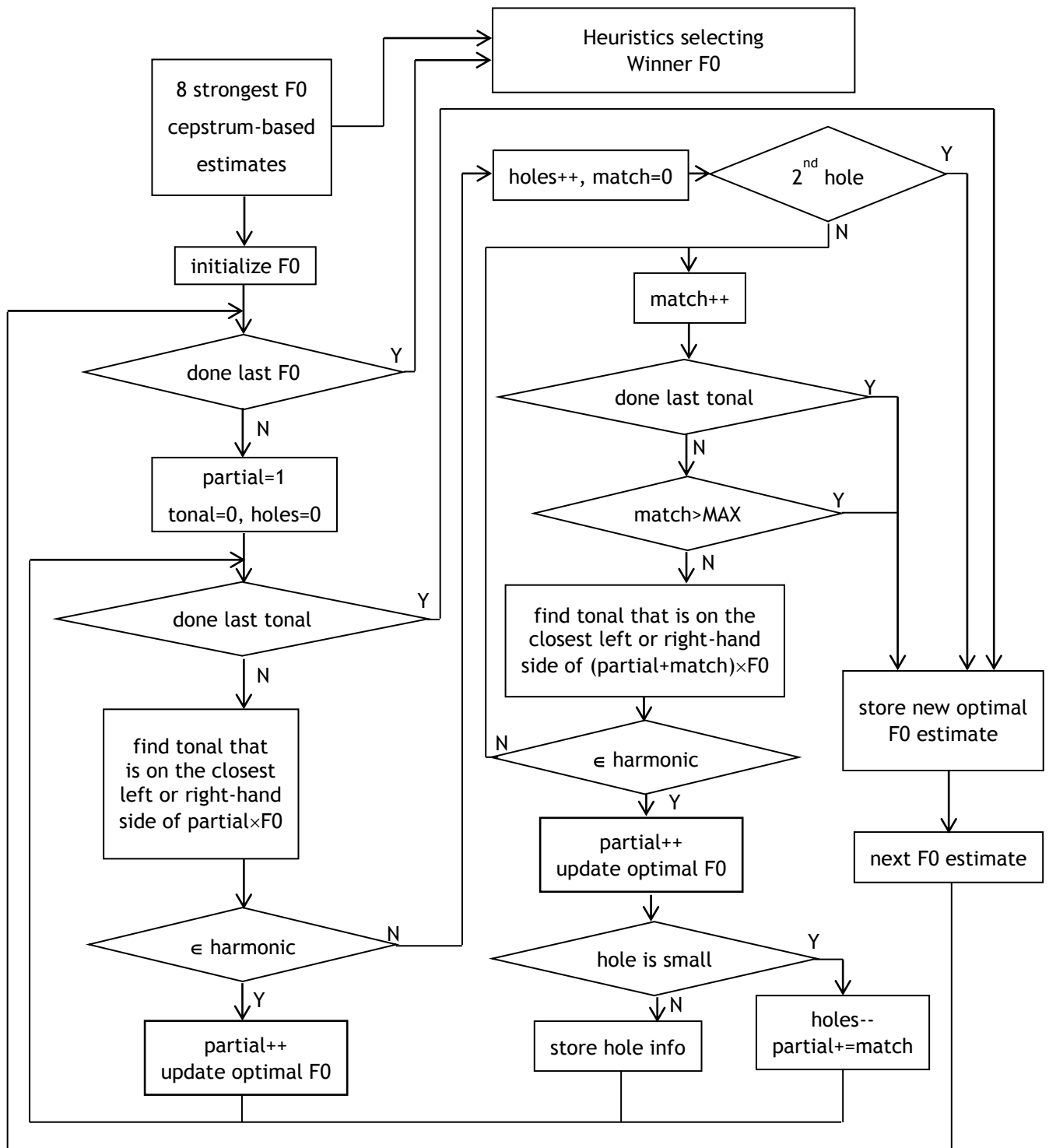


Figura 14 - Fluxograma do método de selecção de F0 por parte do algoritmo SearchTonal.

### 3.3 - Análise no domínio das Frequências

Numa segunda fase de processamento, ilustrado na figura 14, o algoritmo *SearchTonal* utiliza os resultados da análise cepstral para determinar com mais precisão o valor de F0 do sinal, não só através do próprio valor de F0, mas também de harmónicos i.e. parciais da estrutura harmónica do sinal.

# Capítulo 4

## Estimador de Vibrato

O vibrato é uma das características mais importantes num cantor e é um objectivo da dissertação efectuar uma estimação robusta e precisa do mesmo. Tal como já foi descrito no segundo capítulo, o vibrato é caracterizado principalmente por uma variação na frequência fundamental, não sendo apesar disso o único parâmetro que define o vibrato. O objectivo deste capítulo visa demonstrar o funcionamento do algoritmo desenvolvido no âmbito da dissertação, explicando o seu funcionamento de forma detalhada e concisa.

### 4.1 - Definição de Vibrato

O vibrato, tal como foi referido no segundo capítulo, define-se como a variação da frequência fundamental de forma periódica.

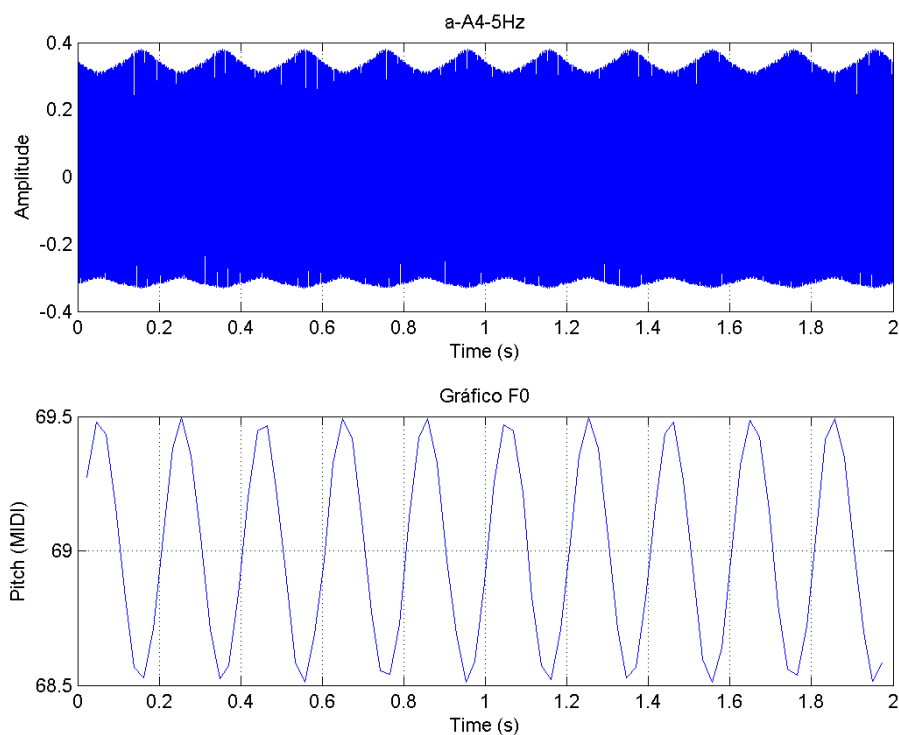


Figura 15 - Voz sintetizada com vibrato.

## 4.2 - Estrutura do algoritmo

A figura 15 ilustra uma voz sintetizada com um valor de  $F_0$  igual a 440Hz, sendo claramente visível a sua variação de forma periódica. A frequência do sinal de  $F_0$  define a frequência do vibrato, que no caso representado na figura 15 é de 5Hz. O vibrato também é caracterizado pela sua extensão, que se define como a amplitude da variação do valor da frequência fundamental. Na ilustração da figura 15 podemos visualizar a extensão do vibrato, que toma valores de  $0,5ST$ .

De forma a efectuar uma caracterização do vibrato é, então, necessário efectuar a sua detecção e extracção dos valores de frequência, extensão e duração. Nas secções seguintes são descritos princípios utilizados e algoritmos desenvolvidos de forma a elaborar o algoritmo estimador do vibrato.

## 4.2 - Estrutura do algoritmo

O algoritmo que estima o vibrato tem como sinal de entrada a frequência fundamental de um sinal, o que implica que o seu bom funcionamento está, por isso, ligado à precisão e robustez do algoritmo que estima  $F_0$ . O algoritmo usado pelo *SingingStudio* de forma a estimar  $F_0$  foi o *SearchTonal*, algoritmo já analisado e discutido no capítulo anterior. O seu desempenho também foi avaliado e será alvo de análise no quinto e sexto capítulo desta dissertação. O diagrama de alto do nível pode ser visualizado na figura 16.

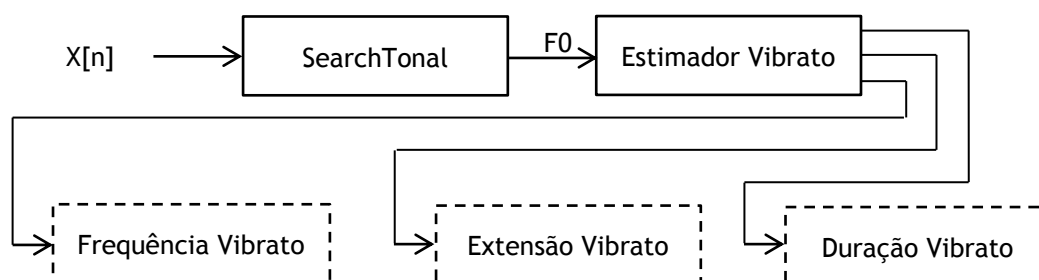


Figura 16 - Implementação do estimador de Vibrato.

Através da figura 16, podemos constatar que o estimador de Vibrato devolve os três parâmetros associados ao vibrato: frequência, extensão e duração. É uma finalidade do algoritmo estimar com precisão estes três parâmetros. Nas secções seguintes, os métodos associados à análise destes parâmetros vão ser abordados um de cada vez para que a sua abordagem seja perceptível.

A figura 17 que se encontra na página seguinte serve para ilustrar os passos de maior relevância associados ao estimador de vibrato.

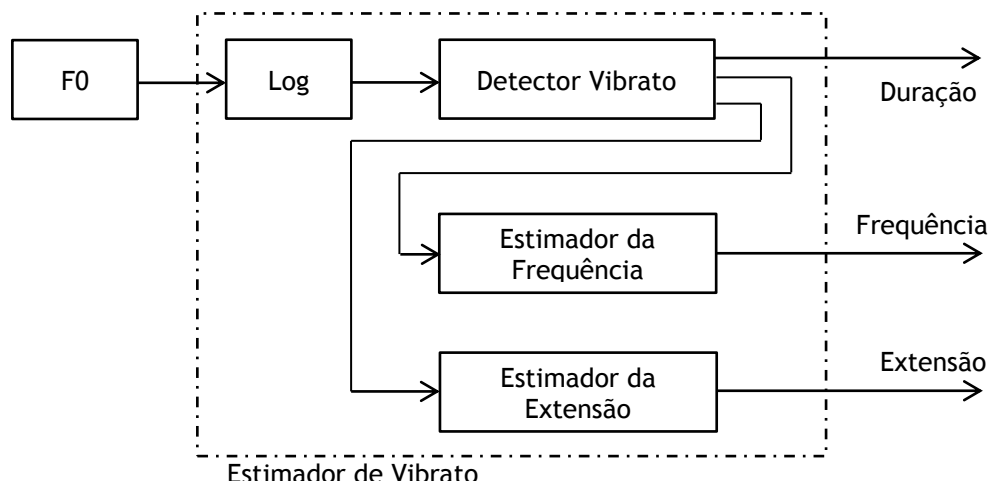


Figura 17 - Diagrama de alto nível do estimador de Vibrato.

Um factor que pode facilmente ser extrapolado da figura 17 é que o detector de vibrato tem uma importância muito elevada pois para além de efectuar a detecção, indica também a duração do vibrato. Se o detector não tiver precisão suficiente, origina erros grosseiros, visto que o algoritmo estaria a tentar determinar os parâmetros do vibrato numa zona de sinal onde este não existe. Também poderia acontecer uma zona do sinal que contém vibrato não ser detectado, o que também originaria uma falha grosseira do algoritmo. Na secção seguinte o funcionamento do detector será descrito pormenorizadamente.

### 4.3 - Detecção e duração do Vibrato

Nesta secção vai ser demonstrado o funcionamento do detector do vibrato, incluindo a fundamentação de escolhas efectuadas no mesmo. A detecção de uma região de vibrato contida no sinal de  $F_0$  é efectuada tendo em conta vários parâmetros para que esta se torne mais robusta perante sinais de diferentes características.

Na figura 18 é ilustrada uma representação do algoritmo de detecção. As partes que o constituem serão abordadas individualmente, recorrendo-se a equações, diagramas e figuras de modo a clarificar o seu funcionamento. De realçar que alguns dos seus conceitos já foram abordados quando o estudo da frequência fundamental, presente no capítulo anterior, e não serão por isso alvo de nova análise.

### 4.3 - Detecção e duração do Vibrato

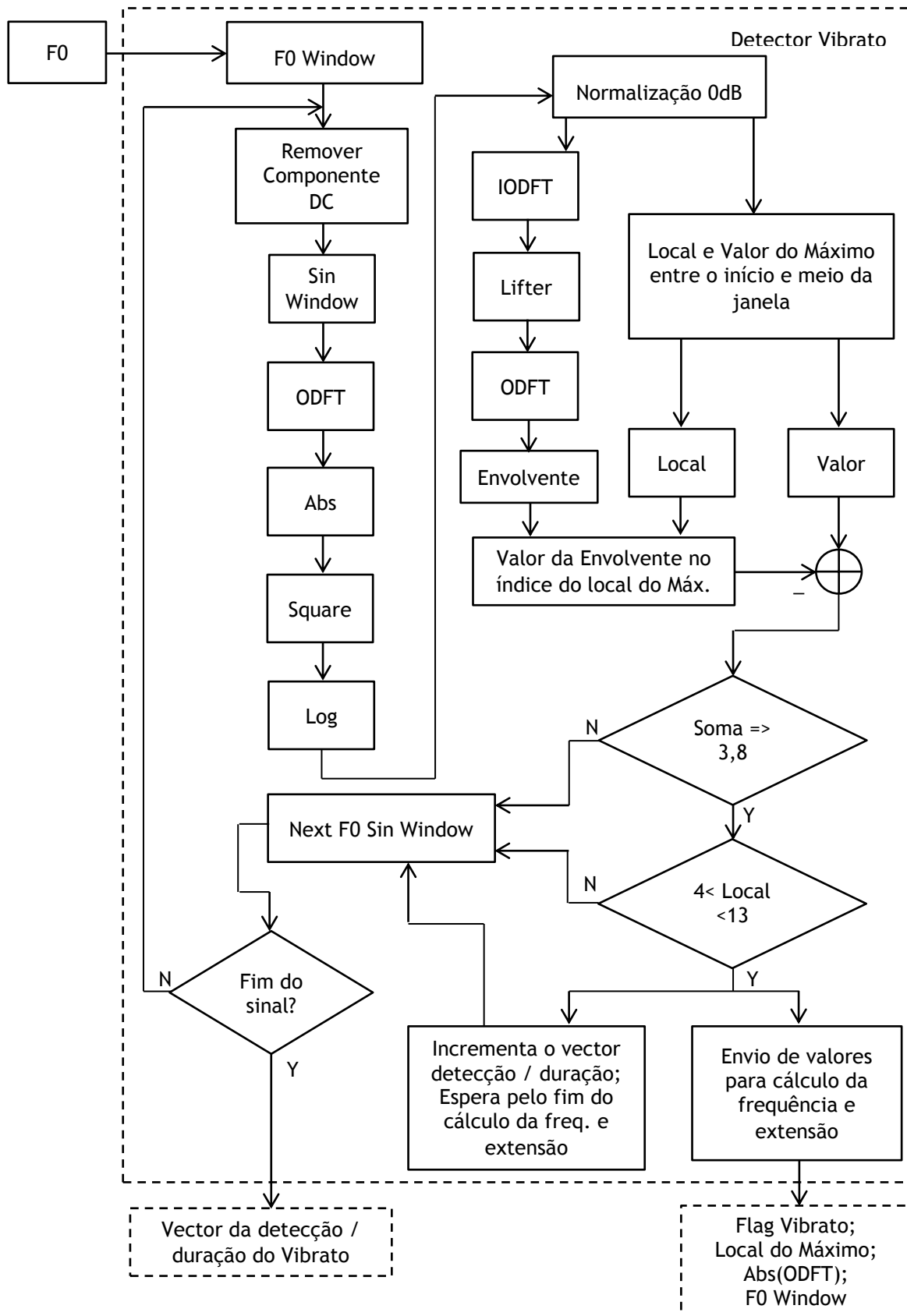
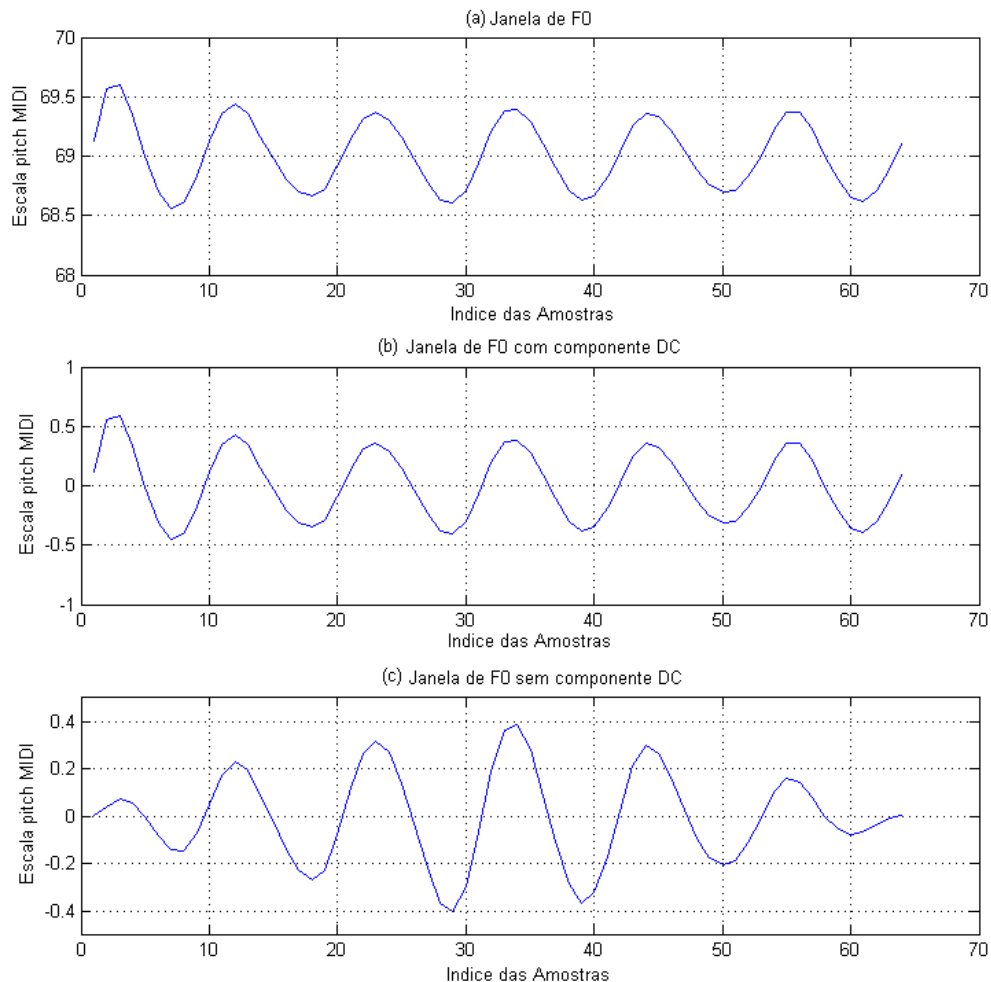


Figura 18 - Fluxograma do algoritmo de detecção do Vibrato.

Apesar do algoritmo não ser complexo, exige a execução de vários passos. A melhor forma de perceber o seu funcionamento é obtida através da visualização de um sinal com Vibrato ao longo do algoritmo de detecção. A figura 19 representa uma janela de comprimento  $N = 64$  sem qualquer alteração(a), após a remoção da sua componente DC (b) e depois de aplicado o conceito de janela seno (c).



**Figura 19** - Evolução da janela de sinal F0.

O sinal de vibrato usado para a figura 19 é de origem sintética, isto é, foi criado a partir de um software que sintetizou a nota cantada e que contam um vibrato caracterizado por uma frequência de 4Hz e uma extensão de 0,5 semitons (ST).

O gráfico (c) da figura anterior ilustra o efeito da aplicação do conceito do janelamento seno. A necessidade de o fazer está relacionada com o facto de ocorrer sobreposição de *frames*, ou seja, o sinal não é simplesmente dividido em pequenas janelas de tamanho  $N = 64$ . Devido ao número limitado de amostras, a percentagem de sobreposição utilizada pelo algoritmo estimador de vibrato foi de 90,6%, o que corresponde a um salto entre frames de  $n = 6$ . Esta elevada percentagem de sobreposição é utilizada devido à quantidade muito reduzida de amostras do sinal de F0.

### 4.3 - Detecção e duração do Vibrato

Na figura 20, podemos visualizar o conceito de sobreposição de janelas utilizado no estimado de vibrato.

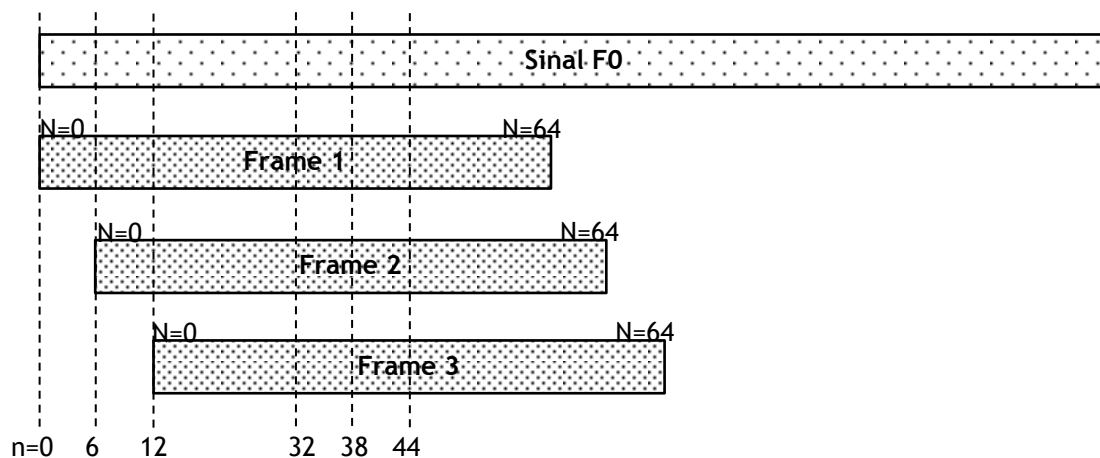


Figura 20 - Sobreposição das janelas de sinal F0.

Devido à sobreposição de janelas, alguma da informação contida em cada uma delas será repetida numa ou em várias janelas. Para que o centro da janela tenha mais peso e relevância nos cálculos a efectuar, é aplicada uma janela seno.

Nos casos em que o algoritmo de estimação de F0 não devolva valores, devido a perturbações muito graves no sinal original, ou devido a transições abruptas do valor de F0, é necessário preencher esses valores devido às perturbações que a sua ausência causa no espectro de potências. Estas perturbações causariam erros na detecção, estimação de extensão e de frequência do vibrato. A forma delineada para ultrapassar este problema é demonstrada na figura 21.

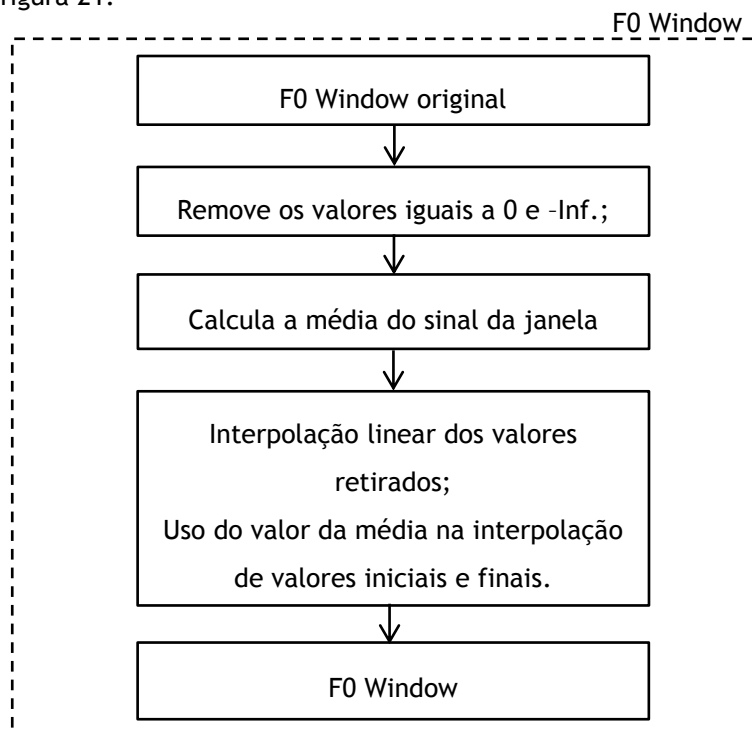


Figura 21 - Algoritmo de tratamento da janela de sinal F0.

O processamento de sinal demonstrado na figura anterior é executado dentro do bloco “F0 Window”, descrito no diagrama do algoritmo de detecção do vibrato. O resultado da sua implementação pode ser visualizado na figura 22 que ilustra um caso em que o algoritmo F0 não retornou alguns valores na janela de sinal, criando descontinuidades no sinal.

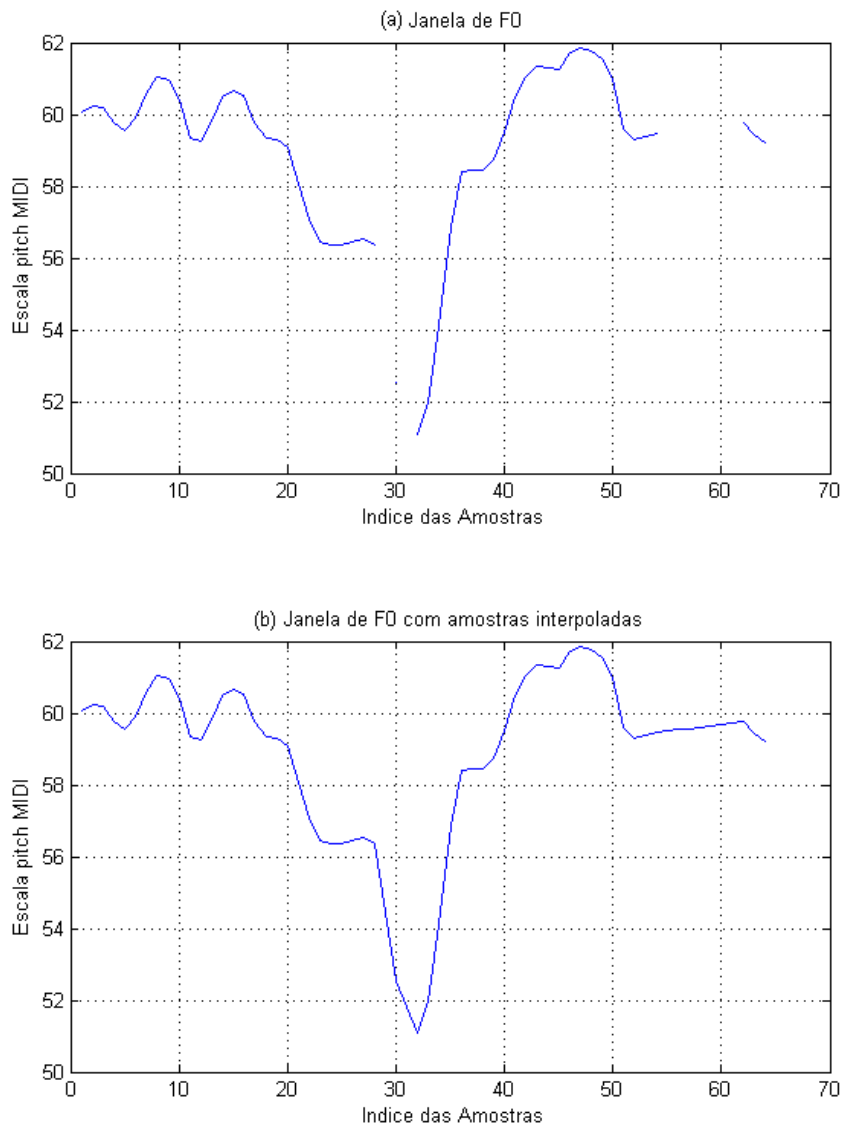


Figura 22 - Tratamento da janela de sinal F0.

O método de interpolação utilizado é bastante simples, pois as falhas que ocorrem no valor de F0 são bastante reduzidas, como poderemos verificar no capítulo relativo aos testes efectuados ao algoritmo de estimação da frequência fundamental *SearchTonal*. O método utilizado para interpolar os valores de sinal  $-\infty$ , que aparecem como descontinuidades no gráfico (a) da figura 22 é o da interpolação linear em que consiste em utilizar o ponto mais próximo conhecido, à direita e à esquerda do valor que se pretende obter, utilizando uma recta que contém os dois valores conhecidos.

### 4.3 - Detecção e duração do Vibrato

Como este caso de descontinuidade apenas acontece em transição abrupta da F0 ou em situações em que o sinal é muito perturbado, e tendo em conta que o número de amostras desconhecidas é normalmente uma apenas, não requer a utilização de métodos de interpolação mais complexos.

Pode ser observado na figura 23 a visualização da aproximação feita através da recta.

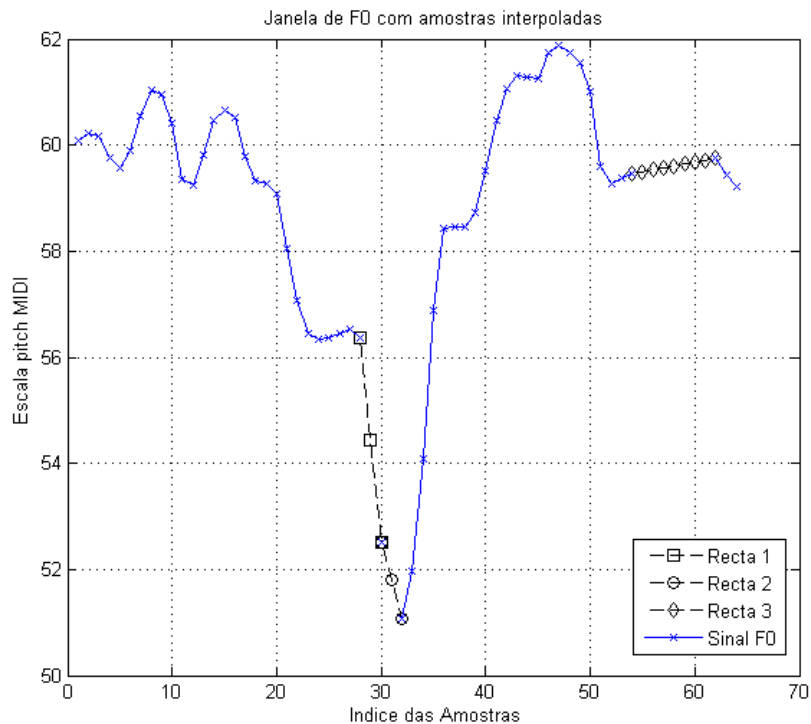


Figura 23 - Visualização da recta utilizada na interpolação linear do cálculo do Valor de F0 desconhecido.

Após o processamento indicado nos passos anteriores, seguida da aplicação de uma janela sinusoidal, os quatro passos seguintes envolvem as operações:

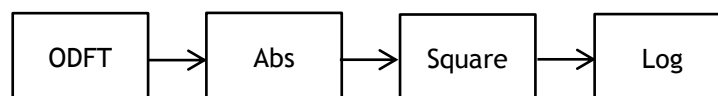
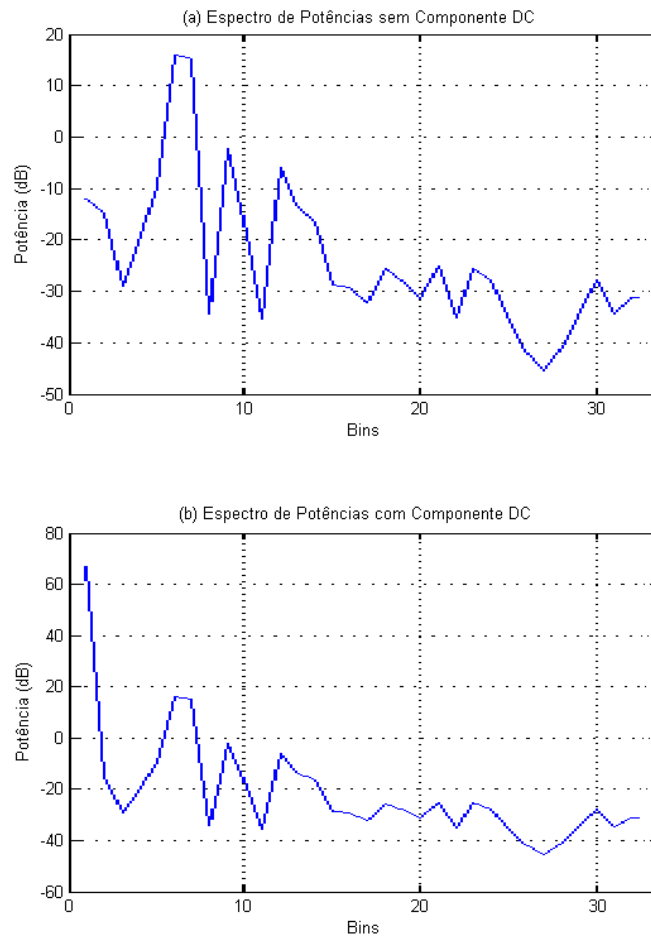


Figura 24 - Cálculo do espectro logarítmico de potências do sinal.

O resultado das operações traduz o espectro logarítmico de potências do sinal.

Uma das razões pela qual é imperativa a remoção da componente DC será ilustrada na figura 25, onde o gráfico (a) é a representação do espectro logarítmico de potências do sinal com a componente DC removida e o gráfico (b) representa o mesmo sinal mas com componente DC.



**Figura 25** - Comparação do espectro logarítmico de potências do sinal com e sem componente DC.

É realmente perceptível na figura 24 o impacto da componente DC no espectro de potência. A remoção torna a identificação do máximo muito mais simples, pois torna-se claramente visível no espectro e não será necessário aplicar um método computacional para encontrar máximos locais, o que introduziria mais um factor de complexidade e, eventualmente, erro.

Esta abordagem enfrentaria um problema sério se o valor máximo se manifestasse nos primeiros dois ou três *bins*, isto porque a remoção da componente DC poderia afectar o real valor do vibrato. Contudo, como sabemos as dimensões utilizadas na transformada de Fourier no cálculo de F0 e do vibrato, bem como a frequência de amostragem do sinal, podemos facilmente obter a região prevista da ocorrência do vibrato. O número de amostras utilizado na transformada de Fourier pelo algoritmo *SearchTonal* é  $N_{FFT} = 1024$ , tendo uma taxa de sobreposição de  $t = 50\% = 0.5$ . Como a frequência de amostragem do sinal é  $F_s = 22050 \text{ Hz}$ , podemos calcular o valor da frequência de amostragem do sinal F0:

$$F0_s = \frac{F_s}{(N_{FFT} \times t)} = \frac{22050}{(1024 \times \frac{1}{2})} \approx 43,07 \text{ Hz} \quad (4.1)$$

### 4.3 - Detecção e duração do Vibrato

Tal como já foi referido em capítulos anteriores, o intervalo do vibrato é definido por um valor máximo de frequência com valor  $F_{v\max} = 8$  Hz, e um valor mínimo de  $F_{v\min} = 4$  Hz. O número de amostras utilizadas na transformada de Fourier associada ao estimador de vibrato é  $N_{v\text{FFT}} = 64$ . Através destes dados é possível calcular o intervalo no espectro de potências esperado para a ocorrência do vibrato:

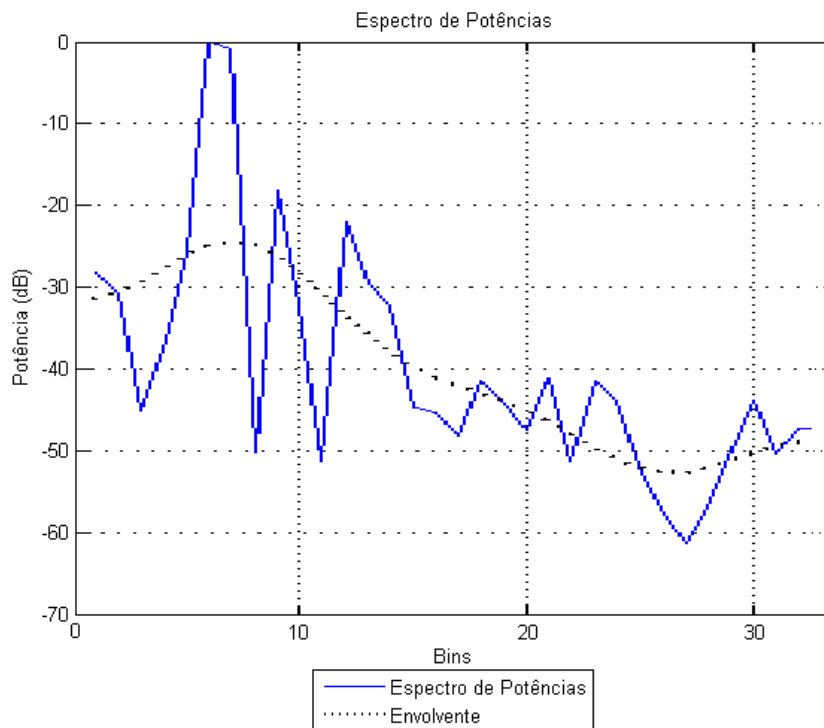
$$Bin_{min} = \frac{F_{v\min} \times N_{v\text{FFT}}}{F0_s} = \frac{4 \times 64}{43,07} = 5,94 \approx 6 \quad (4.2)$$

$$Bin_{max} = \frac{F_{v\max} \times N_{v\text{FFT}}}{F0_s} = \frac{8 \times 64}{43,07} = 11,89 \approx 12 \quad (4.3)$$

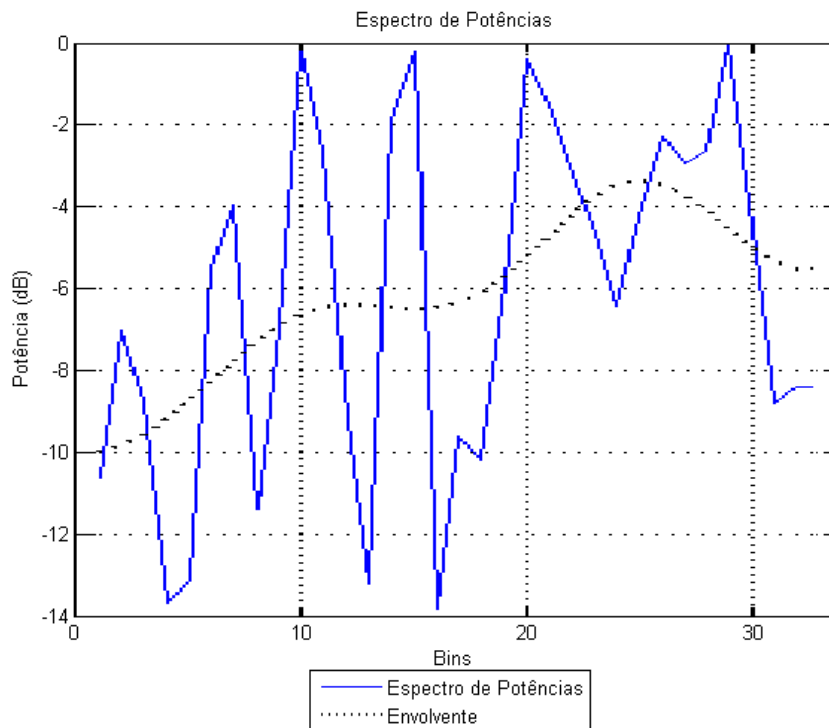
Como podemos verificar, o valor mínimo associado ao vibrato ocorre fora da região crítica, que seriam os primeiros dois a três *bins* mencionados anteriormente.

Uma segunda conclusão que pode ser retirada é que a existência no espectro de um máximo global fora da região de vibrato estimada significa a não presença de vibrato na janela de sinal. Este é um dos métodos utilizados para efectuar a detecção de vibrato através do espectro.

O segundo método utilizado envolve a obtenção da envolvente do espectro de potências normalizado, isto é, com o valor máximo igual a 0dB, e a sua comparação com o valor do espectro original, no local do valor máximo. Apesar da noção ser um pouco confusa, é de fácil percepção quando observada graficamente. Como tal, foram ilustradas diversas situações para demonstrar o seu funcionamento e que se encontram nas figuras 26 e 27.



**Figura 26** - Comparação do espectro logarítmico de potências do sinal com a sua envolvente na presença de vibrato.



**Figura 27** - Comparação do espectro de potências logarítmico do sinal com a sua envolvente na ausência de vibrato.

As figuras 26 e 27 representam o espectro de potências e a respectiva envolvente de dois sinais distintos. Ambos foram sintetizados recorrendo à ferramenta Madde, e a sua frequência fundamental é a mesma, com o valor  $F_0 = 440$  Hz que corresponde à nota A4 na escala musical igualmente temperada. O primeiro sinal contém vibrato, com uma frequência  $F_{V_0} = 4$  Hz, enquanto que o segundo sinal não tem qualquer variação da sua frequência fundamental. Torna-se claro através da comparação de ambas as figuras que os dois métodos de detecção de vibrato, se aplicados isoladamente poderiam não resultar em todas as situações. Contudo, ao aplicarmos os dois em conjunto, não só tornamos o algoritmo de detecção mais imune a perturbações, como o tornamos mais eficiente a nível computacional, visto que não é necessário aplicar métodos de estimação dos máximos locais em todo o espectro.

Quando se verifica a presença de vibrato, o algoritmo activa uma *flag* que significa a presença de vibrato na janela de sinal actual, juntamente com as saídas do valor da localização do máximo local, do vector correspondente à aplicação da  $|ODFT|$  da janela de sinal e do vector de valores da janela de sinal de  $F_0$ . Pode também ser observado na figura correspondente ao diagrama do algoritmo da detecção, que é também incrementado um vector que servirá para indicar a posição no sinal do vibrato e também a sua duração. A saída deste vector para o próximo algoritmo acontecerá quando não existirem mais janelas de sinal  $F_0$  para processar.

## 4.4 - Estimação da Extensão

A estimação do valor da extensão é uma componente importante do algoritmo de estimação do vibrato. Apesar de à partida parecer uma tarefa mais simples e imediata que a detecção, a verdade é que existem pormenores que dificultam essa tarefa. Se voltarmos a observar a figura 18 referente à evolução da janela de sinal  $F_0$ , presente na secção anterior, verificamos que no gráfico (b) que representa a janela de sinal removida da sua componente DC, verificamos que a extensão seria facilmente obtida recorrendo à expressão:

$$E_v = \frac{\Delta F_0}{2} = \frac{F_{0\ max} - F_{0\ min}}{2} \quad (4.4)$$

Contudo existem duas situações em que esta abordagem falha com resultados catastróficos:

- Na presença de mudanças de nota, isto é, quando o valor de  $F_0$  varia bruscamente;
- Quando existe uma perturbação do sinal provocada por uma parte do sinal contida na janela que não é vibrato, principalmente no início e fim da detecção.

A ocorrência destes fenómenos é facilmente perceptível através do uso de representações gráficas do sinal. Na figura 28, é ilustrado o que acontece quando existe uma mudança brusca do valor de  $F_0$ .

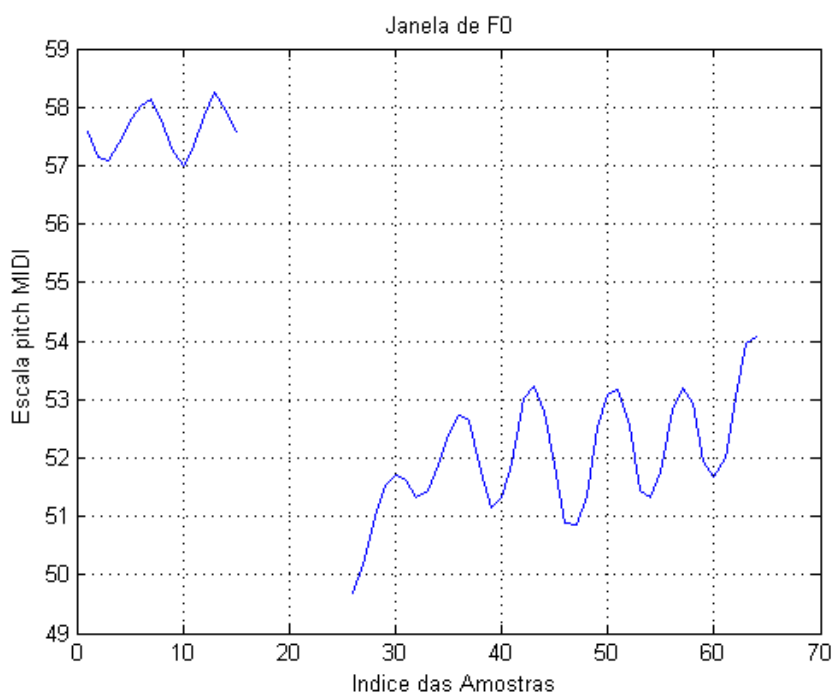


Figura 28 - Janela de sinal  $F_0$  com dois valores distintos.

Na figura 28 é ilustrada uma janela do sinal  $F_0$ , não tendo sido ainda calculados os valores das discontinuidades por meio da interpolação linear já abordada, contendo duas notas musicais. Se aplicássemos a expressão de estimação da extensão do vibrato, o valor devolvido seria aproximadamente  $E_v \approx 4,3$  enquanto que o valor esperado do vibrato na nota centrada

em  $F_0 = 52$  seria de sensivelmente  $E_v \approx 1,1$ , dependendo da amostra da janela considerada. É um erro grosseiro de cerca de 400%.

O segundo caso mencionado anteriormente como possível fonte de erros de estimação da extensão está relacionada com vários factores. A aplicação da janela seno implica um maior peso do valor do sinal no centro da janela, o que confere uma maior robustez na detecção do vibrato. Contudo, existirão caos em que aparecerão regiões da janela com e sem vibrato, principalmente na primeira e ultima janela. Como a sobreposição de janelas é muito elevada, estas perturbações podem ocorrer em mais que uma janela, não se confinando na primeira e última. Este segundo caso, apesar de ter uma natureza diferente do anterior, pode também ele ser ilustrado pela mesma figura.

A solução encontrada para resolver problemas relacionados com os dois acima mencionados, encontra-se descrita no diagrama de blocos da figura 29.

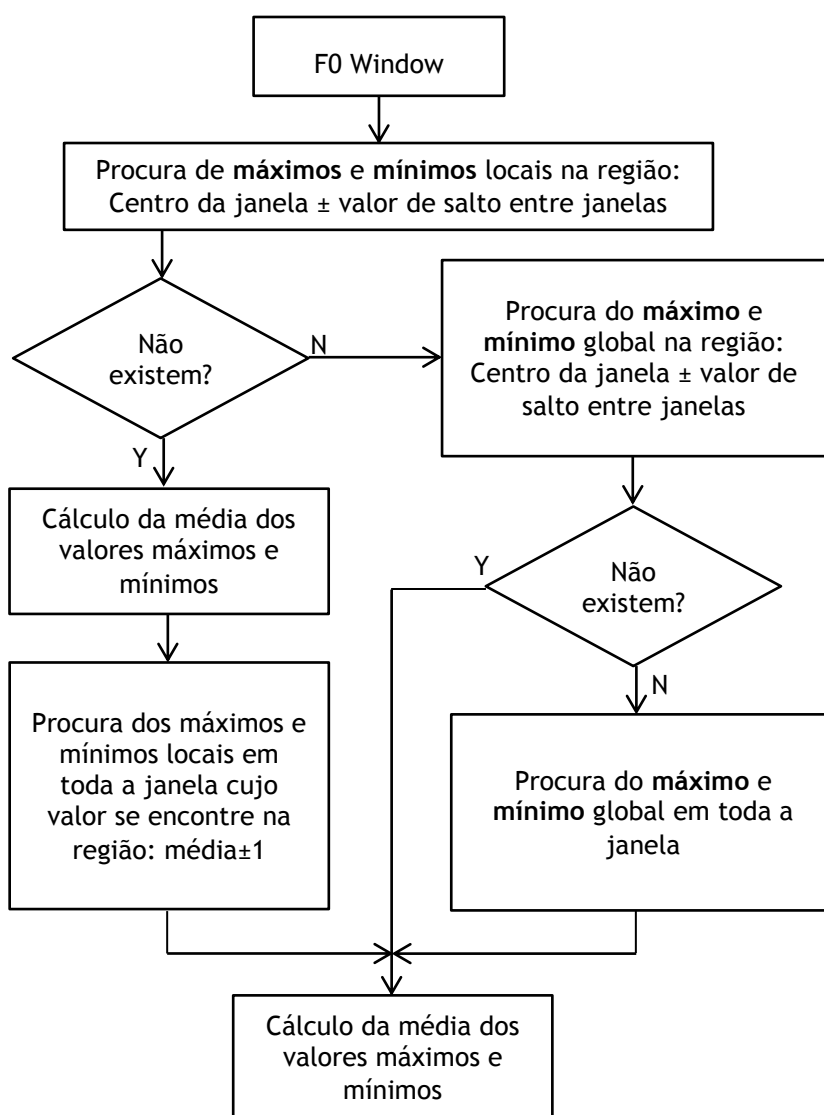


Figura 29 - Fluxograma do algoritmo de extracção do valor da extensão do vibrato.

## 4.5 - Estimação da frequência

Através da média de valores máximos e mínimos, obtemos o valor da extensão do vibrato usando a expressão (4.4).

Existem certas passagens no algoritmo que podem ser de difícil compreensão, todas elas relacionadas com a expressão: “procura de máximos e mínimos locais na região: centro da janela  $\pm$  valor de salto entre janelas”. A figura 30 ilustra o seu significado.

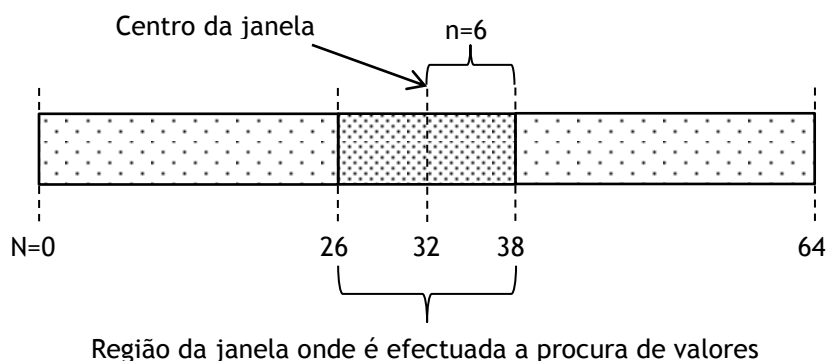


Figura 30 - Região da janela onde é efectuada a procura de máximos e mínimos locais.

A figura 30 ilustra uma janela de sinal  $F_0$  de comprimento  $N$ , em que é representada a região onde é efectuada a procura de valores máximos e mínimos locais. A única relação que não está representada de forma directa é o significado da expressão “valor de salto entre janelas”. Esse valor está representado pela sigla  $n$  e no caso do estimador de vibrato tem o valor  $n = 6$ .

## 4.5 - Estimação da frequência

A estimação de frequência é o parâmetro mais importante que caracteriza o vibrato, o que implica um grau de precisão muito elevado. Este objectivo não torna o algoritmo mais complexo que os usados para extrair os valores dos parâmetros descritos nas secções anteriores, mas necessita de métodos de elevada precisão. O método principal e responsável pela elevada precisão da estimação de frequência é o algoritmo *SearchDeltael* [41].

O algoritmo interpolador *SearchDeltael* é também utilizado pelo estimador da frequência fundamental *SearchTonal* e tem como propósito o de estimar com precisão o valor fraccionário máximo a partir de três amostras do espectro de potências, ou seja, a partir do valor do local do máximo e respectivos valores à direita e esquerda, no espectro de potências, o algoritmo *SearchDeltael* estima o valor do local do máximo até três casas decimais.

O valor utilizado como entrada do algoritmo é uma das saídas do detector de vibrato, o  $|ODFT|$ . A *flag* de vibrato, também fornecida pelo detector de vibrato, é o elemento que acciona o estimador da frequência de vibrato, tal como também acontece com o algoritmo que estima a extensão.

A figura seguinte ilustra o diagrama do algoritmo de estimação da frequência do vibrato, e a respectiva integração do interpolador *SearchDeltael* com os restantes componentes que compõem o estimador.

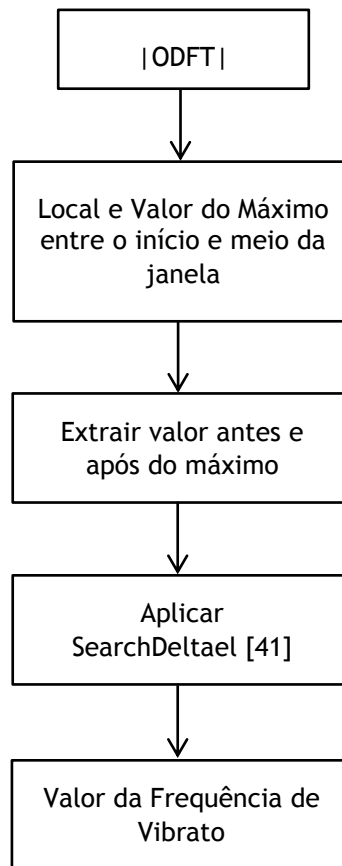


Figura 31 - Algoritmo de estimação do valor da frequência do vibrato.

#### 4.6 - Biofeedback do Vibrato

O *Biofeedback* do vibrato representa o método a utilizar na implementação do algoritmo no *software SingingStudio* para que este seja efectuado de uma forma simples e directa, isto é, de forma a informar o utilizador de uma forma não distractiva mas ao mesmo tempo para que este entenda o seu significado. Quando abrimos um ficheiro no *SingingStudio*, este ilustra a forma de onda da nota cantada, isto é, a frequência fundamental. O vibrato será representado através de uma representação numa cor diferente na região que contem o vibrato. O utilizador ao clicar nessa região obtém uma pequena janela imediatamente acima da região, janela essa de dimensões reduzidas e que contem dois pequenos gráficos correspondentes à frequência e extensão do vibrato. Imediatamente à frente da janela é mostrado os valores médios de ambos os parâmetros, juntamente com a sua duração. A figura 32 ilustra a detecção de uma região de vibrato de uma voz sintetizada.

#### 4.6 - Biofeedback do Vibrato

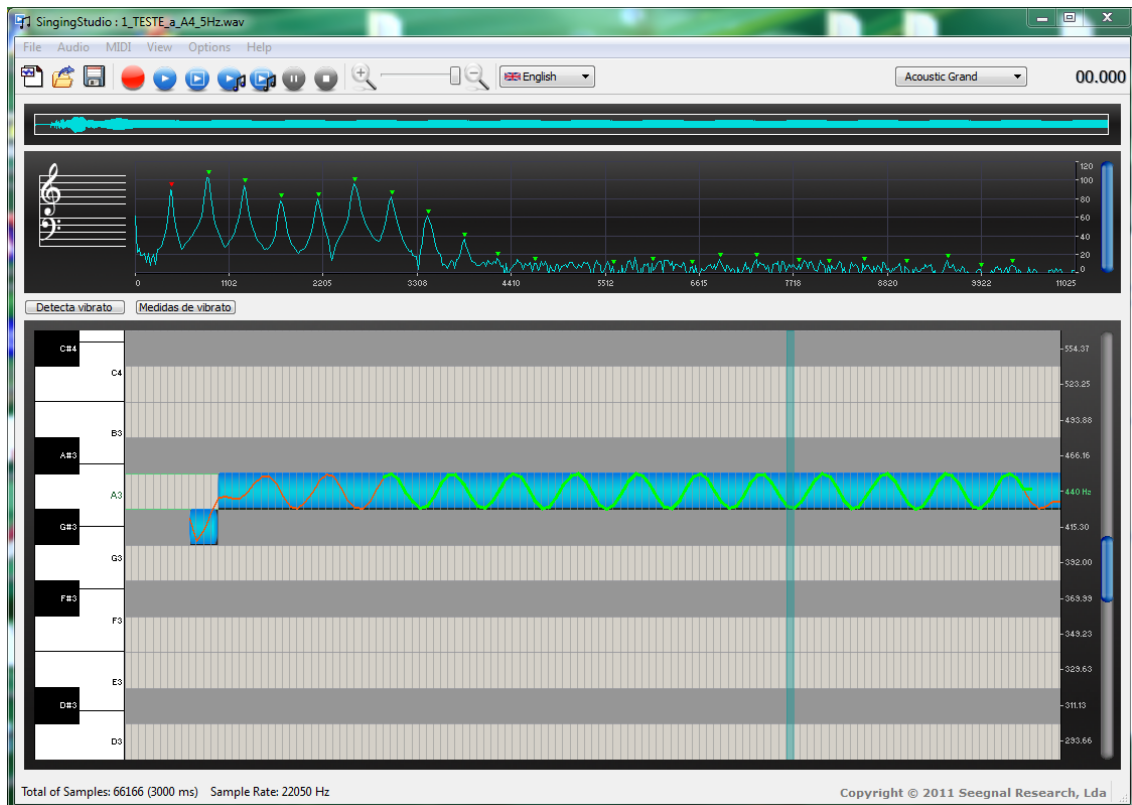


Figura 32 - Detecção de vibrato no *SingingStudio*.

Após a detecção do vibrato, poderíamos obter a informação dos seus parâmetros através de um simples *click* na região pretendida. A informação seria representada com recurso a uma janela semelhante com a ilustrada na figura 33.

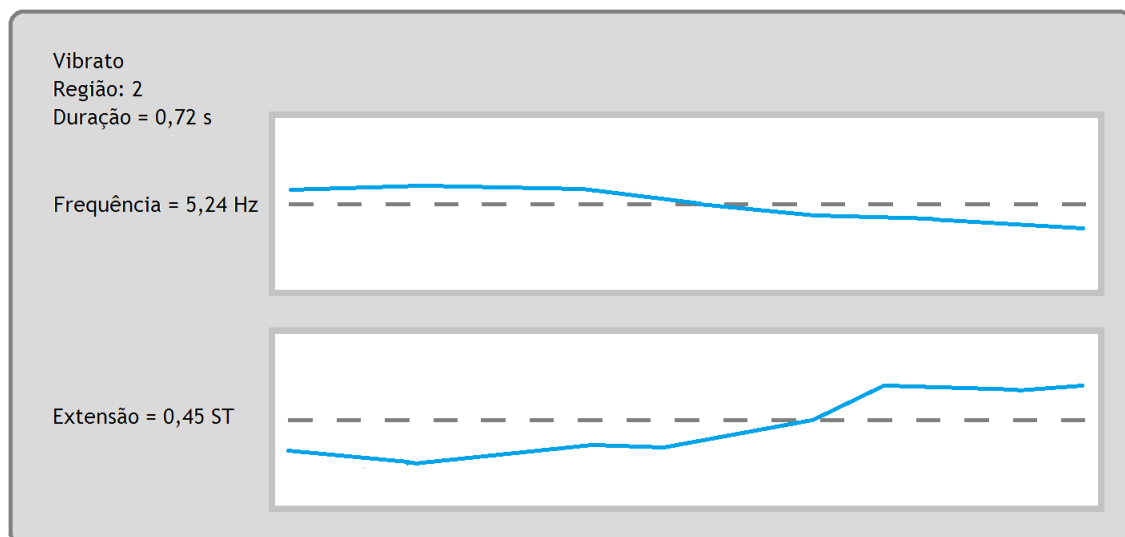


Figura 33 - Representação dos parâmetros do vibrato no *SingingStudio*.

# Capítulo 5

## Planificação de testes

A melhor forma de obter resultados fiáveis, concisos e que representem realmente o funcionamento dos algoritmos no mundo real é obtido através de uma boa planificação de testes. Este capítulo serve para identificar a metodologia utilizada nos testes, ou seja, o seu propósito e os resultados esperados em cada um deles, tendo em vista a análise da estimação da frequência fundamental e do estimador de vibrato. As ferramentas utilizadas são um aspecto importante na planificação dos testes, pois é através destas que será efectuada a avaliação dos algoritmos.

### 5.1 - Ferramentas utilizadas

A escolha das ferramentas utilizadas foi efectuada tendo como base os seguintes princípios:

- A funcionalidade da ferramenta não pode limitar o seu uso em diferentes algoritmos, ou seja, para uma dada função não podem existir várias ferramentas de avaliação dependendo do algoritmo testado;
- As ferramentas têm que possuir um grau de precisão e robustez aceitáveis, de forma a não introduzir erros nos testes dependendo do algoritmo ou da situação de teste.

A plataforma utilizada para testar todos os algoritmos foi o *Matlab* [42]. É uma ferramenta bastante robusta e flexível e que permite a integração de todos os algoritmos de forma eficaz. Outra vantagem apresentada é a variedade de opções de visualização de resultados que esta possibilita, desde exportação de tabelas excel, extracção de ficheiros de áudio ou simples figuras contendo gráficos e espectrogramas que permitem uma rápida avaliação dos algoritmos. Algumas das funcionalidades não são directas, isto é, requerem a elaboração de pequenas funções. Como o *Matlab* é uma ferramenta utilizada com frequência no curso que frequento, os conhecimentos necessários para implementar tais funções já foi adquirido, não apresentado por isso um problema de maior.

Mesmo sendo o *Matlab* uma ferramenta flexível, existe a necessidade de os algoritmos se encontrarem programados de acordo com a sua linguagem. Dos métodos estimadores de F0

## 5.2 - Estimação da Frequência Fundamental

utilizados, apenas um não se encontra programado em linguagem *Matlab*: o algoritmo *SearchTonal*. Contudo, o *Matlab* permite a importação de programas em código C\C++, através do uso de bibliotecas, sendo necessário efectuar uma formatação dos parâmetros de entrada e saída da mesma. Apesar de ser uma tarefa simplista, é um trabalho que requer algum tempo de forma a garantir que a adaptação é executada correctamente, pois de outra forma estaria a ser comprometida o bom funcionamento do algoritmo. Esta tarefa encontra-se presente no plano de trabalhos, que pode ser consultada no primeiro capítulo.

Uma das necessidades dos testes efectuados relaciona-se com a necessidade da produção de vozes sintetizadas de canto, ou seja, vozes cantadas que não são naturais mas que imitem uma pessoa a cantar. A ferramenta utilizada para este fim também tem que respeitar os princípios de escolha delineados no início desta secção, o que implica que a ferramenta tem que poder sintetizar vozes com e sem vibrato e com parâmetros suficientes que permitam a execução de todos os testes. O *software Madde* [30], do ponto de vista funcional, revelou ser bastante completo o que permite o seu uso em todos os testes previstos. Do ponto de vista da qualidade da voz sintetizada, e após uma análise da forma de onda gerada, concluiu-se que também correspondia aos critérios estabelecidos.

A ferramenta utilizada para analisar formas de onda e efectuar a gravação para ficheiros de som das vozes sintetizadas foi o *Adobe Audition*. Trata-se de uma ferramenta conceituada e com vários recursos tendo em vista a análise das formas de onda.

A figura 34 serve para ilustrar a ferramenta *Madde* e a variedade de parâmetros de configuração permitidos.

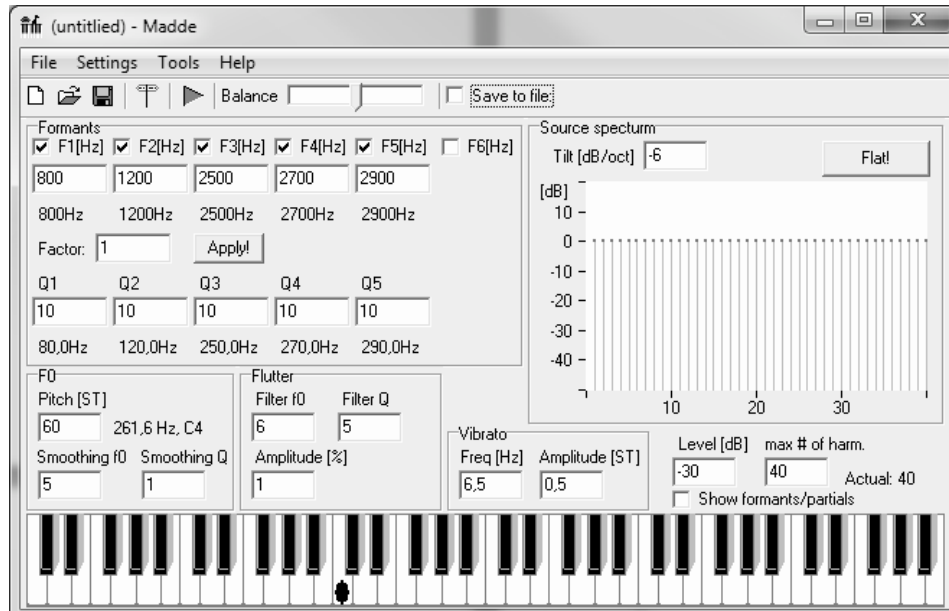


Figura 34 - Software Madde utilizado para sintetizar voz cantada.

## 5.2 - Estimação da Frequência Fundamental

Através dos testes efectuados aos algoritmos utilizados na estimação da frequência fundamental, pretende-se avaliar não só a sua precisão mas também a sua robustez. De

forma a focar cada teste no mínimo de características possíveis, estes foram divididos em quatro partes:

- Teste com vozes sintetizadas;
- Teste com vozes sintetizadas na presença de ruído;
- Teste com segmentos de vozes naturais;
- Teste com amostras de longa duração.

Devido a pequenas variações da frequência fundamental da voz humana, o melhor método para avaliar a precisão com que os algoritmos conseguem estimar F0 é através do recurso a vozes sintetizadas em que não só sabemos à partida a frequência fundamental, como também é garantido que a sua variação é nula. Para que este primeiro teste seja o mais abrangente possível, foram sintetizadas vinte e duas notas contendo desde a nota G2, até à nota G5 da escala igualmente temperada, o que corresponde a uma gama de frequências compreendida de  $F_0 = 98,0 \text{ Hz}$  até  $F_0 = 784,0 \text{ Hz}$ . Foram efectuados testes contendo uma zona estável do sinal das notas, de duração  $t = 2s$ . Devido à quantidade de formas de onda envolvidas no teste, os resultados vão ser apresentados numa tabela, contendo a informação sobre o valor médio e máximo de F0 estimado, bem como o desvio padrão, o erro absoluto e relativo, calculados tendo em conta o valor de F0 previsto para cada nota.

**Tabela 3** - Descrição dos parâmetros utilizados no primeiro teste com vozes sintetizadas dos algoritmos estimadores de F0.

<i>Nome Associado</i>	<i>F0 (Hz)</i>	<i>Nota</i>	<i>Duração (s)</i>	<i>Parâmetros recolhidos</i>
a_98Hz_G2_F0	98,0	G2	2	- Valor médio F0 (Hz); - Valor máximo F0 (Hz); - Desvio padrão (Hz); - Erro absoluto (Hz); - Erro relativo (%).
a_110Hz_A2_F0	110,0	A2	2	
a_123_5Hz_B2_F0	123,5	B2	2	
a_130_8Hz_C3_F0	130,8	C3	2	
a_146_8Hz_D3_F0	146,8	D3	2	
a_164_8Hz_E3_F0	164,8	E3	2	
a_174_6Hz_F3_F0	174,6	F3	2	
a_196Hz_G3_F0	196,0	G3	2	
a_220Hz_A3_F0	220,0	A3	2	
a_246_6Hz_B3_F0	246,6	B3	2	
a_261_6Hz_C4_F0	261,6	C4	2	
a_293_7Hz_D4_F0	293,7	D4	2	
a_329_6Hz_E4_F0	329,6	E4	2	
a_349_2Hz_F4_F0	349,2	F4	2	
a_392Hz_G4_F0	392,0	G4	2	
a_440Hz_A4_F0	440,0	A4	2	
a_493_9Hz_B4_F0	493,9	B4	2	
a_523_3Hz_C5_F0	523,3	C5	2	
a_587_3Hz_D5_F0	587,3	D5	2	
a_659_3Hz_E5_F0	659,3	E5	2	
a_698_5Hz_F5_F0	698,5	F5	2	
a_784Hz_G5_F0	784,0	G5	2	

## 5.2 - Estimação da Frequência Fundamental

O segundo teste efectuado tem como base avaliar em que condições de sinal é possível obter resultados robustos. A melhor forma é adicionar vários valores de ruído gaussiano às vozes sintetizadas no âmbito do teste anterior, obtendo assim formas de onda com valores de  $SNR = 5; 10; 15; 20; 25; 30dB$ . Os resultados serão novamente apresentados recorrendo a uma tabela, mas devido à quantidade de informação nesta contida, será também apresentado um gráfico de barras com todos os valores estimados de  $F_0$ , para cada situação, como forma de resumo. Na tabela 4 pode ser observado um resumo das fontes utilizadas nos testes, e os parâmetros recolhidos da análise do vibrato.

Tal como foi referido, o segundo teste executado tem como base os mesmos ficheiros de vozes sintetizadas descritas na tabela 3, ao qual se adicionou ruído de forma a obter um valor pré-estabelecido da relação sinal-ruído. A tabela 4 mostra os novos mostra os parâmetros que foram alterados em relação ao teste anterior.

**Tabela 4** - Descrição dos parâmetros utilizados no segundo teste com vozes sintetizadas.

<i>Sinal Original</i>	<i>SNR(dB) Novo Sinal</i>	<i>Tipo de Ruído</i>	<i>Parâmetros recolhidos</i>
- Os 22 Sinais utilizados no teste dos Algoritmos estimadores $F_0$ com vozes sintetizadas, sem ruído	30	- Branco Gaussiano	- Valor médio $F_0$ (Hz); - Valor máximo $F_0$ (Hz); - Desvio padrão (Hz); - Erro absoluto (Hz); - Erro relativo (%).
	25		
	20		
	15		
	10		
	5		

Os dois testes seguintes foram já efectuados com recurso a vozes naturais. O terceiro teste foi efectuado utilizando vozes de curta duração, não contendo mais do que uma nota. O seu objectivo é comprovar até que ponto os resultados obtidos com as vozes sintetizadas podem caracterizar o algoritmo e para isso é necessário executar uma comparação com casos simples mas que representem uma situação real. Uma vez que existe agora uma pequena perturbação do valor de  $F_0$ , a comparação dos valores será feito visualmente com recurso a gráficos. No entanto, é de realçar que este teste já não se centra na precisão do algoritmo mas sim no seu funcionamento em situações reais.

**Tabela 5** - Descrição dos parâmetros utilizados no terceiro teste com vozes sintetizadas.

<i>Nome Associado</i>	<i>Tipo de Voz</i>	<i>Nota musical</i>	<i>Vogal</i>	<i>Duração (s)</i>
JMA-cresc-C4-a	Masculina	C4	[a]	5
JMA-cresc-C4-e	Masculina	C4	[e]	4
JMA-cresc-C4-u	Masculina	C4	[u]	4
JMA-cresc-G3-a	Masculina	G3	[a]	5
JMA-cresc-G3-e	Masculina	G3	[e]	6
JMA-cresc-G3-u	Masculina	G3	[u]	5
RE-cresc-A4-a	Feminina	A4	[a]	16
RE-cresc-A4-e	Feminina	A4	[e]	21
RE-cresc-A4-u	Feminina	A4	[u]	13
RE-cresc-D5-a	Feminina	D5	[a]	10
RE-cresc-D5-e	Feminina	D5	[e]	9
RE-cresc-D5-u	Feminina	D5	[u]	9

A tabela 5 apresentada serve para caracterizar as vozes utilizadas no terceiro teste aplicado aos algoritmos estimadores da frequência fundamental.

O quarto e último teste efectuado têm como objectivo verificar o comportamento do algoritmo em sinais naturais de longa duração, isto é, que contenham várias notas. Tal como no teste anterior, os resultados serão obtidos sob a forma de gráficos, onde poderá ser avaliada a sua robustez na presença de variações de F0 de vozes naturais.

**Tabela 6** - Descrição dos parâmetros utilizados no quarto teste com vozes sintetizadas.

<i>Nome Associado</i>	<i>Tipo de Voz</i>	<i>Melodia</i>	<i>Duração (s)</i>
CT2-ppc-m1	Masculina	“N.º 1”	4
CT2-ppc-m2	Feminina	“N.º 1”	5
JF-mem-6-a	Masculina	“N.º 2”	16
JN-mem-6-a	Feminina	“N.º 2”	16
LP-mem-6-a	Feminina	“N.º 2”	16
soprano44	Feminina	“N.º 3”	4
suz32solo	Feminina	“N.º 4”	10
suz32	Feminina	“N.º 4 com batida”	9

Uma particularidade das vozes utilizadas nestes últimos dois testes é a presença de vibrato. É uma situação bastante vantajosa pois permite-nos avaliar a precisão da estimação do valor de F0, que será o sinal utilizado pelo estimador de vibrato para extrair as características do vibrato.

Os resultados dos testes acima descritos encontram-se presentes no capítulo seguinte, divididos por secção de forma a tornar a sua consulta mais acessível.

### 5.3 - Estimador de Vibrato

Neste segundo grupo de testes pretende-se essencialmente avaliar o desempenho do algoritmo de estimação do vibrato. A abordagem efectuada na elaboração dos testes, do ponto de vista do tipo de voz cantada escolhida, foi em muito idêntica à da utilizada na estimação da frequência fundamental, testes esses descritos na secção anterior. Numa primeira fase de testes foram utilizadas vozes sintetizadas, de forma a avaliar a precisão do algoritmo e só numa segunda fase testes foram usadas vozes naturais. Como já foi referido no capítulo do estimador de vibrato, o algoritmo devolve três parâmetros: frequência, extensão e duração do vibrato. Os testes elaborados para avaliar o estimador de vibrato foram concebidos de forma a testar a precisão da extracção de cada um deles, um de cada vez, antes de serem formulados testes mais complexos e abrangentes. Desta forma, obtemos resultados mais claros sobre o desempenho em cada um dos parâmetros.

Os testes efectuados para avaliar a precisão do algoritmo estimador de vibrato foram divididos em quatro partes:

- Teste com ondas sinusoidais;
- Teste com vozes sintetizadas;
- Teste com segmentos de vozes naturais;
- Teste com amostras de longa duração.

### 5.3 - Estimador de Vibrato

Tal como foi referido anteriormente, cada um dos testes tem o seu objectivo principal. Na secção seguinte são detalhados os testes efectuados de forma a testar a precisão da estimação da frequência do vibrato.

#### 5.3.1 - Estimação da Frequência

Dos testes referidos anteriormente, os testes que pretendem avaliar a precisão da estimação da frequência do vibrato são: teste com ondas sinusoidais e teste com vozes sintetizadas. No primeiro teste são criadas nove ondas sinusoidais com período  $T = 2$ , uma frequência compreendida entre  $F_v = 4 \text{ Hz}$  e  $F_v = 8 \text{ Hz}$ , com saltos de frequência  $f = 0,5 \text{ Hz}$ , e uma amplitude  $A = 1$ . Este teste visa elaborar uma primeira avaliação da precisão do algoritmo de extracção da frequência do vibrato, mas de forma ao resultado não ser condicionado devido a um erro que pudesse ocorrer devido ao algoritmo de estimação de  $F_0$ . Desta forma estaríamos somente a testar a precisão do algoritmo estimador de frequência, não esquecendo que se trata apenas de uma sinusoidal e não uma voz sintetizada e sua variação de  $F_0$ . Os resultados serão apresentados sob a forma de uma tabela, que indica o valor de  $F_0$  estimado, o valor médio e máximo da frequência de vibrato, desvio padrão e erro absoluto e relativo, tendo em conta os valores previstos.

O segundo teste é elaborado tendo em conta a utilização de vozes sintetizadas. A frequência fundamental utilizada foi de  $F_0 = 440 \text{ Hz}$ , que corresponde à nota A4, sendo as frequências de vibrato de igual valor às utilizadas nas sinusóides do teste anterior. A extensão do vibrato utilizada é de  $E_v = 0,5 \text{ ST}$  e  $E_v = 1,0 \text{ ST}$ , com uma duração de  $t = 2 \text{ s}$ . O *software* Madde também permite sintetizar a zona de ataque da nota, antes de esta se tornar estável. Para estudar a sua perturbação na estimação do vibrato por parte do algoritmo, também será efectuado um teste, contendo a zona de ataque seguido de uma zona estável, de duração  $t = 3 \text{ s}$ .

Nestes testes, pretende-se avaliar de forma precisa a estimação do valor da frequência do vibrato de forma realista e robusta. Tal como no teste anterior, os resultados são apresentados numa tabela, e o tipo de dados fornecidos idênticos aos do teste anterior. É também efectuada uma ilustração de um caso de forma a avaliarmos graficamente a evolução da frequência do vibrato, bem como o seu espectrograma.

Na tabela 7 é apresentado um resumo dos ficheiros utilizados nos testes associados à avaliação da estimação da frequência do vibrato.

**Tabela 7** - Descrição dos parâmetros utilizados nos primeiros testes associados à avaliação da estimação da frequência do vibrato.

Nome Associado	Tipo de Som	$F_0$ (Hz)	$F$ (Hz) Vibrato	Extensão Vibrato (ST)	Duração	Parâmetros recolhidos
sin_4Hz_05	- Onda Sinusoidal	440	4,0	0,5	T=2	- Valor médio $F_0$ (Hz); - Valor médio $F$ Vibrato (Hz) - Valor máximo $F$ Vibrato (Hz); - Desvio padrão (Hz); - Erro absoluto (Hz); - Erro relativo (%).
sin_4-5Hz_05		440	4,5	0,5		
sin_5Hz_05		440	5,0	0,5		
sin_5-5Hz_05		440	5,5	0,5		
sin_6Hz_05		440	6,0	0,5		
sin_6-5Hz_05		440	6,5	0,5		
sin_7Hz_05		440	7,0	0,5		
sin_7-5Hz_05		440	7,5	0,5		
sin_8Hz_05		440	8,0	0,5		
a_A4_4Hz_05	- Voz Cantada Sintetizada	440	4,0	0,5	t=2s estável; t=3s com ataque no início.	- Valor médio $F_0$ (Hz); - Valor médio $F$ Vibrato (Hz) - Valor máximo $F$ Vibrato (Hz); - Desvio padrão (Hz); - Erro absoluto (Hz); - Erro relativo (%).
a_A4_4-5Hz_05		440	4,5	0,5		
a_A4_5Hz_05		440	5,0	0,5		
a_A4_5-5Hz_05		440	5,5	0,5		
a_A4_6Hz_05		440	6,0	0,5		
a_A4_6-5Hz_05		440	6,5	0,5		
a_A4_7Hz_05		440	7,0	0,5		
a_A4_7-5Hz_05		440	7,5	0,5		
a_A4_8Hz_05		440	8,0	0,5		
a_A4_4Hz_10	- Voz Cantada Sintetizada	440	4,0	1,0	t=2s estável; t=3s com ataque no início.	- Valor médio $F_0$ (Hz); - Valor médio $F$ Vibrato (Hz) - Valor máximo $F$ Vibrato (Hz); - Desvio padrão (Hz); - Erro absoluto (Hz); - Erro relativo (%).
a_A4_4-5Hz_10		440	4,5	1,0		
a_A4_5Hz_10		440	5,0	1,0		
a_A4_5-5Hz_10		440	5,5	1,0		
a_A4_6Hz_10		440	6,0	1,0		
a_A4_6-5Hz_10		440	6,5	1,0		
a_A4_7Hz_10		440	7,0	1,0		
a_A4_7-5Hz_10		440	7,5	1,0		
a_A4_8Hz_10		440	8,0	1,0		

### 5.3.2 - Estimação da Extensão

De forma a testar a precisão do valor estimado da extensão do vibrato, foi elaborado um teste tendo como base as vozes sintetizadas. Neste contexto, foram criadas onze vozes sintetizadas contendo vibrato, todas com a mesma frequência fundamental  $F_0 = 440 \text{ Hz}$ . A frequência de vibrato foi também constante ao longo dos testes, de valor  $F_v = 5 \text{ Hz}$  e a sua duração  $t = 2 \text{ s}$ . A extensão do vibrato utilizada compreende-se entre  $A = 0,2 \text{ ST}$  e  $A = 1,2 \text{ ST}$ , com incrementos de  $A = 0,1 \text{ ST}$ . Os resultados obtidos são expostos numa tabela onde são fornecidos os valores da frequência fundamental estimada, o valor médio e máximo da extensão do vibrato, desvio padrão e erro absoluto e relativo, tendo em conta os valores previstos.

### 5.3 - Estimador de Vibrato

A tabela 8 ilustra as vozes sintetizadas utilizadas nos testes relacionados com o estimador de extensão do vibrato, as suas características, bem como os parâmetros recolhidos.

**Tabela 8** - Descrição dos parâmetros utilizados nos primeiros testes associados à avaliação da estimação da extensão do vibrato.

<i>Nome Associado</i>	<i>F0 (Hz)</i>	<i>F (Hz) Vibrato</i>	<i>Extensão Vibrato (ST)</i>	<i>Duração</i>	<i>Parâmetros recolhidos</i>
a_A4_5Hz_02	440	5,0	0,2	t=2s estável; t=3s com ataque no início.	- Valor médio F0 (Hz);
a_A4_5Hz_03	440	5,0	0,3		- Valor médio Ext. Vibrato (ST)
a_A4_5Hz_04	440	5,0	0,4		- Valor máximo
a_A4_5Hz_05	440	5,0	0,5		
a_A4_5Hz_06	440	5,0	0,6		
a_A4_5Hz_07	440	5,0	0,7		Ext. Vibrato (ST);
a_A4_5Hz_08	440	5,0	0,8		- Desvio padrão (ST);
a_A4_5Hz_09	440	5,0	0,9		- Erro absoluto (Hz);
a_A4_5Hz_10	440	5,0	1,0		- Erro relativo (%).
a_A4_5Hz_11	440	5,0	1,1		
a_A4_5Hz_12	440	5,0	1,2		

#### 5.3.3 - Detecção e estimação da duração

A detecção e estimação da duração do vibrato enquadram-se no terceiro e quarto testes efectuados no âmbito da avaliação do estimador de vibrato: teste com segmento de vozes naturais e teste com amostras de longa duração. Como em ambos os testes os sinais não são apenas compostos por vibrato, poderemos avaliar a precisão da detecção de regiões de vibrato, bem como a extracção do valor da sua duração.

No teste efectuado com vozes naturais, pretende-se avaliar numa curta amostra, e onde não existem variações abruptas de F0, a capacidade do algoritmo em detectar a presença de vibrato, e consequente extracção dos parâmetros associados. Os resultados serão apresentados sobre a forma de gráficos, onde são representados o sinal F0 e correspondente região que contem vibrato, a sua frequência e extensão.

No quarto teste é efectuado uma avaliação utilizando amostras de sinais com variações significativas de F0, com algumas regiões de vibrato. Este último teste tem como objectivo avaliar casos extremos onde poderiam ser detectados falsos vibratos, que poderiam ocorrer principalmente devido à falha do algoritmo que estima o valor da frequência fundamental. Como este é um caso mais genérico, também é efectuado uma avaliação realista do funcionamento do estimador de vibrato e todos os seus algoritmos que o compõem.

As vozes naturais utilizadas no terceiro e quarto testes são as mesmas vozes naturais utilizadas nos testes da estimação da frequência fundamental. As características das vozes encontram-se presentes nas tabelas da secção anterior.

A estimação da duração do vibrato foi representada graficamente nos resultados dos testes, através da delineação no gráfico de F0 da região que contem vibrato. A duração pode ser facilmente inferida tendo em conta o seu início e fim.

# Capítulo 6

## Resultados

Os resultados apresentados neste capítulo foram efectuados tendo por base a planificação de testes elaborada no capítulo anterior, onde foi detalhada toda a informação relevante para a execução destes. A quantidade de informação contida nestes resultados é imensa, sendo necessária uma visualização de resultados úteis, isto é, quando existe informação com uma relevância inferior tendo em conta o objectivo do teste, esta será apresentada em anexo.

### 6.1 - Estimação da Frequência Fundamental

Os algoritmos dos quais se pretende avaliar o seu desempenho na estimação da frequência fundamental são:

- Algoritmo baseado no método *Boersma*;
- Estimador *Yin*;
- Algoritmo *SearchTonal*.

Nas próximas secções serão apresentados os resultados obtidos, tendo em conta o tipo de voz utilizada. As duas próximas secções visam testes com vozes sintetizadas. No capítulo anterior foi demonstrado o propósito destes testes: obter uma medida exacta da precisão dos algoritmos. Através desta medida é então possível executar uma comparação entre os diferentes algoritmos.

#### 6.1.1 - Teste com Vozes Sintetizadas

O resultado dos testes efetuados com vozes sintetizadas, segundo a configuração apresentada no capítulo anterior, encontra-se presente nas tabelas seguintes. As primeiras três tabelas dizem respeito aos resultados obtidos por cada um dos algoritmos testados. Na quarta tabela é apresentada uma comparação dos erros relativos de cada algoritmo, sendo essa comparação também ilustrada graficamente, por meio de uma figura.

## 6.1 - Estimação da Frequência Fundamental

**Tabela 9** - Resultados do algoritmo SearchTonal obtidos com vozes sintetizadas.

<b>Resultados SearchTonal - calculo F0 com vozes sintetizadas</b>					
<i>Nome Associado</i>	<i>média (Hz)</i>	<i>máx. (Hz)</i>	<i>desvio (Hz)</i>	<i>erro abs (Hz)</i>	<i>erro rel (%)</i>
a_98Hz_G2_F0	97,878	100,503	0,865	0,669	0,683%
a_110Hz_A2_F0	109,970	112,351	1,059	0,866	0,787%
a_123_5Hz_B2_F0	123,481	125,437	0,866	0,673	0,545%
a_130_8Hz_C3_F0	130,711	132,806	0,853	0,662	0,506%
a_146_8Hz_D3_F0	146,849	148,573	0,845	0,688	0,469%
a_164_8Hz_E3_F0	164,778	166,728	0,917	0,750	0,455%
a_174_6Hz_F3_F0	174,421	176,241	0,740	0,614	0,352%
a_196Hz_G3_F0	196,084	198,291	0,831	0,663	0,338%
a_220Hz_A3_F0	220,033	222,868	0,970	0,750	0,341%
a_246_9Hz_B3_F0	246,924	249,191	0,893	0,702	0,284%
a_261_6Hz_C4_F0	261,619	263,565	0,867	0,727	0,278%
a_293_7Hz_D4_F0	293,636	295,648	0,822	0,633	0,216%
a_329_6Hz_E4_F0	329,712	331,207	0,736	0,585	0,177%
a_349_2Hz_F4_F0	349,219	350,906	0,810	0,635	0,182%
a_392Hz_G4_F0	392,003	393,454	0,751	0,620	0,158%
a_440Hz_A4_F0	439,973	442,844	0,892	0,682	0,155%
a_493_9Hz_B4_F0	493,880	495,194	0,676	0,569	0,115%
a_523_3Hz_C5_F0	523,136	524,572	0,732	0,593	0,113%
a_587_3Hz_D5_F0	587,277	589,565	1,103	0,896	0,153%
a_659_3Hz_E5_F0	659,244	661,081	0,884	0,731	0,111%
a_698_5Hz_F5_F0	698,334	700,364	0,965	0,761	0,109%
a_784Hz_G5_F0	784,114	786,225	1,098	0,903	0,115%

Tabela 10 - Resultados do algoritmo Yin obtidos com vozes sintetizadas.

<b>Resultados Yin - calculo F0 com vozes sintetizadas</b>					
<i>Nome Associado</i>	<i>média (Hz)</i>	<i>máx. (Hz)</i>	<i>desvio (Hz)</i>	<i>erro abs (Hz)</i>	<i>erro rel (%)</i>
a_98Hz_G2_F0	97,916	100,409	0,829	0,636	0,649%
a_110Hz_A2_F0	109,942	112,194	1,060	0,869	0,790%
a_123_5Hz_B2_F0	123,475	125,372	0,881	0,687	0,557%
a_130_8Hz_C3_F0	130,707	132,839	0,858	0,668	0,511%
a_146_8Hz_D3_F0	146,854	148,485	0,855	0,696	0,474%
a_164_8Hz_E3_F0	164,817	166,700	0,899	0,731	0,444%
a_174_6Hz_F3_F0	174,448	176,369	0,730	0,601	0,344%
a_196Hz_G3_F0	196,087	198,266	0,842	0,668	0,341%
a_220Hz_A3_F0	220,034	222,884	0,990	0,771	0,350%
a_246_9Hz_B3_F0	246,947	249,198	0,892	0,701	0,284%
a_261_6Hz_C4_F0	261,624	263,599	0,883	0,735	0,281%
a_293_7Hz_D4_F0	293,652	295,674	0,828	0,635	0,216%
a_329_6Hz_E4_F0	329,708	331,108	0,736	0,585	0,178%
a_349_2Hz_F4_F0	349,230	350,918	0,809	0,627	0,179%
a_392Hz_G4_F0	392,015	393,487	0,761	0,629	0,160%
a_440Hz_A4_F0	439,975	442,636	0,883	0,672	0,153%
a_493_9Hz_B4_F0	493,824	495,239	0,695	0,590	0,119%
a_523_3Hz_C5_F0	523,193	524,585	0,725	0,586	0,112%
a_587_3Hz_D5_F0	587,280	589,737	1,147	0,935	0,159%
a_659_3Hz_E5_F0	659,275	661,235	0,932	0,759	0,115%
a_698_5Hz_F5_F0	698,261	700,500	1,023	0,834	0,119%
a_784Hz_G5_F0	784,162	786,213	1,101	0,911	0,116%

## 6.1 - Estimação da Frequência Fundamental

Tabela 11 - Resultados do algoritmo Boersma obtidos com vozes sintetizadas.

<b>Resultados boersma - calculo F0 com vozes sintetizadas</b>					
<i>Nome Associado</i>	<i>média (Hz)</i>	<i>máx. (Hz)</i>	<i>desvio (Hz)</i>	<i>erro abs (Hz)</i>	<i>erro rel (%)</i>
a_98Hz_G2_F0	97,959	100,794	0,878	0,671	0,685%
a_110Hz_A2_F0	110,141	112,636	1,105	0,919	0,836%
a_123_5Hz_B2_F0	123,651	125,452	0,879	0,748	0,606%
a_130_8Hz_C3_F0	130,927	133,021	0,900	0,698	0,534%
a_146_8Hz_D3_F0	147,100	149,225	0,879	0,782	0,533%
a_164_8Hz_E3_F0	165,136	167,345	0,981	0,738	0,448%
a_174_6Hz_F3_F0	174,703	176,734	0,867	0,790	0,452%
a_196Hz_G3_F0	196,437	199,072	0,960	0,906	0,463%
a_220Hz_A3_F0	220,692	223,260	1,043	1,156	0,525%
a_246_9Hz_B3_F0	247,737	251,242	1,271	1,459	0,591%
a_261_6Hz_C4_F0	262,554	263,240	1,296	1,587	0,607%
a_293_7Hz_D4_F0	294,883	294,928	0,420	1,279	0,436%
a_329_6Hz_E4_F0	330,268	330,268	0,000	0,641	0,194%
a_349_2Hz_F4_F0	351,316	351,316	0,000	2,088	0,598%
a_392Hz_G4_F0	395,336	395,417	0,751	3,423	0,873%
a_440Hz_A4_F0	443,092	443,092	0,000	3,092	0,703%
a_493_9Hz_B4_F0	492,584	492,584	0,000	1,299	0,263%
a_523_3Hz_C5_F0	527,968	527,968	0,000	4,716	0,901%
a_587_3Hz_D5_F0	587,953	599,773	6,970	5,742	0,978%
a_659_3Hz_E5_F0	670,447	672,996	6,696	12,775	1,938%
a_698_5Hz_F5_F0	695,233	716,748	4,780	4,925	0,705%
a_784Hz_G5_F0	794,196	794,196	0,000	10,205	1,302%

Os resultados obtidos nas tabelas acima apresentadas permitem executar uma avaliação do desempenho dos algoritmos. Através do erro relativo calculado podemos ilustrar o desempenho dos algoritmos graficamente. Pretende-se assim encontrar um meio de comparar os resultados obtidos de uma forma mais simples e fácil, entre os algoritmos de estimação da frequência fundamental. Na figura 35 é representado o gráfico descrito e que compara a performance dos algoritmos.

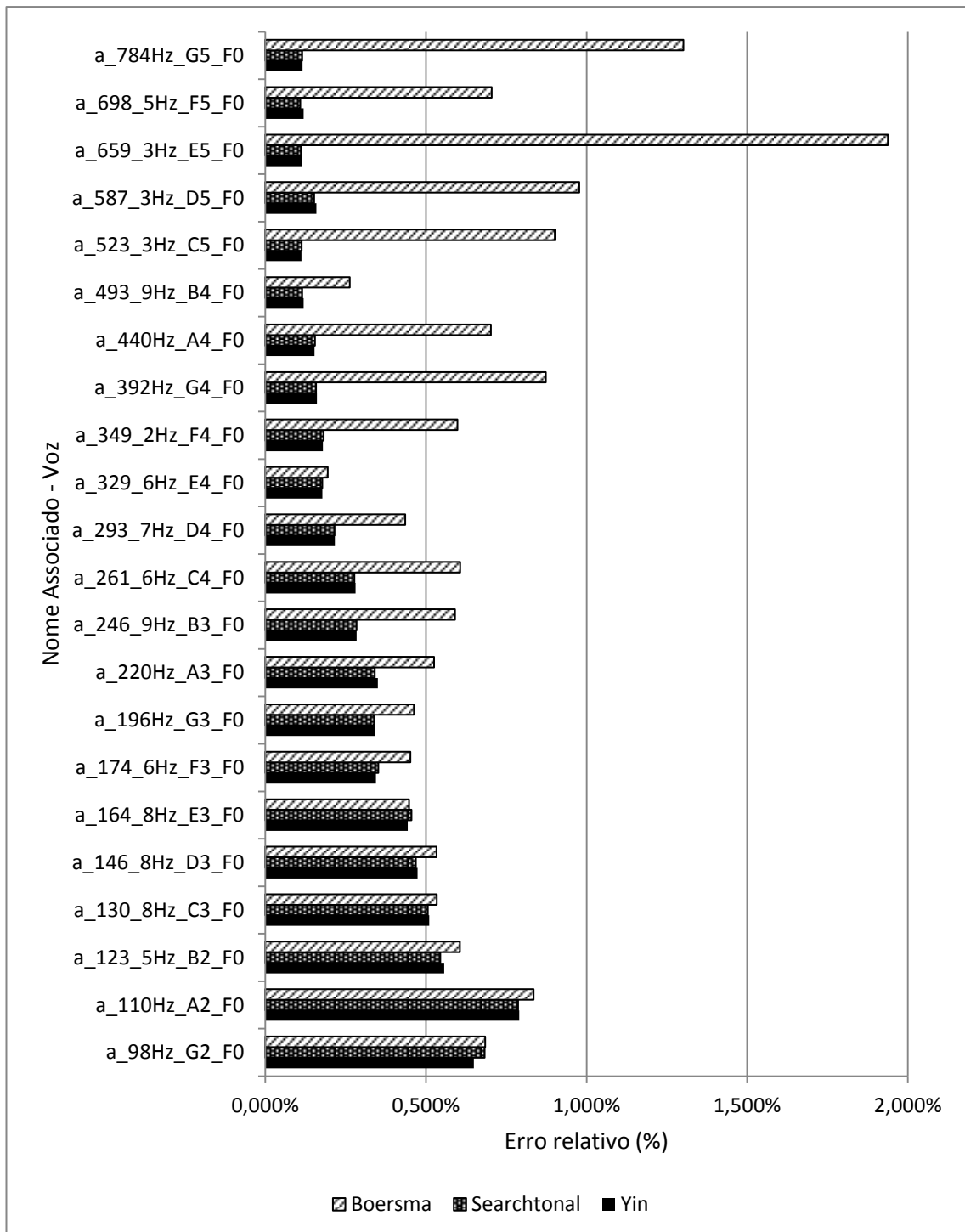


Figura 35 - Gráfico de comparação dos algoritmos de estimação de F0, através do erro relativo dos mesmos.

Os resultados apresentados permitem concluir que o desempenho dos algoritmos *SearchTonal* e *Yin* são claramente superiores ao método *Boersma*, principalmente para valores de F0 superiores a 200Hz. O desempenho dos algoritmos *SearchTonal* e *Yin*, tendo em conta apenas os resultados dos erros relativos apresentados, é bastante similar, sendo a

## 6.1 - Estimação da Frequência Fundamental

diferença entre ambos quase imperceptível. Na próxima secção serão apresentados os resultados obtidos perante os mesmos sinais com adição de ruído.

### 6.1.2 - Teste com Vozes sintetizadas na presença de ruído

Nesta secção, tal como já foi referido, é apresentado o resultado dos algoritmos perante sinais de vozes sintetizadas na presença de ruído. A adição do ruído foi efectuada através do Matlab, sendo depois calculado valor da relação sinal-ruído do sinal para garantir que de facto os valores de SNR são os previstos teoricamente, antes da adição. O resultado pode ser verificado na tabela seguinte.

Tabela 12 - Resultados da adição de ruído às vozes sintetizadas.

<b>SNR (dB) Previsto</b>						
<b>Nome Associado</b>	<b>SNR = 5</b>	<b>SNR = 10</b>	<b>SNR = 15</b>	<b>SNR = 20</b>	<b>SNR = 25</b>	<b>SNR = 30</b>
a_98Hz_G2_F0	5,056	10,012	15,049	20,047	24,975	29,994
a_110Hz_A2_F0	5,006	9,987	14,972	19,980	24,961	30,016
a_123_5Hz_B2_F0	5,028	9,952	15,021	20,006	25,047	29,978
a_130_8Hz_C3_F0	4,970	9,972	15,051	20,008	24,967	30,008
a_146_8Hz_D3_F0	4,998	9,973	14,998	20,021	24,981	30,019
a_164_8Hz_E3_F0	5,017	9,932	15,025	20,020	24,975	29,937
a_174_6Hz_F3_F0	5,020	10,020	14,998	19,954	24,998	30,005
a_196Hz_G3_F0	4,999	10,018	14,994	19,950	25,019	29,995
a_220Hz_A3_F0	5,018	10,023	15,023	19,998	25,054	29,971
a_246_9Hz_B3_F0	5,016	10,015	14,983	19,994	25,010	30,015
a_261_6Hz_C4_F0	5,014	10,035	15,004	19,959	25,016	30,033
a_293_7Hz_D4_F0	4,981	9,956	14,996	19,994	25,039	29,991
a_329_6Hz_E4_F0	5,003	9,991	14,982	19,975	24,996	29,996
a_349_2Hz_F4_F0	5,003	9,959	15,022	20,023	24,996	29,974
a_392Hz_G4_F0	5,015	10,022	15,015	20,040	25,012	30,001
a_440Hz_A4_F0	4,988	9,982	14,984	20,020	25,008	29,987
a_493_9Hz_B4_F0	5,038	9,962	15,018	20,000	25,028	30,015
a_523_3Hz_C5_F0	4,966	10,027	15,015	20,039	24,990	30,021
a_587_3Hz_D5_F0	4,960	10,022	15,020	19,995	25,017	30,028
a_659_3Hz_E5_F0	4,991	9,997	14,962	20,006	25,028	30,015
a_698_5Hz_F5_F0	5,050	10,028	15,004	19,984	25,022	29,996
a_784Hz_G5_F0	4,965	10,041	15,030	19,984	24,980	30,045

Devido à pobre performance revelada pelo algoritmo *Boersma*, particularmente para valores baixos de F0, quando este foi avaliado através das vozes sintetizadas sem ruído, foi decidida a não inclusão do mesmo no teste das mesmas vozes sintetizadas com adição de ruído branco gaussiano.

A quantidade de dados fornecidos pelos testes realizados é cerca de seis vezes maior que a obtida nos testes efectuados sem ruído adicionado às vozes sintetizadas. Para não dificultar a consulta dos dados, as tabelas completas serão apresentadas no anexo A deste documento, sendo apresentada nesta secção um quadro resumo que contém os erros relativos calculados, sob a forma de duas tabelas, correspondentes a cada um dos algoritmos em análise. Também

serão apresentados três gráficos, contendo o desempenho do algoritmo *SearchTonal*, do estimador Yin e uma comparação entre ambos, respectivamente. O terceiro gráfico mencionado está presente no Anexo A deste documento.

**Tabela 13** - Erro relativo do resultado do algoritmo *SearchTonal* obtido através da adição de ruído às vozes sintetizadas.

<b>Resultados SearchTonal - Erro relativo (%)</b>							
<b>Nome Associado</b>	<b>sem ruído</b>	<b>30dB</b>	<b>25dB</b>	<b>20dB</b>	<b>15dB</b>	<b>10dB</b>	<b>5dB</b>
a_98Hz_G2_F0	0,683%	0,681%	0,685%	0,690%	0,691%	0,684%	4,148%
a_110Hz_A2_F0	0,787%	0,790%	0,790%	0,792%	0,796%	0,770%	0,907%
a_123_5Hz_B2_F0	0,545%	0,545%	0,546%	0,542%	0,553%	0,577%	0,635%
a_130_8Hz_C3_F0	0,506%	0,506%	0,504%	0,519%	0,507%	0,531%	0,636%
a_146_8Hz_D3_F0	0,469%	0,467%	0,468%	0,467%	0,470%	0,473%	0,518%
a_164_8Hz_E3_F0	0,455%	0,455%	0,453%	0,456%	0,453%	0,463%	0,503%
a_174_6Hz_F3_F0	0,352%	0,354%	0,353%	0,359%	0,357%	0,361%	0,418%
a_196Hz_G3_F0	0,338%	0,347%	0,339%	0,346%	0,367%	0,932%	0,468%
a_220Hz_A3_F0	0,341%	0,335%	0,346%	0,340%	0,376%	0,952%	0,479%
a_246_9Hz_B3_F0	0,284%	0,277%	0,287%	0,303%	0,312%	0,281%	0,395%
a_261_6Hz_C4_F0	0,278%	0,305%	0,310%	0,315%	0,325%	1,479%	1,062%
a_293_7Hz_D4_F0	0,216%	0,222%	0,202%	0,228%	0,239%	3,770%	0,394%
a_329_6Hz_E4_F0	0,177%	0,179%	0,203%	0,185%	0,216%	5,506%	0,699%
a_349_2Hz_F4_F0	0,182%	0,173%	0,186%	0,214%	0,257%	3,818%	1,908%
a_392Hz_G4_F0	0,158%	0,198%	0,230%	0,276%	1,484%	7,120%	5,704%
a_440Hz_A4_F0	0,155%	0,175%	0,175%	0,825%	6,734%	15,744%	5,139%
a_493_9Hz_B4_F0	0,115%	1,777%	2,746%	6,514%	14,868%	18,155%	12,928%
a_523_3Hz_C5_F0	0,113%	0,209%	0,232%	1,411%	2,697%	16,329%	11,606%
a_587_3Hz_D5_F0	0,153%	0,776%	0,832%	4,472%	11,584%	14,229%	16,615%
a_659_3Hz_E5_F0	0,111%	0,923%	0,980%	3,878%	13,695%	21,816%	18,718%
a_698_5Hz_F5_F0	0,109%	1,081%	1,083%	2,833%	14,078%	29,493%	24,718%
a_784Hz_G5_F0	0,115%	0,597%	1,903%	6,971%	20,763%	30,530%	32,338%

## 6.1 - Estimação da Frequência Fundamental

Tabela 14 - Erro relativo do resultado do estimador Yin obtido através da adição de ruído às vozes sintetizadas.

<b>Resultados Yin - Erro relativo (%)</b>							
Nome Associado	<i>sem ruído</i>	30dB	25dB	20dB	15dB	10dB	5dB
a_98Hz_G2_F0	0,649%	0,649%	0,649%	0,648%	0,649%	0,647%	0,651%
a_110Hz_A2_F0	0,790%	0,791%	0,791%	0,791%	0,790%	0,791%	0,789%
a_123_5Hz_B2_F0	0,557%	0,557%	0,556%	0,555%	0,557%	0,560%	0,561%
a_130_8Hz_C3_F0	0,511%	0,511%	0,511%	0,511%	0,510%	0,509%	0,514%
a_146_8Hz_D3_F0	0,474%	0,474%	0,474%	0,474%	0,474%	0,474%	0,479%
a_164_8Hz_E3_F0	0,444%	0,444%	0,444%	0,445%	0,443%	0,445%	0,456%
a_174_6Hz_F3_F0	0,344%	0,344%	0,344%	0,344%	0,344%	0,344%	0,351%
a_196Hz_G3_F0	0,341%	0,341%	0,341%	0,341%	0,345%	0,342%	0,351%
a_220Hz_A3_F0	0,350%	0,350%	0,350%	0,351%	0,351%	0,353%	0,341%
a_246_9Hz_B3_F0	0,284%	0,284%	0,284%	0,284%	0,283%	0,289%	0,293%
a_261_6Hz_C4_F0	0,281%	0,281%	0,281%	0,282%	0,281%	0,283%	0,293%
a_293_7Hz_D4_F0	0,216%	0,216%	0,217%	0,216%	0,218%	0,217%	0,224%
a_329_6Hz_E4_F0	0,178%	0,178%	0,178%	0,178%	0,179%	0,178%	0,173%
a_349_2Hz_F4_F0	0,179%	0,179%	0,179%	0,179%	0,181%	0,182%	0,197%
a_392Hz_G4_F0	0,160%	0,161%	0,161%	0,162%	0,162%	0,158%	0,168%
a_440Hz_A4_F0	0,153%	0,153%	0,153%	0,152%	0,151%	0,152%	0,157%
a_493_9Hz_B4_F0	0,119%	0,120%	0,120%	0,120%	0,119%	0,123%	0,131%
a_523_3Hz_C5_F0	0,112%	0,112%	0,112%	0,112%	0,113%	0,115%	0,133%
a_587_3Hz_D5_F0	0,159%	0,159%	0,159%	0,159%	0,157%	0,163%	0,164%
a_659_3Hz_E5_F0	0,115%	0,115%	0,115%	0,114%	0,118%	0,117%	0,129%
a_698_5Hz_F5_F0	0,119%	0,120%	0,120%	0,118%	0,120%	0,132%	0,160%
a_784Hz_G5_F0	0,116%	0,117%	0,115%	0,118%	0,124%	0,129%	0,203%

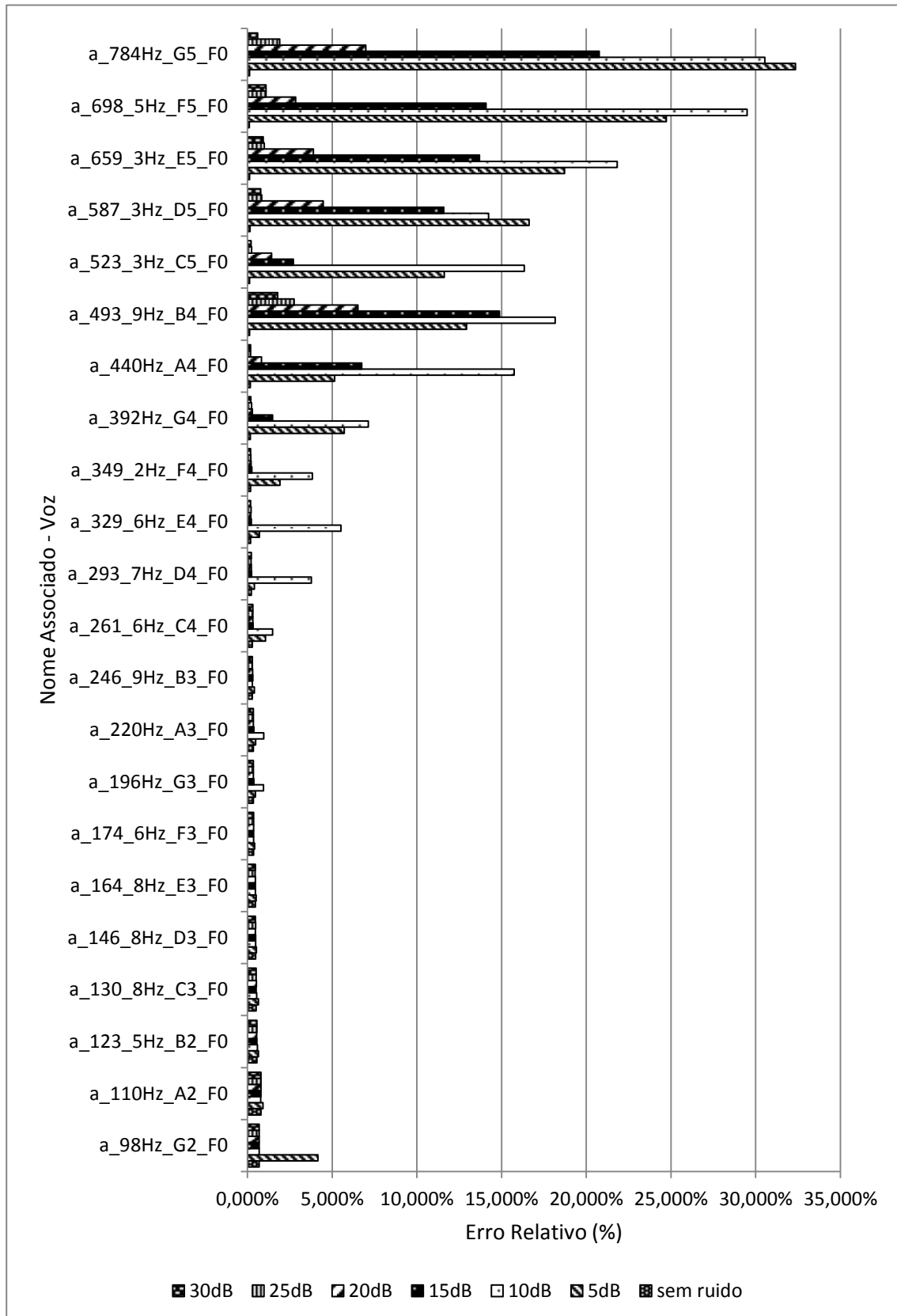


Figura 36 - Gráfico do desempenho de *SearchTonal* na presença de ruído.

## 6.1 - Estimação da Frequência Fundamental

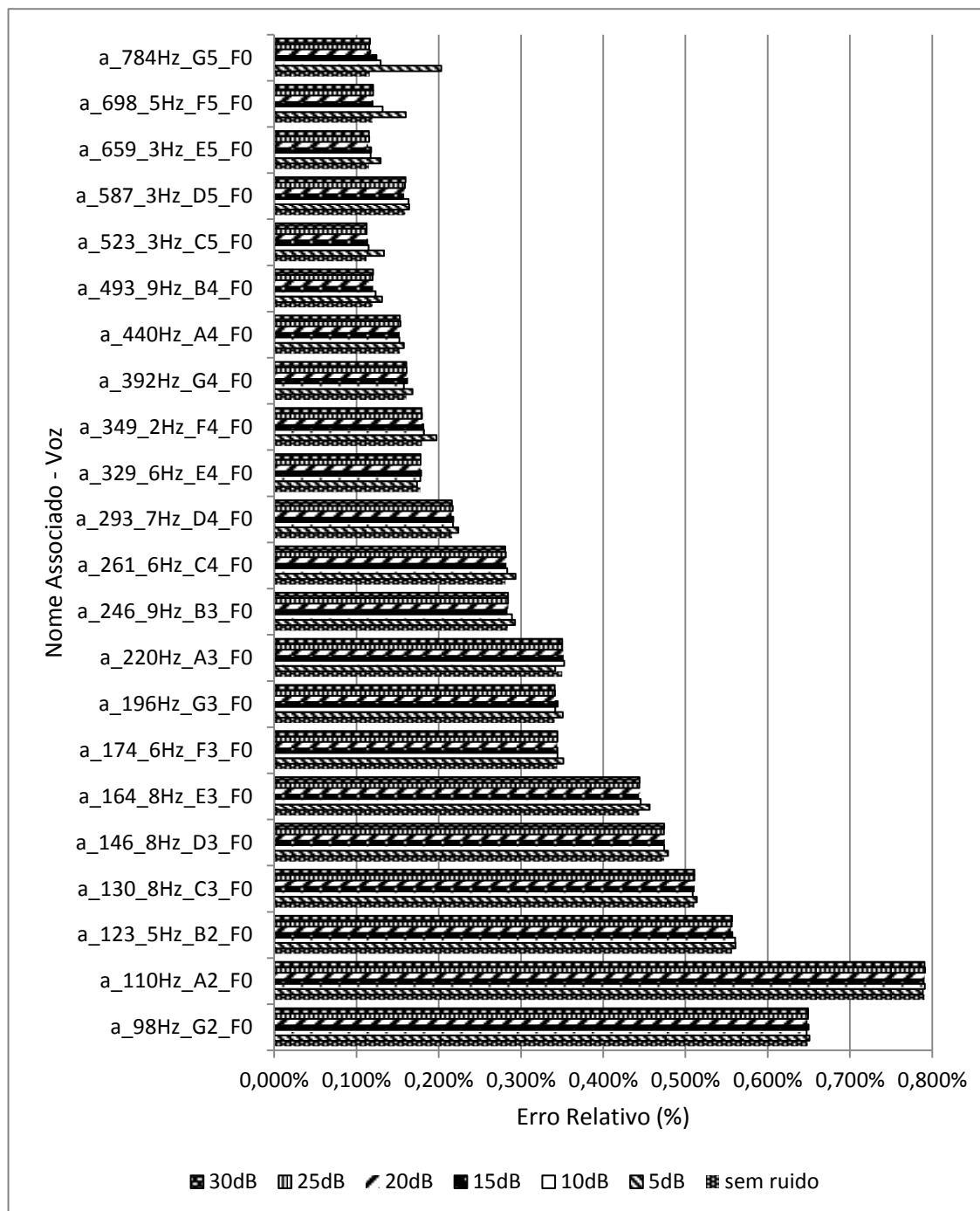


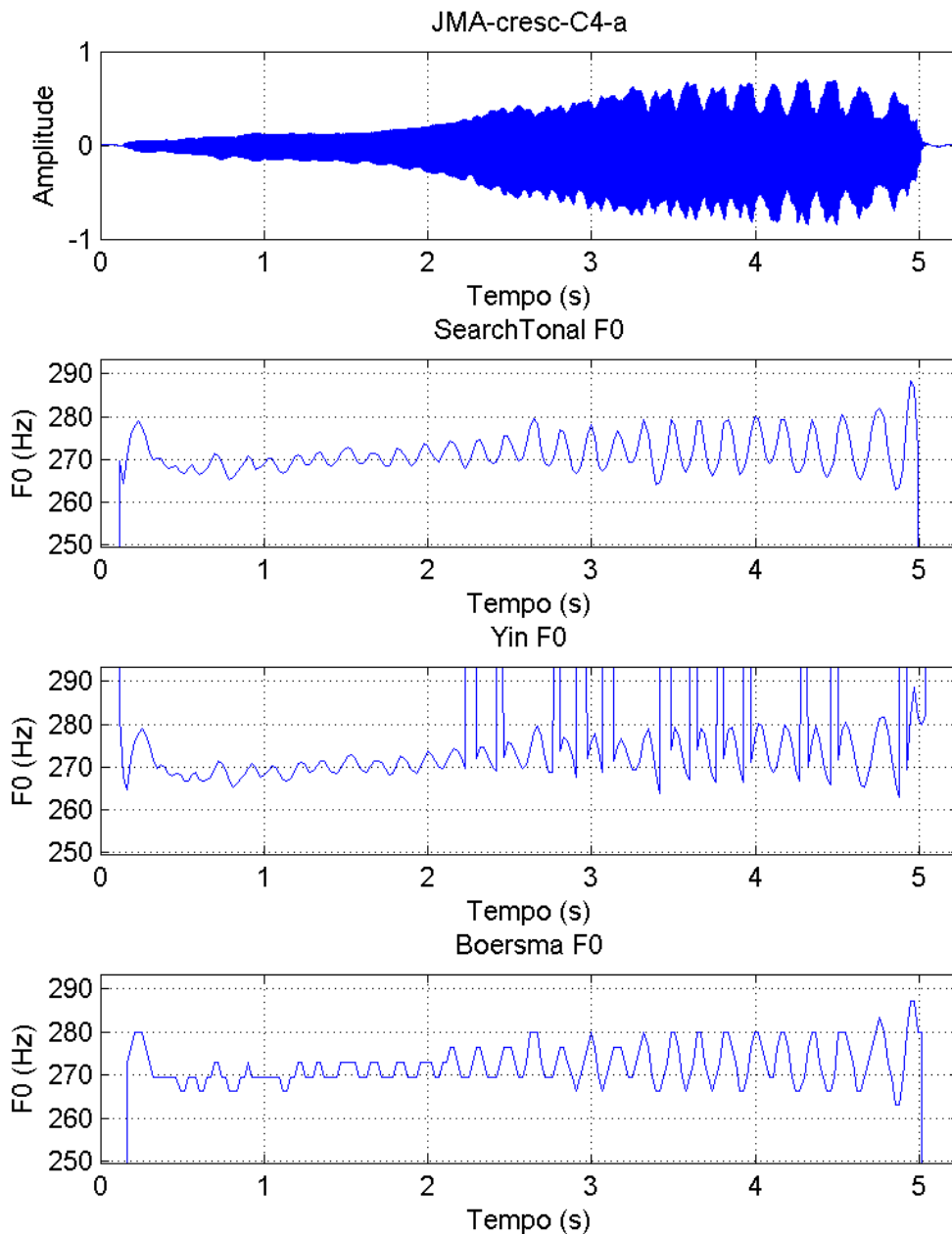
Figura 37 - Gráfico do desempenho do estimador Yin na presença de ruído.

Através das tabelas e gráficos apresentados, é possível efectuar uma avaliação do funcionamento de ambos os algoritmos de sinais de vozes sintetizadas com adição de ruído. O algoritmo *Yin* apresenta resultados sólidos e constantes, independentemente da relação sinal-ruído do sinal utilizado nos testes. O algoritmo *SearchTonal* apresenta dificuldades em estimar o valor de F0 com precisão para valores superiores a 300Hz. Esta perturbação é mais visível, quanto menor for a relação sinal-ruído utilizada nos testes. No anexo A é também apresentado um gráfico que compara os dois algoritmos, através da junção da informação

contida nestes dois gráficos. Contudo, torna-se difícil de efectuar uma boa avaliação devido à densidade de informação, daí não ser apresentado nesta secção.

### 6.1.3 - Teste com segmentos de Vozes Naturais

Nesta secção são apresentados os resultados obtidos na análise dos algoritmos estimadores de F0, através de segmentos de Vozes naturais, sem variações abruptas de F0. Os resultados, tal como já foi referido, serão apresentados graficamente.



**Figura 38** - Gráfico dos desempenhos dos algoritmos estimadores de F0, através da voz natural JMA-cresc-C4-a.

## 6.1 - Estimação da Frequência Fundamental

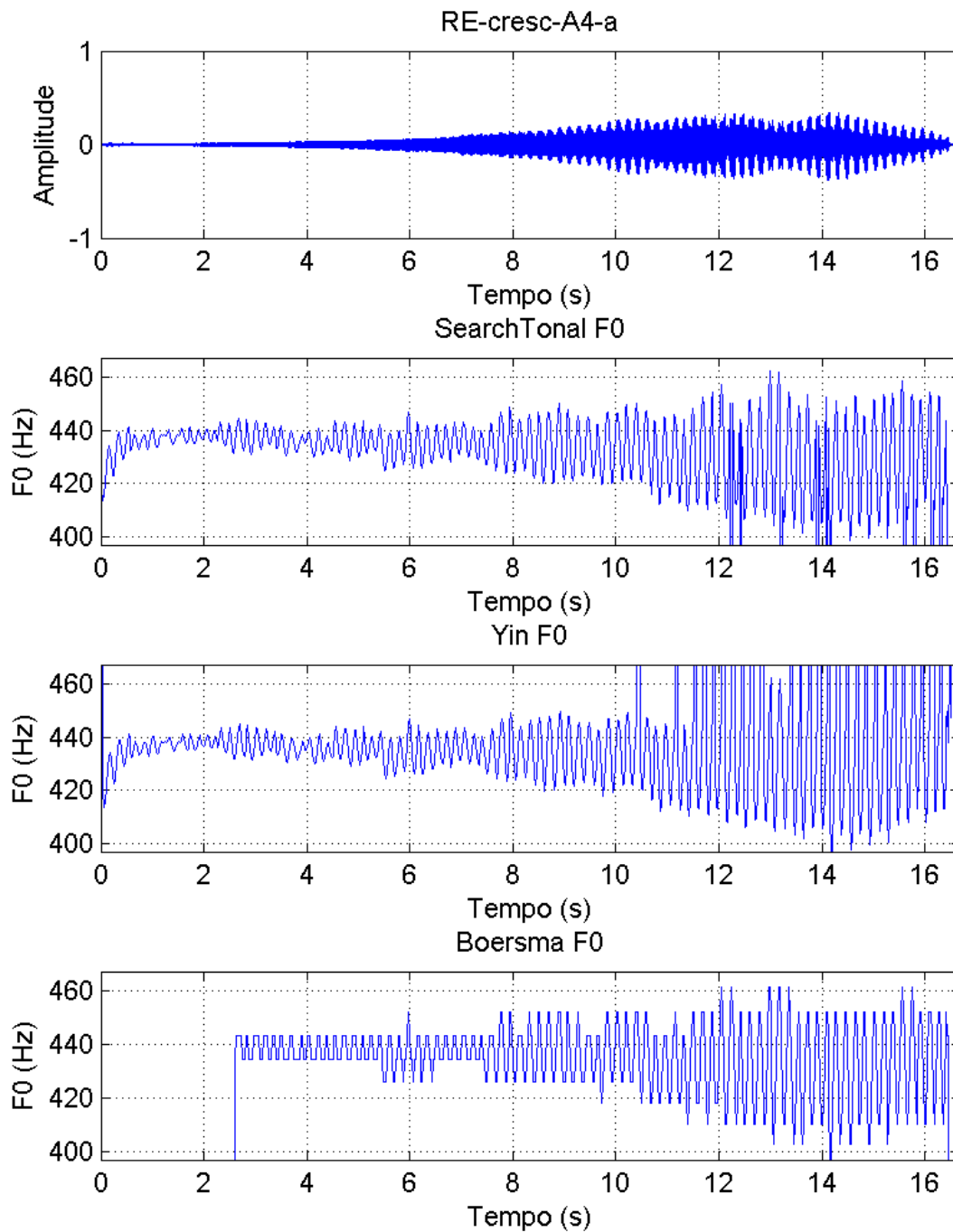
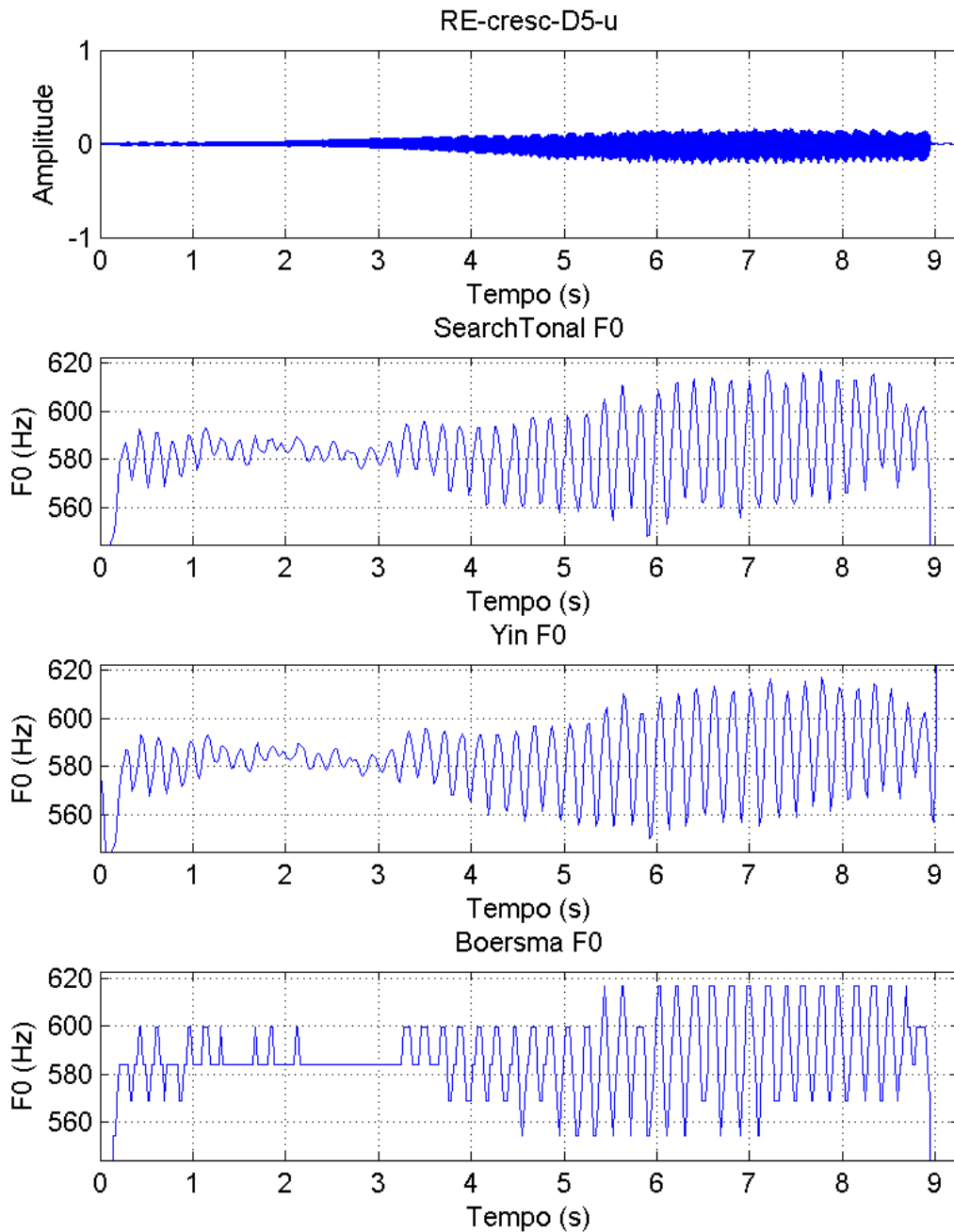


Figura 39 - Gráfico dos desempenhos dos algoritmos estimadores de F0, através da voz natural RE-cresc-A4-a.



**Figura 40** - Gráfico dos desempenhos dos algoritmos estimadores de F0, através da voz natural RE-cresc-D5-u.

As três figuras apresentadas resumem a performance apresentada pelos três algoritmos na estimação de F0 de segmentos de vozes naturais. Os restantes resultados são também relevantes, e por isso encontram-se presentes no Anexo B.

### 6.1.4 - Teste com amostras de longa duração

É apresentado nesta secção os resultados obtidos através da análise dos algoritmos estimadores da frequência fundamental, testados a partir de segmentos de Vozes naturais, com variações abruptas de F0. Os resultados são apresentados graficamente.

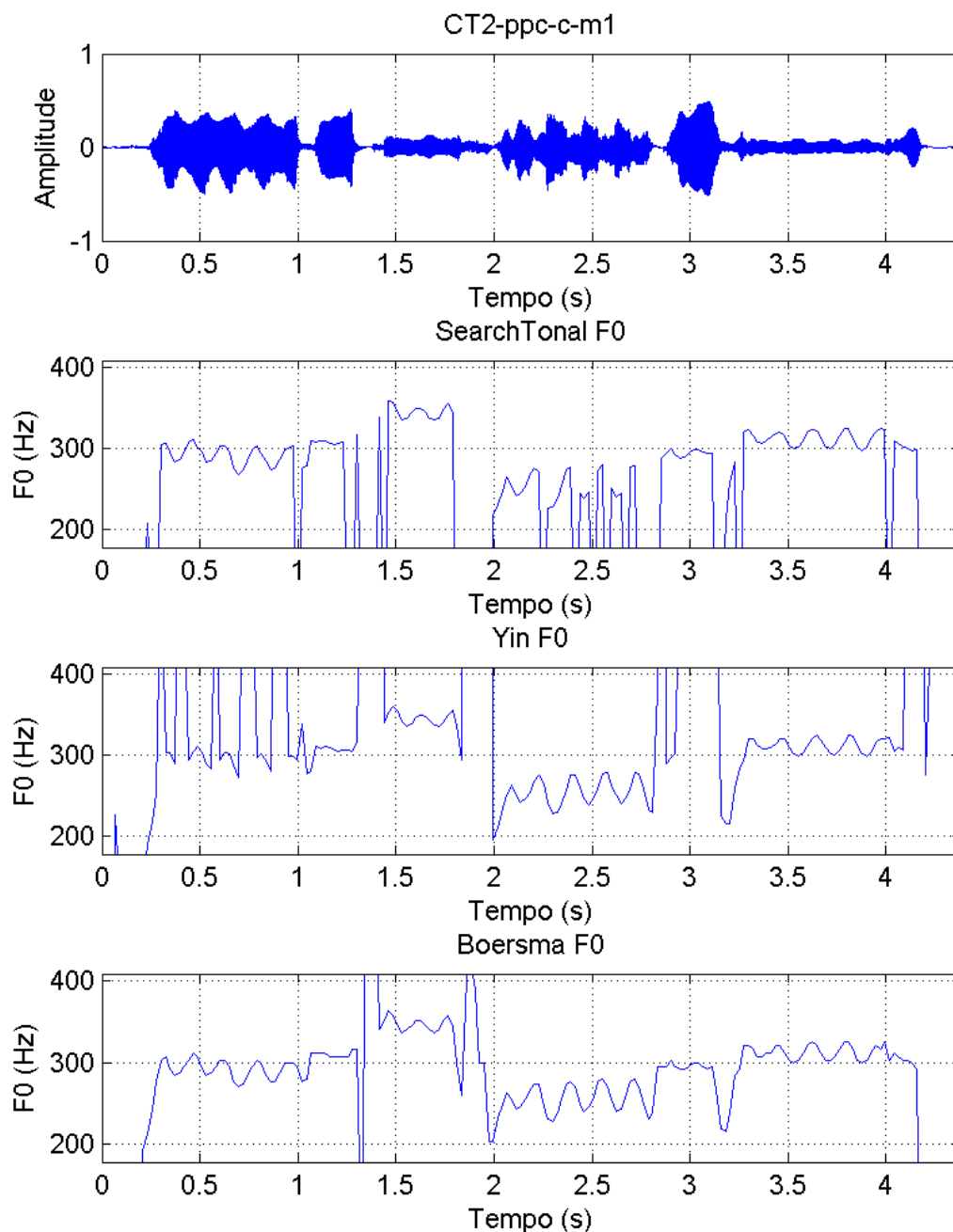


Figura 41 - Gráfico dos desempenhos dos algoritmos estimadores de F0, através da voz natural CT2-ppc-c-m1.

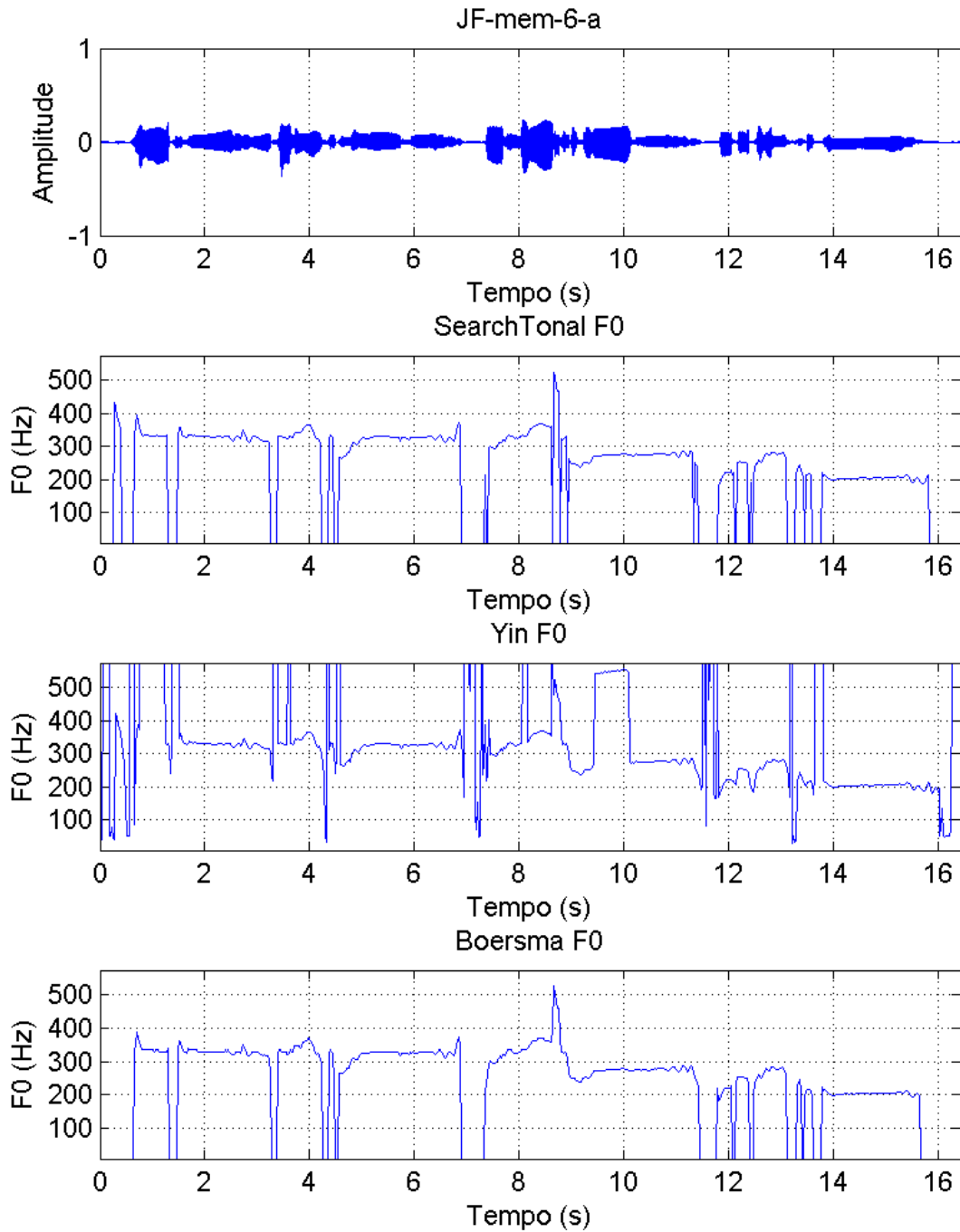


Figura 42 - Gráfico dos desempenhos dos algoritmos estimadores de F0, através da voz natural RE-cresc-A4-a.

## 6.1 - Estimação da Frequência Fundamental

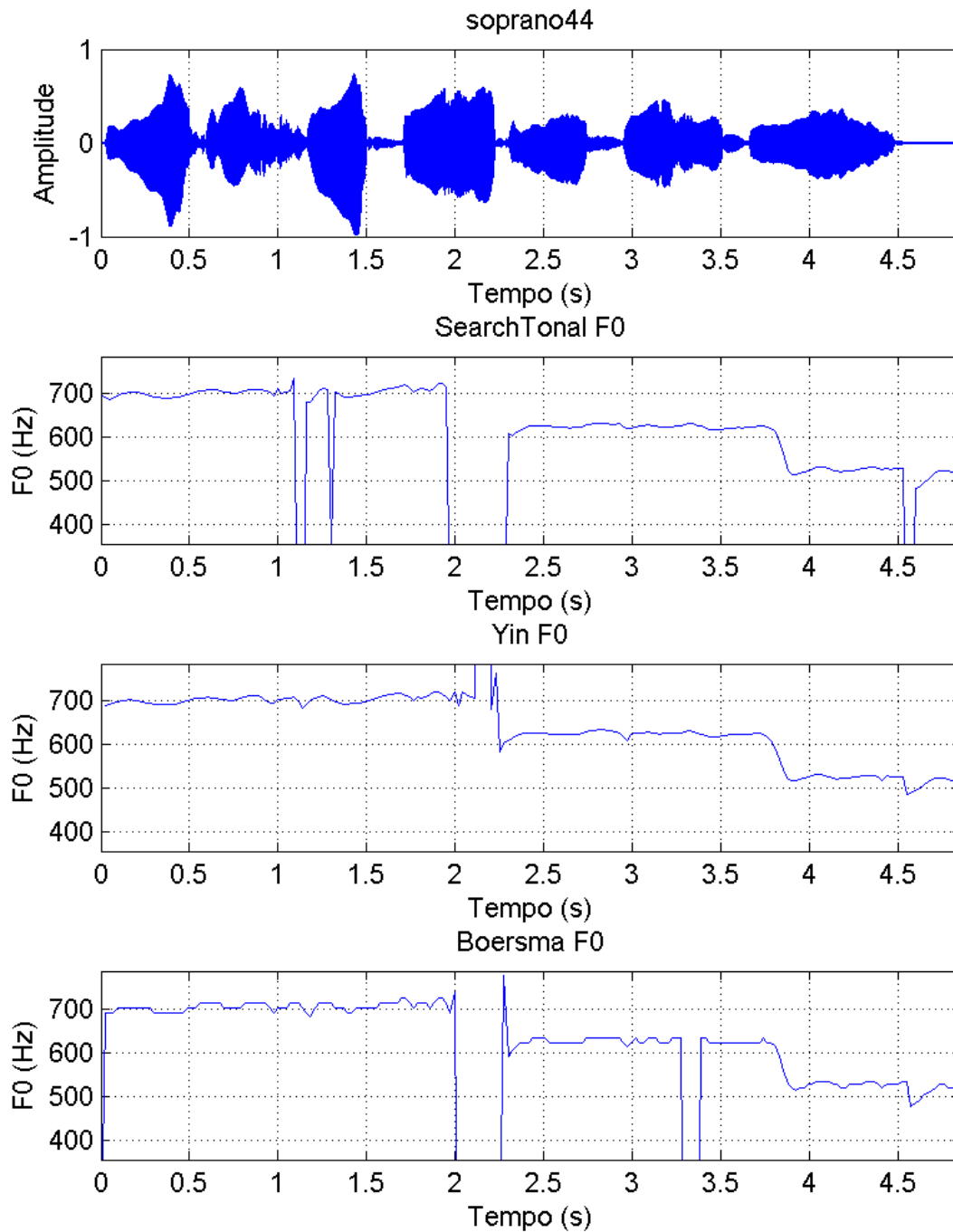


Figura 43 - Gráfico dos desempenhos dos algoritmos estimadores de F0, através da voz natural soprano44.

Tal como no teste elaborado anteriormente, apenas são mostradas as três figuras que ilustram as mais diversas situações encontradas na análise destes três algoritmos. Os resultados dos restantes testes efectuados encontram-se no Anexo C deste documento.

Da análise das figuras 41, 42 e 43, é possível avaliar o desempenho dos algoritmos tendo em conta a sua precisão e robustez. O método *Boesrma* é o que apresenta menor número de falhas na estimação de F0, sendo por isso o mais robusto dos três. Contudo, esta robustez é adquirida em detrimento da precisão, pois dos três algoritmos este é claramente o menos preciso, apresentando inclusive transições de valores de F0 de forma não natural em forma de degraus. Os algoritmos *Yin* e *SearchTonal* apresentam resultados de robustez e precisão muito similares, mas após uma observação mais minuciosa, podemos concluir que o algoritmo *SearchTonal* tem um desempenho superior, principalmente na precisão da estimação dos valores de F0.

## 6.2 - Estimador de Vibrato

Pretende-se nesta secção apresentar os resultados obtidos na avaliação do desempenho do algoritmo estimador de Vibrato. Os algoritmos que compõem o estimador que foram analisados são:

- Algoritmo de estimação da frequência do vibrato;
- Algoritmo que estima o valor da extensão do vibrato;
- Algoritmo que detecta a presença do vibrato e indica a duração deste.

Nas secções seguintes vão ser apresentados os resultados recolhidos, tendo em conta o tipo de voz utilizada em cada teste. As duas primeiras secções dizem respeito aos testes efectuados a partir de vozes sintetizadas. A finalidade de utilizar este tipo de vozes é o de aferir a precisão do algoritmo nas diversas áreas que o caracterizam. Os testes efectuados através de vozes naturais visam reforçar a precisão do algoritmo em situações reais, bem como demonstrar a sua robustez em diversas situações

### 6.2.1 - Teste com Ondas Sinusoidais

Os resultados obtidos no teste efectuado através de ondas sinusoidais vão ser ilustrados através da tabela seguinte.

**Tabela 15** - Resultados obtidos pelo estimador de vibrato nos testes com ondas sinusoidais.

<b>Resultados Calculo Vibrato - Sin Wave</b>						
<i>Nome Associado</i>	<i>F0 (Hz)</i>	<i>F v.(Hz)</i>	<i>F v. máx (Hz)</i>	<i>desvio (Hz)</i>	<i>erro abs (Hz)</i>	<i>erro rel (%)</i>
sin_A4_4Hz	439,999	4,000	4,000	1,86E-04	4,15E-04	0,010
sin_A4_4-5Hz	440,009	4,500	4,500	4,43E-04	1,50E-04	0,003
sin_A4_5Hz	439,999	4,999	5,000	1,71E-05	5,60E-04	0,011
sin_A4_5-5Hz	440,009	5,501	5,501	2,36E-04	6,55E-04	0,012
sin_A4_6Hz	439,999	6,000	6,000	1,08E-04	5,14E-04	0,009
sin_A4_6-5Hz	440,009	6,500	6,501	1,60E-04	2,82E-04	0,004
sin_A4_7Hz	439,999	6,999	6,999	1,38E-05	6,66E-04	0,010
sin_A4_7-5Hz	440,009	7,501	7,501	1,03E-04	6,85E-04	0,009
sin_A4_8Hz	439,999	7,999	7,999	6,00E-05	6,38E-04	0,008

## 6.2.2 - Teste com Vozes Sintetizadas

Nesta secção vão ser apresentados os resultados obtidos usando as vozes sintetizadas descritas no capítulo de planificação de testes, de forma a avaliar a precisão do algoritmo, não só na estimação da frequência, como também da extensão.

Tabela 16 - Resultados obtidos pelo estimador de vibrato nos testes com vozes sintetizadas com uma extensão de Valor 0,5 ST.

<b>Resultados Calculo Vibrato ST=0.5 sem Ataque 2s</b>						
Nome Associado	F0 (Hz)	F v.(Hz)	F v. máx (Hz)	desvio (Hz)	erro abs (Hz)	erro rel (%)
a_A4_4Hz_05	439,886	3,999	4,000	0,001	0,001	0,033
a_A4_4-5Hz_05	439,753	4,498	4,501	0,002	0,003	0,056
a_A4_5Hz_05	439,783	5,001	5,003	0,002	0,002	0,036
a_A4_5-5Hz_05	439,624	5,504	5,505	0,001	0,004	0,075
a_A4_6Hz_05	439,670	5,996	5,998	0,002	0,005	0,075
a_A4_6-5Hz_05	439,528	6,499	6,505	0,005	0,004	0,066
a_A4_7Hz_05	439,665	7,002	7,007	0,007	0,006	0,080
a_A4_7-5Hz_05	439,417	7,501	7,504	0,003	0,002	0,027
a_A4_8Hz_05	439,584	7,995	7,996	0,001	0,005	0,061

Tabela 17 - Resultados obtidos pelo estimador de vibrato nos testes com vozes sintetizadas com uma extensão de Valor 1,0 ST.

<b>Resultados Calculo Vibrato ST=1.0 sem Ataque 2s</b>						
Nome Associado	F0 (Hz)	F v.(Hz)	F v. máx (Hz)	desvio (Hz)	erro abs (Hz)	erro rel (%)
a_A4_4Hz_10	439,794	3,997	3,999	0,001	0,003	0,071
a_A4_4-5Hz_10	439,763	4,499	4,504	0,005	0,004	0,094
a_A4_5Hz_10	439,690	4,996	4,999	0,002	0,004	0,072
a_A4_5-5Hz_10	439,744	5,499	5,502	0,002	0,002	0,031
a_A4_6Hz_10	439,609	6,001	6,002	0,001	0,001	0,019
a_A4_6-5Hz_10	439,222	6,503	6,506	0,002	0,003	0,041
a_A4_7Hz_10	433,985	7,010	7,018	0,009	0,011	0,153
a_A4_7-5Hz_10	438,986	7,502	7,505	0,002	0,002	0,029
a_A4_8Hz_10	439,023	7,983	8,002	0,014	0,018	0,227

**Tabela 18** - Resultados obtidos pelo estimador de vibrato nos testes com vozes sintetizadas com uma frequência de Valor 5,0Hz.

<b>Resultados Calculo Vibrato F0=5.0Hz 2s</b>							
<i>Nome Associado</i>	<i>F0 (Hz)</i>	<i>Fv. (Hz)</i>	<i>Ext (ST)</i>	<i>Ext. máx (ST)</i>	<i>desvio (ST)</i>	<i>erro abs (ST)</i>	<i>erro rel (%)</i>
a_C4_5Hz_02	261,641	5,000	0,190	0,192	0,002	0,010	0,049
a_C4_5Hz_03	261,637	5,003	0,287	0,290	0,002	0,013	0,043
a_C4_5Hz_04	261,627	4,999	0,384	0,388	0,004	0,016	0,039
a_C4_5Hz_05	261,570	4,999	0,483	0,489	0,006	0,017	0,035
a_C4_5Hz_06	261,491	4,998	0,582	0,587	0,004	0,018	0,030
a_C4_5Hz_07	261,520	4,999	0,680	0,687	0,007	0,020	0,028
a_C4_5Hz_08	261,685	5,002	0,780	0,791	0,008	0,020	0,025
a_C4_5Hz_09	261,732	5,002	0,881	0,896	0,014	0,019	0,022
a_C4_5Hz_10	261,721	4,996	0,985	0,996	0,009	0,015	0,015
a_C4_5Hz_11	261,765	5,000	1,081	1,093	0,011	0,019	0,017
a_C4_5Hz_12	261,668	4,996	1,186	1,199	0,011	0,014	0,011

As tabelas 16 e 17 são o resultado dos testes efectuados de forma a determinar a precisão do algoritmo na estimação da frequência do vibrato. Para garantir que estes resultados são independentes da extensão, foram utilizadas dois valores fixos para a extensão, tal como referido na planificação deste teste, onde a tabela 17 apresenta o valor fixo de extensão 0,5ST e a tabela 18 o valor 1,0ST.

Os resultados obtidos são muito importantes e permitem analisar a precisão da estimação da frequência e da extensão de uma forma clara, devido às vozes sintetizadas utilizadas. Tendo em conta a precisão da estimação da frequência do vibrato, podemos verificar que no pior cenário o erro relativo é de 0,153%, sendo a média dos erros relativos de 0,066%, o que nos permite afirmar que este é realmente um método bastante preciso. Através da tabela 18, podemos também avaliar o desempenho da estimação da extensão do vibrato a partir do valor de erro relativo no pior cenário, que é de 0,049%, e da média dos erros relativos, de valor 0,028%. Através destes valores podemos concluir que a estimação da extensão do vibrato é realmente bastante precisa.

No Anexo E deste documento é disponibilizado um conjunto de figuras que correspondem aos gráficos da estimação da frequência do vibrato, bem como da sua extensão. Também são fornecidos os respectivos espectrogramas do sinal, onde é efectuada a marcação do valor da frequência estimada. Essas imagens ilustram os sinais utilizados para elaborar os testes descritos nesta secção e são muito reveladores do próprio funcionamento do algoritmo, pois permitem uma observação no domínio das frequências do sinal de F0, onde é estimado o valor da frequência do vibrato.

### 6.2.3 - Teste com segmentos de Vozes Naturais

A exposição de resultados com base nos testes efectuados da estimação do vibrato, através de vozes naturais, é efectuada da mesma maneira que os testes aplicados aos estimadores de F0.

## 6.2 - Estimador de Vibrato

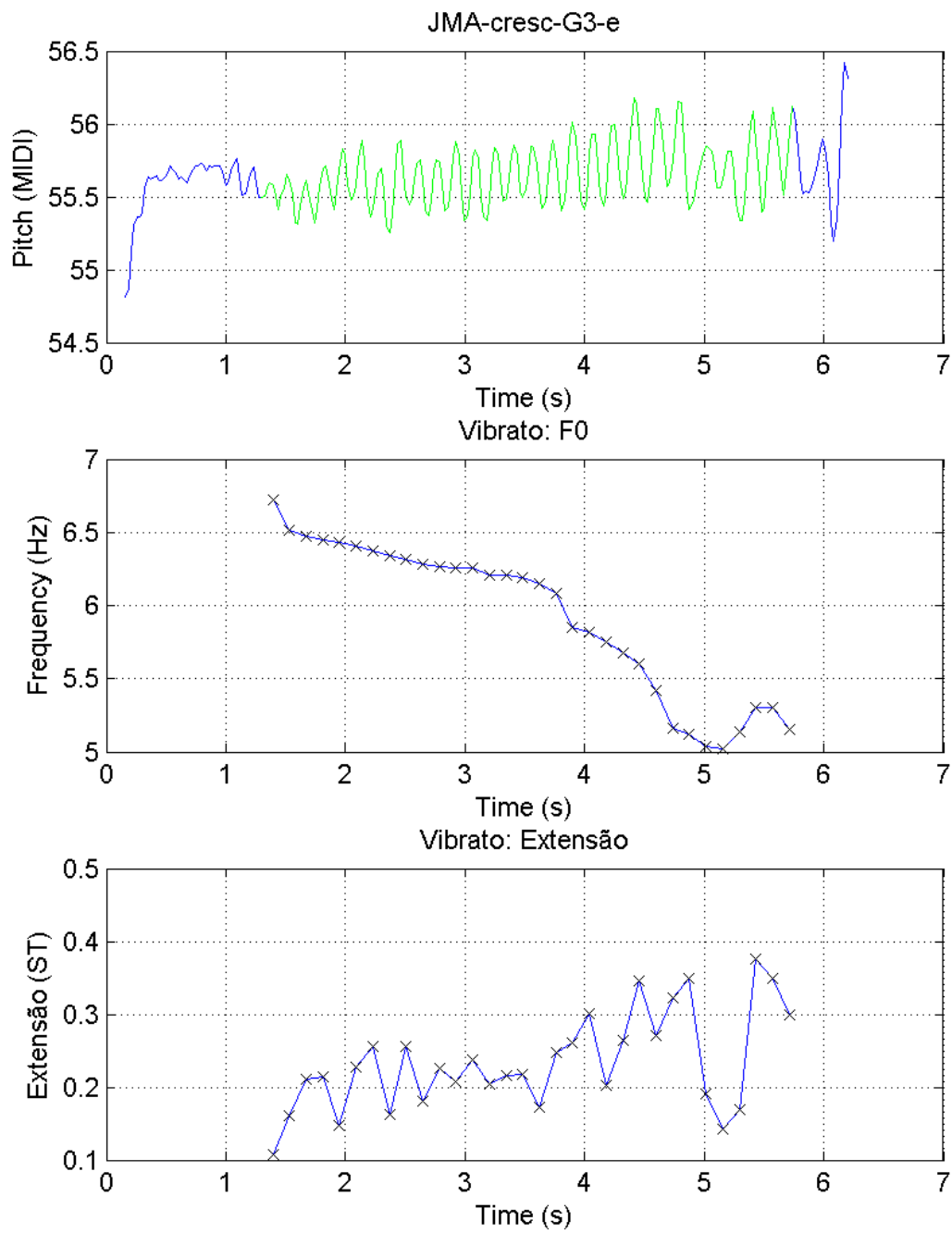


Figura 44 - Desempenho do estimador de vibrato, através da voz JMA-cresc-G3-e.

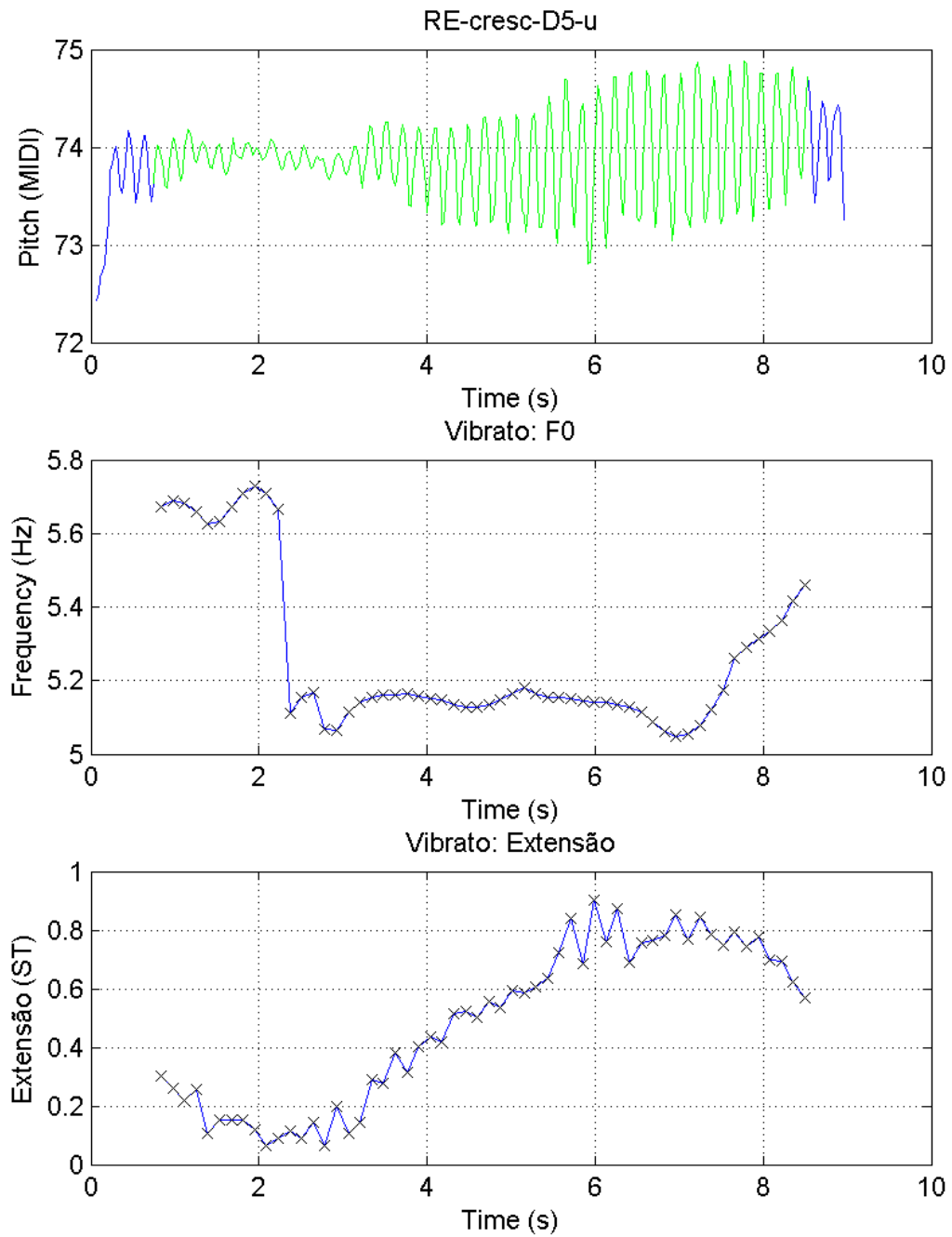


Figura 45 - Desempenho do estimador de vibrato, através da voz RE-cresc-D5-u.

## 6.2 - Estimador de Vibrato

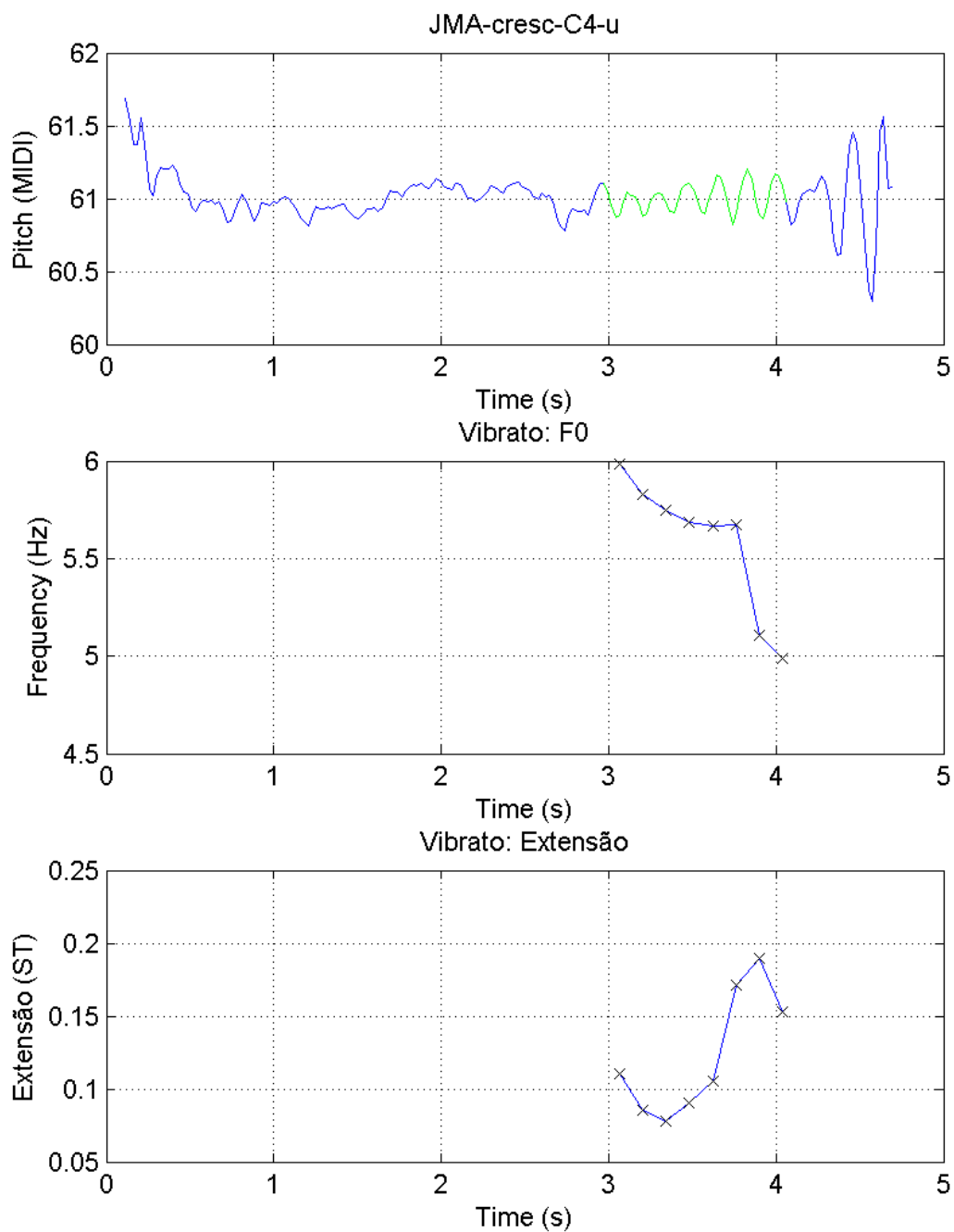


Figura 46 - Desempenho do estimador de vibrato, através da voz JMA-cresc-C4-u.

As três figuras anteriores ilustram os vários resultados das vozes naturais caracterizadas no capítulo anterior no âmbito destes testes. As restantes figuras encontram-se presentes no Anexo F, bem como os espectrogramas respectivos.

### 6.2.4 - Teste com amostras de longa duração

É apresentada nesta secção os resultados obtidos na avaliação do algoritmo de vibrato através de vozes naturais de longa duração e com variações de valores da frequência fundamental. Estes testes representam o cenário mais semelhante quando comparamos com situações de funcionamento normais.

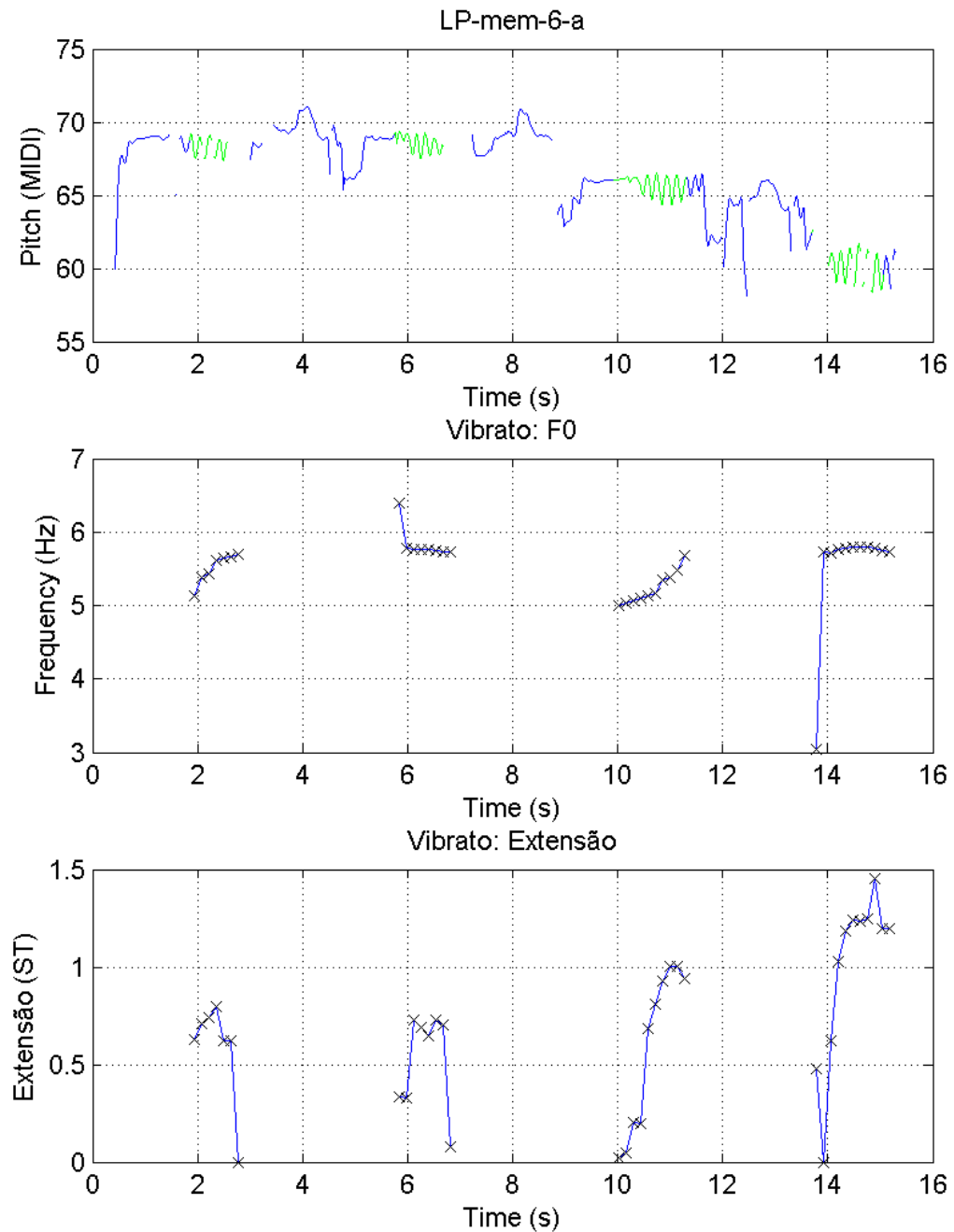


Figura 47 - Desempenho do estimador de vibrato, através da voz LP-cmem-6-a.

## 6.2 - Estimador de Vibrato

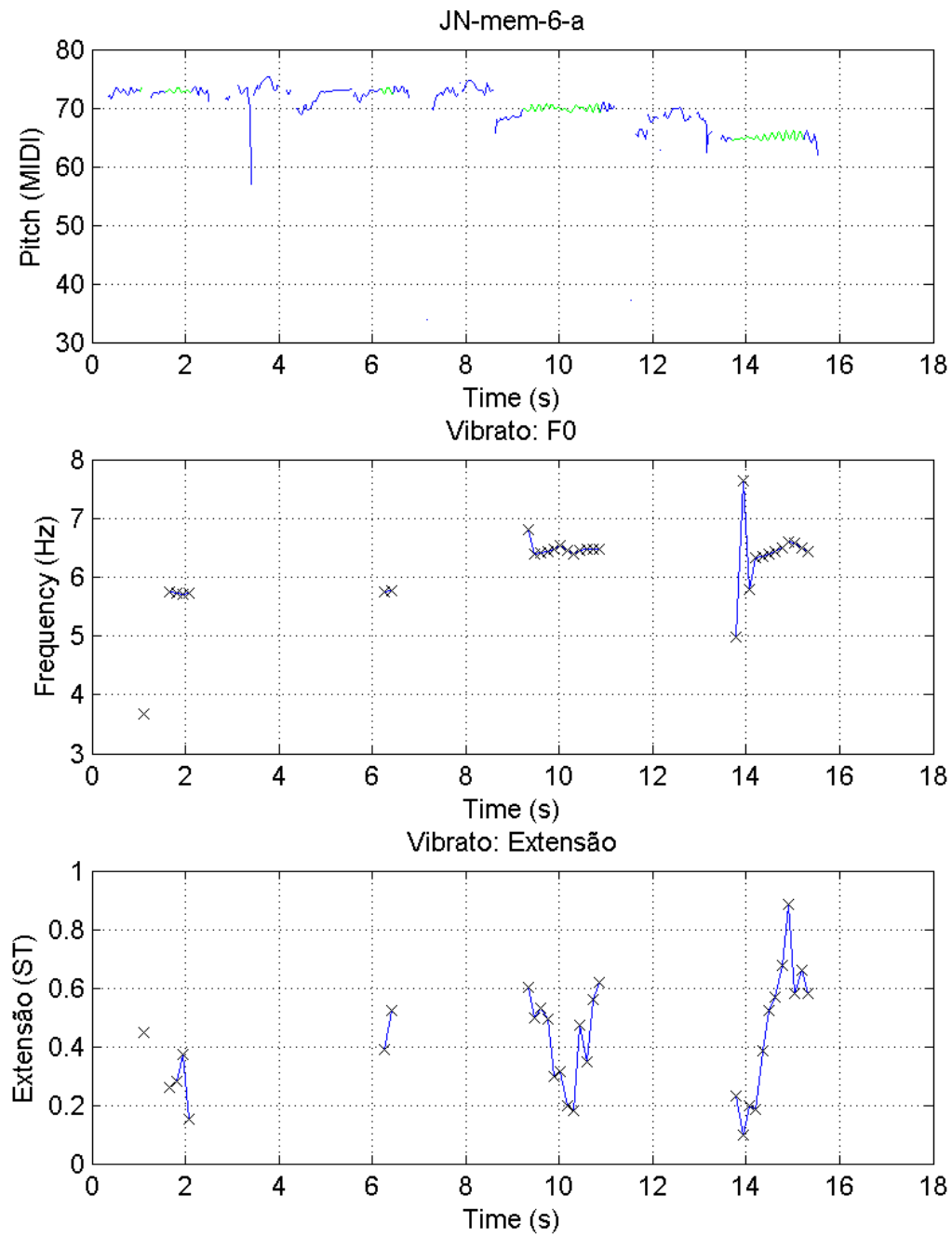


Figura 48 - Desempenho do estimador de vibrato, através da voz JN-cmem-6-a.

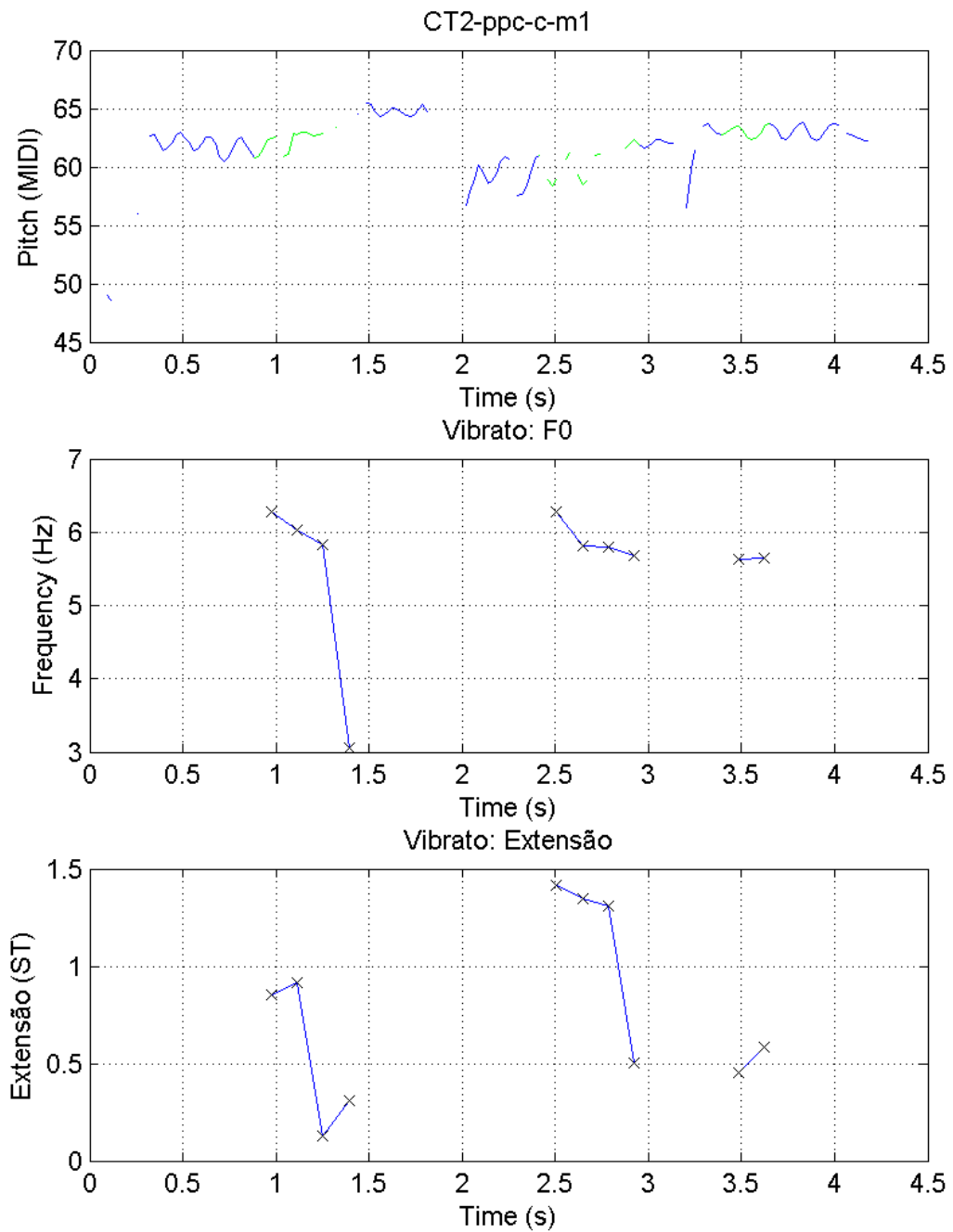


Figura 49 - Desempenho do estimador de vibrato, através da voz CT2-ppc-c-m1.

As restantes figuras encontram-se presentes no Anexo G, bem como os espectrogramas respectivos.

## 6.3 - Discussão de Resultados

Na presente secção será efectuada um breve resumo da discussão de resultados efectuado na exposição dos resultados, abordando a sua relevância e em que condições se aplicam.

### 6.3.1 -Estimação da Frequência Fundamental

Os resultados obtidos na estimação da frequência fundamental são bastante esclarecedores do ponto de vista de performance dos algoritmos. É claramente visível através das tabelas e figuras apresentadas, os pontos fortes e fracos inerentes a cada um dos métodos comparados. A quantidade de informação é também muito completa do ponto de vista de tentar replicar o máximo de situações distintas encontradas em aplicações reais.

Dos resultados obtidos podemos concluir que o método *SearchTonal* é o mais preciso dos três métodos testados, apresentando também uma robustez em condições típicas de sinal muito boa. Contudo, em condições em que a relação sinal-ruído tome valores inferiores a 15dB, o algoritmo tem dificuldade em estimar o real valor de  $F_0$ .

### 6.3.2 -Caracterização do Vibrato

Os resultados que foram obtidos da avaliação do estimador de vibrato são bastante completos e permitem classificar o seu desempenho na estimação de todos os parâmetros que caracterizam o vibrato, bem como o seu funcionamento no seu todo. De facto, podemos com facilidade avaliar a performance do algoritmo na estimação da frequência, na estimação da extensão e da sua correcta detecção e duração, todos eles de forma independente.

A estimação da frequência e extensão do vibrato é efectuada com bastante precisão, e os erros relativos apresentados tomam valores muito satisfatórios, não se prevendo a necessidade de melhoramento num trabalho futuro. A detecção do vibrato, e consequente estimação da duração do vibrato, apresenta também valores de desempenho bastante satisfatórios, sendo o vibrato detectado com sucesso nos testes efectuados. Contudo, existem ainda condições em que dificultam essa detecção, principalmente em regiões de sinal de  $F_0$  com valores em falta, ou seja, situações em que o algoritmo que estima o valor de  $F_0$  não retornou qualquer valor e também em situações em que os valores de  $F_0$  variam de valor bruscamente. O detector de vibrato é, por isso, um algoritmo que pode ainda ser alvo de melhoramentos. Contudo, é aconselhável que a existir um melhoramento, este seja efectuado no algoritmo que estima o valor de  $F_0$ , pois este terá um impacto positivo no funcionamento do algoritmo estimador de vibrato.

# Capítulo 7

## Conclusão

Pretende-se reunir neste capítulo as conclusões retiradas ao longo do desenrolar desta dissertação, bem como dos seus resultados obtidos. O *biofeedback* da voz cantada é um tema muito vasto e com um potencial enorme de evolução. Foi com esta ideia que se partiu para o início desta dissertação, e um vez chegada ao fim, podemos concluir que foram elaborados trabalhos positivos de forma a enriquecer o *biofeedback* da voz cantada, não só sob a forma de análise de novos parâmetros, mas também no melhoramento da análise dos já aplicados.

O trabalho de melhoramentos efectuados foi centrado na performance do algoritmo de estimação da frequência fundamental utilizado no programa *SingingStudio*, o *SearchTonal*. Convém lembrar que este algoritmo é o resultado de vários anos de pesquisa e desenvolvimento, onde sofreu uma evolução contínua e de maturação, facilmente observável nos resultados obtidos. Contudo, foi demonstrado nesta dissertação que existe ainda uma margem de evolução, de forma a o tornar ainda melhor. Na secção seguinte, serão abordadas as conclusões retiradas dos resultados obtidos na estimação da frequência fundamental.

Esta dissertação não só testou alguns métodos de extracção de parâmetros já existentes, como também desenvolveu uma abordagem única na estimação do vibrato. Os resultados obtidos foram positivos sendo agora possível a sua integração na ferramenta *SingingStudio*. As conclusões sobre o seu desempenho em cada um dos parâmetros do vibrato estimado, bem como possíveis direcções de evolução são abordadas na segunda secção deste capítulo.

### 7.1 - Estimação da Frequência Fundamental

Tal como foi mencionado, a informação resultante dos teste aplicados aos três métodos de estimação da frequência fundamental foram muito positivos e permitem avaliar a sua precisão e robustez de forma bastante clara.

Relativamente à precisão, os algoritmos *Yin* e *SearchTonal* tiveram uma prestação consideravelmente superior ao demonstrado pelo algoritmo *Boersma*, ao longo de todos os testes efectuados. Nos testes efectuados com vozes sintetizadas, o algoritmo *Yin* apresenta uma vantagem residual quando comparado com o *SearchTonal*, tendo em conta o erro relativo apresentado. Contudo esta diferença é tão pequena que é mais correcto deduzir que nestas situações ambos os algoritmos têm a mesma performance. Se adicionarmos ruídos a

## 7.2 - Estimador de Vibrato

estes sinais, podemos concluir que o método Yin mantém o seu desempenho, enquanto que o algoritmo *SearchTonal* tem dificuldade em estimar Valores de F0 superiores a  $f \approx 300\text{Hz}$ , com um valor de SNR superior a  $SNR \approx 15\text{dB}$ .

Ao analisarmos os resultados obtidos dos testes efectuados com vozes naturais, obtemos uma imagem mais esclarecedora da performance dos algoritmos. O algoritmo *Boersma* é um método com pouca precisão, quando comparado com os anteriores mas, em condições de sinal com mudanças abruptas de F0, este método é o mais eficaz dos três a devolver um valor de F0, ou seja, nesta condições o algoritmo *Boersma* consegue efectuar uma estimação de F0 embora que pouco precisa, enquanto os restantes algoritmos falham.

Em condições do sinal em que ambos os algoritmos devolvem resultados, podemos verificar que o algoritmo *SearchTonal* é o mais preciso. Embora o algoritmo *Yin* se aproxime, é possível observar em certas zonas do sinal, especialmente nas regiões de vibrato, que o método *SearchTonal* define com mais precisão as mudanças microscópicas de F0, enquanto que o estimador *Yin* tem dificuldade em certas vezes devolver valores na presença de vibrato.

Em resumo, podemos concluir que nas condições normais de funcionamento, o método *SearchTonal* é o mais preciso dos três métodos testados, mas não é o que apresenta maior robustez, quer com sinais na presença de ruído, quer em testes com vozes naturais. Concluímos também que na presença de ruído, em que praticamente os harmónicos são camuflados por este, excepto o valor de F0, o algoritmo *SearchTonal* obtém resultados com um erro relativo elevado. Esta situação piora com o aumento do valor de F0, tal como seria de esperar, devido ao seu método de funcionamento.

O estimador *Yin* é um método bastante preciso, com resultados muito próximos do algoritmo *SearchTonal*. Em relação à sua robustez, este método mostrou-se bastante robusto quando adicionamos ruído a vozes sintetizadas mas nas mudanças abruptas de F0, revelou ser inferior ao algoritmo *Boersma*, e o pior dos três nas vozes naturais com vibrato.

O algoritmo *Boersma* é o exemplo de uma abordagem em que a precisão é preterida em relação à robustez. Este método mostrou-se mais robusto que os restantes nas mudanças abruptas de F0 em vozes reais, mas o seu grau de precisão é fraco, quando comparamos com os outros dois algoritmos e de certa forma até insuficiente.

Na secção referente aos trabalhos futuros é descrito de forma mais precisa os aspectos que se pretendem melhorar no algoritmo *SearchTonal* para que este melhore a sua robustez, quer em sinais com um valor de ruído superior, quer em mudanças de sinal abrupta da frequência fundamental.

## 7.2 - Estimador de Vibrato

A quantidade de informação proveniente dos resultados dos testes efectuados é não só muito positiva como também encorajadora no sentido da implementação deste método no *software SingingStudio*. De realçar que nenhum *software* concorrente consegue executar uma análise do vibrato. Do ponto de vista da estimação do valor da frequência do vibrato podemos concluir que este método é bastante robusto. Quando testado com vozes sintetizadas, o erro relativo médio obtido foi de 0,055%, um resultado que, tendo em conta que o sinal de F0 obtido já provem de um método que introduz um erro, é um resultado bastante positivo. Quando testamos o algoritmo estimador da frequência de vibrato em recurso a vozes naturais verificamos que este método é bastante robusto, não apresentando oscilações abruptas de

valor e mantendo uma forma de onda esperada. Podemos concluir que do ponto de vista da estimação da frequência, este método é bastante preciso.

Os resultados obtidos no âmbito do teste do estimador do valor da extensão do vibrato são também muito positivos. O valor médio do erro relativo, quando testado com vozes sintetizadas, foi de 0,028%, um valor muito positivo. Uma conclusão que podemos retirar prende-se com o facto da relação do erro relativo da extensão, com o erro relativo de  $F_0$ . De certa forma, o erro relativo da extensão do vibrato é uma medida directa da estimação do valor de  $F_0$ , o que permite aferir uma medida de erro do estimador de  $F_0$ . Quando testamos o estimador de extensão do vibrato com vozes naturais, obtemos uma vez mais, resultados que atestam a precisão do estimador de extensão.

A detecção e estimação da duração do vibrato obteve resultados mais modestos, quando comparado com a extracção dos outros parâmetros. De facto, a detecção da existência de vibrato é um dos aspectos chave para a implementação do algoritmo na plataforma *SingingStudio*, e quando testada com vozes naturais com flutuações de  $F_0$ , o detector de vibrato obteve resultados menos positivos. Contudo, em vozes naturais sem variações abruptas de  $F_0$ , as regiões de vibrato foram detectadas com precisão. Uma das conclusões a retirar das diferenças de resultados entre estes dois testes é a da que a detecção de vibrato está intimamente ligada com o bom funcionamento do detector de  $F_0$ . Nas situações em que existem variações do valor estimado de  $F_0$  ou mesmo falha, devido a erros do algoritmo de estimação de  $F_0$ , o detector de vibrato é também afectado, tal como seria de esperar. A reconstrução da janela de  $F_0$  efectuada através de métodos de interpolação minimiza os erros associados a falhas, tal como podemos observar nos resultados obtidos. Também é de realçar a sua performance na não detecção de vibrato em sinais, mesmo este contendo perturbações, de vozes que não contêm vibrato, verificado nos resultados apresentados.

Em todos os testes efectuados que pretendem testar o funcionamento da detecção do vibrato mostram uma impossibilidade de detectar a presença de vibrato nas primeiras e nas últimas 32 amostras. Este problema era esperado pois ocorre devido ao centro da janela aplicada ao sinal não percorrer estas zonas de sinal, ou seja, o vibrato não é detectado nestas zonas pois nunca foi verdadeiramente verificado se este existia nessas regiões. Em situações de teste em que controlamos os sinais, este problema não é relevante. Na secção relacionada com o trabalho futuro, este problema será abordado e uma solução genérica demonstrada.

### 7.3 - Trabalho Futuro

Através dos resultados obtidos podemos planificar futuros objectivos a atingir, devidamente enquadrados no projecto FCT, do qual esta dissertação faz parte. Uma oportunidade de melhorar o algoritmo *SearchTonal* é agora criada, pois sabemos concretamente em que contextos este falha. Existem dois casos concretos que mostram ainda uma margem de melhoramento: na presença de ruído e na falha de valores quando  $F_0$  varia bruscamente. O primeiro ocorre devido à natureza do próprio algoritmo, visto que este utiliza harmónicos presentes relacionados com  $F_0$  para reforçar a sua estimação. A presença de ruído destrói esta abordagem pois camufla os harmónicos e induz o algoritmo em erro.

O trabalho futuro relacionado com a estimação do vibrato passa por um pequeno melhoramento da detecção do vibrato, principalmente quando o sinal de  $F_0$  contém variações abruptas, em que a sua forma acaba por se assemelhar com uma onda triangular ou sinusoidal, o que induz em erro o detector se estas acabarem por ter características

### 7.3 - Trabalho Futuro

semelhantes com as esperadas. O segundo trabalho a realizar é a integração deste algoritmo no *SingingStudio*, que será efectuado ainda no âmbito desta dissertação, durante o período de tempo presente até à sua defesa perante o júri. Tal como foi mencionado anteriormente, nos testes efectuados foi constatada a impossibilidade de detectar a presença de vibrato nas primeiras e nas últimas 32 amostras. Um método eficaz utilizado em situações semelhantes é o de efectuar *padding* no início e no fim do sinal a ser analisado, com amostras de valor médio igual ao calculado nas primeiras e últimas 32 amostras do sinal. Na implementação do estimador no *SingingStudio*, este será o método utilizado.

## Anexo A

Neste anexo são apresentadas as tabelas referentes aos restantes resultados obtidos através dos testes efectuados aos algoritmos de estimação de F0 através do uso de vozes sintetizadas e com ruído.

**Tabela A1** - Resultados obtidos pelo algoritmo *SearchTonal* através de vozes sintetizadas com um SNR=30db.

ficheiro	<b>Resultados Searchtonal - calculo F0 Ruído Branco 30dB</b>				
	média (Hz)	máx. (Hz)	desvio (Hz)	erro abs (Hz)	erro rel (%)
a_98Hz_G2_F0	97,880	100,495	0,862	0,667	<b>0,681%</b>
a_110Hz_A2_F0	109,967	112,345	1,060	0,869	<b>0,790%</b>
a_123_5Hz_B2_F0	123,483	125,427	0,865	0,672	<b>0,545%</b>
a_130_8Hz_C3_F0	130,713	132,819	0,851	0,662	<b>0,506%</b>
a_146_8Hz_D3_F0	146,849	148,571	0,844	0,686	<b>0,467%</b>
a_164_8Hz_E3_F0	164,779	166,718	0,916	0,749	<b>0,455%</b>
a_174_6Hz_F3_F0	174,423	176,245	0,741	0,617	<b>0,354%</b>
a_196Hz_G3_F0	196,088	198,481	0,850	0,681	<b>0,347%</b>
a_220Hz_A3_F0	220,038	222,734	0,973	0,737	<b>0,335%</b>
a_246_9Hz_B3_F0	246,937	249,122	0,885	0,684	<b>0,277%</b>
a_261_6Hz_C4_F0	261,623	263,584	0,971	0,798	<b>0,305%</b>
a_293_7Hz_D4_F0	293,646	295,953	0,849	0,652	<b>0,222%</b>
a_329_6Hz_E4_F0	329,637	331,281	0,769	0,591	<b>0,179%</b>
a_349_2Hz_F4_F0	349,386	351,144	0,725	0,604	<b>0,173%</b>
a_392Hz_G4_F0	392,549	394,441	0,804	0,777	<b>0,198%</b>
a_440Hz_A4_F0	440,322	442,875	0,927	0,769	<b>0,175%</b>
a_493_9Hz_B4_F0	486,502	498,828	44,315	8,778	<b>1,777%</b>
a_523_3Hz_C5_F0	524,091	525,918	1,055	1,095	<b>0,209%</b>
a_587_3Hz_D5_F0	588,164	594,974	4,704	4,556	<b>0,776%</b>
a_659_3Hz_E5_F0	662,025	667,418	5,666	6,087	<b>0,923%</b>
a_698_5Hz_F5_F0	699,940	711,483	7,880	7,553	<b>1,081%</b>
a_784Hz_G5_F0	787,304	817,530	8,590	4,679	<b>0,597%</b>

**Tabela A2** - Resultados obtidos pelo algoritmo *SearchTonal* através de vozes sintetizadas com um SNR=25db.

ficheiro	<b>Resultados Searchtonal - calculo F0 Ruido Branco 25dB</b>				
	média (Hz)	máx. (Hz)	desvio (Hz)	erro abs (Hz)	erro rel (%)
a_98Hz_G2_F0	97,879	100,549	0,865	0,671	<b>0,685%</b>
a_110Hz_A2_F0	109,969	112,330	1,062	0,869	<b>0,790%</b>
a_123_5Hz_B2_F0	123,480	125,418	0,868	0,675	<b>0,546%</b>
a_130_8Hz_C3_F0	130,718	132,827	0,851	0,660	<b>0,504%</b>
a_146_8Hz_D3_F0	146,852	148,584	0,845	0,688	<b>0,468%</b>
a_164_8Hz_E3_F0	164,776	166,750	0,916	0,747	<b>0,453%</b>
a_174_6Hz_F3_F0	174,425	176,176	0,739	0,616	<b>0,353%</b>
a_196Hz_G3_F0	196,089	198,156	0,835	0,665	<b>0,339%</b>
a_220Hz_A3_F0	220,024	222,446	0,965	0,761	<b>0,346%</b>
a_246_9Hz_B3_F0	246,949	249,339	0,913	0,708	<b>0,287%</b>
a_261_6Hz_C4_F0	261,693	263,385	0,981	0,811	<b>0,310%</b>
a_293_7Hz_D4_F0	293,648	295,512	0,768	0,592	<b>0,202%</b>
a_329_6Hz_E4_F0	329,631	331,899	0,831	0,670	<b>0,203%</b>
a_349_2Hz_F4_F0	349,354	351,209	0,863	0,649	<b>0,186%</b>
a_392Hz_G4_F0	392,202	393,882	1,174	0,902	<b>0,230%</b>
a_440Hz_A4_F0	440,434	442,553	0,859	0,768	<b>0,175%</b>
a_493_9Hz_B4_F0	481,647	498,744	52,264	13,564	<b>2,746%</b>
a_523_3Hz_C5_F0	523,853	536,704	2,044	1,213	<b>0,232%</b>
a_587_3Hz_D5_F0	588,419	595,148	4,979	4,889	<b>0,832%</b>
a_659_3Hz_E5_F0	659,077	668,071	6,711	6,460	<b>0,980%</b>
a_698_5Hz_F5_F0	697,827	711,604	7,979	7,561	<b>1,083%</b>
a_784Hz_G5_F0	775,829	819,568	61,061	14,920	<b>1,903%</b>

**Tabela A3** - Resultados obtidos pelo algoritmo *SearchTonal* através de vozes sintetizadas com um SNR=20db.

ficheiro	<b>Resultados searchtonal - calculo F0 Ruido Branco 20dB</b>				
	média (Hz)	máx. (Hz)	desvio (Hz)	erro abs (Hz)	erro rel (%)
a_98Hz_G2_F0	97,873	100,462	0,867	0,676	<b>0,690%</b>
a_110Hz_A2_F0	109,974	112,312	1,061	0,871	<b>0,792%</b>
a_123_5Hz_B2_F0	123,480	125,491	0,867	0,669	<b>0,542%</b>
a_130_8Hz_C3_F0	130,714	132,793	0,862	0,679	<b>0,519%</b>
a_146_8Hz_D3_F0	146,850	148,563	0,845	0,686	<b>0,467%</b>
a_164_8Hz_E3_F0	164,779	166,736	0,917	0,751	<b>0,456%</b>
a_174_6Hz_F3_F0	174,426	176,237	0,750	0,627	<b>0,359%</b>
a_196Hz_G3_F0	196,066	198,129	0,845	0,678	<b>0,346%</b>
a_220Hz_A3_F0	220,002	223,377	0,998	0,747	<b>0,340%</b>
a_246_9Hz_B3_F0	246,953	249,162	0,917	0,747	<b>0,303%</b>
a_261_6Hz_C4_F0	261,593	263,564	0,976	0,824	<b>0,315%</b>
a_293_7Hz_D4_F0	293,688	296,251	0,878	0,668	<b>0,228%</b>
a_329_6Hz_E4_F0	329,658	331,593	0,776	0,611	<b>0,185%</b>
a_349_2Hz_F4_F0	349,359	353,014	1,007	0,746	<b>0,214%</b>
a_392Hz_G4_F0	392,163	394,157	1,323	1,083	<b>0,276%</b>
a_440Hz_A4_F0	437,704	443,439	23,875	3,629	<b>0,825%</b>
a_493_9Hz_B4_F0	463,378	499,865	82,829	32,171	<b>6,514%</b>
a_523_3Hz_C5_F0	517,417	526,259	40,082	7,381	<b>1,411%</b>
a_587_3Hz_D5_F0	565,535	595,854	80,631	26,263	<b>4,472%</b>
a_659_3Hz_E5_F0	641,091	668,341	78,552	25,564	<b>3,878%</b>
a_698_5Hz_F5_F0	685,265	712,952	65,066	19,786	<b>2,833%</b>
a_784Hz_G5_F0	736,642	818,103	132,575	54,649	<b>6,971%</b>

**Tabela A4** - Resultados obtidos pelo algoritmo *SearchTonal* através de vozes sintetizadas com um SNR=15db.

ficheiro	<b>Resultados searchtonal - calculo F0 Ruido Branco 15dB</b>				
	média (Hz)	máx. (Hz)	desvio (Hz)	erro abs (Hz)	erro rel (%)
a_98Hz_G2_F0	97,867	100,639	0,872	0,677	<b>0,691%</b>
a_110Hz_A2_F0	109,967	112,311	1,069	0,875	<b>0,796%</b>
a_123_5Hz_B2_F0	123,476	125,449	0,877	0,683	<b>0,553%</b>
a_130_8Hz_C3_F0	130,701	132,806	0,844	0,664	<b>0,507%</b>
a_146_8Hz_D3_F0	146,855	148,559	0,849	0,689	<b>0,470%</b>
a_164_8Hz_E3_F0	164,779	166,723	0,918	0,747	<b>0,453%</b>
a_174_6Hz_F3_F0	174,409	176,202	0,740	0,624	<b>0,357%</b>
a_196Hz_G3_F0	196,077	198,520	0,881	0,719	<b>0,367%</b>
a_220Hz_A3_F0	220,050	223,130	1,057	0,828	<b>0,376%</b>
a_246_9Hz_B3_F0	246,961	249,944	0,973	0,771	<b>0,312%</b>
a_261_6Hz_C4_F0	261,681	263,878	1,054	0,851	<b>0,325%</b>
a_293_7Hz_D4_F0	293,587	296,379	0,934	0,702	<b>0,239%</b>
a_329_6Hz_E4_F0	329,556	332,820	0,902	0,711	<b>0,216%</b>
a_349_2Hz_F4_F0	349,333	352,002	1,125	0,897	<b>0,257%</b>
a_392Hz_G4_F0	387,258	394,274	29,938	5,817	<b>1,484%</b>
a_440Hz_A4_F0	411,642	447,544	74,326	29,631	<b>6,734%</b>
a_493_9Hz_B4_F0	422,301	500,349	120,092	73,429	<b>14,868%</b>
a_523_3Hz_C5_F0	511,299	537,810	56,006	14,109	<b>2,697%</b>
a_587_3Hz_D5_F0	523,316	595,260	124,224	68,037	<b>11,584%</b>
a_659_3Hz_E5_F0	573,637	668,260	145,135	90,288	<b>13,695%</b>
a_698_5Hz_F5_F0	606,410	714,270	164,402	98,325	<b>14,078%</b>
a_784Hz_G5_F0	630,780	819,966	207,606	162,781	<b>20,763%</b>

**Tabela A5** - Resultados obtidos pelo algoritmo *SearchTonal* através de vozes sintetizadas com um SNR=10db.

ficheiro	<b>Resultados searchtonal - calculo F0 Ruido Branco 10dB</b>				
	média (Hz)	máx. (Hz)	desvio (Hz)	erro abs (Hz)	erro rel (%)
a_98Hz_G2_F0	97,882	100,668	0,859	0,670	<b>0,684%</b>
a_110Hz_A2_F0	109,979	112,303	1,065	0,847	<b>0,770%</b>
a_123_5Hz_B2_F0	123,496	125,767	0,896	0,712	<b>0,577%</b>
a_130_8Hz_C3_F0	130,726	132,901	0,893	0,695	<b>0,531%</b>
a_146_8Hz_D3_F0	146,844	148,333	0,843	0,694	<b>0,473%</b>
a_164_8Hz_E3_F0	164,780	166,796	0,939	0,764	<b>0,463%</b>
a_174_6Hz_F3_F0	174,423	176,319	0,758	0,631	<b>0,361%</b>
a_196Hz_G3_F0	194,870	198,387	10,574	1,827	<b>0,932%</b>
a_220Hz_A3_F0	218,767	222,530	12,019	2,095	<b>0,952%</b>
a_246_9Hz_B3_F0	246,949	249,092	0,870	0,694	<b>0,281%</b>
a_261_6Hz_C4_F0	258,529	264,061	19,965	3,868	<b>1,479%</b>
a_293_7Hz_D4_F0	283,338	296,932	37,810	11,072	<b>3,770%</b>
a_329_6Hz_E4_F0	312,212	332,322	50,988	18,148	<b>5,506%</b>
a_349_2Hz_F4_F0	336,844	352,915	45,160	13,334	<b>3,818%</b>
a_392Hz_G4_F0	365,853	401,067	68,481	27,911	<b>7,120%</b>
a_440Hz_A4_F0	372,417	449,868	103,939	69,273	<b>15,744%</b>
a_493_9Hz_B4_F0	406,240	499,916	134,569	89,662	<b>18,155%</b>
a_523_3Hz_C5_F0	440,435	538,053	125,575	85,443	<b>16,329%</b>
a_587_3Hz_D5_F0	507,424	594,365	138,196	83,568	<b>14,229%</b>
a_659_3Hz_E5_F0	519,976	670,142	187,415	143,822	<b>21,816%</b>
a_698_5Hz_F5_F0	496,362	711,730	196,639	205,997	<b>29,493%</b>
a_784Hz_G5_F0	552,988	817,791	239,130	239,353	<b>30,530%</b>

**Tabela A6** - Resultados obtidos pelo algoritmo *SearchTonal* através de vozes sintetizadas com um SNR=5db.

ficheiro	<b>Resultados searchtonal - calculo F0 Ruído Branco 5dB</b>				
	média (Hz)	máx. (Hz)	desvio (Hz)	erro abs (Hz)	erro rel (%)
a_98Hz_G2_F0	101,218	406,192	31,810	4,065	<b>4,148%</b>
a_110Hz_A2_F0	110,069	113,083	1,223	0,998	<b>0,907%</b>
a_123_5Hz_B2_F0	123,498	126,020	1,017	0,785	<b>0,635%</b>
a_130_8Hz_C3_F0	130,756	133,974	1,047	0,832	<b>0,636%</b>
a_146_8Hz_D3_F0	146,835	149,015	0,937	0,760	<b>0,518%</b>
a_164_8Hz_E3_F0	164,734	167,127	1,014	0,829	<b>0,503%</b>
a_174_6Hz_F3_F0	174,429	177,189	0,924	0,730	<b>0,418%</b>
a_196Hz_G3_F0	196,133	200,579	1,214	0,917	<b>0,468%</b>
a_220Hz_A3_F0	220,051	223,598	1,371	1,055	<b>0,479%</b>
a_246_9Hz_B3_F0	246,969	251,477	1,259	0,975	<b>0,395%</b>
a_261_6Hz_C4_F0	260,085	265,710	14,223	2,778	<b>1,062%</b>
a_293_7Hz_D4_F0	293,670	299,349	1,495	1,157	<b>0,394%</b>
a_329_6Hz_E4_F0	328,976	336,194	12,725	2,304	<b>0,699%</b>
a_349_2Hz_F4_F0	344,212	354,921	29,661	6,662	<b>1,908%</b>
a_392Hz_G4_F0	372,331	408,485	59,198	22,358	<b>5,704%</b>
a_440Hz_A4_F0	420,891	452,568	62,772	22,611	<b>5,139%</b>
a_493_9Hz_B4_F0	432,730	514,683	107,750	63,850	<b>12,928%</b>
a_523_3Hz_C5_F0	465,884	539,533	110,499	60,726	<b>11,606%</b>
a_587_3Hz_D5_F0	493,913	611,827	143,186	97,588	<b>16,615%</b>
a_659_3Hz_E5_F0	540,098	670,825	166,005	123,399	<b>18,718%</b>
a_698_5Hz_F5_F0	530,480	735,233	194,173	172,647	<b>24,718%</b>
a_784Hz_G5_F0	541,467	822,996	237,346	253,526	<b>32,338%</b>

Tabela A7 - Resultados obtidos pelo algoritmo Yin através de vozes sintetizadas com um SNR=30db.

ficheiro	Resultados yin - calculo F0 Ruído Branco 30dB				
	média (Hz)	máx. (Hz)	desvio (Hz)	erro abs (Hz)	erro rel (%)
a_98Hz_G2_F0	97,916	100,410	0,829	0,636	<b>0,649%</b>
a_110Hz_A2_F0	109,942	112,194	1,061	0,870	<b>0,791%</b>
a_123_5Hz_B2_F0	123,475	125,368	0,881	0,687	<b>0,557%</b>
a_130_8Hz_C3_F0	130,707	132,837	0,858	0,668	<b>0,511%</b>
a_146_8Hz_D3_F0	146,854	148,485	0,855	0,696	<b>0,474%</b>
a_164_8Hz_E3_F0	164,817	166,694	0,899	0,732	<b>0,444%</b>
a_174_6Hz_F3_F0	174,448	176,367	0,731	0,601	<b>0,344%</b>
a_196Hz_G3_F0	196,087	198,267	0,842	0,668	<b>0,341%</b>
a_220Hz_A3_F0	220,034	222,876	0,990	0,770	<b>0,350%</b>
a_246_9Hz_B3_F0	246,947	249,189	0,893	0,702	<b>0,284%</b>
a_261_6Hz_C4_F0	261,624	263,595	0,884	0,735	<b>0,281%</b>
a_293_7Hz_D4_F0	293,652	295,673	0,828	0,635	<b>0,216%</b>
a_329_6Hz_E4_F0	329,708	331,114	0,736	0,586	<b>0,178%</b>
a_349_2Hz_F4_F0	349,230	350,934	0,808	0,626	<b>0,179%</b>
a_392Hz_G4_F0	392,016	393,452	0,762	0,630	<b>0,161%</b>
a_440Hz_A4_F0	439,975	442,622	0,883	0,673	<b>0,153%</b>
a_493_9Hz_B4_F0	493,824	495,247	0,696	0,593	<b>0,120%</b>
a_523_3Hz_C5_F0	523,192	524,587	0,725	0,587	<b>0,112%</b>
a_587_3Hz_D5_F0	587,280	589,741	1,149	0,937	<b>0,159%</b>
a_659_3Hz_E5_F0	659,275	661,211	0,931	0,758	<b>0,115%</b>
a_698_5Hz_F5_F0	698,261	700,483	1,026	0,838	<b>0,120%</b>
a_784Hz_G5_F0	784,162	786,254	1,103	0,914	<b>0,117%</b>

Tabela A8 - Resultados obtidos pelo algoritmo Yin através de vozes sintetizadas com um SNR=25db.

ficheiro	Resultados yin - calculo FO Ruído Branco 25dB				
	média (Hz)	máx. (Hz)	desvio (Hz)	erro abs (Hz)	erro rel (%)
a_98Hz_G2_F0	97,916	100,406	0,829	0,636	<b>0,649%</b>
a_110Hz_A2_F0	109,942	112,194	1,061	0,870	<b>0,791%</b>
a_123_5Hz_B2_F0	123,475	125,366	0,881	0,687	<b>0,556%</b>
a_130_8Hz_C3_F0	130,707	132,842	0,857	0,668	<b>0,511%</b>
a_146_8Hz_D3_F0	146,854	148,495	0,855	0,695	<b>0,474%</b>
a_164_8Hz_E3_F0	164,817	166,711	0,899	0,732	<b>0,444%</b>
a_174_6Hz_F3_F0	174,448	176,370	0,730	0,601	<b>0,344%</b>
a_196Hz_G3_F0	196,086	198,265	0,842	0,669	<b>0,341%</b>
a_220Hz_A3_F0	220,034	222,879	0,990	0,770	<b>0,350%</b>
a_246_9Hz_B3_F0	246,947	249,190	0,893	0,702	<b>0,284%</b>
a_261_6Hz_C4_F0	261,624	263,593	0,883	0,736	<b>0,281%</b>
a_293_7Hz_D4_F0	293,652	295,663	0,829	0,637	<b>0,217%</b>
a_329_6Hz_E4_F0	329,708	331,100	0,736	0,586	<b>0,178%</b>
a_349_2Hz_F4_F0	349,230	350,942	0,807	0,627	<b>0,179%</b>
a_392Hz_G4_F0	392,015	393,481	0,762	0,630	<b>0,161%</b>
a_440Hz_A4_F0	439,975	442,602	0,883	0,675	<b>0,153%</b>
a_493_9Hz_B4_F0	493,824	495,215	0,693	0,591	<b>0,120%</b>
a_523_3Hz_C5_F0	523,193	524,583	0,727	0,587	<b>0,112%</b>
a_587_3Hz_D5_F0	587,280	589,751	1,151	0,934	<b>0,159%</b>
a_659_3Hz_E5_F0	659,274	661,265	0,934	0,761	<b>0,115%</b>
a_698_5Hz_F5_F0	698,262	700,560	1,023	0,838	<b>0,120%</b>
a_784Hz_G5_F0	784,161	786,226	1,098	0,904	<b>0,115%</b>

Tabela A9 - Resultados obtidos pelo algoritmo *Yin* através de vozes sintetizadas com um SNR=20db.

ficheiro	Resultados yin - calculo F0 Ruido Branco 20dB				
	média (Hz)	máx. (Hz)	desvio (Hz)	erro abs (Hz)	erro rel (%)
a_98Hz_G2_F0	97,916	100,416	0,829	0,635	<b>0,648%</b>
a_110Hz_A2_F0	109,942	112,191	1,060	0,870	<b>0,791%</b>
a_123_5Hz_B2_F0	123,475	125,358	0,880	0,686	<b>0,555%</b>
a_130_8Hz_C3_F0	130,707	132,829	0,858	0,668	<b>0,511%</b>
a_146_8Hz_D3_F0	146,854	148,488	0,855	0,696	<b>0,474%</b>
a_164_8Hz_E3_F0	164,817	166,714	0,900	0,733	<b>0,445%</b>
a_174_6Hz_F3_F0	174,448	176,385	0,731	0,601	<b>0,344%</b>
a_196Hz_G3_F0	196,087	198,271	0,841	0,668	<b>0,341%</b>
a_220Hz_A3_F0	220,034	222,877	0,991	0,773	<b>0,351%</b>
a_246_9Hz_B3_F0	246,947	249,194	0,894	0,702	<b>0,284%</b>
a_261_6Hz_C4_F0	261,625	263,604	0,887	0,739	<b>0,282%</b>
a_293_7Hz_D4_F0	293,652	295,656	0,825	0,633	<b>0,216%</b>
a_329_6Hz_E4_F0	329,708	331,155	0,737	0,585	<b>0,178%</b>
a_349_2Hz_F4_F0	349,230	350,884	0,808	0,626	<b>0,179%</b>
a_392Hz_G4_F0	392,015	393,493	0,767	0,634	<b>0,162%</b>
a_440Hz_A4_F0	439,974	442,657	0,884	0,669	<b>0,152%</b>
a_493_9Hz_B4_F0	493,827	495,279	0,699	0,593	<b>0,120%</b>
a_523_3Hz_C5_F0	523,193	524,532	0,726	0,588	<b>0,112%</b>
a_587_3Hz_D5_F0	587,278	589,986	1,148	0,933	<b>0,159%</b>
a_659_3Hz_E5_F0	659,274	661,240	0,925	0,752	<b>0,114%</b>
a_698_5Hz_F5_F0	698,259	700,384	1,020	0,824	<b>0,118%</b>
a_784Hz_G5_F0	784,166	786,431	1,115	0,924	<b>0,118%</b>

Tabela A10 - Resultados obtidos pelo algoritmo *Yin* através de vozes sintetizadas com um SNR=15db.

ficheiro	Resultados <i>yin</i> - calculo F0 Ruído Branco 15dB				
	média (Hz)	máx. (Hz)	desvio (Hz)	erro abs (Hz)	erro rel (%)
a_98Hz_G2_F0	97,916	100,405	0,829	0,636	<b>0,649%</b>
a_110Hz_A2_F0	109,941	112,181	1,060	0,869	<b>0,790%</b>
a_123_5Hz_B2_F0	123,475	125,367	0,881	0,688	<b>0,557%</b>
a_130_8Hz_C3_F0	130,707	132,873	0,858	0,667	<b>0,510%</b>
a_146_8Hz_D3_F0	146,854	148,494	0,856	0,696	<b>0,474%</b>
a_164_8Hz_E3_F0	164,817	166,701	0,897	0,729	<b>0,443%</b>
a_174_6Hz_F3_F0	174,448	176,364	0,732	0,601	<b>0,344%</b>
a_196Hz_G3_F0	196,086	198,241	0,843	0,676	<b>0,345%</b>
a_220Hz_A3_F0	220,035	222,902	0,991	0,772	<b>0,351%</b>
a_246_9Hz_B3_F0	246,946	249,286	0,895	0,698	<b>0,283%</b>
a_261_6Hz_C4_F0	261,625	263,641	0,884	0,736	<b>0,281%</b>
a_293_7Hz_D4_F0	293,654	295,612	0,830	0,640	<b>0,218%</b>
a_329_6Hz_E4_F0	329,706	331,140	0,737	0,589	<b>0,179%</b>
a_349_2Hz_F4_F0	349,229	350,968	0,808	0,633	<b>0,181%</b>
a_392Hz_G4_F0	392,018	393,584	0,768	0,634	<b>0,162%</b>
a_440Hz_A4_F0	439,976	442,740	0,890	0,666	<b>0,151%</b>
a_493_9Hz_B4_F0	493,827	495,316	0,695	0,589	<b>0,119%</b>
a_523_3Hz_C5_F0	523,195	524,629	0,731	0,591	<b>0,113%</b>
a_587_3Hz_D5_F0	587,282	589,768	1,141	0,921	<b>0,157%</b>
a_659_3Hz_E5_F0	659,277	661,236	0,949	0,776	<b>0,118%</b>
a_698_5Hz_F5_F0	698,251	700,717	1,035	0,835	<b>0,120%</b>
a_784Hz_G5_F0	784,163	786,826	1,169	0,973	<b>0,124%</b>

Tabela A11 - Resultados obtidos pelo algoritmo *Yin* através de vozes sintetizadas com um SNR=10db.

ficheiro	Resultados yin - calculo F0 Ruído Branco 10dB				
	média (Hz)	máx. (Hz)	desvio (Hz)	erro abs (Hz)	erro rel (%)
a_98Hz_G2_F0	97,916	100,403	0,829	0,634	<b>0,647%</b>
a_110Hz_A2_F0	109,941	112,202	1,060	0,870	<b>0,791%</b>
a_123_5Hz_B2_F0	123,473	125,379	0,883	0,692	<b>0,560%</b>
a_130_8Hz_C3_F0	130,709	132,820	0,856	0,665	<b>0,509%</b>
a_146_8Hz_D3_F0	146,855	148,527	0,857	0,696	<b>0,474%</b>
a_164_8Hz_E3_F0	164,819	166,753	0,902	0,734	<b>0,445%</b>
a_174_6Hz_F3_F0	174,447	176,380	0,730	0,601	<b>0,344%</b>
a_196Hz_G3_F0	196,086	198,277	0,846	0,670	<b>0,342%</b>
a_220Hz_A3_F0	220,035	222,903	0,991	0,776	<b>0,353%</b>
a_246_9Hz_B3_F0	246,946	249,261	0,905	0,714	<b>0,289%</b>
a_261_6Hz_C4_F0	261,617	263,604	0,890	0,741	<b>0,283%</b>
a_293_7Hz_D4_F0	293,659	295,709	0,829	0,638	<b>0,217%</b>
a_329_6Hz_E4_F0	329,706	331,102	0,747	0,586	<b>0,178%</b>
a_349_2Hz_F4_F0	349,237	350,947	0,817	0,636	<b>0,182%</b>
a_392Hz_G4_F0	392,017	393,747	0,752	0,618	<b>0,158%</b>
a_440Hz_A4_F0	439,974	442,674	0,890	0,670	<b>0,152%</b>
a_493_9Hz_B4_F0	493,823	495,192	0,725	0,608	<b>0,123%</b>
a_523_3Hz_C5_F0	523,197	524,622	0,755	0,600	<b>0,115%</b>
a_587_3Hz_D5_F0	587,288	589,644	1,188	0,959	<b>0,163%</b>
a_659_3Hz_E5_F0	659,276	661,235	0,928	0,771	<b>0,117%</b>
a_698_5Hz_F5_F0	698,229	700,786	1,132	0,919	<b>0,132%</b>
a_784Hz_G5_F0	784,156	786,698	1,235	1,015	<b>0,129%</b>

Tabela A12 - Resultados obtidos pelo algoritmo Yin através de vozes sintetizadas com um SNR=5db.

ficheiro	Resultados yin - calculo F0 Ruído Branco 5dB				
	média (Hz)	máx. (Hz)	desvio (Hz)	erro abs (Hz)	erro rel (%)
a_98Hz_G2_F0	97,916	100,355	0,827	0,638	<b>0,651%</b>
a_110Hz_A2_F0	109,941	112,110	1,058	0,868	<b>0,789%</b>
a_123_5Hz_B2_F0	123,473	125,314	0,888	0,692	<b>0,561%</b>
a_130_8Hz_C3_F0	130,707	132,844	0,859	0,672	<b>0,514%</b>
a_146_8Hz_D3_F0	146,844	148,588	0,863	0,703	<b>0,479%</b>
a_164_8Hz_E3_F0	164,819	166,684	0,913	0,752	<b>0,456%</b>
a_174_6Hz_F3_F0	174,446	176,488	0,746	0,614	<b>0,351%</b>
a_196Hz_G3_F0	196,102	198,228	0,851	0,687	<b>0,351%</b>
a_220Hz_A3_F0	220,034	222,994	0,982	0,751	<b>0,341%</b>
a_246_9Hz_B3_F0	246,931	249,424	0,901	0,722	<b>0,293%</b>
a_261_6Hz_C4_F0	261,646	263,632	0,924	0,768	<b>0,293%</b>
a_293_7Hz_D4_F0	293,642	295,581	0,852	0,657	<b>0,224%</b>
a_329_6Hz_E4_F0	329,717	331,542	0,732	0,572	<b>0,173%</b>
a_349_2Hz_F4_F0	349,268	351,092	0,879	0,689	<b>0,197%</b>
a_392Hz_G4_F0	392,036	393,556	0,820	0,659	<b>0,168%</b>
a_440Hz_A4_F0	439,947	442,821	0,958	0,693	<b>0,157%</b>
a_493_9Hz_B4_F0	493,851	495,432	0,787	0,646	<b>0,131%</b>
a_523_3Hz_C5_F0	523,187	525,183	0,851	0,697	<b>0,133%</b>
a_587_3Hz_D5_F0	587,222	589,598	1,189	0,963	<b>0,164%</b>
a_659_3Hz_E5_F0	659,219	661,446	1,039	0,853	<b>0,129%</b>
a_698_5Hz_F5_F0	698,249	701,261	1,343	1,116	<b>0,160%</b>
a_784Hz_G5_F0	784,272	788,562	1,962	1,592	<b>0,203%</b>

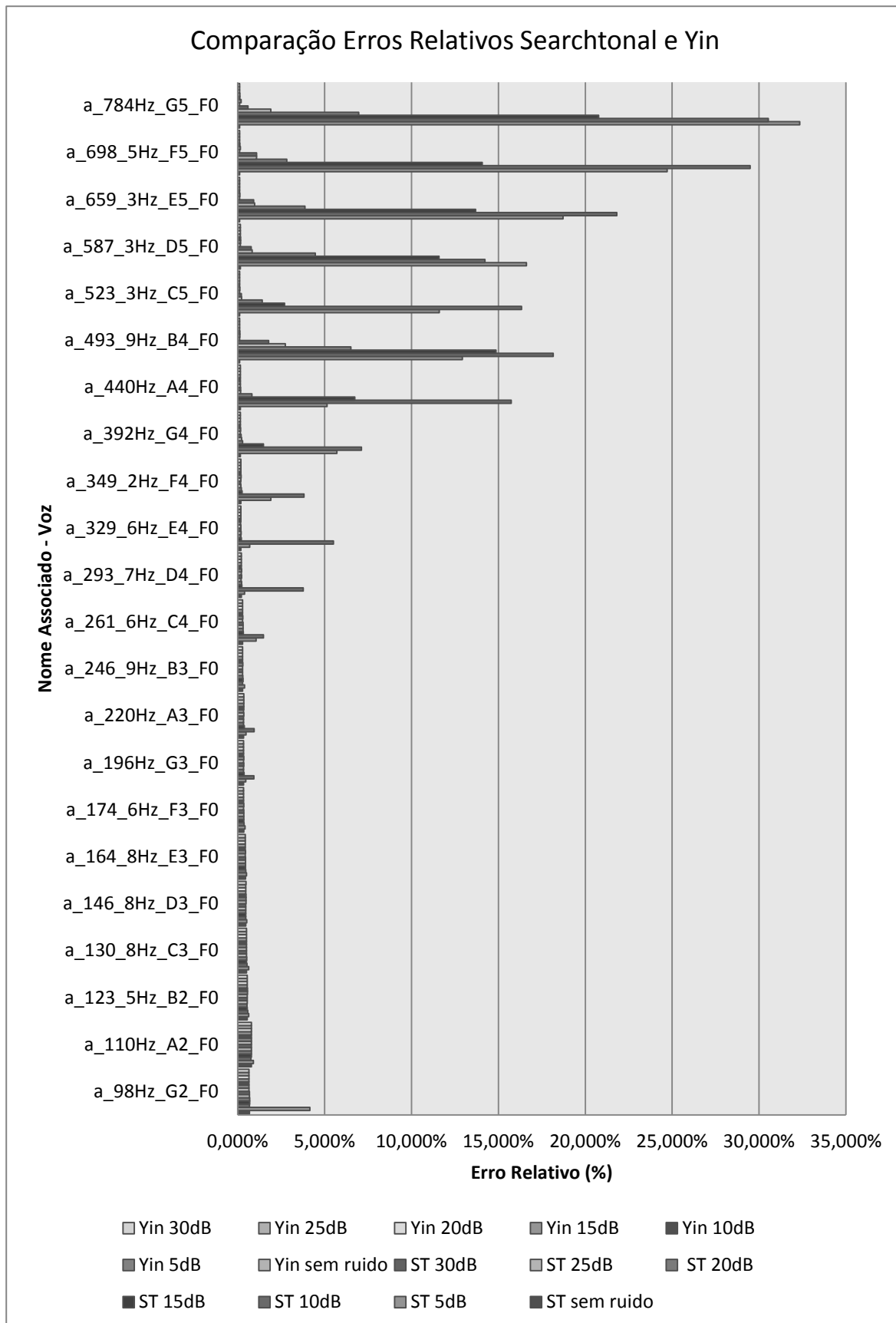
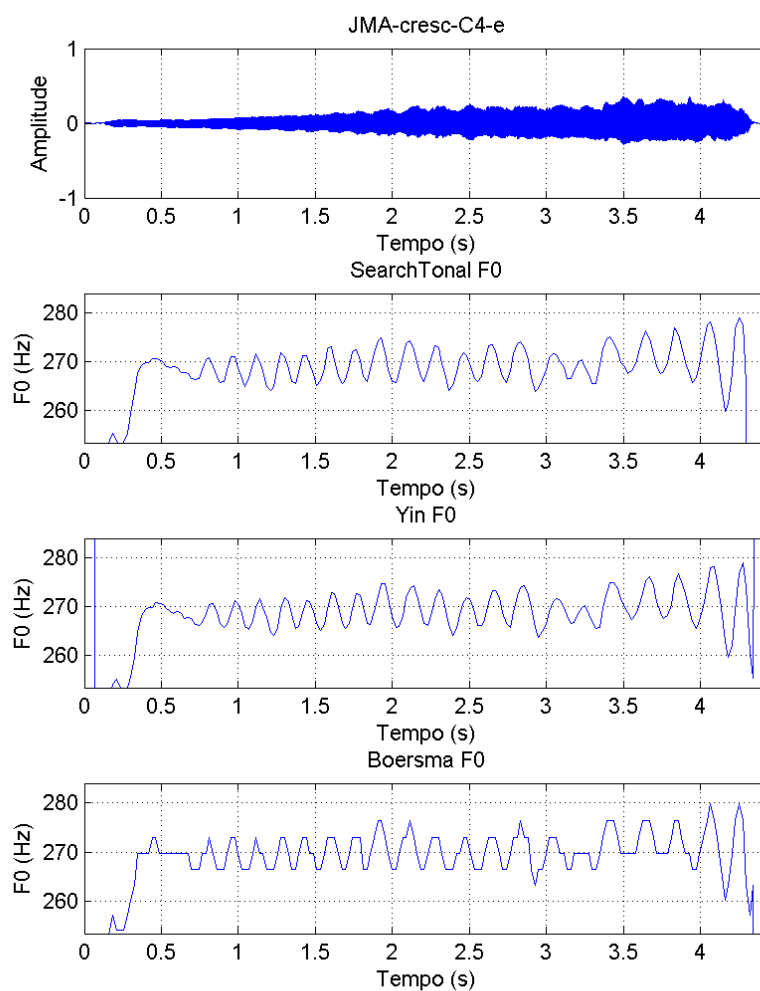


Figura A1 - comparação dos resultados obtidos pelos métodos *SearchTonal* e *Yin*, através de vozes sintetizadas com ruído.



## Anexo B

Neste anexo são apresentadas as figuras referentes aos restantes resultados obtidos através dos testes efectuados aos algoritmos de estimação de F0 através do uso de vozes naturais de curta duração.



**Figura B1** - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz JMA-cresc-C4-e.

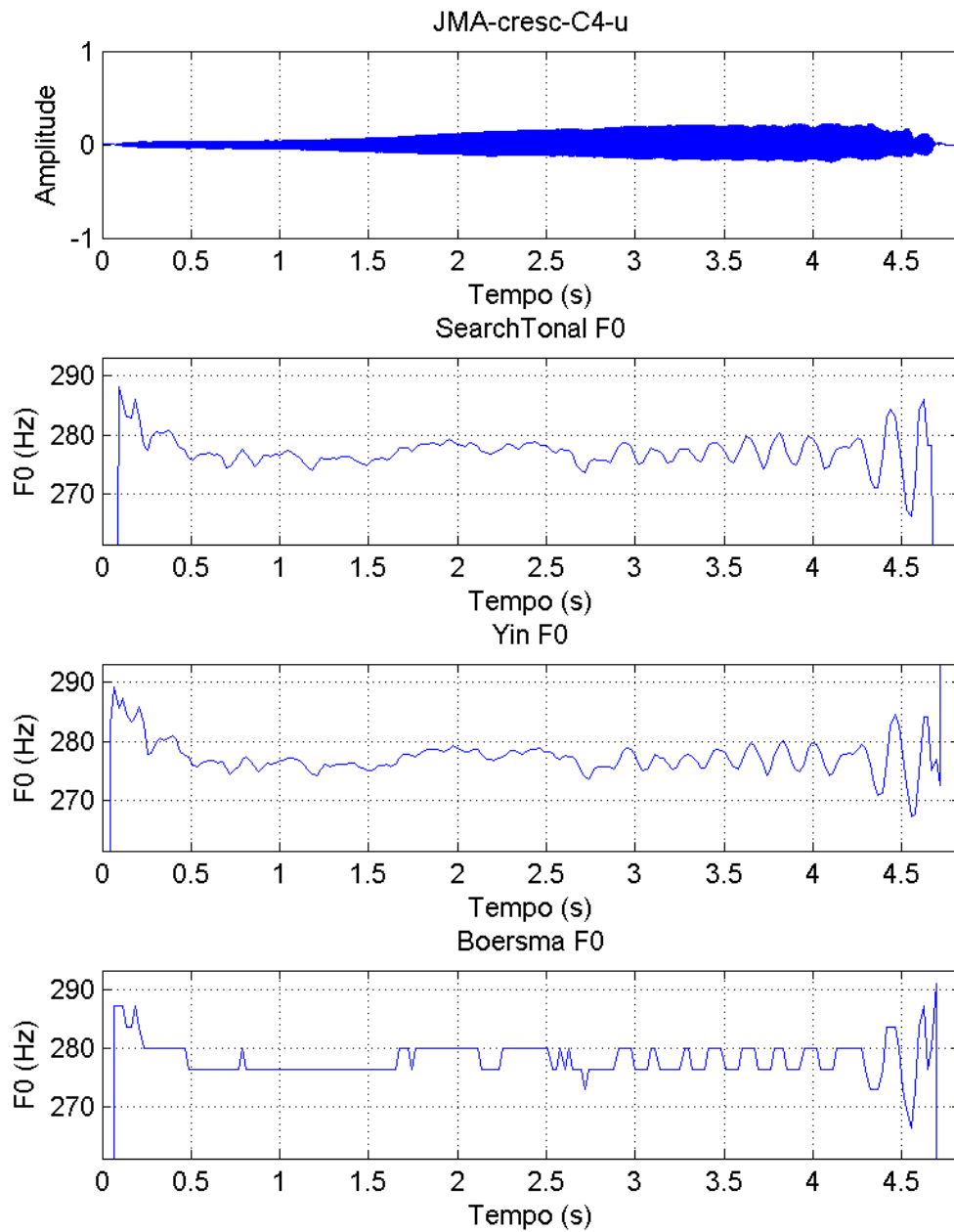
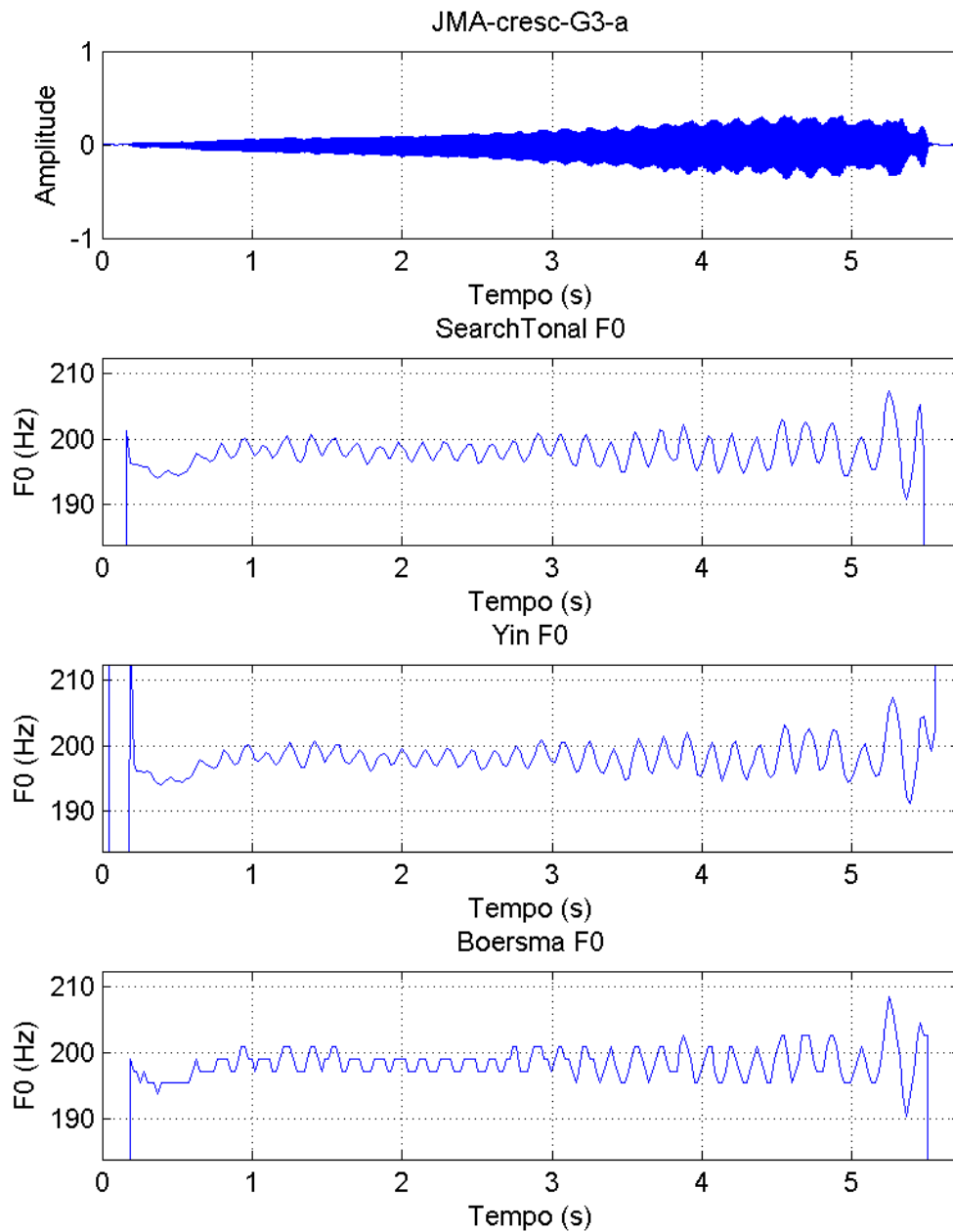


Figura B2 - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz JMA-cresc-C4-u.



**Figura B3** - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz JMA-cresc-G3-a.

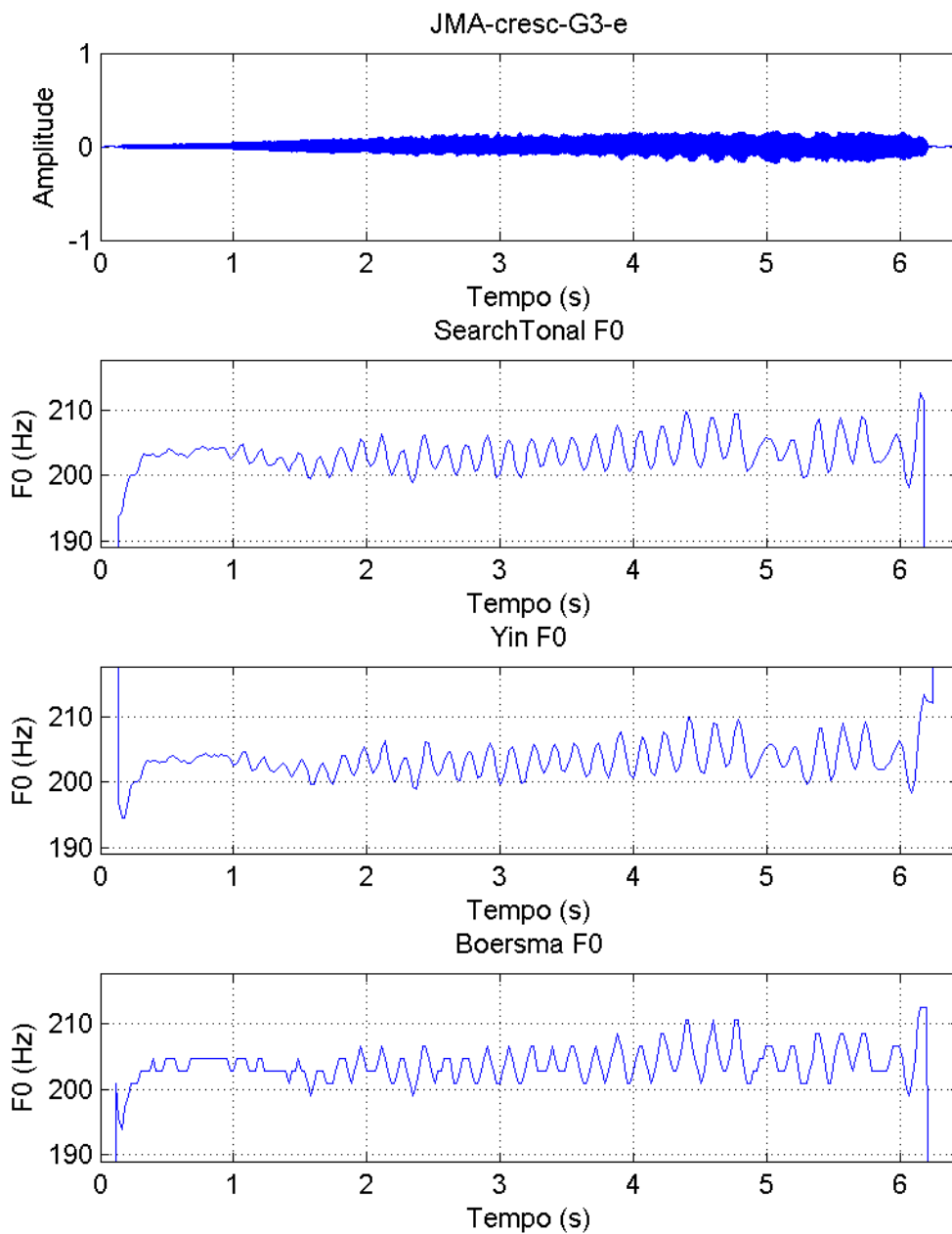


Figura B4 - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz JMA-cresc-G3-e.

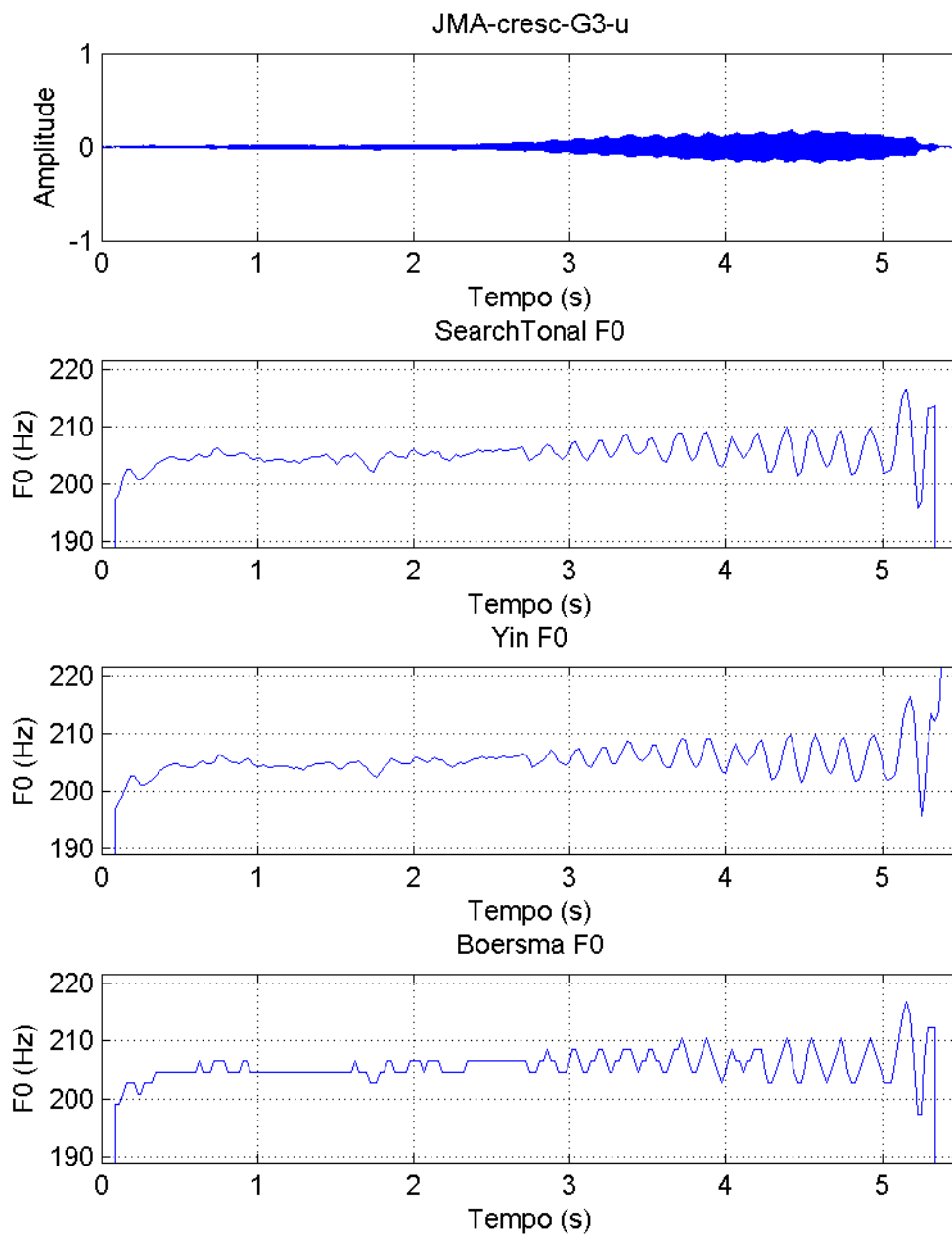


Figura B5 - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz JMA-cresc-G3-u.

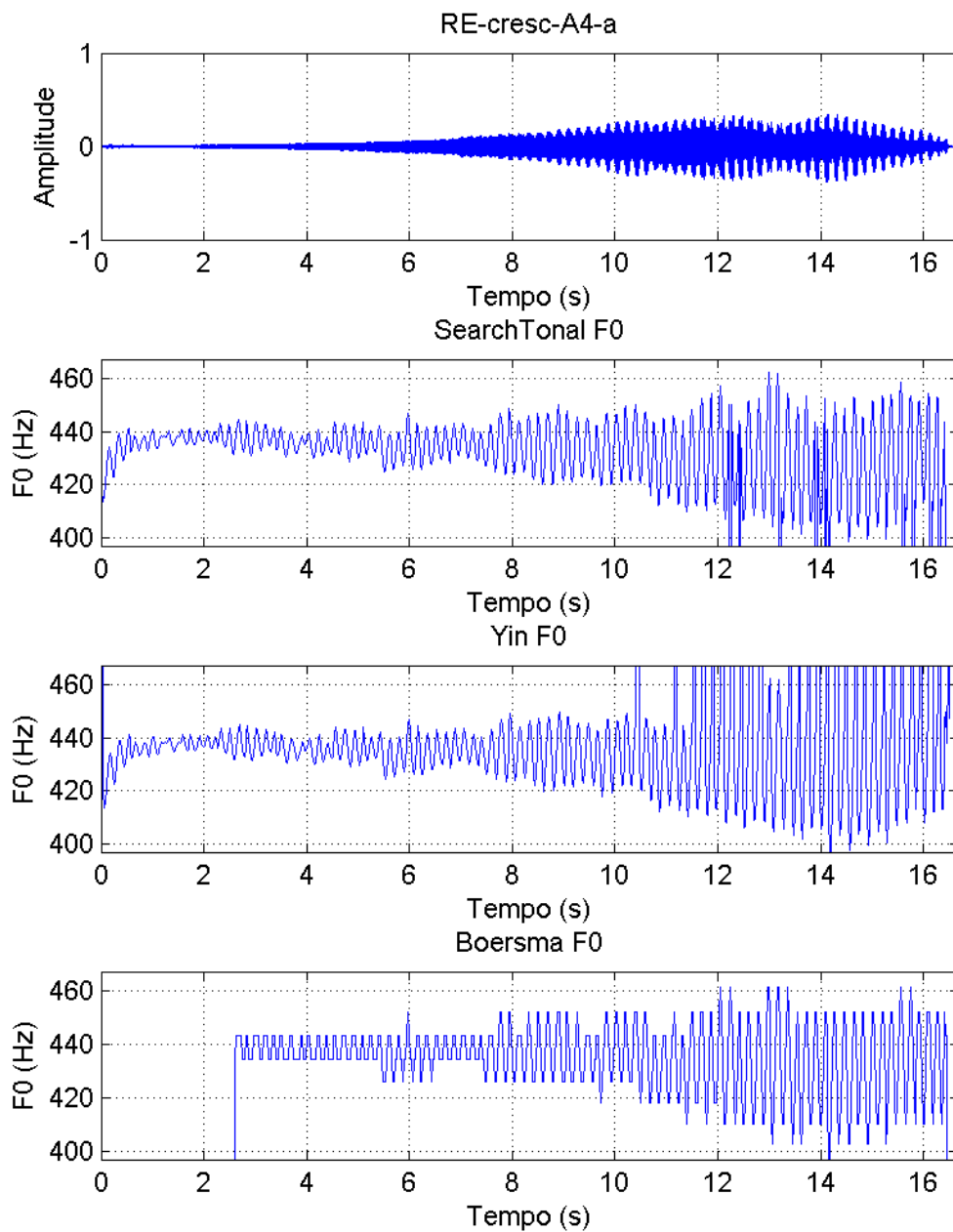


Figura B6 - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz RE-cresc-A4-a.

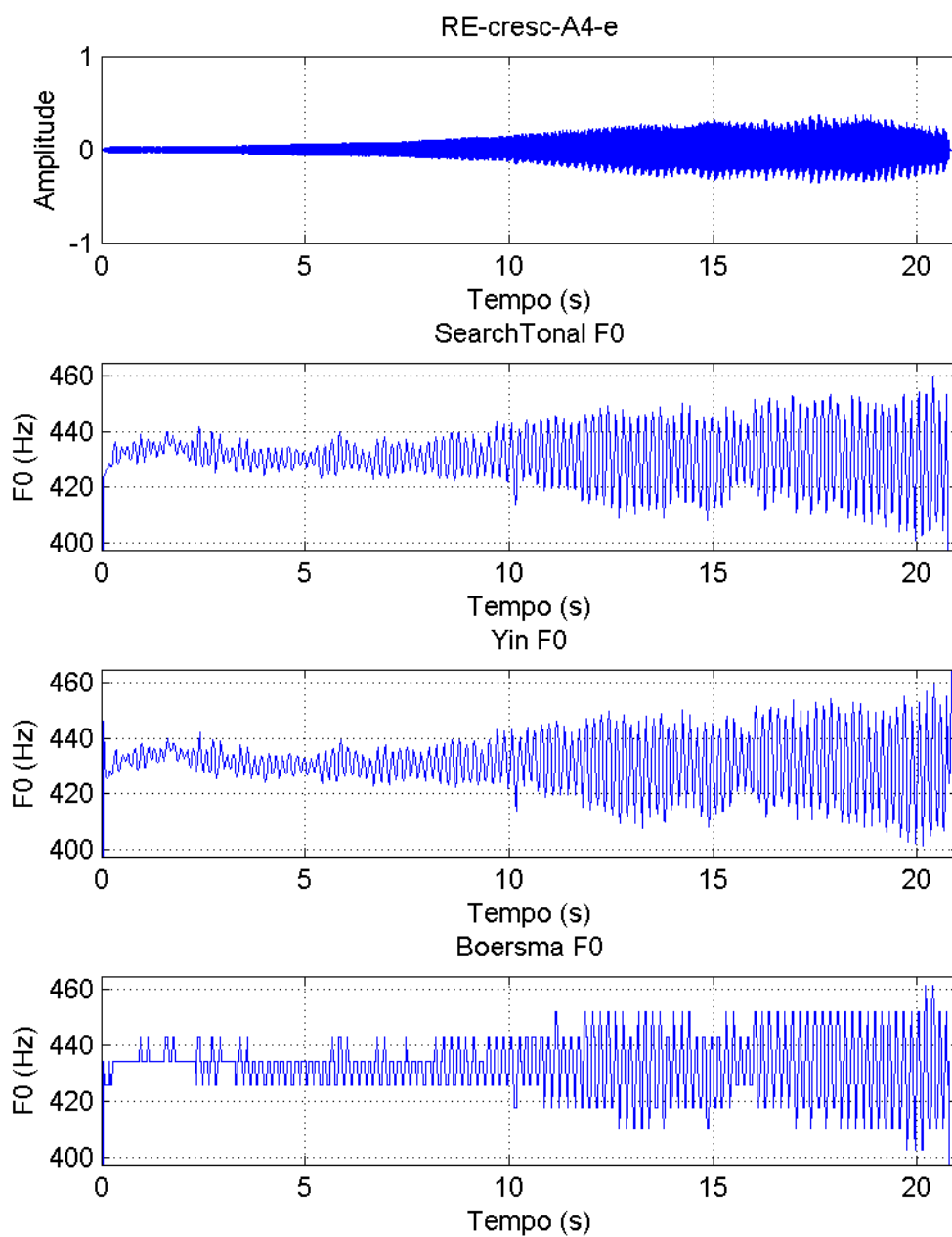


Figura B7 - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz RE-cresc-A4-e.

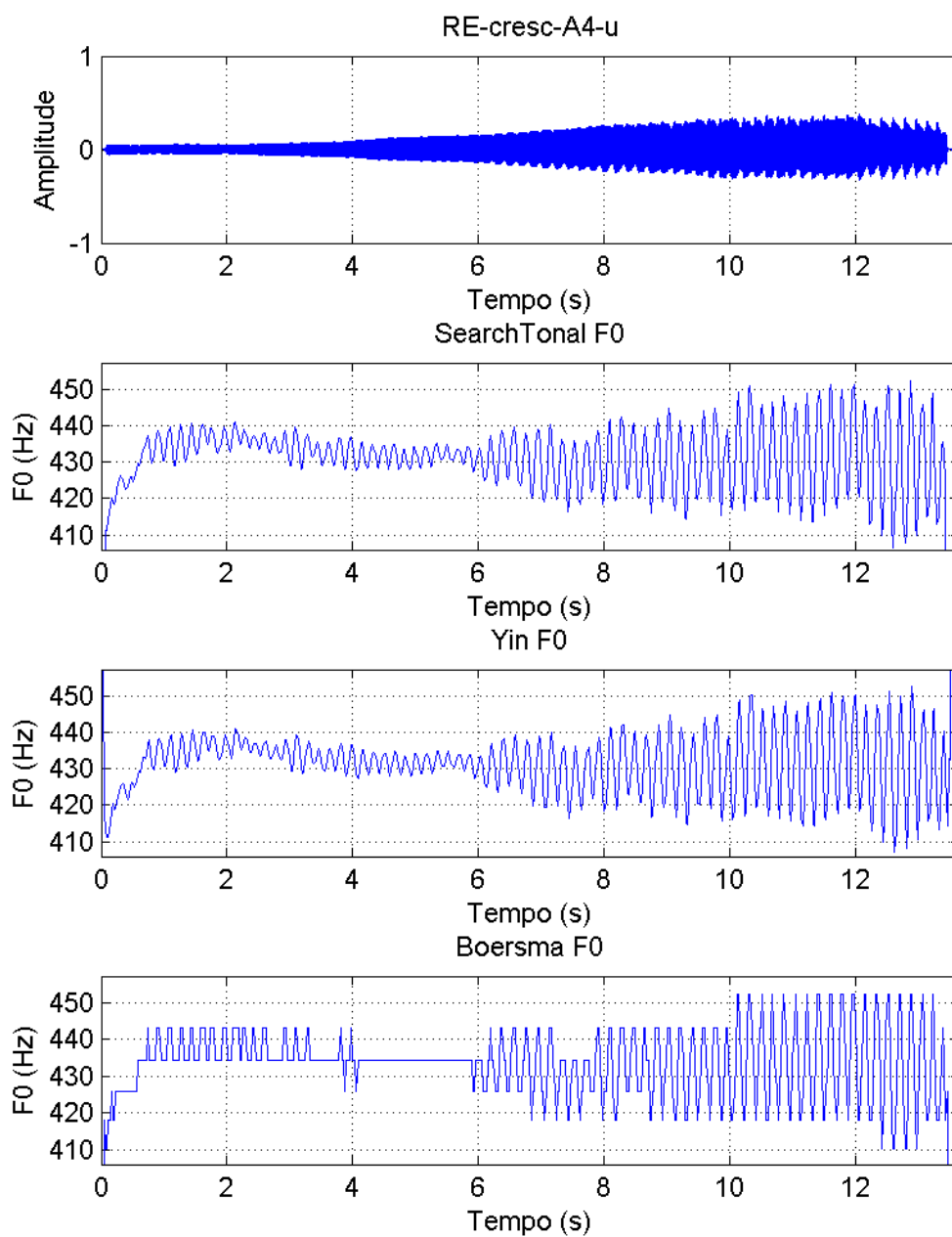


Figura B8 - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz RE-cresc-A4-u.

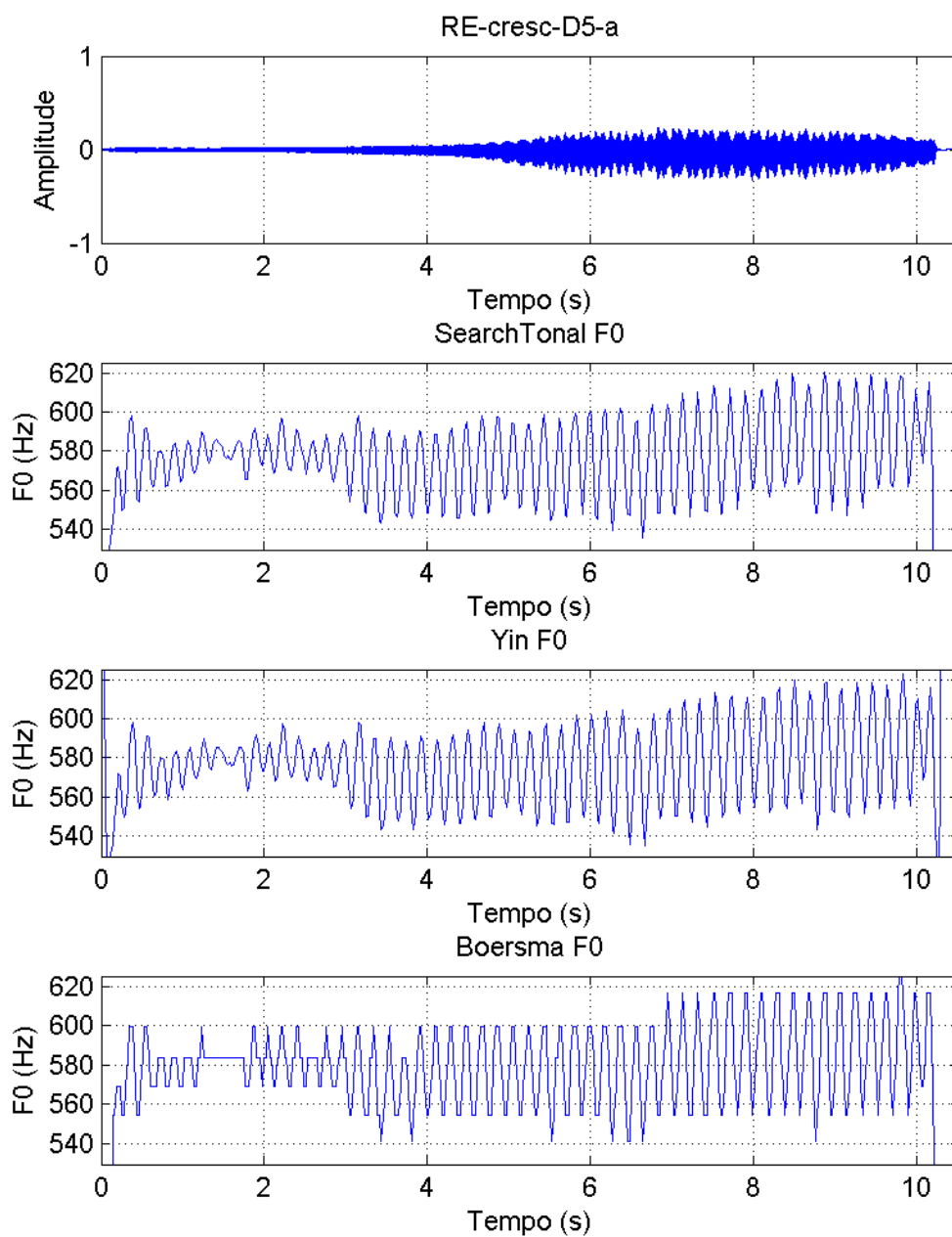


Figura B9 - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz RE-cresc-D5-a.

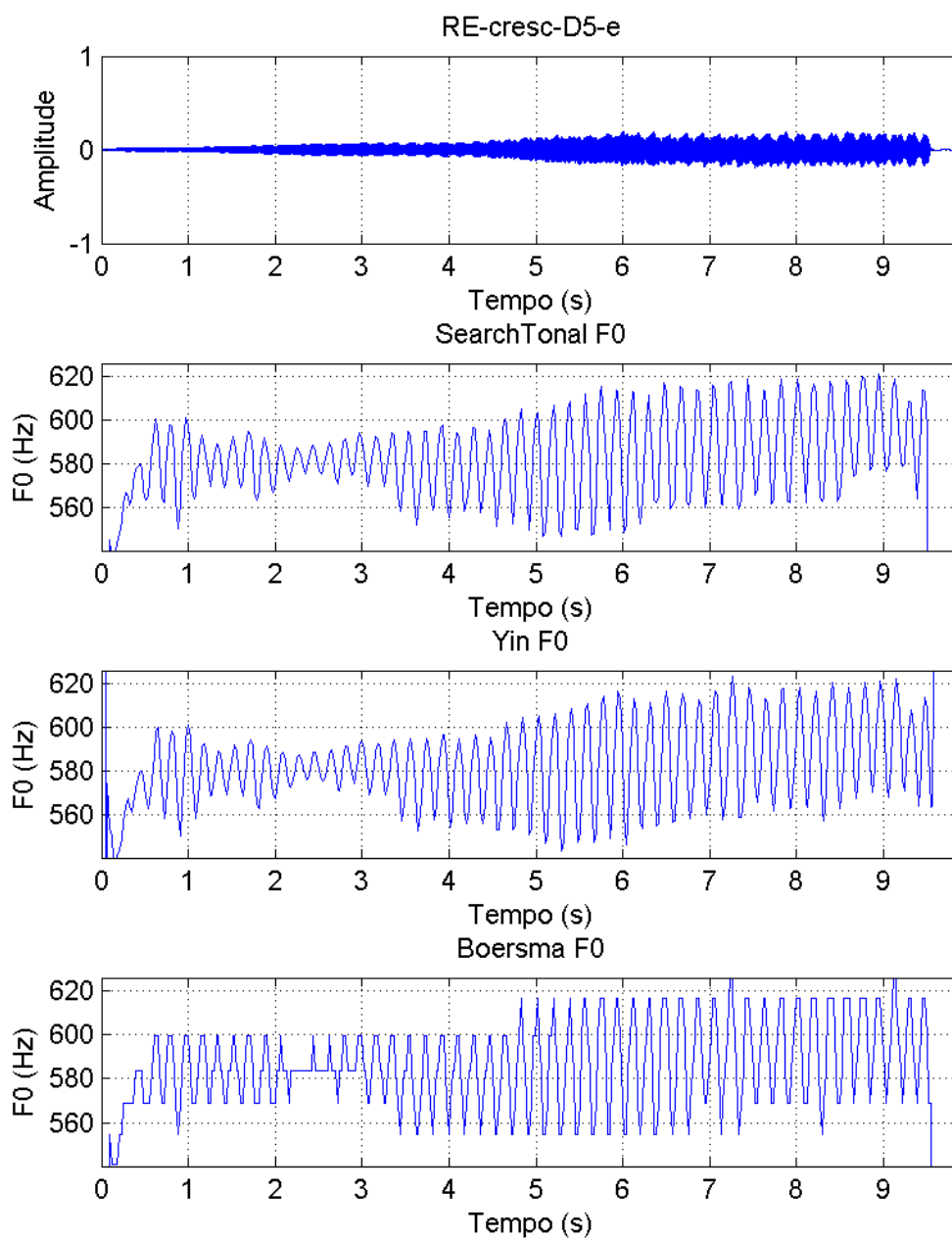
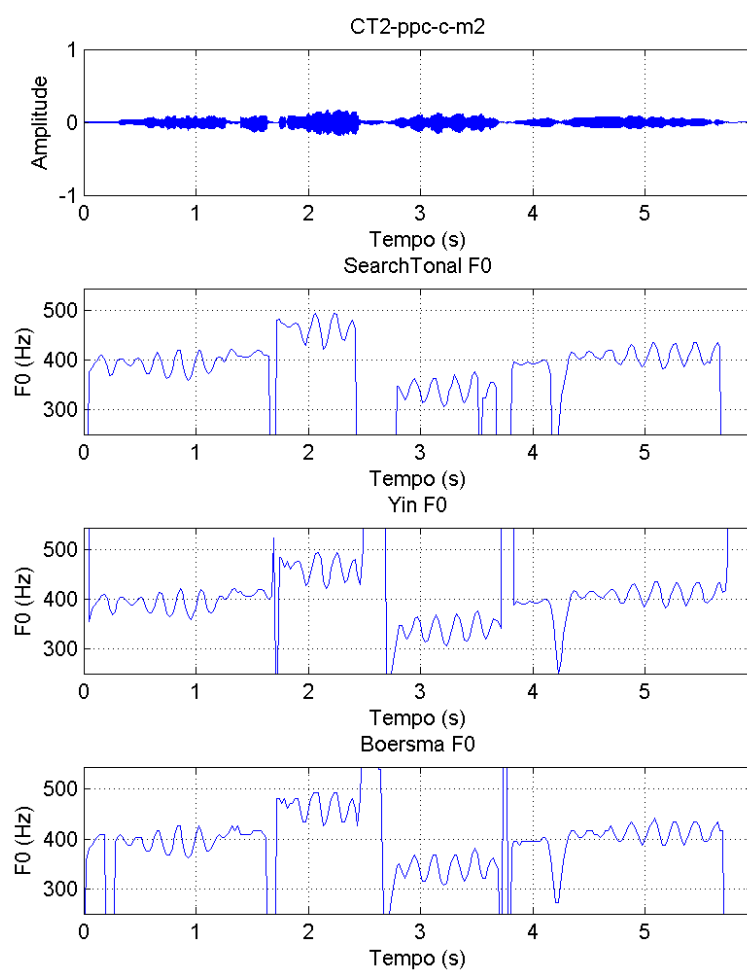


Figura B10 - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz RE-cresc-D5-e.

## Anexo C

Neste anexo são apresentadas as figuras referentes aos restantes resultados obtidos através dos testes efectuados aos algoritmos de estimação de F0 através do uso de vozes naturais de curta duração



**Figura C1** - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz CT2-ppc-c-m2.

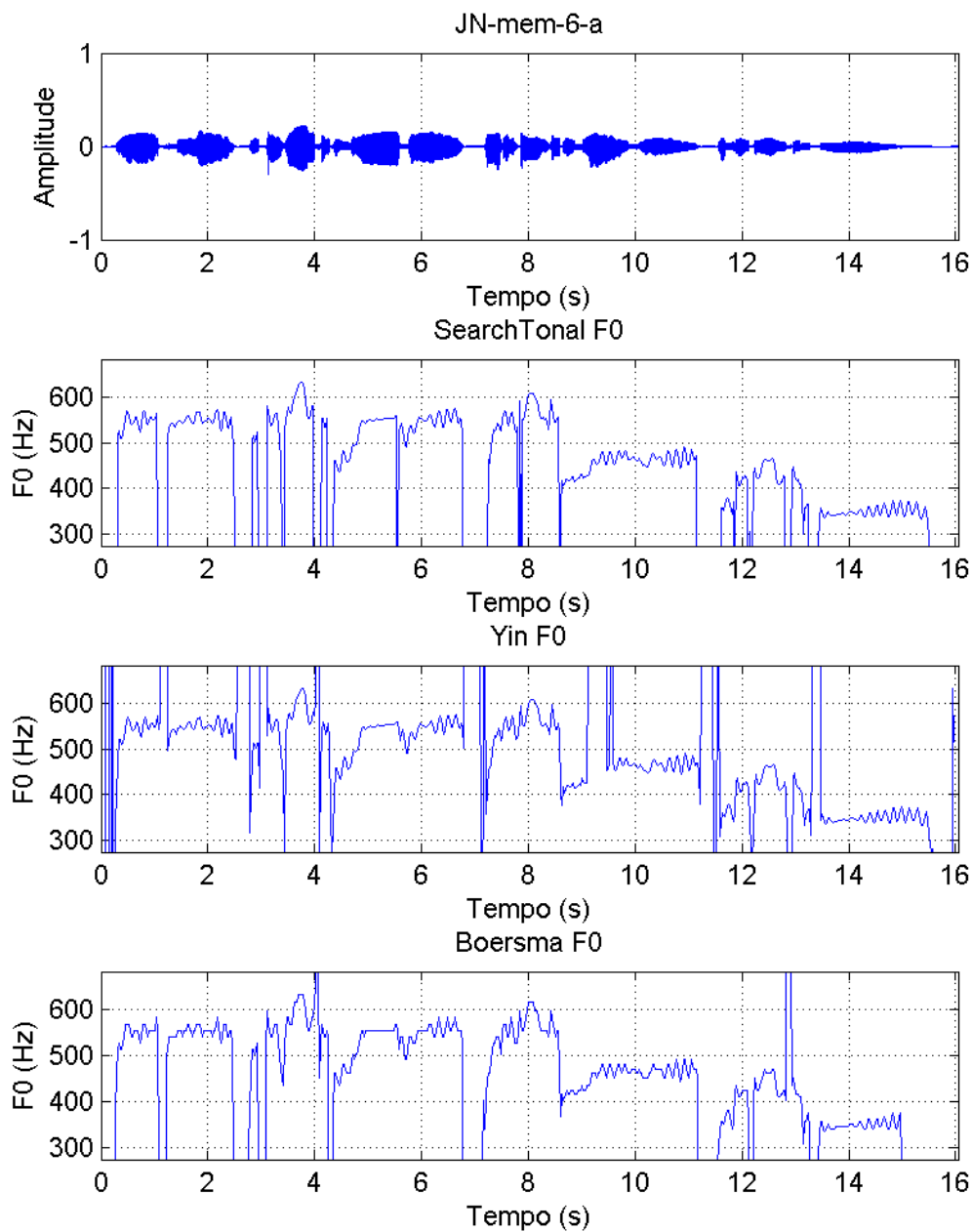


Figura C2 - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz JN-mem-6-a.

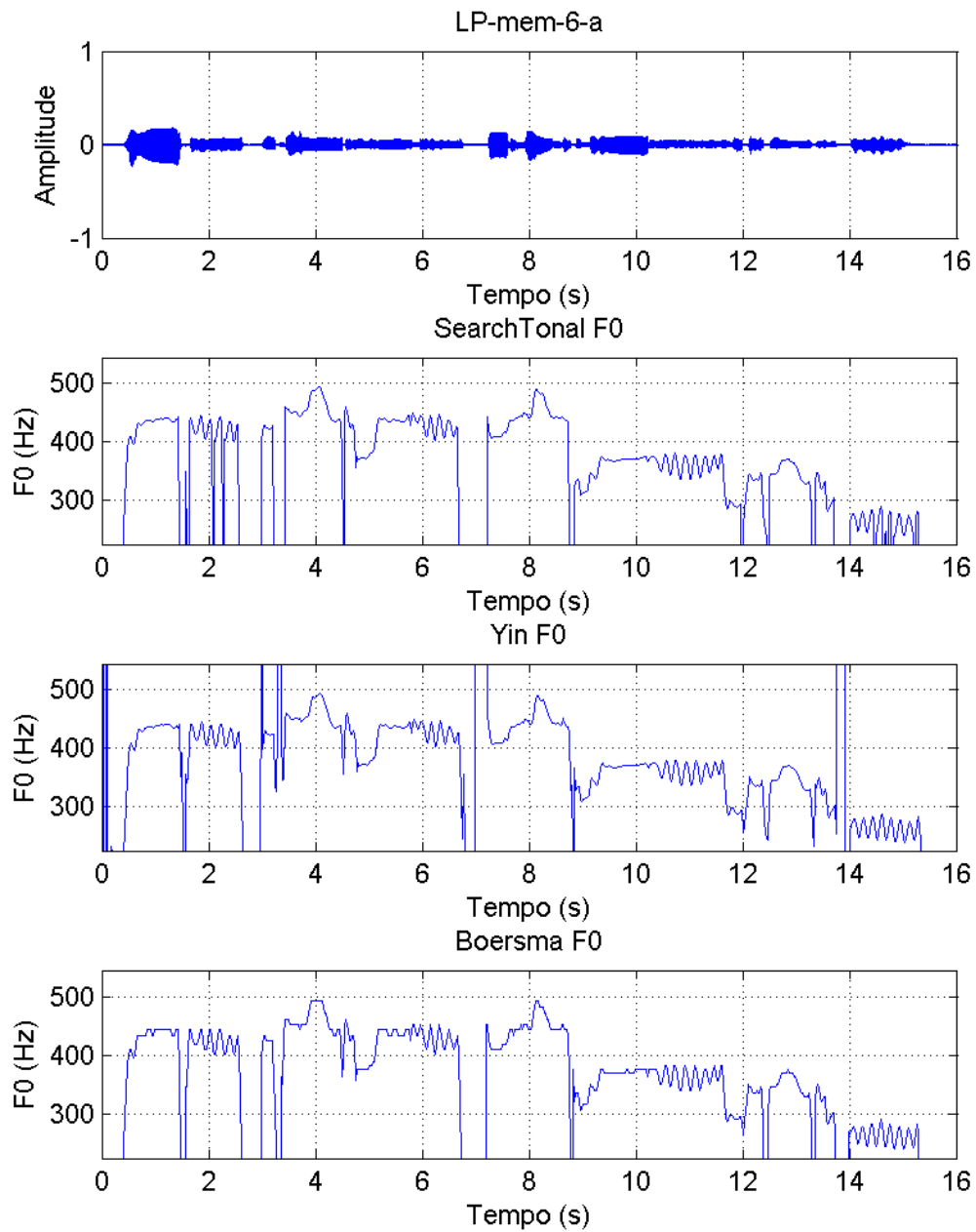


Figura C3 - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz LP-mem-6-a.

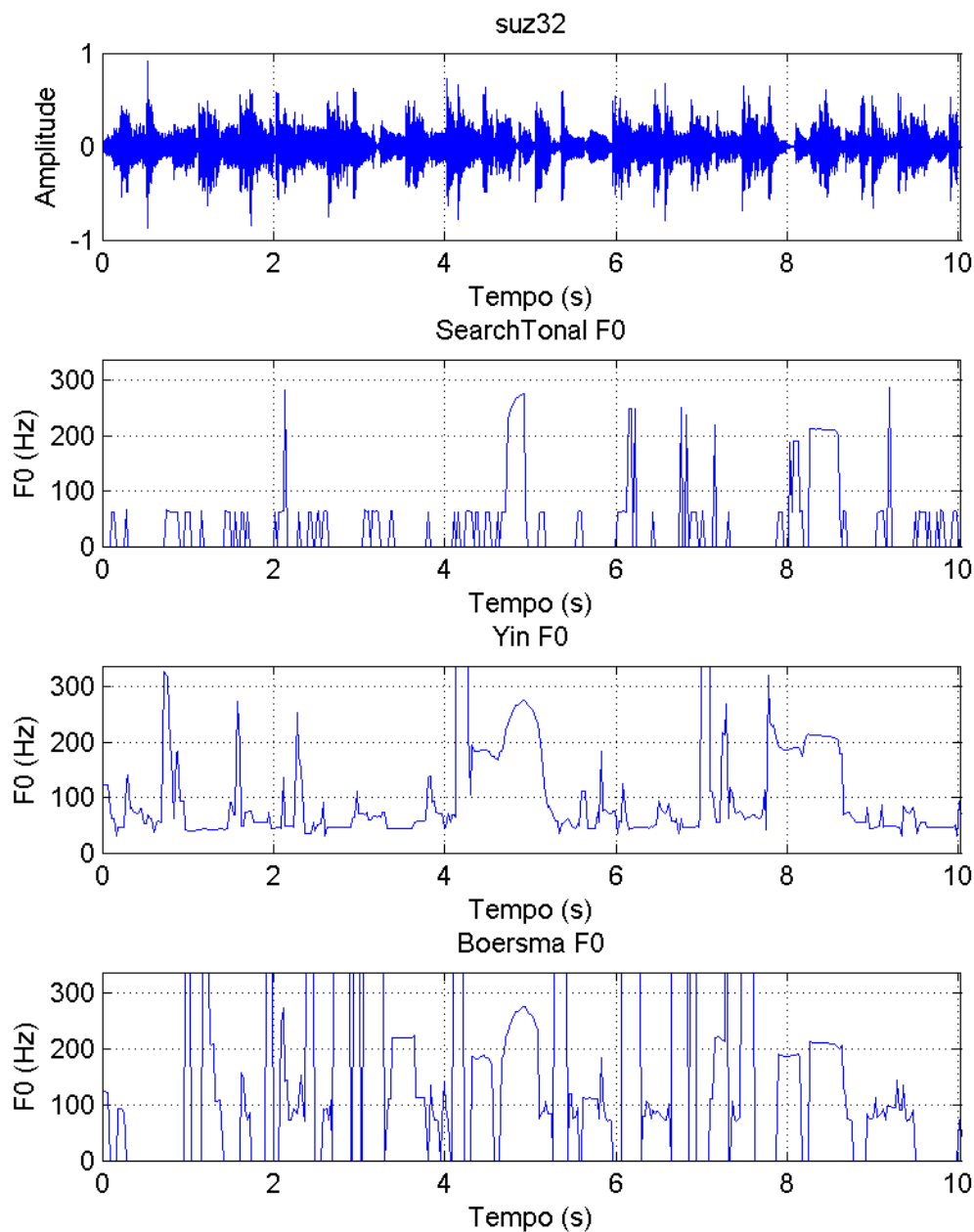


Figura C4 - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz suz32.

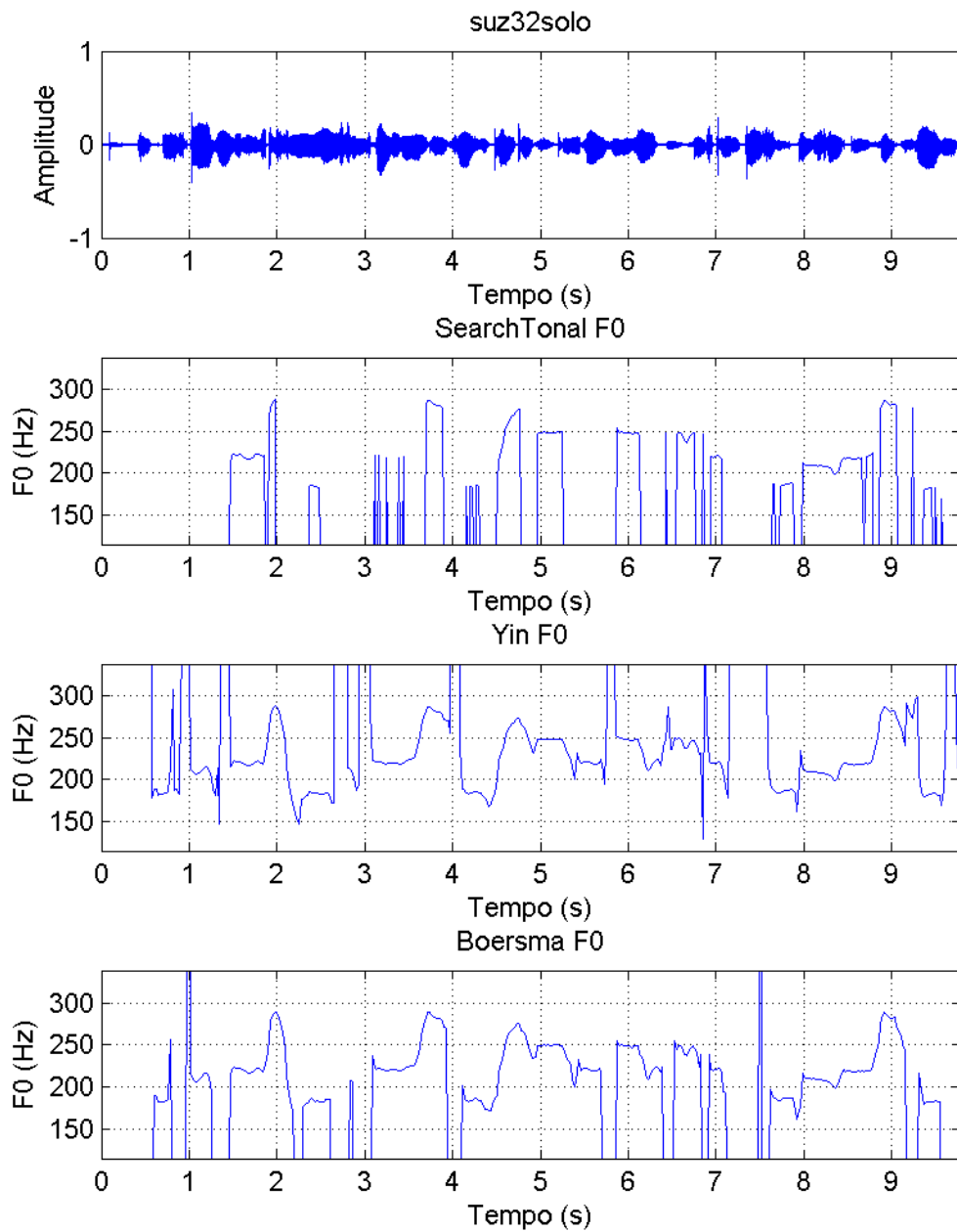


Figura C5 - Comparação dos resultados obtidos pelos métodos de estimação de F0, através da voz suz32solo.



## Anexo D

Neste anexo são apresentadas as tabelas referentes aos restantes resultados obtidos através dos testes efectuados ao algoritmo de estimação do vibrato através do uso de vozes sintetizadas com um valor de ST constante, uma zona de ataque e duração de 2 segundos.

Tabela D1 - Resultado da estimação de vibrato através de uma voz sintetizada com ST=0.5, com zona de ataque e duração 3s.

<b>Resultados Calculo Vibrato ST=0.5 com Ataque 3s</b>						
<i>Nome Associado</i>	<i>F0 (Hz)</i>	<i>F v.(Hz)</i>	<i>F v. máx (Hz)</i>	<i>desvio (Hz)</i>	<i>erro abs (Hz)</i>	<i>erro rel (%)</i>
a_A4_4Hz	408,123	4,001	4,010	0,004	0,002	0,059%
a_A4_4-5Hz	411,621	4,499	4,501	0,002	0,002	0,043%
a_A4_5Hz	411,402	5,001	5,006	0,003	0,002	0,048%
a_A4_5-5Hz	411,459	5,500	5,503	0,002	0,001	0,024%
a_A4_6Hz	407,920	5,998	6,002	0,003	0,003	0,054%
a_A4_6-5Hz	407,956	6,500	6,511	0,008	0,006	0,097%
a_A4_7Hz	414,798	7,001	7,017	0,011	0,009	0,134%
a_A4_7-5Hz	404,740	7,501	7,504	0,002	0,002	0,023%
a_A4_8Hz	408,781	8,000	8,004	0,003	0,003	0,032%

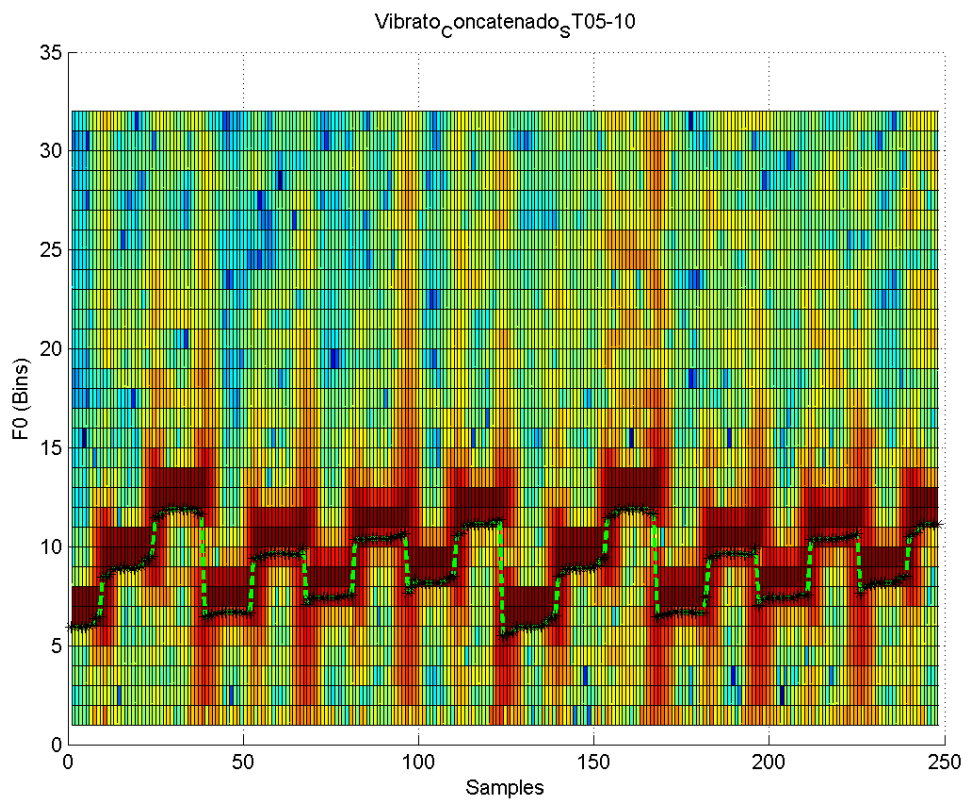
Tabela D2 - Resultado da estimação de vibrato através de uma voz sintetizada com ST=1.0, com zona de ataque e duração 3s.

<b>Resultados Calculo Vibrato ST=1.0 comAtaque 3s</b>						
<i>Nome Associado</i>	<i>F0 (Hz)</i>	<i>F v.(Hz)</i>	<i>F v. máx (Hz)</i>	<i>desvio (Hz)</i>	<i>erro abs (Hz)</i>	<i>erro rel (%)</i>
a_A4_4Hz	414,046	4,000	4,005	0,004	0,003	0,080%
a_A4_4-5Hz	412,731	4,501	4,508	0,005	0,004	0,088%
a_A4_5Hz	418,364	5,004	5,017	0,010	0,009	0,177%
a_A4_5-5Hz	412,830	5,499	5,504	0,003	0,003	0,049%
a_A4_6Hz	410,972	5,999	6,001	0,001	0,001	0,016%
a_A4_6-5Hz	405,539	6,504	6,512	0,006	0,005	0,079%
a_A4_7Hz	396,914	7,003	7,003	0,001	0,003	0,041%
a_A4_7-5Hz	400,542	7,498	7,498	0,000	0,002	0,021%
a_A4_8Hz	397,623	7,988	7,988	0,000	0,012	0,149%



## Anexo E

Neste anexo são apresentadas as figuras referentes aos resultados obtidos dos testes efectuados ao algoritmo de estimação do vibrato através do uso de vozes sintetizadas com um valor de  $ST=05$  e  $ST=1.0$ , e uma frequência de vibrato variável. Os sinais foram concatenados, contendo cada um deles 2 segundos.



**Figura E1** - Visualização do espectro de potências, através da ODFT do sinal Vibrato\_concatenado\_ST05-10.

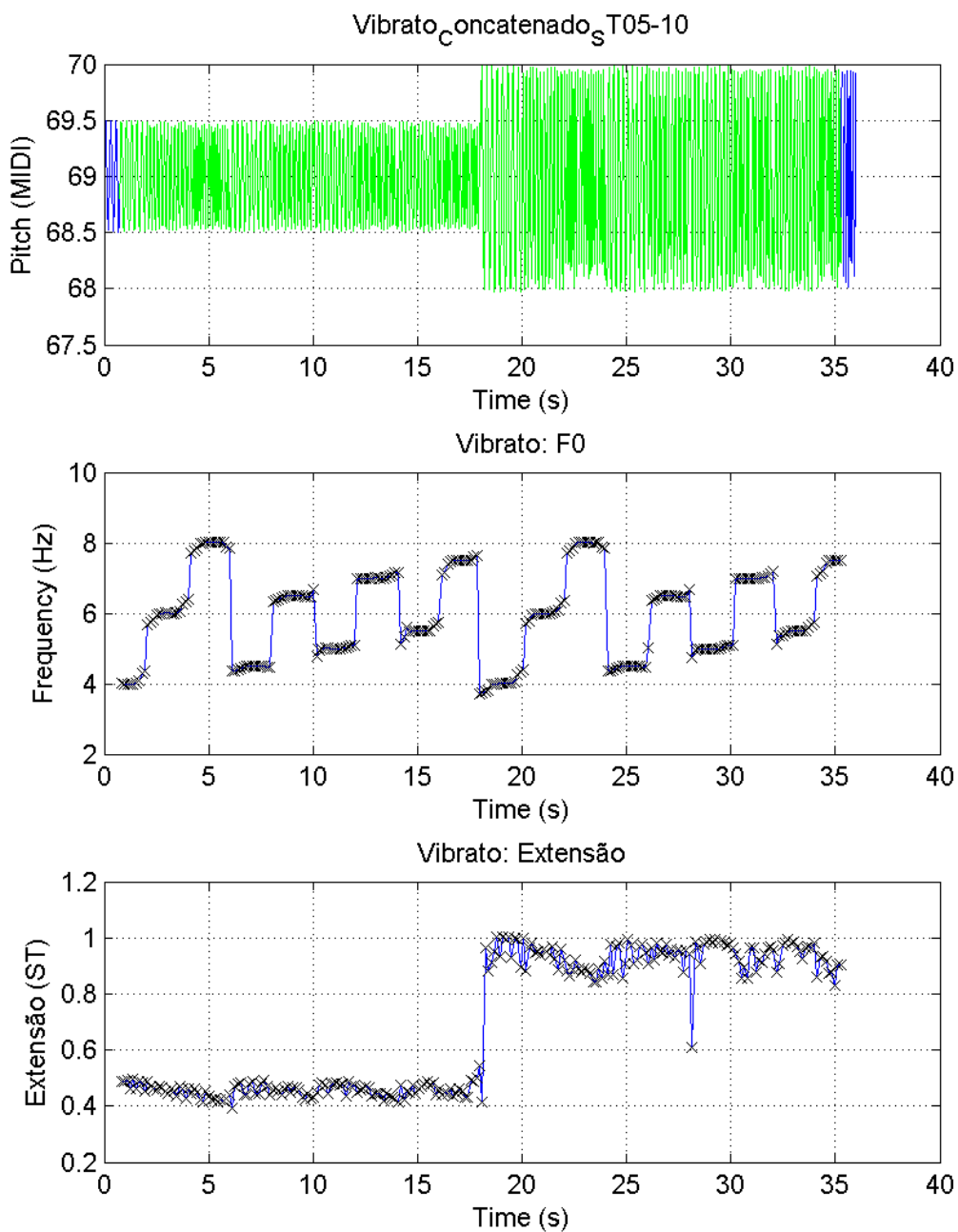


Figura E2 - Visualização da estimação dos parâmetros do vibrato, através do sinal Vibrato\_concatenado\_ST05-10.

## Anexo F

Neste anexo são apresentadas as figuras referentes aos resultados dos testes efectuados ao algoritmo de estimação do vibrato, através do uso de vozes naturais de curta duração e sem variação abrupta de F0.

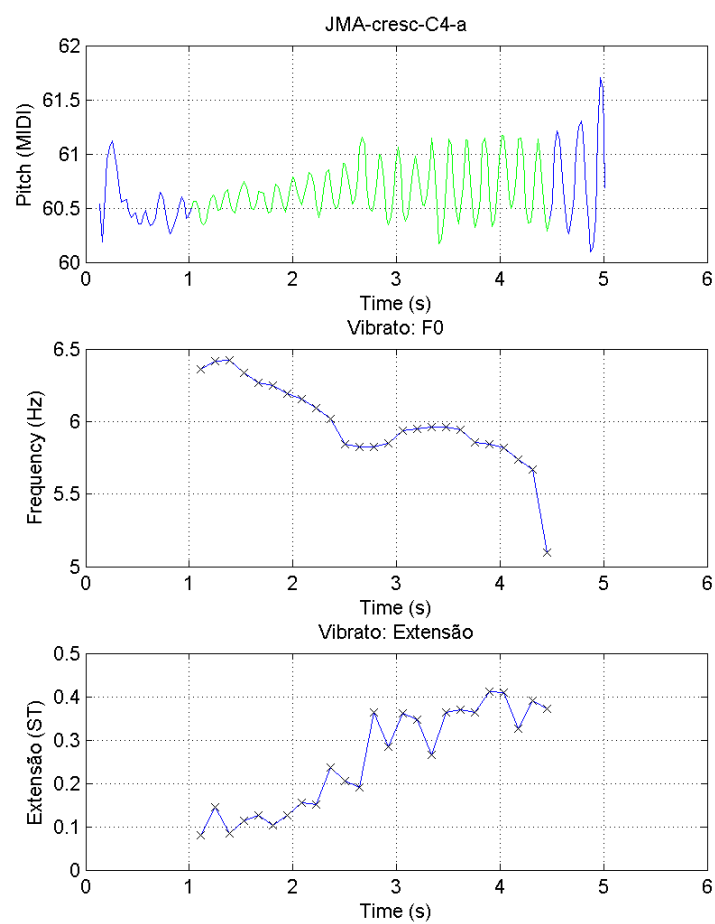


Figura F1 - Visualização da estimação dos parâmetros do vibrato, através do sinal JMA-cresc-C4-a.

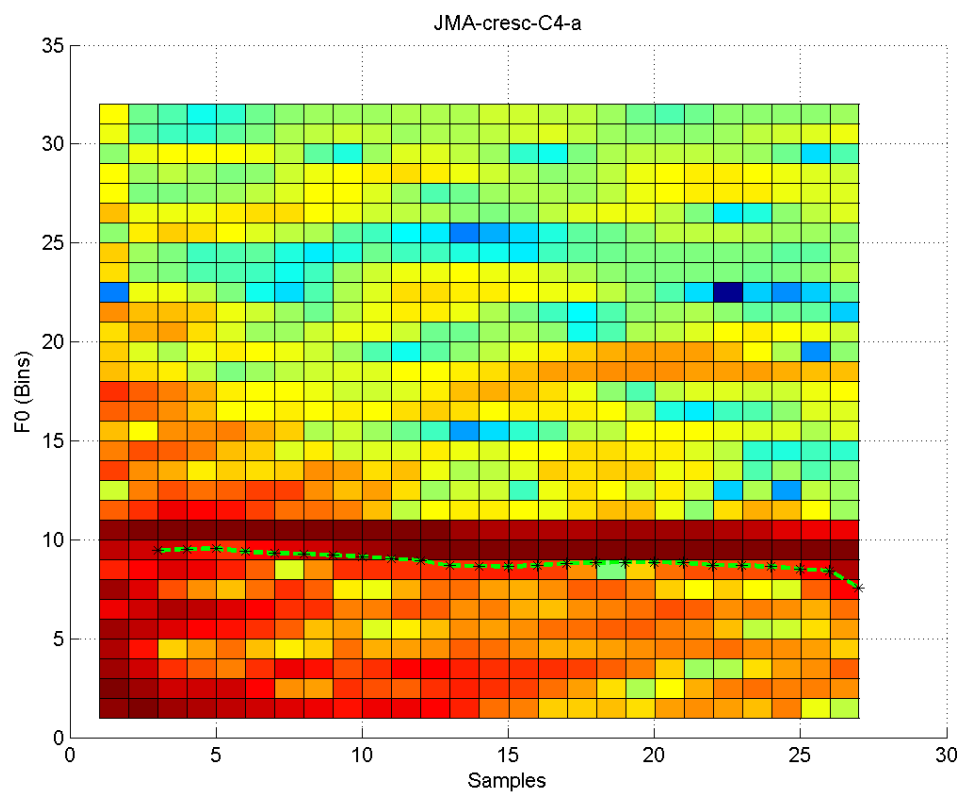


Figura F2 - Visualização do espectro de potências através da ODFT do sinal JMA-cresc-C4-a.

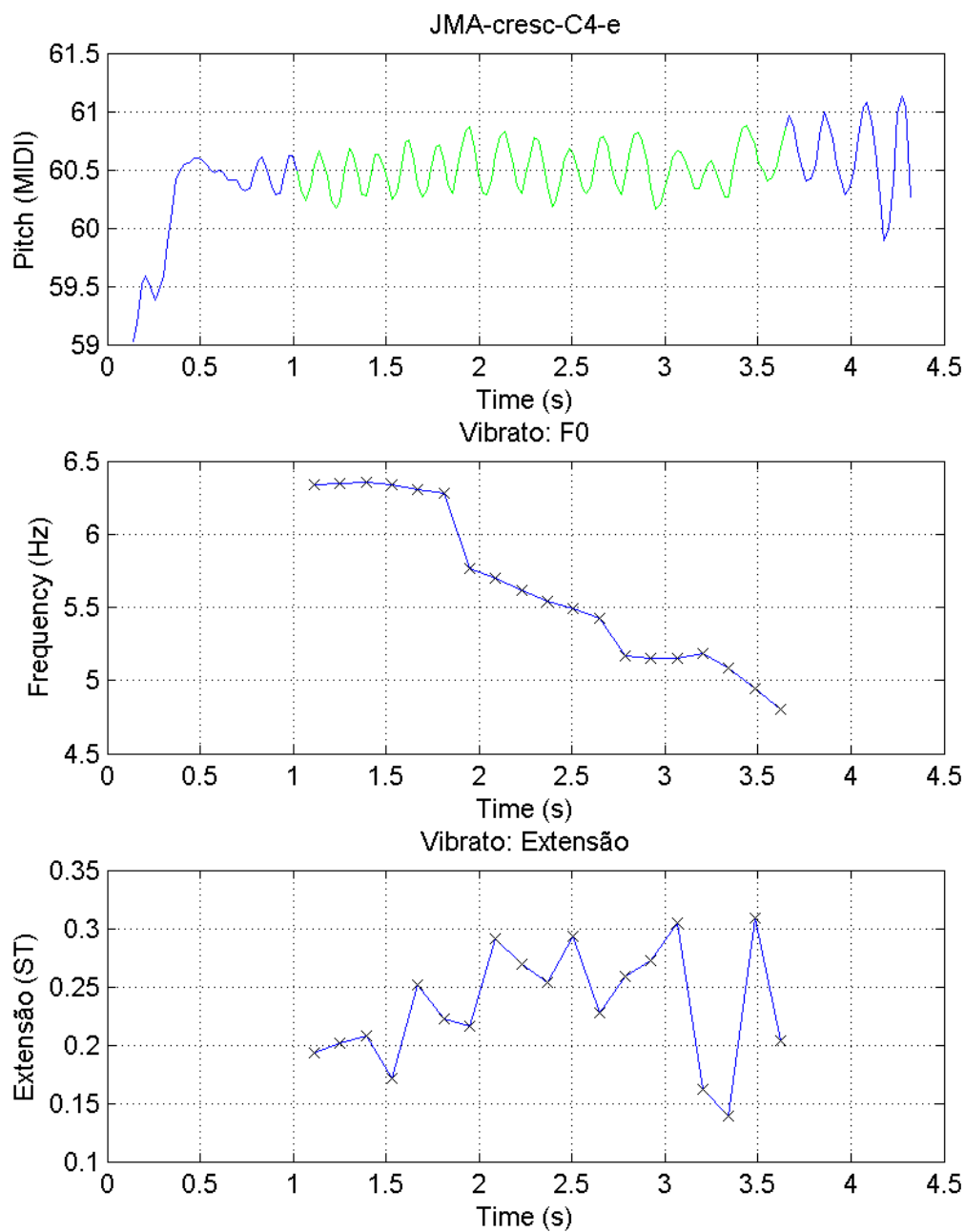


Figura F3 - Visualização da estimação dos parâmetros do vibrato, através do sinal JMA-cresc-C4-e.

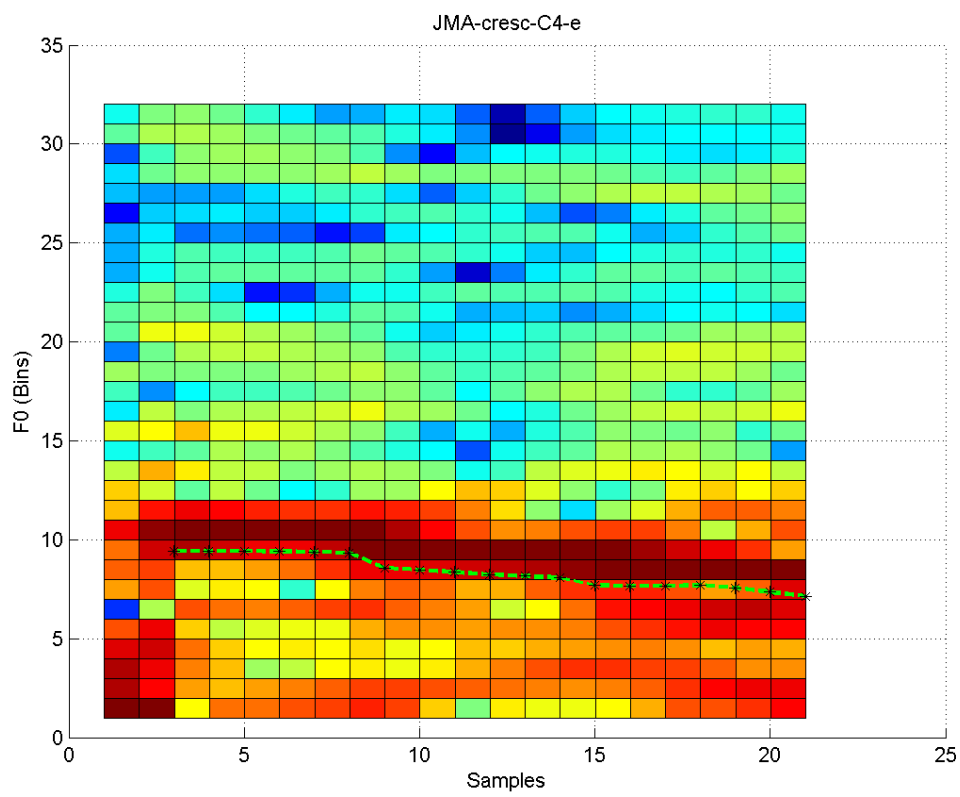


Figura F4 - Visualização do espectro de potências através da ODFT do sinal JMA-cresc-C4-e.

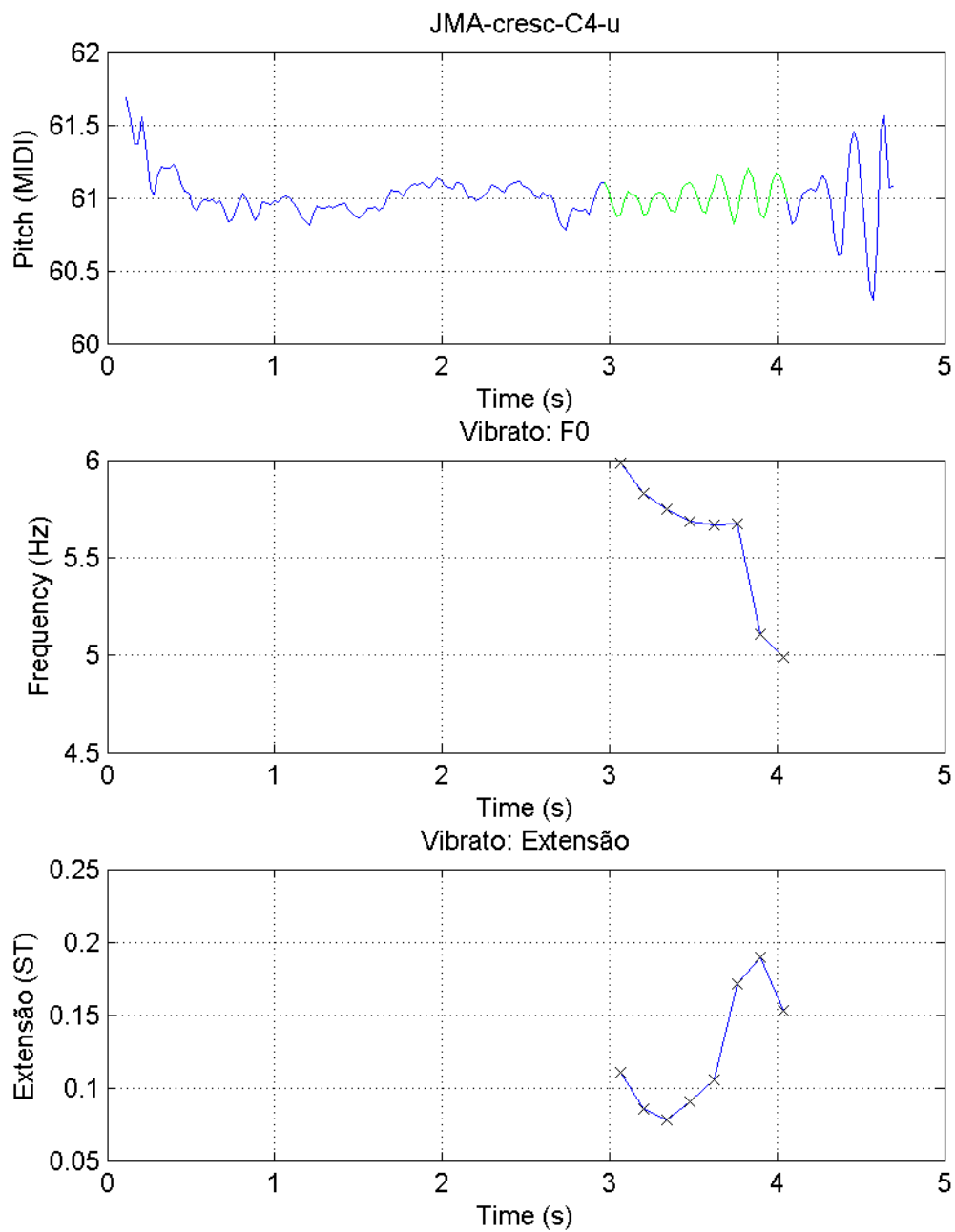


Figura F5 - Visualização da estimação dos parâmetros do vibrato, através do sinal JMA-cresc-C4-u.

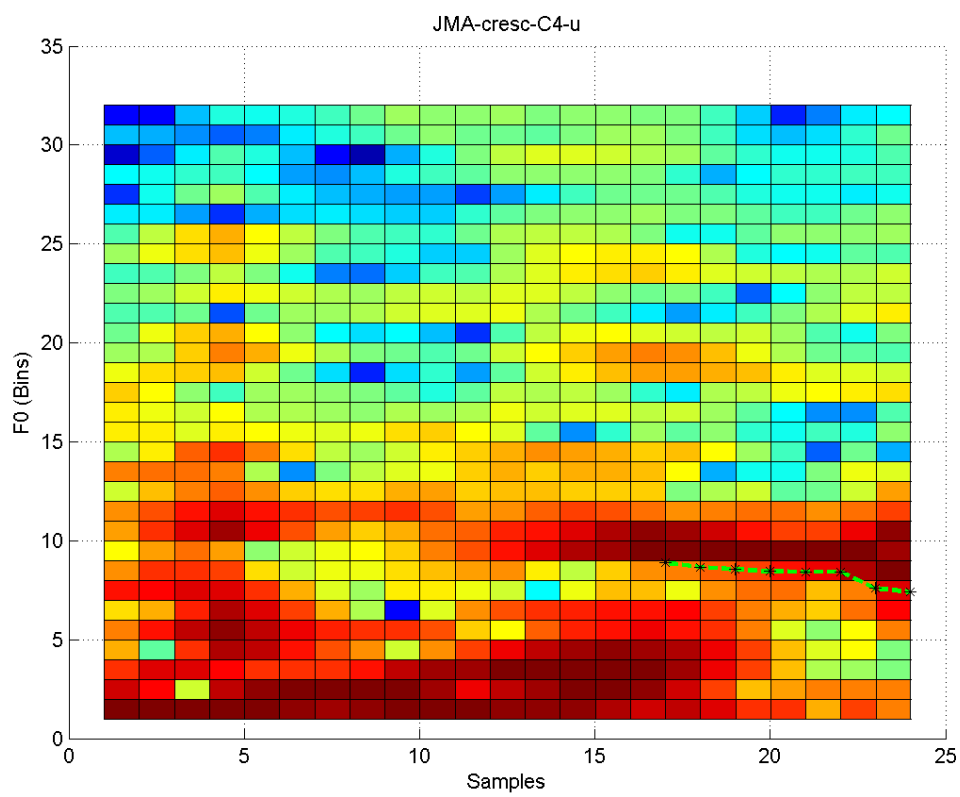


Figura F6 - Visualização do espectro de potências através da ODFT do sinal JMA-cresc-C4-u.

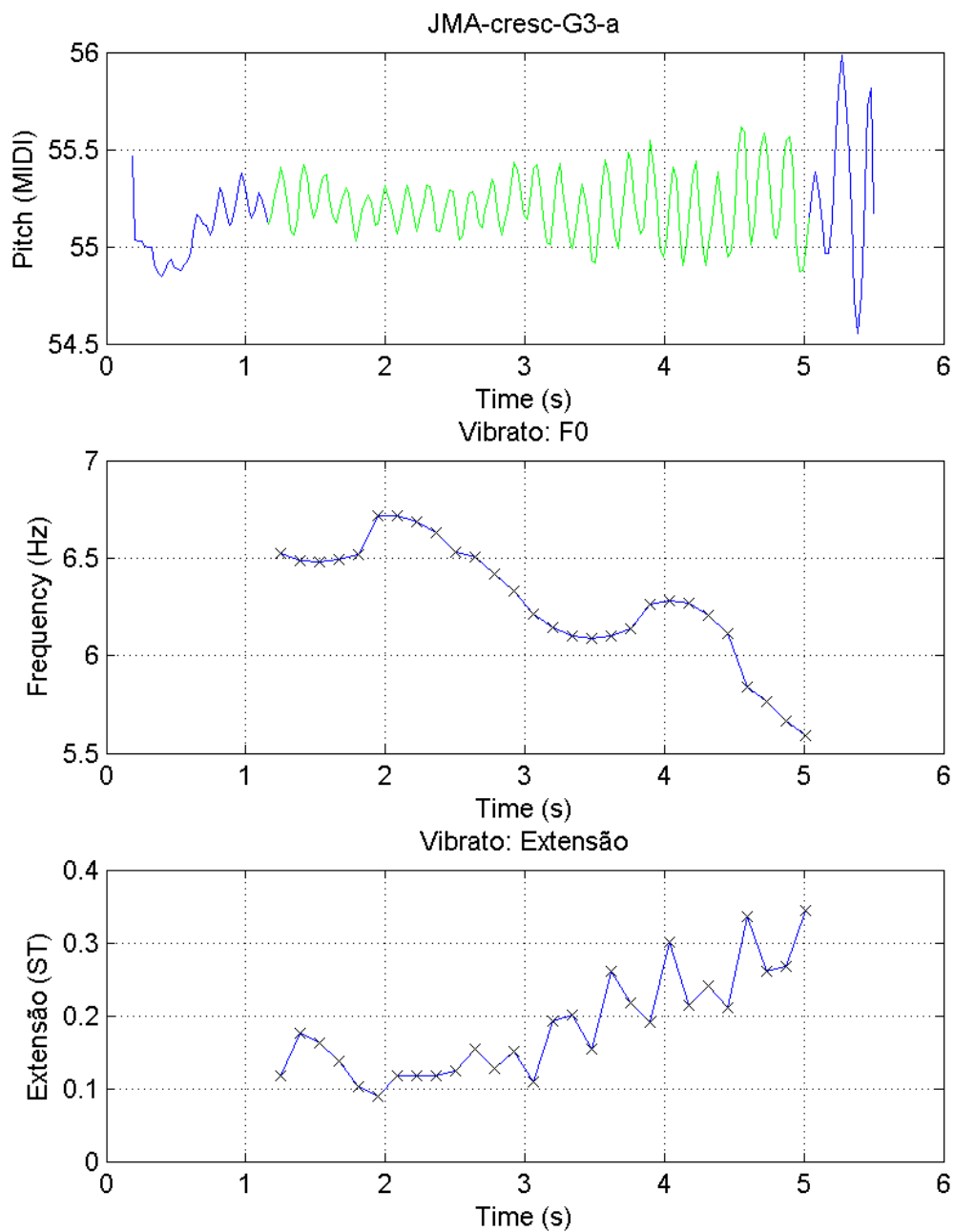


Figura F7 - Visualização da estimação dos parâmetros do vibrato, através do sinal JMA-cresc-G3-a.

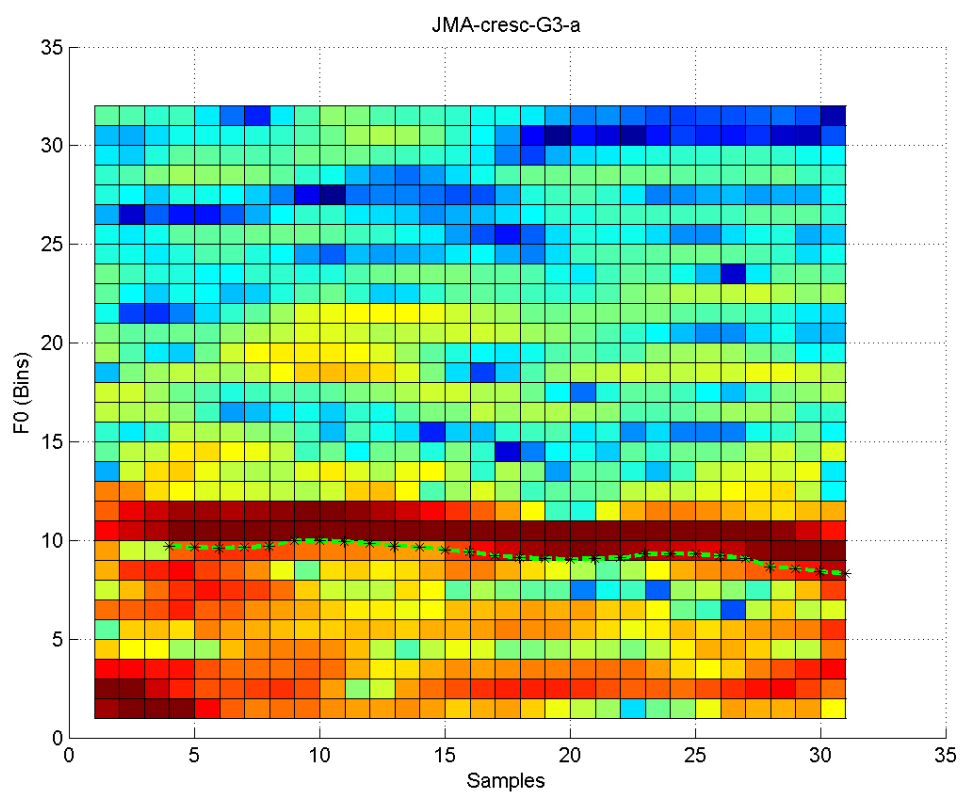


Figura F8 - Visualização do espectro de potências através da ODFT do sinal JMA-cresc-G3-a.

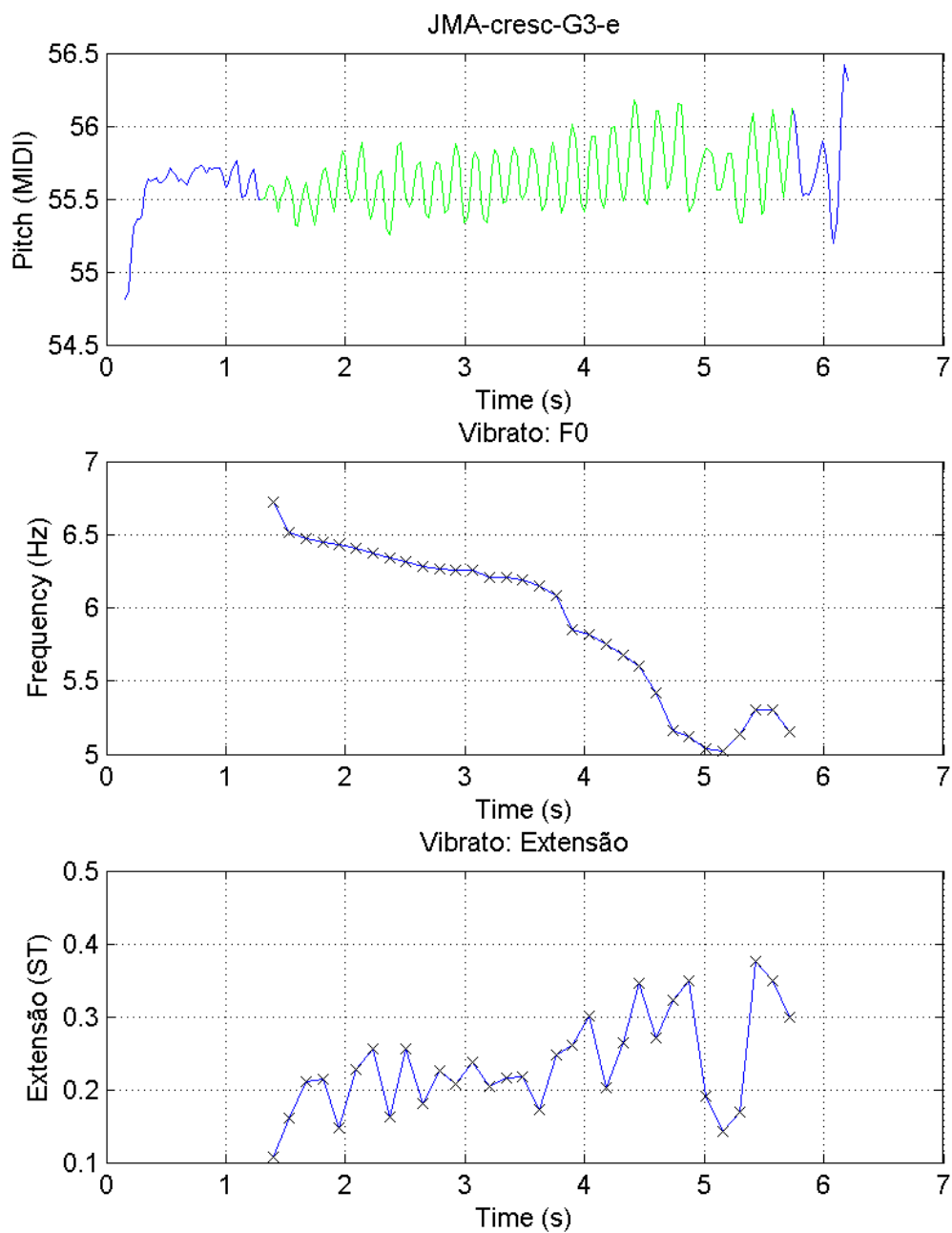


Figura F9 - Visualização da estimação dos parâmetros do vibrato, através do sinal JMA-cresc-G3-e.

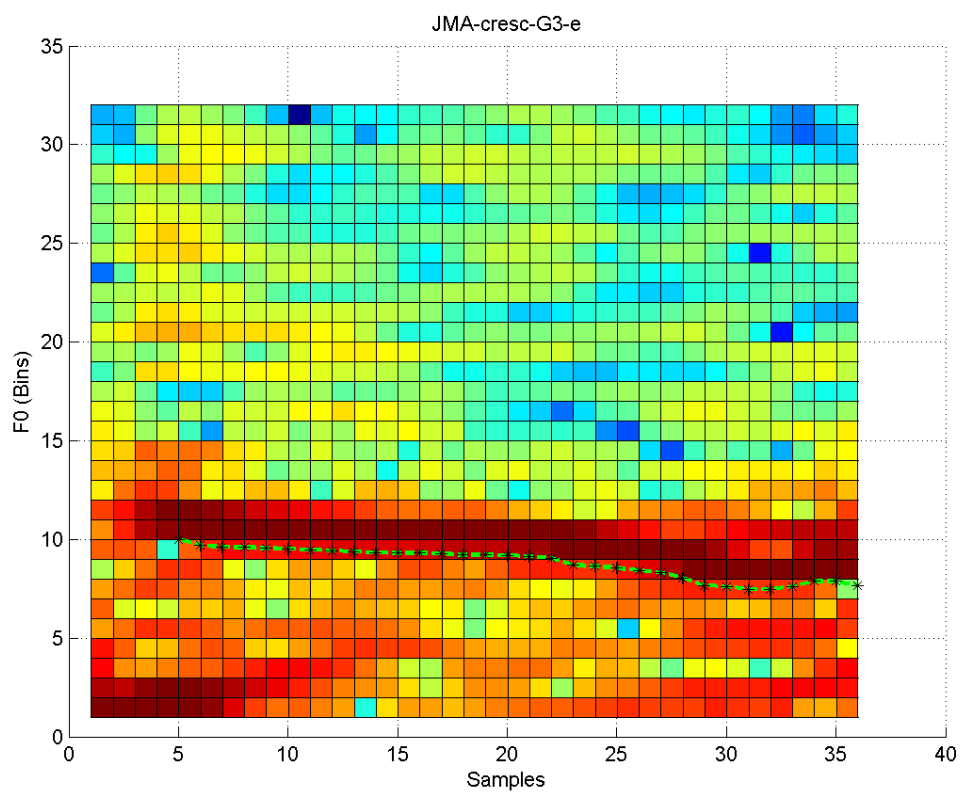


Figura F10 - Visualização do espectro de potências através da ODFT do sinal JMA-cresc-G3-a.

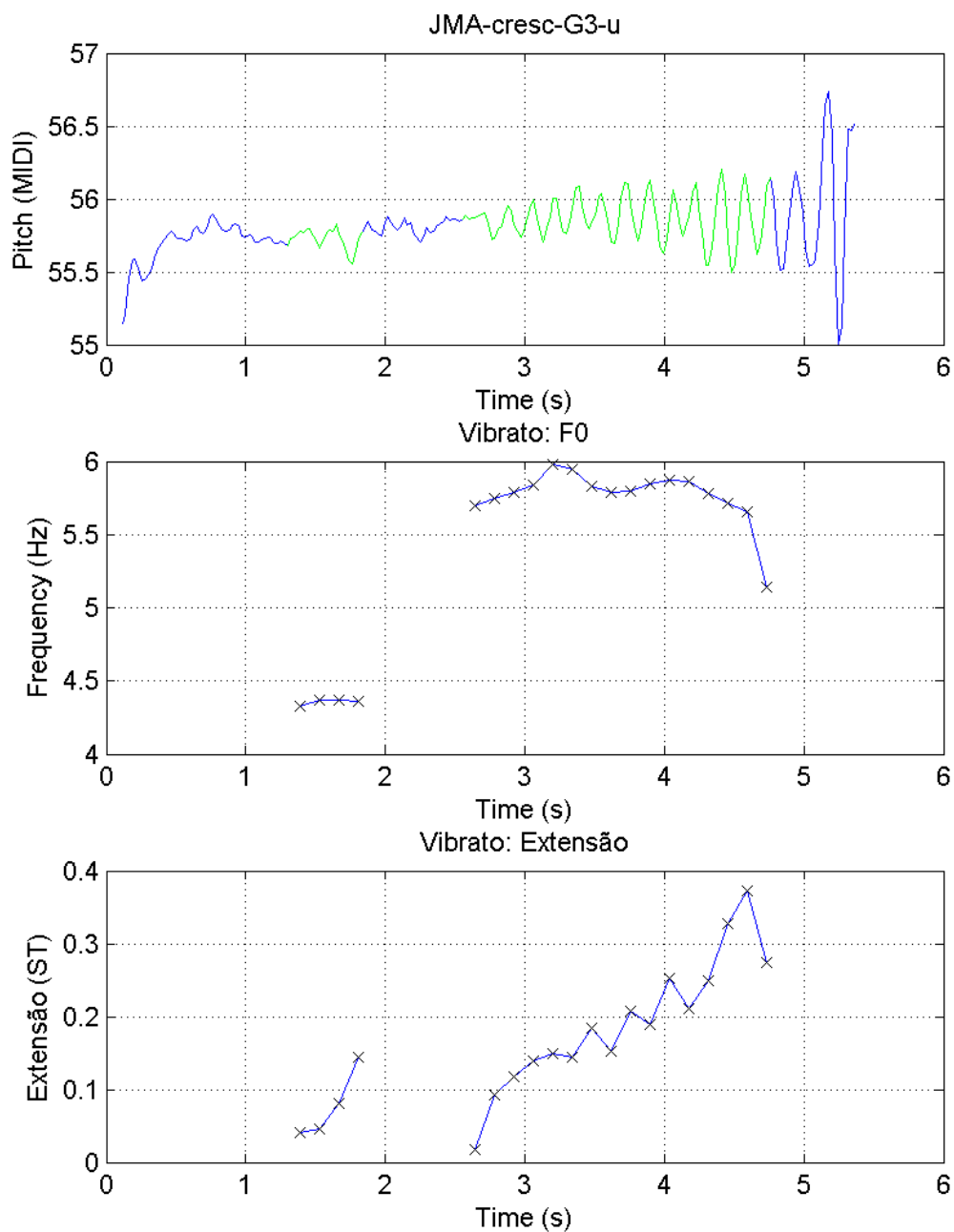


Figura F11 - Visualização da estimação dos parâmetros do vibrato, através do sinal JMA-cresc-G3-u.

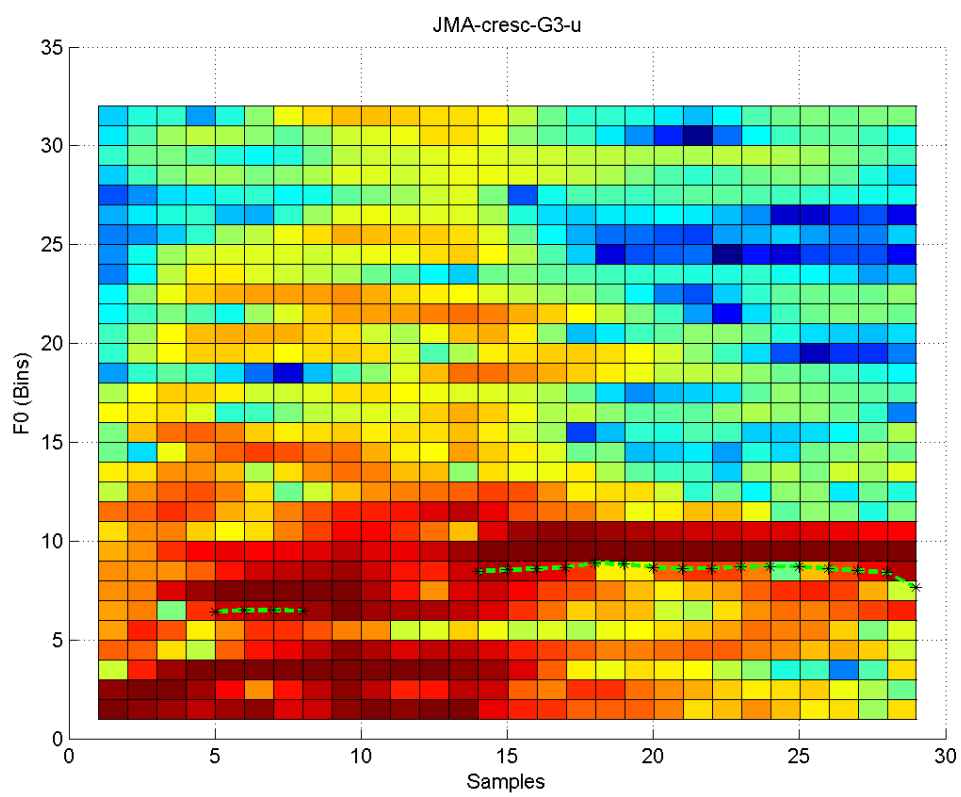


Figura F12 - Visualização do espectro de potências através da ODFT do sinal JMA-cresc-G3-u.

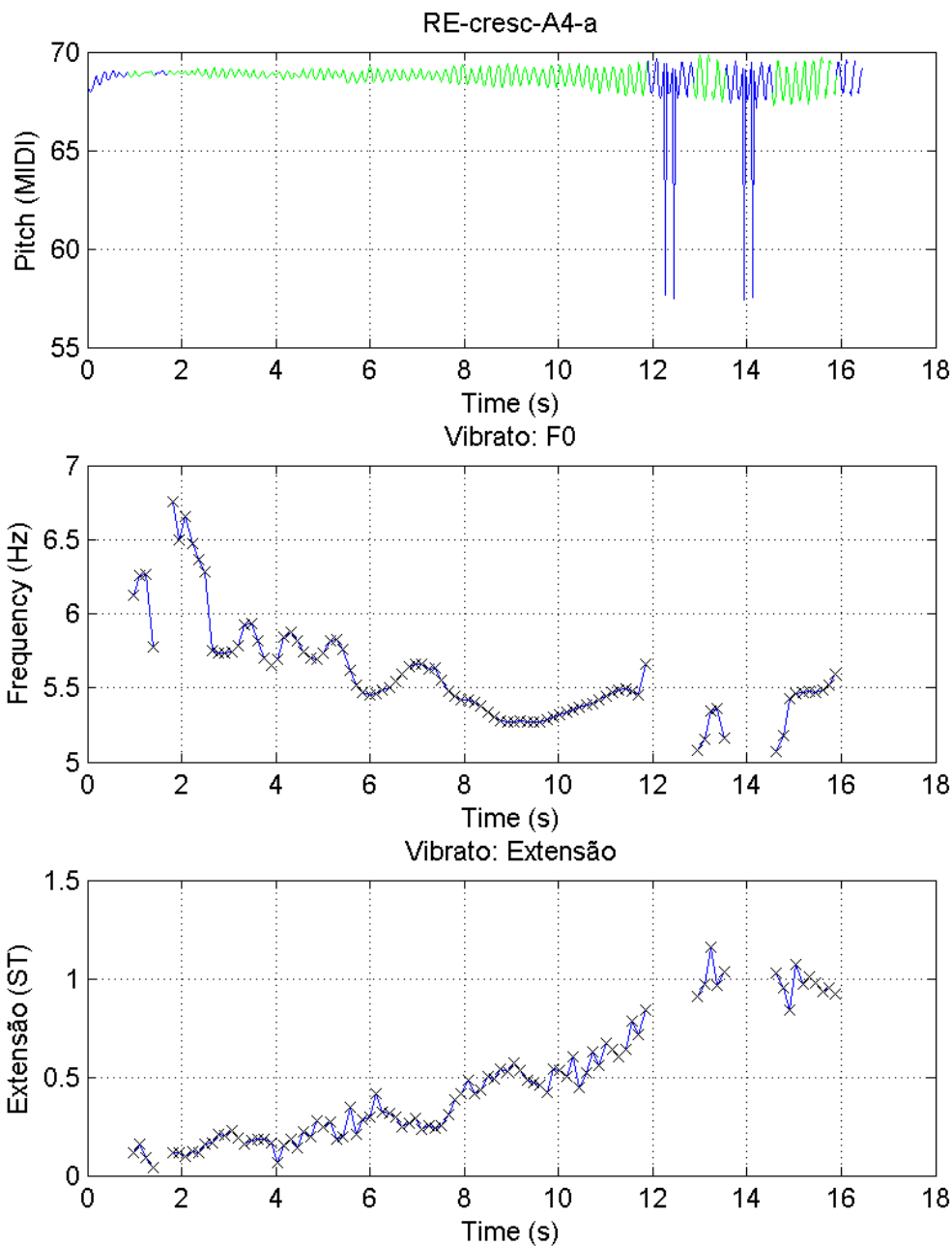


Figura F13 - Visualização da estimação dos parâmetros do vibrato, através do sinal RE-cresc-A4-a.

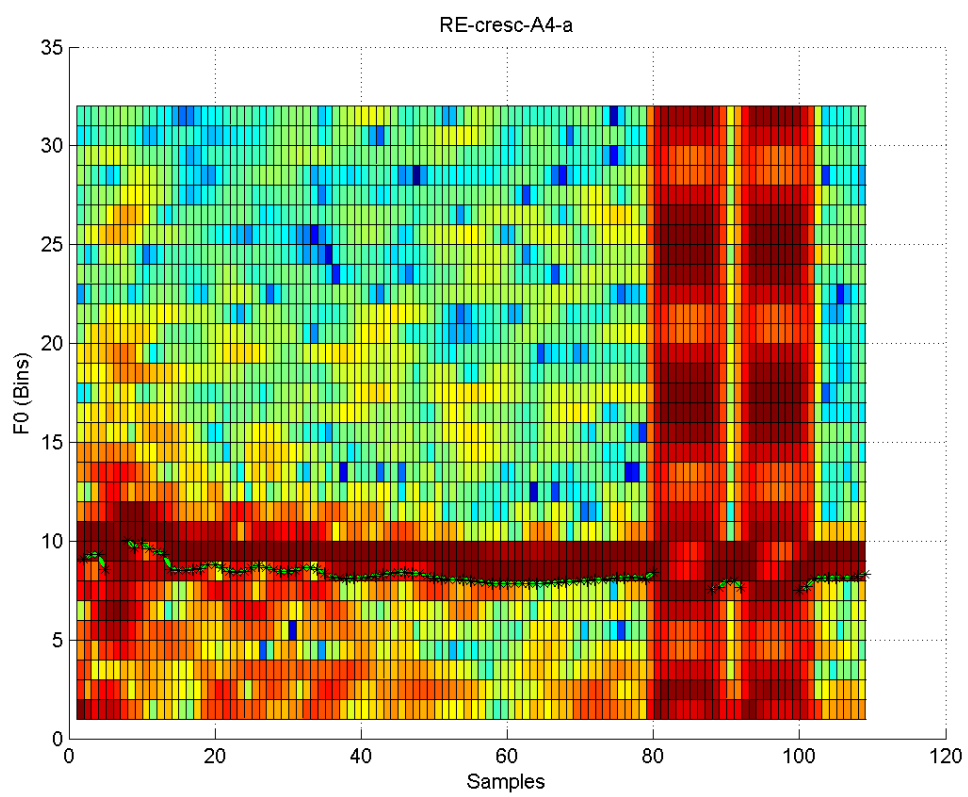


Figura F14 - Visualização do espectro de potências através da ODFT do sinal RE-cresc-A4-a.

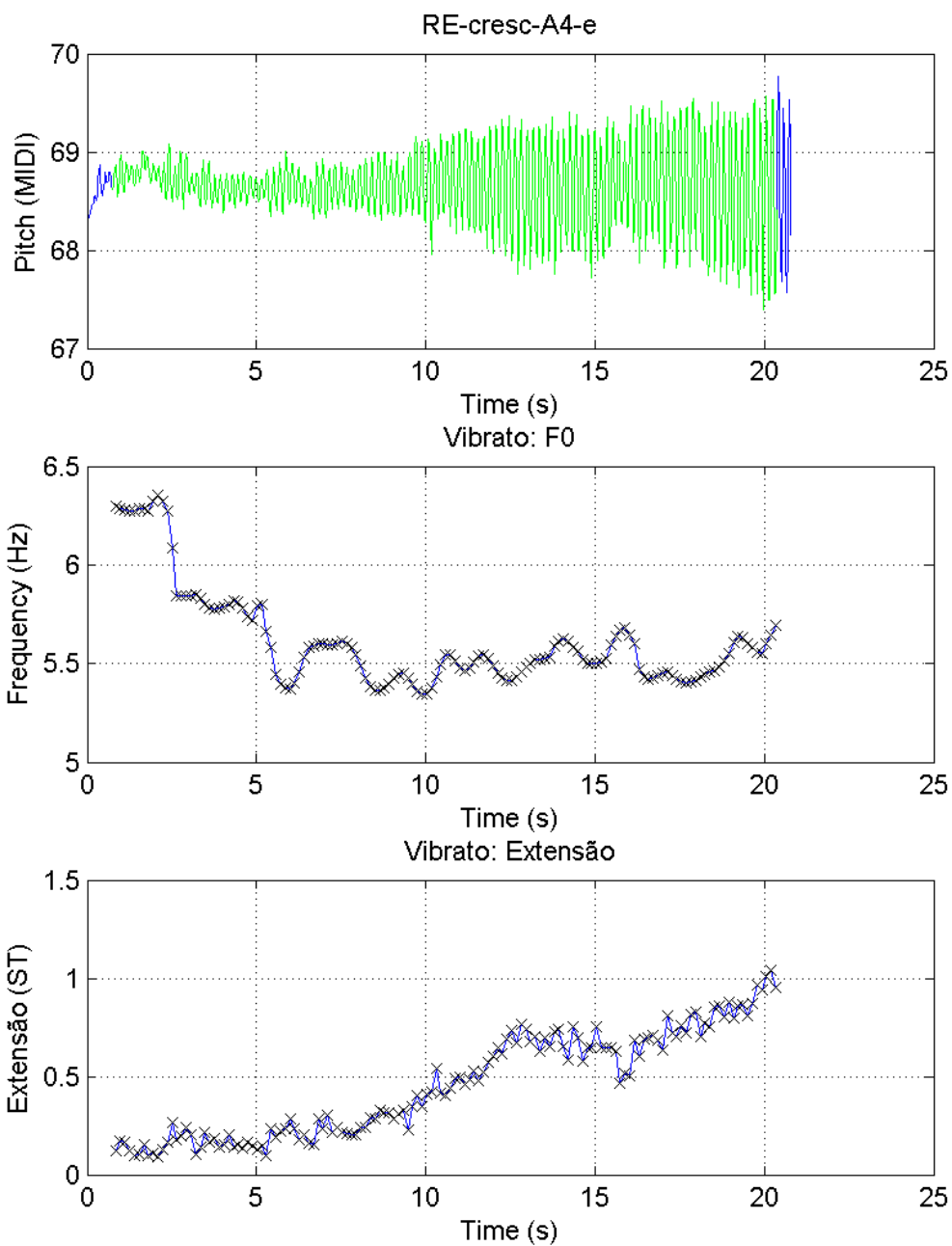


Figura F15 - Visualização da estimação dos parâmetros do vibrato, através do sinal RE-cresc-A4-e.

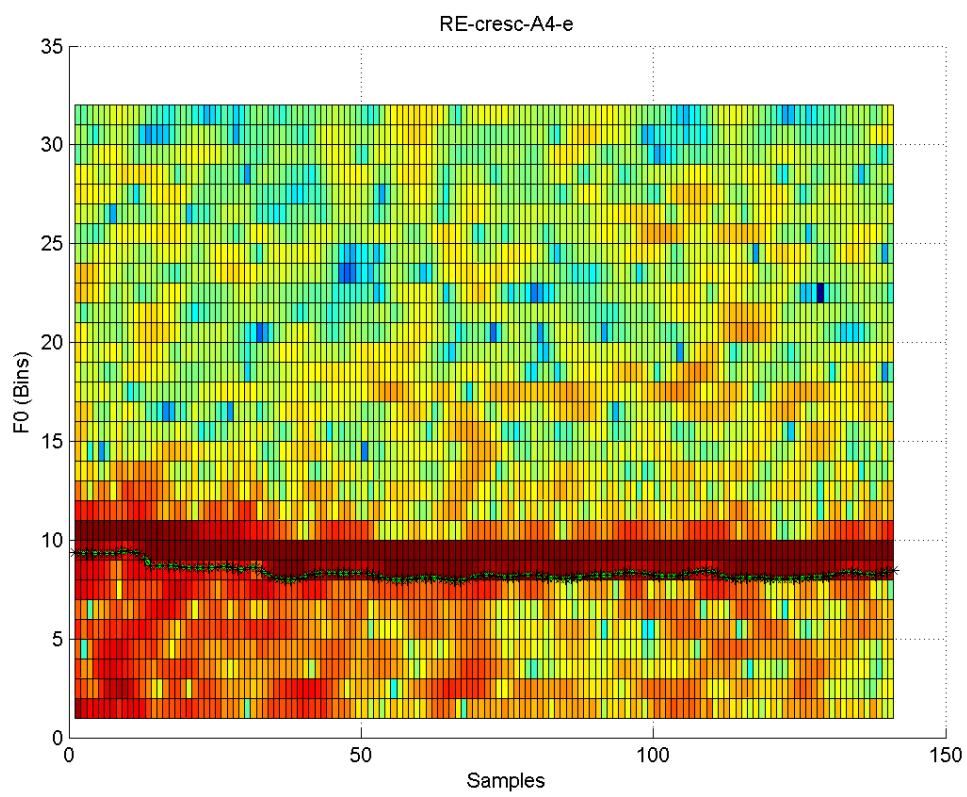


Figura F16 - Visualização do espectro de potências através da ODFT do sinal RE-cresc-A4-e.

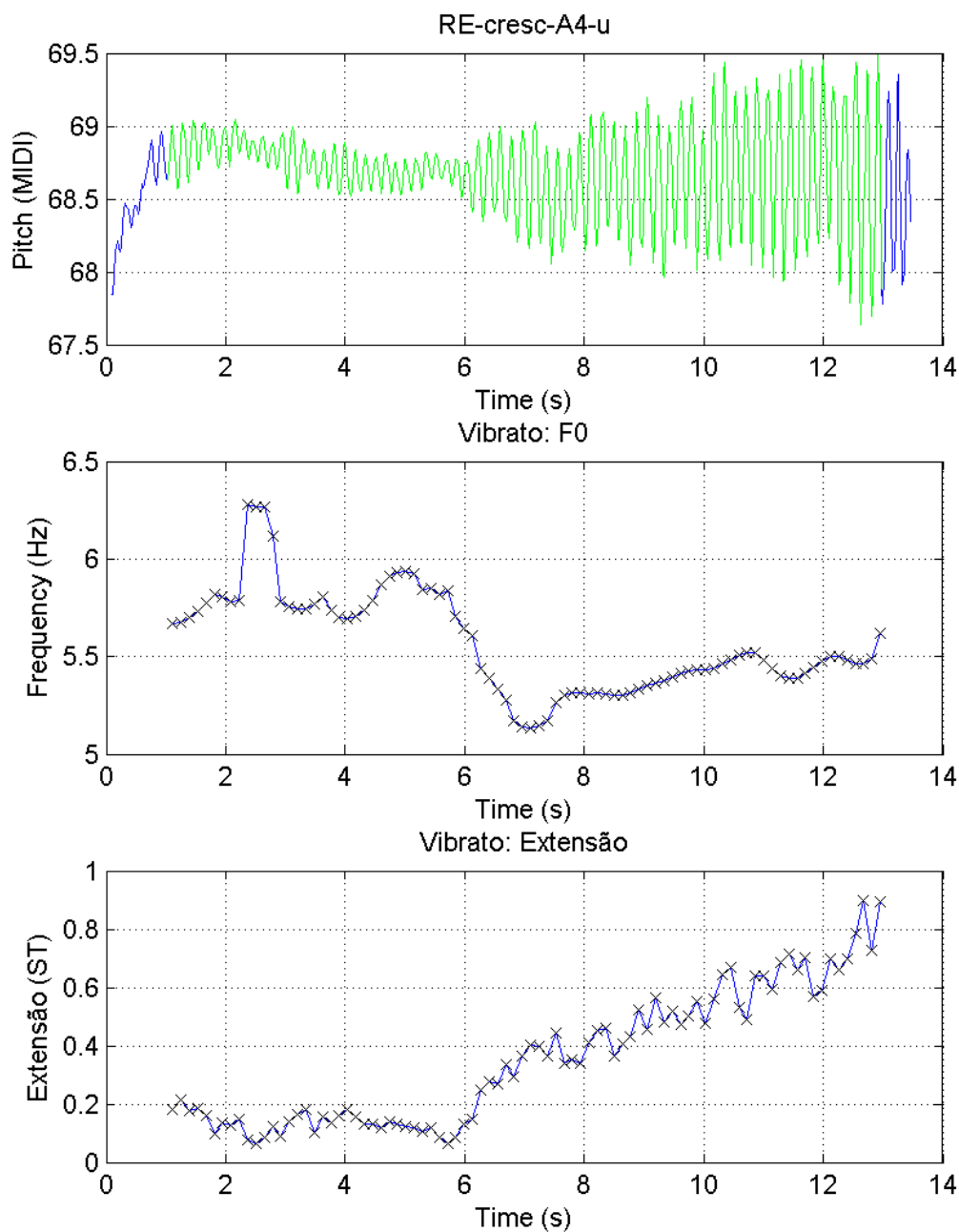


Figura F17 - Visualização da estimação dos parâmetros do vibrato, através do sinal RE-cresc-A4-u.

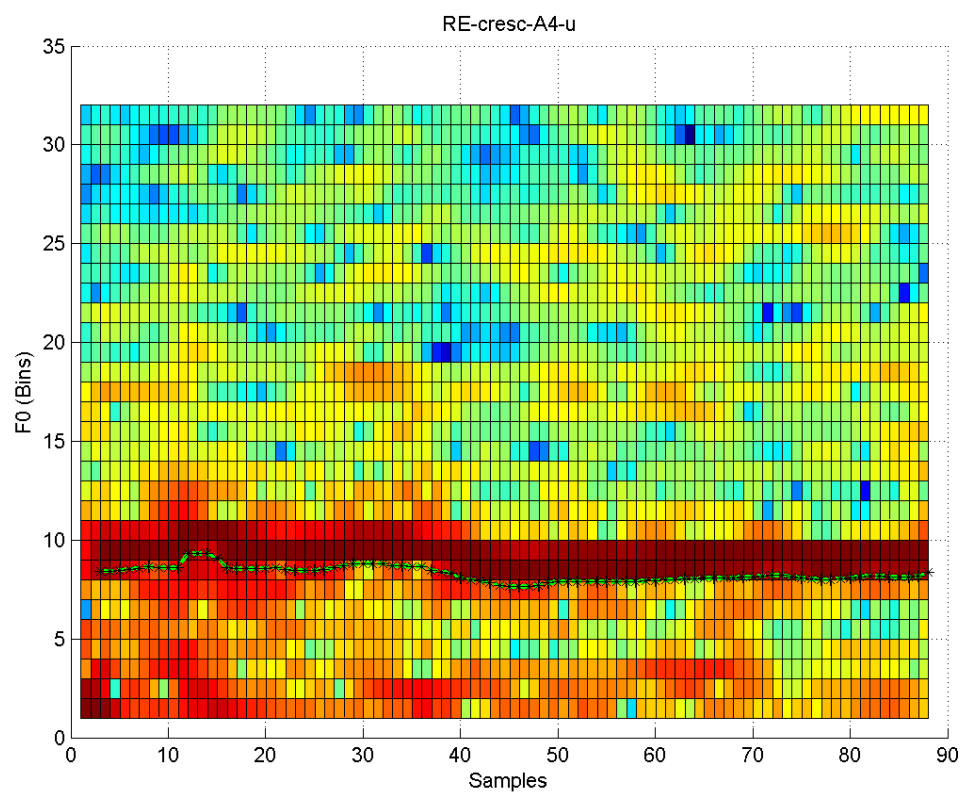


Figura F18 - Visualização do espectro de potências através da ODFT do sinal RE-cresc-A4-u.

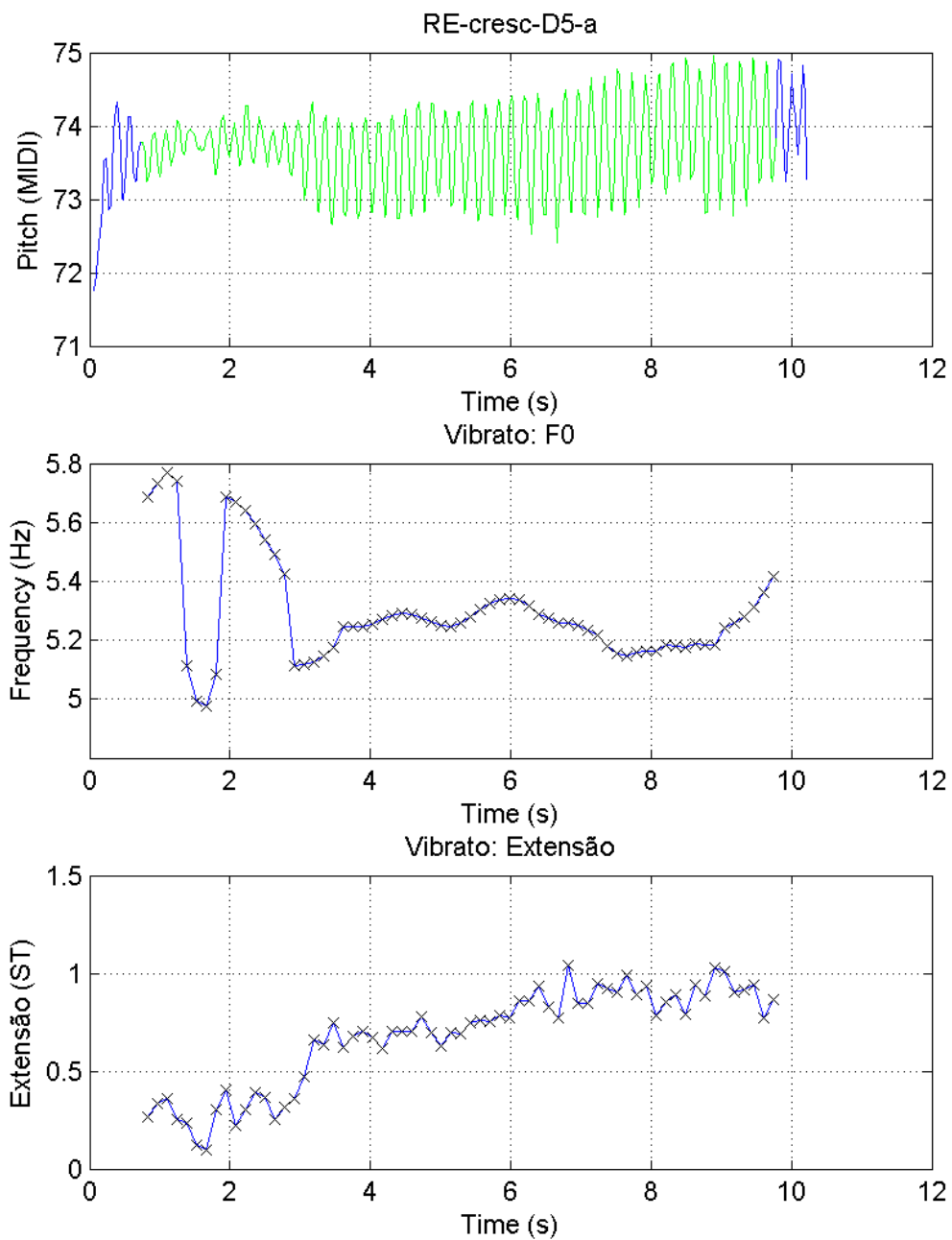


Figura F19 - Visualização da estimação dos parâmetros do vibrato, através do sinal RE-cresc-D5-a.

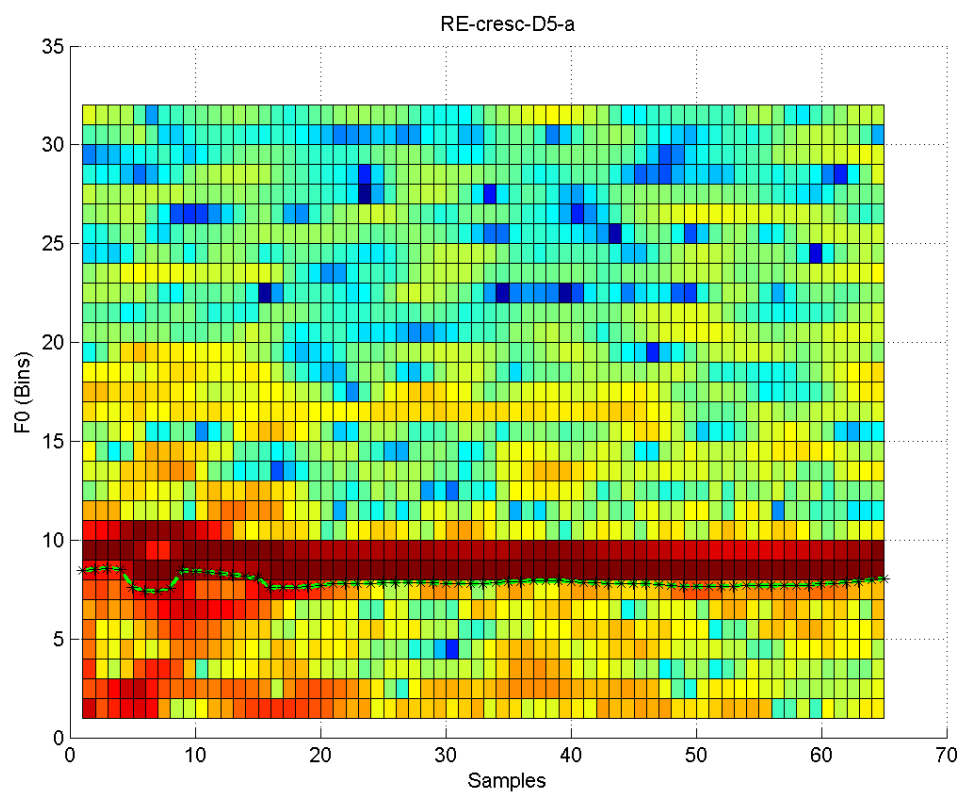


Figura F20 - Visualização do espectro de potências através da ODFT do sinal RE-cresc-D5-a.

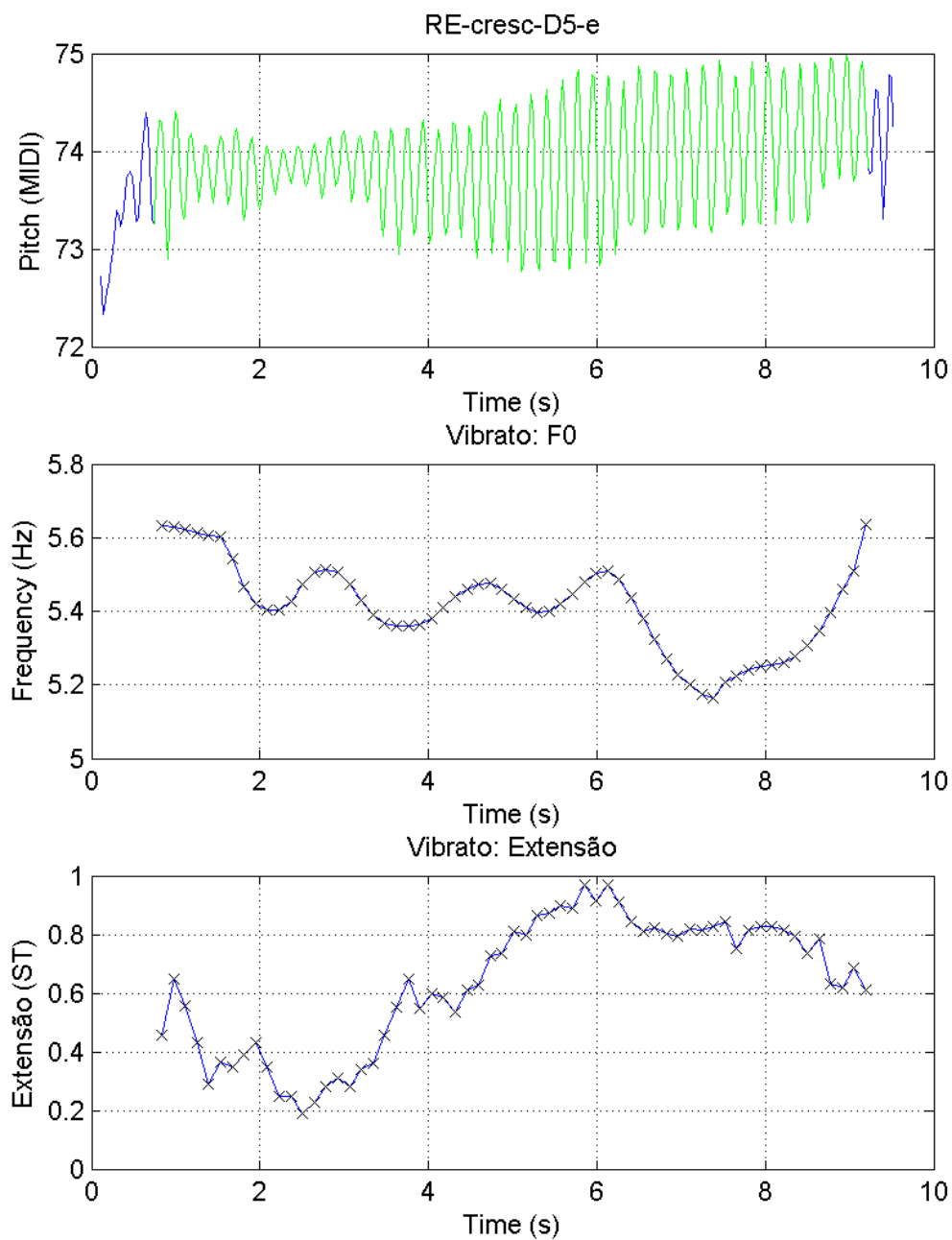


Figura F21 - Visualização da estimação dos parâmetros do vibrato, através do sinal RE-cresc-D5-e.

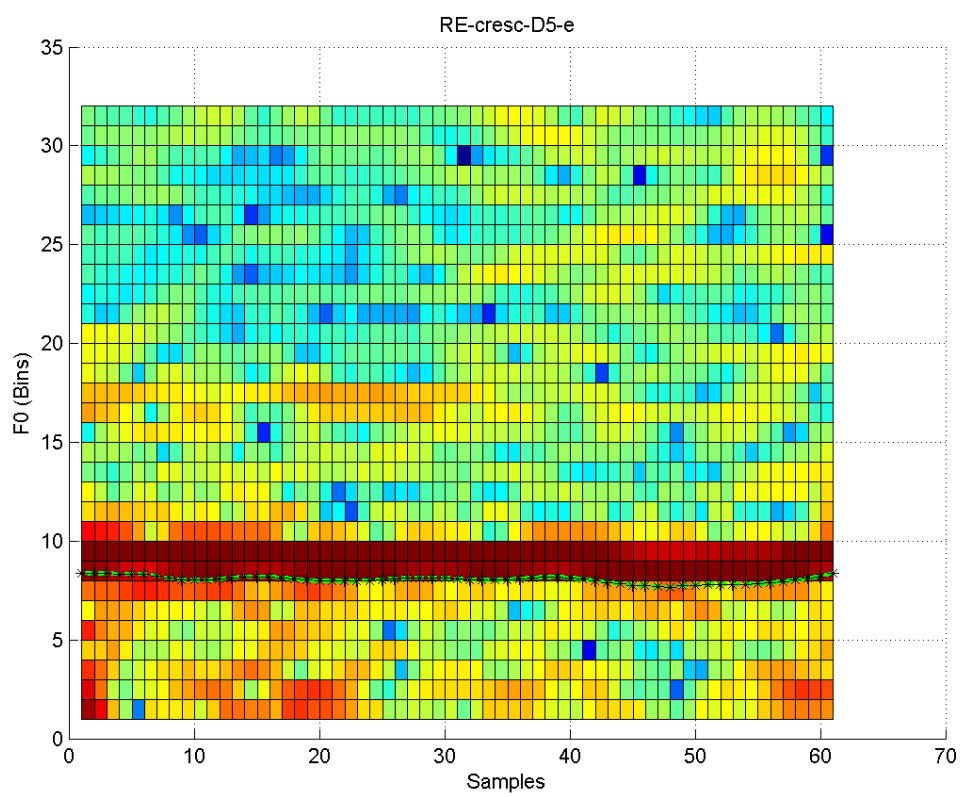


Figura F22 - Visualização do espectro de potências através da ODFT do sinal RE-cresc-D5-e.

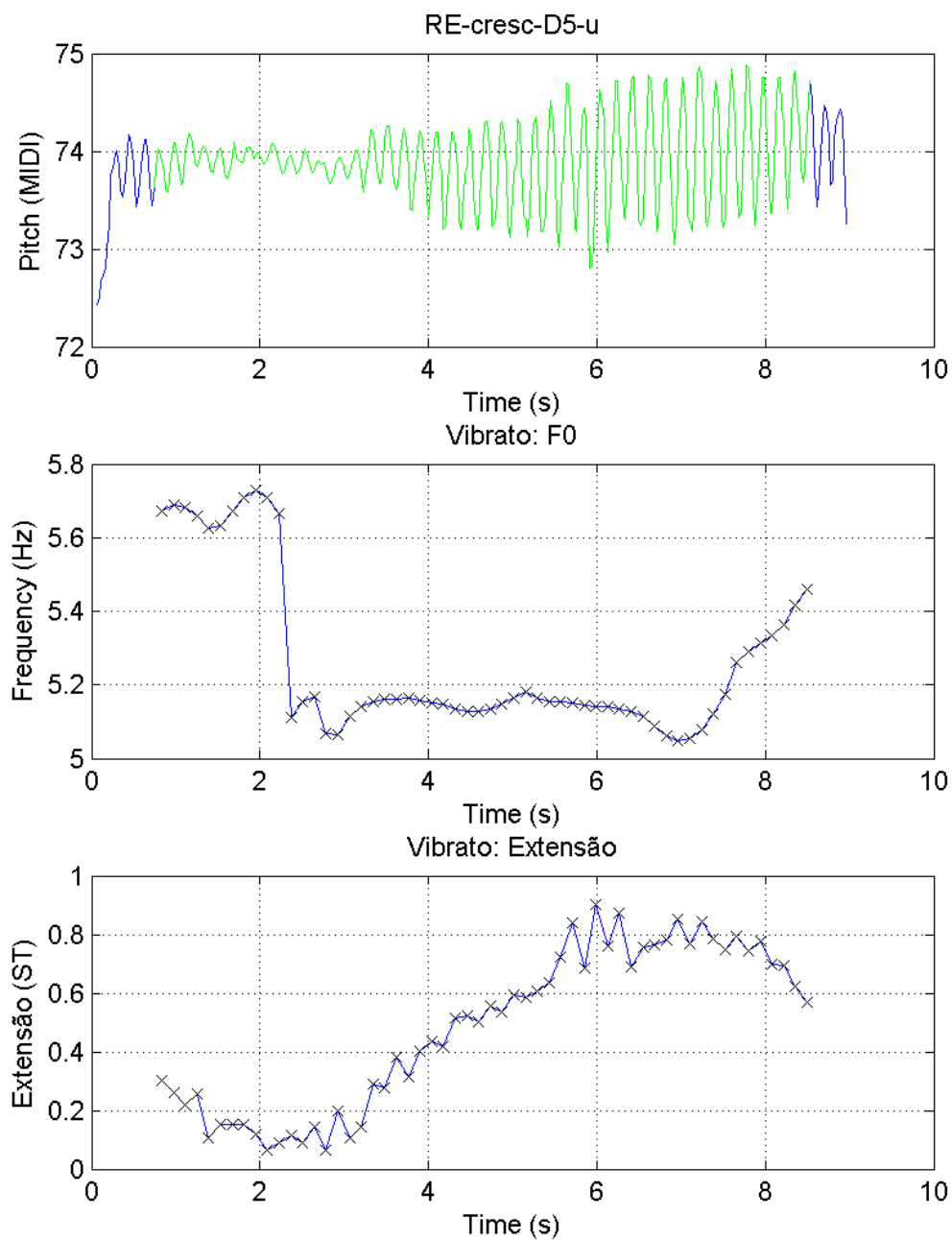


Figura F23 - Visualização da estimação dos parâmetros do vibrato, através do sinal RE-cresc-D5-u.

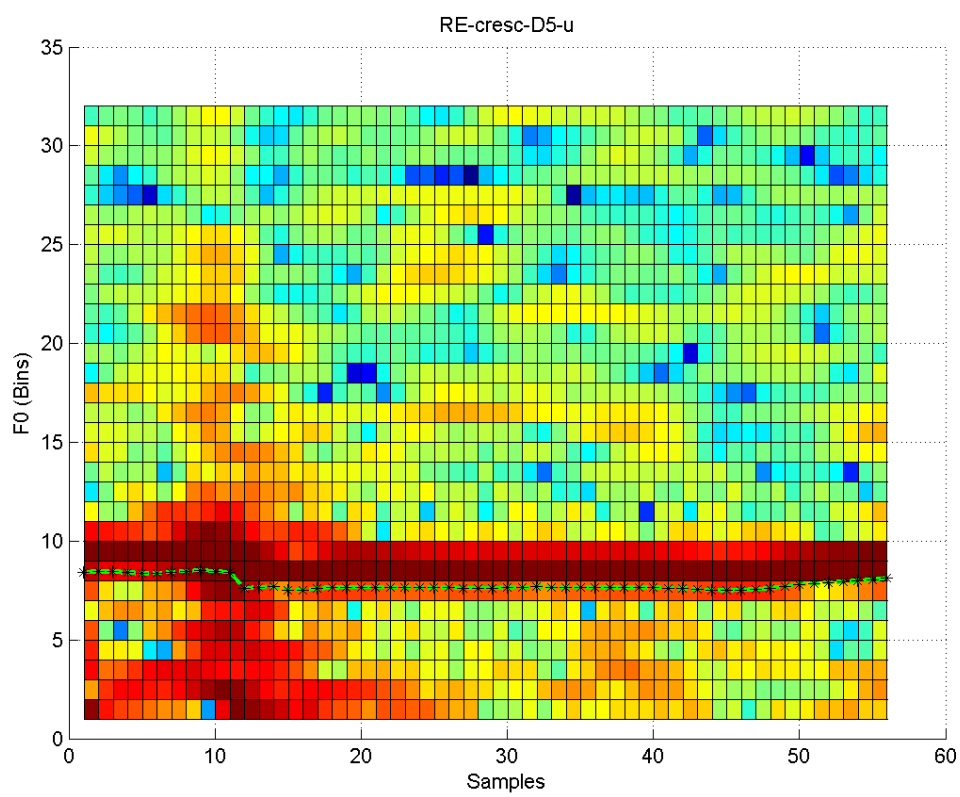


Figura F24 - Visualização do espectro de potências através da ODFT do sinal RE-cresc-D5-u.

## Anexo G

Neste anexo são apresentadas as figuras referentes aos resultados dos testes efectuados ao algoritmo de estimação do vibrato, através do uso de vozes naturais de maior duração e com variação abrupta de F0.

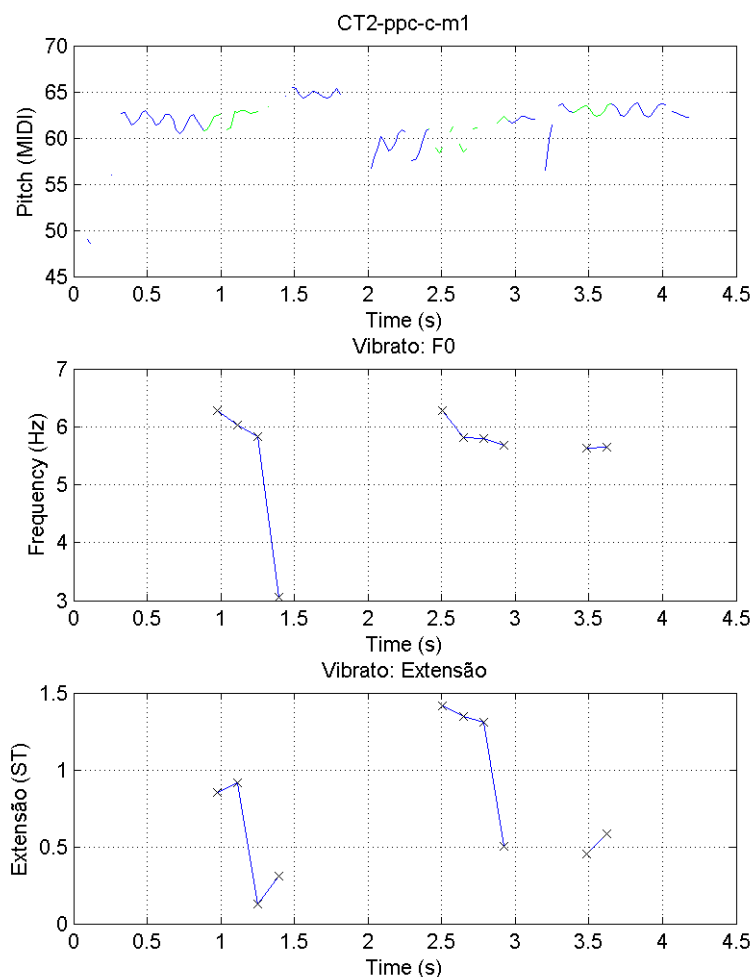


Figura G1 - Visualização da estimação dos parâmetros do vibrato, através do sinal CT2-ppc-c-m1.

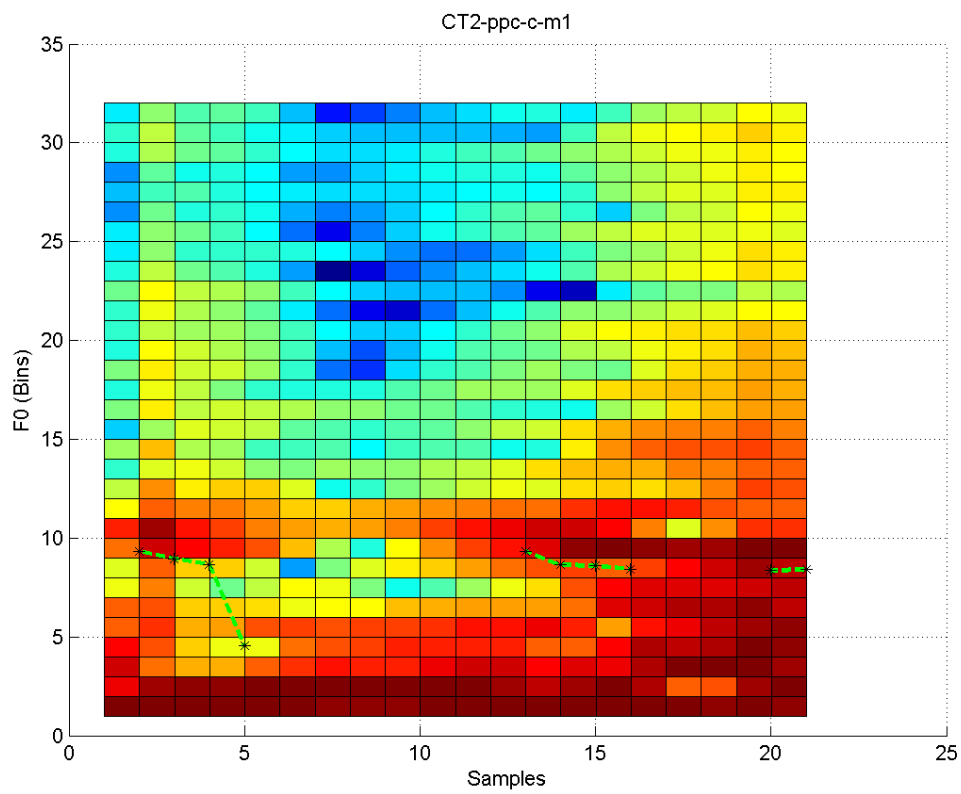


Figura G2 - Visualização do espectro de potências através da ODFT do sinal CT2-ppc-c-m1.

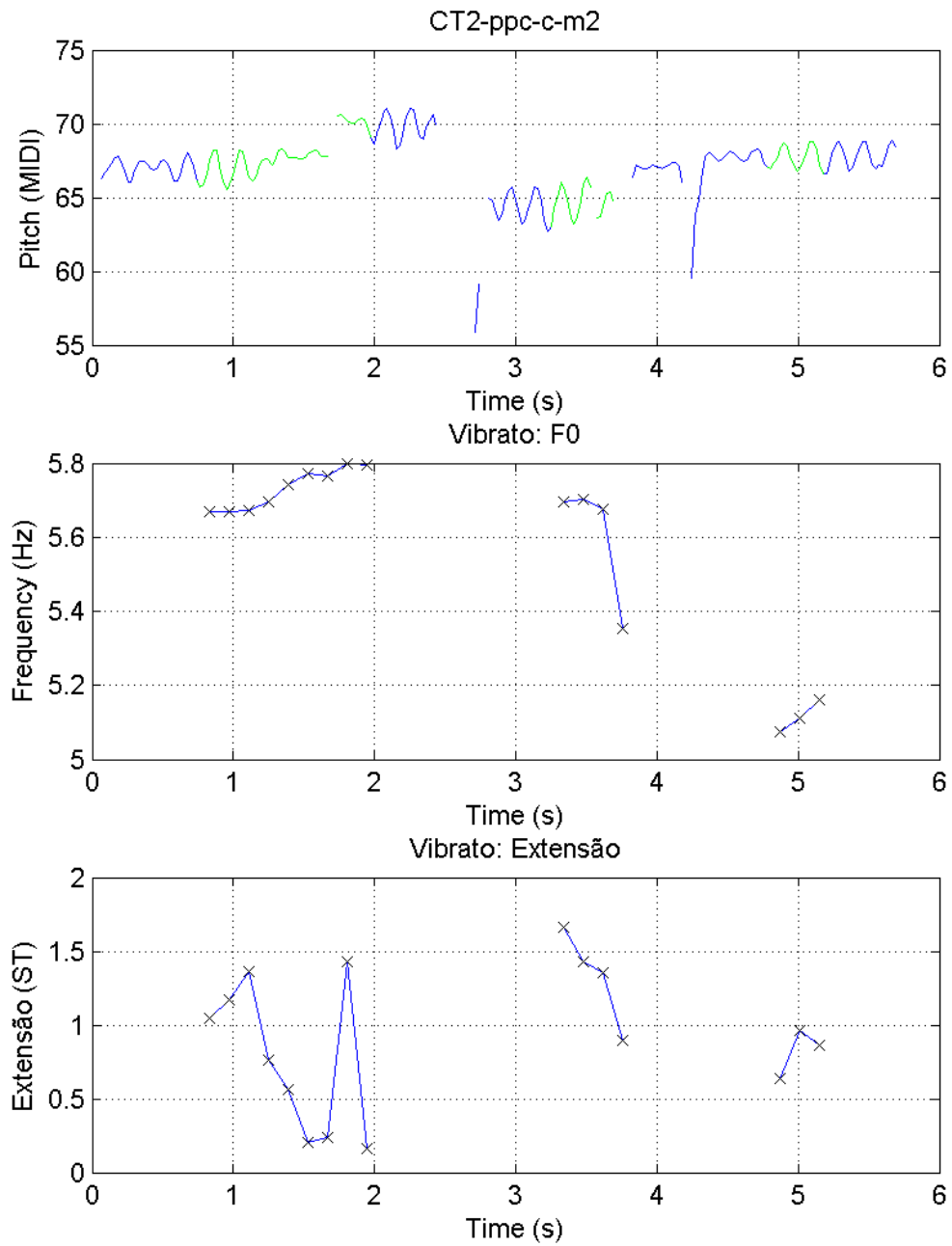


Figura G3 - Visualização da estimação dos parâmetros do vibrato, através do sinal CT2-ppc-c-m2.

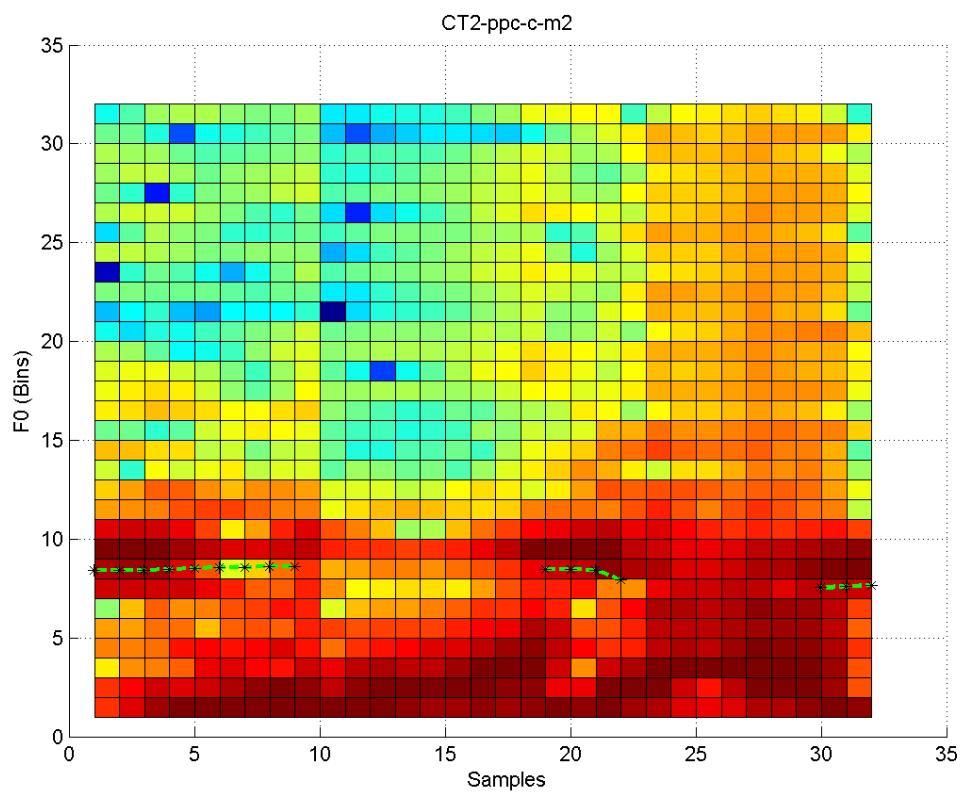


Figura G4 - Visualização do espectro de potências através da ODFT do sinal CT2-ppc-c-m2.

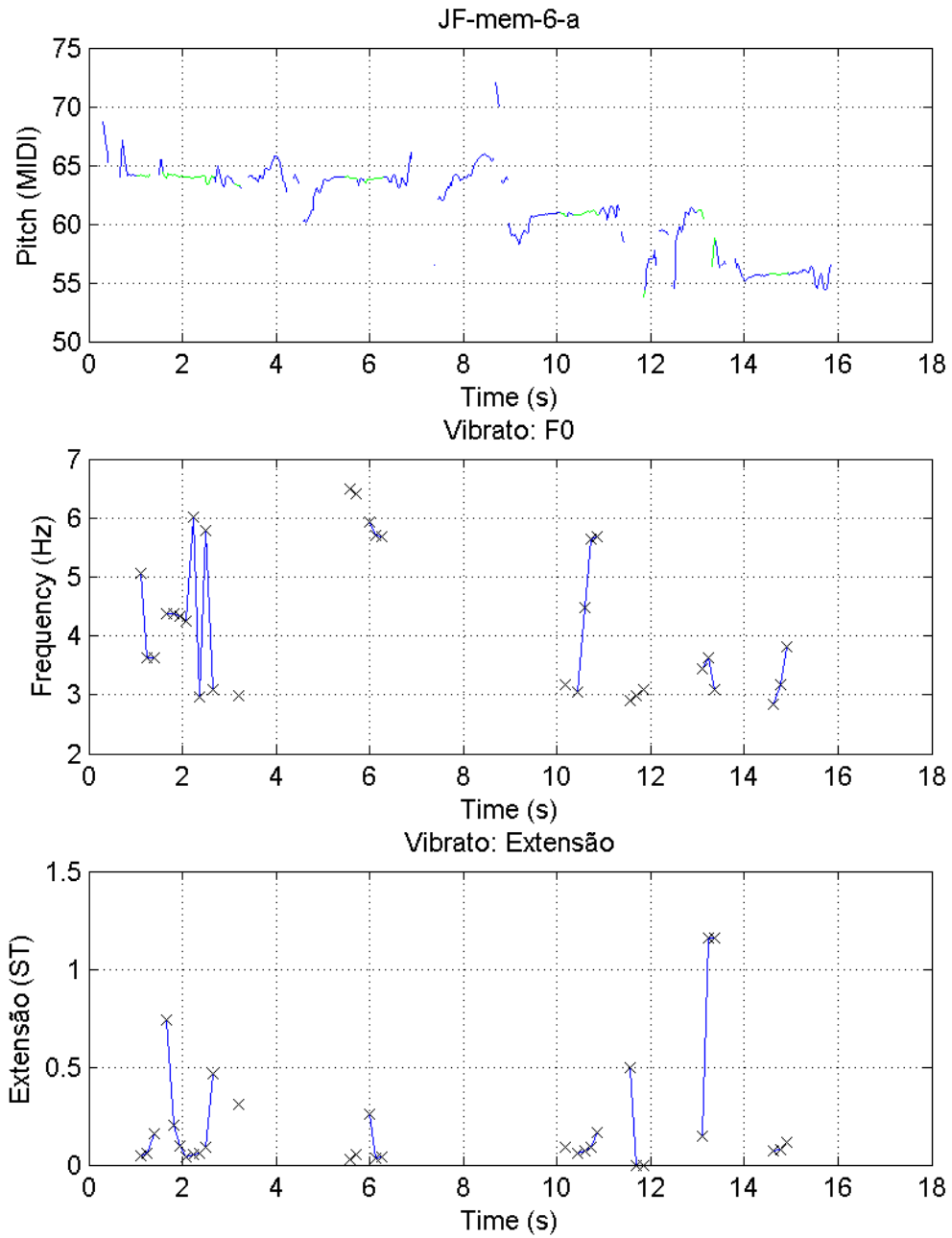


Figura G5 - Visualização da estimação dos parâmetros do vibrato, através do sinal JF-mem-6-a.

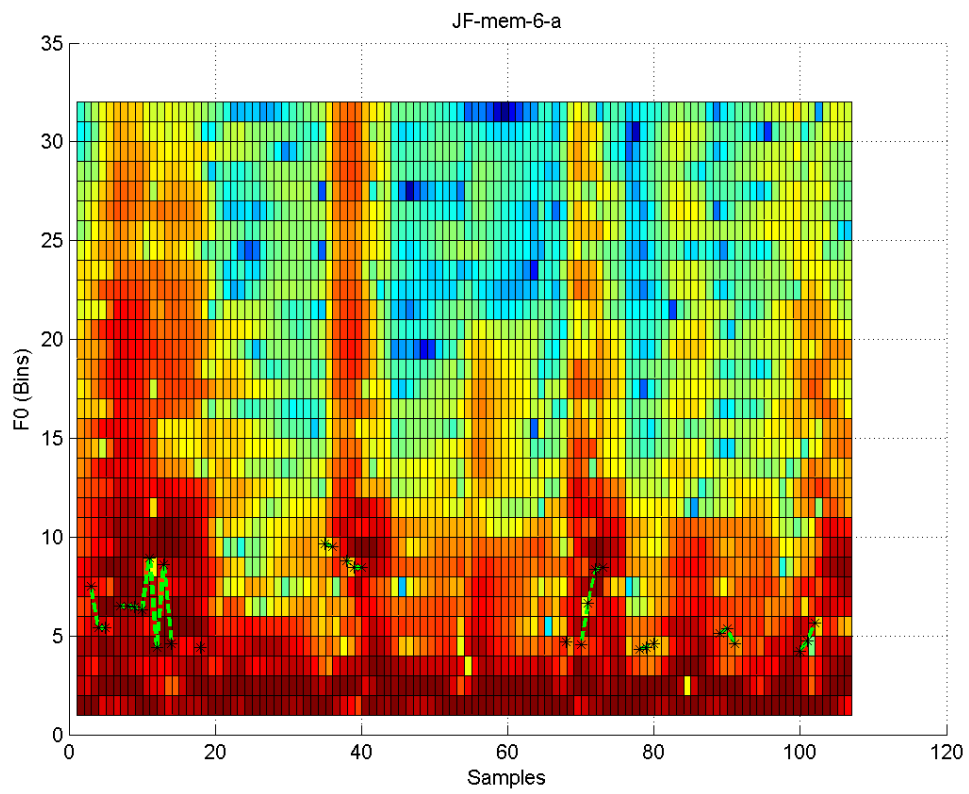


Figura G6 - Visualização do espectro de potências através da ODFT do sinal JF-mem-6-a.

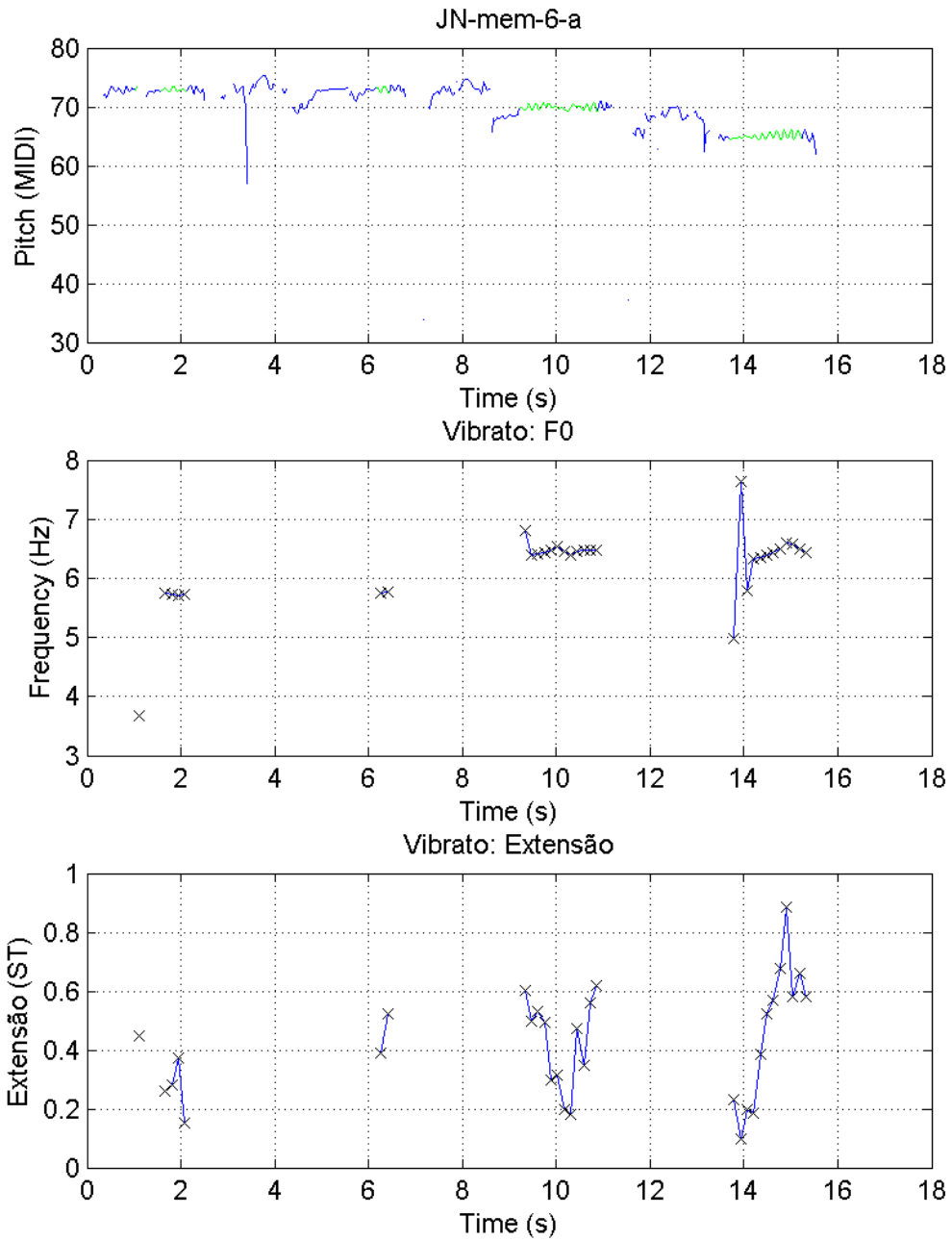


Figura G7 - Visualização da estimação dos parâmetros do vibrato, através do sinal JN-mem-6-a.

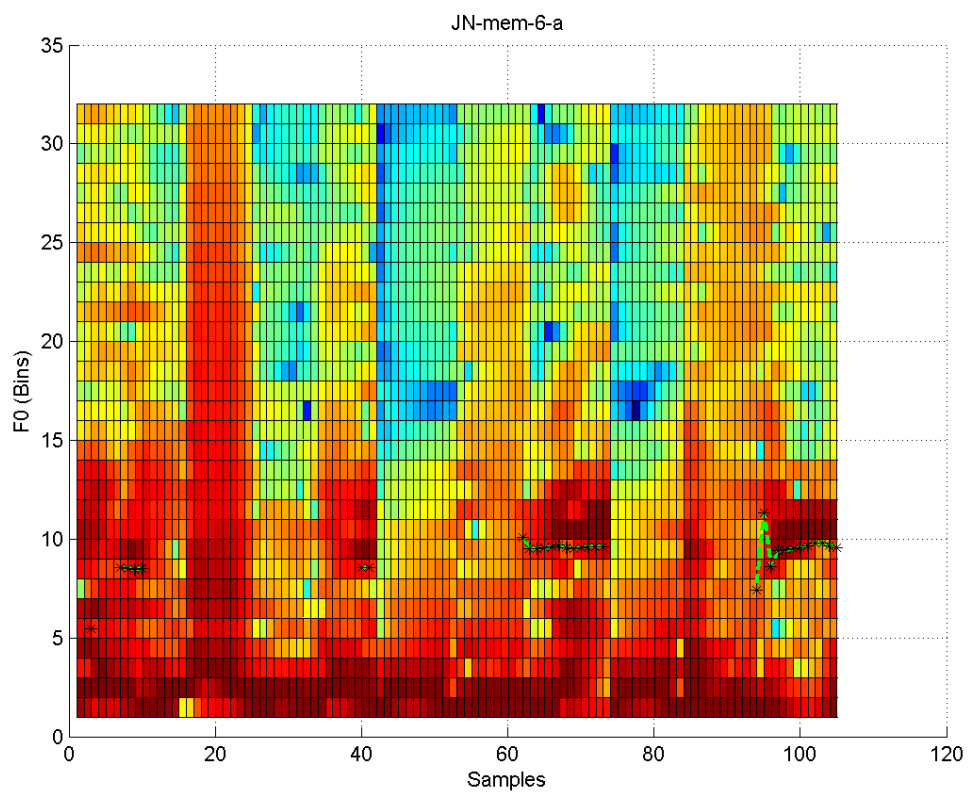


Figura G8 - Visualização do espectro de potências através da ODFT do sinal JN-mem-6-a.

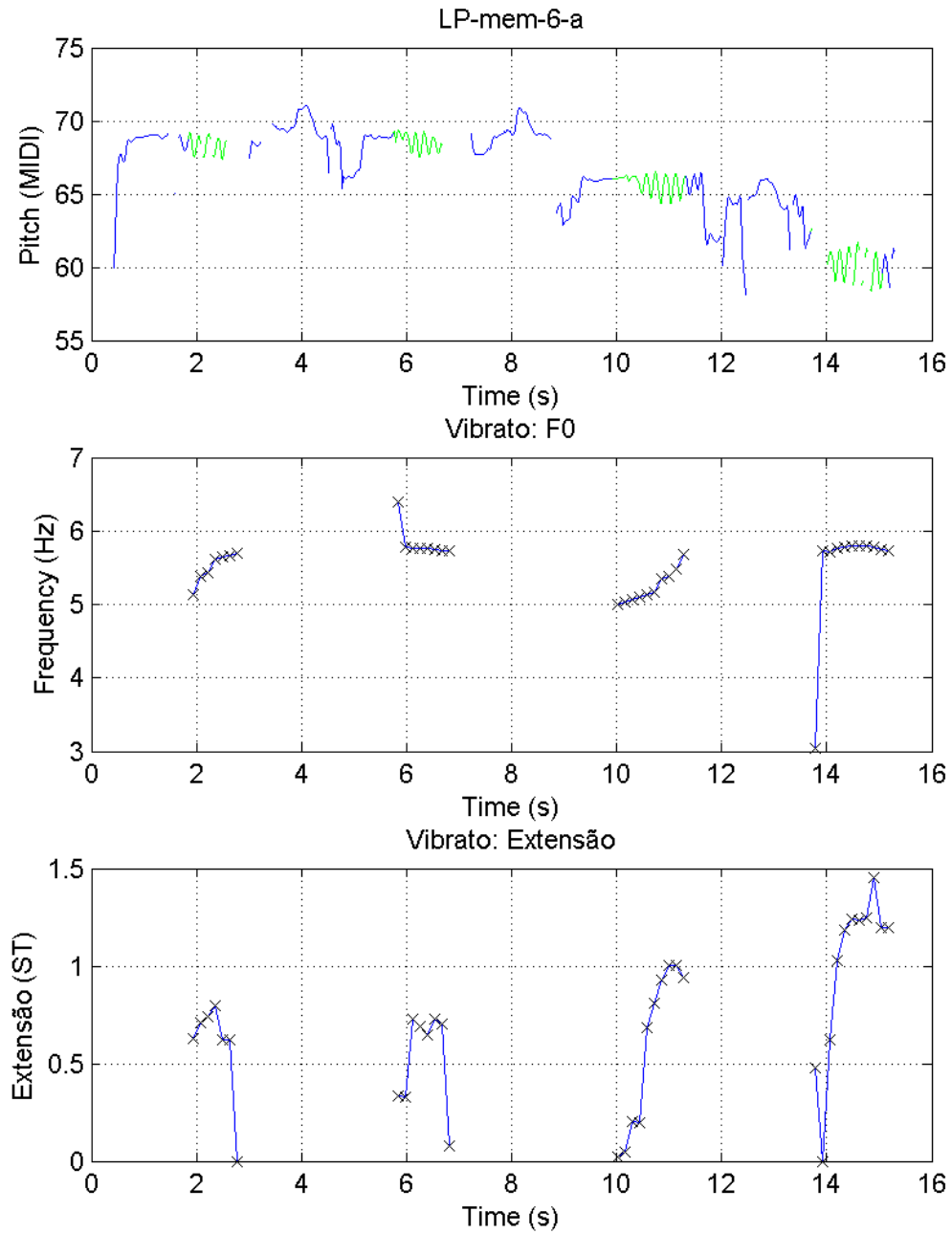


Figura G9 - Visualização da estimação dos parâmetros do vibrato, através do sinal LP-mem-6-a.

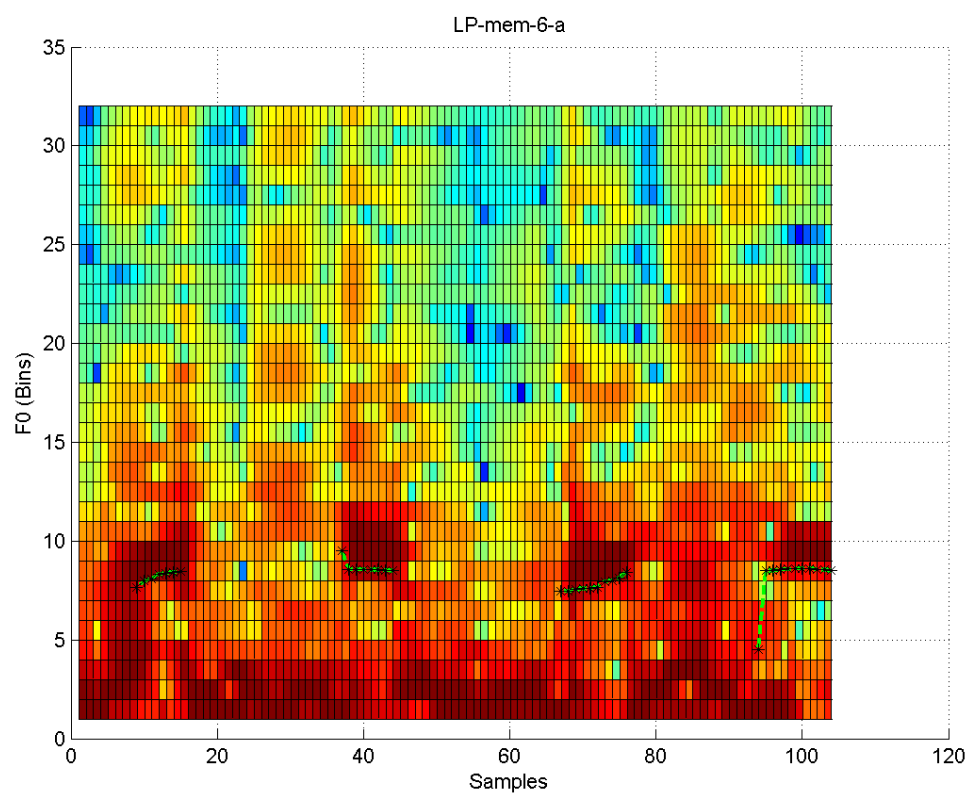


Figura G10 - Visualização do espectro de potências através da ODFT do sinal LP-mem-6-a.

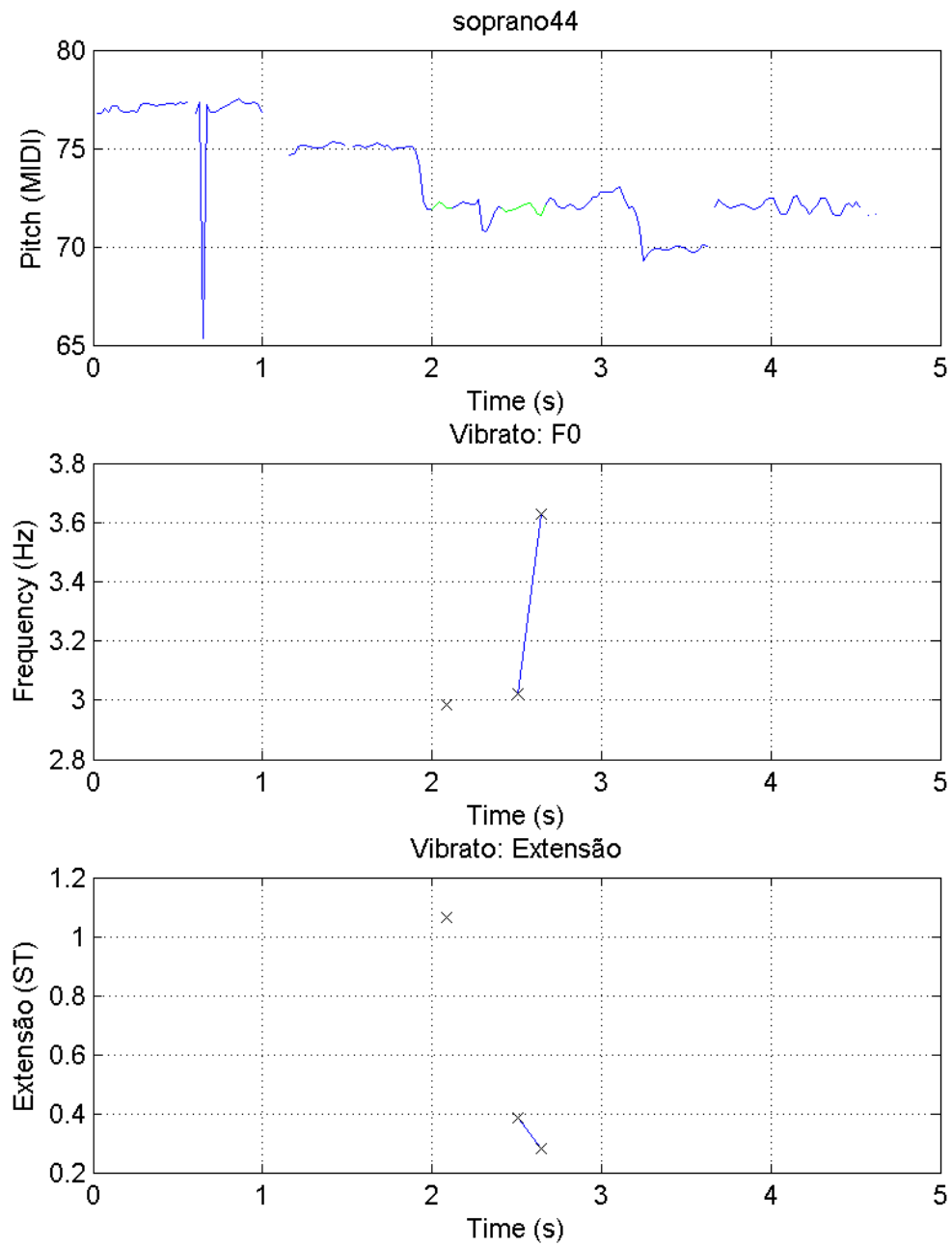


Figura G11 - Visualização da estimação dos parâmetros do vibrato, através do soprano44.

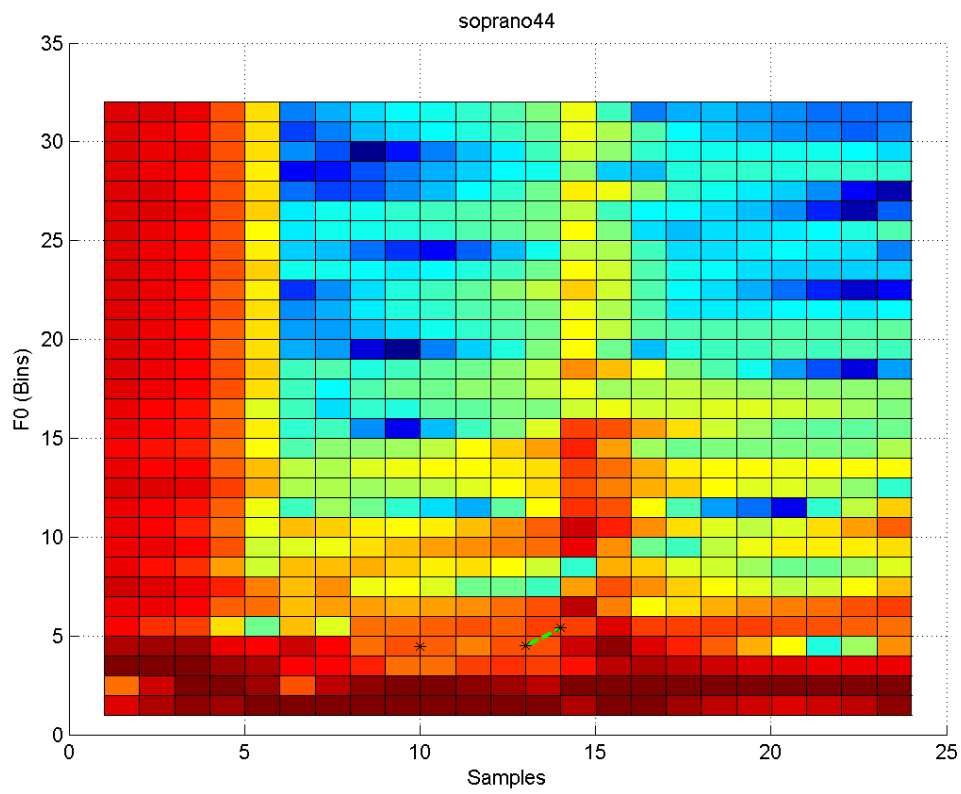


Figura G12 - Visualização do espectro de potências através da ODFT do sinal soprano44.

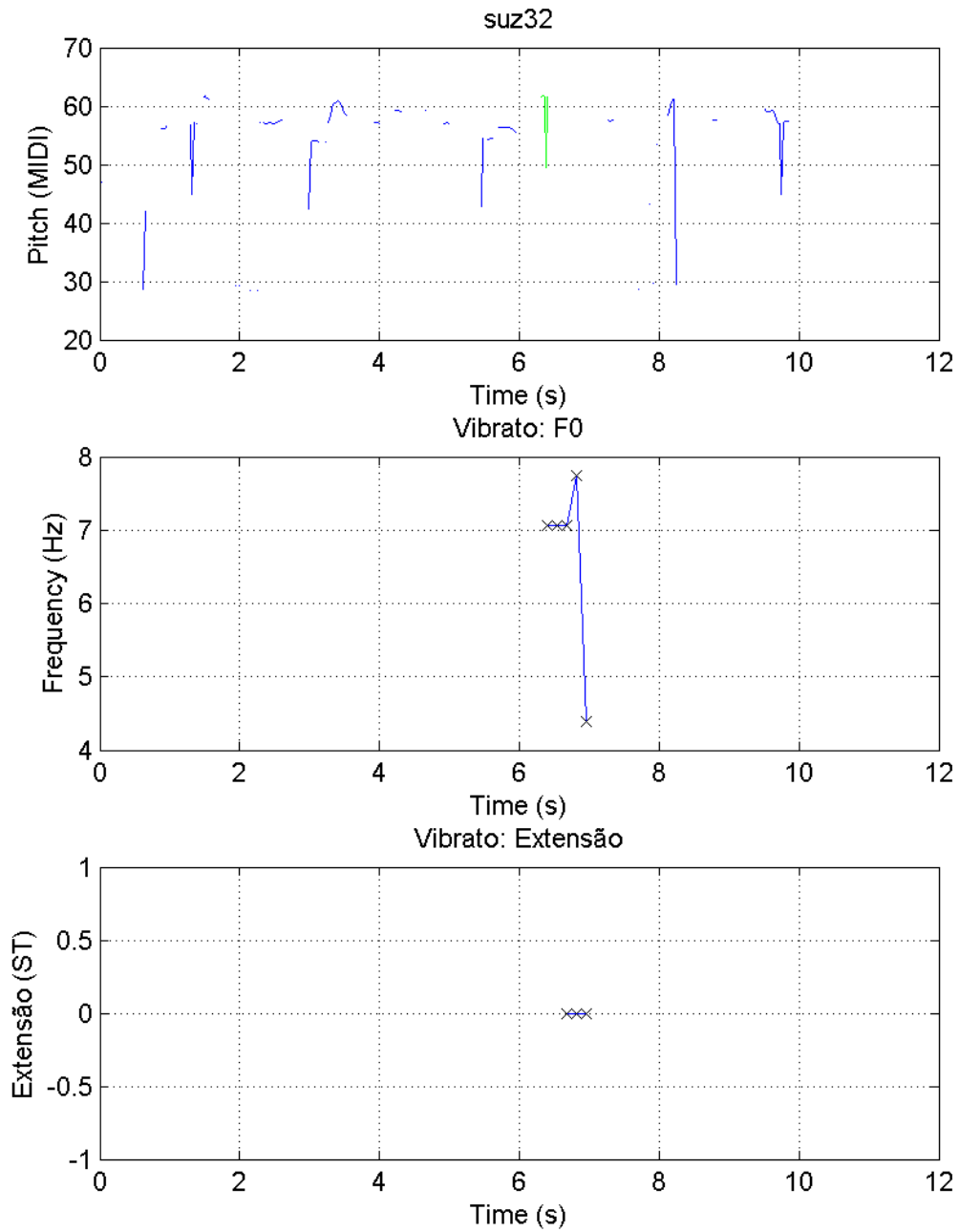


Figura G13 - Visualização da estimação dos parâmetros do vibrato, através do suz32.

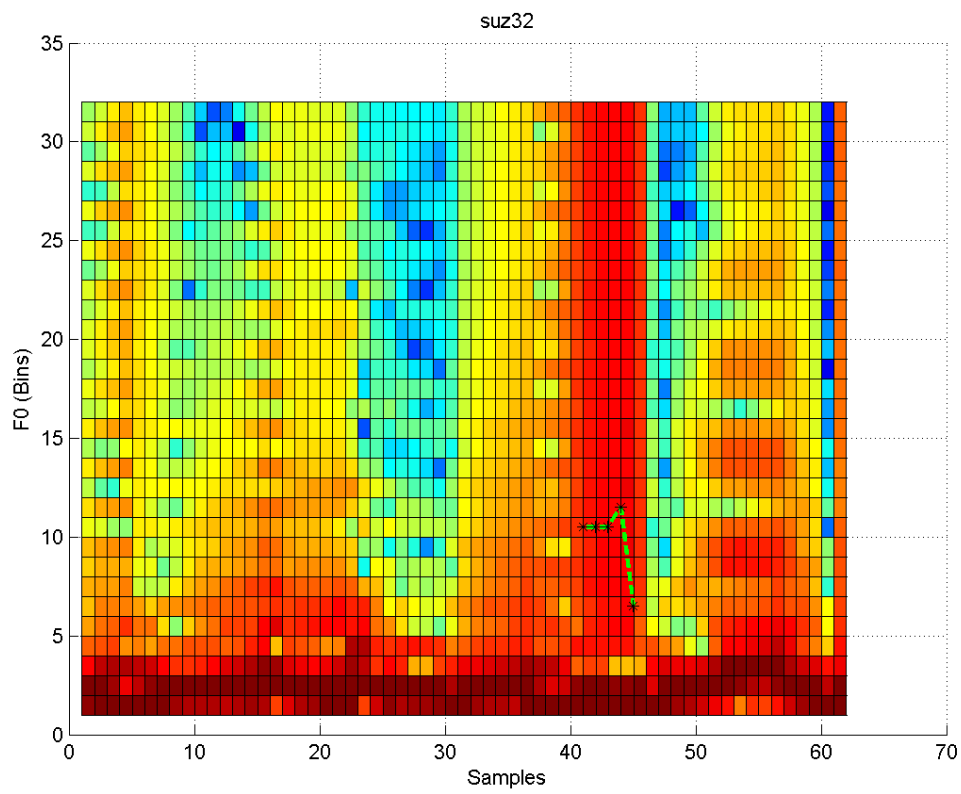


Figura G14 - Visualização do espectro de potências através da ODFT do sinal suz32.

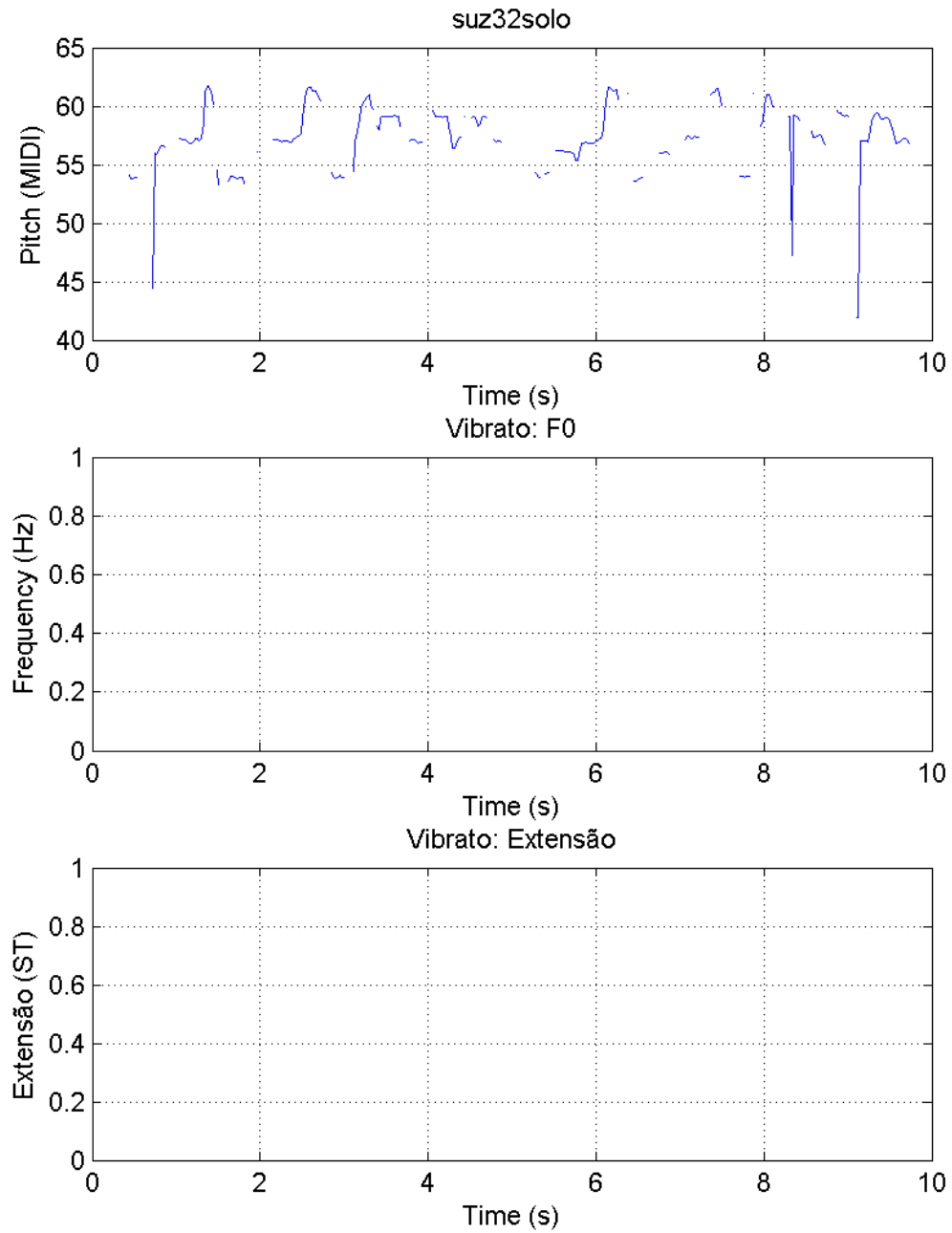


Figura G15 - Visualização da estimação dos parâmetros do vibrato, através do suz32solo.

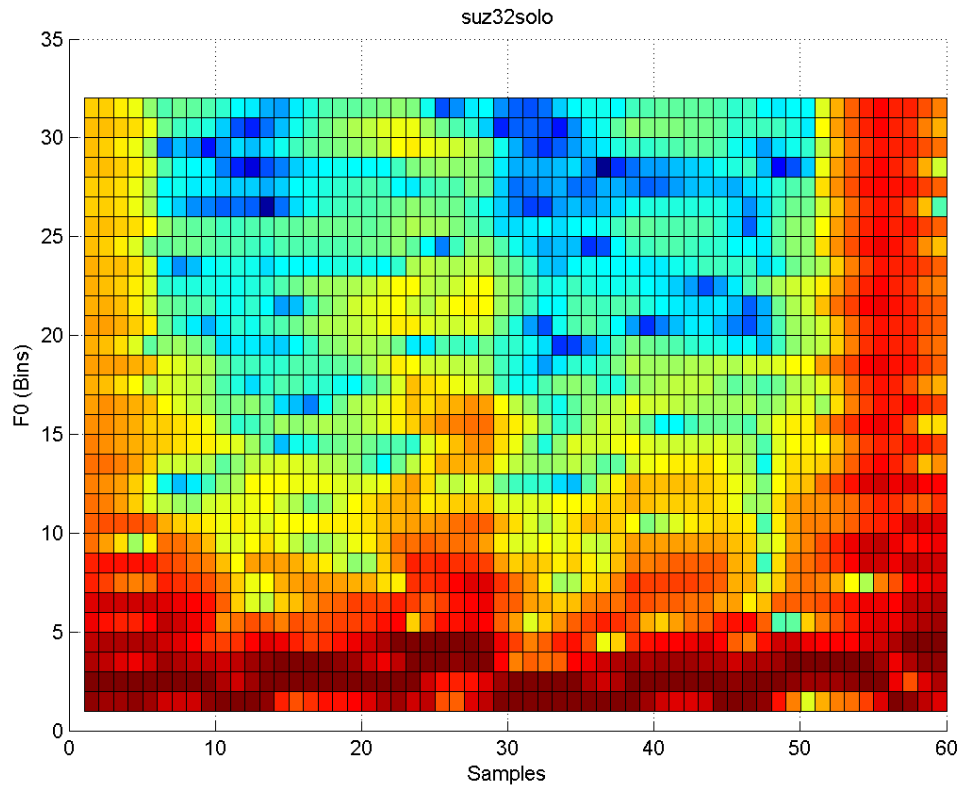


Figura G16 - Visualização do espectro de potências através da ODFT do sinal suz32solo.

## Referências

- [1] SEEgnal *SingingStudio*. Disponível em <http://www.seegnal.pt/products/singingstudio.php>. Acesso em 26/Junho/2011.
- [2] Página da dissertação. Disponível em <http://sites.google.com/site/biofeedback10/>. Acesso em 26/Junho/2011.
- [3] Página da dissertação no siFeup (requer autenticação). Disponível em [https://www.fe.up.pt/si/estagios\\_alunos.entrada?p\\_aluno\\_id=20681](https://www.fe.up.pt/si/estagios_alunos.entrada?p_aluno_id=20681). Acesso em 26/Junho/2011.
- [4] I. Guimarães. “A Ciência e a Arte da Voz Humana”, Escola Superior de Saúde do Alcoitão (ESSA), Alcabideche, 2007
- [5] J. Gauffin & B. Hammarberg, “*Vocal Fold Physiology*”, London, Whurr, 1991
- [6] T. Pereira et. al., “*ABC do corpo humano*”, Selecções do Reader’s Digest, 1989
- [7] Björn E. F. Lindblom and Johan E. F. Sundberg, “*Acoustical Consequences of Lip, Tongue, Jaw, and Larynx Movement, J. Acoust*”. Soc. Am. Volume 50, Issue 4B, 1971
- [8] J. Sundberg, “*The Science of the Singing Voice*”, Dekalb, Northern Illinois University Press, 1987
- [9] Books LLC, “*Human Voice: Vocal Folds, Larynx, Phonation, Voice Analysis, Vocal Loading, Human Pharynx, Glottis, Formant, Voice Frequency, Rever*”, September 2010.
- [10] F. & Y. Rodd-Marling Husler, “*Singing: Physical Nature of the Vocal Organ*”, 2nd Edition, New York, Ebury Press, 1976
- [11] K. Murphy, “*Digital signal processing techniques for application in the analysis of pathological voice and normophonic singing voice*”. Disponível em <http://oa.upm.es/1079/>. Acesso em 26/Junho/2011.

## Referências

- [12] J. W. Cooley & J. W. Tukey, “*An Algorithm for the Machine Calculation of Complex Fourier Series*”, *Math. Comput.* 19, 1965, pp. 297-301.
- [13] P. Alku & T. Backstrom, “*Linear Predictive Method for Improved Spectral Modelling of Lower Frequencies of Speech with Small Prediction Orders*”, *IEEE Tran. Speech Audio Proc.*, Vol. 12, Issue 2, 2004, pp. 93-99.
- [14] I. R. Titze, “*Workshop on Acoustic Voice Analysis*”, Summary Statement, Iowa, 1994.
- [15] R. Fernández-Baillo, “*Aportaciones del Análisis Biomecánico de la Pedagogía Vocal*”, Report, LABAF, Universidad de Alcalá de Henares, 2006.
- [16] I. R. Titze, “*Principles of Voice Production*”, New Jersey, Prentice Hall, 1994.
- [17] I. Arroabarren Alemán, “*Signal Processing Techniques for Singing and Vibrato Modeling*”, PhD Thesis, Universidad Pública de Navarra, 2004.
- [18] A. Loscos, “*Spectral processing of the singing voice*”. Disponível em <http://www.tesisenred.net/TDX-0813107-163951>. Acesso em 26/Junho/2011.
- [19] Jean-Jacques, C. A. Nattiez (Translator), “*Music and discourse: toward a semiology of music*”, Princeton University Press, 1990.
- [20] Sing and See. Disponível em <http://www.singandsee.com/>. Acesso em 26/Junho/2011
- [21] Sing and See no youtube. Disponível em [http://www.youtube.com/watch?v=mm49kFyMafY&feature=player\\_embedded#at=40](http://www.youtube.com/watch?v=mm49kFyMafY&feature=player_embedded#at=40). Acesso em 26/Junho/2011.
- [22] Music MasterWorks. Disponível em <http://www.musicmasterworks.com/>. Acesso em 26/Junho/2011.
- [23] SingStar. Disponível em [http://www.singstar.com/pt\\_PT/singstar.html](http://www.singstar.com/pt_PT/singstar.html). Acesso em 26/Junho/2011.
- [24] Tune in to Reading, a ex empresa carry-a-tune. Disponível em <http://www.tuneintoreading.com/Home>. Acesso em 26/Junho/2011.
- [25] Carry-a-tune. Disponível para compra em <http://www.amazon.com/Carry-A-Tune-Technologies-CAT220-Singing-Coach/dp/B0006NAB8W>. Acesso em 26/Junho/2011.
- [26] Xitona SingingTutor. Disponível em <http://xitona.com/singingtutor.html>. Acesso em 26/Junho/2011.
- [27] Singing Superstar. Disponível em <http://singing-superstar.com/>. Acesso em 26/Junho/2011.

- [28] SETEK Music Tutor, neste momento descontinuado. Disponível em <http://www.sestek.com.tr/eng/index.html>. Acesso em 26/Junho/2011.
- [29] Software Music Tutor. Disponível para compra em <http://music-tutor.sestek.qarchive.org/>. Acesso em 26/Junho/2011.
- [30] Software Madde. Disponível em <http://www.speech.kth.se/music/downloads/smptool/>. Acesso em 26/junho/2011.
- [31] Software Adobe Audition. Disponível em <http://www.adobe.com/products/audition.html>. Acesso em 26/Junho/2011.
- [32] B. Kedem, “*Spectral analysis and discrimination by zero-crossings*”, Proceedings of the IEEE, 74(11):1477-1493, November 1986.
- [33] P. Boersma, “*Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound*”, Proceedings of the Institute of Phonetic Sciences 17: 97-110. University of Amsterdam, 1993.
- [34] A. de Cheveign and H. Kawahara, “*Yin, a fundamental frequency estimator for speech and music*”, Journal of the Acoustical Society of America, 111(4), 2002.
- [35] M. Piszczalski, “*A Computational Model for Music Transcription*”, PhD thesis, University of Stanford, 1986.
- [36] M. Piszczalski and Bernard A. Galler, “*Predicting musical pitch from component frequency ratios*”, Journal of the Acoustical Society of America, 66(3):710-720, September 1979.
- [37] J. E. Lane, “*Pitch detection using a tunable IIR filter*”, Computer Music Journal, 14(3):46-57, Fall 1990.
- [38] B. Kedem, “*Spectral analysis and discrimination by zero-crossings*”, Proceedings of the IEEE, 74(11):1477-1493, November 1986.
- [39] B. P. Bogert, M. J. R. Healy, and J. W. Tukey, “*The Quefreny Alanysis of Time Series for Echoes: Cepstrum, Pseudo Autocovariance, Cross-Cepstrum and Saphe Cracking*”, Proceedings of the Symposium on Time Series Analysis (M. Rosenblatt, Ed) Chapter 15, 209-243. New York: Wiley, 1963.
- [40] E. Geoffriois, “*The multi-lag-window method for robust extended-range  $f_0$  determination*”, in Fourth International Conference on Spoken Language Processing, volume 4, pages 2239-2243, 1996.

## Referências

- [41] A. Ferreira e R. Sousa, "*DFT-based frequency estimation under  $r$  harmonic interference*", in *ISCCSP 2010 - The 4th International Symposium on Communications, Control and Signal Processing 2010*, março de 2010
- [42] Matlab. Disponível em <http://www.mathworks.com/products/matlab/>. Acesso em 26/Junho/2011.