

JOSÉ EDUARDO DE CARVALHO SILVA

**TREMOR E VIBRATO HUMANOS:  
MEDIÇÃO DE TAXA, PROFUNDIDADE E  
REGULARIDADE NO MOVIMENTO DE MEMBROS,  
NA VOZ E NO CONTRABAIXO ACÚSTICO**

Belo Horizonte  
18 de agosto de 2010

UNIVERSIDADE FEDERAL DE MINAS GERAIS  
ESCOLA DE ENGENHARIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**TREMOR E VIBRATO HUMANOS:  
MEDIÇÃO DE TAXA, PROFUNDIDADE E  
REGULARIDADE NO MOVIMENTO DE MEMBROS,  
NA VOZ E NO CONTRABAIXO ACÚSTICO**

Dissertação submetida ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Mestre em Engenharia Elétrica.

Área de Concentração: Engenharia de Computação e Telecomunicações

Linha de Pesquisa: Inteligência Computacional

JOSÉ EDUARDO DE CARVALHO SILVA

Belo Horizonte  
18 de agosto de 2010



UNIVERSIDADE FEDERAL DE MINAS GERAIS

FOLHA DE APROVAÇÃO

Tremor e vibrato humanos:  
medição de taxa, profundidade e regularidade no movimento de  
membros,  
na voz e no contrabaixo acústico

JOSÉ EDUARDO DE CARVALHO SILVA

Dissertação defendida e aprovada pela banca examinadora constituída por:

Prof. Dr. MAURÍLIO NUNES VIEIRA – Orientador  
Departamento de Física / Instituto de Ciências Exatas  
Universidade Federal de Minas Gerais

Prof. Dr. HANI CAMILLE YEHIA – Co-orientador  
Departamento de Engenharia Eletrônica / Escola de Engenharia  
Universidade Federal de Minas Gerais

Prof. Dra. ANA CRISTINA CÔRTEZ GAMA  
Departamento de Fonoaudiologia / Faculdade de Medicina  
Universidade Federal de Minas Gerais

Prof. Dr. CARLOS JÚLIO TIERRA CRIOLLO  
Departamento de Engenharia Elétrica / Escola de Engenharia  
Universidade Federal de Minas Gerais

Prof. Dr. PAULO FERNANDO SEIXAS  
Departamento de Engenharia Eletrônica / Escola de Engenharia  
Universidade Federal de Minas Gerais

Belo Horizonte, 18 de agosto de 2010

# Resumo

Este trabalho investigou o tremor e o vibrato humanos. Foram avaliados algoritmos para a medição da taxa, da profundidade e da regularidade em três tipos de sinal: (i) tremor em membros superiores; (ii) voz falada e cantada e (iii) notas prolongadas no contrabaixo acústico. As análises foram feitas medindo-se flutuações de baixa frequência (4-15 Hz), tanto na série temporal da frequência fundamental ( $F_0$ ) dos sinais quanto nas respectivas envoltórias. Foram investigadas três técnicas para análise de modulações em amplitude e em frequência: sinal analítico, operador de Teager e análise de Fourier. Os dois primeiros apresentaram dificuldades para lidar com sinais multicomponentes, como é o caso da voz e música, enquanto o método de Fourier mostrou-se mais eficiente. Os valores médios obtidos com a análise de Fourier foram: (a) 4,7 Hz (taxa) e 19,83 Hz (profundidade) para tremor em membros; (b) 4,8 Hz (taxa) e 3,98% (profundidade) para tremor em voz falada; (c) 5,7 Hz (taxa) e 3,12% (profundidade) para vibrato em voz cantada; e (d) 5,0 Hz (taxa) e 0,96% (profundidade) para vibrato em contrabaixo. O método de Fourier não se mostrou adequado para rastrear a evolução temporal da taxa e da profundidade das modulações. Para tentar contornar esta limitação, foi proposta uma técnica baseada na função de autocorrelação que mede a similaridade entre ciclos da curva de  $F_0$  ao longo do tempo. Os resultados da aplicação deste método em sinais de tremor/vibrato sugerem que ele possa ser usado para medir a regularidade da frequência fundamental, mas ainda é preciso avaliá-lo melhor em trabalhos futuros.

# Abstract

This work investigated human tremor and vibrato. Algorithms for rate, extent and regularity measurements were evaluated in three signals: (i) tremor in upper limbs; (ii) spoken and sung voice; and (iii) long duration notes in the double bass. The study was carried out by measuring low frequency fluctuations (4-15 Hz) in the fundamental frequency series ( $F_0$ ) and in the signal's envelope. Three techniques for amplitude and frequency modulation analysis were studied: analytic signal, Teager's operator, and Fourier analysis. The first two presented difficulties with multicomponent signals such as voice and music, while the Fourier method was more efficient. The mean values measured with Fourier analysis were (a) 4.7 Hz (rate) and 19.83 Hz (extent) for tremor in limbs; (b) 4.8 Hz (rate) and 3.98% (extent) for tremor in spoken voice; (c) 5.7 Hz (rate) and 3.12% (extent) for vibrato in sung voice; and (d) 5.0 Hz (rate) and 0.96% (extent) for vibrato in double bass. The Fourier method was not suitable to track the modulation rate and extent temporal behavior. To overcome this limitation, a technique was proposed based on the autocorrelation function to measure the similarity between  $F_0$  cycles over time. Results of application of this method to tremor/vibrato signal suggest that it can be used to measure the fundamental frequency regularity, but it still needs to be better evaluated in future works.

# Agradecimentos

Primeiramente, agradeço a Deus por tudo. Ele sempre iluminou meu caminho e me ajudou nas escolhas que fiz durante toda a minha trajetória.

Agradeço também ao meu pai José Eustáquio, à minha mãe Maria da Conceição e ao meu irmão Emílio. Aliás, divido com eles os méritos de mais uma etapa vencida. Sem vocês, não teria conseguido traçar este caminho desde a pré-escola até a pós-graduação. Além deles, agradeço muito aos meus avós Geraldo e Maria do Carmo, Juarez e Luzia. Este trabalho tem muita influência da sabedoria que vocês transmitiram a meus pais e que me foi repassada.

Maurílio e Hani, orientadores deste trabalho, agradeço pela paciência, pelos ensinamentos e pelas horas de dedicação. Um obrigado mais que especial ao Maurílio, quem me motivou a abordar este tema e me mostrou o poder da *simplicidade*. E, também me apresentou diversas direções que foram tão importantes para a minha formação quanto esta dissertação.

Agradeço também aos professores Fausto Borém (Escola de Música/UFMG) e Márcio Moraes (ICB/UFMG) que me ajudaram com valiosas sugestões e explicações. E também aos professores Carlos Júlio e Paulo Seixas, integrantes da banca examinadora, que contribuíram para a melhoria do texto. Além deles, agradeço bastante à professora Ana Cristina Gama (Departamento de Fonoaudiologia/UFMG) que, além de ter fornecido uma base de dados para este trabalho, me ajudou em discussões e esclarecimentos.

Ao pessoal do Cefala, muito obrigado pela companhia ao longo destes anos. Em especial, agradeço ao João Sansão, que, sempre bem humorado, me ajudou em diversos momentos do mestrado.

Por fim, agradeço aos meus colegas e amigos Ana Paula Castro, Camila Cançado, Douglas Pedroso, Henrique Lopes, Hugo Bicalho, Igor Vieira, Jean-Phillipe, João Américo, Marco Aurélio, Maria Cantoni, Sabrina Dornelas, Stefanne Gonçalves, Tiago Zanotelli, Victoria Ponz, Vítor Carneiro.

# Sumário

<b>1</b>	<b>Apresentação</b>	<b>1</b>
1.1	Motivação . . . . .	1
1.2	Objetivos . . . . .	4
1.3	Estrutura do texto . . . . .	5
<b>2</b>	<b>Técnicas de AM e FM na análise de tremor e vibrato</b>	<b>6</b>
2.1	Introdução . . . . .	6
2.2	Representação analítica . . . . .	9
2.2.1	Tempo contínuo . . . . .	9
2.2.2	Tempo discreto . . . . .	11
2.3	Operador de Teager . . . . .	16
2.4	Comparação entre o operador de Teager e o sinal analítico . . . . .	20
2.5	Análise de Fourier . . . . .	20
2.5.1	Medição de $F_0$ . . . . .	21
2.5.2	Análise Espectral . . . . .	23
2.6	Conclusão . . . . .	25
<b>3</b>	<b>Metodologia</b>	<b>26</b>
3.1	Introdução . . . . .	26
3.2	Bases de dados . . . . .	27
3.2.1	Base de dados 1: tremor em membros . . . . .	27
3.2.2	Base de dados 2: vibrato no contrabaixo . . . . .	27
3.2.3	Base de dados 3: perturbação em voz . . . . .	28
3.2.4	Base de dados 4: vibrato no canto . . . . .	28
3.3	Algoritmo . . . . .	28
3.3.1	Extração da fundamental . . . . .	29
3.3.2	Pós-processamento automático . . . . .	30
3.3.3	Inspeção visual da curva de $F_0$ . . . . .	36
3.3.4	Segmentação manual da amostra . . . . .	36
3.3.5	Extração automática da envoltória da amostra . . . . .	37
3.3.6	Análise espectral das curvas de $F_0$ e de envoltória . . . . .	41
3.3.7	Análise automática de regularidade de $F_0$ . . . . .	44

3.4	Calibração . . . . .	51
3.4.1	Envoltória . . . . .	51
3.4.2	Frequência fundamental . . . . .	52
3.4.3	Regularidade . . . . .	53
3.5	Conclusão . . . . .	55
<b>4</b>	<b>Resultados</b>	<b>56</b>
4.1	Introdução . . . . .	56
4.2	Membros . . . . .	56
4.2.1	Exemplo 1: tremor essencial de membros . . . . .	57
4.2.2	Exemplo 2: doença de Parkinson . . . . .	58
4.2.3	Grupo . . . . .	60
4.3	Contrabaixo . . . . .	63
4.3.1	Exemplo 1: mesma nota, diferentes tipos de vibrato . . . . .	63
4.3.2	Exemplo 2: diferentes notas, mesmo tipo de vibrato . . . . .	67
4.3.3	Grupo . . . . .	71
4.4	Tremor Vocal . . . . .	75
4.4.1	Exemplo 1: tratamento de tremor vocal . . . . .	75
4.4.2	Exemplo 2: doença de Parkinson . . . . .	80
4.4.3	Grupo . . . . .	80
4.5	Canto . . . . .	83
4.5.1	Exemplo 1: vibrato no canto lírico . . . . .	83
4.5.2	Exemplo 2: vibrato no canto sertanejo . . . . .	88
4.5.3	Grupo . . . . .	88
<b>5</b>	<b>Conclusão</b>	<b>93</b>
5.1	Sumário . . . . .	93
5.2	Considerações finais . . . . .	94
5.3	Trabalhos Futuros . . . . .	94
	<b>Referências Bibliográficas</b>	<b>96</b>

# Lista de Figuras

1.1	Definição de período, taxa e extensão. . . . .	2
1.2	Modelo para geração de AM no sinal de voz a partir de flutuações de $F_0$ . . . . .	4
2.1	Exemplo de modulação AM. . . . .	7
2.2	Exemplo de modulação FM. . . . .	8
2.3	Exemplo de modulação AM-FM. . . . .	8
2.4	Exemplo de função semelhante ao degrau unitário. . . . .	10
2.5	Função $U_{2\pi}(e^{j\omega})$ . . . . .	12
2.6	Função $U_{2\pi}^0(e^{j\omega})$ . . . . .	13
2.7	Resposta em frequência do filtro usado para gerar o sinal analítico. . . . .	14
2.8	Representação analítica. . . . .	15
2.9	Aplicação do sinal analítico para extração de $F_0$ . . . . .	15
2.10	Gabaritos de filtros. . . . .	16
2.11	Exemplo de funcionamento do algoritmo de Teager. . . . .	17
2.12	DESA-1. . . . .	19
2.13	Aplicação do DESA-1 para extração de $F_0$ . . . . .	19
2.14	Algoritmo SRPD. . . . .	22
2.15	Sinal analítico e operador de Teager aplicados à curva extraída pelo SRPD. . . . .	23
2.16	Análise de Fourier aplicada à curva extraída pelo SRPD. . . . .	24
2.17	Espectrograma de uma amostra de vibrato de longa duração temporal. . . . .	24
2.18	Espectrograma de uma amostra de vibrato de curta duração temporal. . . . .	25
3.1	Tipos de vibrato executados no contrabaixo. . . . .	27
3.2	Descrição geral do método de análise. . . . .	29
3.3	Curva de $F_0$ extraída de vibrato em contrabaixo. . . . .	30
3.4	Estimação de $F_0$ com alguns erros. . . . .	31
3.5	Correção dos erros gerados pelo algoritmo de rastreamento de $F_0$ . . . . .	32
3.6	Distorções provocadas pelo filtro de mediana e suavizador linear. . . . .	33
3.7	Deteção e correção do erro grosseiro. . . . .	33
3.8	Exemplo de erro intermediário. . . . .	34
3.9	Deteção e correção do erro intermediário. . . . .	35
3.10	Exemplo de erro fino. . . . .	35
3.11	Deteção e correção do erro fino. . . . .	36

3.12	Inspeção visual da curva de $F_0$ . . . . .	37
3.13	Segmentação manual. . . . .	37
3.14	Diagrama de blocos da demodulação AM. . . . .	39
3.15	Resposta em frequência do filtro passa-baixas. . . . .	39
3.16	Exemplo de processamento homomórfico. . . . .	40
3.17	Diagrama de blocos da análise de $F_0$ . . . . .	41
3.18	Resposta em frequência do filtro passa-altas. . . . .	41
3.19	Espectro da fundamental de um vibrato produzido por cantor sertanejo. . . . .	42
3.20	Índice de modulação FM do vibrato produzido por cantor sertanejo. . . . .	43
3.21	Envoltória do vibrato produzido por cantor sertanejo. . . . .	44
3.22	Exemplo de séries de $F_0$ regulares e irregulares. . . . .	45
3.23	Exemplos de funções de autocorrelação. . . . .	47
3.24	Diagrama de blocos da análise de regularidade. . . . .	48
3.25	Análise de regularidade – Parte 1. . . . .	49
3.26	Análise de regularidade – Parte 2. . . . .	50
3.27	Sinal usado na calibração da análise de envoltória. . . . .	51
3.28	Resultado da calibração da análise de envoltória. . . . .	52
3.29	Sinal usado na calibração da análise de $F_0$ . . . . .	52
3.30	Resultado da calibração da análise de $F_0$ . . . . .	53
3.31	Calibração da análise de regularidade. . . . .	54
4.1	Série temporal da aceleração da mão de um paciente com tremor essencial. . . . .	57
4.2	Espectro do sinal de aceleração da mão de um paciente com tremor essencial. . . . .	58
4.3	Análise de regularidade do sinal de aceleração da mão de um paciente com tremor essencial. . . . .	58
4.4	Série temporal da aceleração da mão de um paciente que sofre de Parkinson. . . . .	59
4.5	Análise do sinal de aceleração de um paciente com doença de Parkinson. . . . .	59
4.6	Análise de regularidade da série de aceleração da mão de um doente de Parkinson. . . . .	60
4.7	Análise de grupo: base de dados 1. . . . .	60
4.8	Profundidade x taxa: base de dados 1. . . . .	61
4.9	Distribuição da taxa e da profundidade: base de dados 1. . . . .	62
4.10	Comparação entre vibratos tipos I e II no contrabaixo. . . . .	64
4.11	Análise de regularidade da amostra de vibrato tipo I. . . . .	65
4.12	Análise de regularidade da amostra de vibrato tipo II. . . . .	66
4.13	Comparação entre vibratos em duas notas diferentes no contrabaixo. . . . .	68
4.14	Análise de regularidade da amostra de vibrato executado em uma nota grave. . . . .	69
4.15	Análise de regularidade da amostra de vibrato executado em nota aguda. . . . .	70
4.16	Análise de grupo: base da dados 2 classificada pelo registro. . . . .	72
4.17	Profundidade x taxa: base de dados 2 classificada pelo registro. . . . .	73
4.18	Coefficientes de correlação: base de dados 2 classificada pelo registro. . . . .	73

4.19	Distribuição da taxa e da profundidade: base da dados 2 classificada pelo tipo de vibrato. . . . .	74
4.20	Profundidade x taxa: base de dados 2 classificada pelo tipo de vibrato. . . . .	75
4.21	Coefficientes de correlação: base de dados 2 classificada pelo tipo de vibrato. . . . .	76
4.22	Características do tremor antes e depois do tratamento com toxina botulínica. . . . .	77
4.23	Análise de regularidade do tremor vocal antes da medicação. . . . .	78
4.24	Análise de regularidade do tremor vocal após a medicação. . . . .	79
4.25	Características da voz de um paciente com doença de Parkinson. . . . .	80
4.26	Análise de regularidade da voz de um paciente com doença de Parkinson. . . . .	81
4.27	Distribuição da taxa e da profundidade: base da dados 3. . . . .	82
4.28	Profundidade x taxa: base de dados 3 . . . . .	83
4.29	Coefficientes de correlação para a base de dados 3. . . . .	84
4.30	Análise de vibrato no canto lírico. . . . .	85
4.31	Análise de regularidade de vibrato no canto lírico. Amostra 1. . . . .	86
4.32	Análise de regularidade de vibrato no canto lírico. Amostra 2. . . . .	87
4.33	Análise de vibrato no canto sertanejo. . . . .	88
4.34	Espectro do vibrato no canto sertanejo. . . . .	89
4.35	Análise de regularidade de vibrato no canto sertanejo. . . . .	90
4.36	Distribuição da taxa e da profundidade: base da dados 4. . . . .	91
4.37	Profundidade x taxa: base de dados 4 . . . . .	92
4.38	Coefficientes de correlação para a base de dados 4. . . . .	92

# Lista de Tabelas

2.1	DESA-1 e DESA-2 . . . . .	18
4.1	Estatísticas de taxa e profundidade: base de dados 1. . . . .	62
4.2	Estatísticas de taxa e profundidade: base de dados 2 classificada por registro. . .	71
4.3	Estatísticas dos coeficientes de correlação: base de dados 2 classificada por registro.	73
4.4	Estatísticas de taxa e profundidade: base de dados 2 classificada por tipo de vibrato.	75
4.5	Estatísticas de taxa e profundidade: base de dados 3 . . . . .	81
4.6	Estatísticas dos coeficientes de correlação: base de dados 3. . . . .	83
4.7	Estatísticas de taxa e profundidade: base de dados 4 . . . . .	91
4.8	Estatísticas dos coeficientes de correlação: base de dados 4. . . . .	92

# Capítulo 1

## Apresentação

### 1.1 Motivação

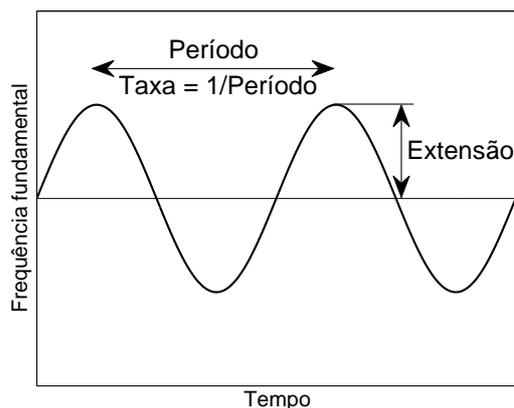
Variações, intencionais ou não, ocorrem na amplitude e/ou frequência de diversos sinais. Isto acontece corriqueiramente na transmissão de informações através da modulação da amplitude ou da frequência da portadora de uma onda de rádio. Instabilidades semelhantes também podem acontecer em movimentos musculares produzidos pelo ser humano durante a fala, canto, ou execução de instrumentos musicais, por exemplo.

A fala resulta da articulação de elementos do aparelho fonatório. A vibração das pregas vocais produz pulsos de ar que se repetem quase periodicamente na chamada frequência fundamental ( $F_0$ ). Dada a forma não senoidal do pulso glótico, o seu espectro é composto por uma série harmônica. Esta série harmônica excita as cavidades do trato vocal que ressoam em determinadas frequências (formantes), sendo finalmente irradiada pelos lábios/narinas. A frequência de vibração média das pregas no homem é de aproximadamente 132 Hz, enquanto, na mulher, de 223 Hz (Peterson e Barney, 1952) e é controlada, principalmente, pela tensão longitudinal das pregas vocais. Perturbações na musculatura laríngea podem causar flutuações em  $F_0$ , como ocorre, por exemplo, no tremor e no vibrato.

O tremor é caracterizado por variações de 1-15 Hz na frequência fundamental e na intensidade da voz causando sensações de flutuação na altura (*pitch*) e no volume sonoro (*loudness*) respectivamente (Titze, 1995; Farinella et al., 2006). Tais flutuações podem ter várias origens, incluindo irregularidades na abdução das pregas vocais, disfunções respiratórias, instabilidade da parede da faringe, tremor da língua, maxilar ou véu palatino (Dromey e Smith, 2008). Geralmente, estão presentes em pacientes que sofrem de distúrbios neurológicos como a doença de Parkinson (Guyton e Hall, 1981).

O vibrato, por sua vez, é um ornamento musical muito explorado por cantores líricos, populares e por instrumentistas. De acordo com Sundberg (1995), há quatro características de interesse, todas relativas à curva de  $F_0$ , também chamada curva melódica (figura 1.1):

- *Taxa* ou *velocidade*: especifica o número de ondulações por segundo.
- *Extensão* ou *profundidade*: amplitude da curva.



**Figura 1.1:** Definição de período, taxa e extensão do vibrato. As mesmas definições são válidas para o tremor.

- *Regularidade:* mede a similaridade entre os ciclos. É considerado um sinal de habilidade, isto é, quanto maior a regularidade, melhor a qualidade do vibrato.
- *Forma da curva:* contorno da curva ao longo do tempo. Normalmente é aceito como senoidal, embora também possa ser triangular ou trapezoidal.

Segundo Seashore (1932, pag. 349), o vibrato é uma pulsação da altura, comumente acompanhada pela pulsação síncrona de volume sonoro e timbre, de tal extensão e taxa que causam uma sensação de riqueza e suavidade aos tons. Normalmente, estas pulsações repetem-se a uma taxa de aproximadamente 5-7 Hz (Sundberg, 1995, pag. 39). No canto ocidental, o vibrato geralmente origina-se de oscilações na tensão das pregas vocais, mas em outras culturas, pode-se encontrar um tipo de vibrato originado pela pulsação da pressão subglótica (Sundberg, 1995).

Os processos de produção do vibrato e do tremor ainda não foram completamente explicados. Titze e colaboradores (Titze et al., 1994; Titze, 1996; Titze et al., 2002) sugerem que o vibrato seja um tremor refinado. Titze et al. (1994) observaram que existe um oscilador periférico envolvendo músculos da região da laringe, mas não concluíram se este oscilador funciona simplesmente como um filtro mecânico, que reduz a variabilidade de frequências injetadas por um oscilador no sistema nervoso central, ou se há algum tipo de realimentação (*reflex loops*) induzindo estas oscilações. O trabalho de Titze (1996) mostrou que o comportamento oscilatório da tensão nas pregas vocais e da frequência fundamental pode ser simulado usando um modelo matemático desta realimentação. Titze et al. (2002) reuniram os principais resultados anteriores e propuseram o modelo de ressonância por reflexo (*reflex resonance*) para tentar explicar o vibrato. De acordo com este modelo, o parâmetro mais influente no vibrato é o ganho do laço de realimentação (*reflex loops*). Um cantor treinado é capaz de controlar o ganho de forma a gerar um vibrato estável. Por outro lado, para o caso do tremor, o autor sugere que o ganho da realimentação seja muito alto e fora do controle do sujeito. Contudo, conforme explicitado pelo próprio autor, este modelo não está completo e não explica, por exemplo, como a taxa do vibrato pode ser alterada através de estímulo

externo.

O trabalho de Leydon et al. (2003) tenta explicar o papel de um estímulo externo sobre o sistema auditivo na produção do vibrato. Um experimento foi realizado com sujeitos sem treinamento em canto e os resultados mostraram que eles produziam vibrato na faixa de aproximadamente 4-7 Hz quando os estímulos apresentados pulsavam nesta mesma faixa de frequência. Entretanto, há trabalhos relatando que em cantores – diferentemente dos não cantores – o vibrato mantém suas características mesmo quando o estímulo auditivo é modificado (Clarkson e Deutsch, 1966; Shipp et al., 1984) ou até mesmo mascarado por ruído (Schultz-Coulon, 1978).

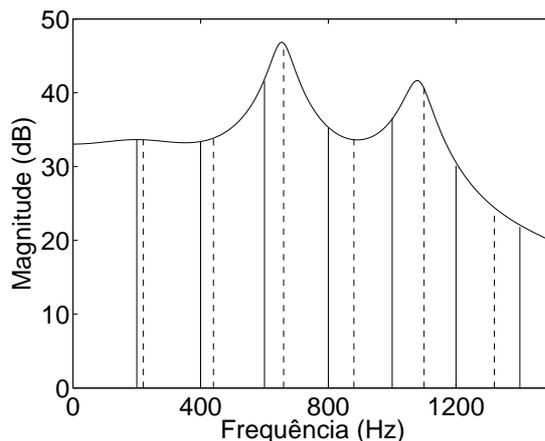
Apesar de ainda não se ter um modelo completo para a produção do tremor e do vibrato, ambos são normalmente definidos como oscilações quase-periódicas e quase-senoidais da frequência fundamental (Hibi e Hirano, 1995; Horii, 1989b; Kreiman et al., 2003). Uma maneira de descrever estas oscilações é através da sua frequência instantânea e da sua envoltória, ou seja, em termos de componentes modulados em frequência (FM) e em amplitude (AM) (Arroabarren et al., 2002). A decomposição de um sinal em partes AM e FM é sabidamente problemática, pois há diversas combinações AM e FM que descrevem o mesmo sinal e o emprego desta representação depende da habilidade de interpretar a informação expressa em seus parâmetros (Arroabarren et al., 2002). No caso da curva de  $F_0$  de um vibrato, ou de um tremor, esta interpretação é simples: as variações na frequência instantânea e envoltória da fundamental são mapeadas, respectivamente, em alterações na taxa e na extensão da curva melódica (figura 1.1).

Na voz, as oscilações em  $F_0$  – e em seus harmônicos – também provocam modulações na amplitude do sinal irradiado. Como mostrado, aparentemente pela primeira vez, por Horii e Hata (1988), grande parte da modulação AM presente no sinal de voz é causada pela interação entre os formantes e o deslocamento em frequência da série harmônica da excitação glotal (figura 1.2). Esta relação de causa e consequência levou Horii (1989a) a questionar o emprego, no canto, das definições de “vibrato”, “vibrato de frequência” e de “vibrato de amplitude” recomendadas pelo *American National Standards Institute* e citadas por Large (1979, pag. 39), segundo a qual, vibrato é

Uma família de efeitos tonais na música que dependem de variações periódicas de uma ou mais características da onda sonora. Nota: Quando características específicas são conhecidas, o termo “vibrato” deve ser modificado de modo a explicitá-las, por exemplo, vibrato de frequência, vibrato de amplitude, vibrato de fase e assim por diante. (Tradução própria)

Segundo Horii (1989a), não há registros de pessoas que emitiram apenas flutuações em amplitude, ou então emitiram AM em um nível muito maior que aquele previsto pela interação entre harmônicos e formantes. Por isso, no canto, não há sentido em separar as definições de “vibrato de frequência” e “vibrato de amplitude”.

O trabalho de Hibi e Hirano (1995) confirmou que as variações da amplitude do sinal de voz são bem explicadas pela interação harmônico-formante e que outros fatores, como vibrações em diversos elementos do trato vocal e variações do espectro do pulso glótico, con-



**Figura 1.2:** Modelo para geração de AM no sinal de voz a partir de flutuações de  $F_0$ . As linhas verticais cheias representam a série harmônica com fundamental em 200 Hz e as linhas verticais tracejadas, a série com fundamental em 220 Hz. Quando a frequência fundamental oscila entre 200 Hz e 220 Hz, a magnitude dos harmônicos da série varia dependendo de sua posição em relação aos formantes. Isto produz flutuações na amplitude do sinal, caracterizando a modulação AM.

tribuem muito pouco para a flutuação da envoltória do sinal. Isto ratifica a justificativa – apresentada por diversos autores (Horii, 1989b; Prame, 1994, 1997; Kreiman et al., 2003; Bretos e Sundberg, 2003; Diaz e Rothman, 2003; de Almeida Bezerra et al., 2009) que analisam apenas a curva de  $F_0$  – de que a informação presente na envoltória do sinal, modulada em amplitude, é redundante.

Nos trabalhos que investigam apenas a curva melódica, a dificuldade de extração da curva de frequência fundamental está sempre presente, apesar de normalmente não ser explicitada. Sabe-se, por exemplo, que não existe algoritmo para extração de frequência fundamental que opere com segurança para voz disfônica. Todavia, a envoltória do sinal, que é extraída de maneira simples (Qi e Weinberg, 1994), parece ter informação suficiente para substituir a curva melódica quando for inviável obtê-la. Neste trabalho, a redundância de informação entre a série de  $F_0$  e a envoltória será estudada com o objetivo de investigar a hipótese de que a envoltória da emissão pode ser usada em lugar da curva de frequência fundamental. Além disso, objetiva-se analisar a regularidade e criar um índice para quantificá-la. Como observado por Sundberg (1995, pag. 39), não há nada que justifique que a regularidade tenha menor relevância perceptiva que a taxa e a extensão. Ainda assim, apenas estes dois últimos parâmetros têm sido extensivamente analisados na literatura.

## 1.2 Objetivos

- Investigar a correlação entre a curva de envoltória e a curva de frequência fundamental da emissão.
- Analisar a regularidade e criar um índice para quantificá-la.

### 1.3 Estrutura do texto

Esta dissertação divide-se em 4 capítulos. O capítulo 2 apresenta três métodos para análise da envoltória e da frequência instantânea de sinais: representação analítica, operador de Teager e análise de Fourier. Discutem-se as vantagens e desvantagens de cada um deles e mostram-se as razões pelas quais o método de Fourier foi o escolhido.

O capítulo 3 detalha as etapas do algoritmo proposto. Neste capítulo também estão descritas as bases de dados usadas no trabalho: (i) tremor em membros, (ii) vibrato no contrabaixo acústico, (iii) voz falada e cantada. Finalmente, no capítulo 4 são discutidos os resultados da aplicação do algoritmo implementado nos sinais de cada uma das bases de dados.

## Capítulo 2

# Técnicas de AM e FM na análise de tremor e vibrato

### 2.1 Introdução

Modulação é o processo de variar um ou mais parâmetros de uma onda, chamada portadora, segundo um sinal modulante, cuja frequência, normalmente, é pelo menos 10 vezes menor que a da portadora (Winholtz e Ramig, 1992). Se, por exemplo, apenas a amplitude da portadora for variada, tem-se a modulação em amplitude (AM); se for apenas a frequência, tem-se a modulação em frequência (FM); se ambos, tem-se a modulação em AM-FM.

A figura 2.1 é um exemplo de modulação em amplitude:

$$x_{am}(t) = [1 + k \cos(2\pi f_m t)] \cos(2\pi f_c t + \theta), \quad (2.1)$$

onde a portadora tem frequência  $f_c$ , amplitude unitária e uma fase arbitrária  $\theta$ ; a modulante tem frequência  $f_m$  e amplitude também unitária; e o coeficiente  $k$  é o índice de modulação AM, calculado pela relação entre os níveis máximo ( $A_{max}$ ) mínimo ( $A_{min}$ ) da envoltória (figura 2.1):

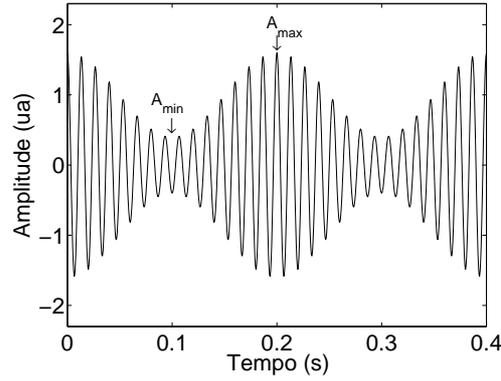
$$k = \frac{A_{max} - A_{min}}{A_{max} + A_{min}}. \quad (2.2)$$

Este índice de modulação,  $k$ , indica o quanto a amplitude da portadora modulada variou em relação ao seu valor nominal, não modulado.

Na modulação FM, a frequência da portadora é o parâmetro que oscila de acordo com a modulante, enquanto a amplitude é mantida constante. Um exemplo da modulação em frequência está ilustrado na figura 2.2 com a mesma portadora e a mesma modulante da figura 2.1. Neste caso, a frequência instantânea  $f_i(t)$  é dada por:

$$f_i(t) = f_c [1 + m \cos(2\pi f_m t)], \quad (2.3)$$

em que, quanto maior o valor do coeficiente  $m$ , maior a extensão de  $f_i(t)$ , a qual é limitada por  $f_c \pm m f_c$ . O produto  $m f_c$  quantifica o máximo desvio causado pela modulante, de frequência



**Figura 2.1:** Exemplo de modulação AM. A frequência da portadora é de 75 Hz e a da modulante, de 5 Hz. O índice de modulação AM vale  $k = 0,6$ .

$f_m$ , sobre a frequência instantânea. O índice de modulação FM,  $\beta$ , é definido por:

$$\beta = \frac{\text{máximo desvio}}{f_m} = \frac{m f_c}{f_m} = \frac{\Delta f_c}{f_m}. \quad (2.4)$$

O sinal modulado em frequência pode ser escrito como:

$$x_{fm}(t) = \cos[2\pi \alpha(t)], \quad (2.5)$$

onde  $\alpha(t) = \int_0^t f_i(\tau) d\tau$ . Substituindo a equação 2.3 em  $\alpha(t)$ , vem:

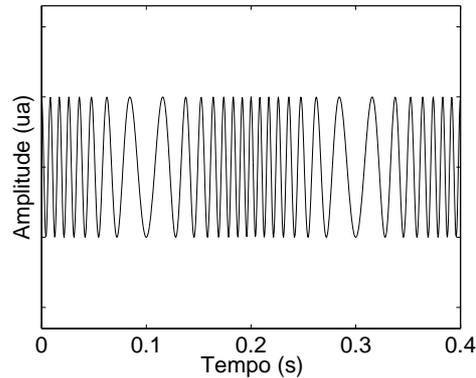
$$\begin{aligned} x_{fm}(t) &= \cos \left[ \int_0^t 2\pi f_i(\tau) d\tau \right], \\ x_{fm}(t) &= \cos \left[ \int_0^t 2\pi f_c d\tau + \int_0^t 2\pi f_c m \cos(2\pi f_m \tau) d\tau \right], \\ x_{fm}(t) &= \cos \left[ 2\pi f_c t + 2\pi \Delta f_c \int_0^t \cos(2\pi f_m \tau) d\tau \right]. \end{aligned} \quad (2.6)$$

O sinal  $x_{fm}(t)$  não carrega nenhuma informação útil na sua amplitude e o sinal  $x_{am}(t)$  não codifica nenhuma mensagem em sua frequência instantânea. Para enviar informação tanto na amplitude quanto na frequência de uma mesma portadora, utiliza-se a modulação AM-FM.

A modulação em AM-FM é exemplificada na figura 2.3. As equações 2.7a e 2.7b descrevem, respectivamente, sinais de tempo contínuo e de tempo discreto modulados desta forma (Potamianos e Maragos, 1994):

$$x(t) = \underbrace{a(t)}_{\text{Parte AM}} \cos \left[ \underbrace{\phi(t)}_{\text{Parte FM}} \right], \quad (2.7a)$$

$$x(n) = \underbrace{a(n)}_{\text{Parte AM}} \cos \left[ \underbrace{\phi(n)}_{\text{Parte FM}} \right]. \quad (2.7b)$$



**Figura 2.2:** Modulação FM. A portadora tem frequência 75 Hz e a modulante, 5 Hz. O índice de modulação FM é  $\beta = 6$ .

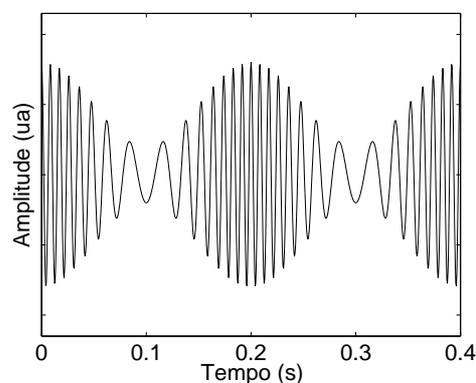
A envoltória do sinal, modulada em amplitude, é dada por  $|a(t)|$ , no caso contínuo, e por  $|a(n)|$ , no discreto. A frequência instantânea (Leon, 1995, cap. 2), modulada em FM, é definida, respectivamente, para os casos contínuo e discreto como:

$$\omega_i(t) = 2\pi f_i(t) \triangleq \frac{d\phi(t)}{dt}, \quad (2.8a)$$

$$\omega_i(n) = 2\pi f_i(n) \triangleq \frac{\Delta\phi(n)}{\Delta n}. \quad (2.8b)$$

onde  $\Delta$  significa uma diferença finita. Note que o símbolo  $\omega_i$  é usado para indicar a frequência instantânea tanto no tempo contínuo quanto no discreto. Quando isto causar alguma dificuldade de compreensão, será explicitada a dependência por  $t$  ou por  $n$ .

A modulação AM-FM, comumente usada em telecomunicações, também tem sido aplicada na análise de voz. Ela foi usada para analisar o tremor vocal (Winholtz e Ramig, 1992), descrever as ressonâncias do trato vocal (Maragos et al., 1993a; Potamianos e Maragos, 1996) e para estudar o vibrato no canto lírico (Arroabarren et al., 2002). Nestes trabalhos, várias técnicas foram usadas para extrair a envoltória do sinal – parte AM – e sua frequência



**Figura 2.3:** Modulação AM-FM. A portadora da modulação em amplitude e em frequência é de 75 Hz e a modulante vale 5 Hz para os dois casos. O índice de modulação AM foi  $k = 0,6$  e o FM,  $\beta = 6$ .

instantânea – parte FM.

Na literatura relacionada, apenas Winholtz e Ramig (1992) usaram circuitos analógicos para medir, de modo automático, os níveis de modulação AM/FM em vozes com tremor. Os outros autores que fizeram medidas automatizadas, usaram a transformada discreta de Fourier (Yair e Gath, 1988; Arroabarren et al., 2002), o sinal analítico (Arroabarren et al., 2002) ou o chamado operador de Teager (Maragos et al., 1993b,a; Potamianos e Maragos, 1996). Além disso, uma grande quantidade de trabalhos foi feita apenas com medidas manuais de taxa e profundidade do vibrato (Horii, 1989b; Prame, 1994, 1997; Bretos e Sundberg, 2003; Howes et al., 2004; de Almeida Bezerra et al., 2009).

O restante deste capítulo apresenta brevemente como o sinal analítico e o operador de Teager estimam a envoltória e a frequência instantânea de um sinal AM-FM e enfatiza o funcionamento da análise de Fourier, que mostrou-se mais adequados aos propósitos do trabalho. As seções são ilustradas com resultados obtidos pelas simulações do autor.

## 2.2 Representação analítica

### 2.2.1 Tempo contínuo

O espectro  $X(j\omega)$  de um sinal real contínuo  $x(t)$  apresenta simetria Hermitiana, ou seja, o espectro de magnitude é par e o de fase é ímpar:

$$\mathcal{F}\{x(t)\} = X(j\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt = \underbrace{\int_{-\infty}^{\infty} x(t) \cos(\omega t) dt}_{\text{par}} - j \underbrace{\int_{-\infty}^{\infty} x(t) \sin(\omega t) dt}_{\text{ímpar}}, \quad (2.9a)$$

$$|X(j\omega)| = |X(-j\omega)|, \quad (2.9b)$$

$$\angle X(j\omega) = -\angle X(-j\omega), \quad (2.9c)$$

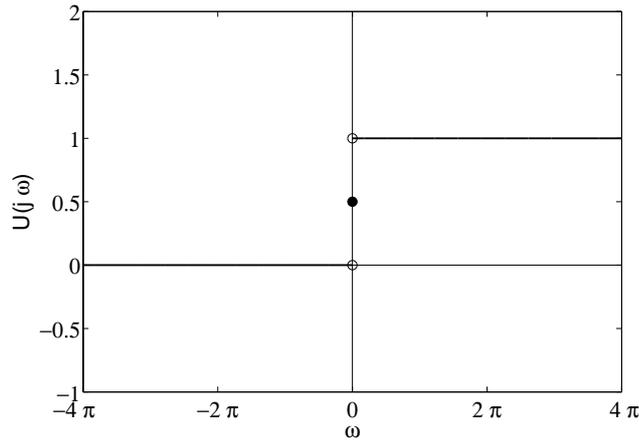
em que  $\mathcal{F}$  denota a transformada de Fourier,  $j = \sqrt{-1}$  é a unidade imaginária,  $|X(j\omega)|$  e  $\angle X(j\omega)$  são, respectivamente, o espectro de magnitude e de fase de  $X(j\omega)$ .

A redundância existente em  $X(j\omega)$  pode ser eliminada sem perda de informação, uma vez que as propriedades de simetria 2.9b e 2.9c permitem recuperar a parte descartada. Por exemplo, pode-se preservar o sinal  $x(t)$  sem nenhuma distorção eliminando-se o espectro da região  $\omega < 0$ , definindo-se:

$$X_+(j\omega) = \begin{cases} 2X(j\omega), & \omega > 0, \\ X(j\omega), & \omega = 0, \\ 0, & \omega < 0. \end{cases} \quad (2.10a)$$

$$= 2U(j\omega) \cdot X(j\omega), \quad (2.10b)$$

onde  $U(j\omega)$  é uma função semelhante ao degrau unitário no domínio da frequência, mostrada na figura 2.4. O fator 2 na equação 2.10 compensa a perda de energia pela eliminação de  $X(j\omega)|_{\omega < 0}$ . Em contrapartida, torna-se necessário manipular um sinal complexo no domínio



**Figura 2.4:** Exemplo de uma função semelhante degrau unitário:  $U(j\omega) = 1$  para  $\omega > 0$ ,  $U(0) = 1/2$  e  $U(j\omega) = 0$  para  $\omega < 0$ .

do tempo correspondente a  $X_+(j\omega)$ :

$$\begin{aligned}
 \mathcal{F}^{-1}\{X_+(j\omega)\} &= \mathcal{F}^{-1}\{2U(j\omega) \cdot X(j\omega)\} \\
 \mathcal{F}^{-1}\{X_+(j\omega)\} &= \mathcal{F}^{-1}\{2U(j\omega)\} * \mathcal{F}^{-1}\{X(j\omega)\} \\
 x_+(t) &= \left[ \delta(t) + j\frac{1}{\pi t} \right] * x(t) \\
 x_+(t) &= x(t) + j\frac{1}{\pi t} * x(t) \\
 x_+(t) &= x(t) + j\hat{x}(t).
 \end{aligned} \tag{2.11}$$

onde o símbolo  $*$  indica a operação de convolução,  $\delta(t)$  é a função Delta de Dirac (Papoulis, 1988, pag. 96) e, como demonstrado em (Papoulis, 1988, pag. 66),

$$\delta(t) + j\frac{1}{\pi t} \xleftrightarrow{\mathcal{F}} 2U(j\omega). \tag{2.12}$$

O sinal complexo  $x_+(t)$ , gerado ao eliminar o espectro da região  $\omega < 0$  do sinal  $x(t)$ , é chamado de representação ou sinal analítico de  $x(t)$ . Note que o sinal  $x(t)$  pode ser recuperado de sua representação analítica simplesmente descartando-se a parte imaginária.

A parte imaginária do sinal analítico,  $\hat{x}(t)$ , também chamada de parte em quadratura, é obtida pela convolução entre  $x(t)$  e o transformador de Hilbert, cuja resposta ao impulso é  $h_{HT} = 1/(\pi t)$  (Oppenheim et al., 1998, cap. 11):

$$\hat{x}(t) = h_{HT}(t) * x(t) = \frac{1}{\pi t} * x(t), \tag{2.13a}$$

$$\hat{X}(j\omega) = \mathcal{F}\{\hat{x}(t)\} = -j \operatorname{sgn}(\omega) X(j\omega), \tag{2.13b}$$

onde a função  $\text{sgn}(\omega)$  é definida como:

$$\text{sgn}(\omega) = \begin{cases} 1, & \omega > 0, \\ 0, & \omega = 0, \\ -1, & \omega < 0. \end{cases} \quad (2.14)$$

É importante enfatizar que  $x_+(t)$  é uma *representação* de  $x(t)$ , ou seja, uma forma diferente de expressar a *mesma* informação contida em  $x(t)$ . Esta representação de  $x(t)$  fornece uma maneira de definir a sua envoltória ( $|a(t)|$ ) e a sua frequência instantânea ( $\omega_i$ ):

$$x_+(t) = x(t) + j\hat{x}(t) = a(t) e^{j\phi(t)}, \quad (2.15a)$$

$$|a(t)| = \sqrt{x^2(t) + \hat{x}^2(t)}, \quad (2.15b)$$

$$\omega_i(t) = \frac{d\phi(t)}{dt} = \frac{d \left[ \arctan \left( \frac{\hat{x}(t)}{x(t)} \right) \right]}{dt}. \quad (2.15c)$$

### 2.2.2 Tempo discreto

No tempo discreto, o sinal analítico é definido de maneira análoga ao caso contínuo. Como detalhado em Oppenheim et al. (1998, seção 11.4), o conceito matemático de analiticidade, a rigor, não se aplica a séries temporais. Mas, pode ser estendido a sinais *discretos complexos* cujas componentes reais e imaginárias se relacionam através de uma convolução discreta similar às relações da transformada de Hilbert. Porém, algumas diferenças merecem destaque:

- O sinal real discreto  $x(n)$ , obtido a partir da amostragem de um sinal real contínuo  $x(t)$  a uma taxa  $f_s$ , tem espectro  $X(e^{j\omega})$  dado por:

$$\mathcal{F}\{x(n)\} = X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x(n)e^{-j\omega n}, \quad (2.16)$$

onde  $\omega = 2\pi \frac{f}{f_s}$  é a frequência normalizada. Além disso, dado que  $x(n)$  é real, a parte do espectro de  $X(e^{j\omega})$  entre  $-\pi < \omega < 0$  contém, a menos de uma defasagem, a mesma informação da região entre  $0 < \omega < \pi$ . Esta redundância é eliminada pela representação analítica.

- O espectro  $X_+(e^{j\omega})$  da representação analítica  $x_+(n)$  de um sinal discreto também é periódico com período  $2\pi$  e é obtido eliminando-se a região entre  $-\pi < \omega < 0$ :

$$X_+(e^{j\omega}) = \begin{cases} 2X(e^{j\omega}), & 0 < \omega < \pi, \\ X(e^{j\omega}), & \omega = 0, \\ 0, & -\pi \leq \omega < 0. \end{cases} \quad (2.17a)$$

$$= 2U_{2\pi}(e^{j\omega}) \cdot X(e^{j\omega}), \quad (2.17b)$$

onde  $U_{2\pi}(e^{j\omega})$  é a função periódica com período igual a  $2\pi$  da figura 2.5.

- A transformada inversa de Fourier de  $U_{2\pi}(e^{j\omega})$  relaciona-se com o espectro da figura

2.6 por uma defasagem:

$$U_{2\pi}(e^{j\omega}) = U_{2\pi}^0(e^{j(\omega-\pi/2)}). \quad (2.18)$$

sendo (Oppenheim et al., 1998, pag. 62):

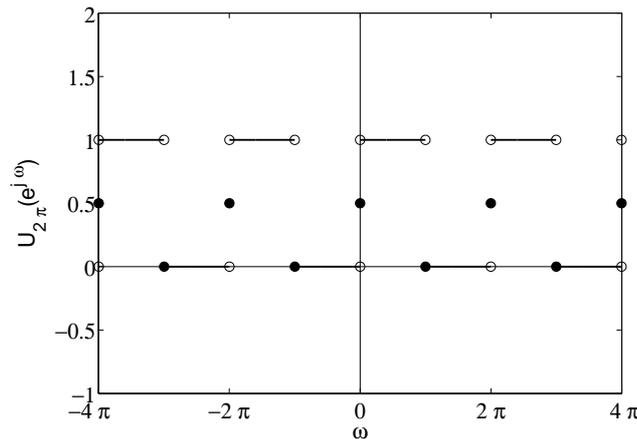
$$u_{2\pi}^0(n) = \frac{\text{sen}(\pi/2 n)}{\pi n}. \quad (2.19)$$

Logo,

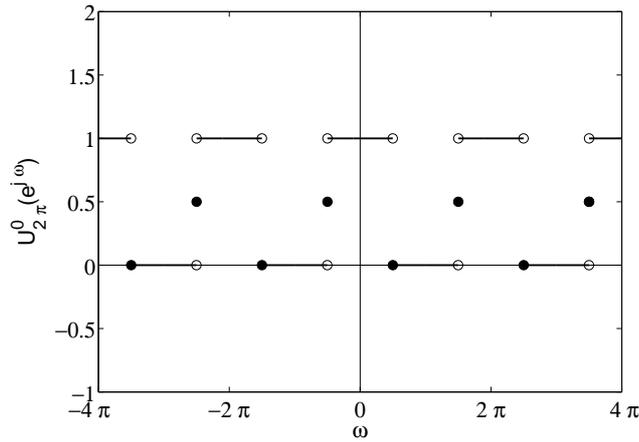
$$\begin{aligned} u_{2\pi}(n) &= \mathcal{F}^{-1}\{U_{2\pi}(e^{j\omega})\} = \mathcal{F}^{-1}\{U_{2\pi}^0(e^{j(\omega-\pi/2)})\} = u_{2\pi}^0(n) \cdot e^{j\pi/2n} \\ &= \frac{\text{sen}(\pi/2 n)}{\pi n} \cdot e^{j\pi/2n} \\ &= \frac{\text{sen}(\pi/2 n)}{\pi n} [\cos(\pi/2n) + j \text{sen}(\pi/2n)] \\ &= \frac{\text{sen}(\pi/2 n) \cdot \cos(\pi/2 n)}{\pi n} + j \frac{\text{sen}^2(\pi/2 n)}{\pi n} \\ &= \frac{1}{2} \frac{\text{sen}(\pi n)}{\pi n} + j \frac{\text{sen}^2(\pi/2 n)}{\pi n} \\ &= \frac{1}{2} \delta(n) + j \frac{\text{sen}^2(\pi/2 n)}{\pi n}. \end{aligned} \quad (2.20)$$

onde foi usado o fato de que  $\frac{\text{sen}(\pi n)}{\pi n} = 0$  para todo  $n \neq 0$  para definir a sequência

$$\delta(n) = \begin{cases} 1, & n = 0, \\ 0, & n \neq 0. \end{cases} \quad (2.21)$$



**Figura 2.5:** Função  $U_{2\pi}(e^{j\omega})$  usada para obter o espectro da representação analítica de um sinal discreto. Devido à periodicidade do espectro do sinal discreto,  $U_{2\pi}(e^{j\omega})$  também é periódica e tem o mesmo período do espectro. Compare as diferenças com a figura 2.4.



**Figura 2.6:**  $U_{2\pi}^0(e^{j\omega}) = 1$  para  $-\pi/2 \leq \omega < \pi/2$  e  $U_{2\pi}^0(e^{j\omega}) = 0$  no restante do período entre  $-\pi$  e  $\pi$ .

- O sinal analítico discreto  $x_+(n)$  será, a partir da equação 2.17b:

$$\begin{aligned}
 \mathcal{F}^{-1}\{X_+(e^{j\omega})\} &= \mathcal{F}^{-1}\{2u_{2\pi}(e^{j\omega}) \cdot X(e^{j\omega})\} \\
 \mathcal{F}^{-1}\{X_+(e^{j\omega})\} &= \mathcal{F}^{-1}\{2u_{2\pi}(e^{j\omega})\} * \mathcal{F}^{-1}\{X(e^{j\omega})\} \\
 x_+(n) &= \left[ \delta(n) + j \frac{2\text{sen}^2(n\pi/2)}{n\pi} \right] * x(n) \\
 x_+(n) &= x(n) + j \frac{2\text{sen}^2(n\pi/2)}{n\pi} * x(n) \\
 x_+(n) &= x(n) + j\hat{x}(n).
 \end{aligned} \tag{2.22}$$

e a parte imaginária  $\hat{x}(n)$  e seu espectro  $\hat{X}(e^{j\omega})$  podem ser escritos como:

$$\hat{x}(n) = x(n) * h_{HT}(n), \tag{2.23a}$$

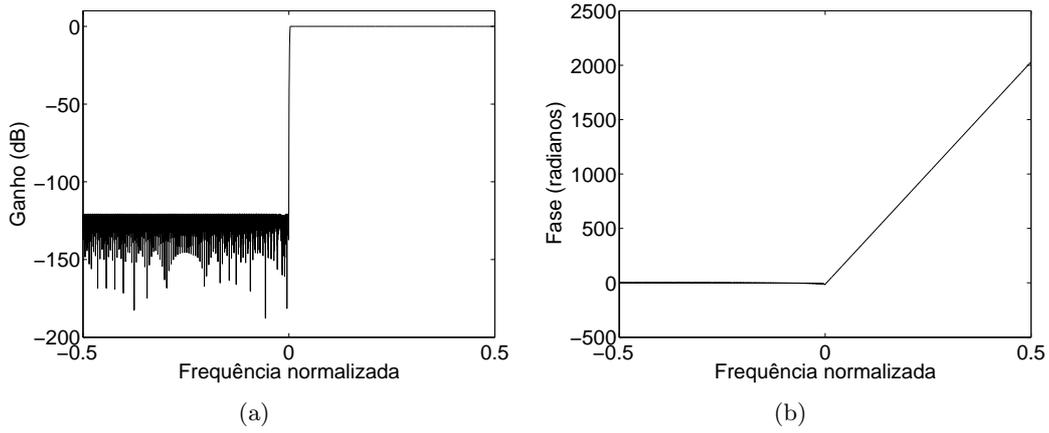
$$h_{HT}(n) = \begin{cases} \frac{2\text{sen}^2(\pi n/2)}{\pi n}, & n \neq 0, \\ 0, & n = 0. \end{cases} \tag{2.23b}$$

$$\begin{array}{c} \uparrow \\ \mathcal{F} \\ \downarrow \end{array}$$

$$\hat{X}(e^{j\omega}) = X(e^{j\omega}) \cdot H_{HT}(e^{j\omega}), \tag{2.23c}$$

$$H_{HT}(e^{j\omega}) = \begin{cases} -j, & 0 \leq \omega < \pi, \\ j, & -\pi \leq \omega < 0. \end{cases} \tag{2.23d}$$

- O filtro FIR (*Finite Impulse Response*) usado para estimar o sinal analítico foi implementado de acordo com o método proposto por Reilly et al. (1994) e está mostrado na figura 2.7.



**Figura 2.7:** O filtro FIR usado para estimar o sinal analítico tem 2604 coeficientes. A sua resposta em frequência tem *ripple* máximo de 0,1% entre  $f = 0,004$  e  $f = 0,496$ . A atenuação na banda de rejeição,  $-0,5 \leq f \leq 0$ , é de pelo menos 120 dB.

• De forma semelhante ao tempo contínuo (equação 2.15), a envoltória  $|a(n)|$  e a frequência instantânea  $\omega_i(n)$  são obtidas escrevendo-se  $x_+(n)$  na forma polar:

$$x_+(n) = a(n) e^{j\phi(n)}, \quad (2.24a)$$

$$|a(n)| = \sqrt{x^2(n) + \hat{x}^2(n)}, \quad (2.24b)$$

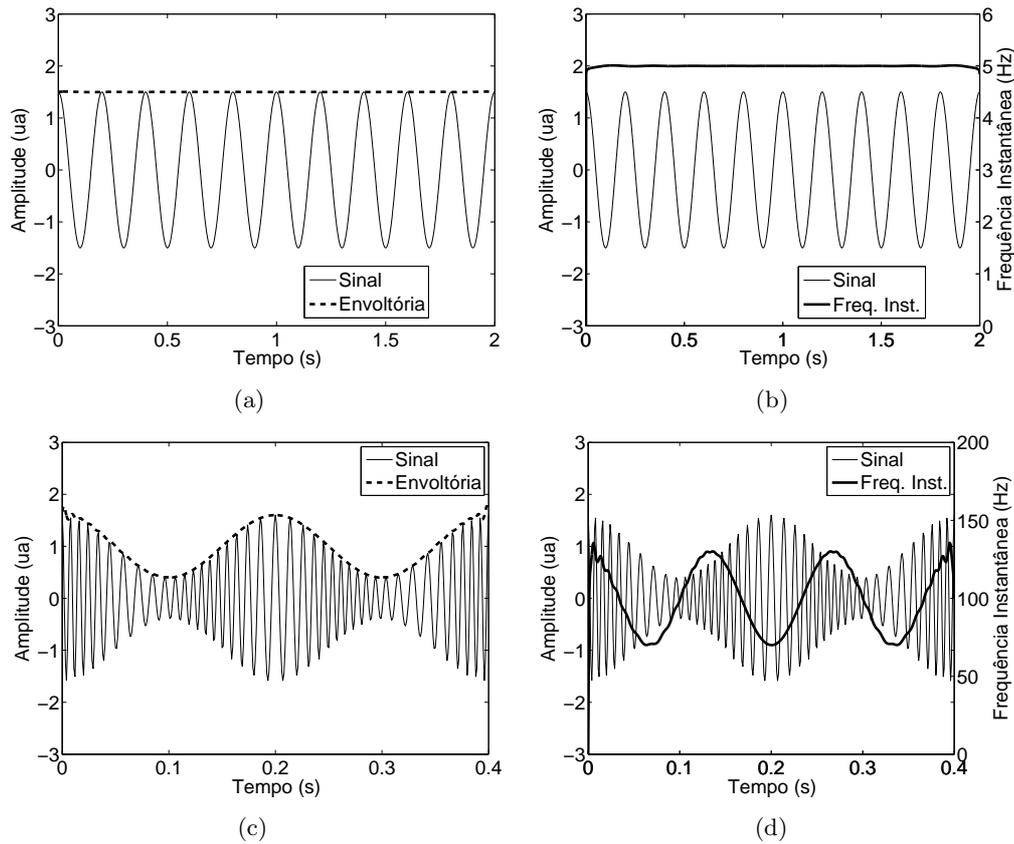
$$\omega_i(n) = \frac{\Delta\phi(n)}{\Delta n} = \frac{\Delta \left[ \arctan \left( \frac{\hat{x}(n)}{x(n)} \right) \right]}{\Delta n}. \quad (2.24c)$$

onde  $\Delta$  indica uma diferença finita.

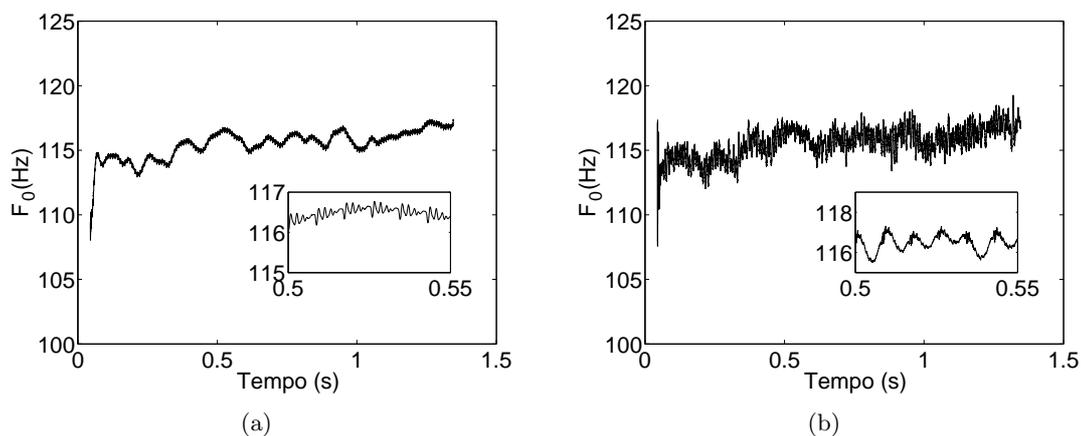
As figuras 2.8(a) e 2.8(b) mostram a envoltória e a frequência instantânea extraída de um sinal senoidal  $x(n) = A \cos(\omega_c n)$  com  $A = 1,5$ . O sinal analítico é  $x_+(n) = A e^{j\omega_c n}$  e as estimativas de  $|a(n)| = A$  e de  $\omega_i = \omega_c$  são exatas. Em 2.8(c) e 2.8(d), mostram-se as estimativas da envoltória e da frequência instantânea de um sinal AM-FM. Neste caso, o sinal analítico fornece estimativas aproximadas. As distorções envolvidas se devem ao fato de o espectro do sinal AM-FM não ser limitado em frequência (Nuttall, 1991). Entretanto, neste exemplo de AM-FM, a maior parte da energia está concentrada em torno da portadora e por isso, há pouca interferência perturbando as estimativas.

O sinal analítico pode ser aplicado para analisar o sinal de voz. Porém, como este sinal é multicomponente (Cohen, 1992), é necessário filtrá-lo para isolar uma região de banda estreita. Mas a escolha dos parâmetros do filtro não é trivial (Potamianos e Maragos, 1994) e influenciam as estimativas produzidas, como mostram as figuras 2.9(a) e 2.9(b). Em todos os dois casos, o sinal analítico foi usado para estimar o valor instantâneo da frequência fundamental de uma voz real, sendo que, para gerar a figura 2.9(a) foi usado o filtro passa-faixa exibido em 2.10(a) e, para gerar 2.9(b) foi usado o filtro passa-baixa da figura 2.10(b).

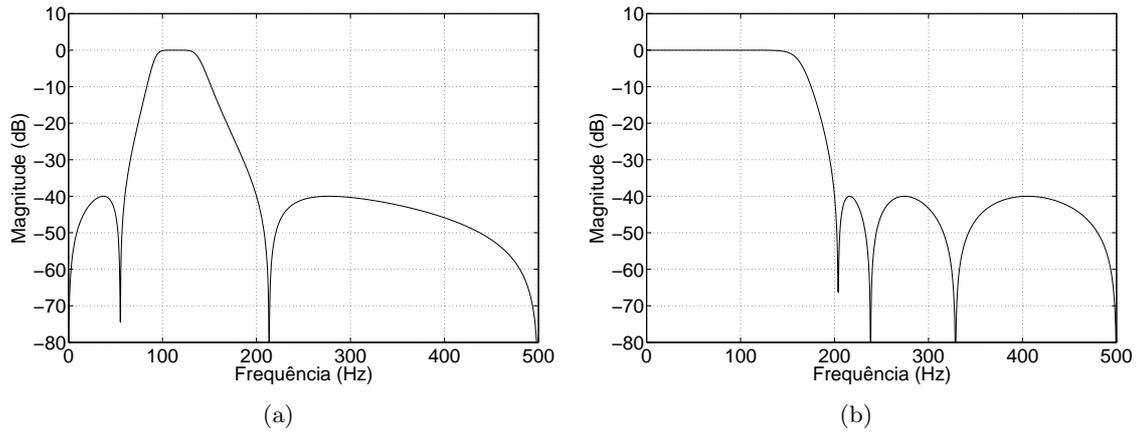
Nestes dois casos a frequência fundamental foi isolada de todos os outros harmônicos. Entretanto, a diferença na largura de banda filtrada gerou resultados diferentes. A estimativa da



**Figura 2.8:** Extração da envoltória e da frequência instantânea usando o sinal analítico. Tanto em (a) quanto em (b), foi usado o sinal  $x(n) = 1,5 \cos(2\pi \frac{5n}{1000})$ . Em (c) e (d), o sinal foi  $x(n) = [1 + 0,6 \cos(2\pi \frac{5n}{1000})] \cos[2\pi \frac{100n}{1000} + 4\pi \frac{7,5n}{1000}]$ .



**Figura 2.9:** Extração de  $F_0$  utilizando o sinal analítico. O mesmo sinal de voz real foi utilizado nas figuras, sendo que, em (a), o sinal foi filtrado com um passa-faixa centrado em torno da frequência fundamental,  $\approx 115$  Hz, e em (b) foi usado um passa-baixa com frequência de corte um pouco acima do valor de  $F_0$ . Os detalhes mostram pormenores das curvas. Os parâmetros dos filtros estão mostrados na figura 2.10.



**Figura 2.10:** (a) Passa-faixa Chebyshev tipo II de terceira ordem com 3 dB de atenuação máxima na banda entre 90 e 130 Hz e, pelo menos 40 dB na faixa de frequências menores que 60 Hz e maiores que 200 Hz. (b) Passa-baixa Chebyshev tipo II de sétima ordem com atenuação máxima de 3 dB na banda até 130 Hz e com pelo menos 40 dB de atenuação a partir de 200 Hz.

frequência instantânea gerada através do sinal analítico é composta pela informação de *todos* os componentes da banda analisada. Por isso, a etapa de filtragem é crítica para garantir que a estimativa gerada não esteja corrompida. Como a frequência fundamental varia ao longo do tempo, é necessária a adaptação do filtro, criando-se dificuldades de implementação (e. g., estabilidade de filtros adaptativos). Um outro método para estimar os componentes AM e FM foi avaliado, como descrito a seguir.

### 2.3 Operador de Teager

O trabalho de Kaiser (1990a) deduz um algoritmo para estimar a energia de um sinal  $x(t)$ , submetido a uma oscilação harmônica, a partir de suas amostras  $x(n)$ , como detalhado a seguir.

Considerando-se um oscilador massa-mola sem atrito, a posição da massa  $m$  ao longo do tempo  $t$  é dada por:

$$x(t) = A \cos(\omega t + \theta), \quad (2.25)$$

em que  $A$  é a amplitude da oscilação,  $\omega$  é a frequência angular de oscilação e  $\theta$  é uma fase inicial arbitrária. A energia total  $E$  deste sistema é a soma da energia potencial da mola e a energia cinética da massa. Pode-se mostrar (Kinsler et al., 1999) que ela vale  $E = \frac{1}{2}m(\omega A)^2$ , ou seja, a energia de um oscilador harmônico simples sem perdas é proporcional ao quadrado da frequência e da amplitude de oscilação.

Considere  $x(n)$  a versão amostrada de  $x(t)$ :

$$x(n) = A \cos(\omega n + \theta), \quad (2.26)$$

onde  $\omega$  é a frequência normalizada do sinal discreto e vale  $\omega = 2\pi f/f_s$ , em que  $0 \leq f < f_s/2$ ,

$f_s$  é a frequência de amostragem,  $n$  é o número da amostra e  $\theta$  é uma fase inicial arbitrária em radianos.

Usando 2.26 e algumas relações trigonométricas, é possível mostrar (Kaiser, 1990a) que  $x(n+1) \cdot x(n-1) = x^2(n) - A^2 \text{sen}^2(\omega)$ . Para pequenos valores de  $\omega$ , vale a aproximação  $\text{sen}(\omega) \approx \omega$ , e:

$$A^2 \omega^2 \approx x^2(n) - x(n+1) \cdot x(n-1). \quad (2.27)$$

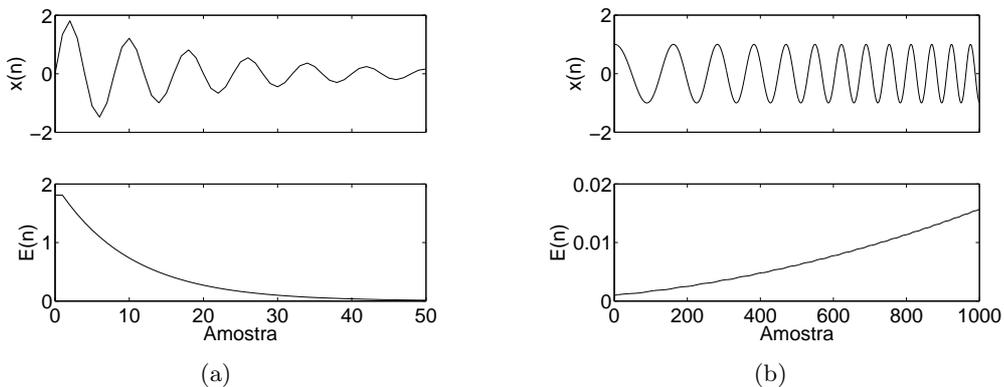
Esta aproximação tem um erro menor que 11% quando  $\omega < \pi/4$ , ou seja,  $f/f_s < 1/8$ . Assim, se o sinal  $x(t)$  for amostrado com pelo menos 8 pontos por ciclo, é possível, dentro deste erro, obter uma estimativa do produto  $(A\omega)^2$  usando apenas 3 amostras de  $x(n)$ . Este produto, como já foi dito, é proporcional à energia do sistema e a expressão abaixo fornece um algoritmo simples para medi-la:

$$\Psi_d[x(n)] = x(n)^2 - x(n+1) \cdot x(n-1) = A^2 \text{sen}^2(\omega) \approx A^2 \omega^2 \propto E(n). \quad (2.28)$$

De acordo com Kaiser (1990a), foi H. M. Teager quem concebeu tal algoritmo. Por isso, ele foi nomeado algoritmo ou operador de Teager, também chamado de operador de energia por outros autores. Teager, porém, não forneceu suas demonstrações nem propriedades e coube a Kaiser (1990a) publicá-las. Em um outro trabalho, Kaiser (1990b) publicou a versão contínua do operador de Teager (equação 2.29):

$$\Psi_c[x(t)] = \left[ \frac{dx(t)}{dt} \right]^2 - x(t) \frac{d^2x(t)}{dt^2} = [\dot{x}(t)]^2 - x(t)\ddot{x}(t). \quad (2.29)$$

Kaiser (1990a) mostrou também que o operador de Teager é capaz de rastrear a energia de um processo que gera um sinal senoidal exponencialmente amortecido, como mostra a figura 2.11(a). Além disso, para sinais com amplitude constante e frequência variável, Kaiser (1990a) demonstrou que o algoritmo estima a sua frequência instantânea, como exemplificado na figura 2.11(b).



**Figura 2.11:** (a) Sinal  $x(n) = 2e^{-0,05n} \text{sen}(n\pi/4)$  na figura de cima. Abaixo, a energia estimada pelo algoritmo de Teager. (b) Sinal com frequência aumentando linearmente de  $\pi/100$  a  $4\pi/100$ . Note que, como a amplitude é constante, a saída do algoritmo de Teager é proporcional ao  $\text{sen}^2(\Omega)$ .

DESA-1	DESA-2
$G(n) = 1 - \frac{\Psi_d[x(n)-x(n-1)] + \Psi_d[x(n+1)-x(n)]}{4\Psi_d[x(n)]}$	
$\Omega_i \approx \arccos(G(n))$	$\Omega_i \approx \arcsen\left(\sqrt{\frac{\Psi_d[x(n+1)-x(n-1)]}{4\Psi_d[x(n)]}}\right)$
$ a(n)  \approx \sqrt{\frac{\Psi_d[x(n)]}{1-G^2(n)}}$	$ a(n)  \approx \frac{2\Psi_d[x(n)]}{\sqrt{\Psi_d[x(n+1)-x(n-1)]}}$

**Tabela 2.1:** Algoritmos para isolar os componentes AM e FM. O significado das variáveis está explicado no texto e  $\Psi_d$  está definido na equação 2.28.

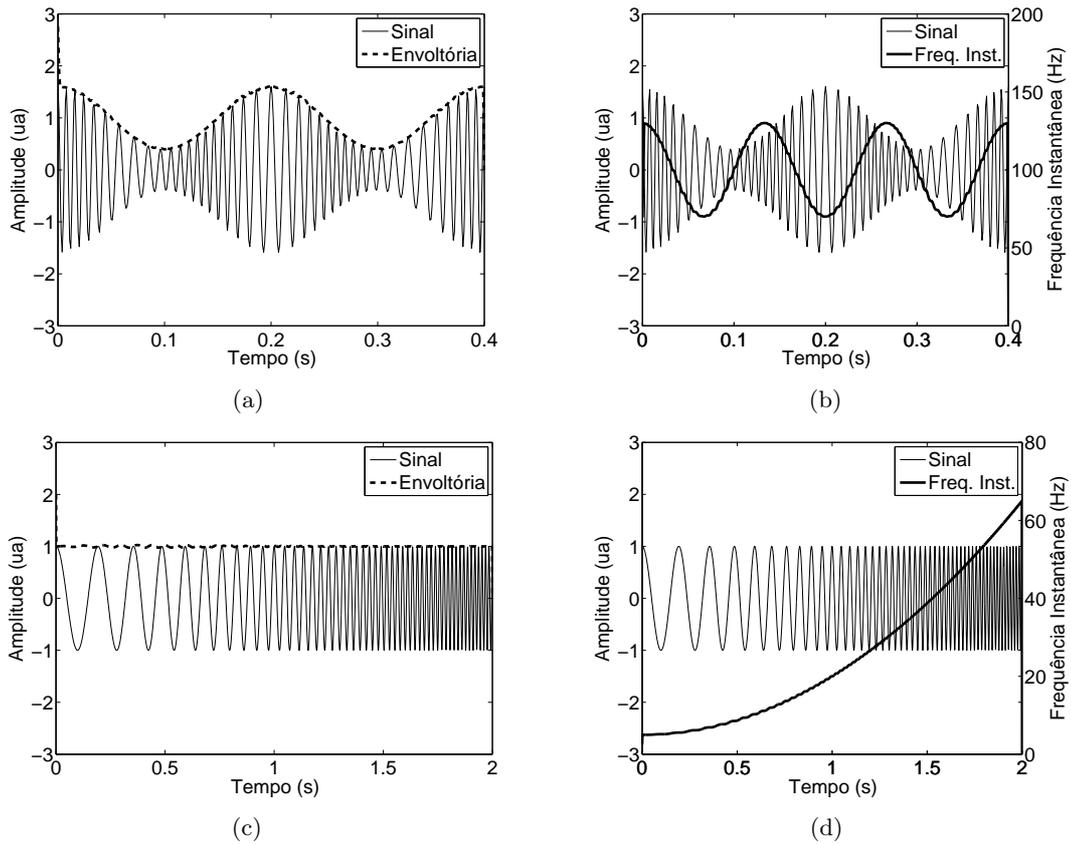
Maragos et al. (1991) foram um dos primeiros a aplicar o operador de Teager à análise do sinal de fala. Os autores filtraram um formante do sinal de fala, modelaram-no usando a modulação AM-FM (equação 2.7) e mostraram que é possível estimar o quadrado do produto entre sua envoltória e sua frequência instantânea usando o algoritmo de Teager (equação 2.28).

Este trabalho foi ampliado em Maragos et al. (1992, 1993b,a). Os autores publicaram algoritmos que usam o operador de Teager para separar a envoltória e a frequência instantânea de um sinal AM-FM. Estes algoritmos foram nomeados algoritmos de separação de energia (ESA – *Energy Separation Algorithm*), sendo que o de tempo contínuo foi chamado CESA e os de tempo discreto receberam nomes de acordo com a aproximação usada para discretizar a derivada. DESA-1 aproxima a derivada usando a diferença entre dois pontos consecutivos,  $dx/dt \approx x(n) - x(n-1)$ , e DESA-2 a aproxima usando a diferença entre três amostras consecutivas,  $dx/dt \approx x(n+1) - x(n-1)$ , conforme a tabela 2.1. De acordo com Maragos et al. (1992) o DESA-1 produz erros um pouco menores que o DESA-2.

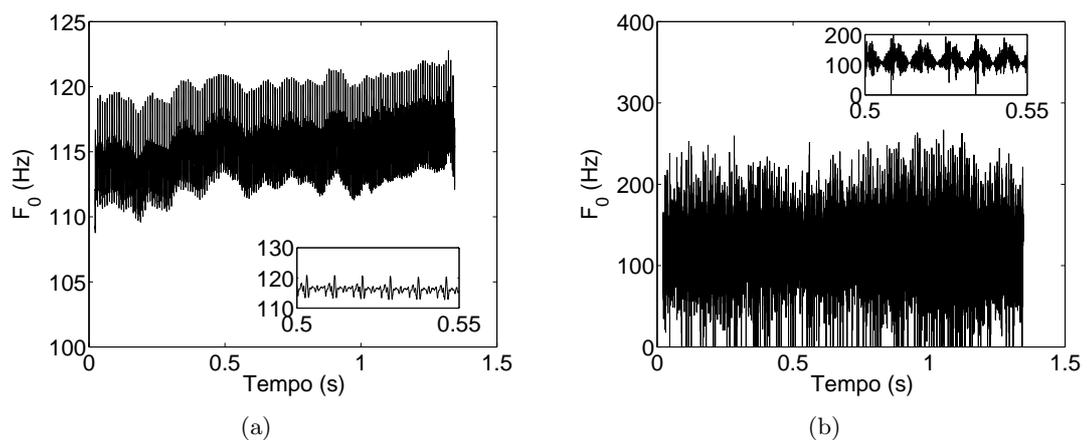
A figura 2.12 ilustra o funcionamento do DESA-1. Note como ele é capaz de acompanhar, em tempo real, as variações na envoltória e na frequência do sinal AM-FM nos dois sinais mostrados. Contudo, estes algoritmos só têm garantia de funcionamento se o sinal AM-FM for de banda estreita (Maragos et al., 1993b). Por isso, antes de aplicar estes algoritmos é necessário filtrar o sinal a fim de limitar sua banda.

A escolha dos parâmetros deste filtro é crucial para garantir que os “DESAs” produzam resultados corretos (Potamianos e Maragos, 1994). As figuras 2.13 mostram a estimativa produzida pelo DESA-1 para a curva de  $F_0$  do mesmo sinal de voz real usado nas figuras 2.9. Na figura 2.13(a), assim como na figura 2.9(a), usou-se o passa-faixa 2.10(a). Na figura 2.13(b), o filtro usado foi o mesmo da figura 2.9(b), ou seja, o passa-baixa exibido na figura 2.10(b). Nos dois casos, não há nenhum múltiplo de  $F_0$  presente, mas a estimativa não é satisfatória em nenhum deles. Além disso, a curva da figura 2.13(a), gerada usando-se o filtro de banda mais estreita, apresenta variância bem menor que a da figura 2.13(b). Isto acontece, porque, assim como na representação analítica, a frequência instantânea concentra a informação de todo a região analisada.

Para rastrear o valor instantâneo da frequência fundamental ou de um determinado formante (Potamianos e Maragos, 1996) é necessário, além de centrar corretamente um filtro



**Figura 2.12:** Extração da envoltória e da frequência instantânea pelo DESA-1 (a) em  $x(n) = [1 + 0,6 \cos(2\pi \frac{5n}{1000})] \cos[2\pi \frac{100n}{1000} + 4\text{sen}(2\pi \frac{7,5n}{1000})]$  e (b) em um sinal de frequência inicial 5 Hz e crescimento quadrático. Ambos amostrados a 1 kHz.



**Figura 2.13:** Extração de  $F_0$  utilizando o DESA-1. O mesmo sinal de voz real foi usado nas figuras, sendo que, em (a), o sinal de voz foi filtrado com um passa-faixa centrado em torno da frequência fundamental e em (b) foi usado um passa-baixa com frequência de corte um pouco acima do valor de  $F_0$ . Os detalhes mostram pormenores das curvas. Os parâmetros dos filtros estão descritos no texto.

passa-banda na região de interesse, definir a sua largura de faixa de modo a minimizar a interferência provocada pelo espectro de regiões vizinhas. Hanson et al. (1993) propuseram um algoritmo que automaticamente ajusta a frequência central do filtro, mas que seleciona manualmente a largura de banda. Potamianos e Maragos (1994) afirmam que a tarefa de determinar corretamente a largura de banda não é trivial. Isto justifica a escolha manual dos parâmetros do filtro em diversos trabalhos (Maragos et al., 1991, 1992, 1993a; Potamianos e Maragos, 1996), constituindo uma importante limitação ao uso automático do método em vozes reais.

## 2.4 Comparação entre o operador de Teager e o sinal analítico

Alguns trabalhos comparam o desempenho de algoritmos que usam o operador de Teager e o sinal analítico para estimar a envoltória e a frequência instantânea de um sinal.

Potamianos e Maragos (1994) consideram o operador de Teager uma melhor escolha em relação ao sinal analítico. Nos resultados apresentados por estes autores, usando sinais sintetizados sem adição de ruído e sinais de voz real filtrada, o operador de Teager foi capaz de gerar resultados melhores ou similares aos do sinal analítico usando apenas 3 amostras consecutivas do sinal para realizar seus cálculos. A representação analítica, por outro lado, precisa de toda a massa de dados para gerar seu resultado.

Já Vakman (1996) mostra, em sinais corrompidos por ruído de banda larga, que o sinal analítico produz melhores estimativas de envoltória e frequência instantânea. Ele ainda observa que a utilização do operador de Teager em tempo contínuo (equação 2.29) é problemática para o caso de sinais com discontinuidades, uma vez que a derivada não é bem definida nestes locais.

As simulações apresentadas nas seções 2.2 e 2.3 confirmam a observação de que o sinal analítico é mais robusto ao ruído que o operador de Teager. As figuras 2.9 e 2.13 mostram o resultado da estimação da curva de  $F_0$  do mesmo sinal de voz real usando o sinal analítico e o operador de Teager. Em todos estes casos, o sinal de voz (multicomponente) foi filtrado de forma a manter apenas o harmônico fundamental e a estimativa das figuras 2.9(a) e 2.13(a) foi melhor que a das figuras 2.9(b) e 2.13(b) devido à menor largura de banda do filtro – o que diminuiu o nível de “ruído” interferindo em  $F_0$ . Além disso, a estimativa de 2.9(a) apresenta variância menor que a de 2.13(a). Como mostrado por Vakman (1996), isto é uma consequência direta da suavização imposta pela representação analítica ao usar toda a informação do sinal para estimar sua frequência instantânea, enquanto o operador de Teager usa apenas 3 amostras consecutivas.

## 2.5 Análise de Fourier

Um requisito comum aos dois métodos investigados anteriormente é o de limitar a banda do sinal que será analisado. Isto é necessário para garantir que a frequência instantânea esti-

mada tenha significado físico (Boashash et al., 1989). Todavia, como ilustrado nos exemplos das figuras 2.9 e 2.13, ajustar os parâmetros do filtro para limitar corretamente uma banda do sinal é uma tarefa difícil. Uma alternativa para isolar a curva de  $F_0$  evitando o emprego de filtros é usar algoritmos de rastreamento da frequência fundamental, ou, simplesmente PDAs (*Pitch Detection Algorithms*), como descrito abaixo.

### 2.5.1 Medição de $F_0$

Os algoritmos de rastreamento da frequência fundamental extraem a frequência fundamental diretamente do sinal original baseando-se em características como:

- Picos da função de autocorrelação (Rabiner e Schafer, 1978, pags 141–149):

$$r_{xx}(K) = \lim_{N \rightarrow \infty} \left\{ \frac{1}{2N+1} \sum_{n=-N}^N x(n)x(n-K) \right\}. \quad (2.30)$$

onde  $x(n)$  é um sinal discreto periódico de valor médio nulo,  $x(n)$ . Mostra-se que, se o período de  $x(n)$  é igual a  $P$ , então

$$r_{xx}(K) = r_{xx}(K + P), \quad (2.31)$$

ou seja, a função de autocorrelação de  $x(n)$  também é periódica e tem o mesmo período de  $x(n)$  (Rabiner e Schafer, 1978). Além disso, (i)  $r_{xx}(K) = r_{xx}(-K)$  e (ii)  $|r_{xx}(K)| \leq r_{xx}(0) \forall K \in \mathbb{Z}$ . Estas propriedades fazem com que o primeiro máximo da função de autocorrelação possa ser usado para estimar o período do sinal  $x(n)$ .

- Vales da função de AMDF (*Average Magnitude Difference Function*), isto é, vales da função de “anti-correlação” (Ross et al., 1974), calculada para  $x(n)$  (definido acima):

$$A(K) = \sum_{n=-\infty}^{\infty} |x(n) - x(n-K)|. \quad (2.32)$$

- Picos do cesptrum (Noll, 1967), onde  $c(n) = |\mathcal{F}\{\log [|X(e^{j\omega})|^2]\}|^2$

Para sinais de voz periódicos com período  $P$ ,  $c(n)$  terá um pico pronunciado em  $n = P$  e picos menores em múltiplos de  $P$  (Oppenheim e Schafer, 1975).

Em aplicações onde a identificação individual de ciclos glóticos é desnecessária, como na codificação de voz ou na linguística, todas as técnicas listadas acima podem apresenta desempenho satisfatório. Entretanto, para seguir os detalhes presentes em cada ciclo da voz, há um certo consenso de que os algoritmos que operam no domínio do tempo, como o de comparação de forma de onda (*wave-matching*), têm melhor desempenho (Vieira, 1997, cap. 5).

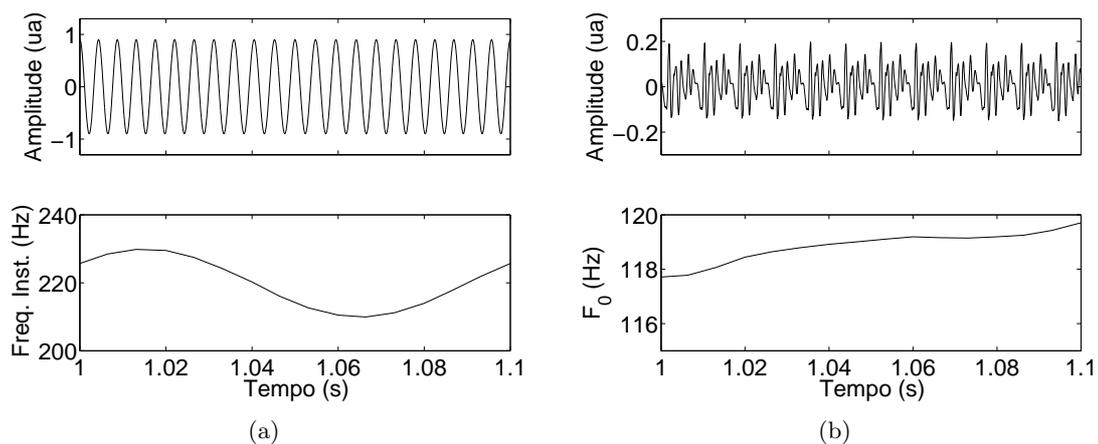
Um exemplo destes algoritmos e que usa a função de autocorrelação é o *Super Resolution Pitch Determination* (SRPD) proposto por Medan et al. (1991). Segundo Medan et al. (1991, pag. 43):

O período verdadeiro  $P_0$  ... é caracterizado por um alto coeficiente de correlação entre dois segmentos de tamanho arbitrário ( $\tau$ ) separados por um intervalo  $P_0$ , desde que  $\tau$  não seja muito grande. Além disso, estimativas de múltiplos ou frações de  $P_0$  também compartilham esta propriedade. Entretanto, os máximos da função de correlação resultantes destas “sub-periodicidades” são diferentes, a menos que elas ocorram exatamente em submúltiplos de  $P_0$  e tenham a mesma magnitude. (Tradução própria)

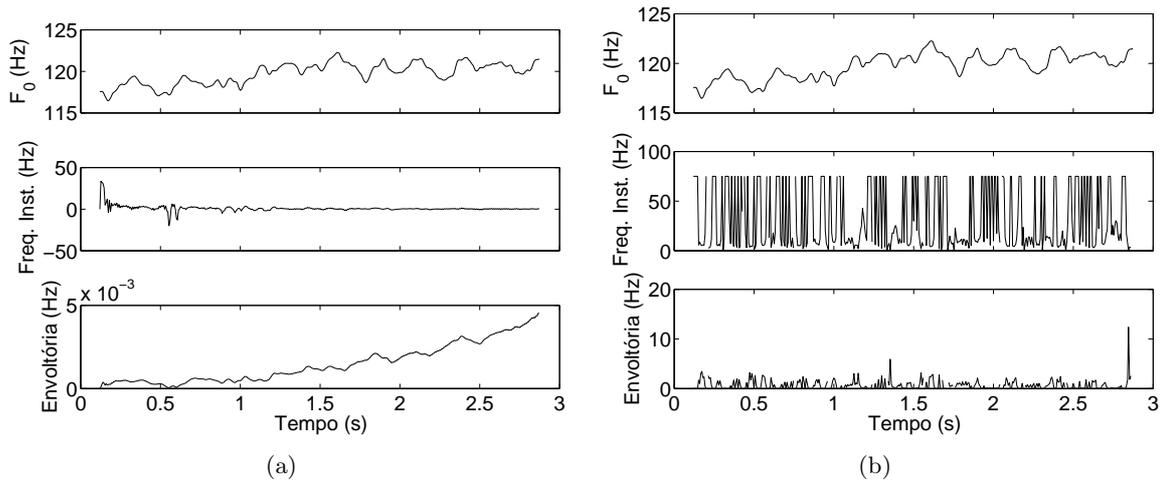
Para eliminar múltiplos ou frações do período verdadeiro, Medan et al. (1991) usam uma abordagem baseada em critérios empíricos. Caso estes critérios não encontrem nenhum valor de  $P_0$ , o segmento analisado é classificado como desvozeado. Na parte final do algoritmo, as estimativas de  $P_0$  são refinadas por meio de interpolação.

O SRPD já foi avaliado independentemente em diversos trabalhos (Bagshaw et al., 1993; Bagshaw, 1994; Vieira, 1997). Os trabalhos mostram que este algoritmo é um dos mais confiáveis extratores de  $F_0$  (Bagshaw, 1994, cap. 7). Além disso, seu código fonte já está implementado em C++ e disponibilizado gratuitamente (CSTR, 2010). Alguns exemplos da aplicação deste código estão mostrados nas figuras 2.14(a) e 2.14(b).

A figura 2.14(a) ilustra a aplicação do SRPD para calcular a frequência instantânea de um sinal FM sintetizado e 2.14(b) exemplifica seu uso na extração da frequência fundamental de um sinal de voz real. O algoritmo calcula o valor da frequência em intervalos regulares de tempo, que, tanto em 2.14(a) quanto em 2.14(b) foi de 20 ms. O resultado exibido nas figuras mostra que o SRPD é capaz de rastrear as flutuações do período do sinal, apesar delas não serem visualmente detectáveis na forma de onda do sinal.



**Figura 2.14:** (a) Sinal modulado em frequência  $x_{fm}(t) = \cos[440\pi t + 1 \cdot \text{sen}(20\pi t)]$  no gráfico superior e curva de frequência instantânea (Freq. Inst.) na parte inferior. (b) Trecho da vogal /a/ na parte superior e curva de  $F_0$  na parte inferior. Nestes exemplos, o SRPD gerou amostras a uma taxa de  $1/20 \text{ ms} = 50 \text{ Hz}$  e, para melhorar a visualização, os resultados foram interpolados para uma taxa de 150 Hz.



**Figura 2.15:** (a) Sinal analítico e (b) operador de Teager aplicados à curva de  $F_0$ .

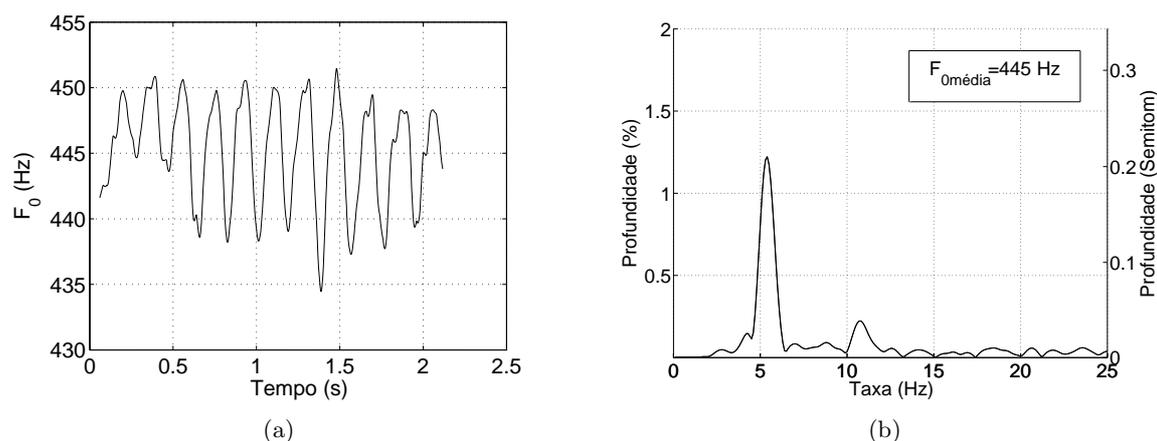
A curva produzida pelo SRPD tem largura de faixa menor que o sinal original, entretanto, não é monocomponente. Por isso a análise feita pelo algoritmo de Teager ou pelo sinal analítico gera resultados de difícil interpretação física. A frequência instantânea de  $F_0$ , que está relacionada à evolução temporal da sua taxa (figura 1.1), apresenta valores negativos quando analisada com o sinal analítico (figura 2.15(a)) e diversos picos quando avaliada com o algoritmo de Teager (figura 2.15(b)). Já a envoltória, relacionada ao comportamento temporal da extensão de  $F_0$  (figura 1.1), tem valor zero em diversos momentos. Como já explicado em seções anteriores, para evitar estes resultados, o sinal de  $F_0$  precisaria ser filtrado antes de aplicá-lo a estas técnicas. Contudo, a filtragem modifica o contorno do sinal, inviabilizando medidas como a de regularidade, que é baseada na forma da onda de  $F_0(t)$ .

### 2.5.2 Análise Espectral

Uma abordagem que permite avaliar a taxa e a extensão da curva melódica sem romper sua forma é a análise de Fourier. Neste método, explora-se o *espectro da frequência fundamental* de forma a obter informações sobre a taxa e a extensão de  $F_0$ .

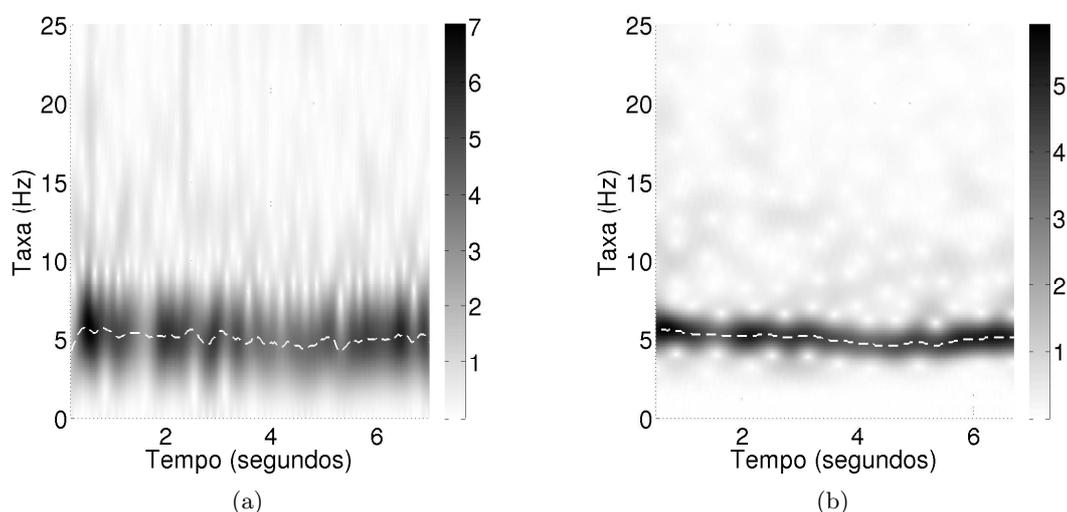
Um exemplo da análise de Fourier está mostrado na figura 2.16. A frequência fundamental (figura 2.16(a)) foi extraída do vibrato produzido em um contrabaixo acústico em torno de uma nota com fundamental em aproximadamente 445 Hz. Uma estimativa manual usando esta curva melódica indica que a taxa do vibrato está em torno de 5 Hz e a profundidade, em aproximadamente  $(\frac{450-440}{2})/(445) \approx 1,1\%$ . O espectro da curva de  $F_0$  (figura 2.16(b)), calculado através da análise de Fourier, que será detalhada na seção 3.3.6, apresenta os valores de profundidade para cada componente do vibrato. Esta curva confirma as estimativas manuais.

Uma desvantagem da análise de Fourier é, em muitos casos, a impossibilidade de expressar o comportamento temporal da taxa e da profundidade. Os resultados da análise de Fourier apontam o valor da taxa e da profundidade presentes em toda a curva de  $F_0$ , e nada

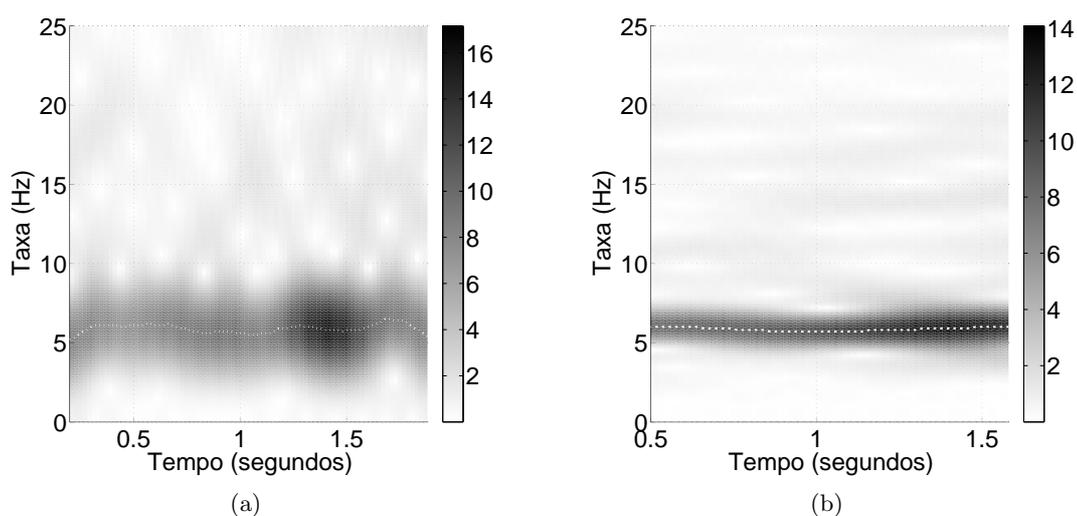


**Figura 2.16:** Análise de Fourier aplicada à curva extraída pelo SRPD. (a) Curva melódica de um vibrato. (b) Espectro da curva melódica.

informam sobre a evolução destes parâmetros ao longo do tempo. Uma alternativa para superar esta desvantagem seria o espectrograma. A figura 2.17 mostra que o espectrograma de banda estreita consegue rastrear a evolução temporal em um vibrato de, aproximadamente, 7 segundos de duração. Contudo, mesmo para amostras musicalmente longas, de 2 segundos de duração, por exemplo, não é possível obter uma resolução no tempo e na frequência satisfatórias (figura 2.18). Por isto, o espectrograma não foi explorado neste trabalho e uma



**Figura 2.17:** Espectrograma de uma amostra de vibrato de longa duração temporal. A escala de tons de cinza indica a profundidade, expressa em unidades arbitrárias. A linha branca tracejada não segue a escala de cinza e indica a taxa de maior extensão em cada instante de tempo. (a) Espectrograma de banda larga (janela de 0,4 segundos). Apesar da resolução temporal ser boa, os artefatos causados pela curta janela de dados contaminam os valores de frequência. (b) Espectrograma de banda estreita (janela de 1 segundo). A resolução temporal é aceitável e há menor interferência de artefatos.



**Figura 2.18:** Espectrograma de uma amostra de vibrato de curta duração temporal. A escala de tons de cinza indica a profundidade, expressa em unidades arbitrárias. A linha branca tracejada não segue a escala de cinza e indica a taxa de maior extensão em cada instante de tempo. (a) Espectrograma de banda larga (janela de 0,4 segundos). Há boa resolução temporal, mas a resolução em frequência é muito pobre. (b) Espectrograma de banda estreita (janela de 1 segundo). A frequência fica mais bem resolvida, mas a perda na resolução temporal dificulta avaliar a evolução da frequência do vibrato ao longo do tempo.

outra técnica, descrita na seção 3.3.7, foi utilizada para analisar o comportamento temporal da taxa e da profundidade.

## 2.6 Conclusão

As técnicas de análise de modulação em amplitude e em frequência representam maneiras de descrever as flutuações presentes em um sinal. Geralmente, estas flutuações contêm informações úteis e precisam ser extraídas. Este capítulo apresentou três métodos que podem ser usados para a extração de flutuações:

1. Sinal analítico;
2. Algoritmo de Teager;
3. Análise de Fourier.

O sinal analítico e o algoritmo de Teager fornecem resultados com alguma interpretação física quando analisam sinais de banda estreita. Por isso, antes de usar estas técnicas em sinais de voz, é necessário filtrá-los e isolar uma pequena faixa de seu espectro. Já o método de Fourier é capaz de lidar com sinais de banda larga. Este método é usado pelo algoritmo apresentado no capítulo 3 e a descrição detalhada de sua implementação está mostrada na seção 3.3.6. O capítulo 3 também apresenta os procedimentos para a medição da curva de  $F_0$ , de calibração da análise espectral e de análise de regularidade.

## Capítulo 3

# Metodologia

### 3.1 Introdução

O algoritmo proposto para analisar as flutuações do vibrato e do tremor está descrito neste capítulo. Este algoritmo é basicamente composto por três grandes blocos de análise, que serão apresentados e avaliados detalhadamente em seções posteriores:

- Extração e análise da curva de frequência fundamental do sinal original;
- Extração e análise da envoltória do sinal original;
- Análise de regularidade da frequência fundamental.

A curva de  $F_0$  é extraída do sinal original usando-se o algoritmo de *Super Resolution Pitch Determination* (SRPD) e analisada com um método baseado na análise de Fourier (Vieira, 1998). A ideia de usar técnicas de Fourier para analisar a série temporal de  $F_0(n)$  não é nova, porém, poucos a empregaram para investigar o tremor e o vibrato. Na literatura relacionada, Yair e Gath (1988) foram os únicos a usá-la para estudar o tremor vocal. Eles determinaram o espectro de  $F_0$  de vários indivíduos afetados pela doença de Parkinson e mostraram que ele pode ser diagnosticado com este espectro. Contudo, não exploraram todas as possibilidades desta técnica, apresentando apenas valores da taxa das perturbações, sem investigar as respectivas extensões. A seção 3.3.6 mostrará como a análise de Fourier pode ser usada, quais os cuidados devem ser tomados para que distorções sejam evitadas e quais os resultados podem ser alcançados.

A envoltória do sinal original, por sua vez, é separada por meio de transformação homomórfica (Qi e Weinberg, 1994) e estudada com a mesma técnica empregada na série de  $F_0$ , ou seja, a análise de Fourier.

A análise de regularidade é feita usando-se a função de autocorrelação (Rabiner e Schafer, 1978). O método descrito na seção 3.3.7 foi desenvolvido neste trabalho para contornar as dificuldades de analisar amostras de curta duração com o espectrograma. Ele se baseia numa quantificação da semelhança entre ciclos consecutivos de  $F_0$  para indicar o quão regular é a curva da frequência fundamental.

Além desses métodos de análise, este capítulo também descreve as bases de dados utilizadas neste trabalho. Elas foram construídas com dados cedidos por vários grupos de pesquisa de diversas áreas e são detalhadas na próxima seção.

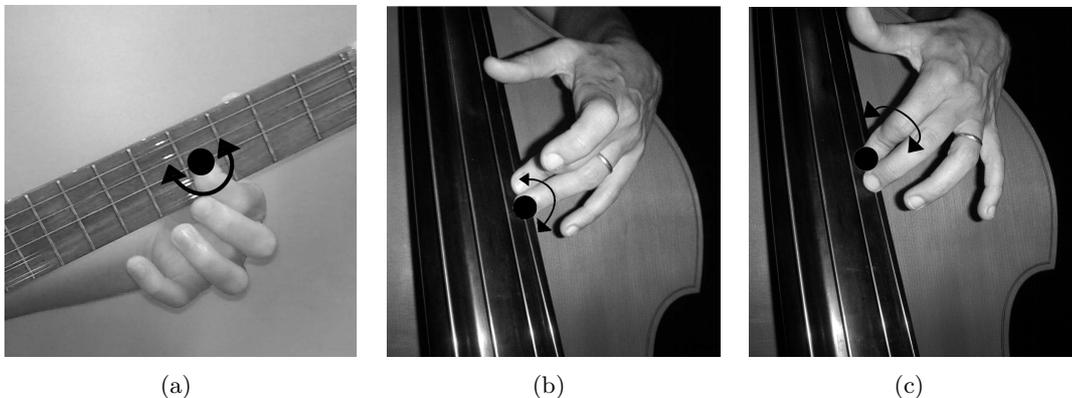
## 3.2 Bases de dados

### 3.2.1 Base de dados 1: tremor em membros

Esta base de dados foi obtida na página do professor Jens Timmer da Universidade de Freiburg (Jens Timmer, 2010). Ela é formada por 10 séries, sendo 5 coletadas de pacientes com tremor essencial e 5 registradas de pessoas com doença de Parkinson. Cada série contém 30 segundos de medidas feitas com acelerômetros presos ao dorso da mão espalmada de cada informante. A taxa de amostragem foi de 1000 Hz em nove delas e de 300 Hz em apenas uma. Não há dados sobre idade/sexo dos pacientes.

### 3.2.2 Base de dados 2: vibrato no contrabaixo

A base de dados utilizada foi a mesma de Borém et al. (2005). Ela é composta por 3 blocos de amostras de vibrato produzidas em notas de registros diferentes do contrabaixo acústico: grave (110 Hz), médio (220 Hz) e agudo (440 Hz). Em cada um destes registros há 30 repetições de vibrato do tipo I, 30 de vibrato tipo II e 30 de vibrato tipo III (figura 3.1). Todas as amostras foram produzidas por um só instrumentista e amostradas a 22050 Hz.



**Figura 3.1:** Tipos de vibrato. O círculo preto indica o dedo que está apoiado no braço do instrumento e as setas indicam sua movimentação durante a produção do vibrato. (a) Vibrato tipo I. Apenas um dedo é usado. (b) Vibrato tipo II. Há dois dedos envolvidos nesta técnica de produção de vibrato. Um dedo prende a corda na nota alvo, enquanto o outro, posicionado em uma região mais grave, auxilia o movimento e, idealmente, não toca a corda. (c) Vibrato tipo III. Este vibrato também usa dois dedos e a sua diferença para o tipo II é o posicionamento do dedo auxiliar. No tipo III, o dedo auxiliar fica posicionado em uma região mais aguda do instrumento.

### 3.2.3 Base de dados 3: perturbação em voz

A base de dados foi cedida pela professora Ana Cristina C. Gama do departamento de fonoaudiologia da UFMG. Esta conjunto de dados é formado por dois grupos:

- 4 emissões de vogal /a/ e 1 emissão de vogal /e/ de pacientes com tremor vocal confirmado perceptivamente. Duas das amostras da vogal /a/ foram produzidas por um mesmo paciente, antes e depois de ser tratado com Propranolol 40 mg (seção 4.4.1). Todas amostradas a 11025 Hz e gravadas em consultório fonoaudiológico.
- 32 emissões de vogal /a/ de pacientes com doença de Parkinson, mas sem avaliação perceptiva do tremor. Este grupo de amostras foi coletado a taxas de 11025, 22050 e 44100 Hz no ambulatório de Fonoaudiologia do Hospital das Clínicas da UFMG.

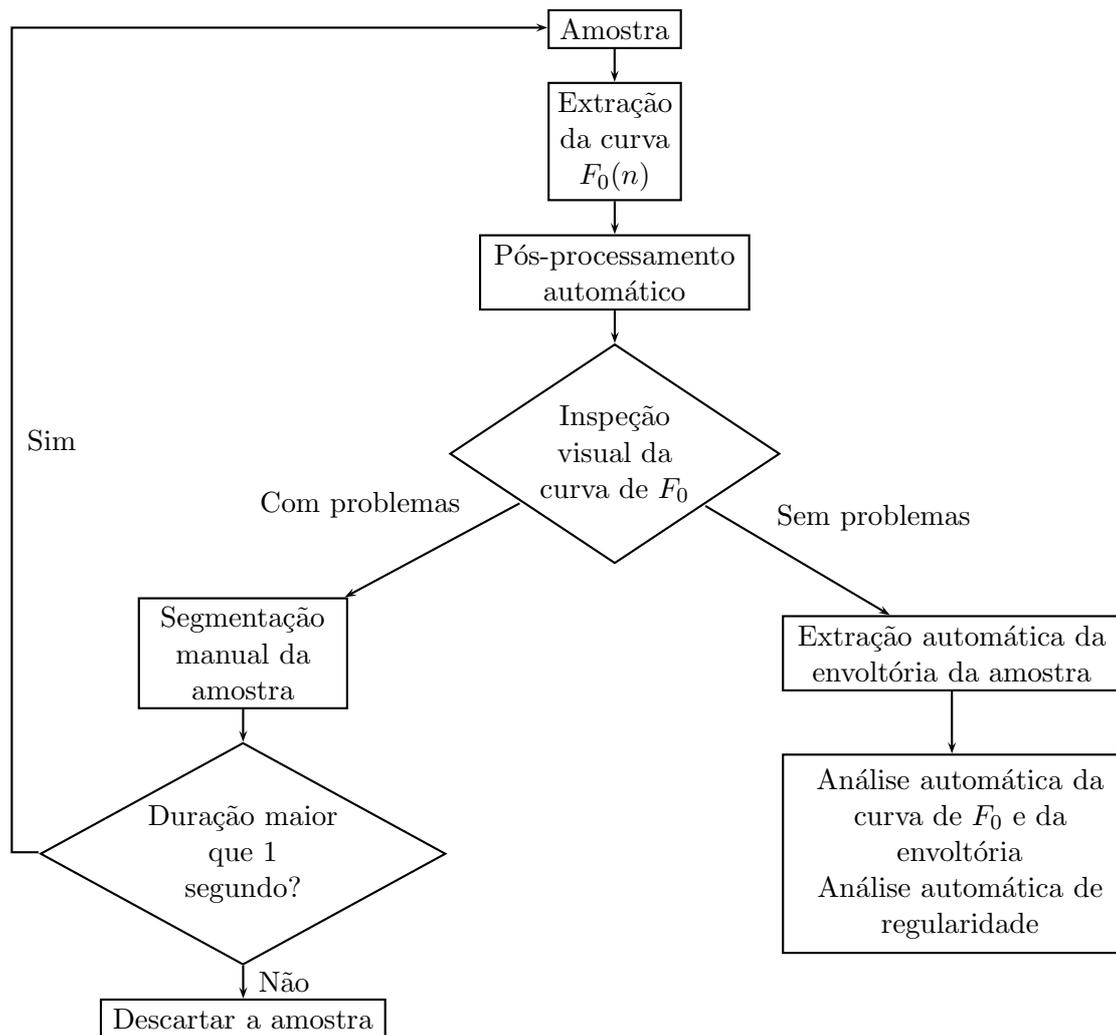
### 3.2.4 Base de dados 4: vibrato no canto

- A base de dados de canto lírico foi obtida em (Rothman, 2010). Todas as amostras de vibrato deste conjunto foram extraídas de trechos de músicas. Foram utilizadas 11 emissões, sendo 8 da vogal /a/ e 3 da vogal /e/, todas sem acompanhamento instrumental e amostradas a 22050 Hz.
- A base de dados de canto sertanejo foi composta com 15 exemplares do conjunto utilizado por Rosa (2003). Estas amostras foram produzidas por onze cantores sertanejos enquadrados na categoria semiprofissional (Rosa, 2003) que cantaram o trecho padrão da música “Galopeira” de Maurício Cardozo Ocampo. Todas as emissões de vibrato que formaram esta base de dados foram produzidas sobre vogal /a/ e amostradas a 22050 Hz.

## 3.3 Algoritmo

O funcionamento do algoritmo está resumido na figura 3.2. Cada etapa é brevemente descrita abaixo e será detalhada em seções posteriores.

- **Extração da fundamental:** Executa-se o algoritmo SRPD sobre o sinal acústico a fim de extrair  $F_0$ .
- **Pós-processamento:** A curva extraída pelo SRPD é processada para correção de pequenos erros na curva da frequência fundamental.
- **Inspeção visual curva de  $F_0$ :** Verifica-se visualmente se a curva de  $F_0$  não apresenta quebras.
- **Se a curva de  $F_0$  apresenta problemas:**
  - **Segmentar sinal:** Fragmenta-se manualmente o sinal original de modo a eliminar as regiões que provocaram erros do SRPD. Caso o sinal segmentado tenha mais que 1 segundo de duração, ele é aproveitado; caso contrário, o fragmento é descartado.



**Figura 3.2:** Descrição geral do método de análise.

- **Se a curva de  $F_0$  não apresenta problemas:**
  - **Extração da envoltória:** Extrai-se a envoltória do sinal acústico utilizando-se transformação homomórfica.
  - **Análise da frequência fundamental e da envoltória:** Utiliza-se a transformada de Fourier para analisar a curva de  $F_0$  e a envoltória do sinal original.
  - **Análise de regularidade:** Analisa-se a semelhança entre ciclos consecutivos da curva de  $F_0$ .

### 3.3.1 Extração da fundamental

O executável do SRPD foi instalado juntamente com outros programas do pacote *festival*, versão beta 1.96 (CSTR, 2010), adquiridos gratuitamente a partir do repositório padrão do

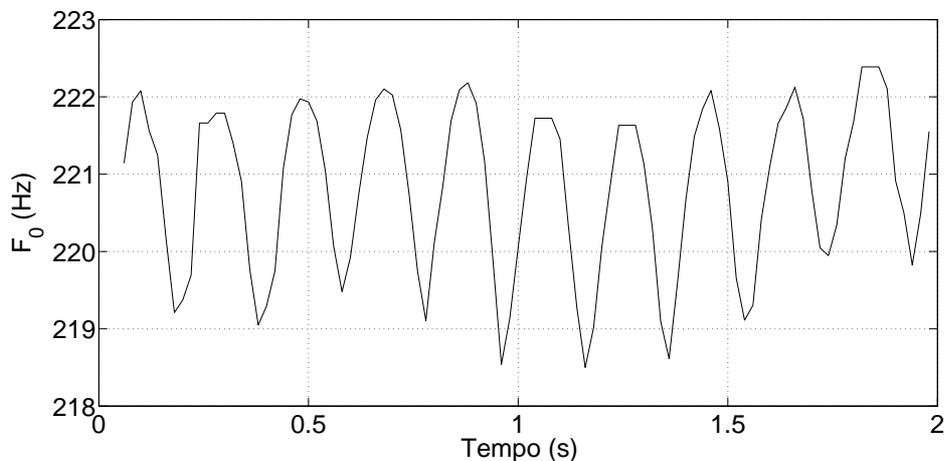
sistema operacional Ubuntu 10.04 (Ubuntu, 2010). A configuração dos parâmetros usados por este executável foi:

- L** : Realiza uma filtragem passa-baixa no sinal de entrada do algoritmo.
- lpfilter 1000** : A frequência de corte do filtro passa-baixa foi de 1000 Hz.
- d 1** : Não houve decimação do sinal de entrada, logo, o fator de decimação usado foi 1.
- P** : Refina a extração de  $F_0$  (Bagshaw et al., 1993).
- fmin 80** : Configura o menor valor permitido para  $F_0$  em 80 Hz.
- fmax 800** : Configura o maior valor permitido para  $F_0$  em 800 Hz.
- shift 0.02** : Taxa de amostragem do SRPD foi de  $1/0,02 = 50$  Hz.

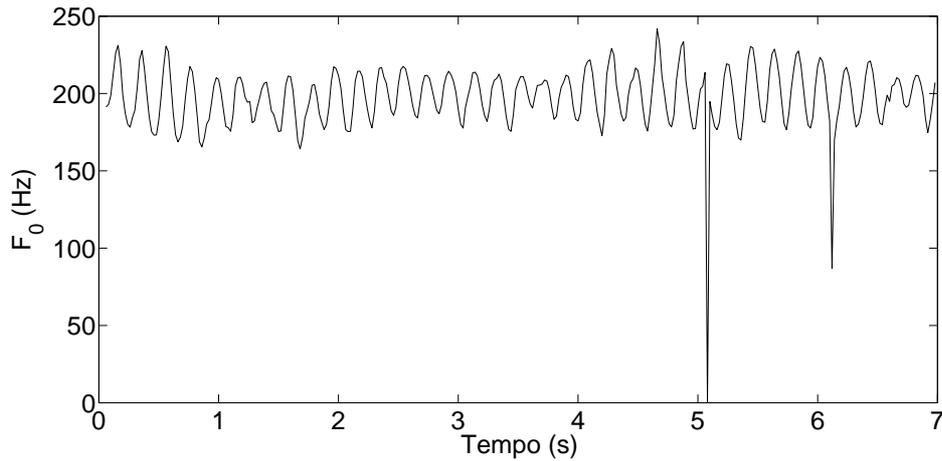
Os valores de  $fmin$  e  $fmax$  foram escolhidos para abranger toda a grande faixa de  $F_0$  nas bases de dados estudadas (voz, canto, contrabaixo). A taxa de amostragem foi configurada em 50 Hz por ser suficiente para investigar as flutuações de interesse, que oscilam em torno de 4 a 15 Hz. Um exemplo de resultado produzido nesta etapa está mostrado na figura 3.3. Ela foi produzida a partir de uma amostra de um vibrato executado em um contrabaixo acústico. Nota-se uma diferença na duração dos picos e dos vales de  $F_0$ , sugerindo que o instrumentista tenha executado um movimento assimétrico.

### 3.3.2 Pós-processamento automático

É comum a curva de frequência fundamental estimada pelo algoritmo de extração de  $F_0$  apresentar alguns erros. Estes erros são consequências de instabilidades no sinal original que não foram adequadamente rastreadas. Na figura 3.4, por exemplo, o algoritmo falha grosseiramente em torno de 5 e de 6 segundos. Erros como este devem ser tratados em uma etapa de pós-processamento.



**Figura 3.3:** Exemplo de curva de  $F_0$  extraída de um vibrato produzido em contrabaixo acústico.



**Figura 3.4:** Estimação de  $F_0$  com alguns erros. Erros grosseiros são vistos em torno de 5 e de 6 segundos.

Uma solução comumente empregada para eliminar estas falhas é a associação entre um filtro de mediana e um suavizador linear (Rabiner et al., 1975; Bagshaw et al., 1993). O filtro de mediana percorre o sinal de  $F_0$ , amostra por amostra, substituindo cada valor pela mediana entre a amostra atual e seus vizinhos. Usualmente, é utilizada a mediana de 3 pontos (Rabiner et al., 1975):

$$\text{mediana}(a, b, c) = \begin{cases} a, & b \leq a \leq c \text{ ou } c \leq a \leq b, \\ b, & a \leq b \leq c \text{ ou } c \leq b \leq a, \\ c, & b \leq c \leq a \text{ ou } a \leq c \leq b. \end{cases} \quad (3.1)$$

O suavizador linear tem a função de reduzir as irregularidades preservando o padrão de comportamento da curva. Normalmente, este suavizado é implementado através de uma janela de Hanning de 3 pontos, cujos coeficientes são (Rabiner et al., 1975):

$$w_{\text{HANN}}(n) = \begin{cases} 0,25, & n = 0, \\ 0,5, & n = 1, \\ 0,25, & n = 2. \end{cases} \quad (3.2)$$

A figura 3.5 mostra o resultado da aplicação destas técnicas para eliminar os problemas da figura 3.4. Em um primeiro momento, o filtro de mediana e o suavizador linear parecem resolver os problemas resultantes de extração de  $F_0$ . Entretanto, uma investigação mais cautelosa mostra que estas técnicas de processamento podem afetar a amplitude das variações de  $F_0$ . Isto está exemplificado na figura 3.6. A ação do filtro de mediana ceifa os picos e os vales. Além disso, o suavizador linear elimina detalhes que podem ser importantes na análise de regularidade do sinal.

Uma alternativa *empírica* foi então desenvolvida com o objetivo de remover os erros presentes em  $F_0$  sem provocar as distorções do filtro de mediana e do suavizador linear em toda a duração da curva de  $F_0$ . Ela é dividida em três partes, descritas a seguir.

### I) Tratamento de erros grosseiros

Erros onde  $F_0$  é estimada como múltiplo ou submúltiplo do valor real são comuns (Bagshaw, 1994). Para tratar estes erros, a curva de  $F_0$  é percorrida automaticamente, amostra por amostra, verificando se o valor de cada entrada da série de  $F_0$  é maior ou igual ao dobro do valor da amostra imediatamente anterior, ou então, se é menor ou igual à metade do valor da amostra imediatamente anterior. Em caso afirmativo, a amostra atual é substituída pelo valor da mediana entre ela e as amostras imediatamente anterior e imediatamente posterior. O quadro abaixo resume o pseudocódigo desta etapa e a figura 3.7 mostra o resultado de sua aplicação aos problemas ilustrados anteriormente.

**Entrada:** Série temporal de  $F_0$  com  $N$  amostras.  
**Saída:** Série de  $F_0$  com erros grosseiros tratados.

**para**  $n \leftarrow 2$  **até**  $N - 1$  **faça**

se  $F_0(n) \geq 2F_0(n - 1)$  **ou**  $F_0(n) \leq 1/2F_0(n - 1)$  **então**

$F_0(n) \leftarrow \text{mediana}[F_0(n - 1), F_0(n), F_0(n + 1)]$  ;

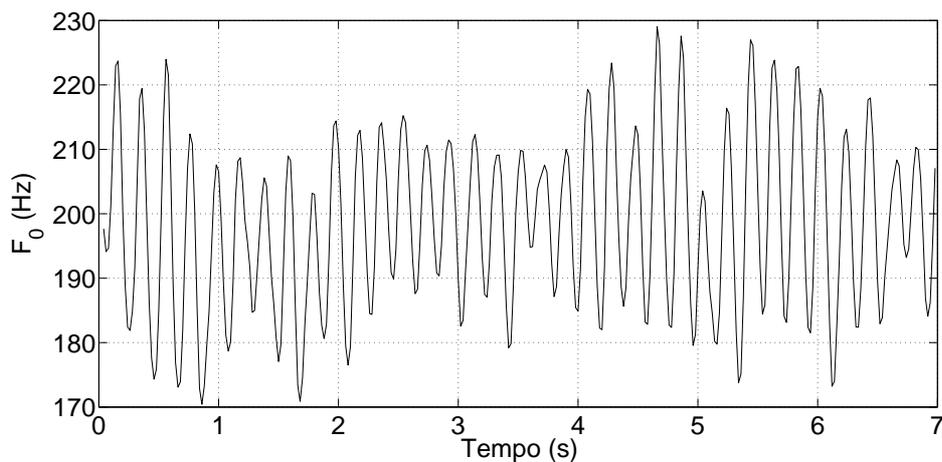
**fim**

**fim**

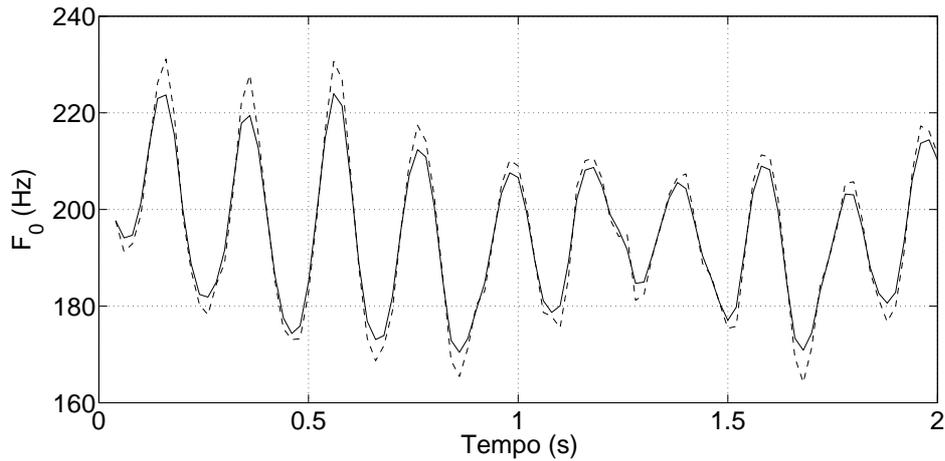
**Algoritmo 1:** Tratamento de erros grosseiros

### II) Tratamento de erros intermediários

Alguns dos erros cometidos pelo algoritmo de extração de  $F_0$  são caracterizados por “espículas” (*spikes*) como as destacadas na figura 3.8. Estes erros são chamados de intermediários e não são corrigidos na etapa anterior porque não têm amplitude suficiente para satisfazerem o critério de erro grosseiro, isto é, não são múltiplos ou submúltiplos do valor real de  $F_0$ .



**Figura 3.5:** Correção dos erros gerados pelo algoritmo de rastreamento de  $F_0$  e mostrados na figura 3.4. Foi usado o filtro de mediana da equação 3.1 associado ao suavizador linear da equação 3.2. A escala vertical desta figura é diferente daquela usada na figura 3.4.



**Figura 3.6:** Trecho entre 0 e 2 segundos da figura 3.5 evidenciando as distorções provocadas pelo filtro de mediana e suavizador linear. As linhas tracejada e cheia representam, respectivamente, o sinal de  $F_0$  antes e depois de serem processados com estas técnicas.

A detecção dos erros intermediários é baseada na diferença entre valores consecutivos da série de  $F_0$ :

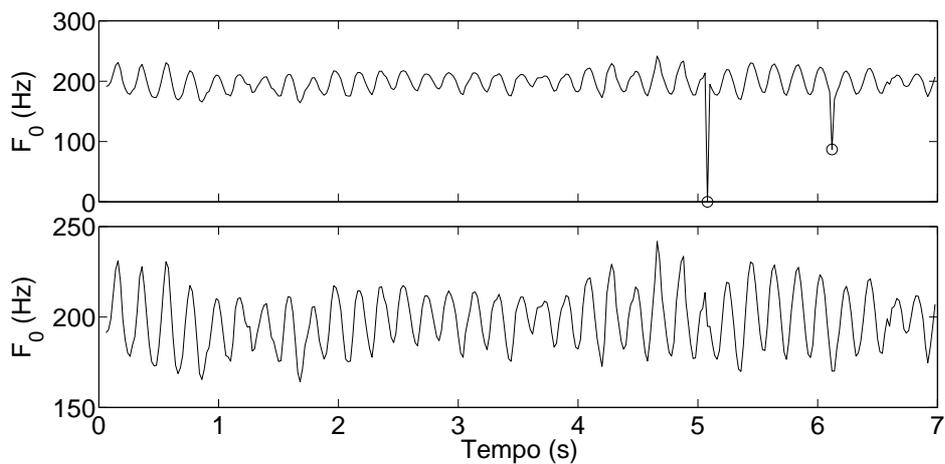
$$d(n) = F_0(n+1) - F_0(n), \quad (3.3)$$

para  $n = 1, 2, \dots, N-1$ , onde  $N$  é o número de amostras da série de  $F_0$ . Amostras de  $F_0$  são identificadas como erros intermediários quando fazem o valor correspondente em  $d(n)$  ultrapassar um limite:

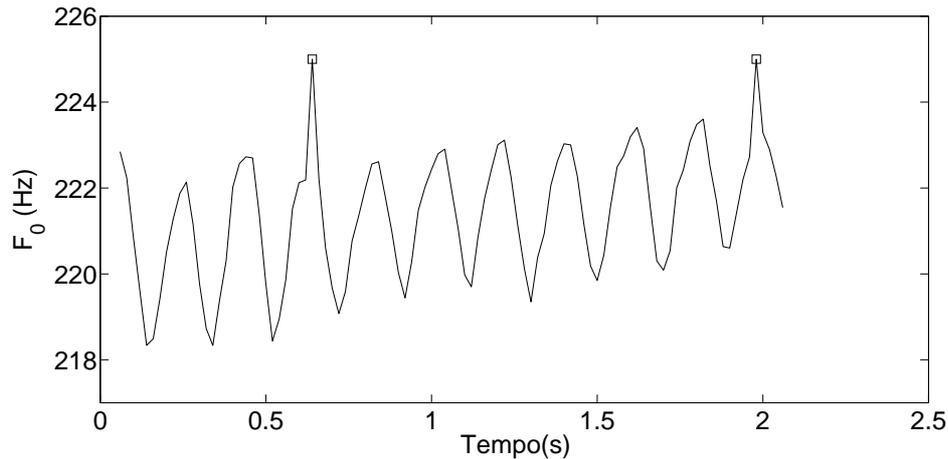
$$|d(n)| > m_d + 2s_d, \quad (3.4)$$

sendo  $m_d$  e  $s_d$  a mediana e o desvio padrão de  $d(n)$  respectivamente.

O tratamento deste tipo de erro é o mesmo da etapa anterior. A entrada  $F_0(n)$  marcada como erro intermediário é substituída pela mediana entre  $F_0(n-1)$ ,  $F_0(n)$ ,  $F_0(n+1)$ . A figura



**Figura 3.7:** Primeira etapa: detecção e correção do erro grosseiro. As circunferências marcam as amostras erradas e que foram corrigidas no gráfico inferior. Compare com as figuras 3.4 e 3.5 sabendo que as escalas verticais das figuras estão diferentes.



**Figura 3.8:** Exemplo de erro intermediário. Os quadrados destacam falhas que não são resolvidas na etapa de tratamento de erros grosseiros.

3.9 mostra o resultado desta etapa de tratamento sobre a série de  $F_0$  do painel inferior na figura 3.7. Observa-se que os dois picos marcados entre 4 e 5 segundos são corrigidos. Além disso, nota-se que algumas amostras, como as duas marcadas antes de 1 segundo, são classificadas, equivocadamente, como erro intermediário. Isto é uma consequência do valor calculado para o limite da equação 3.4. Contudo, como estas amostras não estão localizadas em extremos da curva, elas não são alteradas pela mediana.

```

Entrada: Série temporal de  $F_0$  com  $N$  amostras.
Saída: Série de  $F_0$  com erros intermediários tratados.

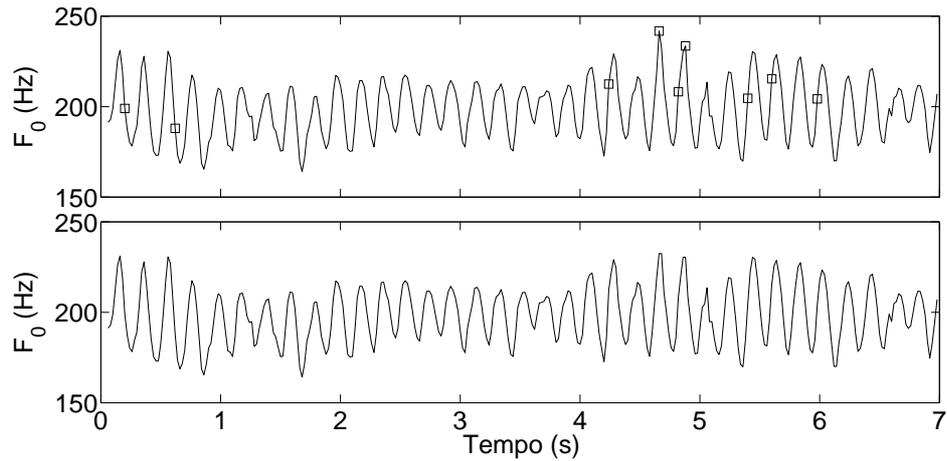
 $d \leftarrow$  vetor com  $N-1$  entradas ;
para  $n \leftarrow 2$  até  $N$  faça
  |  $d(n-1) \leftarrow F_0(n) - F_0(n-1)$  ;
fim
 $m_d \leftarrow \text{mediana}(\underline{r})$  ;
 $s_d \leftarrow \text{std}(\underline{r})$ ;
 $l \leftarrow m_d + 2s_d$ ;
para  $n \leftarrow 1$  até  $N-2$  faça
  | se  $|d(n)| > l$  então
  | |  $F_0(n+1) \leftarrow \text{mediana}[F_0(n), F_0(n+1), F_0(n+2)]$  ;
  | fim
fim

```

**Algoritmo 2:** Tratamento de erros intermediários.

### III) Tratamento de erros finos

O objetivo desta etapa é suprimir o tipo de erro mostrado na figura 3.10. Este erro é caracterizado, em linhas gerais, por certos “desvios” do comportamento da curva de  $F_0$ . O ponto da esquerda marcado com um triângulo na figura 3.10 mostra que a frequência fundamental estava aumentando seu valor quando deu um salto grande de frequência em



**Figura 3.9:** Segunda etapa: detecção e correção do erro intermediário. O erro grosseiro já foi corrigido e os quadrados marcam as amostras corrompidas pelo erro intermediário e que estão corrigidas no gráfico inferior. Compare com a figura 3.7.

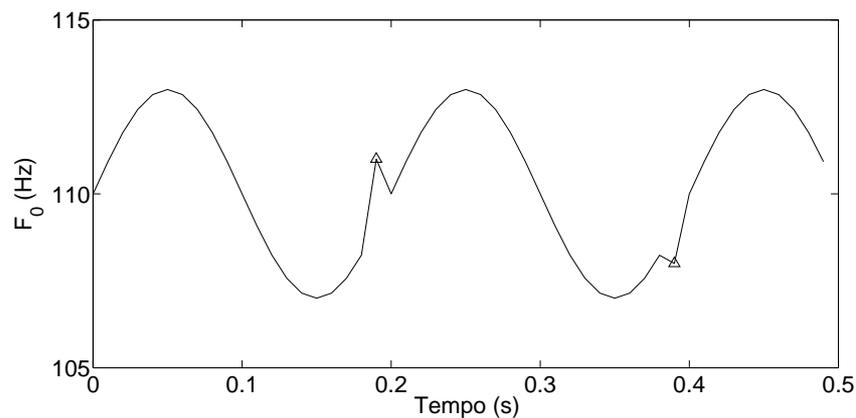
uma *única* amostra para, na amostra seguinte, diminuir seu valor e retomar o ritmo de crescimento anterior.

Este erro não é eliminado pelos critérios anteriores. A estratégia para eliminá-lo é percorrer todo o vetor  $d(n)$  da equação 3.3 verificando-se:

$$\text{sgn}(d(n-1)) \neq \text{sgn}(d(n)) \neq \text{sgn}(d(n+1)), \quad (3.5)$$

sendo que  $\text{sgn}(x)$  foi definido na equação 2.14.

Em caso afirmativo, detecta-se o salto. Para suprimi-lo, usa-se a mesma ideia apresentada acima: substitui-se a amostra problemática pela mediana entre ela e seus vizinhos imediatos. O resultado final está mostrado na figura 3.11.



**Figura 3.10:** Exemplo de erro fino. Os triângulos marcam um tipo de erro que não atende os critérios dos erros grosseiro e intermediário.

**Entrada:** Série de  $F_0$  com  $N$  e  $r(n) = F_0(n) - F_0(n - 1)$  com  $N - 1$  amostras.

**Saída:** Série de  $F_0$  com erros finos tratados.

```

para  $n \leftarrow 2$  até  $N - 2$  faça
  | se  $\text{sgn}[d(n - 1)] \neq \text{sgn}[d(n)]$  e  $\text{sgn}[d(n)] \neq \text{sgn}[d(n + 1)]$  então
  | |  $F_0(n + 1) \leftarrow \text{mediana}[F_0(n), F_0(n + 1), F_0(n + 2)]$ ;
  | fim
fim

```

**Algoritmo 3:** Tratamento de erros finos.

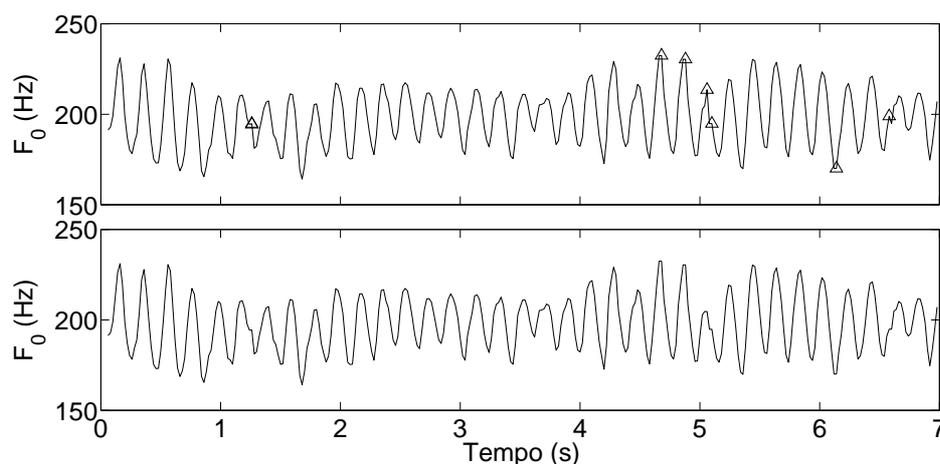
### 3.3.3 Inspeção visual da curva de $F_0$

O método apresentado acima é empírico e, apesar de geralmente resolver os problemas, não corrige *todos* os casos. Por isso, o resultado final é inspecionado a fim de verificar se a curva está ou não em condições de ser analisada. É importante enfatizar que isto é feito porque o objetivo não é de realizar uma análise totalmente automática – o que não é necessariamente simples em vozes disfônicas –, mas é garantir que, dado conjuntos de sinais devidamente preparados, a análise esteja correta.

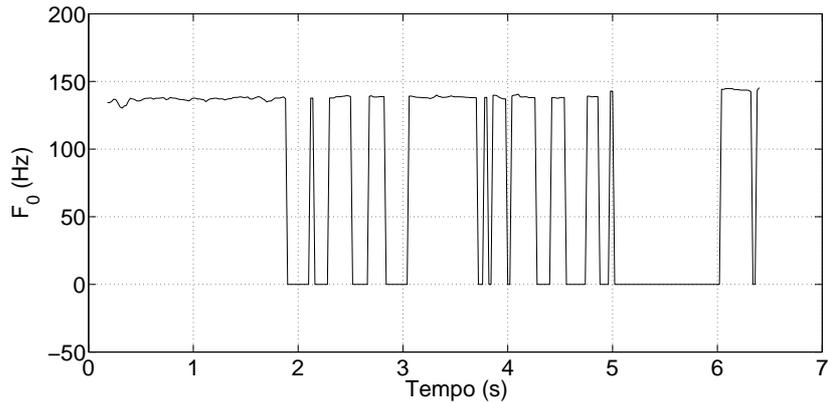
A figura 3.11 inferior exemplifica uma curva que teve todos os seus erros eliminados na etapa de pós-processamento e está pronta para ser analisada. Já a curva da figura 3.12 ainda apresenta diversos erros, mesmo já tendo sido pós-processada.

### 3.3.4 Segmentação manual da amostra

Caso a inspeção visual de  $F_0$  ainda indique a existência de erros, segmenta-se manualmente o sinal original descartando-se todo o trecho que causou problemas na detecção automática de  $F_0$ . Alguns exemplos são dados na figura 3.13. Como visto, caso o sinal segmentado tenha mais de 1 segundo de duração, ele era mantido para análise de Fourier e de regularidade da



**Figura 3.11:** Última etapa: detecção e correção do erro fino. Os erros grosseiro e intermediário já foram corrigidos e os triângulos marcam as amostras corrompidas pelo erro fino e que estão corrigidas no gráfico inferior.

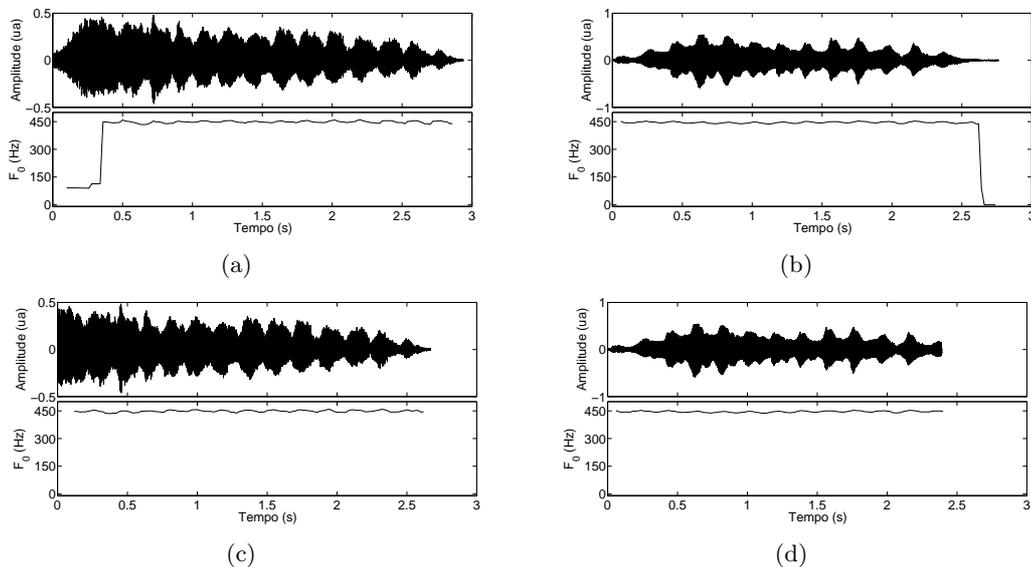


**Figura 3.12:** Curva de  $F_0$  extraída de uma amostra de voz real ainda apresenta diversos erros, mesmo já tendo sido pós-processada. A solução foi segmentar a amostra e descartar a parte fragmentada não corrigida pelo pós-processamento ( $t > 1,8$  s).

curva de  $F_0$ . Antes de detalhar estas análises, será descrito, abaixo, o processo de extração da envoltória do sinal para análise AM.

### 3.3.5 Extração automática da envoltória da amostra

A extração da envoltória do sinal de áudio original só é realizada depois que a inspeção final da curva de  $F_0$  não aponta nenhum problema. Desta forma, garante-se que tanto a envoltória quanto a série de  $F_0$  sejam extraídas do mesmo segmento de áudio e que as



**Figura 3.13:** Amostras de vibrato produzidos em um contrabaixo acústico. A extração de  $F_0$  apresentou falhas em (a) devido a instabilidades do sinal de áudio e em (b) por causa da pequena amplitude do sinal no final da amostra. Após descartar aproximadamente 30 ms do início da figura (a) e do fim do sinal da figura (b), o algoritmo de rastreamento de frequência fundamental consegue operar normalmente, como indicado em (c) e (d).

informações destas duas curvas possam ser comparadas.

Para extrair a envoltória do sinal de voz foi usado o algoritmo proposto por Qi e Weinberg (1994). Este algoritmo descreve o sinal de voz através de um modelo de modulação em amplitude definido como:

$$s_{vAM}(n) = s_e(n) \cdot s_v(n), \quad (3.6)$$

onde  $s_e(n)$  é o sinal da envoltória e  $s_v(n)$  é o sinal de voz sem perturbações na amplitude.

A multiplicação de  $s_e(n)$  por  $s_v(n)$  na equação 3.6 pode ser transformada em soma através da função logaritmo.

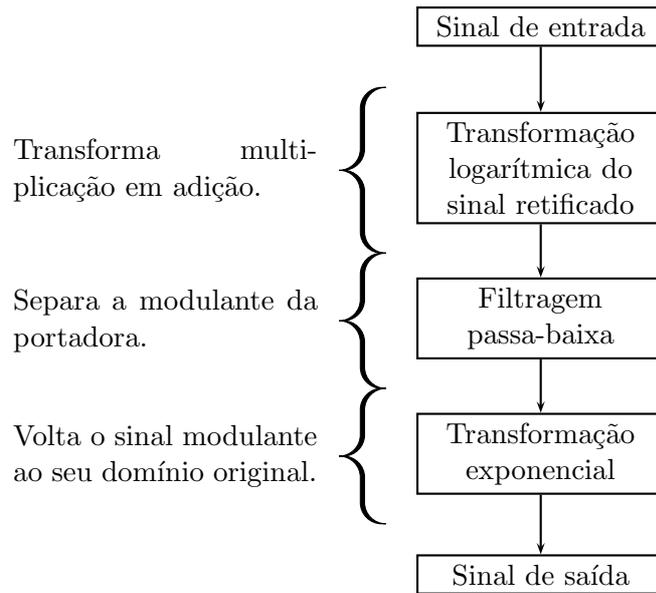
$$\begin{aligned} |s_{vAM}(n)| &= |s_e(n) \cdot s_v(n)| \\ |s_{vAM}(n)| &= |s_e(n)| \cdot |s_v(n)| \\ \log [|s_{vAM}(n)|] &= \log [|s_e(n)| \cdot |s_v(n)|] \\ \log [|s_{vAM}(n)|] &= \log [|s_e(n)|] + \log [|s_v(n)|]. \end{aligned} \quad (3.7)$$

Na equação 3.7, o espectro da envoltória aparece somado ao espectro do sinal de voz; não mais convoluído como na equação 3.6. Assim, desde que os espectros de  $\log [|s_e(n)|]$  e  $\log [|s_v(n)|]$  ocupem faixas diferentes no espectro, é possível separá-los através de uma filtragem.

Para contornar problemas com argumentos negativos e nulos em  $\log(\ )$ , o sinal  $s_{vAM}$  é retificado e os zeros restantes são substituídos por uns. A troca de 0 por 1 em uma amostra de 16 bits muda em  $1/2^{16} \approx 1,53 \cdot 10^{-5}$  alguns poucos pontos da portadora, com efeito desprezível na envoltória do sinal (Qi e Weinberg, 1994).

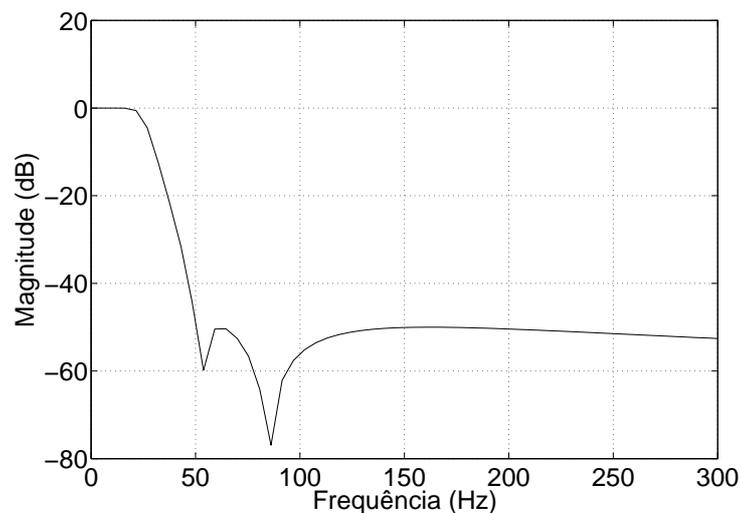
Esta estratégia para isolar a envoltória e o sinal de voz faz parte de uma técnica mais geral, chamada processamento homomórfico (Oppenheim e Schaffer, 1975, cap. 10). Nesta técnica, uma relação não linear entre variáveis é mapeada para outro domínio de forma que, no novo domínio, as relações entre variáveis sejam lineares e possam ser usadas técnicas de processamento linear. No algoritmo de Qi e Weinberg (1994), mostrado na figura 3.14, o mapeamento não linear é realizado por um logaritmo, o processamento é uma filtragem passa-baixas e o mapeamento de retorno é feito com uma função exponencial.

Um exemplo de funcionamento do algoritmo de Qi e Weinberg (1994) está mostrado na figura 3.16. O sinal de voz sintetizado da figura 3.16(a) foi modulado em AM, com índice de modulação  $k = 0,4$ , por uma senóide de 10 Hz. O conteúdo espectral da modulante aparece replicado ao lado de cada um dos harmônicos do sinal de voz na figura 3.16(d) e pode ser recuperado como visto acima. Para tal, primeiramente aplica-se o logaritmo sobre o sinal modulado da figura 3.16(c), fazendo com que os espectros de  $s_e(n)$  e de  $s_v(n)$  apareçam em diferentes faixas de frequência (figura 3.16(f)). Para extrair o espectro da envoltória filtra-se o sinal  $\log [|s_{vAM}(n)|]$  com um passa-baixas. A figura 3.15 mostra a resposta em frequência do passa-baixas utilizado para este fim. A resposta de fase do filtro não foi mostrada por ser irrelevante, uma vez que a filtragem foi feita sem atraso de fase (Mitra, 1998, seções 4.3.5 e 8.2), usando a função *filtfilt* do Matlab. Uma vez que o espectro da envoltória está isolado,

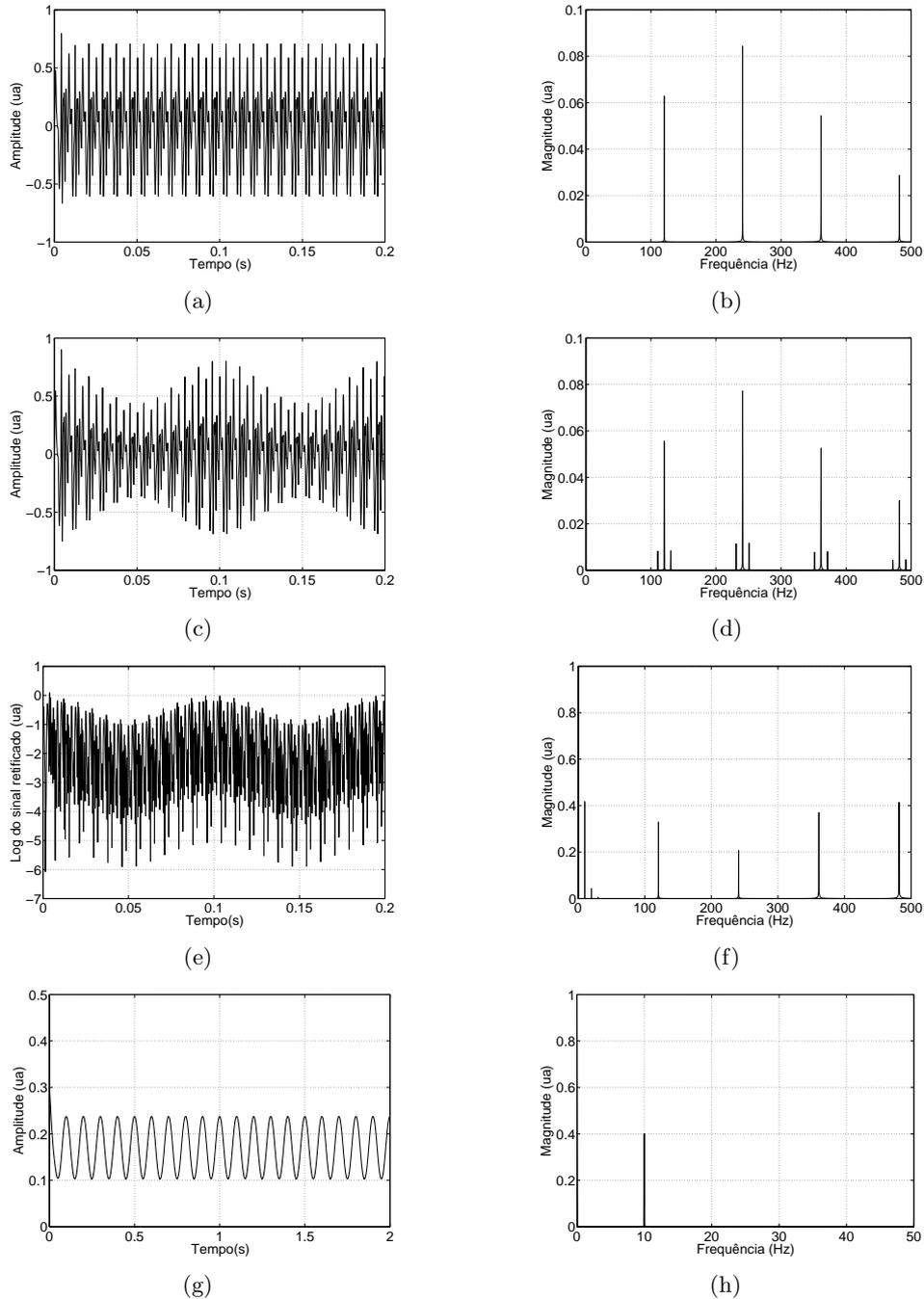


**Figura 3.14:** Diagrama de blocos da demodulação AM usando transformação homomórfica.

deve-se transformá-lo de volta à sua forma original. Este é o último passo do algoritmo e consiste em realizar uma transformação exponencial a fim de desfazer os efeitos da operação logarítmica realizada na primeira etapa. As figuras 3.16(g) e 3.16(h) mostram o resultado final do algoritmo.



**Figura 3.15:** Resposta em frequência de um filtro passa-baixas de Chebyshev tipo II de quinta ordem, com atenuação de pelo menos 50 dB a partir de 50 Hz.

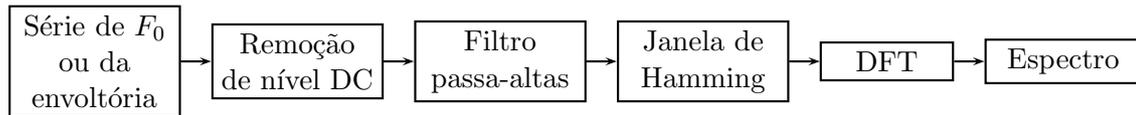


**Figura 3.16:** Sinal de voz (a) sintetizado sem nenhum tipo de perturbação e (c) modulado em amplitude por uma senóide de 10 Hz de frequência e usando-se um índice de modulação  $k = 0,4$ . O espectro (b) do sinal sem perturbação é formado pela série harmônica com fundamental em 120 Hz. No espectro (d) do sinal modulado em AM, cada componente da série harmônica se comporta como uma portadora modulada em amplitude. (e) Logaritmo do sinal modulado retificado. (f) O espectro da envoltória aparece isolado do espectro da série harmônica do pulso glótico. Compare com a figura (b). (g) Envoltória extraída do sinal modulado através da transformação homomórfica. (h) Espectro da envoltória normalizado pelo nível médio, com a raia na frequência da modulante, 10 Hz.

### 3.3.6 Análise espectral das curvas de $F_0$ e de envoltória

#### Curva de $F_0$

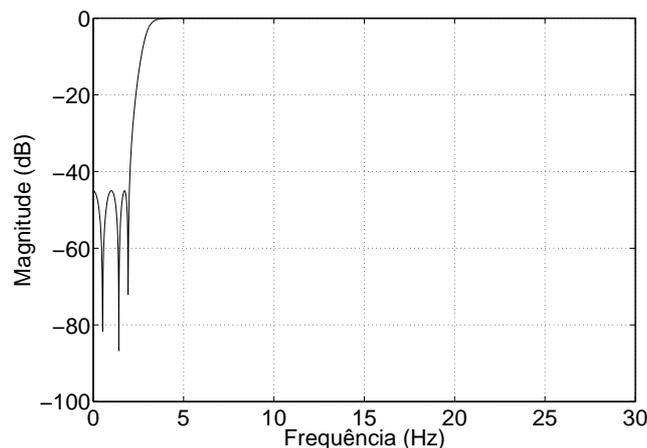
A análise de Fourier das séries de  $F_0(n)$  e da envoltória  $s_e(n)$  tem por objetivo determinar a amplitude de cada componente espectral nestes sinais. Para isto, estas séries são manipuladas com a transformada discreta de Fourier (DFT). Antes porém, elas são pré-processadas (figura 3.17) para minimizar o surgimento de franjas espectrais (artefatos).



**Figura 3.17:** Diagrama de blocos da análise de  $F_0$ . Os quatro primeiro blocos fazem parte do pré-processamento para minimizar os artefatos da transformada discreta de Fourier.

Primeiramente, remove-se o valor médio (DC) da curva analisada, uma vez que ele, geralmente, tem magnitude muito maior que as outras flutuações. Em segundo lugar, flutuações lentas, fora da faixa de interesse, são eliminadas através de uma filtragem passa-altas (figura 3.18) sem distorção de fase (Mitra, 1998, seções 4.3.5 e 8.2). No último passo antes da DFT, o sinal é multiplicado por uma janela de Hamming para suavizar transições abruptas de amplitude nas bordas.

Na etapa de DFT, caso a série temporal tenha menos que 500 pontos, realiza-se *zero-padding*. Isto garante uma resolução espectral mínima de 0,1 Hz, pois tanto a curva de  $F_0$  quanto a de envoltória estão sempre amostradas a 50 Hz. O cálculo da DFT é realizado



**Figura 3.18:** Resposta em frequência de um filtro passa-altas de Chebyshev tipo II de sexta ordem, com atenuação de pelo menos 45 dB para frequências menores que 2 Hz. Usado para remover as flutuações de baixa frequência na série estudada.

diretamente a partir de sua definição:

$$X(k) = \frac{1}{N} \sum_{n=0}^{n=N-1} x(n) e^{-j2\pi \frac{k\Delta f}{f_s} n}, \quad k = 0, 1, \dots, N-1, \quad (3.8)$$

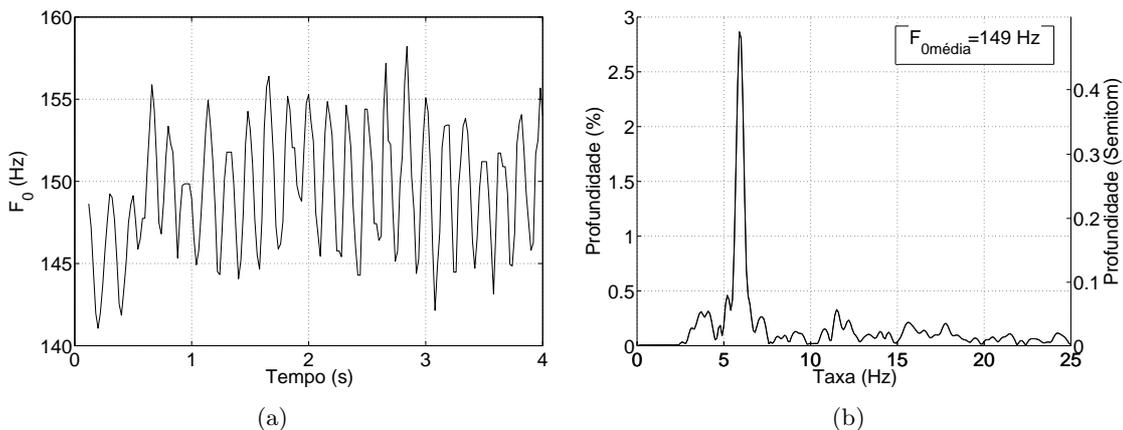
onde  $\Delta f = 0,1$  Hz,  $f_s = 50$  Hz é a frequência de amostragem da série  $x(n)$ , a qual contém  $N$  amostras. Este cálculo poderia ser feito usando-se o algoritmo de FFT (*Fast Fourier Transform*). Entretanto, como as séries de interesse têm, em geral, aproximadamente 2 segundos de duração, a DFT é calculada com poucos pontos e a FFT não produz um ganho substancial de tempo de processamento.

O espectro de  $F_0(n)$  expressa a amplitude de cada um dos componentes da frequência fundamental. A figura 3.19 mostra um exemplo para o caso de um vibrato produzido por um cantor sertanejo. Observa-se que neste espectro há, basicamente, uma só raia localizada em torno de 6 Hz e com quase meio semitom ( $\approx 3\%$ ) de extensão. Isto reflete o fato que a curva de  $F_0$  é predominantemente uma senóide de frequência 6 Hz com amplitude de aproximadamente  $5/149 \approx 3\%$ , onde 149 Hz é o valor médio de  $F_0$  no intervalo.

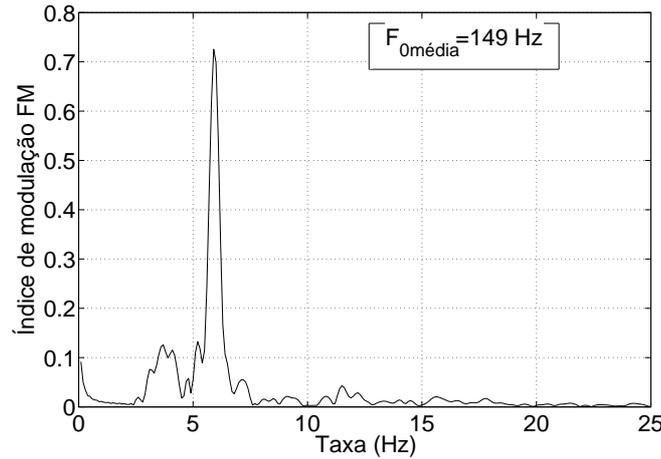
A curva de  $F_0(n)$  também pode ser descrita por meio de modulação em frequência (FM), usando a análise de Fourier. Neste modelo, a frequência média de  $F_0$  faz o papel de portadora que, modulada em frequência, oscila de acordo com o sinal modulante. A relação entre a amplitude da modulante e a sua frequência define o índice de modulação FM,  $\beta$  (equação 2.4). No caso da figura 3.19, por exemplo, o sinal modulante é composto basicamente por uma só frequência, em torno de 6 Hz, e faz a portadora de 149 Hz variar entre 155 e 145 Hz, de modo que, de forma aproximada:

$$\beta_6 = \frac{\Delta f}{f_m} \Big|_{f_m=6} \approx \frac{155-145}{6} \approx 0,8. \quad (3.9)$$

Na equação acima, o índice de modulação FM foi escrito com o subscrito 6 para indicar



**Figura 3.19:** Espectro da fundamental extraída de um sinal de vibrato de um cantor sertanejo. (a) Curva de  $F_0$ . (b) Espectro de  $F_0$ . O eixo vertical indica amplitude (i.e., extensão) de cada componente harmônico (i.e., taxa) descrito no eixo horizontal.



**Figura 3.20:** Índice de modulação FM para cada componente do vibrato da figura 3.19. O pico corresponde à divisão da extensão de aproximadamente 5 Hz pela taxa de 6 Hz, ou seja,  $5/6 \approx 0,8$ . A elevação na região onde a taxa é próxima de 0 é consequência da divisão realizada para calcular o índice FM.

que este é o valor de  $\beta$  da modulante de componente 6 Hz. Os índices das outras frequências que compõem a curva da figura 3.19(a) de maneira análoga, ou seja,

$$\beta_i = \frac{\Delta f}{f_m} \Big|_{f_m=i} = \frac{\text{amplitude da } i\text{-ésima componente}}{\text{frequência da } i\text{-ésima componente}} \quad (3.10)$$

e estão mostrados na figura 3.20.

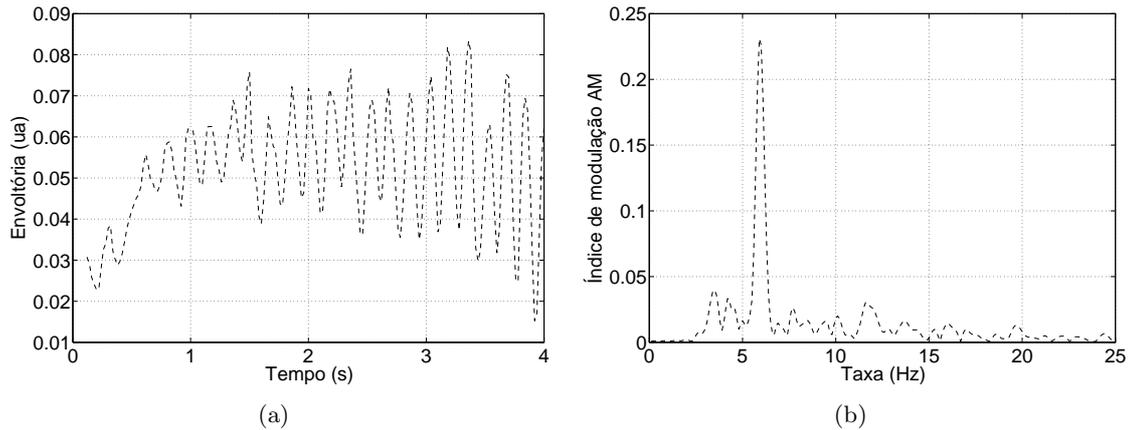
### Envoltória

O espectro da envoltória exhibe a amplitude de cada componente harmônica que modula o contorno do sinal de voz. Este espectro é pouco explorado na literatura porque, apoiando-se no modelo harmônico-formante (Horii e Hata, 1988), muitos autores estudam apenas a curva de  $F_0$  e aceitam que a informação da amplitude da envoltória é redundante. Porém, dada a dificuldade de extração de  $F_0$ , especialmente em voz disfônica, este espectro será investigado com o objetivo de saber se ele tem informações que possam ser usadas quando o espectro da fundamental não estiver disponível.

A maneira escolhida para representar as informações da amplitude de cada raia do espectro da envoltória foi através do índice de modulação AM,  $k$  (equação 2.2). Este índice mede a variação provocada na envoltória do sinal em relação a amplitude do sinal não modulado, ou seja:

$$k_i = \frac{\text{amplitude da } i\text{-ésima componente}}{\text{amplitude do sinal não modulado}}. \quad (3.11)$$

A figura 3.21 mostra a envoltória e o índice AM de cada um de seus componentes do mesmo sinal cuja curva de  $F_0$  foi analisada na figura 3.19. Nota-se por simples comparação entre as figuras 3.21(b) e 3.20 que existem semelhanças entre os índices de modulação AM e FM, que serão exploradas mais à frente.



**Figura 3.21:** (a) Envolvória de um vibrato produzido por um cantor sertanejo. (b) Índice de modulação AM para cada componente harmônico da envoltória.

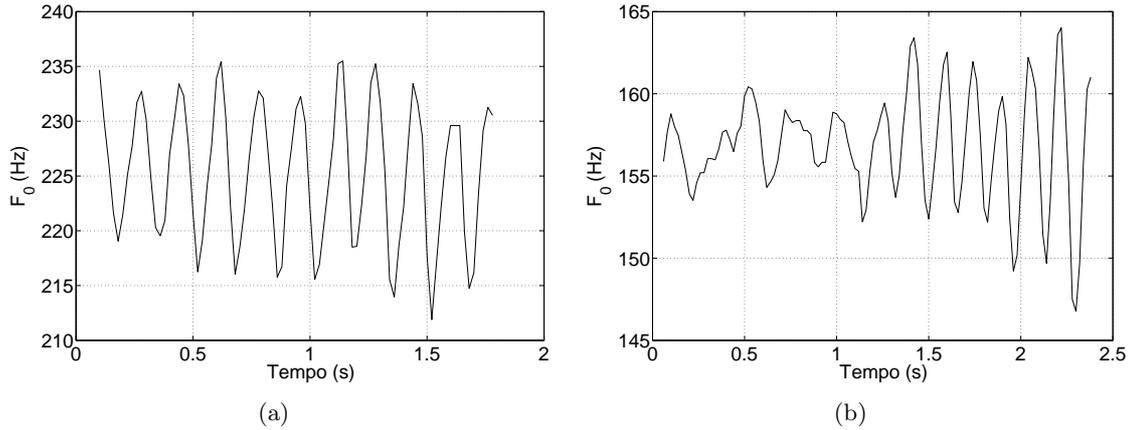
### 3.3.7 Análise automática de regularidade de $F_0$

#### Função de autocorrelação

Como discutido anteriormente, o método de análise da seção 3.3.6 é eficiente para analisar o espectro das séries da frequência fundamental. Porém, não diz nada sobre o comportamento temporal dos componentes harmônicos de  $F_0$ . O espectrograma seria capaz de resolver esta dificuldade, mas, como mostrado na seção 2.5, ele não apresenta bom desempenho com sinais de “curta duração”, que são a grande maioria dos sinais de interesse.

A análise do comportamento temporal da curva de  $F_0$  tem o objetivo de investigar sua *regularidade* (figura 3.22). Este aspecto é pouco estudado, mesmo sendo relevante perceptivamente (Sundberg, 1995). Verifica-se, por exemplo que, quanto mais regular é a curva de  $F_0$  do vibrato, melhor a sua qualidade estética (Diaz e Rothman, 2003). Os poucos trabalhos que publicaram maneiras automatizadas de medir a regularidade usaram o sinal analítico (Arroabarren et al., 2002) ou filtragem LPC, *Linear Predictive Coding* (Diaz e Rothman, 2003). Em todos os casos os autores sempre estudaram a variação instantânea da taxa e da profundidade de  $F_0$ . Além disso, reportaram resultados apenas para amostras de vibrato no canto lírico.

Neste trabalho, deseja-se investigar, de forma quantitativa, a regularidade em uma classe maior de sinais, incluindo tremor na voz, vibrato no canto sertanejo e lírico. Para isto, *não* será medida a evolução temporal da taxa e da profundidade de  $F_0$ , pois, ao fazer este tipo de medição, assume-se, implicitamente, que a forma da curva de  $F_0$  seja senoidal. Esta hipótese é geralmente aceita para o caso de vibrato em cantores líricos de renome, mesmo que algumas publicações apontem algumas controvérsias (Horii, 1989b). Todavia, não há evidências que indiquem que a forma da onda da frequência fundamental no tremor patológico ou no vibrato de cantores inexperientes seja senoidal. Por isso, o método proposto “quantifica” a regularidade realizando medidas diretamente sobre a forma da onda de  $F_0(n)$  por meio da função de autocorrelação (FAC).



**Figura 3.22:** (a) Série de  $F_0$  com comportamento quase regular do início ao fim. (b) Série de  $F_0$  irregular inicialmente. A partir de aproximadamente 1,5 passa a ter um comportamento mais estável, aumentado a semelhança entre ciclos consecutivos. As escalas verticais são diferentes.

A função de autocorrelação é uma medida de semelhança entre as amostras de uma série temporal em função da separação  $K$  entre elas. Para o caso de uma série discreta  $x(n)$ , a FAC é matematicamente definida por:

$$\begin{aligned} r_{xx}(K) &= \mathcal{E}\{[x(n) - \mu][x(n - K) - \mu]\} \\ &= \lim_{N \rightarrow \infty} \left\{ \frac{1}{2N + 1} \sum_{n=-N}^N [x(n) - \mu][x(n - K) - \mu] \right\}, \end{aligned} \quad (3.12)$$

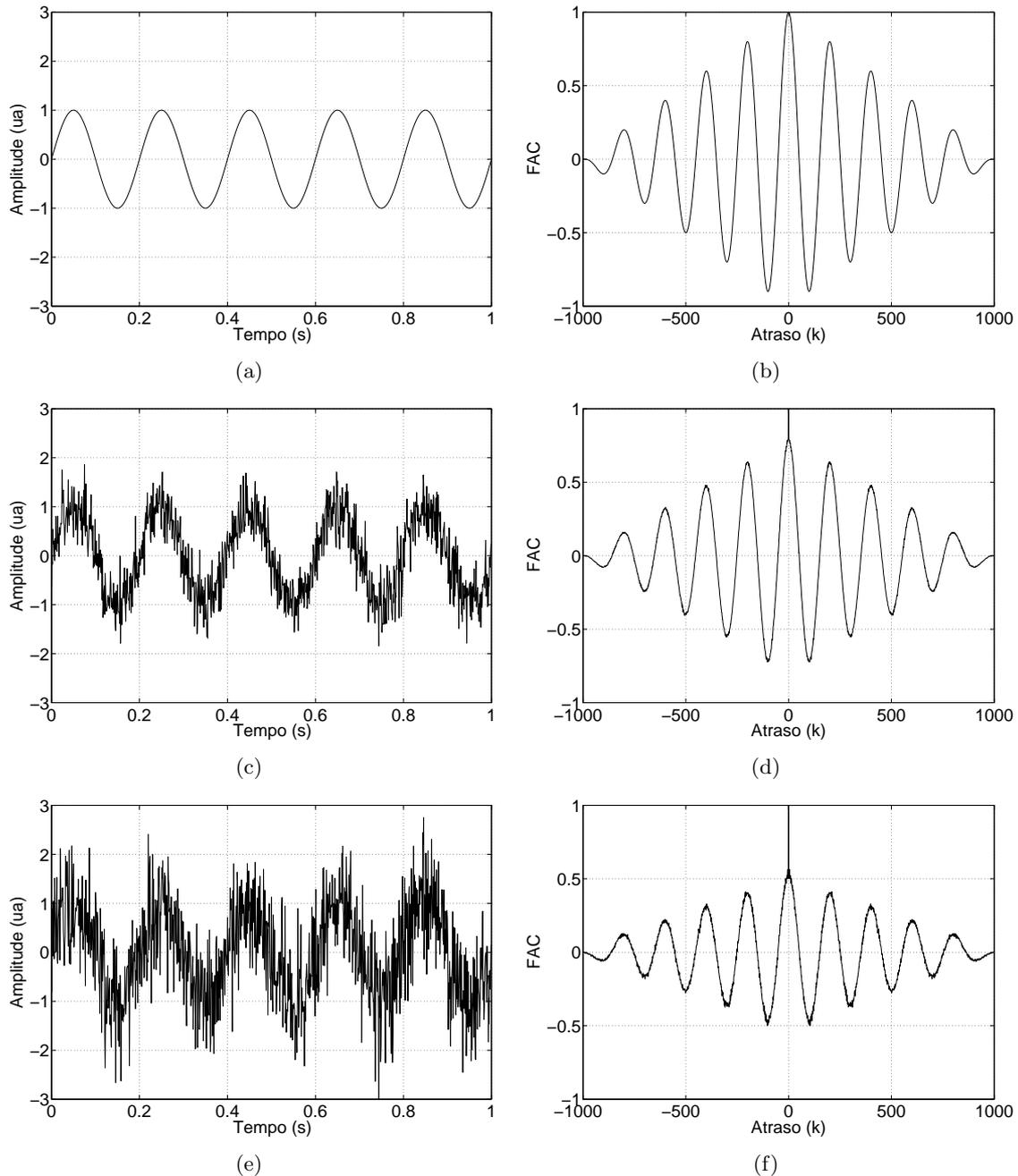
em que  $K$  é o chamado “atraso”,  $\mu$  é o valor esperado de  $x(n)$  e assume-se que  $x(n)$  seja ergódico, de modo a substituir a esperança matemática  $\mathcal{E}$  pela média temporal.

Quando a série  $x(n)$  é finita, de tamanho  $N$ , e sua média  $\mu$  não é conhecida, a FAC é aproximada por:

$$\tilde{r}_{xx}(K) = \frac{1}{N} \sum_{n=1}^{N-K} [x(n) - \bar{x}(n)][x(n - K) - \bar{x}(n)], \quad K = 0, 1, \dots, N - 1 \quad (3.13)$$

onde  $\bar{x}(n) = \frac{1}{N} \sum_{n=1}^N x(n)$  estima a média  $\mu$  e  $\tilde{r}_{xx}(K) = \tilde{r}_{xx}(-K)$ . Um exemplo é dado na figura 3.23(b).

O valor máximo de  $\tilde{r}_{xx}(K)$  acontece em  $K = 0$ . É comum normalizar a expressão 3.13 por este valor, a fim de mapear todos os valores da FAC na faixa  $[-1, 1]$ . Uma autocorrelação de +1 representa um casamento perfeito em fase entre  $x(n)$  e  $x(n - K)$ , ou seja, as formas de onda de  $x(n)$  e de  $x(n - K)$  são iguais. Já a correlação de -1 indica um casamento perfeito em anti-fase, isto é,  $x(n) = -x(n + K)$ . O valor 0 indica que os sinais não tem semelhanças e valores entre 0 e 1 e entre -1 e 0 indicam que o casamento entre os sinais não é perfeito, mas eles têm algumas semelhanças.



**Figura 3.23:** (a) Sinal senoidal de 5 Hz amostrado a 1 kHz. (b) FAC do sinal senoidal de 5 Hz. Quando o atraso é igual ao período da onda, senos em fase são sobrepostos e a FAC é máxima. Quando o atraso é de meio período, as duas ondas estão em anti-fase e a FAC é mínima. Quando o atraso é de um quarto de ciclo, os dois sinais são ortogonais e a FAC vale 0. A FAC estimada tem um aspecto triangular porque o estimador usado (equação 3.13) é tendencioso. (c) Sinal senoidal de 5 Hz com ruído aditivo branco gaussiano, relação sinal ruído de 6 dB. (d) FAC do sinal senoidal corrompido pelo ruído. Apesar do ruído corromper a forma do sinal, ainda há algumas semelhanças entre os ciclos adjacentes. A FAC quantifica esta semelhança residual. (e) Sinal senoidal de 5 Hz com ruído aditivo branco gaussiano, relação sinal ruído de 0 dB. (f) A presença mais forte do ruído branco diminui ainda mais as semelhanças existentes no sinal senoidal original, fazendo o valor da FAC diminuir.

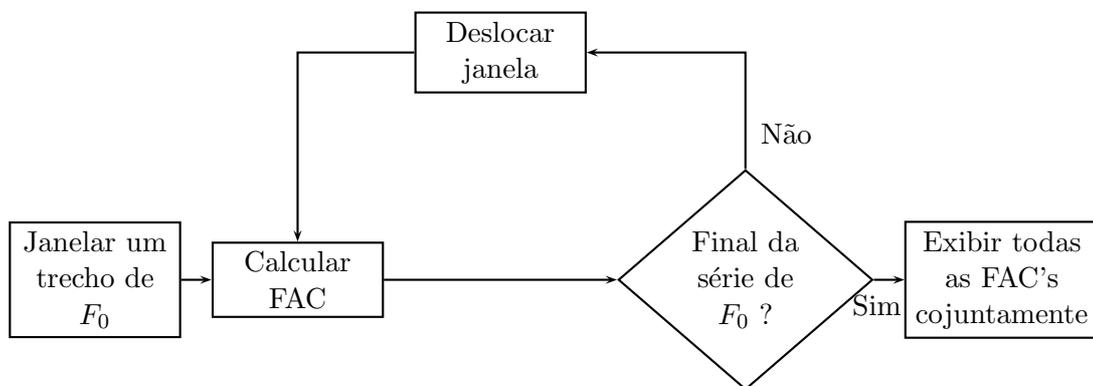
Como mostrado em diversos trabalhos da área (e.g. Papoulis (1988, pag. 361)), a estimativa da FAC usando a equação 3.13 é tendenciosa, como ilustrado na figura 3.23(b). Apesar disso, este estimador é bastante empregado, pois, além de ser consistente ( $\tilde{r}_{xx}(K) \rightarrow r_{xx}(K)$  quando  $N \rightarrow \infty$ ), apresenta variância menor que a versão não-tendenciosa.

A figura 3.23(d) mostra que a FAC do sinal senoidal severamente corrompido por ruído é periódica com a mesma frequência do seno. Na verdade, esta propriedade da função de autocorrelação não se limita a senóides e a FAC de qualquer sinal periódico é também periódica e com o mesmo período (Papoulis, 1988). Esta propriedade, juntamente com o fato de a FAC conseguir detectar periodicidades e fornecer um índice que é função da semelhança entre trechos do sinal são explorados pelo algoritmo proposto para análise de regularidade.

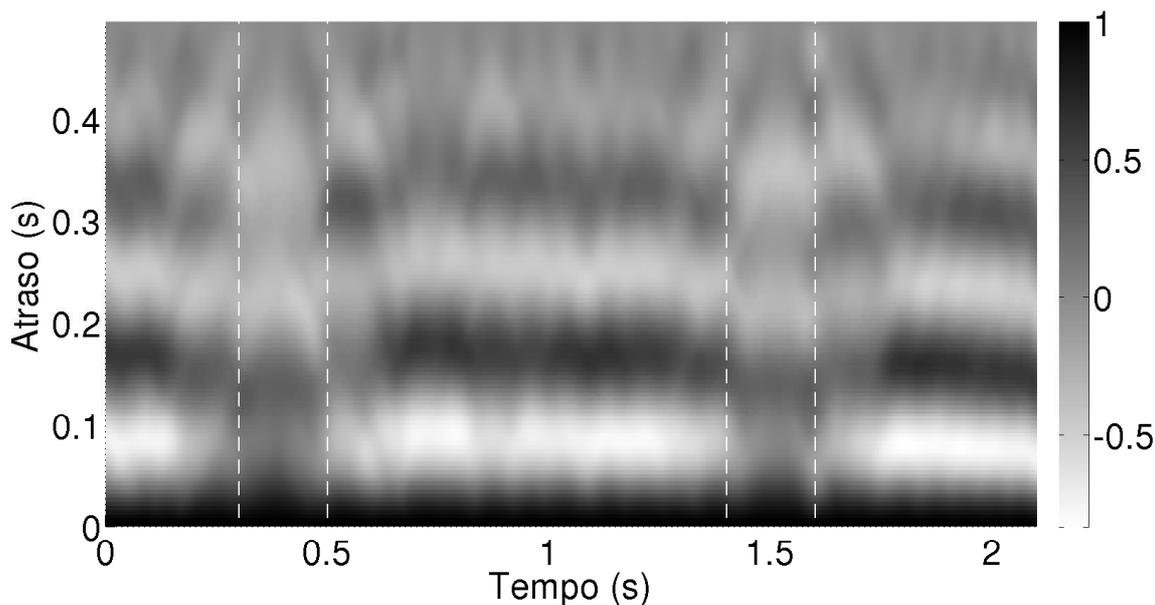
### Algoritmo de regularidade

O primeiro estágio do algoritmo de regularidade (figura 3.24) é separar um trecho da série de  $F_0$  com uma janela retangular. A função desta janela é isolar alguns ciclos da série de  $F_0$  para, em uma etapa seguinte, calcular a FAC do trecho janelado e medir a similaridade entre os ciclos. O tamanho da janela deve atender à seguinte relação de compromisso: se ela for muito curta, apenas alguns pontos são cercados e não há comparação entre ciclos adjacentes; se for muito grande, vários ciclos são janelados e os detalhes do comportamento temporal são perdidos. Neste trabalho, o valor adotado para o tamanho da janela foi de 500 ms porque: (i) as taxas do vibrato e do tremor são, normalmente, maiores que 4 Hz, de modo que a janela englobará pelo menos 2 ciclos; (ii) há um detalhamento adequado da evolução temporal, já que dificilmente esta janela englobará 4 ciclos ou mais, pois as taxas de vibrato e tremor geralmente são menores que 8 Hz.

Depois de calcular a FAC de um trecho, a janela é deslocada de um determinado número de amostras e a nova região janelada é analisada. Foi adotada uma estratégia conservadora e o deslocamento da janela foi feito ponto por ponto. Como os valores da série de  $F_0$  não alteram bruscamente de uma amostra para outra, é esperado que haja pequena variação entre as FACs de janelas adjacentes.



**Figura 3.24:** Diagrama de blocos da análise de regularidade.



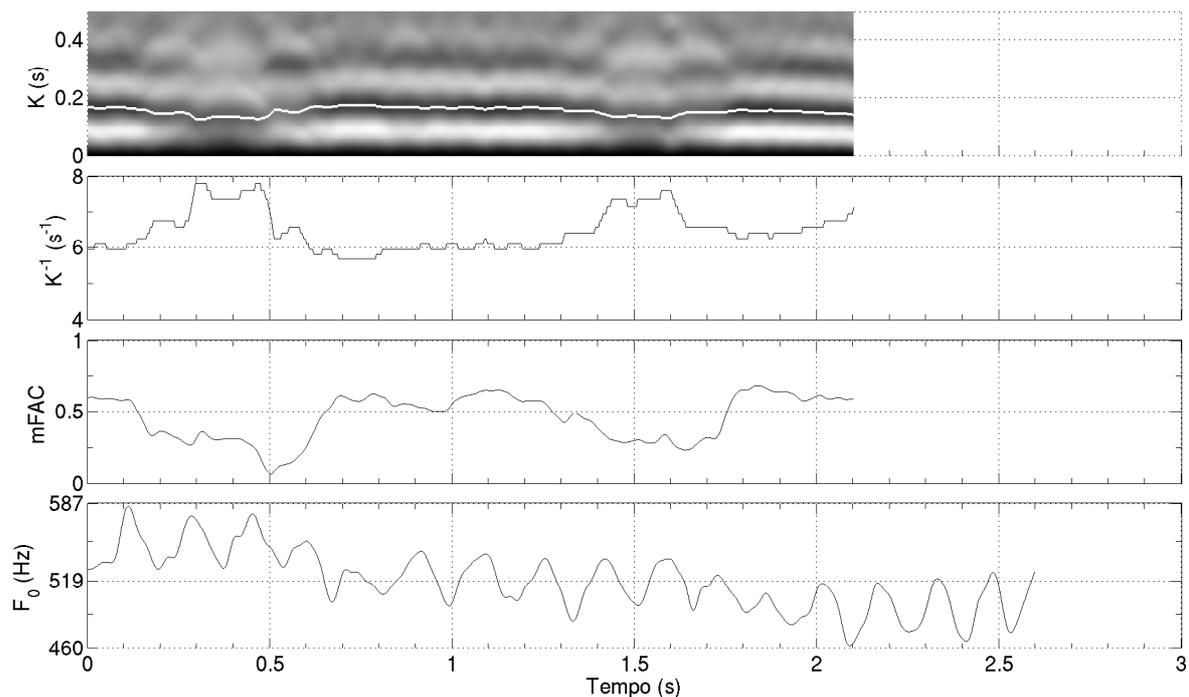
**Figura 3.25:** Resultado produzido pelo algoritmo de análise de regularidade. Para cada ponto  $n$  do eixo de tempo, plota-se o resultado da FAC da janela que cobre o trecho entre  $n$  e  $n + 500$  ms. O valor da autocorrelação é dado pela escala de tons cinza e o atraso é medido em segundos. As regiões marcadas estão relacionadas a comportamentos irregulares da curva de  $F_0$ .

O processo de janelar, calcular FAC e deslocar janela é repetido até atingir o final da série de  $F_0$ . Depois disso, todas as FACs calculadas são exibidas conjuntamente.

O método está ilustrado na figura 3.25, com a análise de regularidade de uma amostra de vibrato de um cantor lírico. A figura mostra máximos (faixa horizontal escura) próximos aos atrasos  $K$  de 0,2 s e 0,4 s durante praticamente todo o tempo. Isto é um indicativo de que os ciclos da série de  $F_0$  têm duração aproximadamente constante, de  $\approx 0,2$  s. Há, também, duas regiões delimitadas por linhas verticais tracejadas, próximas a 0,4 s e 1,5 s, em que as FACs têm máximos menos evidentes. Estas regiões apontam instantes em que ciclos adjacentes da série de  $F_0$  têm menor semelhança entre si, indicando uma irregularidade momentânea na curva melódica.

O comportamento temporal do máximo da FAC de  $F_0$  carrega informações sobre a regularidade da frequência fundamental. Para explicitar este fato, o primeiro máximo com atraso maior que zero ao longo do tempo na figura 3.25 foi destacado com uma linha branca na figura 3.26. Além disso, esta linha foi traçada separadamente em outros dois painéis, um com a informação do máximo das FACs (mFAC) ao longo do tempo e outro com a evolução temporal do inverso do atraso,  $K^{-1}$ , medido em  $s^{-1}$ .

Na figura 3.26, a variação do valor da mFAC e de  $K^{-1}$  ao longo do tempo  $t$  é explicada pelas diferenças na forma da onda da frequência fundamental. Inicialmente, a mFAC começa valendo pouco mais que 0,5, indicando uma boa correlação entre os primeiros 500 ms da curva melódica. Neste trecho, a curva melódica tem 3 ciclos espaçados quase uni-



**Figura 3.26:** Resultado produzido pelo algoritmo de análise de regularidade. O painel superior mostra a mesma imagem da figura 3.25 com a linha branca destacando o primeiro máximo de atraso não nulo em cada FAC. Esta linha está traçada separadamente nos dois painéis abaixo, um para o valor do máximo das FACs (mFAC) e outro exibindo o valor  $K^{-1}$ . O painel inferior mostra a curva de  $F_0$  que foi analisada.

formemente, de modo que  $K^{-1} \approx 3/0,5 = 6 \text{ s}^{-1}$ . Porém, por volta de  $t = 200 \text{ ms}$ , o máximo da função de autocorrelação começa a diminuir, indicando que uma janela da curva melódica com baixa semelhança entre ciclos começa a ser analisada. Em torno de  $t = 400 \text{ ms}$ ,  $K^{-1} \approx 8 \text{ s}^{-1}$ , mostrando que o espaçamento entre ciclos consecutivos diminuiu para aproximadamente  $1/8 = 0,125$  segundos, ou seja, a taxa aumenta. O valor mínimo de mFAC ocorre em  $t = 500 \text{ ms}$ , quando o trecho de  $F_0$  entre 500 ms e 1 segundo é investigado. Este trecho da curva melódica é caracterizado por um comportamento irregular, como pode ser constatado no painel inferior. Já a parte da curva da frequência fundamental entre 1 e 1,5 segundos volta a ter ciclos mais regulares, por isso, o valor da autocorrelação volta a aumentar. A janela de  $F_0$  entre 1,5 e 2 segundos tem ciclos um pouco mais irregulares, causando a diminuição da mFAC. Além disso, este trecho engloba aproximadamente 4 ciclos da frequência fundamental, fazendo com que  $K^{-1}$  valha quase  $8 \text{ s}^{-1}$ . A análise termina por volta dos 2 segundos, ou seja, 0,5 segundos antes do fim da série de  $F_0$ .

Tendo sido feita uma descrição detalhada dos algoritmos a serem empregados no próximo capítulo, serão realizados, antes, testes de calibração.

## 3.4 Calibração

A etapa de calibração tem o objetivo de verificar o correto funcionamento dos algoritmos apresentados neste capítulo. Para isto, o método de análise da frequência fundamental e da envoltória (seção 3.3.6), o extrator de envoltória (seção 3.3.5) e o método de análise de regularidade (seção 3.3.7) foram sintonizados com o auxílio de sinais sintetizados. O funcionamento do algoritmo extrator de  $F_0$ , SRPD (seção 3.3.1), não foi investigado já que a implementação usada neste trabalho já foi testada e validada independentemente (Bagshaw, 1993, 1994).

### 3.4.1 Envoltória

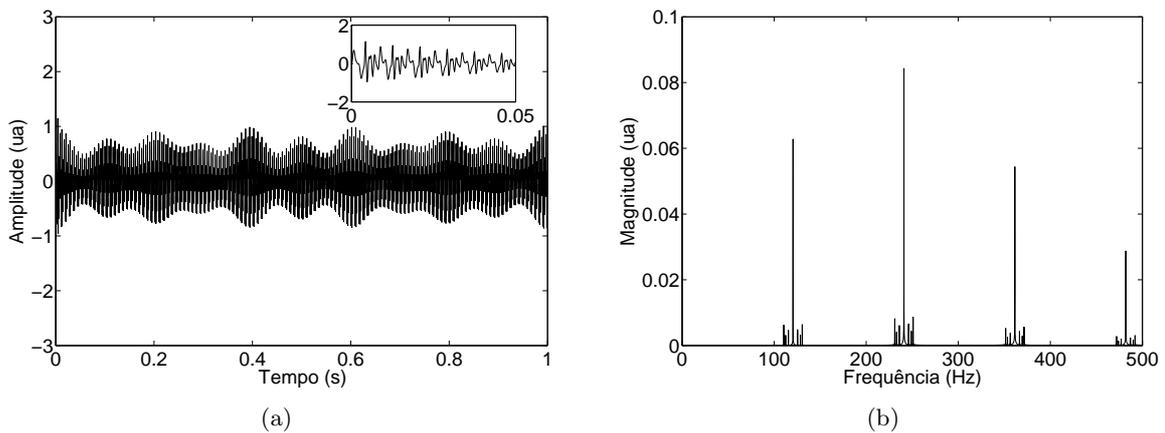
Um sinal de voz sintetizado e modulado em amplitude foi usado para avaliar a análise da envoltória no algoritmo da seção 3.3.5 e também o algoritmo extrator de envoltória.

O sinal de voz  $s_{\text{voz}}(n)$  foi sintetizado processando um trem de pulsos glóticos com fundamental em 120 Hz em um filtro que representa o trato vocal durante a emissão do /a/ (Sansão, 2009, cap. 2). A forma de onda típica desta vogal pode ser vista no detalhe da figura 3.27(a).

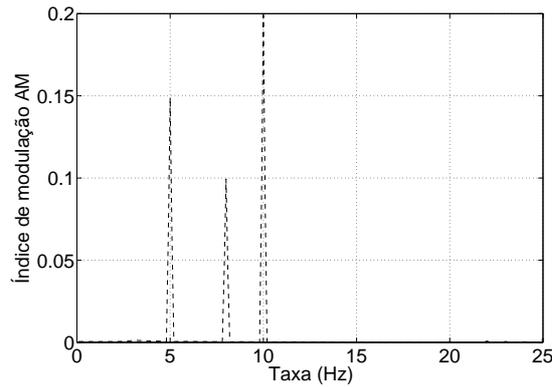
A modulação AM imposta a este sinal foi:

$$s_{AM}(n) = \left[ 1 + 0,15 \cos\left(\frac{2\pi 5n}{22050}\right) + 0,1 \cos\left(\frac{2\pi 8n}{22050}\right) + 0,2 \cos\left(\frac{2\pi 10n}{22050}\right) \right] \cdot s_{\text{voz}}(n).$$

A flutuação de amplitude da vogal é mostrada na figura 3.27(a), e um detalhe do espectro, na figura 3.27(b). Este sinal tem uma série harmônica com fundamental em 120 Hz na função



**Figura 3.27:** Sinal usado para fazer a calibração da análise de envoltória (a) no domínio do tempo e (b) no domínio da frequência.



**Figura 3.28:** Resultado da calibração da análise de envoltória.

de portadoras e modulantes em 5, 8 e 10 Hz, com os respectivos índices de modulação AM:

$$\begin{aligned} k_5 &= 0,15, \\ k_8 &= 0,1, \\ k_{10} &= 0,2. \end{aligned}$$

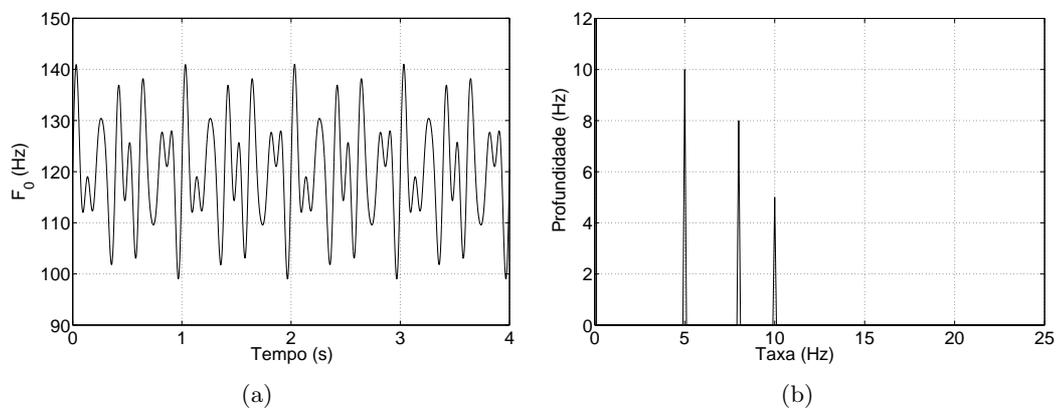
A comparação entre estes índices e o valor calculado pelo algoritmo (figura 3.28) ilustra o funcionamento correto do algoritmo.

### 3.4.2 Frequência fundamental

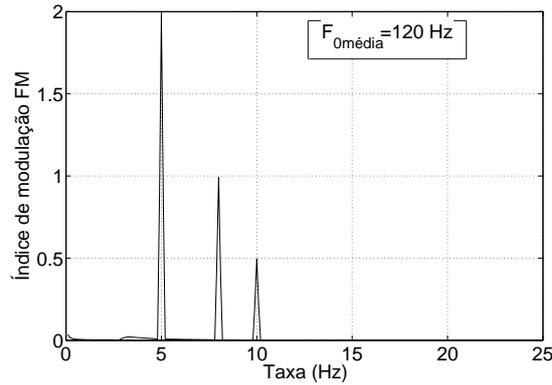
Para sintonizar a parte de análise da frequência fundamental do algoritmo explicado na seção 3.3.6, um sinal de  $F_0$  modulado em frequência foi sintetizado:

$$F_0(n) = s_{FM}(n) = 120 + 10 \operatorname{sen} \left( \frac{2\pi 5n}{22050} \right) + 8 \operatorname{sen} \left( \frac{2\pi 8n}{22050} \right) + 5 \operatorname{sen} \left( \frac{2\pi 10n}{22050} \right).$$

A portadora de 120 Hz foi modulada pelas frequências de 5, 8 e 10 Hz com os índices de



**Figura 3.29:** Sinal usado para fazer a calibração da análise de  $F_0$  (a) no domínio do tempo e (b) no domínio da frequência.



**Figura 3.30:** Resultado da calibração da análise de  $F_0$ .

modulação FM:

$$\begin{aligned}\beta_5 &= \frac{20\pi}{10\pi} = 2, \\ \beta_8 &= \frac{16\pi}{16\pi} = 1, \\ \beta_{10} &= \frac{10\pi}{20\pi} = 0,5.\end{aligned}$$

O resultado da análise está mostrado na figura 3.30, ilustrando o funcionamento correto desta parte do algoritmo.

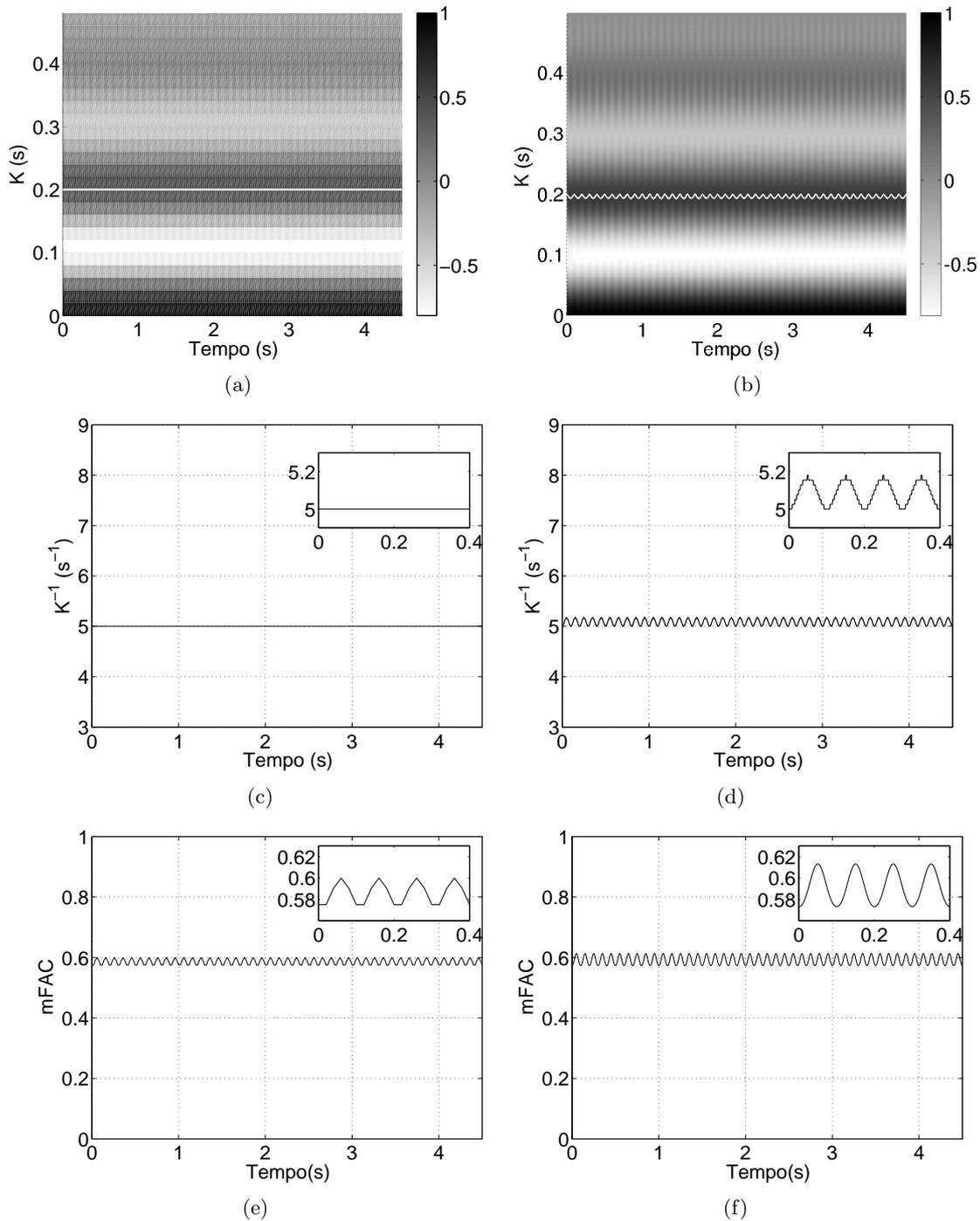
### 3.4.3 Regularidade

A parte de análise de regularidade foi calibrada com o auxílio de um sinal sintetizado para a frequência fundamental:

$$F_{0reg} = s_{reg}(n) = 120 + 10 \cdot \cos\left(\frac{2\pi 5n}{50}\right).$$

Este sinal foi primeiramente amostrado a 50 Hz durante 5 segundos. As figuras 3.31(a), 3.31(c) e 3.31(e) mostram o resultado da análise de regularidade deste sinal. A existência de transições abruptas na figura 3.31(a) motivou interpolar este sinal para uma taxa mais alta e analisá-lo outra vez. O fator de interpolação escolhido foi 20 e as figuras 3.31(b), 3.31(d) e 3.31(f) mostram o resultado da análise do sinal interpolado para 1000 Hz.

Em todo o conjunto de figuras 3.31, o fato que mais chama a atenção é a presença de flutuações nos valores de  $K^{-1}$  e mFAC (figuras 3.31(e), 3.31(d) e 3.31(f)). Tais oscilações não deveriam existir e, teoricamente haveria apenas uma linha horizontal. Porém, como já dito na seção 3.3.7, a função de autocorrelação é apenas *aproximada* (equação 3.13). Esta aproximação apresenta erros que aumentam a medida que a janela de sinal diminui. Com o tamanho de 500 ms, escolhido para janelar o sinal, os erros da aproximação são da ordem de 5 %. Este erro é aceitável nas análises que seguirão e o estudo de possíveis alternativas para sua redução é deixado para trabalhos futuros.



**Figura 3.31:** Calibração da análise de regularidade usando o sinal senoidal da equação 3.4.3. Nas figuras (a), (c) e (e), a senóide tem 50 Hz de amostragem. Nas figuras (b), (d) e (f), a taxa de amostragem é de 1000 Hz. A caixa dentro das últimas exibe os detalhes das curvas.

### 3.5 Conclusão

Este capítulo apresentou um algoritmo para análise das flutuações presentes nos sinais de vibrato e de tremor vocal. Foram descritas técnicas para investigar a curva de  $F_0$  e também a envoltória do sinal acústico. Além disso, foi apresentado um método para estudar a regularidade da frequência fundamental e todos estes métodos foram calibrados. A discussão dos resultados de aplicação destas técnicas nas amostras disponíveis nas bases de dados será apresentada no próximo capítulo.

# Capítulo 4

## Resultados

### 4.1 Introdução

Os resultados da aplicação do algoritmo descrito no capítulo anterior sobre as quatro bases de dados utilizadas serão apresentados a seguir. Cada seção deste capítulo apresenta a análise de uma das bases de dados da seguinte forma:

1. Exemplos individuais: algumas amostras da base de dados estudada serão analisadas a fim de apontar detalhes presentes no conjunto estudado.
2. Análise de grupo: todas as amostras da base de dados são analisadas a fim de obter um comportamento geral do conjunto de dados investigado.

Na parte de exemplos individuais serão apresentadas a forma de onda de  $F_0$ , a envoltória do sinal original e os respectivos espectros, expressos tanto na forma de índice de modulação quanto de profundidade em função da taxa. Vale a pena dizer que todos os valores de profundidade apresentados neste capítulo se referem à amplitude (figura 1.1) e não ao valor pico-a-pico. Também será apresentada a análise de regularidade destes casos.

As seções de análise de grupo apresentarão a distribuição da taxa e da profundidade, e também mostrarão como usar a redundância de informações existentes na envoltória e na série de  $F_0$ .

O capítulo começa apresentando a análise de tremor em membros, depois trata do vibrato no contrabaixo e termina investigando o sinal de voz com perturbações, o sinal de fala e os cantos lírico e sertanejo.

### 4.2 Membros

Tremor em membros do corpo humano denota uma oscilação involuntária dos braços ou pernas do corpo. Conforme exemplificado na seção 2 de Timmer et al. (2000), a taxa desta perturbação está na mesma faixa de frequências do tremor vocal, ou seja, aproximadamente entre 4 e 15 Hz, motivando seu estudo com o algoritmo descrito no capítulo 3. A presente

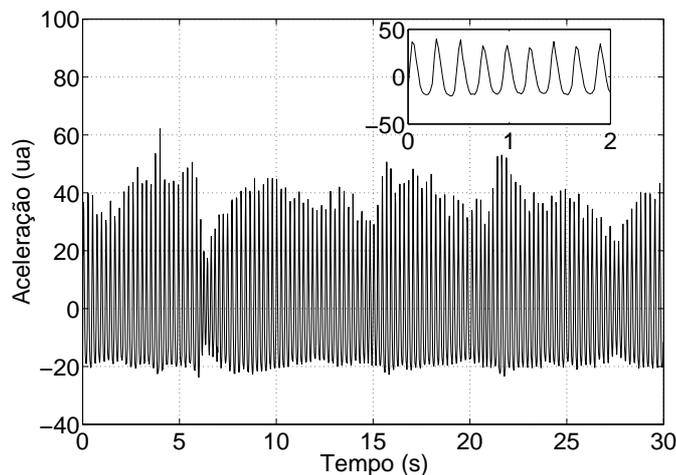
seção apresenta o resultado da aplicação de tal algoritmo nas séries da base de dados 1 (seção 3.2.1).

#### 4.2.1 Exemplo 1: tremor essencial de membros

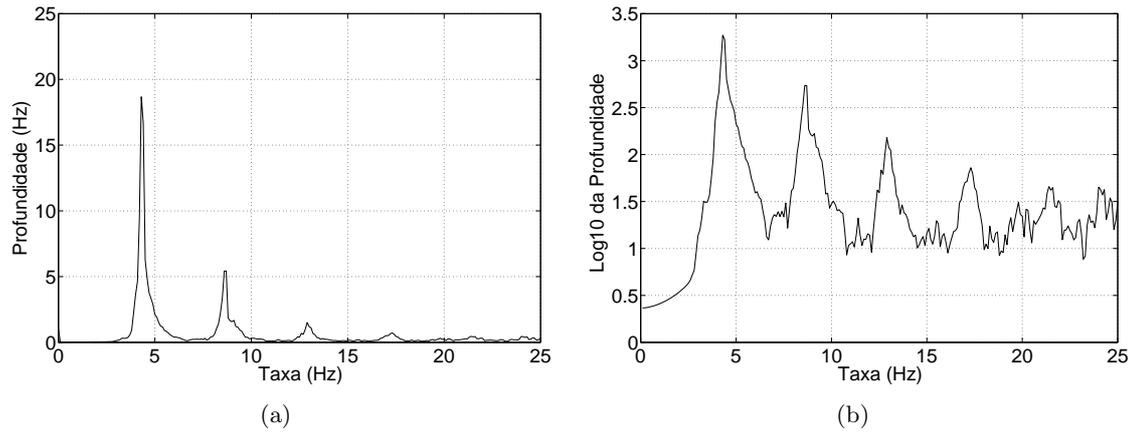
A figura 4.1 mostra um exemplo de série de um paciente com tremor essencial de membros, isto é, de um paciente com tremor de origem não diagnosticada. Assim como todas as outras séries desta base de dados, esta curva foi decimada e sua frequência de amostragem reduzida para 50 Hz. Como a informação destes sinais está concentrada em torno de uma dezena de hertz, a taxa de amostragem de 50 Hz é suficiente.

A curva mostrada na figura 4.1 é quase periódica, com período de aproximadamente 4 Hz e tem uma extensão média de 25 unidades arbitrárias (ua) de aceleração. Além disso, apresenta assimetrias e, como consequência, tem um espectro constituído por uma série de múltiplos do harmônico fundamental. A figura 4.2 evidencia a presença de múltiplos e também mostra que a contribuição dos harmônicos de ordem mais elevada é baixa. A figura 4.2(b), que contém a mesma informação da figura 4.2(a), está representada com escala vertical logarítmica a fim de evidenciar a série de múltiplos do harmônico fundamental.

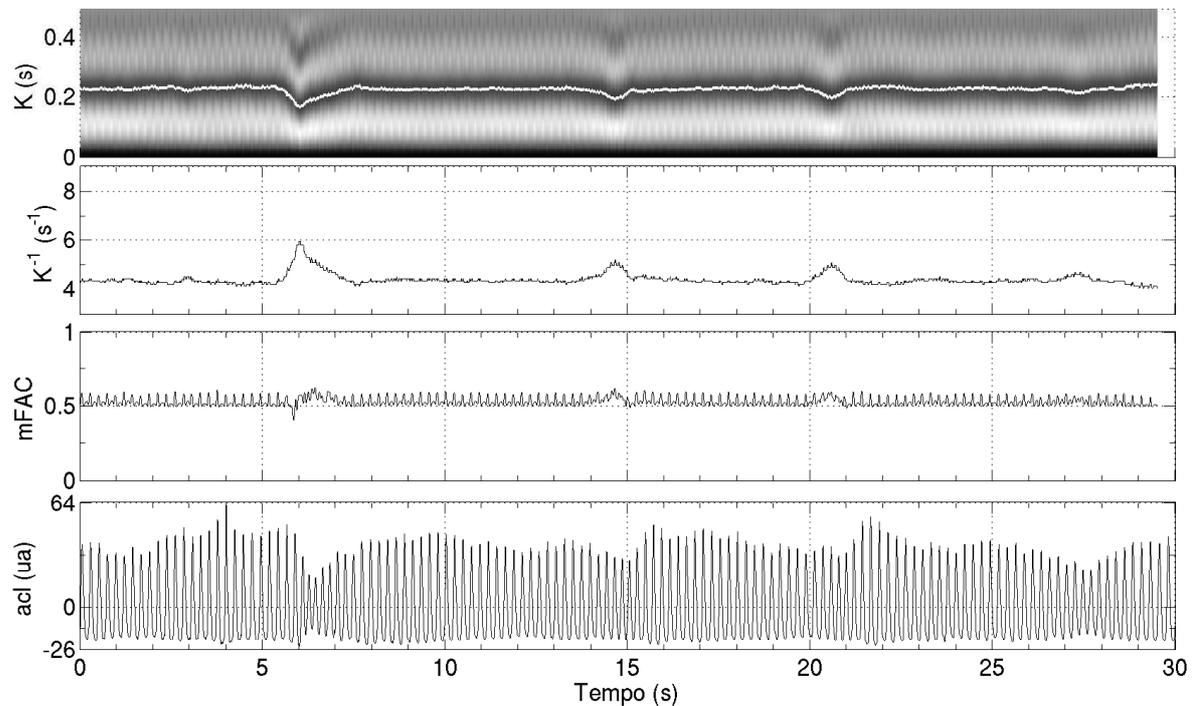
A análise de regularidade desta série está mostrada na figura 4.3. De acordo com o painel inferior desta figura, a série da aceleração é bastante regular até aproximadamente 7 segundos. Em todo este trecho, os valores da mFAC e de  $K^{-1}$  se mantêm, basicamente, constantes. Em torno de 7 a 8 segundos, a instabilidade presente na curva da aceleração é detectada pela mudança no comportamento de  $K^{-1}$  e também por uma pequena alteração do valor da mFAC. Outras perturbações também são detectadas em torno de 15, 22 e 28 segundos, todas menos intensas que aquela próxima a 7 segundos.



**Figura 4.1:** Série temporal da aceleração da mão de um paciente com tremor essencial. Um detalhe do comportamento da série entre 0 e 2 segundos está mostrado no canto superior direito.



**Figura 4.2:** Profundidade x taxa do sinal de aceleração da mão de um paciente com tremor essencial em (a) com a escala vertical linear e em (b) com a escala vertical logarítmica. Como a média do sinal de aceleração é nula, não há sentido em apresentar a profundidade em porcentagem da média. Por isso, a escala da figura (a) é indicada em Hz.

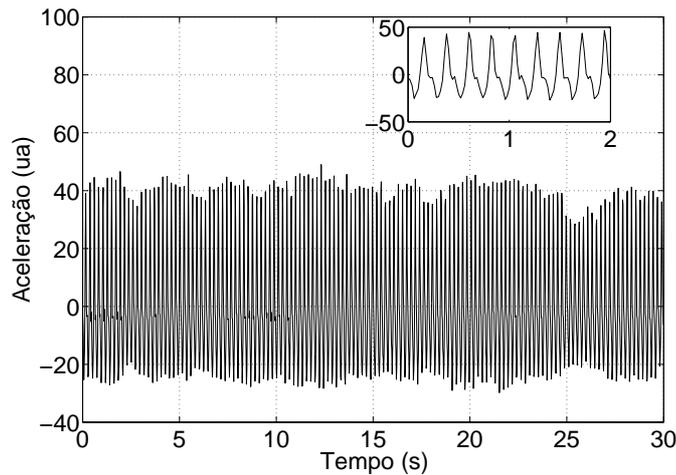


**Figura 4.3:** Análise de regularidade do sinal de aceleração (acl) da mão de um paciente com tremor essencial.

#### 4.2.2 Exemplo 2: doença de Parkinson

A figura 4.4 mostra uma série temporal da aceleração da mão de um paciente que sofre de doença de Parkinson. Neste exemplo, a taxa e a profundidade da perturbação estão em torno de 4 Hz e 30 ua, respectivamente.

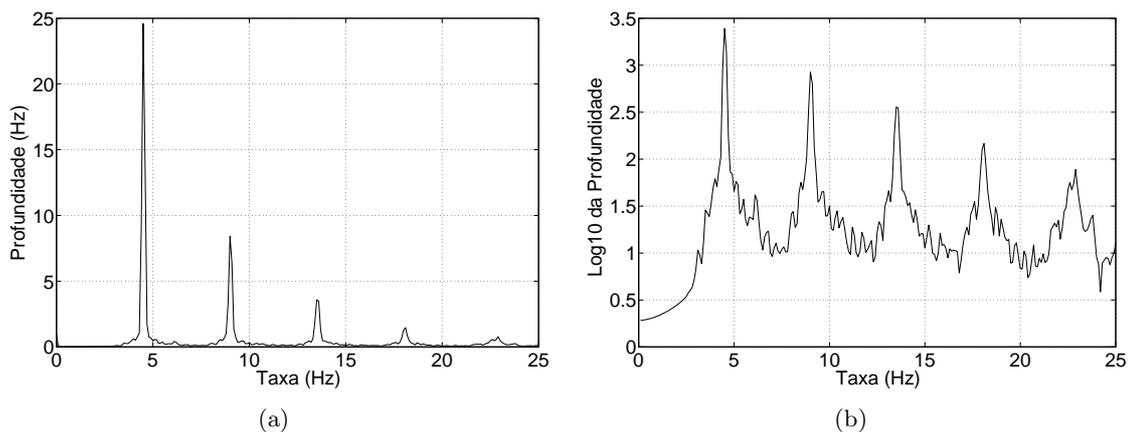
Comparando-se as figuras 4.1 e 4.4, nota-se que a curva de aceleração deste indivíduo



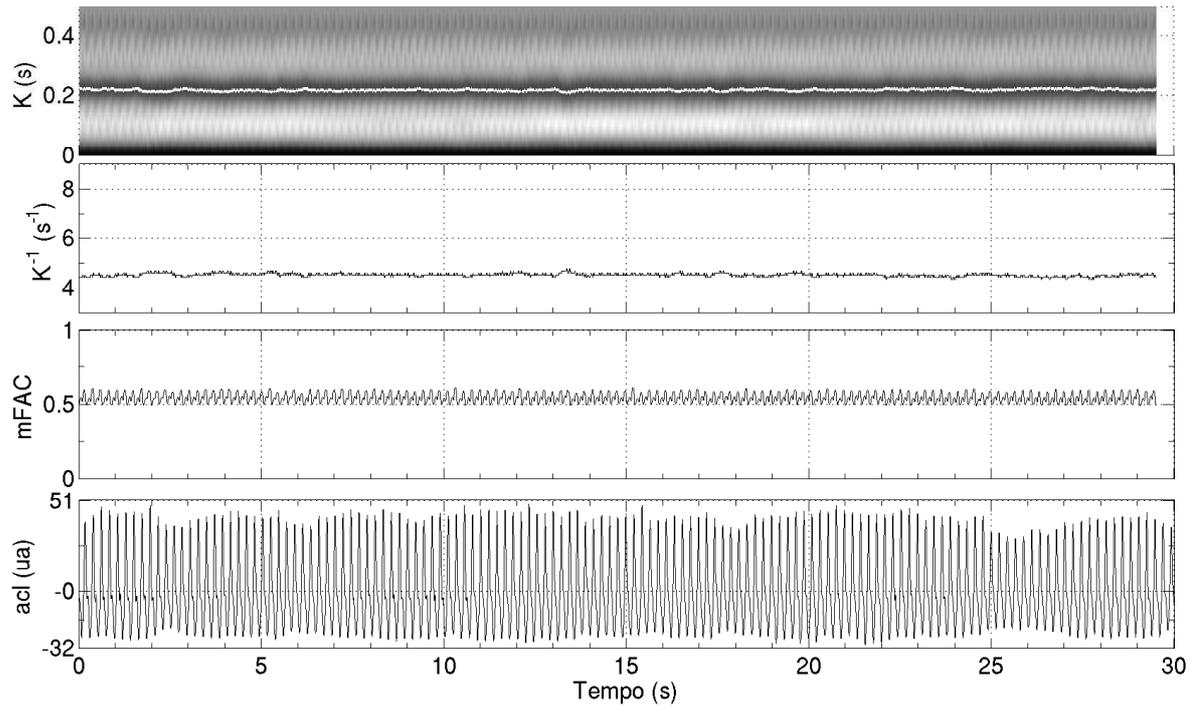
**Figura 4.4:** Série temporal da aceleração da mão de um paciente que sofre de doença de Parkinson. Um detalhe do comportamento da série entre 0 e 2 segundos está mostrado no canto superior direito. Compare com o caso do tremor essencial mostrado na figura 4.1.

afetado por Parkinson apresenta mais nuances que a do indivíduo do exemplo 1. Por consequência, as amplitudes dos componentes harmônicos da série deste indivíduo (figura 4.5) são maiores que a do paciente do exemplo 1.

A curva de aceleração deste paciente com doença de Parkinson é mais regular, ao longo do tempo, que aquela do paciente com tremor essencial. A figura 4.6 não contém as quebras presentes no exemplo de tremor essencial (figura 4.3). Isto sugere que o tremor deste indivíduo com doença de Parkinson seja mais “constante” que o do indivíduo do exemplo 1, o qual parece ser caracterizado por algumas variações abruptas.



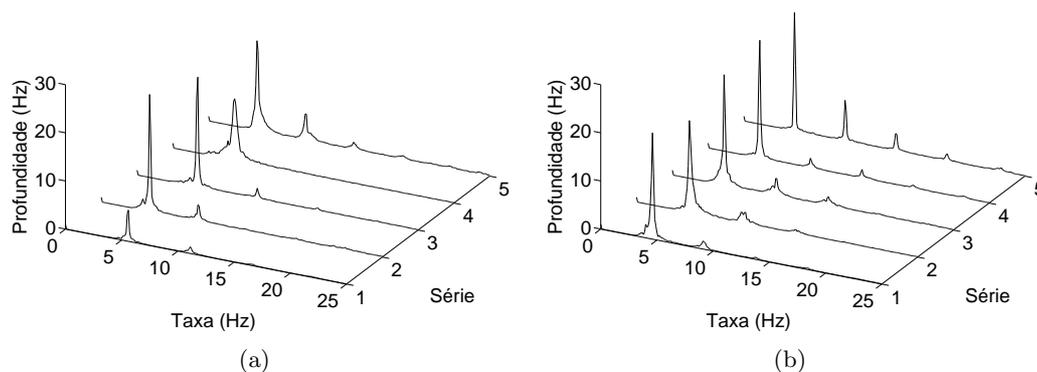
**Figura 4.5:** (a) Análise do sinal de aceleração de um paciente com doença de Parkinson com escala vertical linear. (b) A série de múltiplos do harmônico fundamental fica mais evidente na escala vertical logarítmica. Como a média do sinal de aceleração é nula, não há sentido em apresentar a profundidade em porcentagem da média. Por isso, a escala da figura (a) é indicada em Hz. Compare com o caso do tremor essencial da figura 4.2.



**Figura 4.6:** Análise de regularidade da série da aceleração (acl) da mão de um paciente sofrendo de doença de Parkinson. Compare com a figura 4.3.

### 4.2.3 Grupo

A figura 4.7 reúne os valores da taxa e da profundidade de todas as séries desta base de dados. Conforme é observado, os espectros de todos aqueles afetados pela doença de Parkinson têm, praticamente, os mesmos valores de extensão e taxa, ao contrário do que acontece com os espectros dos indivíduos com que sofrem de tremor essencial. Esta homogeneidade entre os espectros do tremor de informantes atingidos pela doença de Parkinson parece indi-



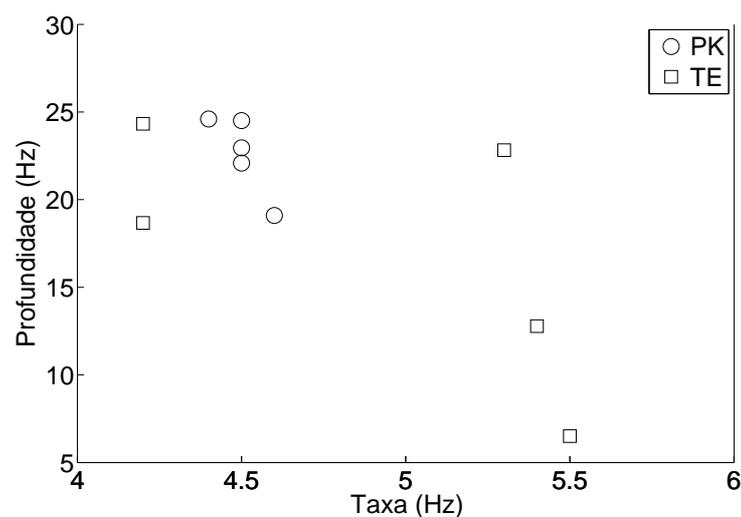
**Figura 4.7:** (a) Análise de todas as séries de indivíduos com tremor essencial. O exemplo 1 foi feito com a série de número 5. (b) Análise de todas as séries de indivíduos com doença de Parkinson. O exemplo 2 foi feito com a série de número 5.

car a existência de uma mesma base de transtorno neurológico em todos os pacientes deste grupo. Por outro lado, nos indivíduos com diagnóstico de tremor essencial, as variações entre os espectros de cada paciente sugere a existência de diferentes transtornos neurológicos responsáveis por estas oscilações.

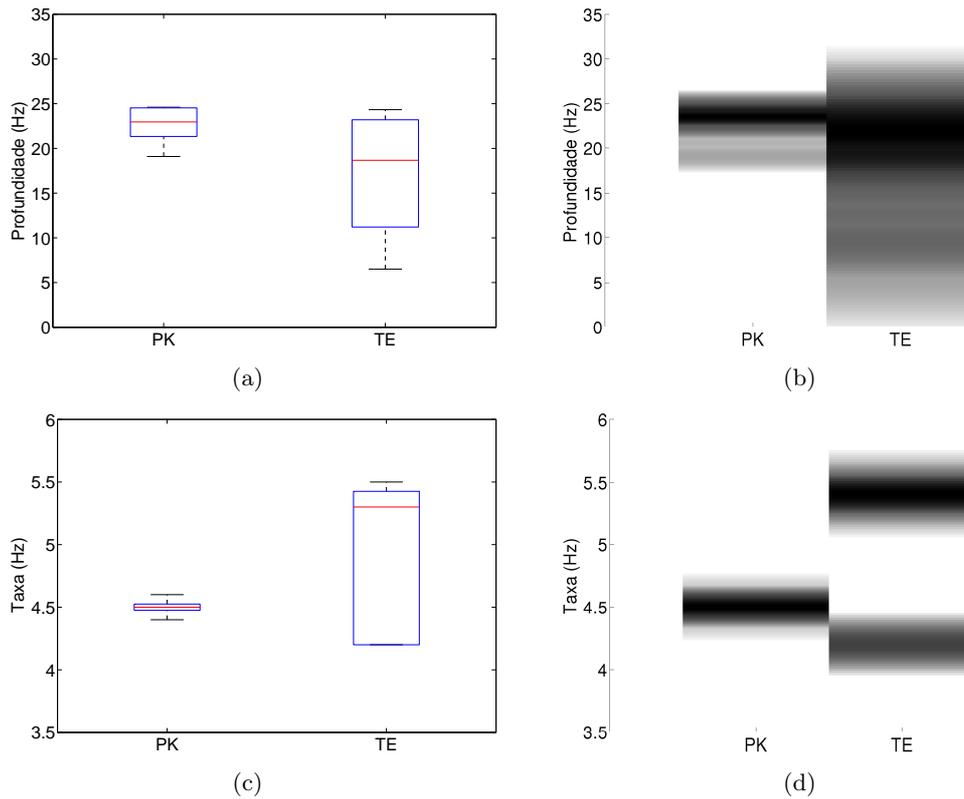
A figura 4.8 reúne os valores dos máximos globais de cada um dos espectros da figura 4.7. Este gráfico, apesar de não informar nada sobre os múltiplos do harmônico fundamental de cada série, é muito útil para investigar o comportamento geral entre os grupos. Nele, observa-se, por exemplo, que a taxa do grupo diagnosticado com tremor essencial está distribuída em duas regiões distintas, enquanto a taxa do grupo de pacientes com doença de Parkinson está concentrada em torno de um só valor.

As figuras 4.9(a) – (d) foram criadas para facilitar a visualização dos dados da figura 4.8. A imagens 4.9(a) e 4.9(b) exibem apenas a distribuição da profundidade em cada grupo, e a figuras 4.9(c) e 4.9(d), somente a distribuição da taxa. É interessante notar como as informações das figuras do *boxplot* e da distribuição das amostras se complementam. Comparando as imagens 4.9(c) e 4.9(d), percebe-se que é simples localizar o valor da mediana, dos quartis, do máximo e do mínimo dos dois grupos na figura 4.9(c). Todavia, a diferença de comportamento entre as duas distribuições só fica clara com a figura 4.9(d), que evidencia a distribuição unimodal para a taxa do grupo afetado por Parkinson e a distribuição bimodal para a taxa do grupo com tremor essencial.

Esta diferença de distribuições da taxa entre os dois grupos viola pressupostos para aplicação dos testes de ANOVA (normalidade das distribuições) e de Kruskal-Wallis (distribuições quaisquer, mas de mesmo tipo; Kohler (1988)). Quanto à profundidade, admitiu-se uma mesma forma de distribuição deste parâmetro nos grupos de Parkinson e de tremor essencial (figura 4.9(b)). Para este caso foi executado o teste de Kruskal-Wallis, que retornou  $p = 0,12$ . Normalmente este valor de  $p$  é julgado alto e aceito como um indicativo de



**Figura 4.8:** Profundidade x taxa de todas as amostras de tremor essencial (TE) e doença de Parkinson (PK).



**Figura 4.9:** Distribuição da taxa e da profundidade dos grupos afetados por tremor essencial (TE) e doença de Parkinson (PK). (a) e (b) Distribuição da profundidade em cada um dos grupos (c) e (d) Distribuição da taxa em cada um dos grupos. Os tons de cinza nas figuras (b) e (d) indicam a frequência estatística das amostras, sendo tanto maior quanto mais escura for a região.

ausência de diferenças estatisticamente significativas. Entretanto, é importante enfatizar que cada grupo é formado por apenas 5 amostras, uma quantidade pequena para concluir que não existam diferenças estatísticas significativas entre populações de indivíduos afetados pela doença de Parkinson e pelo tremor essencial.

Por fim, a tabela 4.1 reúne os valores da média, da mediana e do desvio padrão da taxa e da profundidade de cada um dos grupos. Nota-se que, nos dois grupos, o desvio padrão da profundidade é maior que o da taxa, apontando que há uma estabilidade maior da frequência que da amplitude do movimento nos dois grupos. Estes valores serão usados em comparações nas análises das próximas seções.

Grupo	Taxa			Profundidade		
	Média (Hz)	Mediana (Hz)	Desvio Padrão (Hz)	Média (Hz)	Mediana (Hz)	Desvio Padrão (Hz)
Tremor Essencial ( $n = 5$ )	4,9	5,3	0,7	17,02	18,67	7,39
Parkinson ( $n = 5$ )	4,5	4,5	0,1	22,65	22,93	2,25

**Tabela 4.1:** Estatísticas de taxa e profundidade: base de dados 1.

## 4.3 Contrabaixo

### 4.3.1 Exemplo 1: mesma nota, diferentes tipos de vibrato

A figura 4.10 mostra uma análise feita em duas amostras da base de dados 2 (seção 3.2.2). A coluna da esquerda refere-se a uma emissão produzida com vibrato tipo I (figura 3.1(a)) e fundamental em 220 Hz e a coluna da direita, a uma emissão com mesma fundamental, mas produzida com vibrato tipo II (figura 3.1(b)).

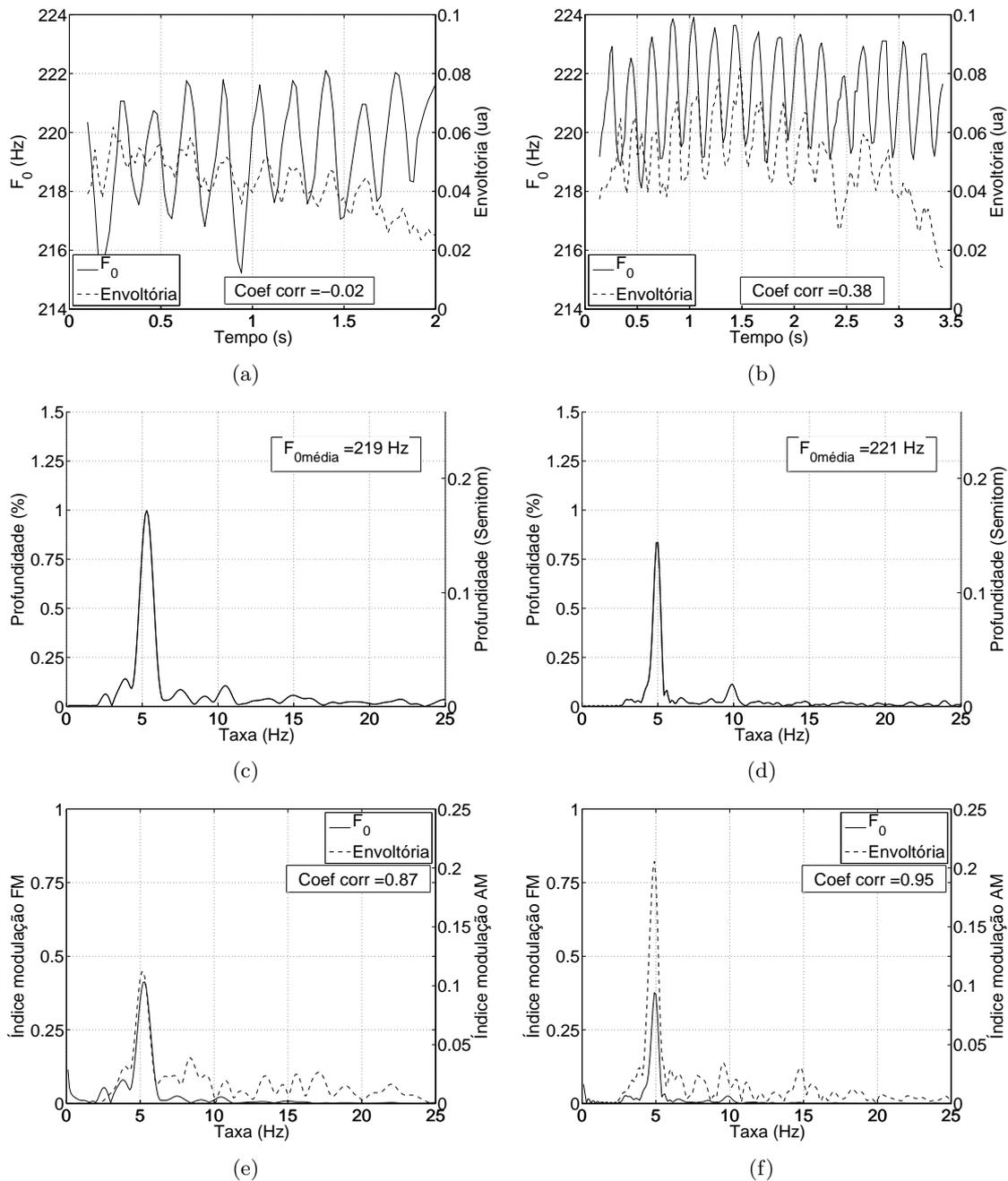
Observando as figuras 4.10(a) e 4.10(b), nota-se que, no vibrato tipo I, a curva de  $F_0$  excursions em torno do valor da nota alvo, enquanto, no vibrato tipo II, a excursão de  $F_0$  está centrada em uma frequência ligeiramente maior que a da nota alvo. Supondo que nos dois casos o instrumentista marca corretamente a posição da nota alvo no braço do instrumento, estas figuras sugerem que a utilização do dedo auxiliar, no vibrato tipo II, altera o movimento executado pelo instrumentista provocando um aumento na altura da nota tocada. Além disso, observa-se que a curva de  $F_0$  está em fase com a envoltória nos dois casos. No caso de instrumentos musicais, o modelo de harmônico-formante da figura 1.2 pode ser aplicado desde que a curva de formante do trato vocal seja substituída pela curva de ressonâncias do instrumento.

As figuras 4.10(c) e 4.10(d) mostram que a taxa e a profundidade das duas amostras é aproximadamente a mesma. Percebe-se também a existência de um segundo harmônico nos dois casos e que ele está mais bem definido na figura da direita. Isto aconteceu porque a amostra de áudio usada para gerar o gráfico da direita era mais longa que a outra amostra de vibrato.

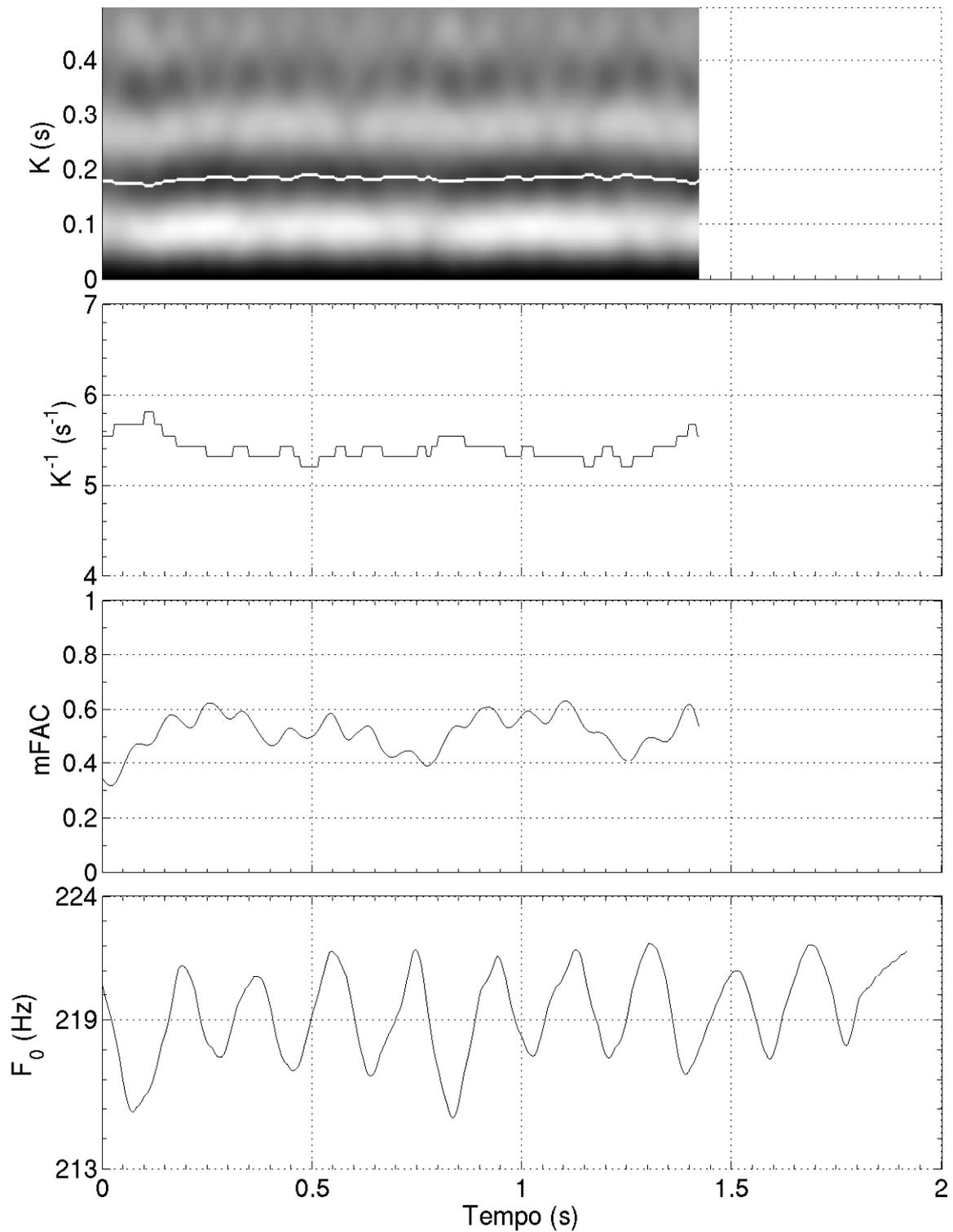
Os espectros da envoltória e da curva de  $F_0$  das duas gravações estão representados nas figuras 4.10(e) e 4.10(f) através dos índices de modulação AM e FM. O mais surpreendente nestas figuras é, sem dúvida, a diferença entre os espectros das envoltórias. O índice de modulação AM da amostra de vibrato tipo II é, praticamente, duas vezes maior que o da outra amostra, indicando uma maior flutuação da envoltória quando o instrumentista usa a técnica II. Em um julgamento realizado pelo autor, há diferença perceptiva entre as duas amostras e elas deverão ser exploradas em trabalhos futuros.

Por último, as figuras 4.11 e 4.12 trazem a análise de regularidade destas amostras de vibrato. Nota-se primeiramente que o valor médio de  $K^{-1}$  é maior para a amostra de vibrato tipo I que a tipo II. De maneira correspondente, a taxa do vibrato tipo I é um pouco maior que a taxa do vibrato tipo II, indicando que  $K^{-1}$  pode ser usado como um indicativo da taxa do sinal. Quanto à curva de mFAC, que está relacionada à similaridade entre ciclos vizinhos de  $F_0$ , percebem-se algumas diferenças entre as curvas do tipo I e do tipo II. Para o caso I, mFAC começa valendo aproximadamente 0,4 e aumenta depois que o primeiro mínimo de  $F_0$  deixa de ser analisado, por volta de 150 ms. Em torno de 800 ms e de 1,25 s esta curva tem novos mínimos de destaque, indicando que os trechos de 800 ms a 1,3 s, e de 1,25 s a 1,75 s da série de  $F_0$  englobam ciclos com comportamento irregular. Para a amostra tipo II, a única região que apresentou uma quebra de comportamento ocorreu por volta de 2,4 s. Por isso sua mFAC tem um mínimo em torno de 2,1 s. É importante observar que, neste exemplo,

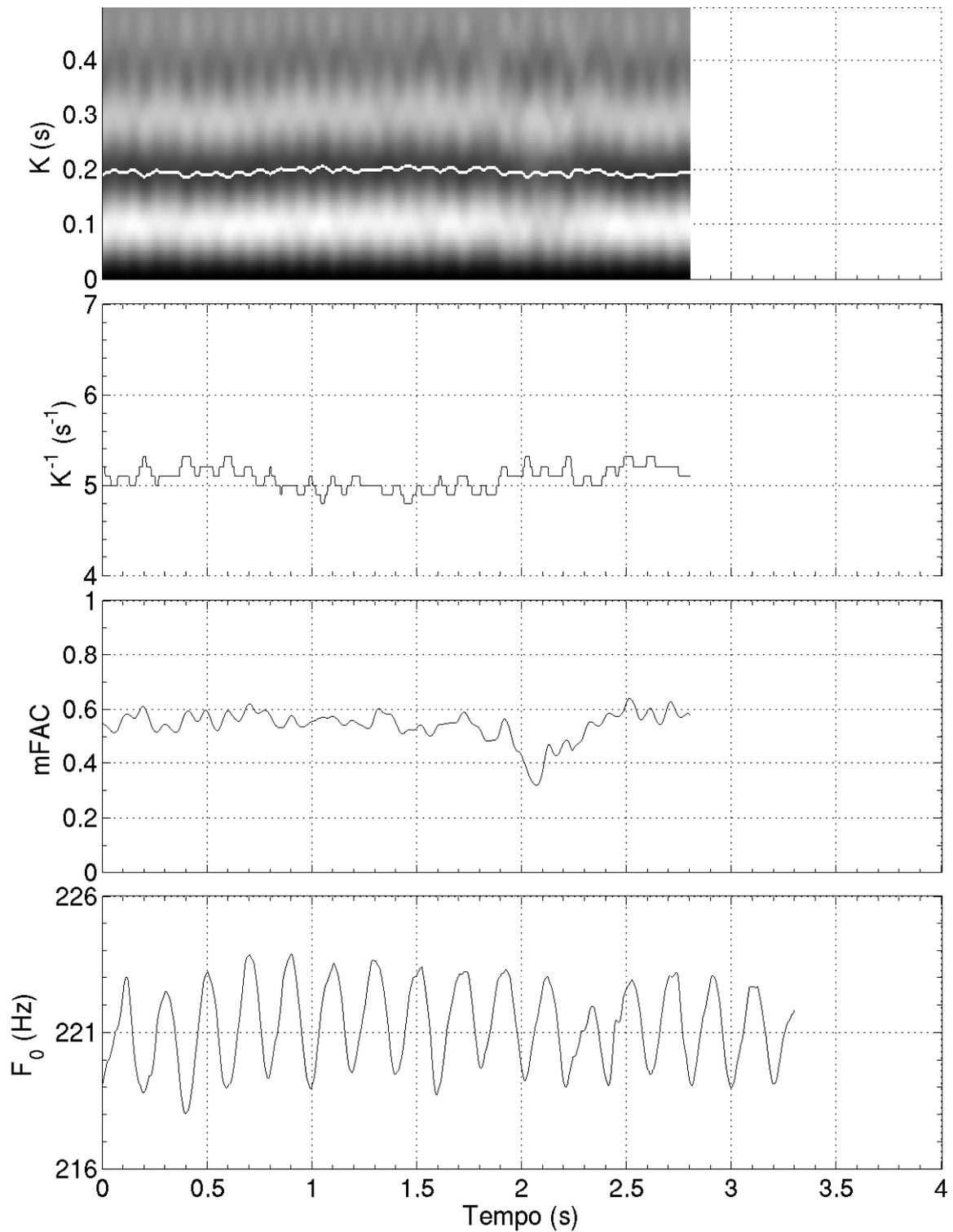
não se pode afirmar que a amostra de vibrato tipo II é mais regular que a amostra tipo I baseando-se simplesmente no valor médio destas curvas de mFAC. Como o estimador usado para calcular a função de autocorrelação (equação 3.13) é tendencioso, tal comparação levaria a equívocos. Uma provável alternativa, a ser estudada em trabalhos futuros, é construir a curva de mFAC usando o estimador não-tendencioso para a função de autocorrelação.



**Figura 4.10:** Comparação entre vibratos tipo I (1 dedo), na coluna da esquerda e tipo II (2 dedos), na coluna da direita. A mesma nota (220 Hz) foi tocada nos dois casos.



**Figura 4.11:** Análise de regularidade da amostra de vibrato tipo I. A curva de  $F_0$  é a mesma da figura 4.10(a), interpolada para melhorar a visualização de detalhes.



**Figura 4.12:** Análise de regularidade da amostra de vibrato tipo II. A curva de  $F_0$  é a mesma da figura 4.10(b), interpolada para melhorar a visualização de detalhes.

### 4.3.2 Exemplo 2: diferentes notas, mesmo tipo de vibrato

Neste exemplo, duas amostras da base de dados 2 (seção 3.2.2) de vibrato tipo I produzidas em diferentes registros do contrabaixo serão estudadas.

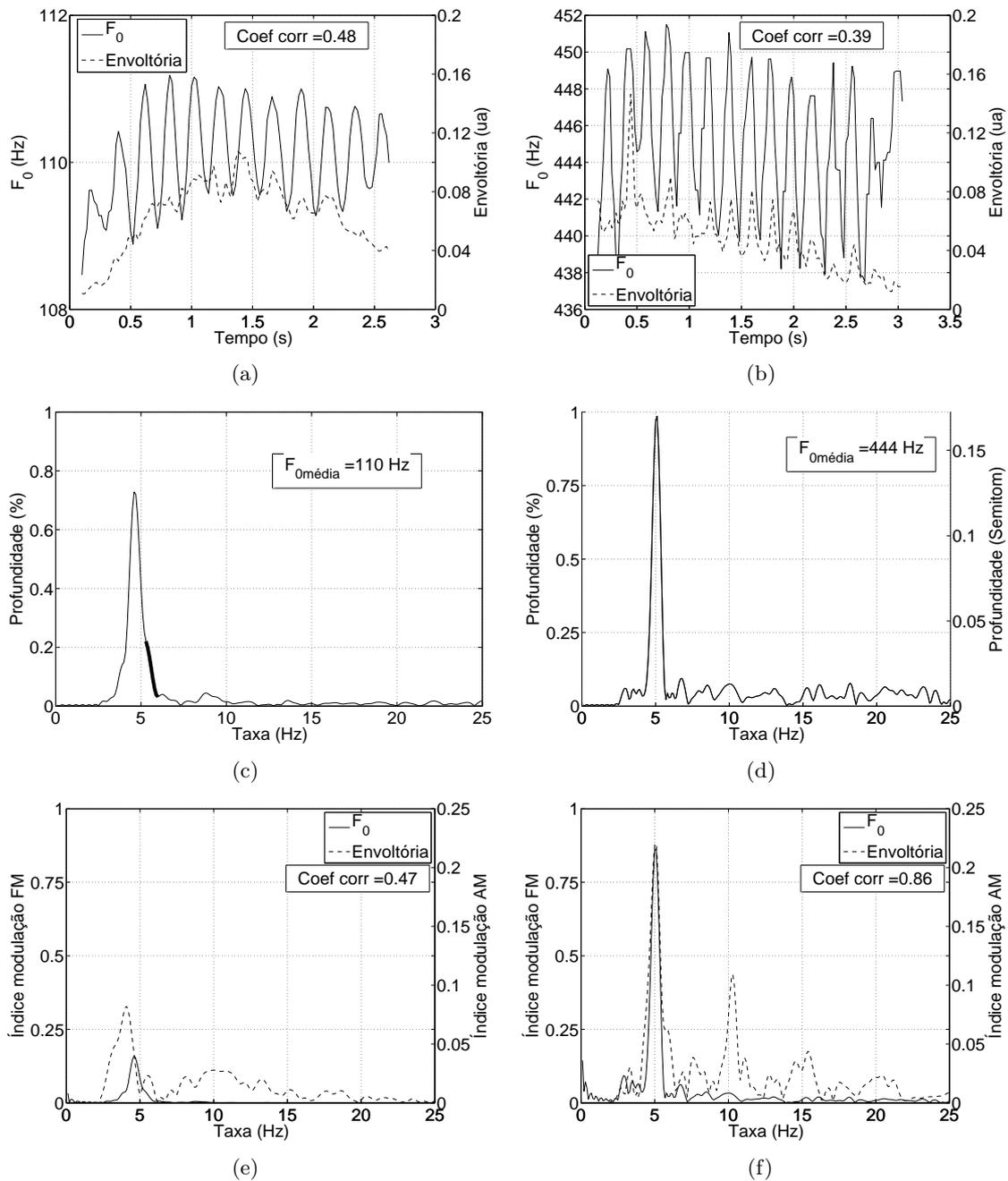
A curva de  $F_0$  e a envoltória dos vibratos executados em torno de uma nota grave e de uma nota aguda são mostradas, respectivamente, nas figura 4.13(a) e 4.13(b). Note, nos dois casos, que a curva de frequência fundamental é mais instável no início da execução do vibrato. De acordo com Borem (2010), esta variação inicial pode ser uma consequência do movimento balístico (Lage et al., 2002) feito pelo instrumentista ao começar o vibrato. O estudo das implicações deste fato sobre a percepção do vibrato foge do escopo deste trabalho e será explorado em trabalhos futuros.

O espectro de  $F_0$  e da envoltória das duas amostras está mostrado nas figura 4.13(c) e 4.13(d). É evidente que o vibrato executado no registro mais alto, deste exemplo, tem taxa e profundidade maiores que o vibrato executado no grave. Percebe-se também que o espectro da figura 4.13(c) parece não ter uma única componente harmônica na região de aproximadamente 5 Hz, destacada na figura. Parece haver, além da componente de maior profundidade e localizada em torno de  $\approx 4,8$  Hz, uma raia em 5,1 Hz e com uma profundidade de aproximadamente 0,2. Contudo, como a amostra é de curta duração, não é possível resolver o eixo de frequência a um nível suficiente para observar tais detalhes. Como será mostrado mais adiante, a curva de  $K^{-1}$  da figura 4.14, obtida na análise de regularidade, consegue resolver melhor esta outra raia.

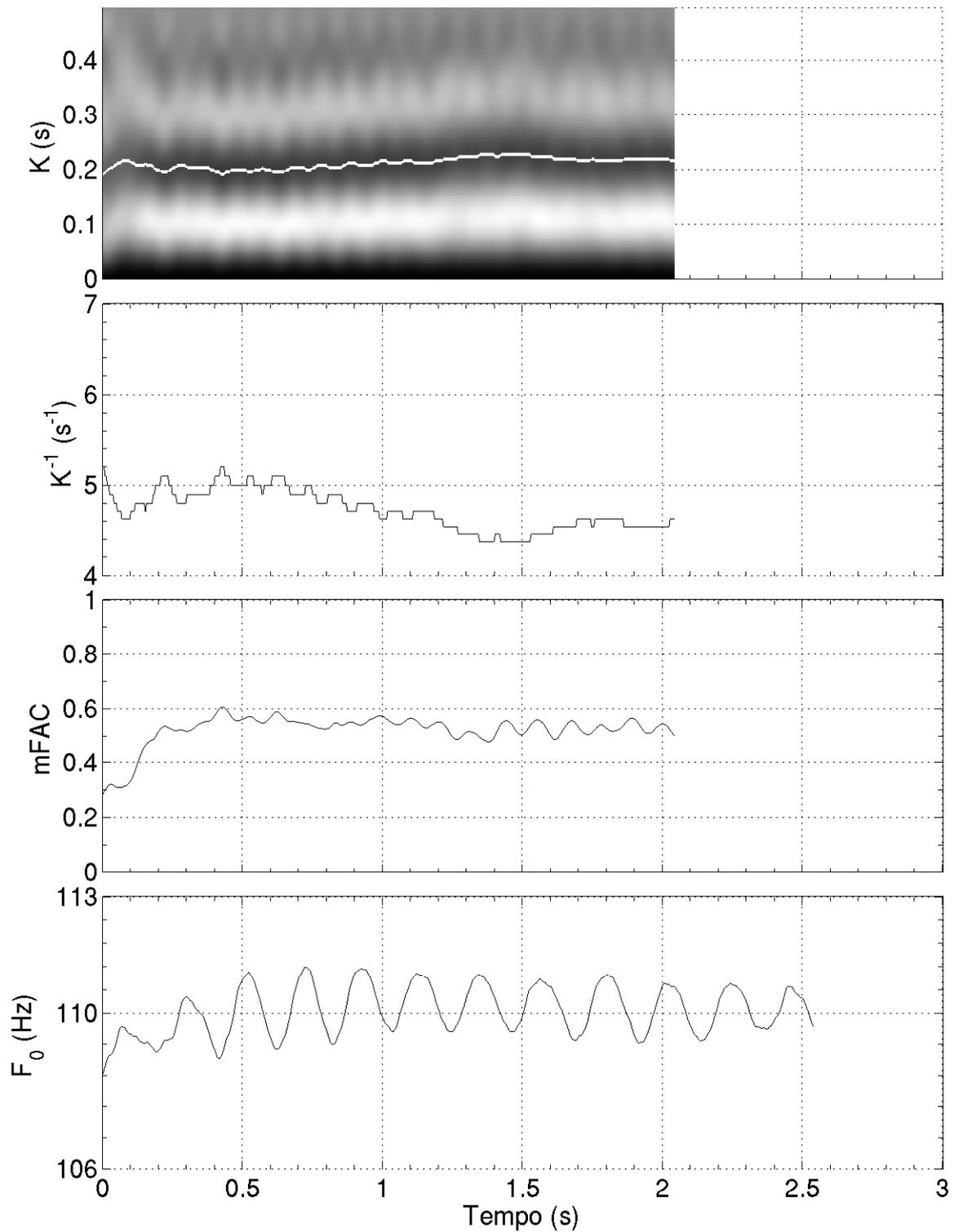
Observando as figuras 4.13(e) e 4.13(f), que exibem os índices de modulação AM e FM das duas amostras de vibrato, percebe-se que, além do espectro de  $F_0$ , o espectro da envoltória também tem maior magnitude na nota mais aguda. Além disso, nota-se na figura 4.13(f) a presença destacada de uma série harmônica no índice AM do vibrato produzido sobre a nota mais aguda. Este mesmo comportamento parece também acontecer no vibrato sobre a nota grave, sugerindo que a técnica empregada para produzir o vibrato tipo I produza uma envoltória rica em harmônicos, ou seja, o movimento executado nesta técnica não tem um comportamento harmônico simples.

Por fim, a análise de regularidade deste exemplo é apresentada nas figuras 4.14 e 4.15. Como esperado, devido ao início um pouco irregular das duas amostras, suas curvas de mFAC começam com vales e só estabilizam em torno de mFAC = 0,5 por volta de 0,2 segundos. No caso da amostra executada sobre a nota aguda, mFAC também tem uma variação considerável em sua parte final, devido à presença de alguns ciclos menos regulares/semelhantes em  $F_0$  a partir de  $t = 2,3$  segundos. As curvas de  $K^{-1}$  das duas amostras de vibrato apresentam dois platôs ao longo do tempo. Na curva da vibrato produzido sobre a nota grave,  $K^{-1} \approx 5 \text{ s}^{-1}$  no início e, depois de aproximadamente 1,3 segundos,  $K^{-1} \approx 4,5 \text{ s}^{-1}$ . Isto sugere que a taxa deste sinal não seja composta por uma só componente, justificando a presença de uma raia na região destacada da figura 4.13(c). No caso do vibrato produzido sobre a nota aguda,  $K^{-1}$  começa valendo  $5,5 \text{ s}^{-1}$ , reduz para  $5 \text{ s}^{-1}$  e volta para  $5,5 \text{ s}^{-1}$  no final. Isto sugere que, no caso da nota aguda, a taxa do vibrato tenha mais de uma componente, que não está

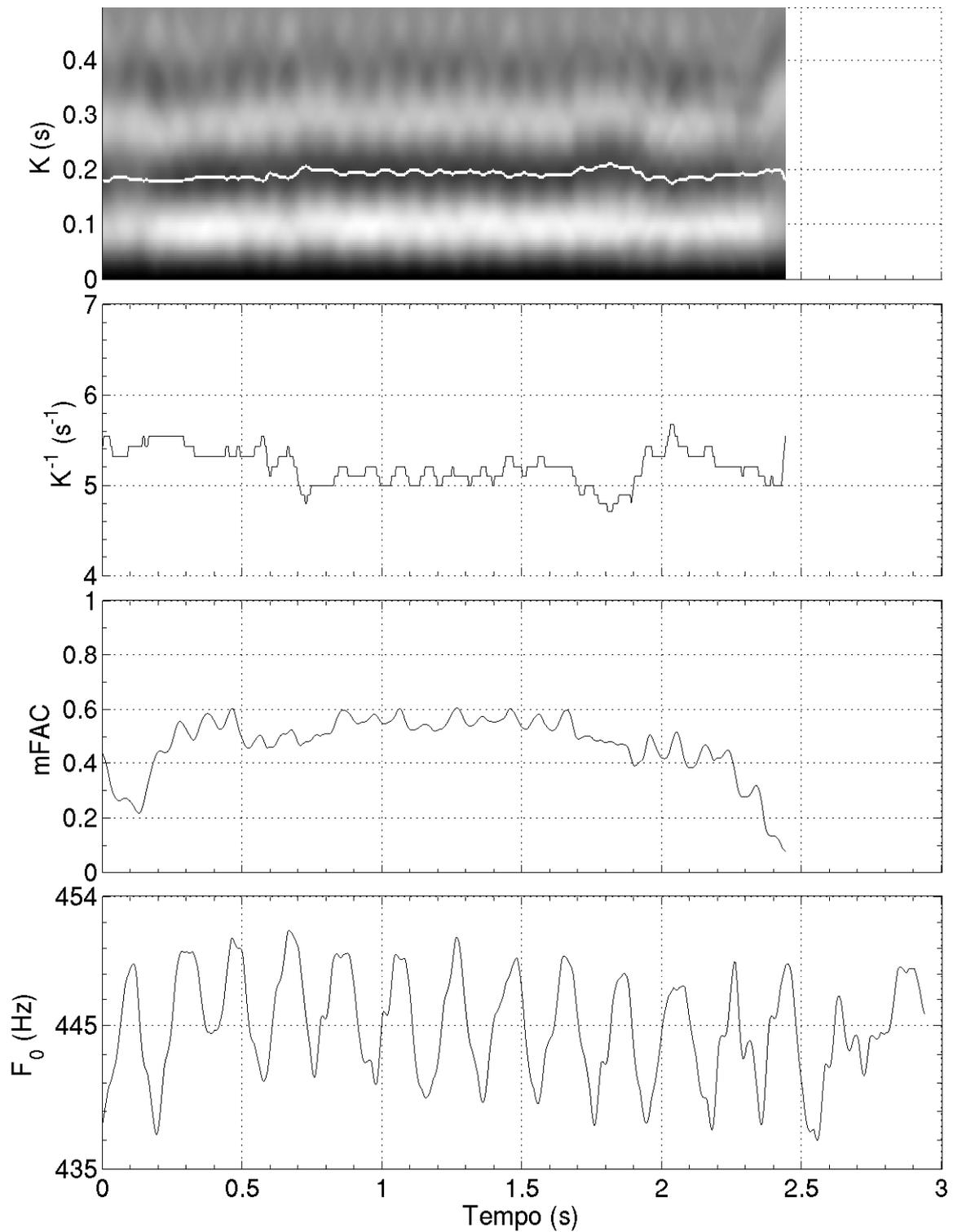
evidenciada na figura 4.13(d) devido às limitações da análise de Fourier causadas pela curta duração da amostra analisada.



**Figura 4.13:** Comparação entre vibratos em uma nota com fundamental de 110 Hz, na coluna da esquerda e em uma nota com fundamental em 440 Hz, na coluna da direita. Nos dois casos foi usado o vibrato tipo I. A razão do destaque em parte da figura 4.13(c) é comentada no texto.



**Figura 4.14:** Análise de regularidade da amostra de vibrato executado sobre uma nota grave (110 Hz). A curva de  $F_0$  é a mesma da figura 4.13(a), interpolada para melhorar a visualização de detalhes.



**Figura 4.15:** Análise de regularidade da amostra de vibrato executado sobre uma nota aguda (440 Hz). A curva de  $F_0$  é a mesma da figura 4.13(b), interpolada para melhorar a visualização de detalhes.

### 4.3.3 Grupo

Para esta base de dados, foram feitas duas análises de grupo. Na primeira delas, as amostras foram divididas por registro (grave, médio ou agudo) e na outra, classificadas de acordo com o tipo de vibrato (I, II ou III). O objetivo desta divisão é investigar a influência do tipo de registro e do tipo do vibrato na taxa e na profundidade. É importante lembrar que todas as amostras desta base de dados foram produzidas por um só instrumentista e os resultados mostrados nesta seção não devem ser generalizados.

#### 4.3.3.1 Divisão por registro

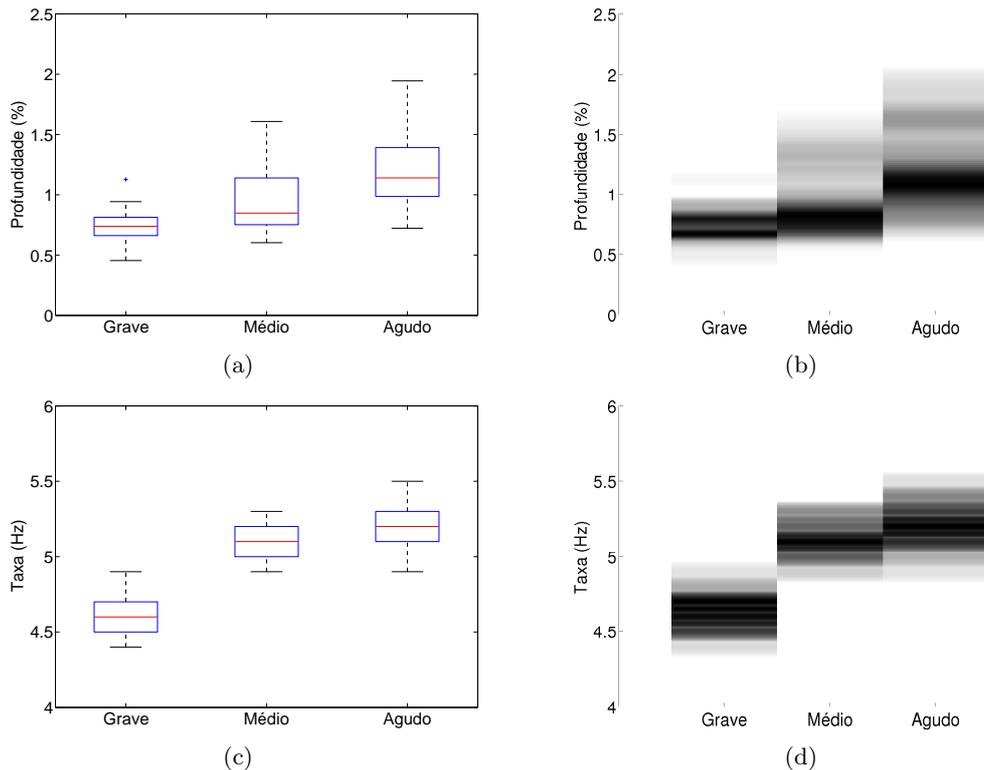
A tabela 4.2 traz um resumo da taxa e da profundidade do vibrato em função da altura do registro. Nota-se que o valor médio da taxa está na mesma faixa de valores do tremor em membros (tabela 4.1), sugerindo a existência de um sistema “biológico” comum por trás destes dois tipos de movimento. Percebe-se que, à medida que a frequência do registro aumenta, a taxa e a profundidade do vibrato também aumentam. Segundo Borem (2010), este comportamento acontece não só no contrabaixo, mas em toda a família de instrumentos de corda, e.g., viola, violino, violoncelo. A razão para isto, explica Borem (2010), está nas diferentes amplitudes e velocidades dos movimentos realizados para geração de vibrato em cada registro. Em tons graves, o movimento é amplo e lento; em tons agudos, o movimento é curto e rápido. Uma hipótese, a ser explorada em trabalhos futuros, é verificar se o produto amplitude x velocidade do movimento na produção do vibrato tende ao mesmo valor nos 3 registros, sugerindo que a “energia” gasta pelo instrumentista é a mesma nos três casos.

A distribuição da taxa e da profundidade dos grupos grave, médio e agudo está exibida na figura 4.16. No painel 4.16(b) nota-se uma assimetria nas distribuições da profundidade dos grupos médio e agudo. Para o grupo grave, observa-se uma distribuição mais simétrica e com duas regiões de concentração mais pronunciadas. Além disso, o painel 4.16(d) mostra que a diferença entre as taxas máxima e mínima é praticamente a mesma em todos os grupos e vale 0,5 Hz.

A figura 4.17 mostra todos os dados em um gráfico de profundidade versus taxa. Neste gráfico observa-se uma transição suave tanto na taxa quanto na profundidade do vibrato à medida em que o instrumentista aumenta a altura do registro. Além disso, nota-se que o controle sobre a profundidade do vibrato é menor em notas mais altas. Isto é justificado

Registro	Taxa			Profundidade		
	Média (Hz)	Mediana (Hz)	Desvio Padrão (Hz)	Média (%)	Mediana (%)	Desvio Padrão (%)
Grave ( $n = 90$ )	4,6	4,6	0,1	0,74	0,74	0,11
Médio ( $n = 90$ )	5,1	5,1	0,1	0,94	0,85	0,26
Agudo ( $n = 90$ )	5,2	5,2	0,1	1,21	1,14	0,31

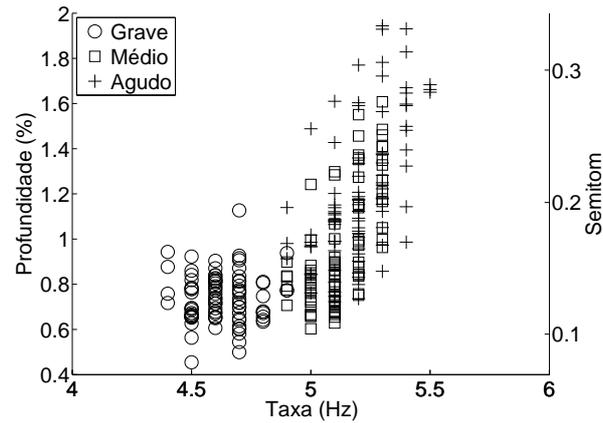
**Tabela 4.2:** Estatísticas de taxa e profundidade: base de dados 2 classificada por registro.



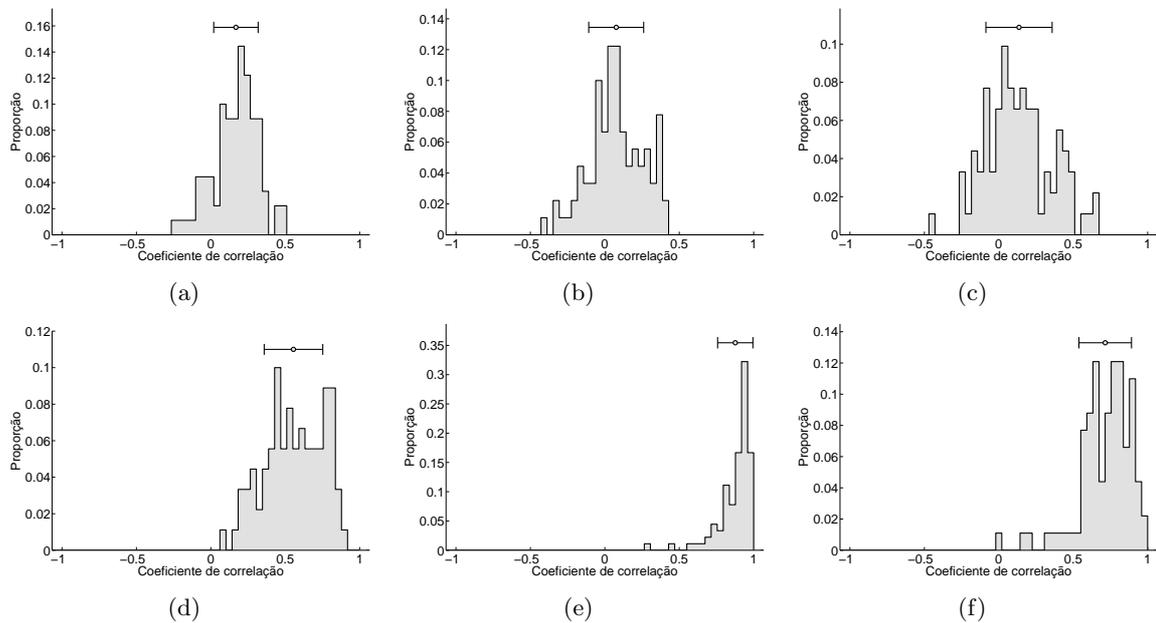
**Figura 4.16:** Distribuição da taxa e da profundidade das amostras da base da dados 2, organizadas por tipo de registro. Grave, Médio e Agudo indicam os registros do contrabaixo onde o vibrato foi produzido. (a) e (b) Distribuição da profundidade em cada um dos grupos (c) e (d) Distribuição da taxa em cada um dos grupos. Os tons de cinza nas figuras (b) e (d) indicam a frequência estatística das amostras, sendo tanto maior quanto mais escura for a região.

pela maior dificuldade no controle da amplitude do movimento na produção do vibrato em registros altos, já que, nestas regiões, pequenos desvios na trajetória “ideal” do dedo sobre a nota alvo causam maiores variações na altura do vibrato que em registros graves.

A figura 4.18 mostra, na linha superior, as distribuições dos coeficientes de correlação entre as curvas da frequência fundamental e da envoltória e, na linha inferior, as distribuições dos coeficientes de correlação entre os índices de modulação FM e AM para as amostras de cada um dos três registros. O coeficiente de correlação entre os índices AM e FM normalmente está entre 0,5 e 0,7 em todos os três registros (tabela 4.3), mostrando a concordância da informação expressa por eles. Por outro lado, a forma da onda da envoltória não representa, geralmente, uma boa alternativa para estimar parâmetros da curva de  $F_0$ . Segundo a tabela 4.3, a correlação entre estas duas séries é geralmente baixa, em torno de 0,1 e pode ser explicada pelo modelo entre ressonantes e harmônicos. Segundo este modelo, é possível que com um aumento de  $F_0$ , a intensidade do sinal diminua (correlação negativa entre  $F_0$  e a envoltória), ou aumente (correlação positiva entre  $F_0$  e a envoltória), ou ainda, que permaneça praticamente constante (correlação praticamente nula entre  $F_0$  e a envoltória). Em todos estes casos, entretanto, a envoltória carrega alguma informação das variações de  $F_0$  e, por isso, o espectro destas duas séries apresenta maior correlação.



**Figura 4.17:** Profundidade x taxa de todas as amostras da base de dados 2. A resolução escolhida para a escala da taxa foi de 0,1 Hz. Por isso faixas verticais são vistas.



**Figura 4.18:** Histograma dos coeficientes de correlação entre as séries de  $F_0$  e envoltória para vibrato produzido no registro (a) grave, (b) médio e (c) agudo. Histograma dos coeficientes de correlação entre os índices de modulação AM e FM para vibrato produzido no registro (d) grave (e) médio e (f) agudo. O símbolo acima de cada distribuição indica a média, cercada por um desvio padrão.

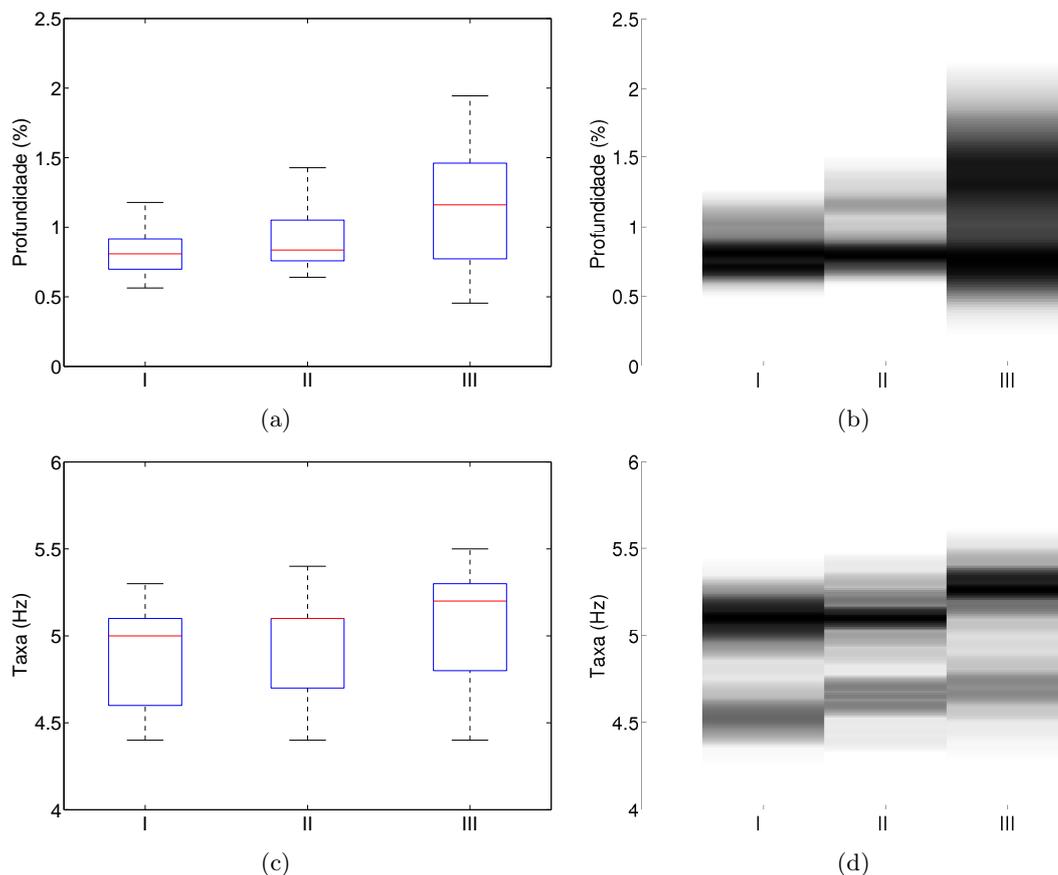
Registro	Correlação entre curvas			Correlação entre índices de modulação		
	Média	Mediana	Desvio Padrão	Média	Mediana	Desvio Padrão
Grave ( $n = 90$ )	0,17	0,19	0,15	0,55	0,56	0,20
Médio ( $n = 90$ )	0,08	0,07	0,18	0,88	0,92	0,12
Agudo ( $n = 90$ )	0,14	0,11	0,22	0,72	0,75	0,18

**Tabela 4.3:** Estatísticas dos coeficientes de correlação: base de dados 2 classificada por registro.

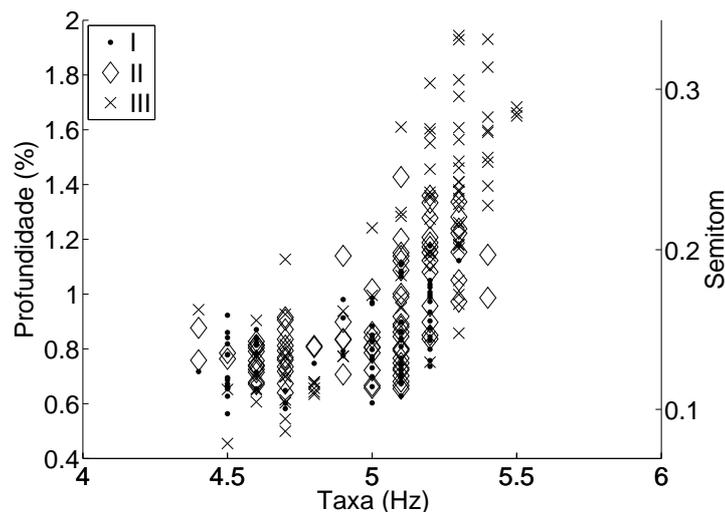
### 4.3.3.2 Divisão por tipo de vibrato

As distribuições da taxa e da profundidade em função do tipo de vibrato estão mostradas na figura 4.19. O painel 4.19(b) indica que há uma profundidade preferencial, por volta de 0,8%, em todos os três tipos de vibrato. Também se vê neste painel que a distribuição da profundidade é assimétrica em todos os tipos de vibrato. Em relação à distribuição da taxa, o painel 4.19(d) mostra a existência de um comportamento bimodal, com as maiores frequências estatísticas em torno de 4,6 Hz e de 5,2 Hz.

A figura 4.20 apresenta a taxa em função da profundidade de todas as amostras, classificadas de acordo com o tipo de vibrato. Observa-se que apenas amostras de vibrato tipo III (figura 3.1) tiveram profundidade maior que 1,5%. Interessantemente, nota-se também que o vibrato tipo III foi o que apresentou as menores profundidades. Em relação à taxa percebe-se algo parecido, isto é, usando o vibrato tipo III o instrumentista produziu amostras com as menores e com as maiores taxas de toda a base de dados. Este espalhamento de valores de taxa e profundidade gerado com o vibrato III está quantificado na coluna “desvio padrão”



**Figura 4.19:** Distribuição da taxa e da profundidade dos indivíduos da base de dados 2. I, II e III indicam, respectivamente, o grupo de emissões produzidas com vibrato dos tipos I, II e III. (a) e (b) Distribuição da profundidade em cada um dos grupos. (c) e (d) Distribuição da taxa em cada um dos grupos. Os tons de cinza nas figuras (b) e (d) indicam a frequência estatística das amostras, sendo tanto maior quanto mais escura for a região.



**Figura 4.20:** Profundidade por taxa de todas as amostras da base de dados 2. A resolução escolhida para a escala da taxa foi de 0,1 Hz. Por isso faixas verticais são vistas.

Tipo de vibrato	Taxa			Profundidade		
	Média (Hz)	Mediana (Hz)	Desvio Padrão (Hz)	Média (%)	Mediana (%)	Desvio Padrão (%)
I ( $n = 90$ )	4,9	5,0	0,3	0,83	0,81	0,15
II ( $n = 90$ )	5,0	5,1	0,3	0,91	0,84	0,20
III ( $n = 90$ )	5,1	5,2	0,3	1,15	1,15	0,40

**Tabela 4.4:** Estatísticas de taxa e profundidade: base de dados 2 classificada por tipo de vibrato.

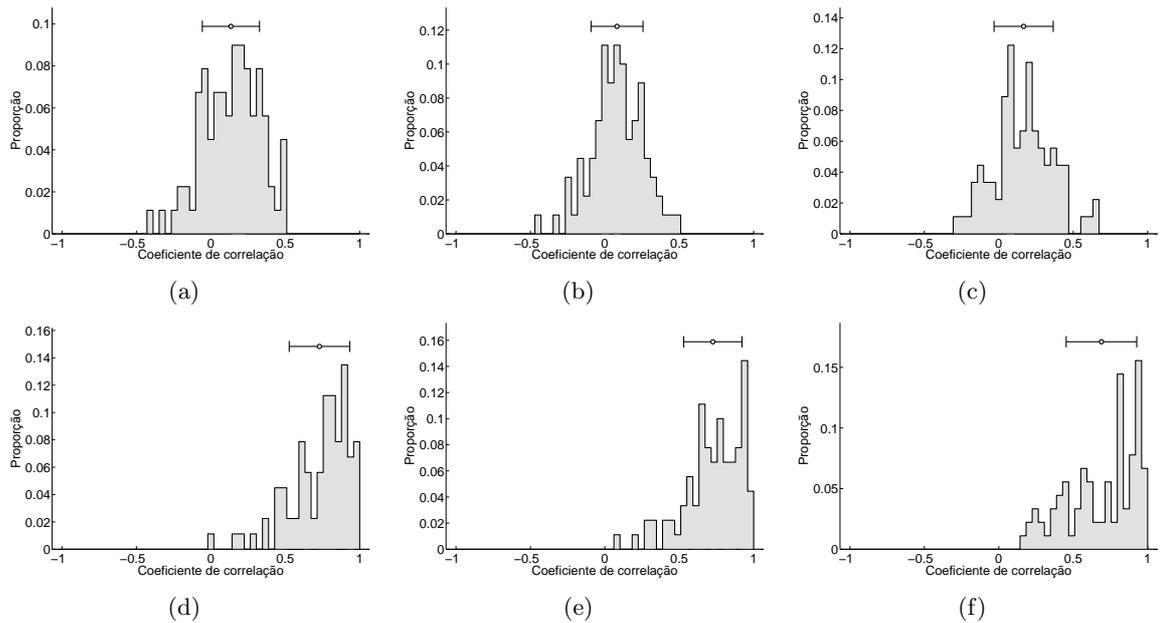
da tabela 4.4, que também apresenta outras estatísticas para cada grupo de vibrato. Uma possível justificativa para a maior variância da taxa e da profundidade com o vibrato tipo III está na dificuldade de execução do movimento. Ao contrário dos outros dois tipos de vibrato, ao executar o tipo III, o instrumentista não pode deixar o dedo auxiliar tocar a corda. Caso isto ocorra, a frequência da nota emitida aumenta e o vibrato fica mais amplo, explicando o aumento de sua profundidade. Para evitar encostar o dedo na corda, o instrumentista pode encurtar o movimento do dedo sobre a nota alvo, gerando um vibrato de pequena profundidade.

Por fim, a figura 4.21 mostra a distribuição dos coeficientes de correlação quando as amostras são separadas pelo tipo de vibrato. Observam-se distribuições parecidas com as da figura 4.18 e as explicações dadas para aquele caso se aplicam aqui.

## 4.4 Tremor Vocal

### 4.4.1 Exemplo 1: tratamento de tremor vocal

O caso analisado é de um paciente com diagnóstico confirmado de tremor vocal, virgem de tratamento, e que estava medicado há 2 meses com Propranolol 40 mg (2x ao dia). Duas



**Figura 4.21:** Histograma dos coeficientes de correlação entre as séries de  $F_0$  e envoltória para (a) vibrato tipo I (b) vibrato tipo II e (c) vibrato tipo III. Histograma dos coeficientes de correlação entre os índices de modulação AM e FM para (d) vibrato tipo I (e) vibrato tipo II e (f) vibrato tipo III. O símbolo acima de cada distribuição indica a média, cercada por um desvio padrão.

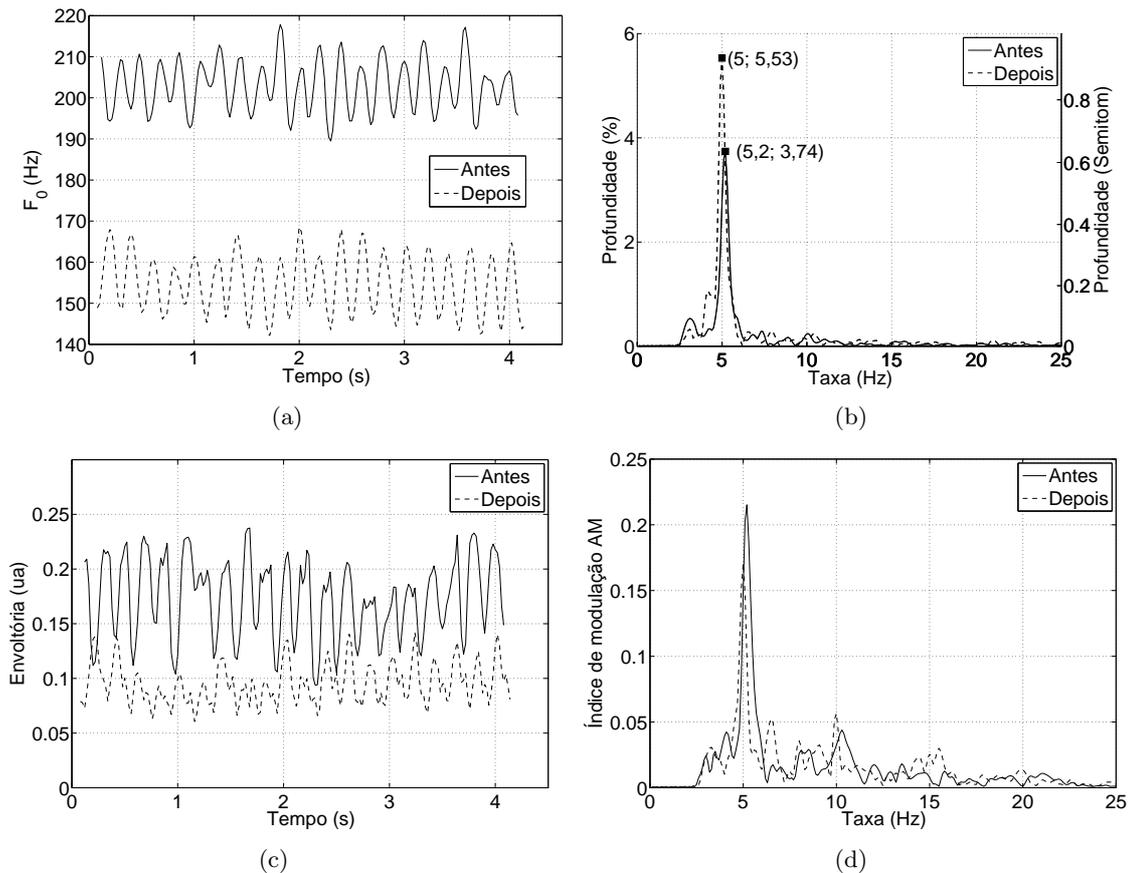
emissões da vogal /a/, uma antes e outra depois do tratamento, foram selecionadas da base de dados 3 (seção 3.2.3) para investigar as alterações provocadas pela medicação.

A figura 4.22(a) mostra a frequência fundamental das emissões anterior e posterior à aplicação do medicamento. De acordo com esta figura, a forma das séries de frequência e amplitude das oscilações parece não variar por causa do tratamento. A principal diferença entre as duas curvas é o seu valor médio. Antes do tratamento,  $F_{0média} \approx 203$  Hz e, depois,  $F_{0média} \approx 155$  Hz.

Na figura 4.22(b), o aparente aumento no valor da extensão do tremor depois do tratamento se destaca. Isto é, em grande parte, provocado pela redução do valor médio da frequência fundamental após o tratamento e não devido a um aumento significativo da amplitude das oscilações. A amplitude de aproximadamente 8 Hz do tremor, vista na figura 4.22(a), equivalia a  $8/203 \approx 4\%$  de  $F_{0média}$  antes da medicação e passou a representar  $8/155 \approx 5\%$  de  $F_{0média}$  depois da injeção da toxina. Além disso, nota-se também nesta figura que a taxa do tremor diminuiu um pouco; fato que dificilmente é percebido diretamente na forma da curva de  $F_0$ .

A taxa de oscilação do sinal da envoltória também diminuiu, como está mostrado nos gráficos 4.22(c) e 4.22(d). Além disso, comparando as figuras 4.22(b) e 4.22(d), nota-se que a taxa da envoltória apresentou o mesmo comportamento da taxa da curva de frequência fundamental tanto antes quanto depois da medicação.

A análise da regularidade nas amostras está mostrada nas figuras 4.23 e 4.24 para as emissões anterior e posterior ao tratamento, respectivamente. As curvas de  $K^{-1}$ , que estão



**Figura 4.22:** Características do tremor antes e depois do tratamento com toxina botulínica. Nestes gráficos, excepcionalmente, a linha cheia indica medidas na amostra coletada antes do tratamento e linha tracejada, medidas na amostra de depois do tratamento. No restante deste capítulo, a linha cheia está relacionada a medidas feitas na curva de  $F_0$  e a linha tracejada, a medidas na envoltória.

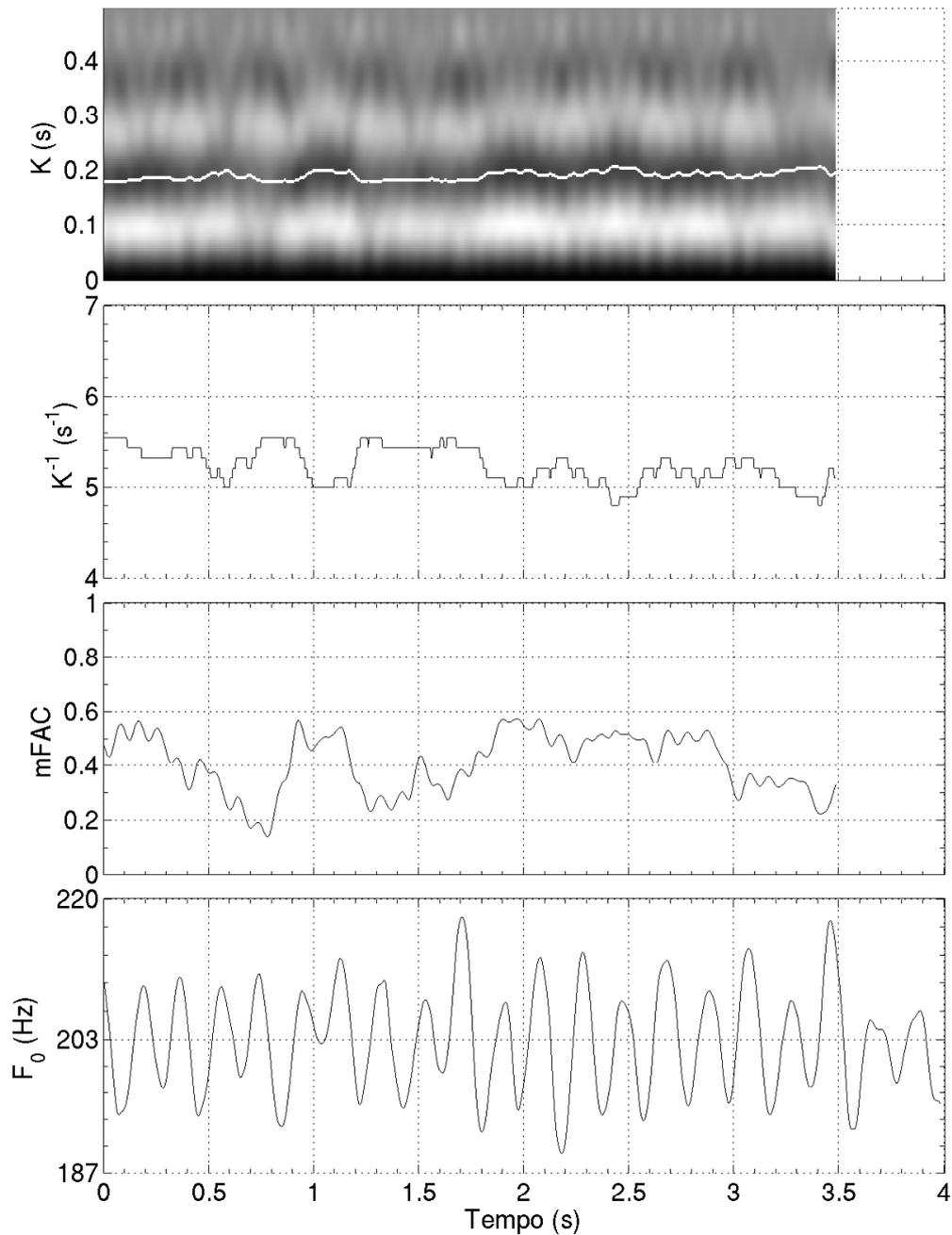
relacionadas ao comportamento instantâneo da taxa de  $F_0$ , têm excursões semelhantes nos dois casos, mas as médias são diferentes. Antes e depois do tratamento, elas valem, respectivamente,  $5,0 \text{ s}^{-1}$  e  $5,2 \text{ s}^{-1}$ , os mesmos valores das taxas mostradas na figura 4.22(b). Já a curva de mFAC, relacionada à semelhança entre ciclos vizinhos, apresenta diferenças nos dois casos. Observa-se que, na pré-medicação, a curva de mFAC varia entre aproximadamente 0,1 a 0,6, enquanto na curva de depois do tratamento, esta faixa é reduzida para 0,4 até 0,6, apontando uma melhoria da regularidade na emissão pós-tratamento.

Primeiramente, observa-se que o comportamento da curva de frequência fundamental da emissão pós-tratamento é mais regular que aquela da emissão pré-tratamento, pois apresenta curvas de  $K^{-1}$  e mFAC um pouco mais planas. Isto é consequência do comportamento da curva de  $F_0$  vista no painel inferior de cada uma destas figuras. Os ciclos consecutivos da curva para o caso pós-medicação são mais parecidos entre si que no caso pré-medicação.

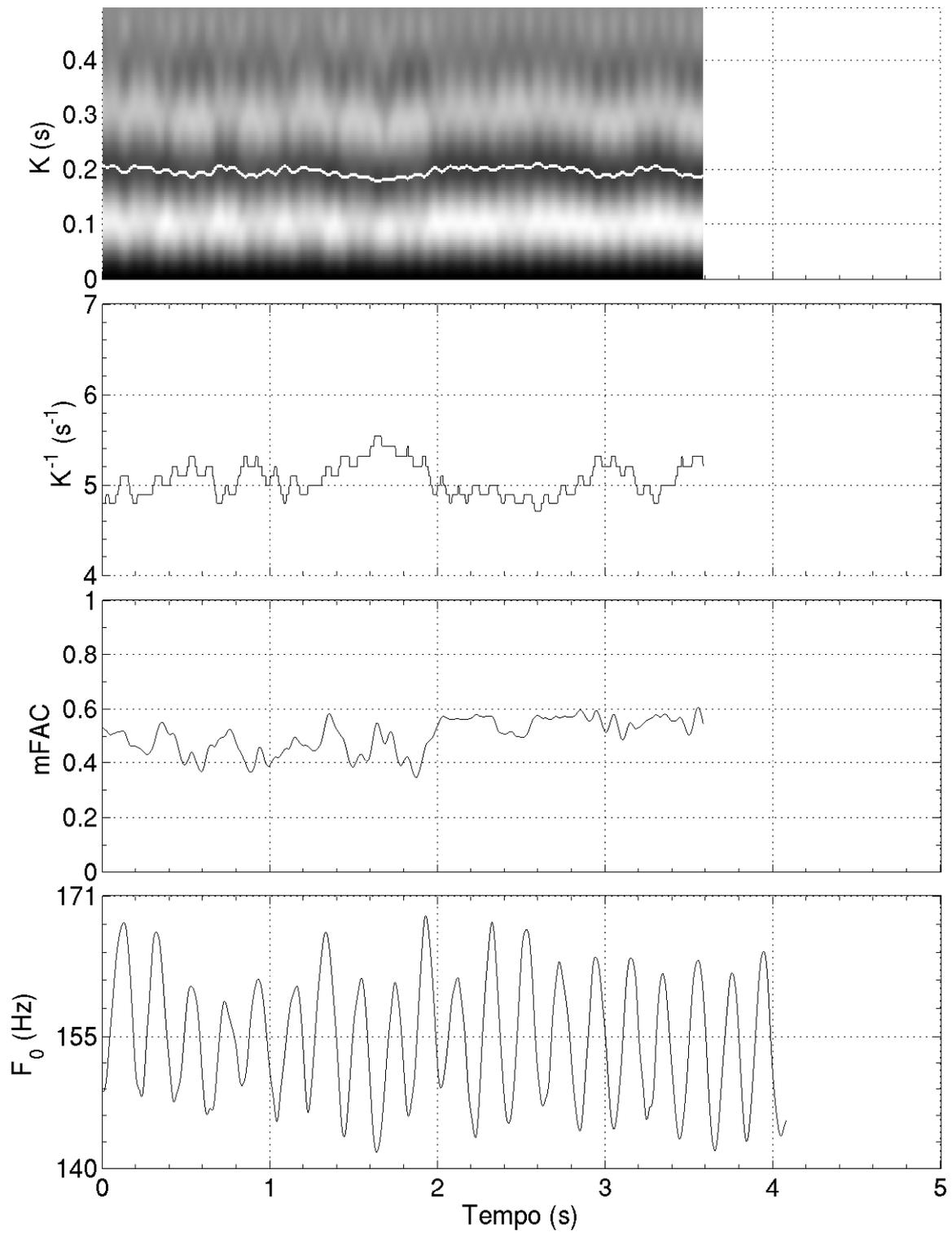
Esta análise mostrou que a medicação diminuiu o valor médio da frequência fundamental. Porém, praticamente não interferiu nas características, i.e. taxa, profundidade, do tremor. Uma discussão sobre possíveis explicações biológicas para estas observações foge do escopo

deste trabalho.

Por último, é importante ressaltar que esta análise foi feita com uma só amostra de antes do tratamento e uma só amostra pós-tratamento. Este estudo com um único exemplar é pouco significativo para avaliar a eficiência do medicamento. Contudo, é suficiente para mostrar que a ferramenta desenvolvida neste trabalho pode auxiliar no acompanhamento do tratamento de tremor vocal.



**Figura 4.23:** Análise de regularidade do tremor vocal antes da medicação. A série de  $F_0$  é a mesma da curva de linha cheia na figura 4.22(a), interpolada para melhorar a visualização de detalhes.



**Figura 4.24:** Análise de regularidade do tremor vocal após a medicação. A série de  $F_0$  é a mesma da curva de linha tracejada na figura 4.22(a), interpolada para melhorar a visualização de detalhes.

#### 4.4.2 Exemplo 2: doença de Parkinson

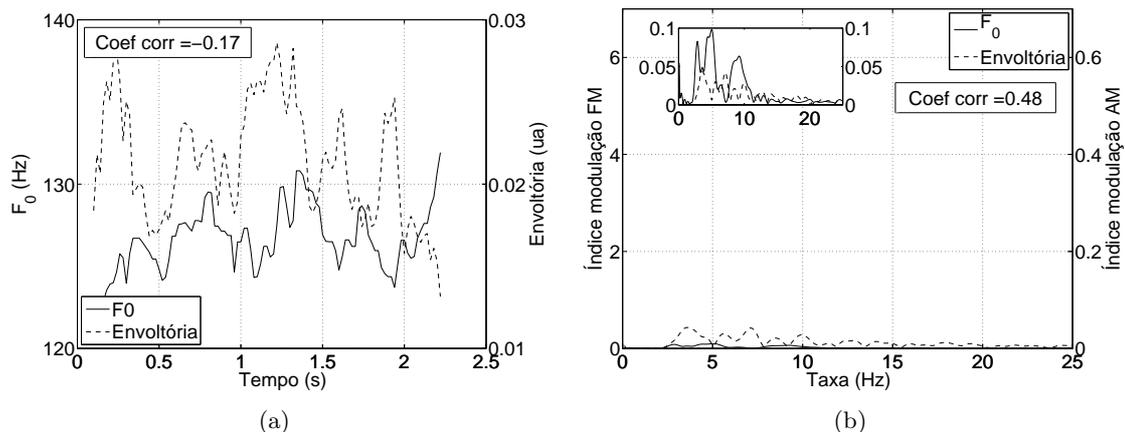
A emissão de um indivíduo com doença de Parkinson da base de dados 3 foi analisada nesta seção. Como já dito na seção 3.2.3, os indivíduos desta base de dados não têm avaliação perceptiva de tremor vocal. A amostra selecionada foi julgada perceptivamente pelo autor como livre de tremor vocal. Ela foi escolhida a fim de ilustrar as diferenças entre o sinal de voz com tremor diagnosticado no exemplo anterior e o sinal de voz julgado sem tremor.

A figura 4.25(a) mostra a série de  $F_0$  e a envoltória da emissão escolhida. Nota-se que as duas curvas flutuam, grosseiramente, a uma taxa de 2 ciclos por segundo. Esta componente não está na faixa de frequências de interesse e não é considerada na análise. Além dessa oscilação, aparentemente não há perturbações, na faixa do tremor, como aquelas evidentes na figura 4.22(a). O gráfico 4.25(b) confirma este fato. Há pequenas elevações em torno de 5 e de 10 Hz, mas suas respectivas extensões são muito baixas, não chegando sequer a 1 Hz.

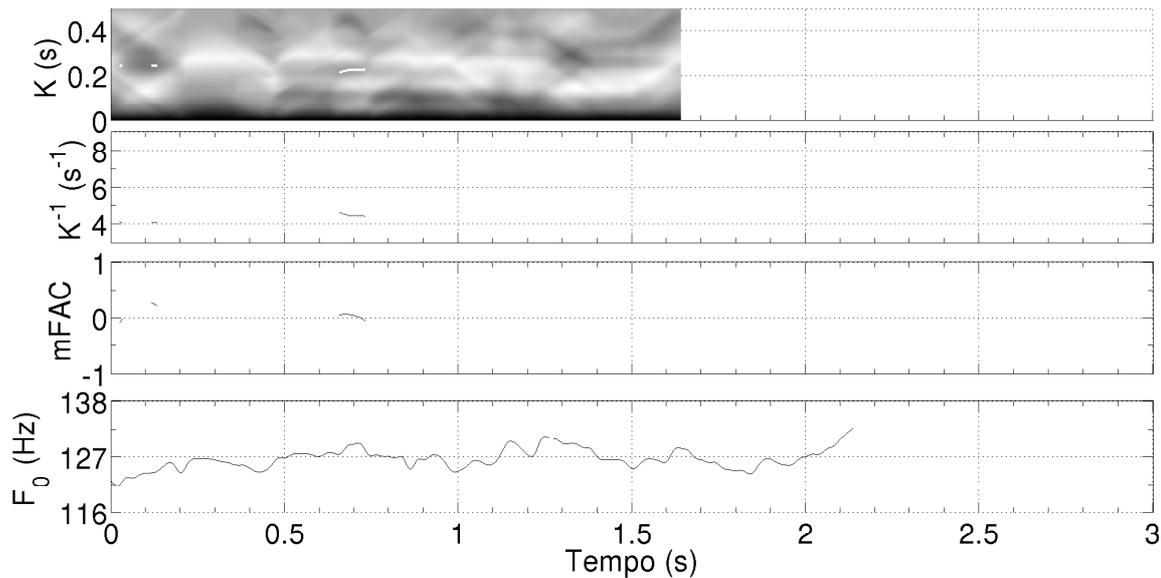
A regularidade deste sinal é analisada na figura 4.26. Como a curva de  $F_0$ , mostrada no painel inferior desta figura, não apresenta periodicidades dentro da faixa de frequência de interesse, nada é rastreado pelo algoritmo. Isto indica que a emissão usada neste exemplo não é perturbada por tremor vocal, confirmando o julgamento perceptivo do autor. Isso sugere que, além de poder ser aplicada para acompanhar a evolução do tratamento do tremor, esta técnica também poderá auxiliar no diagnóstico desta perturbação. Estas possíveis aplicações serão assuntos de trabalhos futuros.

#### 4.4.3 Grupo

A figura 4.27 mostra a distribuição da taxa e da profundidade dos dois grupos da base de dados 3. As distribuições de taxa e profundidade do grupo formado por indivíduos afetados por Parkinson são assimétricas e as medianas de ambas são menores que as respectivas medianas do outro grupo. Além disso, observa-se que os dois grupos têm comportamentos



**Figura 4.25:** Características da voz de um paciente com doença de Parkinson. Na figura (b) há detalhes que mostram os pormenores das curvas representadas. O eixo vertical desta figura foi ajustado de modo a facilitar comparações com outros resultados apresentados no texto.



**Figura 4.26:** Análise de regularidade da voz de um paciente com doença de Parkinson. A curva de  $F_0$  é a mesma da figura 4.25(a), interpolada para melhorar a visualização de detalhes.

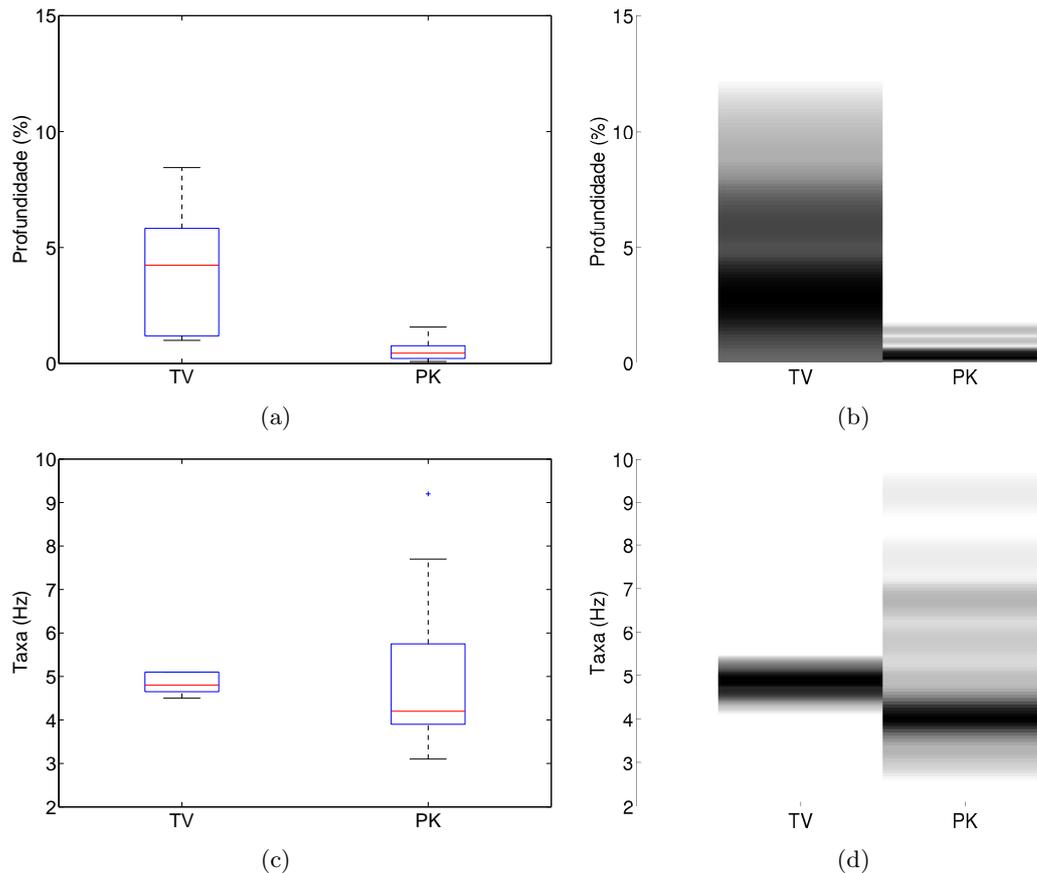
contrários: enquanto o grupo de tremor vocal tem a taxa concentrada e a profundidade distribuída, o grupo com doença de Parkinson tem a taxa dispersa e a profundidade concentrada. Veja na tabela 4.5 um resumo dos valores das médias, medianas e desvios padrão da taxa e da profundidade de cada um dos grupos desta base de dados.

Os valores da tabela 4.5 foram calculados com todos os indivíduos de cada grupo, sem separá-los por sexo. A média e a mediana da taxa no grupo com tremor vocal estão um pouco abaixo dos valores publicados por Ramig e Shipp (1987) e por Winholtz e Ramig (1992). Ramig e Shipp (1987) encontraram, em um trabalho com 9 pacientes, uma taxa média de 6,77 Hz com 2,63 Hz de desvio padrão e Winholtz e Ramig (1992), medindo em 6 homens e 6 mulheres, encontraram, respectivamente, 5,5 Hz e 5,8 Hz para o valor mediano da taxa. Os valores da profundidade do grupo com tremor vocal estão de acordo com os valores de Ramig e Shipp (1987), que mediram 4,25% com desvio padrão de 0,52% e de Winholtz e Ramig (1992), que mediram 4,8% para homens e 7,4% para mulheres.

Os resultados da tabela 4.5 para o grupo de indivíduos com doença de Parkinson não foram comparados com outros trabalhos da literatura porque a grande parte das emissões deste grupo não parece, para o autor, ter tremor vocal, mas sim outro tipo de disfonia. O

Grupo	Taxa			Profundidade		
	Média (Hz)	Mediana (Hz)	Desvio Padrão (Hz)	Média (%)	Mediana (%)	Desvio Padrão (%)
Tremor Vocal ( $n = 5$ )	4,8	4,8	0,3	3,98	4,24	3,05
Parkinson ( $n = 32$ )	5,6	4,2	1,4	0,55	0,45	0,43

**Tabela 4.5:** Estatísticas de taxa e profundidade: base de dados 3.

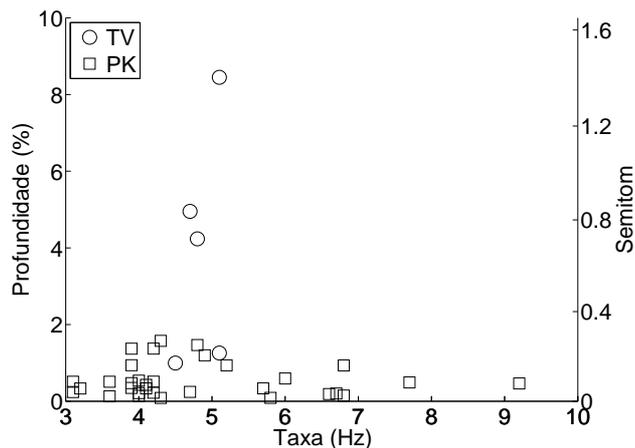


**Figura 4.27:** Distribuição da taxa e da profundidade dos indivíduos da base de dados 3. TV indica o grupo de informantes com tremor vocal diagnosticado e PK indica o grupo de indivíduos com doença de Parkinson. (a) e (b) Distribuição da profundidade em cada um dos grupos (c) e (d) Distribuição da taxa em cada um dos grupos. Os tons de cinza nas figuras (b) e (d) indicam a frequência estatística das amostras, sendo tanto maior quanto mais escura for a região.

presente trabalho somente apresenta uma hipótese para explicar a forma da distribuição da taxa e da profundidade deste grupo, sendo que um estudo mais detalhado com estas amostras é deixado para um trabalho futuro.

A forma da distribuição da taxa e da profundidade no grupo composto por indivíduos com doença de Parkinson pode ser explicada pela presença de diversas emissões julgadas, pelo autor, como livres de tremor vocal. Nestas emissões, a ausência de tremor faz com que a profundidade das oscilações na faixa entre 4 e 15 Hz seja pequena (veja um exemplo na figura 4.25(b)), justificando o que está mostrado nas figuras 4.27(a) e 4.27(b). Além disso, devido à ausência de perturbações destacadas na faixa 4–15 Hz, a taxa de maior amplitude nas emissões deste grupo não ficou concentrada em uma só raia (figura 4.28), explicando a dispersão mostrada nas figuras 4.27(c) e 4.27(d).

A figura 4.28 mostra a distribuição da profundidade em função da distribuição da taxa nos dois grupos. É evidente nesta figura que o tremor está concentrado em torno da taxa de 5 Hz, enquanto as emissões dos pacientes com doença de Parkinson estão concentradas em torno da profundidade de menos de 1%. O resultado do teste Kruskal-Wallis sobre estes grupos



**Figura 4.28:** Profundidade x taxa de todas as amostras da base de dados 3.

motrou que há diferença estatística entre estes grupos quando se compara a profundidade ( $p \approx 10^{-7}$ ) mas, o mesmo não ocorre em relação à taxa ( $p = 0,98$ ).

Por último, a figura 4.29 apresenta a distribuição dos coeficientes de correlação entre as curvas de frequência fundamental e envoltória e também entre os índices de modulação FM e AM. Os resultados estão resumidos na tabela 4.6. Nota-se que, assim como para a base de dados 2, a média distribuição dos índices de correlação entre as curvas de  $F_0$  e de envoltória é menor que a média da distribuição dos coeficientes de correlação entre os índices AM e FM.

Grupo	Correlação entre curvas			Correlação entre índices de mod.		
	Média	Mediana	Desvio Padrão	Média	Mediana	Desvio Padrão
Tremor Vocal ( $n = 5$ )	0,41	0,55	0,36	0,78	0,88	0,18
Parkinson ( $n = 32$ )	0,02	-0,02	0,39	0,61	0,61	0,12

**Tabela 4.6:** Estatísticas dos coeficientes de correlação: base de dados 3.

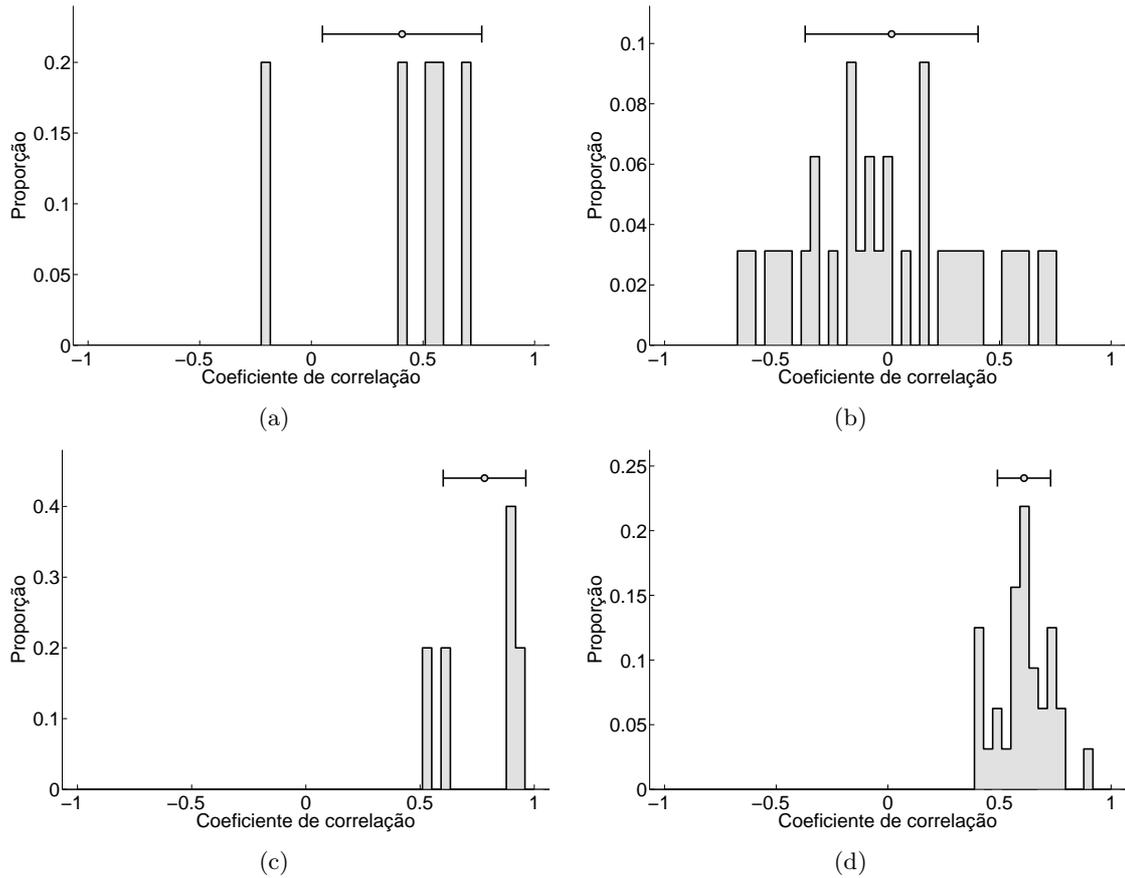
## 4.5 Canto

### 4.5.1 Exemplo 1: vibrato no canto lírico

As colunas da esquerda e da direita da figura 4.30 exibem análises de duas amostras de vibrato produzidas por diferentes cantores da base de dados 4.

Neste exemplo, o que mais chama a atenção é o comportamento das envoltórias das duas amostras. Observando as figuras 4.30(a) e 4.30(b), percebe-se que as curvas de  $F_0$  estão centradas em torno de frequências um pouco diferentes e possuem praticamente a mesma taxa e profundidade; já as séries das envoltórias aparentam ter pouco em comum. No vibrato da coluna da direita não há uma relação tão clara entre as curvas da envoltória e de  $F_0$  como a existente na emissão da coluna esquerda, na qual as duas séries estão em fase.

O espectro destas amostras, ilustrados nas figuras 4.30(e) e 4.30(f) pelos índices de modulação AM e FM, esclarecem as diferenças vistas ao longo do contorno de  $F_0$  e da envoltória.



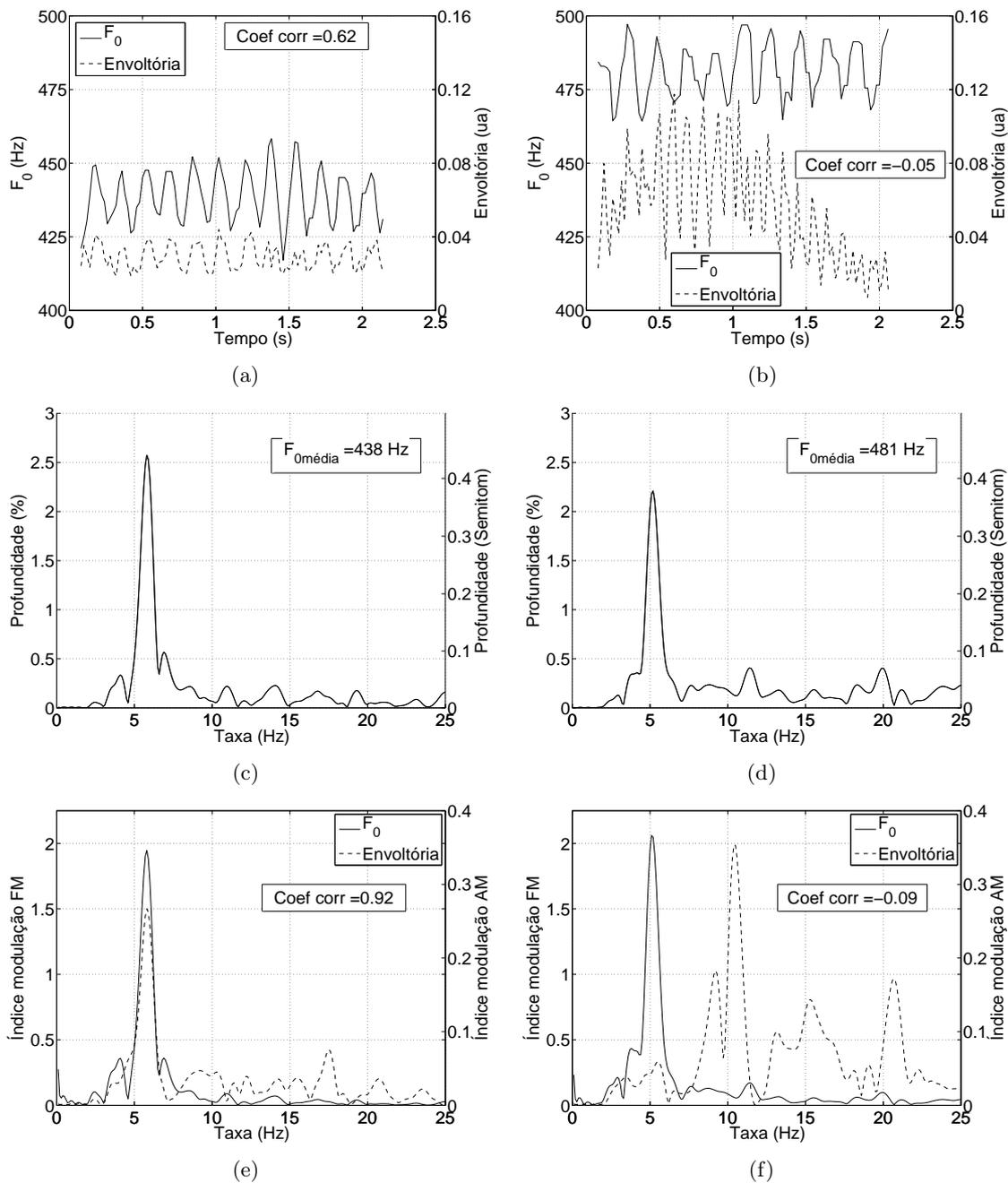
**Figura 4.29:** Histograma dos coeficientes de correlação entre as séries de  $F_0$  e envoltória para o grupo com (a) tremor vocal confirmado perceptivamente e (b) doença de Parkinson. Histograma dos coeficientes de correlação entre os índices de modulação AM e FM para o grupo com (c) tremor vocal confirmado perceptivamente e (d) doença de Parkinson. O símbolo acima de cada distribuição indica a média, cercada por um desvio padrão.

A amostra de vibrato da coluna da esquerda apresenta flutuações de frequência e de amplitude na mesma taxa, porém, o mesmo não acontece com a amostra da coluna da direita. A diferença entre a taxa da oscilação de  $F_0$  e a da envoltória também pode ser explicada pela interação harmônico-formante (figura 1.2), mas merece estudo detalhado (veja Horii e Hata (1988, pags. 305-306)).

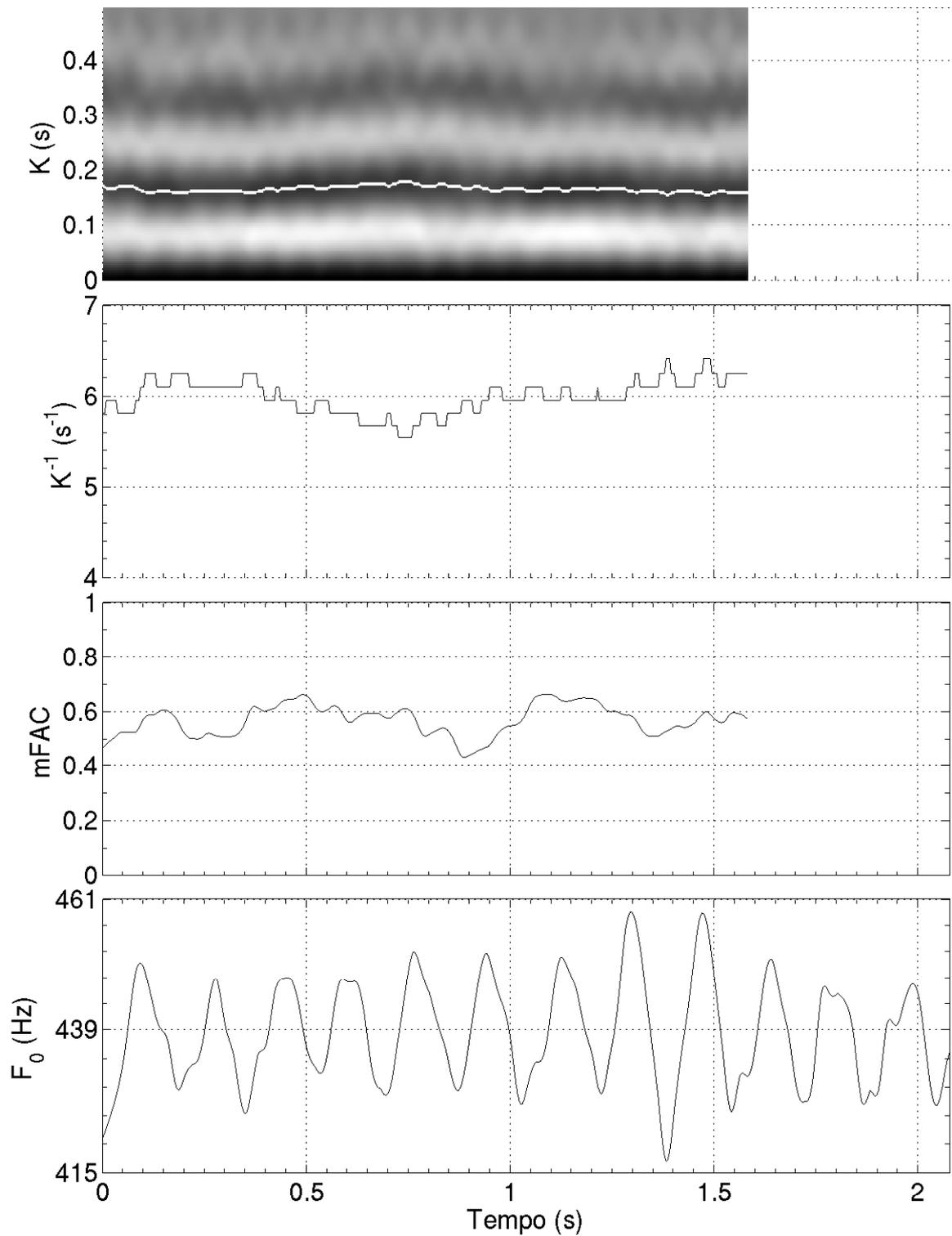
Testes informais de escuta realizados com um músico profissional indicaram que esta diferença entre os espectros da envoltória e de  $F_0$  não eram percebidos. O participante escutou as duas amostras analisadas neste exemplo e afirmou não perceber uma diferença de duas vezes entre as taxas dos vibratos. Um estudo mais detalhado sobre a relação entre as taxas da envoltória e de  $F_0$  sobre a percepção foge do escopo deste trabalho e é sugerido como trabalho futuro.

As figuras 4.31 e 4.32 trazem a análise de regularidade destas amostras de vibrato. Assim como aconteceu em exemplos de outras bases de dados, a amostra com maior taxa apresentou também maior valor médio de  $K^{-1}$ . Além disso, é interessante notar que mesmo na presença

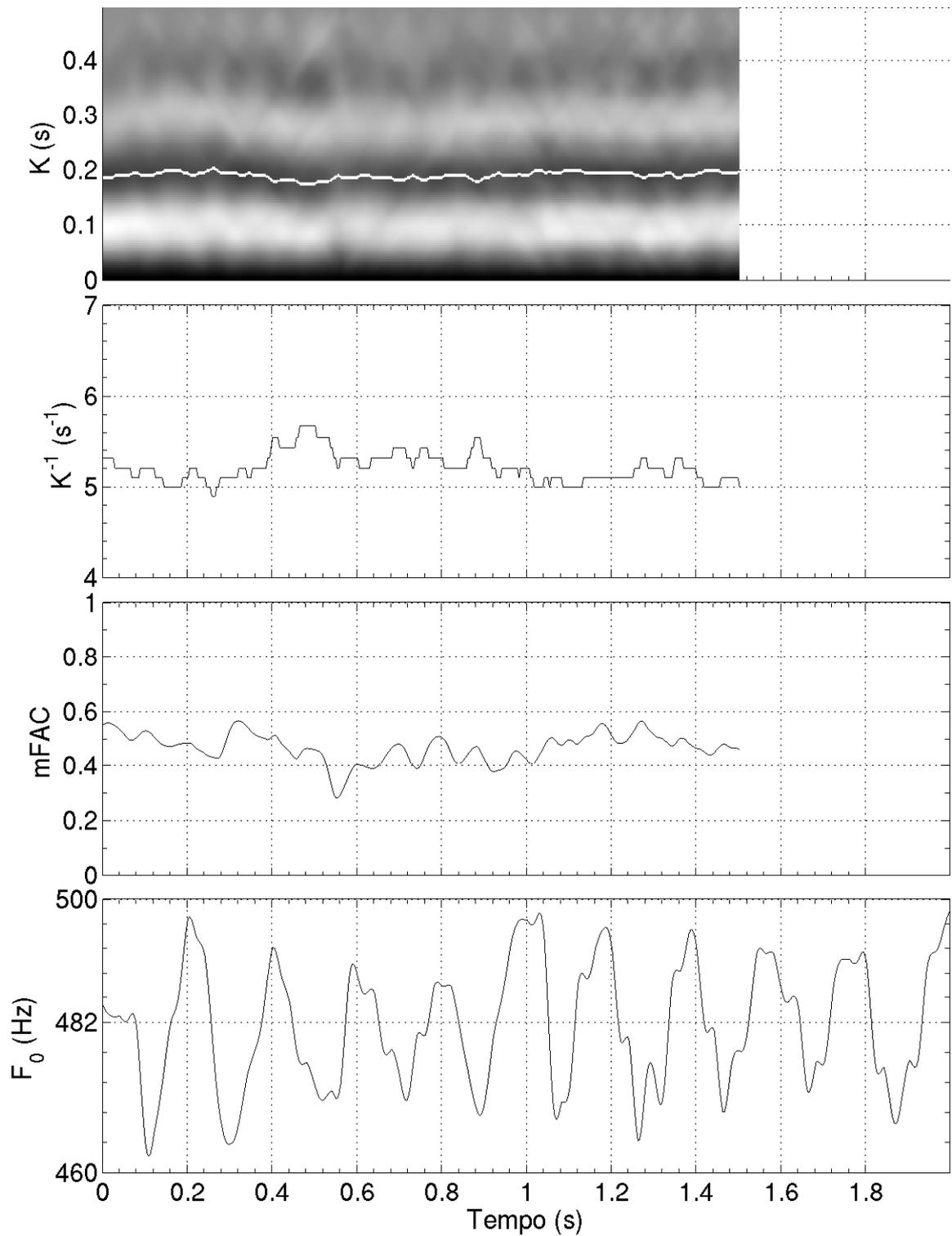
de harmônicos no contorno de  $F_0$ , as curvas de mFAC não apresentam quebras, como as da figura 4.3, indicando que os ciclos adjacentes são similares.



**Figura 4.30:** Análise de vibrato no canto lírico. (a), (c) e (e) são do mesmo cantor, chamado, no texto, de cantor I. (b), (d) e (f) são todos do cantor II. Observe que as escalas verticais de (a) e (b) são diferentes. As escalas de (c) – (f) foram escolhidas para facilitar a comparação com outras figuras deste capítulo.



**Figura 4.31:** Análise de regularidade de vibrato no canto lírico. A curva de  $F_0$  é a mesma da figura 4.30(a), interpolada para melhorar a visualização de detalhes.



**Figura 4.32:** Análise de regularidade de vibrato no canto lírico. A curva de  $F_0$  é a mesma da figura 4.30(b), interpolada para melhorar a visualização de detalhes.

### 4.5.2 Exemplo 2: vibrato no canto sertanejo

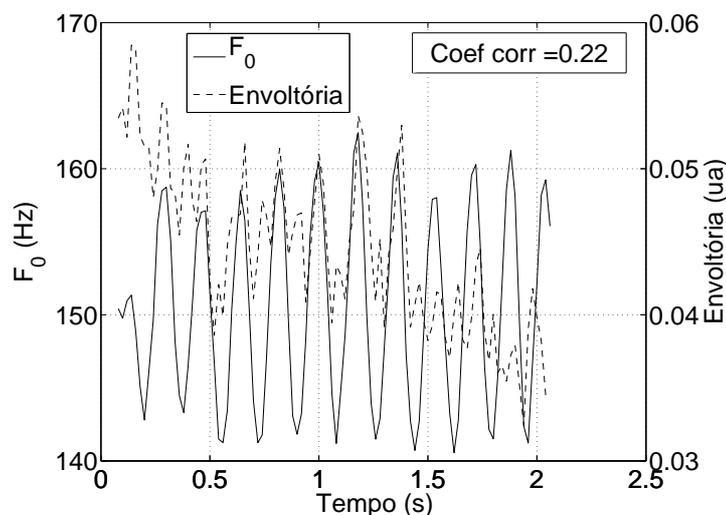
A envoltória e a curva de  $F_0$  de vibrato produzido por um cantor sertanejo estão mostradas na figura 4.33. Nota-se que as flutuações na série da frequência fundamental se mantêm em torno de  $F_{0média} = 150$  Hz durante todo trecho, enquanto as da envoltória acompanham uma reta de inclinação negativa. Esta diminuição da envoltória sugere que a intensidade da emissão diminui com o passar do tempo, o que é esperado, pois, este vibrato foi produzido em um momento que precede silêncio na canção.

A profundidade desta amostra de vibrato é maior que a das emissões no canto lírico mostradas no exemplo anterior (figuras 4.30(c) e 4.30(d)). No caso do canto sertanejo, o vibrato foi produzido sobre  $F_{0média} = 150$  Hz, com aproximadamente  $9/150 = 6\%$  de profundidade (figura 4.34(a)). Já nas amostras de canto lírico, (i)  $F_{0média} = 438$  Hz e profundidade de  $\approx 10/438 \approx 2\%$ ; (ii)  $F_{0média} = 481$  Hz com profundidade  $\approx 10/481 \approx 2\%$ . Em relação ao índice FM, nota-se um comportamento oposto, isto é, o índice FM desta amostra de vibrato no canto sertanejo vale  $\approx 9/5,3 \approx 1,7$  (figura 4.34(b)), menor que os dos cantores líricos, que valem  $\approx 10/5 = 2\%$  (figuras 4.30(e) e 4.30(f)). A consequência destes resultados sobre a percepção não foi explorada neste trabalho e é sugerida como estudo futuro.

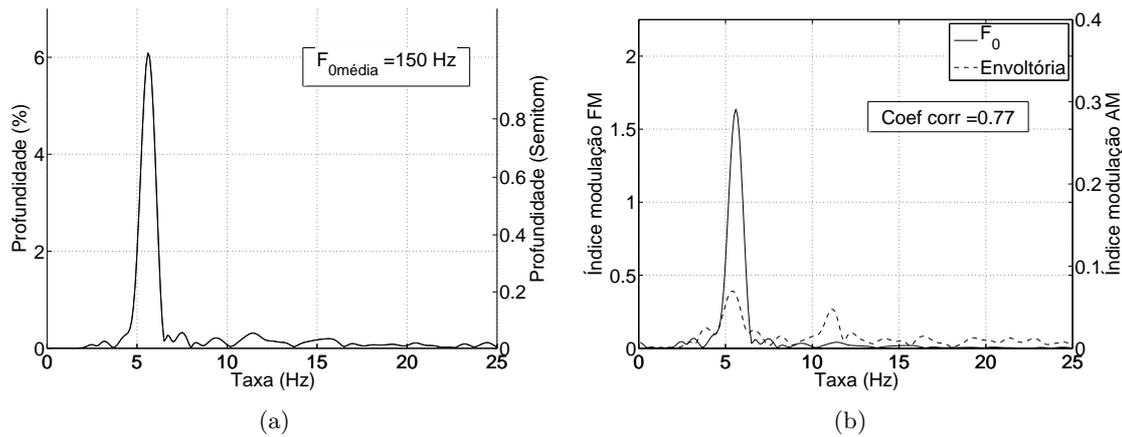
A regularidade da curva de  $F_0$  desta amostra de vibrato é investigada na figura 4.35. Nota-se que o cantor emite ciclos bem similares, o que é quantificado em curvas praticamente planas de  $K^{-1}$  e de mFAC. Aparentemente, esta série é mais regular que a dos cantores líricos mostradas no exemplo anterior. Entretanto, perceptivamente para o autor ela soa menos agradável.

### 4.5.3 Grupo

O comportamento da extensão e da taxa do vibrato entre as amostras desta base de dados está mostrado na figura 4.36. Nota-se uma menor dispersão dos valores de taxa e de



**Figura 4.33:** Vibrato produzido por um cantor sertanejo da base de dados 4.

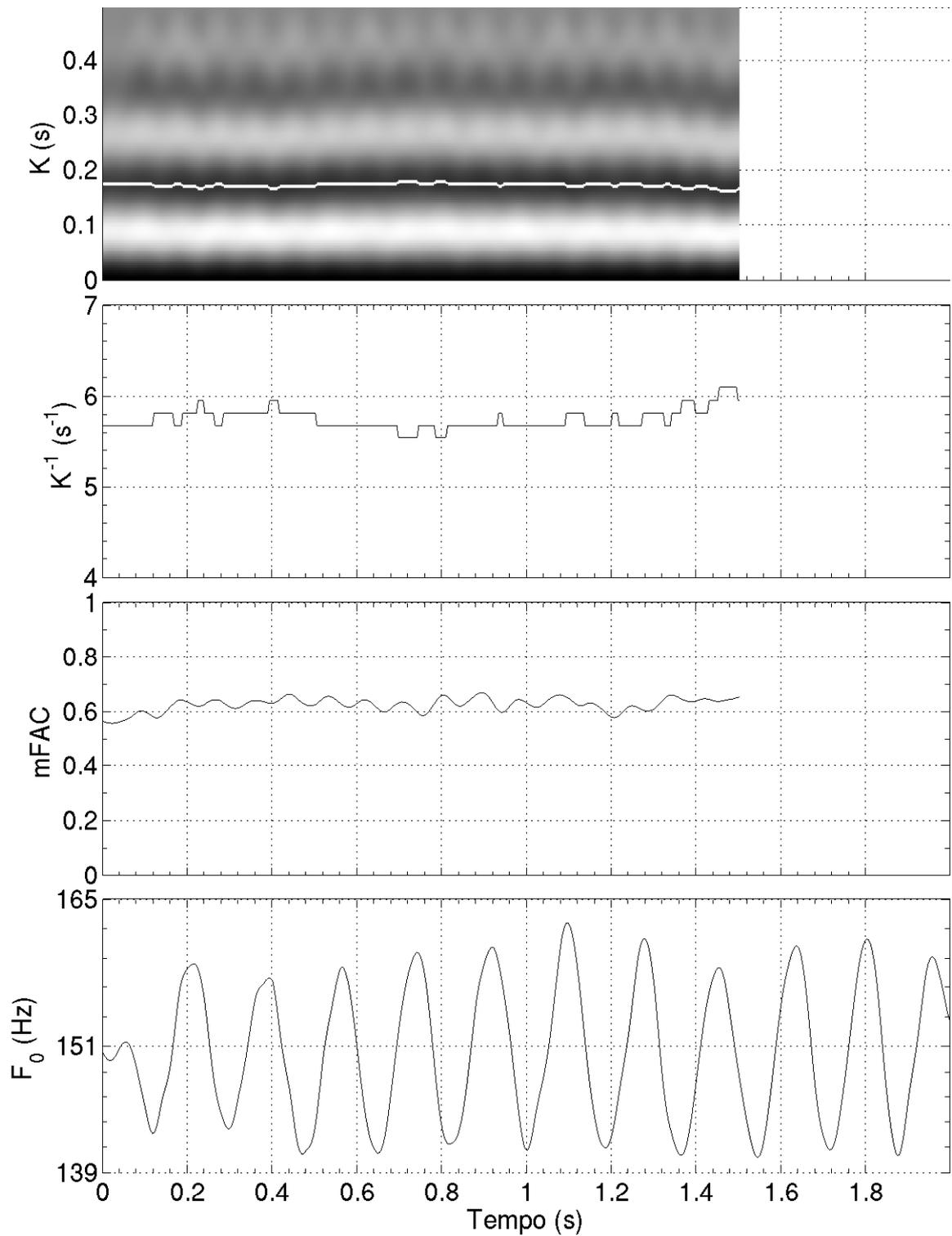


**Figura 4.34:** Espectro de uma amostra de vibrato no canto sertanejo.

profundidade do grupo de cantores líricos em relação ao de cantores sertanejos. Além disso, observa-se uma maior simetria das distribuições de taxa e profundidade no grupo de cantores sertanejo. Observa-se também a presença de um *outlier* no grupo de cantores sertanejo. Este *outlier* foi estudado no segundo exemplo desta base de dados e, como mostrado, a razão do alto valor percentual de profundidade foi o fato dele produzir o vibrato em um registro grave, i.e., 150 Hz.

Os valores de média, mediana e desvio padrão destas distribuições estão resumidos na tabela 4.7. A média de 5,4 Hz para a taxa do vibrato no canto lírico está dentro da faixa esperada por Sundberg (1995). Ela foi um pouco menor que a taxa média de 6 Hz obtida por Prame (1994) através de medidas manuais em um grupo de 10 cantores profissionais; porém, um pouco maior que a taxa média de Bretos e Sundberg (2003), de aproximadamente 5 Hz, medida de maneira automática em 8 estudantes de canto. Quanto ao valor da profundidade neste grupo, o valor médio encontrado, 3,4%, foi maior que o de Prame (1997), 2,1%, e menor que o de Howes et al. (2004), 3,6%, ambos medidos manualmente. Contudo, manteve-se na faixa de valores esperados no canto lírico (Sundberg, 1995). Em relação ao grupo de cantores sertanejo, o valor médio da taxa está dentro do intervalo de 5,2 Hz a 7,8 Hz medido por Rosa (2003) e também dentro do intervalo medido por de Almeida Bezerra et al. (2009), de 5,00 Hz a 6,56 Hz. O valor médio da profundidade ficou dentro da faixa encontrada por Rosa (2003), 0,05% a 4,75%, mas não daquela publicada por de Almeida Bezerra et al. (2009), 1,58% a 2,82%.

A figura 4.37 apresenta conjuntamente taxa e profundidade dos casos estudados. Note que os valores máximo e mínimo da taxa e da profundidade acontecem em amostras do grupo de cantores sertanejo, sugerindo um menor controle do vibrato entre os indivíduos deste grupo. O teste de Kruskal-Wallis retornou  $p = 0,06$  para a taxa e  $p = 0,32$  para a profundidade. Normalmente, estes valores seriam julgados como indicativos de ausência de diferenças estatisticamente significativas. Porém, como a quantidade de amostras nas bases de dados é pequena, pode-se aceitar a existência de diferenças significativas, especialmente em relação à taxa.

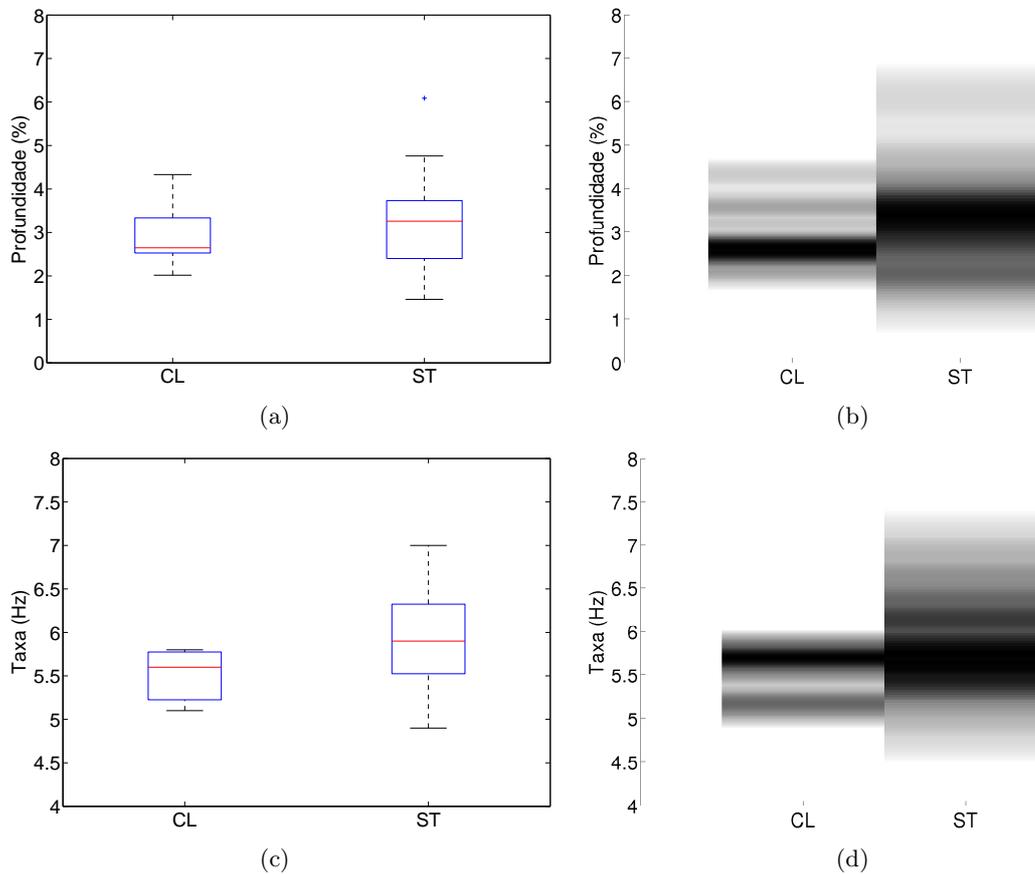


**Figura 4.35:** Análise de regularidade de vibrato no canto sertanejo. A curva de  $F_0$  é a mesma da figura 4.34(a), interpolada para melhorar a visualização de detalhes.

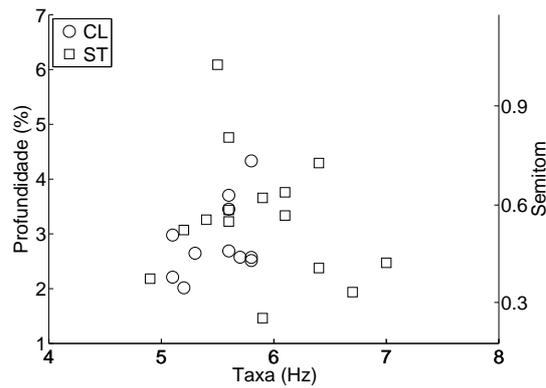
As distribuições dos índices de correlação entre as curvas de  $F_0$  e envoltória e entre os índices de modulação AM e FM estão mostradas, respectivamente em 4.38(a) e 4.38(c), para o grupo de canto lírico e nas figuras 4.38(b) e 4.38(d), para o grupo de canto sertanejo. As estatísticas resumidas na tabela 4.8 mostram que as informações contidas no espectro de  $F_0$  e da envoltória das amostras desta base de dados também compartilham semelhanças, fato semelhante ao relatado para as bases de dados 2 e 3.

Grupo	Taxa			Profundidade		
	Média (Hz)	Mediana (Hz)	Desvio Padrão (Hz)	Média (%)	Mediana (%)	Desvio Padrão (%)
Canto Lírico ( $n = 11$ )	5,5	5,6	0,3	2,88	2,65	0,69
Sertanejo ( $n = 15$ )	5,9	5,9	0,6	3,29	3,26	1,18

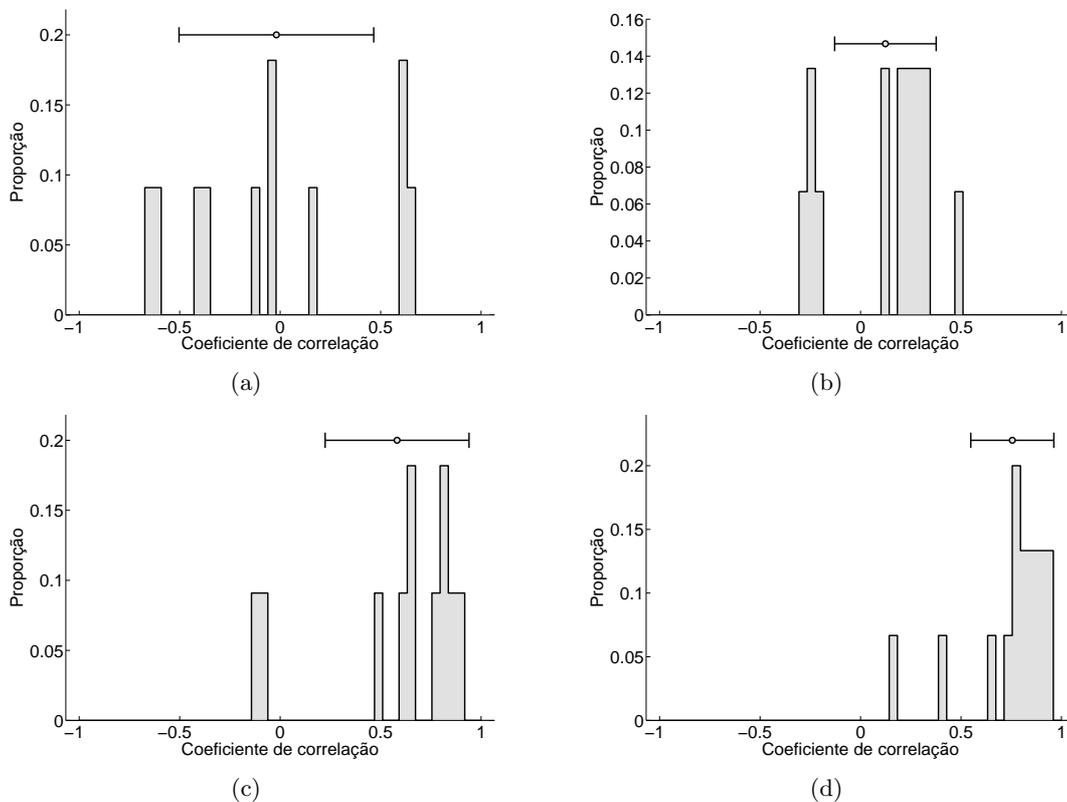
**Tabela 4.7:** Estatísticas de taxa e profundidade: base de dados 4.



**Figura 4.36:** Distribuição da taxa e da profundidade dos indivíduos da base da dados 4, sendo que CL se refere ao grupo de cantores lírico e ST, ao grupo de cantores sertanejo. (a) e (b) Distribuição da profundidade (c) e (d) Distribuição da taxa. Os tons de cinza nas figuras (b) e (d) indicam a frequência estatística das amostras, sendo tanto maior quanto mais escura for a região.



**Figura 4.37:** Profundidade x taxa de todas as amostras da base de dados 4. CL denota o grupo de cantores líricos e ST, o grupo de cantores sertanejo.



**Figura 4.38:** Histograma dos coeficientes de correlação entre as séries de  $F_0$  e envoltória para o grupo (a) de canto lírico e (b) de canto sertanejo. Histograma dos coeficientes de correlação entre os índices de modulação AM e FM para o grupo de (c) canto lírico e (d) de canto sertanejo. O símbolo acima de cada distribuição indica a média, cercada por um desvio padrão.

Grupo	Correlação entre curvas			Correlação entre índices de mod.		
	Média	Mediana	Desvio Padrão	Média	Mediana	Desvio Padrão
Canto Lírico ( $n = 11$ )	-0,02	-0,05	0,48	0,58	0,66	0,36
Sertanejo ( $n = 15$ )	0,13	0,22	0,25	0,76	0,83	0,21

**Tabela 4.8:** Estatísticas dos coeficientes de correlação: base de dados 4.

# Capítulo 5

## Conclusão

### 5.1 Sumário

Este trabalho pode ser, simplificadaamente, dividido em três partes: (i) estudo de técnicas para análise de demodulação AM/FM, (ii) avaliação, implementação e calibração dos algoritmos de análise e (iii) aplicação dos algoritmos em diversas bases de dados relacionadas ao tremor e vibrato.

Na primeira parte (capítulo 2), o sinal analítico, o operador de Teager e a análise de Fourier foram investigados. O estudo mostrou que os resultados gerados pelo sinal analítico e pelo operador de Teager na análise de oscilações em um dado sinal só têm significado físico se este sinal for monocomponente. Como este não é o caso da voz, nenhum destes dois métodos foi adotado.

A análise de Fourier foi a alternativa escolhida. Na segunda etapa do trabalho (capítulo 3), ela foi implementada e calibrada. Ainda nesta etapa implementou-se um algoritmo de análise de regularidade baseado na função de autocorrelação.

A terceira etapa (capítulo 4) consistiu da aplicação dos algoritmos implementados em sinais de tremor em membros, vibrato no contrabaixo, voz falada e cantada. Em relação às medidas de taxa e profundidade, os valores médios encontrados para o grupo formado por indivíduos com tremor em membros foram, respectivamente, 4,7 Hz e 19,83 Hz. Na base de dados do contrabaixo, a média foi de 5,0 Hz para taxa e de 0,96% para a profundidade. No caso da voz falada, a taxa média foi de 4,8 Hz e a profundidade média, de 3,98% entre os que foram avaliados, perceptivamente, com tremor e de 5,6 Hz/0,55% entre aqueles não avaliados e, para o canto, 5,7 Hz foi a taxa média e 3,12%, a profundidade média.

Quanto à regularidade, diversos exemplos mostraram que a técnica proposta é capaz de produzir indicativos da taxa instantânea e da semelhança entre ciclos vizinhos da curva de  $F_0$ . Porém, este método ainda precisa ser mais estudado e validado em trabalhos futuros.

## 5.2 Considerações finais

Durante a etapa de análise de resultados, alguns comportamentos existentes nos sinais chamaram a atenção do autor, especialmente aqueles das seções 4.3.1 e 4.5.1.

No primeiro caso, um instrumentista executou dois tipos diferentes de vibrato sobre a mesma nota em um contrabaixo acústico e, interessantemente, a diferença mais clara entre os dois estava no espectro da envoltória. Como mostrado na seção 4.3.1, a diferença entre os picos dos espectros de envoltória era de praticamente 100%, enquanto os picos dos espectros de  $F_0$  variavam menos de 25%. Ao escutar as duas amostras de vibrato, o autor e também alguns outros indivíduos perceberam diferenças que pareciam estar relacionadas à profundidade do sinal.

No outro exemplo, um cantor lírico produziu um vibrato com taxa de modulação da envoltória duas vezes maior que a taxa da curva de  $F_0$ . Horii e Hata (1988) já tinha mostrado que isto era perfeitamente natural. Contudo, nenhum autor da literatura relacionada investigou um caso como este (talvez, porque diversos deles estudam apenas a curva da frequência fundamental). Em alguns testes informais realizados, o fato de a taxa da envoltória ser uma ou duas vezes o valor da taxa da frequência fundamental não alterou a percepção da taxa do vibrato.

Algumas questões surgiram a partir deste ponto: qual a importância da modulação presente na envoltória de sinais complexos, como o de voz e o de música, sobre a percepção? Qual a influência de uma série de  $F_0$  composta por diversos harmônicos em relação a outra monocomponente sobre a percepção? Será que a regularidade de  $F_0$  é o aspecto determinante na qualidade do vibrato, como apontado por vários autores, ou será que a combinação entre determinados níveis de modulação da envoltória e de modulação de  $F_0$  é a chave para que o vibrato seja percebido musicalmente como “bom”?

Estas indagações foram deixadas para serem analisadas em trabalhos futuros. A contribuição deste trabalho foi criar uma ferramenta de medida para ser utilizada em diversas áreas de pesquisa, e.g., engenharia, música, fonoaudiologia, neurociência, a fim de que questões abrangendo mais de uma área de conhecimento pudessem ser tratadas adequadamente.

Os resultados apresentados mostram as limitações (e.g., resolução em frequência dependente da duração da amostra) e as potencialidades (e.g., capacidade da análise de regularidade de rastrear detalhes) desta ferramenta. Contudo, é importante enfatizar que os resultados obtidos são preliminares, suficientes apenas para validar a ferramenta desenvolvida, não tendo sido ainda realizada uma análise profunda dos fenômenos usados como exemplos.

## 5.3 Trabalhos Futuros

Alguns assuntos não foram abordados neste trabalho e são deixados como sugestões para pesquisas futuras:

1. Explorar técnicas de análise tempo-frequência para analisar os sinais de tremor/vibrato, sabendo de suas dificuldades de interpretação física (Mellody e Wakefield, 2000).

2. Investigar mais detalhadamente o método de análise de regularidade e analisar o seu comportamento quando outros estimadores para a função de autocorrelação são usados.
3. Explorar o método de análise de regularidade proposto no diagnóstico e no acompanhamento do tratamento de perturbações em voz.
4. Estudar a relação entre as características medidas (taxa, profundidade, regularidade) e as percebidas nos sinais de tremor/vibrato.
5. Desenvolver um método para estudar o início (*onset*) e o fim (*offset*) do sinal de tremor/vibrato. Geralmente, estes trechos apresentam instabilidades e são descartados no método proposto.

# Referências Bibliográficas

- Arroabarren, I.; Zivanovic, M.; Bretos, J.; Ezcurra, A. e Carlosena, A. (2002). Measurement of vibrato in lyric singers. *IEEE Transactions on Instrumentation and Measurement*, 51(4):660–665.
- Bagshaw, P. (1993). An investigation of acoustic events related to sentential stress and pitch accents, in English. *Speech Communication*, 13(3-4):333–342.
- Bagshaw, P. (1994). *Automatic prosody analysis*. PhD thesis, University of Edinburgh, Scotland, UK.
- Bagshaw, P.; Hiller, S. M. e Jack, M. A. (1993). Enhanced pitch tracking and the processing of f0 contours for computer aided intonation teaching. In *in Proceedings of the 3rd European Conference on Speech Communication and Technology*, pp. 1003–1006.
- Boashash, B.; Lovell, B. C. e Kootsookos, P. J. (1989). Time-Frequency Signal Analysis and Instantaneous Frequency Estimation: Methodology, Relationships and Implementations. In *International Symposium on Circuits and Systems*, pp. 1237–1242.
- Borem, F. (2010). Comunicação pessoal.
- Borém, F.; Lage, G.; Vieira, M. N. e Barreiros, J. P. (2005). As informações sensoriais (audição, tato e visão) no controle da afinação não-temperada do contrabaixo acústico. In *15 Congresso da ANNPOM (Associação Nacional de Pós-Graduação em Música)*, Rio de Janeiro, volume 1, pp. 291 – 294. ISCA.
- Bretos, J. e Sundberg, J. (2003). Measurements of vibrato parameters in long sustained crescendo notes as sung by ten sopranos. *Journal of Voice*, 17(3):343–352.
- Clarkson, J. e Deutsch, J. (1966). Effect of threshold reduction on the vibrato. *Journal of Experimental Psychology*, 71(5):706–710.
- Cohen, L. (1992). What is a multicomponent signal? In *IEEE International Conference on Acoustics, Speech, and Signal Processing, 1992. ICASSP-92.*, volume 5.
- CSTR (2010). Festival speech synthesis system – the centre for speech technology research, university of edinburgh. <http://www.cstr.ed.ac.uk/projects/festival/download.html> [último acesso em 22 de junho de 2010].

- de Almeida Bezerra, A.; Cukier-Blaj, S.; Duprat, A.; Camargo, Z. e Granato, L. (2009). The Characterization of the Vibrato in Lyric and Sertanejo Singing Styles: Acoustic and Perceptual Auditory Aspects. *Journal of Voice*, 23(6):666–670.
- Diaz, J. e Rothman, H. (2003). Acoustical comparison between samples of good and poor vibrato in singers. *Journal of Voice*, 17(2):179.
- Dromey, C. e Smith, M. (2008). Vocal tremor and vibrato in the same person: acoustic and electromyographic differences. *Journal of Voice*, 22(5):541–545.
- Farinella, K.; Hixon, T.; Hoit, J.; Story, B. e Jones, P. (2006). Listener Perception of Respiratory-Induced Voice Tremor. *American Journal of Speech-Language Pathology*, 15(1):72.
- Guyton, A. e Hall, J. (1981). Textbook of medical physiology. *Philadelphia, Pa*, pp. 650–651.
- Hanson, H.; Maragos, P. e Potamianos, A. (1993). Finding speech formants and modulations via energy separation: with application to a vocoder. In *1993 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1993. ICASSP-93.*, volume 2.
- Hibi, S. e Hirano, M. (1995). Voice quality variations associated with vibrato. In Fujimura, O. e Hirano, M., editores, Vocal Fold Physiology – Voice Quality Control, chapter 11, pp. 189–202. Singular Publishing Group, Inc.
- Horii, Y. (1989a). Acoustic analysis of vocal vibrato: A theoretical interpretation of data. *Journal of Voice*, 3(1):36–43.
- Horii, Y. (1989b). Frequency Modulation Characteristics of Sustained /a/ Sung in Vocal Vibrato. *Journal of Speech, Language and Hearing Research*, 32(4):829.
- Horii, Y. e Hata, K. (1988). A note on phase relationships between frequency and amplitude modulations in vocal vibrato. *Folia phoniatrica*, 40(6):303.
- Howes, P.; Callaghan, J.; Davis, P.; Kenny, D. e Thorpe, W. (2004). The relationship between measured vibrato characteristics and perception in Western operatic singing. *Journal of Voice*, 18(2):216–230.
- Jens Timmer (2010). Data analysis and modeling of dynamic processes in the life sciences. [http://webber.physik.uni-freiburg.de/~jeti/path\\_tremor/](http://webber.physik.uni-freiburg.de/~jeti/path_tremor/) [último acesso em 16 de julho de 2010].
- Kaiser, J. F. (1990a). On a simple algorithm to calculate the ‘energy’ of a signal. In *International Conference on Acoustics, Speech, and Signal Processing, ICASSP-90*, pp. 381–384.
- Kaiser, J. F. (1990b). On Teager’s energy algorithm and its generalization to continuous signals. In *Proc. 4th IEEE Digital Signal Processing Workshop*.

- Kinsler, L. E.; Frey, A. R.; Coppens, A. B. e Sanders, J. V. (1999). *Fundamentals of Acoustics*. John Wiley Sons, Inc., 4 edio.
- Kohler, H. (1988). *Essentials of statistics*. Scott Foresman/Addison-Wesley.
- Kreiman, J.; Gabelman, B. e Gerratt, B. (2003). Perception of vocal tremor. *Journal of Speech, Language and Hearing Research*, 46(1):203.
- Lage, G.; Borém, F.; Benda, R. e Moraes, L. (2002). Aprendizagem motora na performance musical: reflexões sobre conceitos e aplicabilidade. *Revista Per Musi*, 5(6):14–37.
- Large, J. (1979). An air flow study of vocal vibrato. In *Transcripts of the eighth symposium on care of the professional voice*. New York: The Voice Foundation part I, pp. 39–45.
- Leon, C. (1995). *Time-frequency analysis: theory and applications*. New York: Prentice Hall.
- Leydon, C.; Bauer, J. e Larson, C. (2003). The role of auditory feedback in sustaining vocal vibrato. *Journal of the Acoustical Society of America*, 114:1575–1581.
- Maragos, P.; Kaiser, J. F. e Quatieri, T. F. (1992). On separating amplitude from frequency modulations using energy operators. In *Proc IEEE Int. Conf. ASSP, San Francisco, CA*, pp. 1–4.
- Maragos, P.; Kaiser, J. F. e Quatieri, T. F. (1993a). Energy separation in signal modulations with application to speech analysis. *IEEE Transactions on signal processing*, 41(10):3024–3051.
- Maragos, P.; Kaiser, J. F. e Quatieri, T. F. (1993b). On amplitude and frequency demodulation using energy operators. *IEEE Transactions on signal processing*, 41(4):1532–1550.
- Maragos, P.; Quatieri, T. F. e Kaiser, J. F. (1991). Speech nonlinearities, modulations, and energy operators. In *Acoustics, Speech, and Signal Processing, 1991. ICASSP-91., 1991 International Conference on*, pp. 421–424.
- Medan, Y.; Yair, E. e Chazan, D. (1991). Super resolution pitch determination of speech signals. *IEEE Transactions on Signal Processing*, 39(1):40–48.
- Mellody, M. e Wakefield, G. (2000). The time-frequency characteristics of violin vibrato: Modal distribution analysis and synthesis. *Journal of the Acoustical Society of America*, 107:598.
- Mitra, S. (1998). *Digital signal processing: a computer-based approach*. McGraw-Hill New York.
- Noll, A. M. (1967). Cepstrum pitch determination. *Journal of the Acoustical Society of America*, 41(2):293–309.

- Nuttall, A. (1991). Complex Envelope Properties, Interpretation, Filtering, and Evaluation. *NUSC Technical Report 8827*.
- Oppenheim, A. e Schafer, R. (1975). *Digital Signal Processing*. Englewood Clis. NJ, Prentice Hall.
- Oppenheim, A.; Schafer, R. e Buck, J. (1998). *Discrete-time signal processing*. Prentice Hall Englewood Cliffs, NJ, 2 edio.
- Papoulis, A. (1988). *Signal Analysis*. McGraw-Hill International Editions, international edio.
- Peterson, G. e Barney, H. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24(2):175–184.
- Potamianos, A. e Maragos, P. (1994). A comparison of the energy operator and the Hilbert transform approach to signal and speech demodulation. *Signal Processing*, 37(1):95–120.
- Potamianos, A. e Maragos, P. (1996). Speech formant frequency and bandwidth tracking using multiband energy demodulation. *Journal of the Acoustical Society of America*, 99:3795–3806.
- Prame, E. (1994). Measurements of the vibrato rate of ten singers. *Journal of the Acoustical Society of America*, 96(4):1979–1984.
- Prame, E. (1997). Vibrato extent and intonation in professional Western lyric singing. *Journal of the Acoustical Society of America*, 102:616.
- Qi, Y. e Weinberg, B. (1994). A homomorphic method for measuring secondary amplitude modulation. *Journal of speech and hearing research*, 37(4):800.
- Rabiner, L.; Sambur, M. e Schmidt, C. (1975). Applications of a nonlinear smoothing algorithm to speech processing. *IEEE Transactions on Acoustics Speech and Signal Processing*, 23(6):552–557.
- Rabiner, L. e Schafer, R. (1978). *Digital processing of speech signals*. Prentice-Hall, NJ.
- Ramig, L. e Shipp, T. (1987). Comparative measures of vocal tremor and vocal vibrato. *Journal of Voice*, 1(2):162–167.
- Reilly, A.; Frazer, G. e Boashash, B. (1994). Analytic signal generation-tips and traps. *IEEE Transactions on Signal Processing*, 42(11):3241–3245.
- Rosa, L. L. C. (2003). Vibrato sertanejo: análise acústica e correlatos fisiológicos no trato vocal. Master's thesis, Faculdade de Medicina-USP.
- Ross, M.; Shaffer, H.; Cohen, A.; Freudberg, R. e Manley, H. (1974). Average magnitude difference function pitch extractor. *IEEE transactions on acoustics, speech and signal processing*, 22(5):353–362.

- Rothman, H. (2010). Website. <http://www.clas.ufl.edu/users/hrothman/pub/> [último acesso em 20 de julho de 2010].
- Sansão, J. P. H. (2009). Medida da relação harmônico/ruído em vozes disfônicas pelo processamento digital de imagens espectrográficas. Master's thesis, PPGEE-UFMG.
- Schultz-Coulon, H. (1978). The neuromuscular phonatory control system and vocal function. *Acta Oto-Laryngologica*, 86(1-6):142–153.
- Seashore, C. (1932). *The vibrato*. Iowa State Univ. Press.
- Shipp, T.; Sundberg, J. e Haglund, S. (1984). A model of frequency vibrato. In *Transcripts of the Thirteenth Symposium: Care of the Professional Voice*.
- Sundberg, J. (1995). Acoustic and psychoacoustic aspects of vocal vibrato. In Dejonckere, P.; Hirano, M. e Sundberg, J., editores, *Vibrato*, chapter 2, pp. 35–62. Singular Publishing Group, Inc.
- Timmer, J.; Haussler, S.; Lauk, M. e Lucking, C. (2000). Pathological tremors: Deterministic chaos or nonlinear stochastic oscillators? *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 10:278.
- Titze, I. (1995). Workshop on acoustic voice analysis. Summary statement. *National Center for Voice and Speech, Denver, CO*, p. 36.
- Titze, I. (1996). Coupling of neural and mechanical oscillators in control of pitch, vibrato, and tremor. *Vocal fold physiology: Controlling complexity and chaos*, pp. 47–62.
- Titze, I.; Solomon, N.; Luschi, E. e Hirano, M. (1994). Interference between normal vibrato and artificial stimulation of laryngeal muscles at near-vibrato rates. *Journal of Voice*, 8(3):215–223.
- Titze, I.; Story, B.; Smith, M. e Long, R. (2002). A reflex resonance model of vocal vibrato. *Journal of the Acoustical Society of America*, 111:2272.
- Ubuntu (2010). Ubuntu 10.04. <http://www.ubuntu.com/desktop/get-ubuntu/download> [último acesso em 22 de junho de 2010].
- Vakman, D. (1996). On the analytic signal, the Teager-Kaiser energy algorithm, and other methods for defining amplitude and frequency. *IEEE Transactions on signal processing*, 44(4):791–797.
- Vieira, M. (1997). *Automated Measures of Dysphonias and the Phonatory Effects of Asymmetries in the Posterior Larynx*. PhD thesis, University of Edinburgh.
- Vieira, M. N. (1998). Algoritmo para extração do espectro do tremor e do vibrato na voz humana. Technical Report FAPEMIG – TEC 1143/98, Departamento de Física/ICEx/UFMG.

- 
- Winholtz, W. e Ramig, L. (1992). Vocal tremor analysis with the vocal demodulator. *Journal of Speech, Language and Hearing Research*, 35(3):562–573.
- Yair, E. e Gath, I. (1988). On the use of pitch power spectrum in the evaluation of vocal tremor. *Proceedings of the IEEE*, 76(9):1166–1175.