

MINISTÉRIO DA DEFESA  
EXÉRCITO BRASILEIRO  
DEPARTAMENTO DE CIÊNCIA E TECNOLOGIA  
INSTITUTO MILITAR DE ENGENHARIA  
CURSO DE MESTRADO EM ENGENHARIA ELÉTRICA

DANIEL PATRICIO NICOLALDE RODRÍGUEZ

AUTENTICAÇÃO DE ÁUDIO DIGITAL BASEADA NA PRESENÇA DA  
FREQUÊNCIA DA REDE ELÉTRICA

Rio de Janeiro  
2010

**INSTITUTO MILITAR DE ENGENHARIA**

**DANIEL PATRICIO NICOLALDE RODRÍGUEZ**

**AUTENTICAÇÃO DE ÁUDIO DIGITAL BASEADA NA PRESENÇA DA  
FREQUÊNCIA DA REDE ELÉTRICA**

Dissertação de Mestrado apresentada ao Curso de Mestrado em Engenharia Elétrica do Instituto Militar de Engenharia, como requisito parcial para obtenção do título de Mestre em Ciências em Engenharia Elétrica.

Orientador: Prof. José Antonio Apolinário Jr. - D. Sc.,  
Co-orientador: Prof. Luiz Wagner P. Biscainho - D. Sc.,

Rio de Janeiro  
2010

c2010

INSTITUTO MILITAR DE ENGENHARIA  
Praça General Tibúrcio, 80-Praia Vermelha  
Rio de Janeiro-RJ CEP 22290-270

Este exemplar é de propriedade do Instituto Militar de Engenharia, que poderá incluí-lo em base de dados, armazenar em computador, microfilmear ou adotar qualquer forma de arquivamento.

É permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do(s) autor(es) e do(s) orientador(es).

621.38154 Nicolalde R., D. P.  
R696a

Autenticação de Áudio Digital baseada na Presença da Frequência da Rede Elétrica/ Daniel Patricio Nicolalde Rodríguez.

– Rio de Janeiro: Instituto Militar de Engenharia, 2010.  
77 p.:il.

Dissertação: (mestrado) – Instituto Militar de Engenharia, Rio de Janeiro, 2010.

1. Áudio Digital. 2. Autenticação de áudio. I. Título.  
II. Instituto Militar de Engenharia.

CDD 621.38154

INSTITUTO MILITAR DE ENGENHARIA

DANIEL PATRICIO NICOLALDE RODRÍGUEZ

AUTENTICAÇÃO DE ÁUDIO DIGITAL BASEADA NA PRESENÇA DA  
FREQUÊNCIA DA REDE ELÉTRICA

Dissertação de Mestrado apresentada ao Curso de Mestrado em Engenharia Elétrica do Instituto Militar de Engenharia, como requisito parcial para obtenção do título de Mestre em Ciências em Engenharia Elétrica.

Orientador: Prof. José Antonio Apolinário Jr. - D. Sc.,

Co-orientador: Prof. Luiz Wagner P. Biscainho - D. Sc.,

Aprovada em 26 de Fevereiro de 2010 pela seguinte Banca Examinadora:

---

Prof. José Antonio Apolinário Jr. - D. Sc., do IME - Presidente

---

Prof. Luiz Wagner P. Biscainho - D. Sc., da COPPE/UFRJ

---

Prof. Ernesto Leite Pinto - D. C., do IME

---

Marco Antonio Macciola Rodrigues - D. Sc., do CEPTEL

Rio de Janeiro  
2010

À minha família querida, que fiquei sem ver mais de dois anos por me encontrar fora do Equador.

## AGRADECIMENTOS

A Deus, ser supremo e conservador do mundo, por me ajudar a conservar uma vida digna e produtiva.

Aos meus pais e irmãos, que sempre me ajudaram e apoiaram na realização dos meus objetivos.

Ao Prof. José Antonio Apolinário Jr. e ao Prof. Luiz Wagner Pereira Biscainho, pela dedicação outorgada durante a realização deste trabalho.

A todos os professores, funcionários e alunos do IME, especialmente da SE-3.

A este lindo país, Brasil, e em especial ao povo carioca por ter feito a minha vida prazenteira durante a minha estadia aqui. Atualmente já me sinto brasileiro, ainda que não seja de raízes, ao menos sim de coração.

A todos meus familiares e amigos que ao longo da minha vida me ajudaram na realização dos meus objetivos.

“O futuro pertence àqueles que acreditam na beleza dos seus sonhos.”  
Anna Eleanor Roosevelt.

## SUMÁRIO

LISTA DE ILUSTRAÇÕES .....	8
LISTA DE TABELAS .....	9
LISTA DE ABREVIATURAS E SÍMBOLOS .....	10
<b>1 INTRODUÇÃO .....</b>	<b>13</b>
1.1 Objetivo da Dissertação .....	14
1.2 Contribuições da Dissertação .....	14
1.3 Organização da Dissertação .....	15
<b>2 AUTENTICAÇÃO DE ÁUDIO COMO FERRAMENTA DE AUXÍLIO À FONÉTICA FORENSE .....</b>	<b>16</b>
2.1 Introdução .....	16
2.2 Atuação da Fonética Forense .....	16
2.3 Exame de Autenticação .....	16
2.3.1 Autenticação de Áudio na Forma Analógica .....	17
2.3.2 Autenticação de Áudio na Forma Digital .....	18
2.4 ENF na Autenticação de Áudio Digital .....	18
<b>3 ESTIMANDO A FREQUÊNCIA E A FASE .....</b>	<b>20</b>
3.1 Introdução .....	20
3.2 Método da DFT .....	20
3.3 Método da DFT <sup>1</sup> .....	21
3.4 Experimentos Sintéticos .....	24
3.5 Considerações Finais .....	29
<b>4 AUTENTICAÇÃO DE ÁUDIO DIGITAL COM BASE NA MU- DANÇA DE FASE DO SINAL DA REDE ELÉTRICA .....</b>	<b>31</b>
4.1 Introdução .....	31
4.2 Experimentos Preliminares .....	31
4.2.1 Informação Visual da Medida da Fase .....	33
4.3 Método Proposto .....	35

4.3.1	Detecção Automática da ENF .....	35
4.3.2	Mecanismo Visual .....	38
4.3.3	Mecanismo Automático .....	39
4.4	Resumo .....	41
<b>5</b>	<b>AVALIAÇÃO DA TÉCNICA PROPOSTA .....</b>	<b>42</b>
5.1	Introdução .....	42
5.2	Criação das Bases de Dados .....	42
5.2.1	BASE ESPANHOLA .....	42
5.2.2	BASE CARIOCA 1 .....	44
5.2.3	BASE CARIOCA 2 .....	44
5.3	Resultados Experimentais .....	45
5.3.1	Obtenção de limiares para a detecção automática da ENF .....	45
5.3.2	Exemplos do Mecanismo Visual .....	45
5.3.3	Avaliação do Mecanismo Automático .....	51
5.4	Efeitos de Degradações na Detecção .....	55
5.4.1	Efeito de Ruído de Fundo .....	57
5.4.2	Efeito de Saturações .....	60
5.4.3	Efeito da Variação da ENF .....	61
5.5	Resumo .....	64
<b>6</b>	<b>CONSIDERAÇÕES FINAIS .....</b>	<b>66</b>
6.1	Conclusões .....	67
6.2	Propostas para trabalhos futuros .....	68
<b>7</b>	<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>70</b>

## LISTA DE ILUSTRAÇÕES

FIG.3.1	Interpolação linear no argumento de $X'(k)$ para a obtenção do valor aproximado de $\theta$ .....	25
FIG.3.2	Espectro de uma de tom de 60,98 Hz janelado. ....	26
FIG.3.3	Estimação de fase de um tom artificial de 60,98 Hz. (a) Sinal; (b) Fase estimada. ....	28
FIG.4.1	Tom de 60 Hz editado com pequenas descontinuidades de fase. (a) Sinal editado e as suas mudanças de fase; (b) Sinal filtrado. ....	32
FIG.4.2	Tom de 60 Hz editado com consideráveis descontinuidades de fase. (a) Sinal editado e as suas mudanças de fase; (b) Sinal filtrado. ....	33
FIG.4.3	Estimação de fase para os casos: tom com mudanças leves de fase e tom com mudanças consideráveis de fase. ....	34
FIG.4.4	Curvas de estimação de fase para diferentes valores de mudanças de fase num tom de 60 Hz editado. ....	36
FIG.4.5	Esquema de segmentação em blocos do sinal de áudio. ....	39
FIG.5.1	Histograma normalizado da distribuição das mudanças de fase nos sinais editados. (a) BASE ESPANHOLA; (b) BASE CARIOCA 1; (c) BASE CARIOCA 2. ....	43
FIG.5.2	Histograma normalizado da relação $R_{ENF}$ . (a) Para 50,0 Hz; (b) Para 60,0 Hz. ....	46
FIG.5.3	Visualização de uma eliminação de fragmento (Sinal da BASE ESPANHOLA). (a) Sinal Original; (b) Sinal Editado; (c) Estimação de fase usando DFT; (d) Estimação de fase usando $DFT^1$ . ....	47
FIG.5.4	Visualização de uma inserção de fragmento (Sinal da BASE ESPANHOLA). (a) Sinal Original; (b) Sinal Editado; (c) Estimação de fase usando DFT; (d) Estimação de fase usando $DFT^1$ . ....	48
FIG.5.5	Visualização de uma eliminação de fragmento (Sinal da BASE CARIOCA 1). (a) Sinal Original; (b) Sinal Editado; (c) Estimação de fase usando DFT; (d) Estimação de fase usando $DFT^1$ . ....	49
FIG.5.6	Visualização de uma inserção de fragmento (Sinal da BASE CARIOCA 2). (a) Sinal Original; (b) Sinal Editado; (c) Estimação	

	de fase usando DFT; (d) Estimação de fase usando DFT <sup>1</sup> . . . . .	50
FIG.5.7	Histograma normalizado da distribuição da medida característica, $F$ . (a) BASE ESPANHOLA; (b) BASE CARIOCA 1; (c) BASE CARIOCA 2. . . . .	55
FIG.5.8	Curva DET: $P_M$ em função de $P_F$ . (a) BASE ESPANHOLA; (b) BASE CARIOCA 1; (c) BASE CARIOCA 2. . . . .	56
FIG.5.9	Exemplo de utilização do algoritmo de VAD. . . . .	57
FIG.5.10	Magnitude da Resposta em frequência da função de transferência $H(z)$ para os casos de RUÍDO 2 e RUÍDO 3. . . . .	59
FIG.5.11	Efeito de ruído: Erro (EER) de detecção na autenticação de áudio em função da SNR. . . . .	60
FIG.5.12	Exemplo de um sinal com 3% das amostras (correspondentes à região ativa de voz) saturadas. . . . .	61
FIG.5.13	Efeito de saturação: Erro (EER) de detecção na autenticação de áudio em função de percentagem de saturação. . . . .	62
FIG.5.14	Histograma normalizado da distribuição das estimativas da ENF (sinais não editados). (a) BASE ESPANHOLA; (b) BASE CARIOCA 1; (c) BASE CARIOCA 2. . . . .	63
FIG.5.15	Histograma normalizado da distribuição do desvio-padrão das estimativas da ENF em cada sinal do banco de dados (sinais não editados). (a) BASE ESPANHOLA; (b) BASE CARIOCA 1; (c) BASE CARIOCA 2. . . . .	64

## LISTA DE TABELAS

TAB.3.1	Avaliação estatística das estimativas de fase e frequência usando os métodos DFT e DFT <sup>1</sup> . . . . .	29
TAB.5.1	Avaliação do mecanismo automático de autenticação de áudio para BASE ESPANHOLA. . . . .	52
TAB.5.2	Avaliação do mecanismo automático de autenticação de áudio para BASE CARIOCA 1. . . . .	53
TAB.5.3	Avaliação do mecanismo automático de autenticação de áudio para BASE CARIOCA 2. . . . .	54

## LISTA DE ABREVIATURAS E SÍMBOLOS

### ABREVIATURAS

AR	-	<i>Autoregressive model</i>
ARMA	-	<i>Autoregressive moving average model</i>
DET	-	<i>Detection error tradeoff</i>
DFT	-	<i>Discrete Fourier transform</i>
EER	-	<i>Equal error rate</i>
ENF	-	<i>Electric network frequency</i>
FIR	-	<i>Finite impulse response</i>
HMM	-	<i>Hidden Markov model</i>
NB	-	<i>Narrow band</i>
PDA	-	<i>Personal digital assistants</i>
SNR	-	<i>Signal-to-noise ratio</i>
VAD	-	<i>Voice active detection</i>
WB	-	<i>Wide band</i>

### SÍMBOLOS

$e_f$	-	erro das estimativas de frequência num tom.
$e_\phi$	-	erro das estimativas de fase num tom.
$\bar{e}_f$	-	valor médio do erros das estimativas de freq. para varios tons.
$\bar{e}_\phi$	-	valor médio dos erros das estimativas de fase para varios tons.
$F$	-	medida característica para detectar a edição de áudio.
$f_{\text{downsample}}$	-	frequência de sub-amostragem.
$f_{\text{DFT}}$	-	frequência estimada quando usado o método da DFT.
$f_{\text{DFT}^1}$	-	frequência estimada quando usado o método da DFT <sup>1</sup> .
$f_s$	-	frequência de amostragem.
$f_{\text{tone}}$	-	frequência real do tom.
$\hat{f}(n_b)$	-	frequência estimada no $n_b$ -ésimo bloco.
$H(z)$	-	função de transferência do filtro.
$H_{\text{ENF}}$	-	hipótese de presença da ENF no sinal de áudio.
$H_{\sim\text{ENF}}$	-	hipótese de ausência da ENF no sinal de áudio.

$H_E$	- hipótese de que o sinal de áudio digital seja editado.
$H_O$	- hipótese de que o sinal de áudio digital seja original.
$\hat{H}$	- decisão final (sinal original ou editado).
$k$	- índice associado com os valores discretos da DFT.
$k_{\text{DFT}^1}$	- valor correspondente a $f_{\text{DFT}^1}$ no domínio de $k$ .
$k_{\text{high}}$	- inteiro mais próximo maior ou igual a $k_{\text{DFT}^1}$ .
$k_{\text{low}}$	- inteiro mais próximo menor ou igual a $k_{\text{DFT}^1}$ .
$k_{\text{peak}}$	- índice inteiro associado com o valor máximo de $ X(k) $ .
$M$	- tamanho da janela (número de amostras).
$m_{\hat{\phi}}$	- valor médio de $\hat{\phi}'(n_b)$ .
$n$	- índice associado com os valores discretos das amostras do sinal.
$n_1$	- duração de bloco (ciclos da ENF nominal).
$n_2$	- superposição de bloco (ciclos da ENF nominal) com o seguinte.
$n_{\text{add}}$	- ruído com valor médio zero.
$n_{\text{background}}$	- ruído de fundo.
$n_b$	- índice associado com os blocos do sinal.
$N_{\text{Block}}$	- número total de blocos.
$N_{\text{DFT}}$	- número de pontos da DFT.
$P_{\text{background}}$	- potência média do ruído de fundo.
$P_D$	- probabilidade de detecção dos sinais editados.
$P_F$	- probabilidade de falso alarme na procura de sinais editados.
$P_M$	- probabilidade de perda para sinais editados.
$P_{\text{voice}}$	- potência média total nas regiões de voz ativa.
$R_{\text{ENF}}$	- relação usada para determinar a presença da ENF.
$s$	- sinal de voz puro.
$s_{\text{NB}}$	- sinal filtrado com banda estreita.
$s_{\text{tone}}(n)$	- tom simples (uma só componente de frequência).
$s_{\text{WB}}$	- sinal filtrado com banda com banda mais larga.
$s'_{\text{tone}}(n)$	- derivada aproximada de $s_{\text{tone}}(n)$ .
$\text{SNR}_{\text{original}}$	- razão sinal-ruído original.
$w(n)$	- janela usada.
$x(n)$	- versão janelada de $s_{\text{tone}}(n)$ .
$X(k)$	- DFT de $x(n)$ .

$\mathbf{x}'(n)$	-	versão janelada de $s'_{\text{tone}}(n)$ .
$\mathbf{X}'(\mathbf{k})$	-	DFT de $x'(n)$ .
$\gamma$	-	limiar para determinar se um sinal é original ou editado.
$\gamma_{\text{ENF}}$	-	limiar usada para determinar a presença da ENF.
$\theta$	-	fase inicial de $s'_{\text{tone}}$ .
$\phi_0$	-	fase inicial de $s_{\text{tone}}$ .
$\phi_{\text{DFT}}$	-	fase estimada quando usado o método da DFT.
$\phi_{\text{DFT}^1}$	-	fase estimada quando usado o método da DFT <sup>1</sup> .
$\phi(\mathbf{n}_b)$	-	fase original no $n_b$ -ésimo bloco.
$\hat{\phi}(\mathbf{n}_b)$	-	fase estimada no $n_b$ -ésimo bloco.
$\hat{\phi}'(\mathbf{n}_b)$	-	variação de $\hat{\phi}(\mathbf{n}_b)$ do bloco anterior para o bloco atual.

## RESUMO

O objetivo principal da análise de autenticação de áudio é avaliar a integridade de gravações (analógicas ou digitais) para determinar se alguma parte do sinal foi eliminada, adicionada ou alterada de algum modo. Este trabalho propõe um método para verificar se uma gravação de áudio digital foi adulterada ou não. O método se baseia nas mudanças detectadas na fase do sinal da rede elétrica (ENF), usualmente embutido nos sinais de áudio quando a gravação é realizada com um equipamento conectado a uma tomada.

O método é dividido em três partes. A primeira delas consiste num mecanismo automático para detectar a presença da ENF nas gravações. A segunda parte é uma técnica que pode auxiliar na localização visual dos pontos de edição, bem como na inferência dos tipos de edição: inserção ou eliminação de fragmento no/do sinal de áudio. Para localizar os pontos de edição, é necessário identificar as mudanças abruptas na fase da ENF presente no sinal gravado. Para estimação da fase, foram usadas duas técnicas: uma baseada na transformada discreta de Fourier (DFT) de tempo curto, outra numa análise de Fourier de alta precisão que também utiliza a DFT da primeira derivada do sinal. A última parte do método proposto apresenta um mecanismo automático para discriminar entre sinais originais e editados com base numa relação de decisão aplicada sobre uma medida característica que computa a descontinuidade da fase da ENF.

Apresentam-se os fundamentos teóricos e questões práticas de implementação do método proposto. Finalmente, seu desempenho é avaliado em três *corpora* de sinais de áudio reais que foram digitalmente editados. Dois deles apresentam sinais gravados em condições favoráveis e o outro, sinais com ocorrência de saturações e ruído de fundo pronunciado.

## ABSTRACT

The main goal of audio authenticity analysis is to assess the integrity of (analogue or digital) recordings in order to determine if any part of it has been deleted, added or altered in any way. This work proposes a method that verifies whether or not a digitally recorded audio signal has been adulterated. The method is based on detecting subtle phase changes of the electric network frequency (ENF) signal, usually embedded in audio signals whenever the recording is carried out with the equipment connected to an outlet.

The method is divided in three parts. The first one is a mechanism to detect automatically the presence of ENF in the recorded material. The second part is a technique which might help visually locating edition points as well as inferring the type of edition: insertion or removal of audio pieces. To locate the edition points, it is necessary to identify abrupt phase changes in the ENF present in the recorded signal. Two techniques were used to carry out the phase estimation process, one based on the short-time discrete Fourier transform (DFT), the other one based on a high-precision Fourier analysis that also uses the DFT of the first derivative of the signal. The last part of the proposed method describes a mechanism to discriminate automatically between original and edited signals, based on a decision relation applied to a feature measure which computes the discontinuity of the ENF phase.

Theoretical background is presented along with practical implementation issues related to the proposed method. Finally, its performance is evaluated over three *corpora* composed of real audio signals which have been digitally edited. Two of them contain signals recorded under favorable conditions, and the last one, signals exhibiting saturations and/or high-level of background noise.

# 1 INTRODUÇÃO

No mundo atual, grande parte da informação já é criada, capturada, transmitida, armazenada e processada em forma digital, vindo a ser manipulada pelas pessoas em sua vida diária (para fins de armazenamento, comunicações, criptografia de dados, etc.). É, portanto, importante para a preservação da harmonia na sociedade assegurar que esta manipulação não se dê para fins ilegais.

A ciência forense, entre inúmeras outras tarefas, se encarrega de determinar a legitimidade da informação no contexto da Justiça, utilizando métodos científicos para responder as questões de interesse legal com fundamento nas investigações feitas. Uma análise forense basicamente identifica fontes de evidência que possam colaborar com o sistema legal, preserva e analisa as evidências, e apresenta o resultado da análise.

Quando se está lidando com evidência digital na análise forense, é interessante considerar os seguintes aspectos (DELP, 2009):

- Os dados digitais são representados mediante uma sequência de bits e não possuem claramente propriedades da sua autenticidade ou origem;
- Existem diversos dispositivos para criar e armazenar informação digital. Estes dispositivos podem ser câmaras, gravadores de áudio, PDA (*personal digital assistants*), telefones celulares e computadores. Verificar a veracidade da informação obtida destes dispositivos representa um grande desafio;
- Às vezes, a evidência digital passa por diferentes canais e pode se encontrar espalhada por múltiplos dispositivos e em diferentes formatos;
- Os dados digitais são manipuláveis. Existem ferramentas que podem ser usadas para ocultar ou eliminar parte da evidência digital.

Levando em conta as considerações acima citadas, um dos desafios dos pesquisadores é criar técnicas científicas que permitam verificar a autenticidade das evidências quando se trata de dados digitais. Desde este ponto de vista, o estudo forense digital é multidisciplinar; ele envolve, entre outras possíveis áreas, engenharia (processamento de sinais), ciência da computação e justiça criminal.

Considerando que as evidências sejam gravações digitais de áudio, os especialistas (peritos) realizam uma análise de áudio forense. (RUMSEY, 2008) apresenta algumas aplicações envolvidas com a análise de áudio forense. Estas ferramentas incluem: eliminação computacional de voz (necessária para decidir se uma voz é de uma determinada pessoa ou existem outras vozes), quantificação da voz do locutor (necessário para identificação automática de locutor mediante extração de características de vozes gravadas) e autenticação de áudio.

(MAHER, 2009) e (RUMSEY, 2008) apresentam novos trabalhos baseados na autenticação de áudio digital empregando técnicas baseadas no critério da presença do sinal da rede elétrica. O sinal da rede elétrica, conhecido como ENF (*Electric Network Frequency*), usualmente está presente nas gravações onde o equipamento encontra-se conectado a uma tomada. A idéia fundamental destes trabalhos é determinar o lugar e o instante em que foi realizada a gravação.

O foco de nosso trabalho, que também utiliza o critério da ENF para a autenticação de áudio digital, é determinar se um sinal foi ou não editado.

## 1.1 OBJETIVO DA DISSERTAÇÃO

O objetivo principal desta dissertação é propor uma ferramenta para autenticação de áudio digital com aplicação forense. Esta ferramenta utiliza o critério da presença da ENF nas gravações para determinar se um sinal foi ou não editado por meio de cortes ou inserções de fragmentos no sinal. Este trabalho é fundamentado em determinar se existem mudanças abruptas da fase da ENF como sintoma de que um sinal foi editado. São usadas técnicas de estimação de fase para determinar o comportamento de fase da ENF.

## 1.2 CONTRIBUIÇÕES DA DISSERTAÇÃO

Com este trabalho contribui-se com a aplicação dos fundamentos de processamento digital de sinais ao campo da Perícia Fonética, que, tal como qualquer outro ramo da perícia criminal, é uma base decisória que pode direcionar a investigação policial e o processo criminal. A ferramenta de autenticação de áudio proposta pode servir como ajuda para os peritos quando se trata de evidências na forma de áudio digital.

A principal contribuição científica deste trabalho é um método de autenticação de

áudio baseado na detecção de mudanças abruptas na fase da ENF provocadas pelas edições feitas no sinal, incluindo uma formulação para determinação da fase de um sinal (com uma componente significativa de frequência) a partir de uma análise de Fourier de alta precisão.

Finalmente, o interesse do IME em contribuir para a perícia fonética se justifica pela aplicabilidade do trabalho na criminalística militar (nos Batalhões de Polícia do Exército) e em ações de inteligência e de Guerra Eletrônica (Análise de Sinais).

### 1.3 ORGANIZAÇÃO DA DISSERTAÇÃO

Após esta introdução geral, no Capítulo 2 aborda-se a fonética forense como ferramenta da criminalística, mencionando-se o exame de autenticação como uma de suas vertentes. Neste capítulo, expressa-se a importância do exame de autenticação de áudio, mencionando como é feita a autenticação na forma analógica e na forma digital. Finalmente, comenta-se como trabalham os sistemas de transmissão elétrica e como a ENF aparece embutida nos sinais de áudio.

O Capítulo 3 aborda duas técnicas para estimação de frequência e fase de um tom simples (assumindo que o sinal de rede elétrica pode ser considerado como tal): a primeira, baseada na Transformada Discreta de Fourier (DFT) de tempo curto; a segunda, baseada numa análise de Fourier de alta precisão (DFT<sup>1</sup>) que também usa a DFT da primeira derivada do sinal. Estas técnicas são testadas em tons gerados com frequência aleatória.

O Capítulo 4 apresenta a técnica proposta de autenticação de áudio digital com base na mudança de fase da rede elétrica. Esta técnica inclui um procedimento para detecção automática da presença da ENF nas gravações; um procedimento de auxílio à localização de pontos de edição pela observação visual de mudanças na fase da rede elétrica (estimadas pelas técnicas descritas no Capítulo 3); e um procedimento automático para discriminar entre sinais originais e sinais editados mediante uma relação de decisão.

No Capítulo 5, a técnica de autenticação de áudio proposta é avaliada em gravações digitalmente editadas. Estes sinais pertencem a três bases de sinais de voz (*corpora*): duas delas gravadas em condições favoráveis (com baixo ruído de fundo e sem saturações), e uma terceira gravada sob condições menos controladas (sinais com ruído de fundo elevado e/ou presença de saturações). Adicionalmente, são analisados efeitos de degradações como ruído de fundo, saturações e variação da ENF sobre a detecção.

Finalmente, no Capítulo 6, fazem-se as considerações finais sobre o trabalho realizado,

que incluem as conclusões e as propostas para trabalhos futuros.

## 2 AUTENTICAÇÃO DE ÁUDIO COMO FERRAMENTA DE AUXÍLIO À FONÉTICA FORENSE

### 2.1 INTRODUÇÃO

No campo da criminalística, as provas físicas servem como evidência para reconstituição dos fatos que aconteceram em determinado caso, podendo assim auxiliar na definição da inocência ou culpa dos implicados durante uma investigação pericial. Estas provas podem ser marcas de sangue, resíduos de pólvora, compostos químicos, impressões digitais, documentos pessoais, etc.

Para o caso da Fonética Forense as provas periciais serão conversações gravadas diretamente de um microfone ou oriundas de ligações telefônicas. As situações mais típicas, onde as gravações constituem um elemento importante no processo investigativo, são os casos de suborno, extorsão, sequestro, chantagem, etc. (BRAID, 2003).

### 2.2 ATUAÇÃO DA FONÉTICA FORENSE

A fonética forense utiliza a descrição, a classificação e a transcrição da fala para fins jurídicos e cíveis. As áreas de atuação da fonética no âmbito forense são: verificação de locutor, melhoria do sinal de áudio degradado, transcrição fonográfica e exame de autenticação (BUTCHER, 2002).

A verificação de locutor visa a determinar se a voz de uma determinada pessoa (suspeito) corresponde a uma voz conhecida, já gravada anteriormente. Na melhoria do áudio, o objetivo é melhorar a qualidade da gravação filtrando um ruído não correlacionado (ruído ambiente, ar-acondicionado, uma canção etc.) do sinal original. A transcrição fonográfica transforma o conteúdo do áudio em documento escrito. O exame de autenticação será analisado na seção seguinte.

### 2.3 EXAME DE AUTENTICAÇÃO

Uma das principais preocupações na análise das gravações é determinar se a gravação possui o conteúdo original ou se foi alterada. A análise de autenticidade é uma atividade complexa que requer um exame detalhado da gravação de áudio (COOPER, 2006). O

exame tem que explicar os motivos de qualquer tipo de anormalidade encontrada; podem-se citar como exemplo as descontinuidades na forma de onda do sinal gravado.

Antigamente, os meios comumente usados para as gravações eram as fitas de áudio (cassete), ou seja, meios analógicos. Atualmente, com o uso da tecnologia digital, é comum usar meios digitais para fazer as gravações. Desta forma, as alterações feitas podem ser dos tipos analógico ou digital.

### 2.3.1 AUTENTICAÇÃO DE ÁUDIO NA FORMA ANALÓGICA

As principais técnicas para autenticação na forma analógica podem ser encontradas em (KOENING, 1990), (DEAN, 1991):

- **Exame crítico do perito ao ouvir a gravação:** A idéia é, mediante um sistema de reprodução de som, ouvir a gravação cuidadosamente. Tem-se que ouvir a gravação o número de vezes necessário para identificar posições de vários eventos ou anomalias que podem, por sua vez, requerer uma análise mais detalhada, possivelmente do tipo instrumental.

O objetivo é tentar identificar, com um ouvido bem treinado, dentre outras coisas: variações dos sons no ruído de fundo, as regiões sem presença de áudio, as mudanças na qualidade de áudio, as características de voz não naturais, os sons particulares de efeitos de pausa, qualquer tipo de ruído gerado eletronicamente, etc.

- **Inspeção física:** O objetivo é achar indícios de alterações, catalogar danos produzidos por usos anormais e manipulação inapropriada. Para isto, o mais comum é examinar a parte exterior para procurar marcas de falsificação, tais como evidências de tampa levantada ou enroscamento do cassete. Também é feita uma análise na fita magnética para verificar consistência no tamanho, na largura, etc. Igualmente, são verificadas perdas de óxido, rachaduras profundas, e estiramento de bordas da fita, entre outras coisas.
- **Exames instrumentais:** Correspondem à análise do sinal no domínio do tempo, sobre a forma de onda; no domínio da frequência, mediante análise espectral; e no domínio da frequência em função do tempo, com ajuda de espectrogramas ou outras distribuições espectrais.

- **Comparação de conteúdo:** Relação de descrições de testemunhas e suspeitos com o conteúdo da fita e comparação com exemplos similares.

### 2.3.2 AUTENTICAÇÃO DE ÁUDIO NA FORMA DIGITAL

Com o uso da tecnologia digital, alterar, editando de alguma maneira, o conteúdo dos sinais de áudio é, hoje em dia, uma atividade considerada simples. Isso aumenta as exigências sobre a perícia fonética, no sentido de determinar a autenticidade das gravações de áudio para aceitá-las como provas em procedimentos legais (KOENING, 1990), (BRIXEM, 2008).

Para o caso de áudio digital, o uso do processamento digital de sinais constitui uma ferramenta adicional que ajuda na avaliação de sua autenticidade. Mesmo considerando que existe no mercado um *software* comercial (Spe, 2005) que avalia a autenticidade de áudio digital, a literatura técnica associada, principalmente sob a perspectiva de processamento digital de sinais, é pobre. Adicionalmente, (KOENING, 2009) expressa que é importante desenvolver métodos para autenticar gravações digitais para a prática no laboratório forense (na atualidade grande parte das evidências se encontram em formatação digital).

Em (COOPER, 2006) é proposta uma técnica de autenticação de áudio digital baseada em verificar se o conteúdo de uma fita analógica foi digitalmente editado antes de ser copiado na fita, o que a desqualificaria como prova física. A ideia é detectar vestígios da interface analógica empregada na gravação da fita adulterada.

Em (SANDERS, 2008), (GRIGORAS, 2005) e (COOPER, 2008), encontram-se já métodos para autenticação de áudio baseados no fato de que a frequência da rede elétrica (ENF- Electric Network Frequency) está presente na grande maioria das gravações de interesse. Este fato faz da ENF um parâmetro potencialmente importante para a autenticação de áudio.

## 2.4 ENF NA AUTENTICAÇÃO DE ÁUDIO DIGITAL

Os sistemas de transmissão de energia elétrica trabalham conectados a subestações de energia numa rede de alta tensão para, posteriormente, mediante transformadores de distribuição, reduzir sua tensão e fazê-la chegar aos nossos lares. A maior quantidade de potência introduzida dentro da rede elétrica é provida por turbinas que trabalham como

geradores de corrente alternada. Desta forma, a velocidade de rotação da turbina é que determina a ENF (COOPER, 2008).

O padrão utilizado para o valor nominal da ENF é de 50 Hz ou 60 Hz. Citamos como exemplos de utilização de 50 Hz os países europeus e alguns países de América do Sul, tais como Argentina, Bolívia, Chile, Uruguai e Paraguai. Por outro lado, são exemplos de utilização de 60 Hz: Equador, Peru, Venezuela, Colômbia, Brasil e Estados Unidos.

Seguindo os critérios de qualidade em projetos de uma rede elétrica, busca-se evitar a perda de sincronismo nas unidades geradoras de energia, e assim manter a voltagem e a frequência dentro de limites aceitáveis. Com estas premissas, considera-se o comportamento da ENF estável, especialmente nas regiões mais desenvolvidas (grandes cidades), onde existe um rigoroso controle. Um exemplo é o caso do sistema de transmissão de energia elétrica da Inglaterra e do País de Gales, administrado pela “*National Grid Company*”, onde a ENF, cujo valor nominal é de 50 Hz, com variações limitadas à faixa de  $\pm 0,2$  Hz em torno deste valor nominal (COOPER, 2008).

Considerando a influência do campo eletromagnético, irradiado basicamente por todo tipo de equipamento elétrico conectado à rede elétrica, a ENF está embutida na maioria das gravações. Pela provável presença e relativa estabilidade, a ENF constitui uma importante ferramenta para a autenticação de áudio.

(SANDERS, 2008), (GRIGORAS, 2005) e (COOPER, 2008) utilizam espectrogramas ou estimativas da ENF por blocos para ter uma representação do comportamento desta frequência no sinal gravado. Considerando-se que são armazenados dados de diferentes regiões com as informações da ENF proveniente de uma tomada, pode-se comparar os dados de um sinal com os dados armazenados e determinar o lugar e o instante em que a gravação foi realizada.

O presente trabalho se baseia na utilização de valores estimados da fase do sinal da rede elétrica, embutido nas gravações digitais. Assume-se aqui que não está disponível um banco de dados com informação da ENF. Assumindo que a ENF não varia muito rapidamente, as mudanças abruptas de fase do sinal da rede elétrica são então usadas como sintomas de que um sinal de áudio foi editado.

Considerando-se o sinal da rede elétrica um tom simples, dois métodos para estimar a frequência e a fase deste tom são apresentados no seguinte capítulo. O primeiro deles é baseado na Transformada Discreta de Fourier e o outro baseado em análise de Fourier de alta precisão.

## 3 ESTIMANDO A FREQUÊNCIA E A FASE

### 3.1 INTRODUÇÃO

Considerando o sinal da rede elétrica um tom simples (sinal cossenoidal), o nosso interesse neste trabalho é achar o valor da sua fase e da sua frequência. Em outras palavras, o que se procura é o valor exato do pico no espectro da frequência e o seu correspondente argumento. (BISCAINHO, 2007) apresenta alguns métodos para estimação de frequência e fase. O algoritmo mais simples é baseado na busca dos máximos da Transformada Discreta de Fourier (DFT) de tempo curto. Um outro algoritmo é baseado na interpolação parabólica do pico a partir dos valores discretos da DFT no seu entorno. Uma outra alternativa é a utilização de uma janela cujo lóbulo principal no espectro tenha a forma de um triângulo. Desta maneira, o ponto de intersecção (no espectro do sinal janelado) entre as linhas que juntam o vértice do triângulo determina a frequência do pico.

Existem outros métodos, paramétricos, como modelos auto-regressivos (AR), modelos auto-regressivos com média móvel (ARMA), etc. Também existem métodos estatísticos que usam modelos ocultos de Markov (HMM) definindo estados que interpretam todas as possíveis continuidades de picos entre todos os picos de blocos sucessivos; a complexidade deste algoritmo é muito elevada.

Neste capítulo, são abordados dois métodos para estimação de frequência e fase: o já mencionado método baseado na Transformada Discreta de Fourier (DFT) de tempo curto e o método baseado na análise de Fourier de alta precisão (denominado DFT<sup>1</sup>).

### 3.2 MÉTODO DA DFT

Primeiramente, definimos  $s_{\text{tone}}(n)$  como um tom simples, considerado como o nosso sinal de interesse, composto de  $M$  amostras. A versão janelada de  $s_{\text{tone}}(n)$  será  $x(n) = s_{\text{tone}}(n)w(n)$ , onde  $w(n)$  representa a janela usada. Neste trabalho, foi usada uma janela de Hann<sup>1</sup>. Subsequentemente,  $X(k)$  corresponde à *DFT* de  $x(n)$ , com  $N_{\text{DFT}} \geq M$  pontos na sua representação espectral.

O nosso objetivo é encontrar o valor máximo de  $|X(k)|$ . Considerando  $k_{\text{peak}}$  como o

---

<sup>1</sup>Implementada com a função “*hanning*” de Matlab®

índice inteiro associado com o valor máximo de  $|X(k)|$ , o valor da frequência estimada é:

$$f_{\text{DFT}} = k_{\text{peak}} \frac{f_s}{N_{\text{DFT}}}, \quad (3.1)$$

onde  $f_s$  é a frequência de amostragem de  $s_{\text{tone}}(n)$ . Uma consideração importante ao método é que  $f_{\text{DFT}}$  pode somente assumir  $\frac{N_{\text{DFT}}}{2}$  valores discretos no intervalo  $0 \leq f_{\text{DFT}} < \frac{f_s}{2}$ . Desta maneira, a resolução dos valores de  $f_{\text{DFT}}$  é  $\frac{f_s}{N_{\text{DFT}}}$ . Isto significa que, usando-se um valor maior de  $N_{\text{DFT}}$ , aumenta-se a probabilidade de  $f_{\text{DFT}}$  assumir um valor mais próximo ao valor real da frequência. A desvantagem é que, à medida que  $N_{\text{DFT}}$  aumenta, a carga computacional no processo de cálculo também cresce. Finalmente, a fase associada a  $s_{\text{tone}}(n)$  é o argumento de  $X(k_{\text{peak}})$ :

$$\phi_{\text{DFT}} = \arg [X(k_{\text{peak}})]. \quad (3.2)$$

### 3.3 MÉTODO DA DFT<sup>1</sup>

O método proposto em (DESAINTE-CATHERINE, 2000), denominado DFT<sup>1</sup>, melhora a estimativa de frequência de um tom com o uso da derivada do sinal. Este método é comumente usado para extrair parâmetros em modelos espectrais de sons. No processo de estimação, este método utiliza a DFT de tempo curto da derivada de primeira ordem do sinal, obtendo uma melhor precisão na busca do pico do espectro do sinal (o valor real da frequência).

O processo básico para estimar a frequência é detalhado passo a passo:

(a) Computar a derivada aproximada de  $s_{\text{tone}}(n)$ :

$$s'_{\text{tone}}(n) = f_s [s_{\text{tone}}(n) - s_{\text{tone}}(n-1)]. \quad (3.3)$$

(b) Obter a versão janelada (janela de Hann) de  $s_{\text{tone}}(n)$  e  $s'_{\text{tone}}(n)$ :

$$\begin{aligned} x(n) &= s_{\text{tone}}(n)w(n), \\ x'(n) &= s'_{\text{tone}}(n)w(n). \end{aligned}$$

(c) Obter a DFT de  $N_{\text{DFT}}$  pontos de  $x(n)$  e  $x'(n)$ . Estas transformadas serão denotadas como  $X(k)$  e  $X'(k)$ , respectivamente.

(d) Computar  $|X(k)|$  e  $|X'(k)|$ , assim como  $k_{\text{peak}}$  (v. Seção 3.2).

(e) Multiplicar  $|X'(k)|$  pelo fator

$$F(k) = \frac{\pi k}{N_{\text{DFT}} \operatorname{sen}\left(\frac{\pi k}{N_{\text{DFT}}}\right)}.$$

Este fator serve para compensar o efeito da inclinação espectral provocado pela derivada no domínio discreto. Nesse ponto, temos:  $\text{DFT}^0[k] = |X(k)|$  e  $\text{DFT}^1[k] = F(k)|X'(k)|$ .

(f) Finalmente, o valor estimado da frequência é:

$$f_{\text{DFT}^1} = \frac{1 \text{ DFT}^1[k_{\text{peak}}]}{2\pi \text{ DFT}^0[k_{\text{peak}}]}. \quad (3.4)$$

Este valor, segundo (DESAINTE-CATHERINE, 2000), pode somente ser considerado válido se  $(k_{\text{peak}} - \frac{1}{2}) < k_{\text{DFT}^1} < (k_{\text{peak}} + \frac{1}{2})$ , onde  $k_{\text{DFT}^1} = \frac{N_{\text{DFT}} f_{\text{DFT}^1}}{f_s}$ .

O mecanismo proposto em (DESAINTE-CATHERINE, 2000) serve para calcular o valor da frequência de um tom simples (este método também pode ser usado para calcular os harmônicos significativos dos sinais) com base na Transformada de Fourier e na derivada do sinal. Como complemento ao método original, que não descreve o cômputo da fase, é proposto aqui um método para calcular a fase de um tom simples.

Inicialmente, podemos modelar  $s_{\text{tone}}(t)$  como  $s_{\text{tone}}(t) = d(t) + s(t)$  (DESAINTE-CATHERINE, 2000), onde  $d(t)$  e  $s(t)$  representam a parte determinística (componente de banda estreita) e a parte estocástica (componente de banda larga) de  $s_{\text{tone}}(t)$ , respectivamente. Para o presente trabalho, é considerada somente a parte determinística. A partir deste modelo, considerando o caso de um tom simples (o sinal com uma única componente de frequência e com uma parcela estocástica de ruído com uma amplitude muito menor),  $s_{\text{tone}}(t)$  pode ser expresso como:

$$s_{\text{tone}}(t) = a \cos(\phi(t)). \quad (3.5)$$

Neste caso a amplitude “ $a$ ” é considerada um valor constante. Adicionalmente,  $\phi(t)$  pode ser expresso por:

$$\phi(t) = \Omega t + \phi_0,$$

onde  $\phi_0$  é a fase inicial do tom e  $\Omega = 2\pi f_{\text{tone}}$ .  $f_{\text{tone}}$  é a frequência real de  $s_{\text{tone}}(t)$ . Para o caso do domínio discreto,  $s_{\text{tone}}(n)$  é:

$$s_{\text{tone}}(n) = a \cos(\omega_0 n + \phi_0), \quad (3.6)$$

onde  $\omega_0 = \frac{2\pi f_{\text{tone}}}{f_s}$ .

Conseqüentemente, a EQ. 3.3 pode ser expressa como:

$$\frac{s'_{\text{tone}}(n)}{af_s} = [\cos(\omega_0 n + \phi_0) - \cos(\omega_0 n - \omega_0 + \phi_0)]. \quad (3.7)$$

A EQ. 3.7 pode ainda ser expressa como<sup>2</sup>:

$$\begin{aligned} \frac{s'_{\text{tone}}(n)}{af_s} &= [\cos(\phi_0) - \cos(\phi_0 - \omega_0)] \cos(\omega_0 n) \\ &\quad - [\sin(\phi_0) - \sin(\phi_0 - \omega_0)] \sin(\omega_0 n). \end{aligned} \quad (3.8)$$

Adicionalmente, devido ao fato que a derivada de um cossenóide é outra cossenóide com a mesma frequência, a EQ. 3.8 pode ser representada por:

$$\frac{s'_{\text{tone}}(n)}{af_s} = C \cos(\omega_0 n + \theta), \quad (3.9)$$

onde  $C$  é um valor constante e  $\theta$  é a fase de  $s'_{\text{tone}}$ . Da EQ. 3.9, pode-se obter:

$$\frac{s'_{\text{tone}}(n)}{af_s} = C \cos(\theta) \cos(\omega_0 n) - C \sin(\theta) \sin(\omega_0 n). \quad (3.10)$$

Comparando a EQ. 3.8 com a EQ. 3.10, pode-se afirmar que:

$$C \cos(\theta) = \cos(\phi_0) - \cos(\phi_0 - \omega_0), \quad (3.11)$$

$$C \sin(\theta) = \sin(\phi_0) - \sin(\phi_0 - \omega_0). \quad (3.12)$$

Dividindo a EQ. 3.12 pela EQ. 3.11, nós obtemos:

$$\text{tg}(\theta) = \frac{\sin(\theta)}{\cos(\theta)} = \frac{\sin(\phi_0) - \sin(\phi_0 - \omega_0)}{\cos(\phi_0) - \cos(\phi_0 - \omega_0)}. \quad (3.13)$$

A EQ. 3.13 pode ser escrita como:

$$\text{tg}(\theta) = \frac{\sin(\phi_0) [1 - \cos(\omega_0)] + \cos(\phi_0) \sin(\omega_0)}{\cos(\phi_0) [1 - \cos(\omega_0)] - \sin(\phi_0) \sin(\omega_0)}. \quad (3.14)$$

Dividindo o numerador e o denominador da EQ. 3.14 por  $\cos(\phi_0)$ , temos:

$$\text{tg}(\theta) = \frac{\text{tg}(\phi_0) [1 - \cos(\omega_0)] + \sin(\omega_0)}{1 - \cos(\omega_0) - \text{tg}(\phi_0) \sin(\omega_0)}. \quad (3.15)$$

Isolando  $\text{tg}(\phi_0)$  da EQ. (3.15), obtêm-se a seguinte expressão:

$$\text{tg}(\phi_0) = \frac{\text{tg}(\theta) [1 - \cos(\omega_0)] + \sin(\omega_0)}{1 - \cos(\omega_0) - \text{tg}(\theta) \sin(\omega_0)}. \quad (3.16)$$

---

<sup>2</sup>É utilizada a fórmula  $\cos(a \pm b) = \cos(a) \cos(b) \pm \sin(a) \sin(b)$ .

Renomeando  $\phi_0$  como  $\phi_{\text{DFT}^1}$ , temos que o valor desejado da fase de  $s_{\text{tone}}$  será dado por:

$$\phi_{\text{DFT}^1} = \text{arctg} \left\{ \frac{\text{tg}(\theta) [1 - \cos(\omega_0)] + \text{sen}(\omega_0)}{1 - \cos(\omega_0) - \text{tg}(\theta) \text{sen}(\omega_0)} \right\}. \quad (3.17)$$

Nessa expressão, o valor de  $\omega_0$  pode ser obtido como  $\omega_0 \approx \frac{2\pi f_{\text{DFT}^1}}{f_s}$ .

Já o valor de  $\theta$  é aproximado por interpolação linear do argumento de  $X'(k)$ . São considerados os valores de:

$$\begin{aligned} k_{\text{low}} &= \text{floor} [k_{\text{DFT}^1}], \\ k_{\text{high}} &= \text{ceil} [k_{\text{DFT}^1}], \end{aligned}$$

onde  $\text{floor} [k_{\text{DFT}^1}]$  aproxima o valor de  $k_{\text{DFT}^1}$  pelo inteiro mais próximo menor ou igual a  $k_{\text{DFT}^1}$ ; e  $\text{ceil} [k_{\text{DFT}^1}]$  aproxima o valor de  $k_{\text{DFT}^1}$  pelo inteiro mais próximo maior ou igual a  $k_{\text{DFT}^1}$ . A posição de  $(k_{\text{DFT}^1}, \arg [X'(k_{\text{DFT}^1})])$  pode ser obtida por interpolação linear entre os pontos  $(k_{\text{low}}, \arg [X'(k_{\text{low}})])$  e  $(k_{\text{high}}, \arg [X'(k_{\text{high}})])$ . O valor desse último argumento será o valor utilizado de  $\theta$ :

$$\theta \approx \arg [X'(k_{\text{DFT}^1})] \approx \arg [X'(k_{\text{low}})] + \frac{k_{\text{DFT}^1} - k_{\text{low}}}{k_{\text{high}} - k_{\text{low}}} \{ \arg [X'(k_{\text{high}})] - \arg [X'(k_{\text{low}})] \}. \quad (3.18)$$

Na FIG. 3.1, é apresentada esta interpolação linear necessária para obter o valor aproximado de  $\theta$ .

Na EQ. 3.17, pode-se ter duas respostas para  $\phi_{\text{DFT}^1}$ : se  $\text{arctg}(\phi_{\text{DFT}^1})$  é um valor positivo,  $\phi_{\text{DFT}^1}$  pode estar localizado no primeiro ou no terceiro quadrante do sistema cartesiano de duas dimensões; por outro lado, se  $\text{arctg}(\phi_{\text{DFT}^1})$  é um valor negativo,  $\phi_{\text{DFT}^1}$  pode estar no segundo ou no quarto quadrante. Utilizamos então o valor de  $\phi_{\text{DFT}}$  (valor estimado pelo método da DFT) como uma referência, ou seja, usamos o valor de  $\phi_{\text{DFT}^1}$  que seja mais próximo a  $\phi_{\text{DFT}}$ .

### 3.4 EXPERIMENTOS SINTÉTICOS

Para uma primeira avaliação dos métodos de estimação discutidos, realizaram-se alguns experimentos artificiais. Inicialmente, foi feita uma análise para o caso particular de um tom de áudio fixo e posteriormente uma análise estatística sobre tons gerados com frequência aleatória.

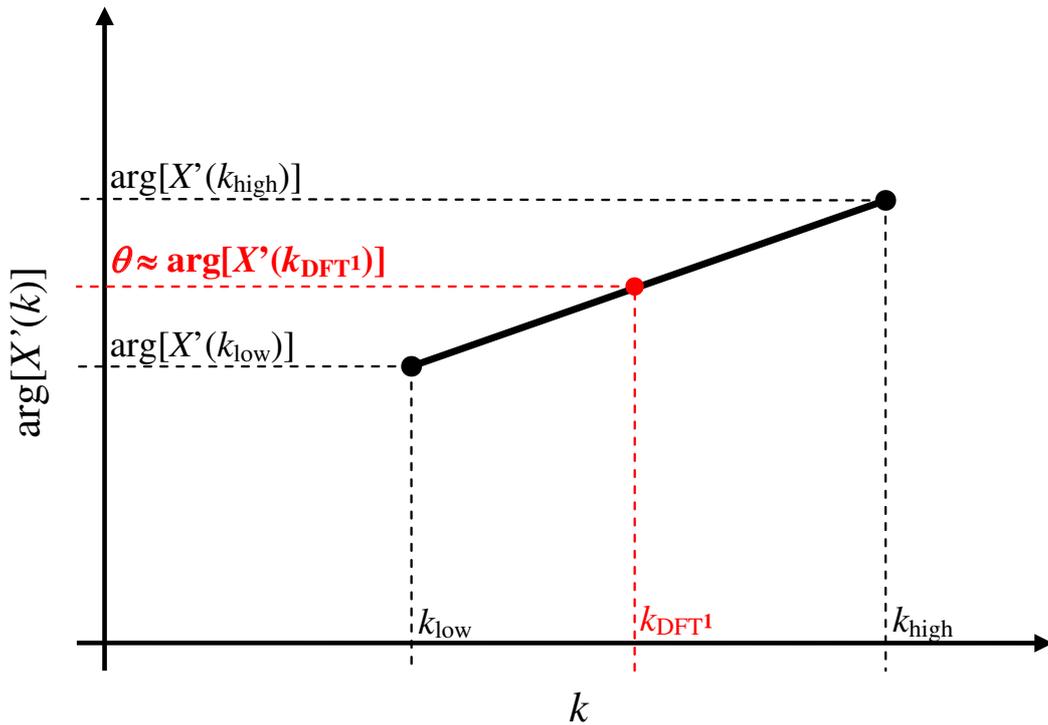


FIG. 3.1: Interpolação linear no argumento de  $X'(k)$  para a obtenção do valor aproximado de  $\theta$

O primeiro caso de análise é baseado num tom gerado computacionalmente e cuja frequência é de 60,98 Hz. A frequência de amostragem usada é de 1200 Hz.

Como ajuda visual para a interpretação do espectro desse tom, apresentamos na FIG. 3.2 um pedaço do espectro do tom janelado com 500 amostras. A janela usada é a janela de Hann. A FIG. 3.2 apresenta o espectro real  $|X(e^{j\omega})|_{\text{dB}}$  assim como o espectro baseado na DFT com  $N_{\text{DFT}} = 2000$  pontos. Como se pode observar, usando a DFT no processo de estimação não se pode conseguir a posição exata do pico do espectro do sinal.

Para testar os dois métodos de estimação propostos, encontramos as 100 primeiras estimativas de fase e frequência do mencionado tom. Estas estimativas foram feitas em blocos segmentados do sinal original. O tamanho do bloco é de 200 amostras com superposição de 199 amostras. Isto significa que realizamos uma estimação de frequência e fase amostra a amostra (*sliding window*).

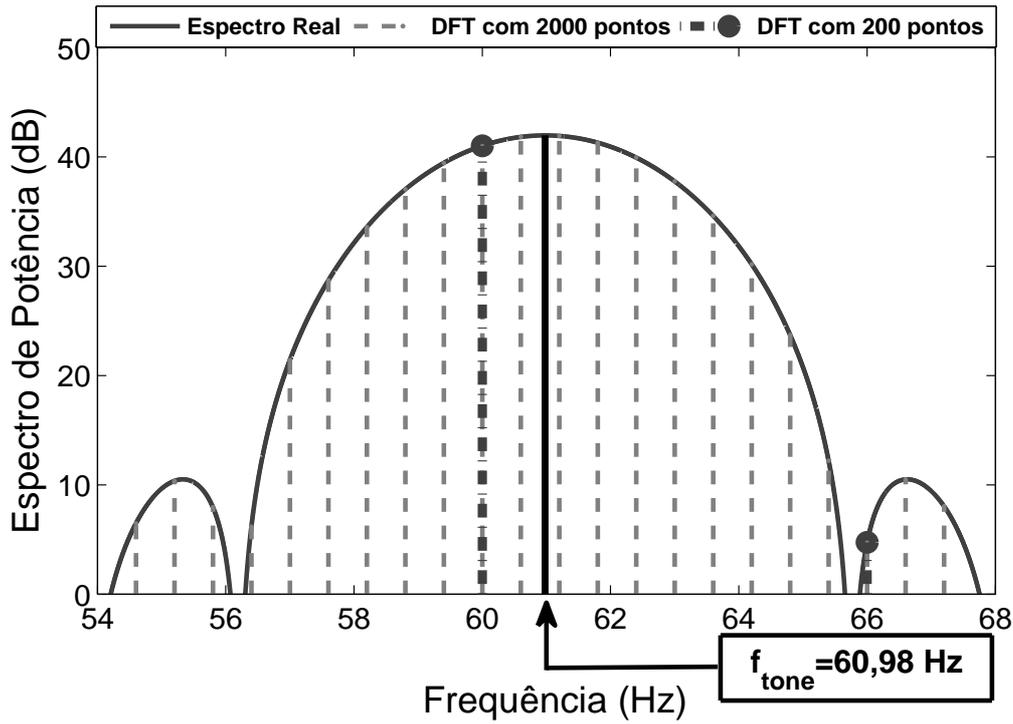


FIG. 3.2: Espectro de uma de tom de 60,98 Hz janelado. A frequência de amostragem usada é de 1200 Hz. A janela usada é de Hann com 500 amostras.

Usando o método da DFT com  $N_{\text{DFT}} = 200$  no processo de estimação de frequência, o valor achado nas 100 estimativas de frequência foi de 60,0 Hz. Usando o mesmo método de estimação de frequência com  $N_{\text{DFT}} = 2000$  pontos, o valor achado nas 100 estimativas de frequência foi de 61,2 Hz. Por outro lado, usando o método da DFT<sup>1</sup> com  $N_{\text{DFT}} = 200$ , os valores da estimativa de frequência possuem um valor médio de 60,9717 com desvio padrão de 0,0025 Hz. Usando DFT<sup>1</sup> com  $N_{\text{DFT}} = 2000$ , os valores da estimativa de frequência possuem um valor médio de 60,9818 com desvio padrão de 0,0032 Hz.

Para medir o erro das estimativas de frequência de uma maneira percentual, definimos  $e_f$  como o valor médio do erro das estimativas de frequência obtidas. A expressão matemática para  $e_f$  é:

$$e_f = \frac{1}{N_{\text{Block}}} \sum_{n_b=1}^{N_{\text{Block}}} \left[ \frac{|\hat{f}(n_b) - f_{\text{tone}}|}{f_{\text{tone}}} \cdot 100\% \right], \quad (3.19)$$

onde  $\hat{f}(n_b)$  representa a frequência estimada no  $n_b$ -ésimo bloco,  $f_{\text{tone}}$  representa a fre-

quência real do tom (valor nominal) e  $N_{\text{Block}}$  é o número total de blocos. Para o nosso caso,  $N_{\text{Block}} = 100$ .

O valor de  $e_f$  para as estimativas de frequência usando o método da DFT foi de 1,61% para o caso de  $N_{\text{DFT}} = 200$  e de 0,36% para  $N_{\text{DFT}} = 2000$ . Para o método da DFT<sup>1</sup>, o valor obtido para  $e_f$  foi de 0,014% para  $N_{\text{DFT}} = 200$  e de 0,005% para  $N_{\text{DFT}} = 2000$ . Com estes valores, podemos afirmar que os erros obtidos com o método da DFT<sup>1</sup> são substancialmente menores que os erros obtidos com o método da DFT.

Por outro lado, na FIG. 3.3 são apresentadas as estimativas de fase para do tom de 60,98 Hz anteriormente analisado, obtidas pelo método da DFT e pelo método da DFT<sup>1</sup>, nos dois casos com  $N_{\text{DFT}} = 200$ . Visualmente, na FIG. 3.3, pode-se apreciar a maior exatidão da fase obtida com o método da DFT<sup>1</sup>, se comparada com a fase original.

Para medir o erro das estimativas de fase, definimos  $e_\phi$  como o valor médio do valor absoluto da diferença entre as estimativas obtidas e os valores originais. A expressão matemática para  $e_\phi$  é:

$$e_\phi = \frac{1}{N_{\text{Block}}} \sum_{n_b=1}^{N_{\text{Block}}} \left| \hat{\phi}(n_b) - \phi(n_b) \right|, \quad (3.20)$$

onde  $\hat{\phi}(n_b)$  representa a fase estimada no  $n_b$ -ésimo bloco,  $\phi(n_b)$  representa a fase original do tom e  $N_{\text{Block}}$  representa o número total de blocos.

O valor de  $e_\phi$  para as estimativas de fase usando o método de DFT foi de 29,2530 graus para o caso de  $N_{\text{DFT}} = 200$  e de 6,5670 graus para  $N_{\text{DFT}} = 2000$ . Para o método da DFT<sup>1</sup>, o valor obtido para  $e_\phi$  foi de 0,2464 grau para  $N_{\text{DFT}} = 200$  e de 0,0912 grau para  $N_{\text{DFT}} = 2000$ . Tal como no caso das estimativas de frequência, pode-se afirmar que com o uso da DFT<sup>1</sup> se conseguem os melhores valores estimados.

Para uma avaliação estatística dos dois métodos (DFT e DFT<sup>1</sup>), foram criados mil tons com frequências aleatórias variando (com distribuição uniformemente) entre 59,0 Hz e 61,0 Hz. Para estes tons, foram obtidas as estimativas de fase e frequência e seus erros de estimação:  $e_f$  e  $e_\phi$  (como para o caso do tom de 60,98 Hz). Os erros para a nossa avaliação estatística serão os valores médios dos erros de estimação para os tons gerados. Estes erros foram achados para diferentes números de pontos da DFT ( $N_{\text{DFT}}$ ) e tamanhos de janela ( $M$ ). A TAB. 3.1 apresenta um resumo da avaliação estatística realizada.

A TAB. 3.1 demonstra que quando o número de pontos da DFT aumenta (considerando um tamanho de janela constante), as estimativas de fase e frequência melhoram nos dois

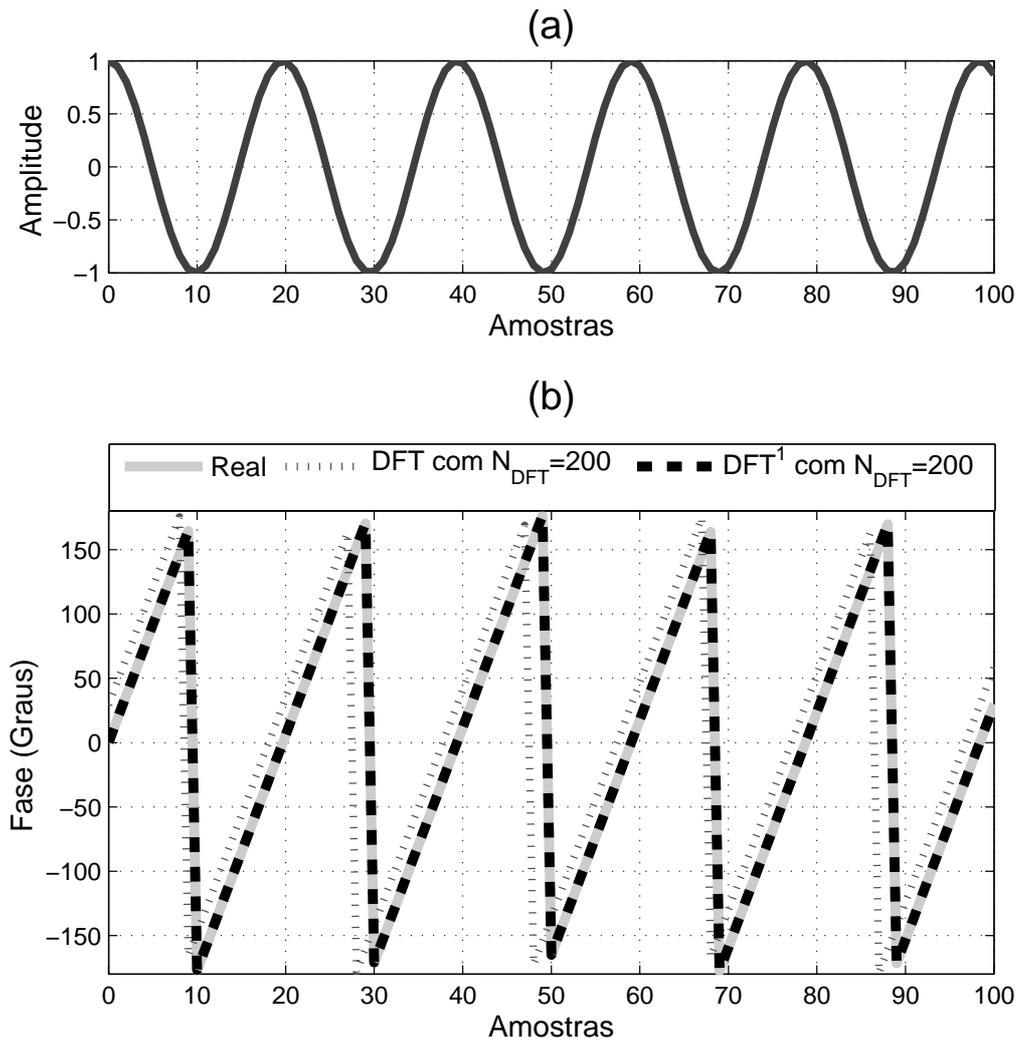


FIG. 3.3: Estimação de fase de um tom artificial de 60,98 Hz. A frequência de amostragem usada é de 1200 Hz. (a) Sinal; (b) Fase estimada.

métodos. Isto se justifica, pois com uma maior quantidade de pontos (raias) na representação discreta do espectro, a probabilidade de ter uma resposta mais acurada aumenta. Adicionalmente, no método da DFT<sup>1</sup>, comparando com o método da DFT, observa-se uma melhoria substancial nas estimativas de fase e frequência (considerando o mesmo  $N_{DFT}$ ). Esta melhoria pode ser constatada verificando-se que as estimativas do método da DFT<sup>1</sup> com menor número de pontos da DFT ( $N_{DFT} = 200$ ) já são melhores que as estimativas do método da DFT com maior número de pontos ( $N_{DFT} = 20000$ ). O método

TAB. 3.1: Avaliação estatística das estimativas de fase e frequência usando os métodos DFT e DFT<sup>1</sup>. O experimento foi levado a cabo com mil tons gerados com frequências aleatórias (com distribuição uniforme) variando entre 59,0 e 61,0 Hz.  $\bar{e}_f$  e  $\bar{e}_\phi$  representam os valores médios dos erros na estimação de frequência e de fase para os tons gerados.

Método	$M$ (amostras)	$N_{\text{DFT}}$ (pontos)	$\bar{e}_f$ (%)	$\bar{e}_\phi$ (graus)
DFT	60	200	0,8306	4,4086
		2000	0,2696	1,4309
		20000	0,0679	0,3606
	100	200	0,8189	7,2944
		2000	0,2688	2,3949
		20000	0,0261	0,2329
	200	200	0,8180	14,6496
		2000	0,2588	4,6359
		20000	0,0246	0,4411
DFT <sup>1</sup>	60	200	0,0559	0,2907
		2000	0,0543	0,2831
		20000	0,0538	0,2802
	100	200	0,0138	0,1217
		2000	0,0131	0,1160
		20000	0,0130	0,1152
	200	200	0,0024	0,1221
		2000	0,0024	0,0559
		20000	0,0015	0,0436

da DFT<sup>1</sup> utiliza nas suas operações duas DFTs de tamanho menor além de outras operações simples. Como um exemplo, podemos afirmar que, usando o método da DFT<sup>1</sup> com  $N_{\text{DFT}} = 200$ , a carga computacional é aproximadamente cinquenta vezes menor que a do método da DFT com  $N_{\text{DFT}} = 20000$ . Comparando os resultados, vemos que usando o método da DFT<sup>1</sup> obtemos erros menores com menor carga computacional (menor  $N_{\text{DFT}}$ ).

### 3.5 CONSIDERAÇÕES FINAIS

Neste capítulo foram apresentadas duas técnicas para estimação de frequência e fase para um tom simples. A primeira é baseada na Transformada Discreta de Fourier (DFT) de tempo curto; e a outra, denominada DFT<sup>1</sup>, é baseada num mecanismo de alta precisão que utiliza além da DFT de tempo curto a primeira derivada do sinal. Com esta técnica de alta precisão, foi possível obter maior acurácia na localização do pico do espectro

do sinal (frequência do sinal). O mecanismo proposto por (DESAINTE-CATHERINE, 2000) considera somente a estimativa dos valores dos harmônicos de frequência de um sinal (DESAINTE-CATHERINE, 2000); neste capítulo foi proposto um complemento a este método para determinar o valor da fase de um sinal que possui um harmônico significativo de frequência (um tom simples por exemplo).

A técnica de estimação de fase proposta, bem como as de estimação de frequência, foram avaliadas e verificou-se que aquelas baseadas na DFT<sup>1</sup> são muito mais precisas que as baseadas na DFT.

No capítulo seguinte é proposto um algoritmo para detectar a autenticidade de áudio com base na mudança de fase do sinal da rede elétrica, quase sempre embutido nas gravações. Considerando que o sinal da rede elétrica pode ser visto como um tom simples, serão utilizadas as técnicas de estimação de fase propostas neste capítulo.

## 4 AUTENTICAÇÃO DE ÁUDIO DIGITAL COM BASE NA MUDANÇA DE FASE DO SINAL DA REDE ELÉTRICA

### 4.1 INTRODUÇÃO

Como mencionado no Capítulo 2, a frequência da rede elétrica (ENF - *Electric Network Frequency*) encontra-se embutida na maioria das gravações. Este fenômeno é devido ao campo eletromagnético gerado por um equipamento elétrico conectado à rede (COOPER, 2008). Considerando que em edição de áudio digital pode-se fazer recortes e/ou inserções no conteúdo do sinal, os mesmos são também aplicados sobre a ENF embutida.

A idéia neste trabalho é determinar as mudanças de fase na ENF devidas às edições feitas no sinal de áudio. O método proposto é baseado em (NICOLALDE, 2009a) e (NICOLALDE, 2009b).

### 4.2 EXPERIMENTOS PRELIMINARES

Em nosso primeiro experimento, geramos um sinal senoidal artificial de 60,0 Hz, com taxa de amostragem de 11025 Hz (frequência típica de amostragem em sinais de áudio, 44100/4 Hz). Posteriormente, editamos este sinal de tal forma que dois intervalos ou fragmentos do sinal fossem trocados de lugar, provocando quatro possíveis mudanças de fase (os pontos onde foram feitas as mudanças de fase serão denominados  $P1$ ,  $P2$ ,  $P3$  e  $P4$ ). O sinal original, assim como o sinal editado, foram sub-amostrados para a taxa de 1000 Hz, com o objetivo de diminuir a carga computacional do processo. Depois da sub-amostragem, um filtro FIR passa-banda, de fase linear e corte muito agudo, foi aplicado aos sinais original e editado. O filtro passa-banda foi centrado em 60,0 Hz e projetado com uma largura de banda de 1,4 Hz, realizado com 10000 coeficientes numa implementação de fase zero por filtragem direta e reversa (função “*filtfilt*” do Matlab®).

A FIG. 4.1 apresenta o efeito do filtro passa banda num tom de 60 Hz editado com leves mudanças de fase. Nesta figura, pode-se observar que:

- a) os pontos  $P1$ ,  $P2$ ,  $P3$ , e  $P4$  apresentam pequenas variações de fase;
- b) depois de se filtrar o tom editado, aparece o fenômeno de redução de amplitude no

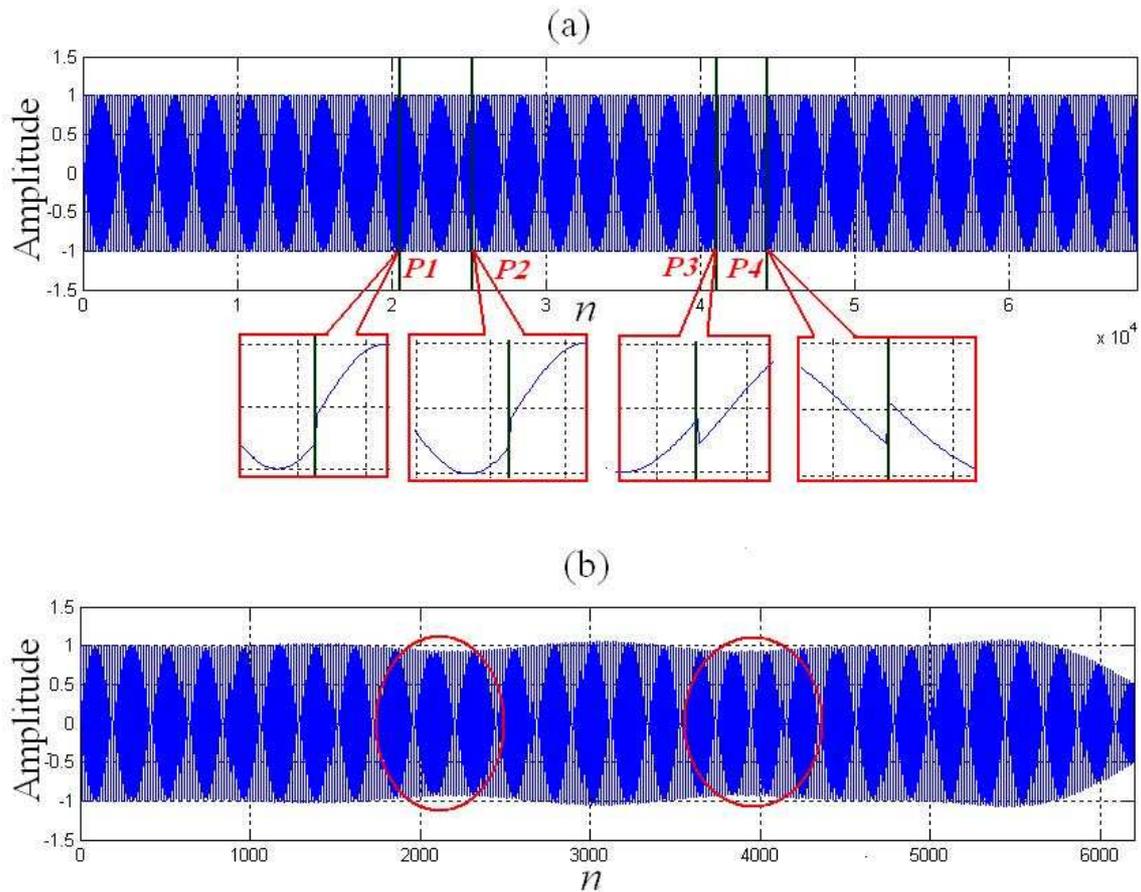


FIG. 4.1: Tom de 60 Hz editado com pequenas discontinuidades de fase. (a) Sinal editado e as suas mudanças de fase ( $f_s = 11025$  Hz); (b) Sinal filtrado ( $f_s = 1000$  Hz).

sinal filtrado na proximidade dos pontos de edição.

A FIG. 4.2 apresenta o efeito do filtro passa-banda num tom de 60 Hz editado com consideráveis mudanças de fase. Nesta figura, pode-se observar que:

- os pontos  $P1$ ,  $P2$ ,  $P3$ , e  $P4$  apresentam maiores variações de fase;
- depois de se filtrar o tom editado, aparece com maior realce o efeito da “modulação” em amplitude nas proximidades dos pontos de edição (a energia contida nas harmônicas foi filtrada), o que claramente sugere que o sinal foi editado.

Dos experimentos mostrados anteriormente, pode-se perceber que a filtragem (necessária para considerar somente as componentes da ENF) suaviza o efeito da mudança de fase. Isto se deve a alguns componentes de frequência (harmônicos) necessários para descrever

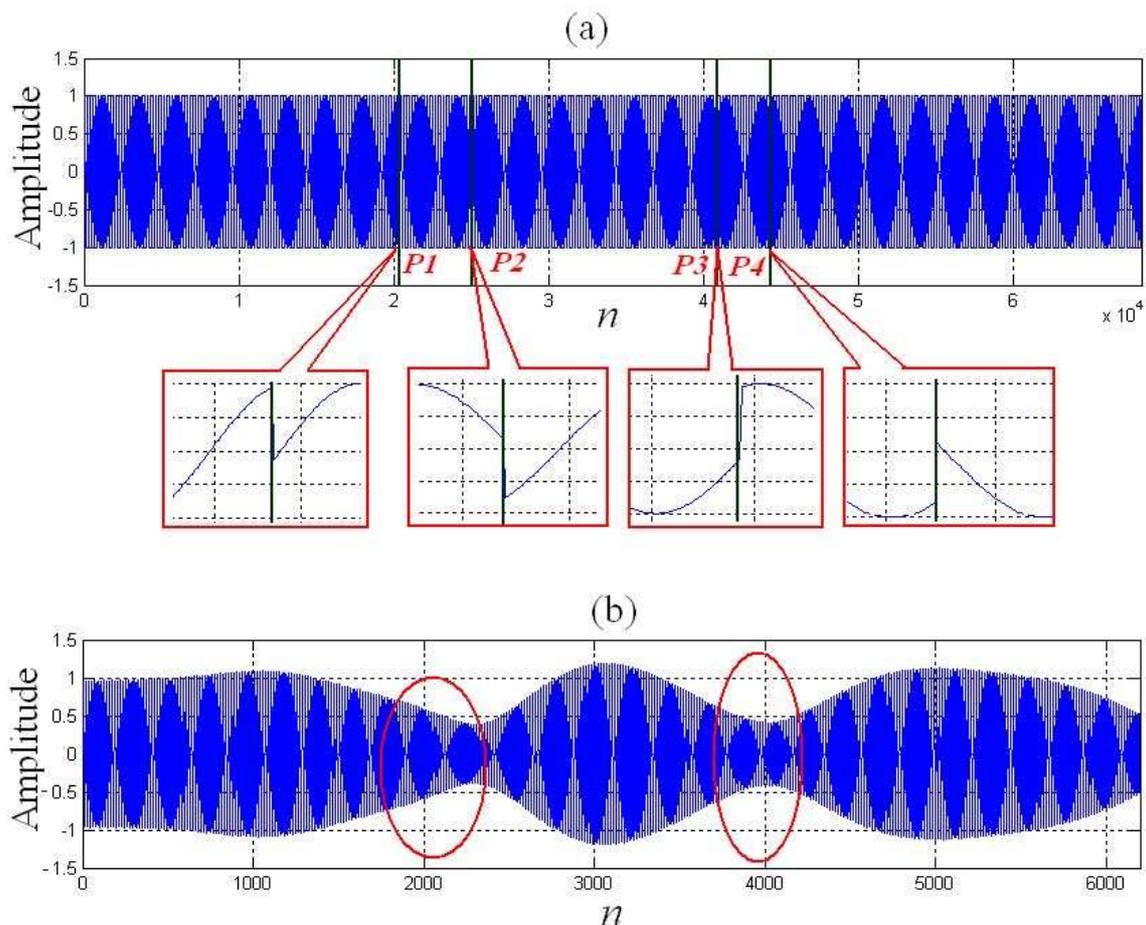


FIG. 4.2: Tom de 60 Hz editado com consideráveis descontinuidades de fase. (a) Sinal editado e as suas mudanças de fase ( $f_s = 11025$  Hz); (b) Sinal filtrado ( $f_s = 1000$  Hz).

a mudança brusca de fase serem eliminados pela filtragem passa-banda. Adicionalmente, para um caso como o da FIG. 4.1, é mais complicado verificar o efeito da filtragem no sinal editado devido às pequenas variações de fase que apresenta. Esta situação particular (como da FIG. 4.1) é considerada uma das mais difíceis para a técnica proposta.

Por outro lado, mesmo uma pessoa com conhecimento técnico, se não tiver conhecimento prévio deste assunto específico, não evitará deliberadamente criar modificações abruptas na fase da ENF durante a edição do áudio.

### 4.2.1 INFORMAÇÃO VISUAL DA MEDIDA DA FASE

Em lugar de usar informação no domínio do tempo (efeito de “modulação”) resultante da filtragem passa-banda como sintoma de que um sinal foi editado, é conveniente medir a fase da ENF, com o objetivo de melhor identificar as descontinuidades causadas pelas edições.

O procedimento inicial utilizado para obter a medida de fase foi a aplicação da DFT (Transformada Discreta de Fourier) de blocos janelados do sinal filtrado (com tamanho correspondente a 3 ciclos do tom de 60 Hz—50 milissegundos) ao longo do tempo. A janela aplicada foi a de Hann. Aqui, o uso da DFT é suficiente, considerando que estamos trabalhando com um tom artificial cujo pico no espectro do sinal (posicionado na frequência do tom) coincide com uma das componentes do espectro da DFT—cujo número de pontos foi escolhido de forma que uma de suas raias correspondesse exatamente a 60 Hz. Finalmente, o valor em graus do ângulo da DFT nesse ponto é a fase da ENF.

As curvas de estimação de fase para os casos de mudanças leves e mudanças consideráveis de fase são apresentadas na FIG. 4.3. Claramente, percebe-se que a edição é mais visível e a medida aproxima melhor o caso real no caso de mudanças consideráveis de fase. Adicionalmente, é importante considerar que o tempo de duração dos fragmentos de inserção é curto (fragmento entre  $P1$  e  $P2$  aproximadamente de 475 ms e o fragmento entre  $P3$  e  $P4$  aproximadamente de 400 ms) e com isso o efeito da filtragem (o grande número de coeficientes do filtro FIR) afeta a visibilidade das transições de fase do tom, ainda mais quando as mudanças são leves.

No próximo experimento, com o objetivo de analisar quantitativamente o efeito da mudança de fase no caso mais simples de um tom de 60,0 Hz editado, removeu-se um trecho deste tom de forma que a cossenóide sofresse uma determinada descontinuidade de fase. Com a finalidade de simular uma situação real, o processo de estimação de fase foi feito depois de se filtrar este tom. A FIG. 4.4 apresenta as curvas de estimação de fase para diferentes valores de mudanças de fase entre  $0^\circ$  e  $180^\circ$ . O ponto de edição,  $k = 4000$ , é o mesmo para todos os casos. A frequência de amostragem usada foi  $f_s = 1000$  Hz.

## 4.3 MÉTODO PROPOSTO

Com base nos experimentos artificiais anteriormente apresentados e considerando que o sinal da rede elétrica está quase sempre embutido nos sinais gravados, é proposto um

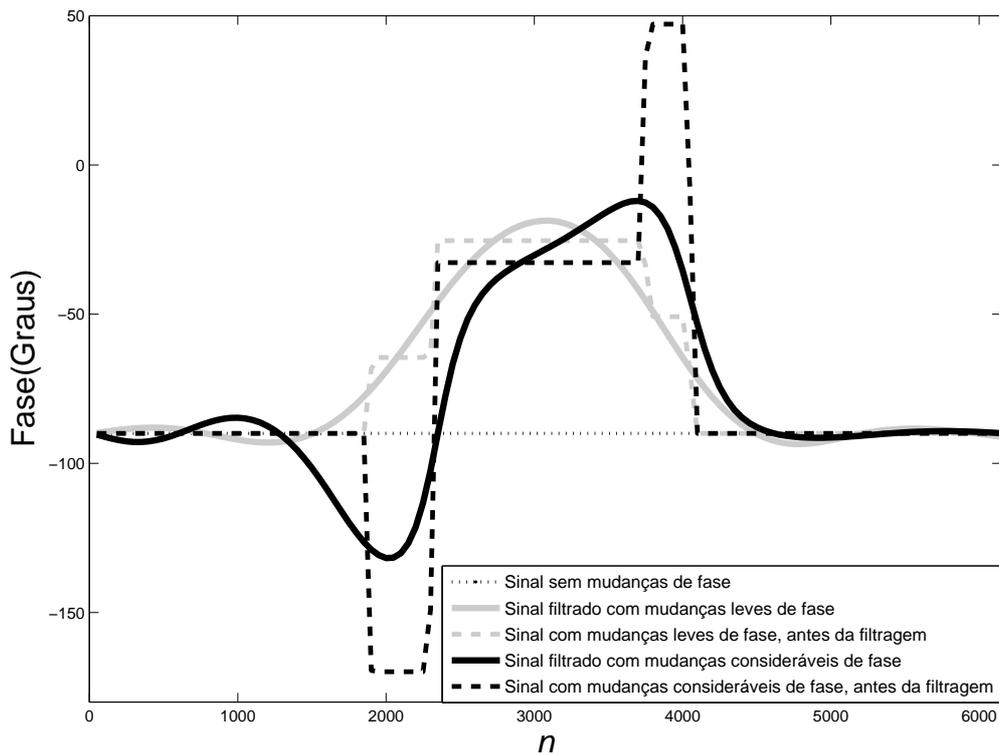


FIG. 4.3: Estimação de fase para os casos: tom com mudanças leves de fase (FIG. 4.1) e tom com mudanças consideráveis de fase (FIG. 4.2). A curva pontilhada corresponde ao tom não editado. Note que, neste exemplo, temos duas inserções de trechos no sinal que podem melhor ser percebidas pelas curvas tracejadas.

método que tenta detectar as mudanças abruptas na fase deste sinal.

O método pode ser dividido em três partes. A primeira delas mostra um mecanismo de detecção automática para verificação da existência da ENF no sinal de áudio. A segunda parte apresenta um mecanismo visual que permite observar o comportamento da fase estimada do sinal da rede elétrica. Finalmente, a terceira parte aborda um mecanismo para discriminação automática entre sinais originais e editados por meio de uma relação de decisão.

#### 4.3.1 DETECÇÃO AUTOMÁTICA DA ENF

O primeiro passo na nossa proposta de autenticação de áudio digital é detectar a presença da ENF no sinal de áudio analisado. O método mais simples é a verificação da

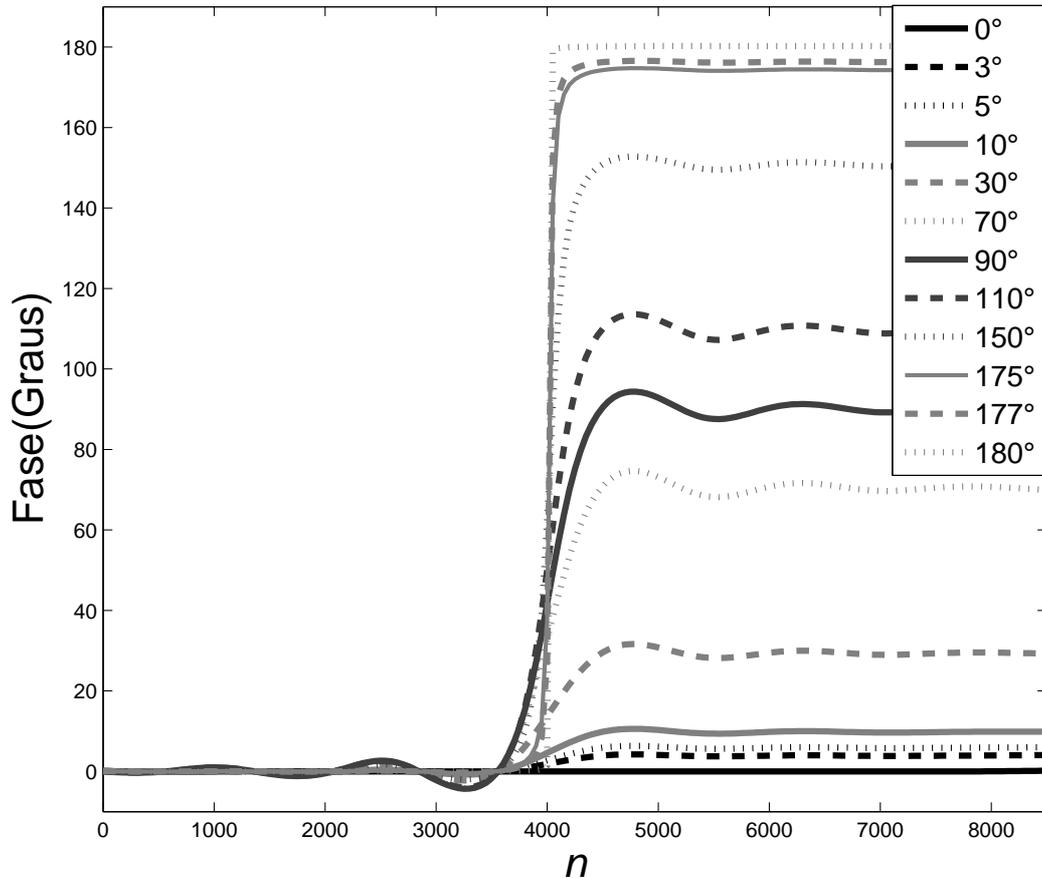


FIG. 4.4: Curvas de estimação de fase para diferentes valores de mudanças de fase num tom de 60 Hz editado (uma porção do sinal em  $k = 4000$  foi removido).

raia da componente espectral no valor da ENF, seja esta 50 ou 60 Hz. Contudo, é sempre importante parametrizar precisamente a presença deste sinal e desta maneira chegar-se a uma detecção automática.

O mecanismo de detecção automática da ENF utilizado neste trabalho é explicado a seguir:

- a) Realizar uma sub-amostragem<sup>3</sup> do sinal de áudio para a frequência  $f_{\text{downsample}}$  de modo a diminuir a carga computacional no processamento. A frequência  $f_{\text{downsample}}$

---

<sup>3</sup>O processo de sub-amostragem foi feito usando a função *resample* do Matlab<sup>®</sup>, que aplica um filtro FIR (passa-baixas) anti-aliasing durante o processo de sub-amostragem.

pode, por exemplo, adotar valores entre 1000 Hz e 1500 Hz.

- b) Aplicar ao sinal sub-amostrado um filtro passa-banda, muito estreito e de fase linear centrado no valor nominal da ENF e com largura de faixa de 0,6 Hz. Nas simulações realizadas neste trabalho, o filtro utilizado foi um FIR de 10000 coeficientes, sendo a filtragem realizada de modo a não introduzir retardo em relação ao sinal original (função *filtfilt* do Matlab<sup>®</sup>). Denominaremos este sinal filtrado de  $s_{NB}(n)$ .
- c) Aplicar ao sinal sub-amostrado um filtro passa-banda similar ao do item b), mas com largura de faixa de 5 Hz. Denominaremos este sinal filtrado como  $s_{WB}(n)$ .
- d) Determinar a relação  $R_{ENF}$ :

$$R_{ENF} = 10 \log \left\{ \frac{\text{var} [s_{NB}(n)]}{\text{var} [s_{WB}(n)]} \right\}. \quad (4.1)$$

Esta relação não é outra coisa que a relação logarítmica das variâncias dos sinais filtrados com banda estreita de 0,6 Hz e de banda mais larga de 5 Hz. Com base nesse precedente, quando existe a ENF nas gravações,  $R_{ENF}$  assume valores próximos a 0 dB (a componente da ENF é a mais significativa no sinal, mesmo considerando uma faixa de 0,6 Hz ou 5 Hz). Por outro lado, quando a ENF é ausente,  $R_{ENF}$  assume valores afastados de 0 dB (como não existe uma componente significativa no espectro, indiscutivelmente a variância ou potência do sinal que possui banda maior vai ser maior). É importante considerar que  $R_{ENF}$  adota valores negativos.

- e) Definir se a ENF está presente ou ausente com base na seguinte relação de decisão:

$$R_{ENF} \underset{H_{\sim ENF}}{\overset{H_{ENF}}{\gtrless}} \gamma_{ENF}. \quad (4.2)$$

onde  $H_{ENF}$  e  $H_{\sim ENF}$  representam as hipóteses de presença e ausência da ENF no sinal de áudio digital, respectivamente.  $\gamma_{ENF}$  é o limiar utilizado para a decisão final. Com base na relação exposta na EQ. 4.2, quando  $R_{ENF}$  adota valores maiores do que  $\gamma_{ENF}$  decide-se que a ENF está presente no sinal de áudio.

Para determinar o valor de  $\gamma_{ENF}$  é necessária a preparação de um banco de dados com sinais que possuam a ENF e com sinais que não possuam a ENF no espectro. Posteriormente, calculam-se os valores de  $R_{ENF}$  para todos os sinais. Subsequentemente, determinam-se os histogramas separados dos valores de  $R_{ENF}$  para sinais que possuam e não possuam a ENF no espectro. Finalmente, escolhe-se um valor de

$\gamma_{\text{ENF}}$  que esteja entre os histogramas (dos valores de  $R_{\text{ENF}}$ ) para sinais que possuem a ENF e para sinais que não a possuem. Quando não existe superposição, o valor  $\gamma_{\text{ENF}}$  pode ser o valor médio entre o valor máximo de  $R_{\text{ENF}}$  para os sinais que não possuem a ENF e o valor mínimo de  $R_{\text{ENF}}$  para os sinais que possuem a ENF.

Na Seção 5.3.1, são apresentados histogramas da relação  $R_{\text{ENF}}$  para sinais que possuem ENF e para sinais que carecem dela. Adicionalmente são apresentados valores do limiar de  $\gamma_{\text{ENF}}$  para os casos da ENF ser 50 Hz e 60 Hz.

Uma vez que é comprovada a presença da ENF, pode-se continuar com os seguintes (dois) mecanismos para autenticação de áudio.

#### 4.3.2 MECANISMO VISUAL

O mecanismo visual é explicado passo a passo:

- a) Realizar uma sub-amostragem do sinal de áudio para a frequência  $f_{\text{downsample}}$ , que pode, por exemplo, adotar valores entre 1000 Hz e 1500 Hz. O objetivo desta sub-amostragem é diminuir a carga computacional no processamento. Como sugestão, recomendamos usar os valores para  $f_{\text{downsample}}$  de 1000 Hz ou 1200 Hz, dependendo de o valor nominal da ENF ser 50 Hz ou 60 Hz, respectivamente. Com estes valores de frequência, pode-se trabalhar com um número exato de amostras por ciclo do valor nominal da ENF (20 amostras, para os dois casos), o que facilita o processo proposto.
- b) Aplicar ao sinal sub-amostrado o mesmo filtro passa banda do item b) da Seção 4.3.1; mas com largura de faixa sugerida (baseada em nossos experimentos) entre 0,6 e 1,4 Hz, dependendo da tolerância permitida da ENF (tal tolerância pode ser obtida com a companhia fornecedora de energia elétrica ou experimentalmente).
- c) Segmentar o sinal filtrado em blocos de duração de  $n_1$  ciclos da ENF nominal com sobreposição de  $n_2$  ciclos com respeito ao bloco seguinte. Sugere-se usar valores de  $n_1$  entre 3 e 10 ciclos da ENF, sendo  $n_2 < n_1$  obrigatoriamente. Após este processo, o sinal é segmentado em  $N_{\text{Block}}$  blocos. A idéia é obter  $N_{\text{Block}}$  estimativas de fase ao longo do tempo. Na FIG. 4.5, consideram-se 3 ciclos de ENF nominal com sobreposição de 2 ciclos.

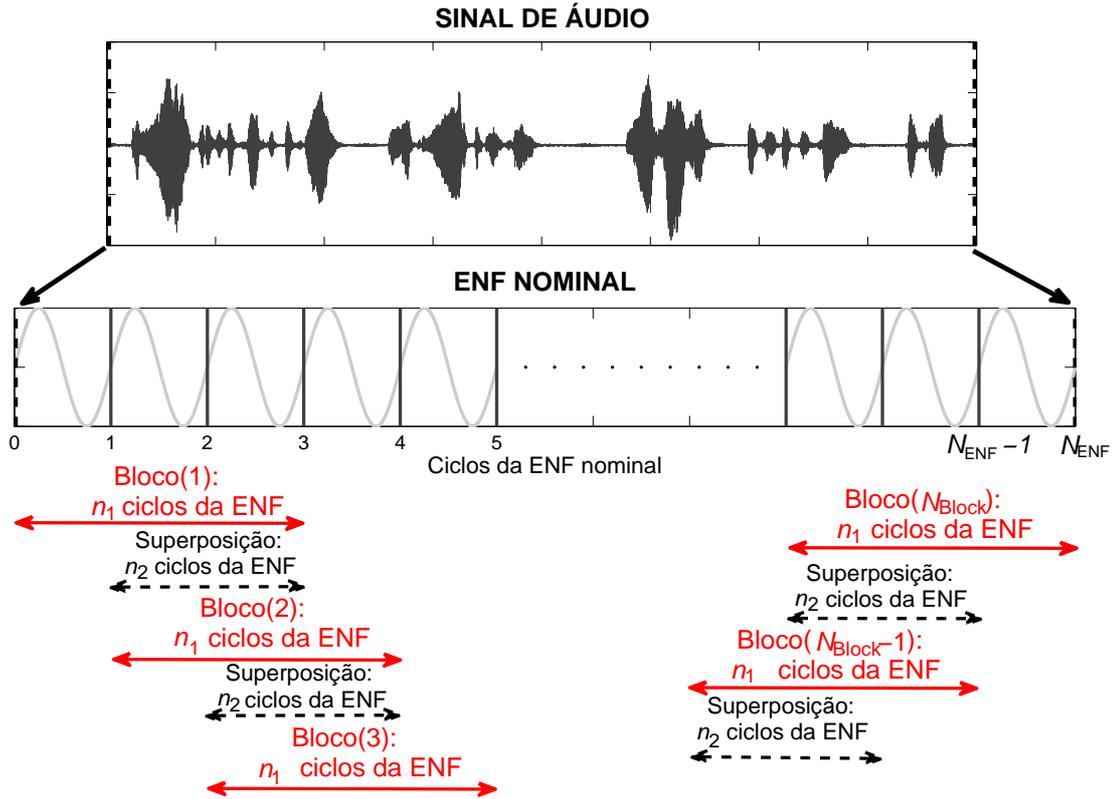


FIG. 4.5: Esquema de segmentação em blocos do sinal de áudio.  $N_{\text{ENF}}$  representa o número total de ciclos da ENF nominal contidos no sinal de áudio e  $N_{\text{Block}}$ , o número total de blocos segmentados.

- d) Estimar a fase de cada bloco segmentado usando os métodos da DFT (Seção 3.2) ou o método da DFT<sup>1</sup> (Seção 3.3). Definimos a  $\hat{\phi}(n_b)$  como a fase estimada do bloco correspondente ao índice  $n_b$ .

Na Seção 5.3.2, são apresentados alguns exemplos do procedimento visual para a detecção de edição de áudio enunciada nesta Seção.

### 4.3.3 MECANISMO AUTOMÁTICO

Com o objetivo de discriminar automaticamente sinais editados dos sinais originais, é necessário determinar uma medida (*feature*) que caracterize as mudanças abruptas de fase do sinal da rede elétrica, embutido no sinal de áudio e relacionado com o nosso método de detecção de edição.

Primeiramente, definimos  $\hat{\phi}'(n_b)$  como a variação da fase estimada  $\hat{\phi}(n_b)$  do sinal da rede elétrica do bloco anterior para o bloco atual:

$$\hat{\phi}'(n_b) = \hat{\phi}(n_b) - \hat{\phi}(n_b - 1). \quad (4.3)$$

O objetivo de computar-se  $\hat{\phi}'(n_b)$  é verificar as variações temporais no comportamento da curva de  $\hat{\phi}(n_b)$ . Quando um sinal não foi editado, espera-se que  $\hat{\phi}'(n_b)$  se assemelhe a uma linha reta com valor constante ao longo do tempo, dada pela inclinação da curva de estimação de  $\hat{\phi}(n_b)$ . Quando o sinal foi editado, tem-se a ocorrência de picos positivos e/ou negativos como resultado de mudanças de fase positivas e/ou negativas. Quanto maior é a amplitude do pico, mais abrupta é a mudança de fase. Adicionalmente, estes picos podem ajudar na localização aproximada dos pontos onde o sinal foi editado.

Definimos o valor médio de  $\hat{\phi}'(n_b)$  por:

$$m_{\hat{\phi}'} = \frac{1}{N_{\text{Block}} - 1} \sum_{n_b=2}^{N_{\text{Block}}} \hat{\phi}'(n_b). \quad (4.4)$$

Posteriormente, define-se a medida característica  $F$  para avaliar a edição de áudio com base na fase da ENF como:

$$F = 100 \log \left\{ \frac{1}{N_{\text{Block}} - 1} \sum_{n_b=2}^{N_{\text{Block}}} \left[ \hat{\phi}'(n_b) - m_{\hat{\phi}'} \right]^2 \right\} \quad (4.5)$$

A finalidade de subtrair o valor médio de  $\hat{\phi}'(n_b)$  é não considerar o valor médio da inclinação da curva de estimação de fase da ENF, provocada pelas variações (mais lentas) da ENF em sinais não editados. O logaritmo é utilizado somente para facilitar a visualização de  $F$  na decisão entre os sinais editados e os originais.

Para o processo de detecção, definimos as hipóteses  $H \in \{H_O, H_E\}$ , em que  $H_O$  e  $H_E$  representam as hipóteses de que o sinal de áudio digital seja original e editado, respectivamente (VAN TREES, 1968). Posteriormente, para o procedimento de decisão, determinamos a relação de decisão dada por

$$F \underset{H_O}{\overset{H_E}{\gtrless}} \gamma, \quad (4.6)$$

onde  $\gamma$  corresponde ao limiar para a decisão final,  $\hat{H}$ . Com valores de  $F$  superiores a  $\gamma$ , decide-se que o sinal de áudio digital foi editado. Subsequentemente, definimos  $P_D$  como a probabilidade de detecção dos sinais editados (quando se decide que são sinais editados e de fato o são),  $P_F$  como probabilidade de falso alarme na procura de sinais

editados (quando se decide que são sinais editados mas de fato são originais) e  $P_M$  como probabilidade de perda para sinais editados (quando se decide que são sinais originais mas de fato são editados).  $P_D$ ,  $P_F$  e  $P_M$  são determinadas por:

$$P_D = P(\hat{H} = H_E | H_E) = P(F > \gamma | H_E), \quad (4.7)$$

$$P_F = P(\hat{H} = H_E | H_O) = P(F > \gamma | H_O), \quad (4.8)$$

$$P_M = P(\hat{H} = H_O | H_E) = P(F < \gamma | H_E). \quad (4.9)$$

Adicionalmente, tem-se que:

$$P_D = 1 - P_M. \quad (4.10)$$

Buscamos na detecção, achar o valor de  $\gamma$  que resulte no maior valor de  $P_D$  (diminuindo o valor de  $P_F$ ). Para estabelecer este limiar na prática, é necessária a preparação de um banco de dados com sinais de áudio digital editados para serem comparados com os originais. Realiza-se a avaliação do método sobre esse banco de dados variando-se  $\gamma$  e construindo-se uma curva de  $P_M$  em função de  $P_F$ , conhecida como curva DET (*Detection Error Tradeoff*) (MARTIN, 1997). Nesta curva, o ponto em que  $P_M = P_F$  é conhecido como de EER (*Equal Error Rate*). O valor de  $\gamma$  correspondente à EER é usado como o ponto de operação quando consideramos que ambos os erros têm igual importância. Mesmo que numa dada aplicação o ponto de operação seja escolhido diferente do EER, seu valor dá uma idéia do desempenho global do método.

Na Seção 5.3.3 é feita uma avaliação deste mecanismo automático de detecção de edição de áudio para três bancos de dados que possuem sinais de áudio com as suas correspondentes versões editadas (recortes ou inserções de fragmento no sinal).

#### 4.4 RESUMO

Neste capítulo apresentou-se um método de autenticação de áudio com base na mudança de fase da ENF, que se encontra embutida na maioria das gravações. Primeiramente foi apresentado um mecanismo para a detecção automática da ENF. Subsequentemente, um mecanismo que fornece informações visuais para detecção de sinais editados. Estas informações baseiam-se na verificação das mudanças de fase para localização dos pontos de edição. Além disso, o cálculo de uma medida característica (*feature*) definida sobre os parâmetros estimados pelo método possibilita que, mediante uma relação de decisão, se discrimine de forma automática entre sinais originais e editados.

Todo o procedimento de autenticação de áudio exposto tem como fundamento básico experimentos realizados sobre tons artificiais, o que significa condições ideais. No capítulo seguinte será realizada a avaliação do método em gravações reais que foram digitalmente editadas.

## 5 AVALIAÇÃO DA TÉCNICA PROPOSTA

### 5.1 INTRODUÇÃO

Para avaliar o método proposto, autenticação de áudio com base na mudança de fase da ENF, aplicou-se a técnica em gravações digitalmente editadas. Para obter um *corpus* razoável, para posterior avaliação da técnica proposta, procedeu-se a editar sinais provenientes de três bases (*corpora*) em forma digital. Na seção seguinte explica-se como foram criadas essas bases.

### 5.2 CRIAÇÃO DAS BASES DE DADOS

Para avaliar a técnica criaram-se três bases de dados, os quais serão denominadas (para posterior referência textual) como: BASE ESPANHOLA, BASE CARIOCA 1 e BASE CARIOCA 2.

#### 5.2.1 BASE ESPANHOLA

Os sinais originais deste *corpus* são provenientes de duas bases públicas em castelhano, AHUMADA e GAUDI, obtidas via <http://atvs.ii.uam.es/databases.jsp> (ORTEGA-GARCIA, 2000). Todos os sinais utilizados desta base podem ser considerados livres de saturações, apresentam baixo ruído de fundo e possuem a componente da ENF no espectro. Por ser a base proveniente da Espanha, a ENF nominal dos sinais é de 50 Hz (característica dos países europeus). Os locutores utilizados foram 25 homens e 25 mulheres. Cada locutor produziu 2 sinais, cada sinal tendo duração entre 15 e 30 segundos. Dispõe-se, portanto, de 100 sinais de áudio.

Posteriormente, dois sinais por locutor são editados: do primeiro sinal é eliminado um fragmento, provocando uma mudança de fase (um ponto de edição); no segundo sinal é inserido um fragmento provocando duas mudanças de fase (dois pontos de edição). Esta inserção foi feita com um fragmento proveniente do mesmo sinal, evitando-se mudanças de espectro do tempo curto produzidas por diferenças na frequência de amostragem, que tornariam a detecção mais fácil. Deve-se observar que todas as edições foram feitas sem que se tomasse qualquer cuidado com as modificações da fase da ENF nos sinais, tentando-

se emular o que faria uma pessoa que desconhecesse o efeito da ENF na edição de áudio. O histograma da distribuição das mudanças de fase dos sinais editados da BASE ESPANHOLA é apresentado na FIG. 5.1(a). Ele indica uma distribuição aproximadamente uniforme entre  $-180^\circ$  e  $+180^\circ$ . Ao final, dispõe-se de um banco de dados com 100 sinais originais e 100 sinais editados.

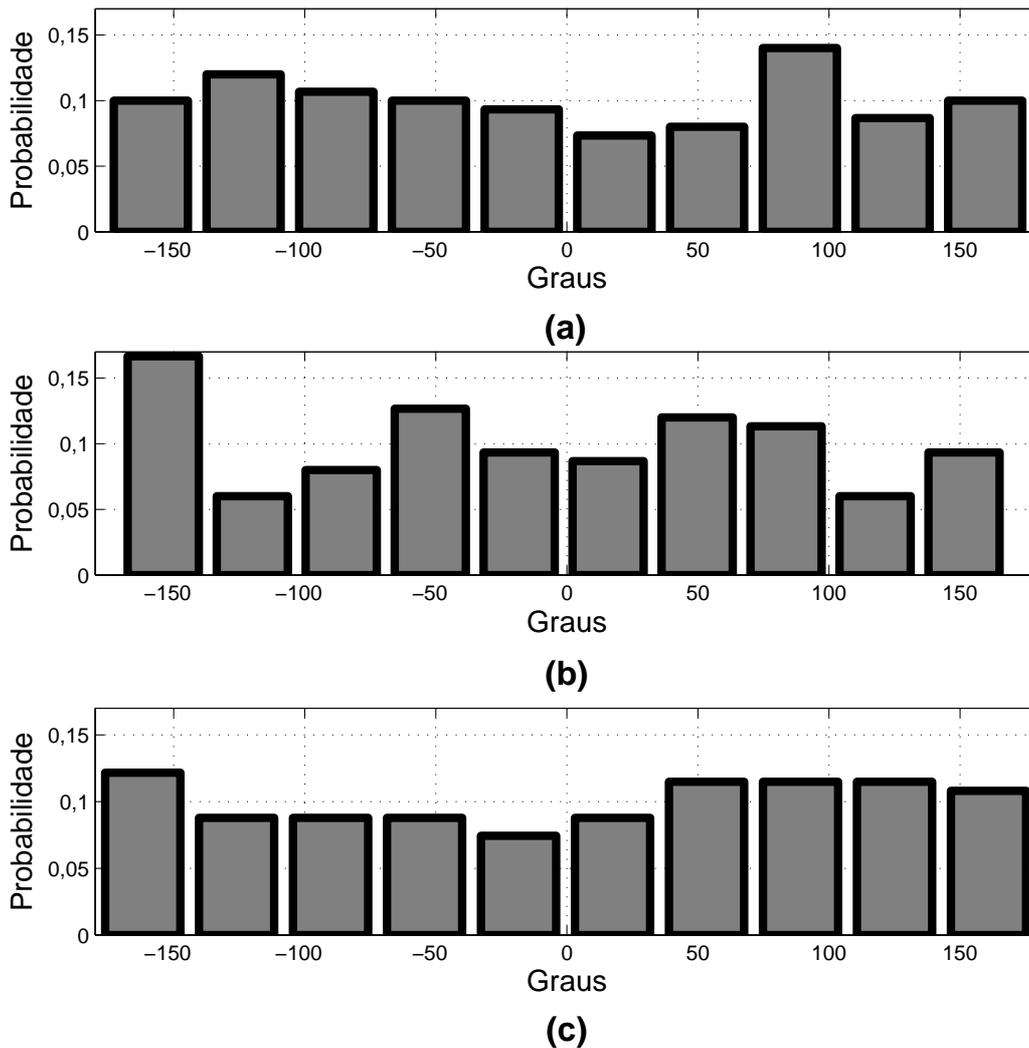


FIG. 5.1: Histograma normalizado da distribuição das mudanças de fase nos sinais editados. (a) BASE ESPANHOLA; (b) BASE CARIOCA 1; (c) BASE CARIOCA 2.

### 5.2.2 BASE CARIOCA 1

Os sinais originais desta base foram provenientes de duas ligações telefônicas feitas no Rio de Janeiro, RJ. A ENF nominal dos sinais é de 60 Hz (característica do Brasil). Foram usados dois locutores (um homem e uma mulher) neste banco. Cada locutor gravou aproximadamente 1 hora de sinal de voz numa ligação local dentro da cidade. De cada gravação se extraíram 50 sinais entre 25 e 30 segundos. Dispõe-se, portanto, de 100 sinais de áudio. Todos os sinais utilizados da base estão livres de saturações, apresentam baixo ruído de fundo e possuem a componente da ENF no espectro.

Posteriormente, os 50 sinais por locutor foram editados. Nos primeiros 25 sinais, foi eliminado um fragmento, provocando uma mudança de fase (um ponto de edição). Nos outros 25 sinais, foi inserido um fragmento provocando duas mudanças de fase (dois pontos de edição). Como no caso anterior, esta inserção foi feita com um fragmento proveniente do mesmo sinal, e todas as edições foram feitas sem que se tomasse qualquer cuidado com as mudanças da fase da ENF nos sinais. O histograma da distribuição das mudanças de fase dos sinais editados desta base é apresentado na FIG. 5.1(b) que, novamente, indica uma distribuição aproximadamente uniforme entre  $-180^\circ$  e  $+180^\circ$ . Ao final, dispõe-se de um banco de dados com 100 sinais originais e 100 sinais editados.

### 5.2.3 BASE CARIOCA 2

Os sinais originais desta base foram também provenientes de ligações telefônicas feitas no Rio de Janeiro, RJ. A ENF nominal dos sinais é, portanto, de 60 Hz. Os locutores utilizados foram 50 homens e 50 mulheres. Cada locutor produziu um sinal de duração entre 15 e 30 segundos. Dispõe-se, portanto, de 100 sinais de áudio. Além disso, há sinais desta base que apresentam saturações e/ou ruído de fundo elevado. A base, assim, descreve situações onde as gravações foram realizadas em condições desfavoráveis.

As edições nesta base foram feitas da mesma maneira que na base em castelhano: em 50 sinais (pertencentes a 25 homens e 25 mulheres) foi feito um corte de fragmento por frase, e nos 50 sinais restantes uma inserção do fragmento. O histograma da distribuição das mudanças de fase dos sinais editados desta base é apresentado na FIG. 5.1(c), que mais uma vez indica uma distribuição aproximadamente uniforme entre  $-180^\circ$  e  $+180^\circ$ . Ao final, dispõe-se de um banco de dados com 100 sinais originais e 100 sinais editados.

### 5.3 RESULTADOS EXPERIMENTAIS

Nesta seção, serão apresentados resultados da técnica de autenticação de áudio detalhada no Capítulo 4, todos com base nos sinais dos três bancos de dados mencionados na Seção 5.2. Primeiramente, serão apresentados os resultados para a detecção automática da ENF (Seção 4.3.1); posteriormente, alguns exemplos do mecanismo visual (Seção 4.3.2) para edição de áudio; e, finalmente, uma avaliação do mecanismo automático (Seção 4.3.3) de detecção.

#### 5.3.1 OBTENÇÃO DE LIMIARES PARA A DETECÇÃO AUTOMÁTICA DA ENF

Para a detecção automática da ENF, é necessário, como foi explicado na Seção 4.3.1, determinar do valor do limiar  $\gamma_{\text{ENF}}$  (na relação de decisão determinada pela EQ. 4.2).

Para o caso quando o valor da ENF é de 50 Hz, utilizamos os sinais da BASE ESPANHOLA como o grupo de sinais que a possuía e utilizamos os sinais da BASE CARIOCA 2 como o grupo de sinais que não a possuía. Na FIG. 5.2(a), são apresentados os histogramas dos valores da relação  $R_{\text{ENF}}$  para sinais com a ENF de 50 Hz e para sinais sem a ENF de 50 Hz. Adicionalmente, pode-se observar a localização do limiar de decisão  $\gamma_{\text{ENF}} = -5,61$ . Este limiar encontra-se numa região entre os histogramas, que não possuem superposição.

Para o caso quando o valor da ENF é de 60 Hz, utilizamos os sinais da BASE CARIOCA 2 como o grupo de sinais que a possuíam e utilizamos os sinais da BASE ESPANHOLA como o grupo de sinais que não a possuíam. Na FIG. 5.2(b), são apresentados os histogramas dos valores da relação  $R_{\text{ENF}}$  para sinais com a ENF de 60 Hz e para sinais sem a ENF de 60 Hz. Adicionalmente, pode-se observar a localização do limiar de decisão  $\gamma_{\text{ENF}} = -5,88$ . Como no caso anterior, este limiar encontra-se numa região entre os histogramas, que não possuem superposição.

Tanto para o caso da ENF de 50 quanto de 60 Hz, conseguem-se valores do limiar de decisão que permitem discriminar sinais que a possuem com os sinais que não a possuem. Na prática, o valor de  $\gamma_{\text{ENF}}$  pode ser o mesmo para os dois casos.

#### 5.3.2 EXEMPLOS DO MECANISMO VISUAL

Apresentamos a seguir quatro exemplos do mecanismo visual para detecção de edição de áudio como proposto na Seção 4.3.2. Nos dois primeiros exemplos, são considerados sinais da BASE ESPANHOLA; para o terceiro e quarto exemplos, são considerados sinais

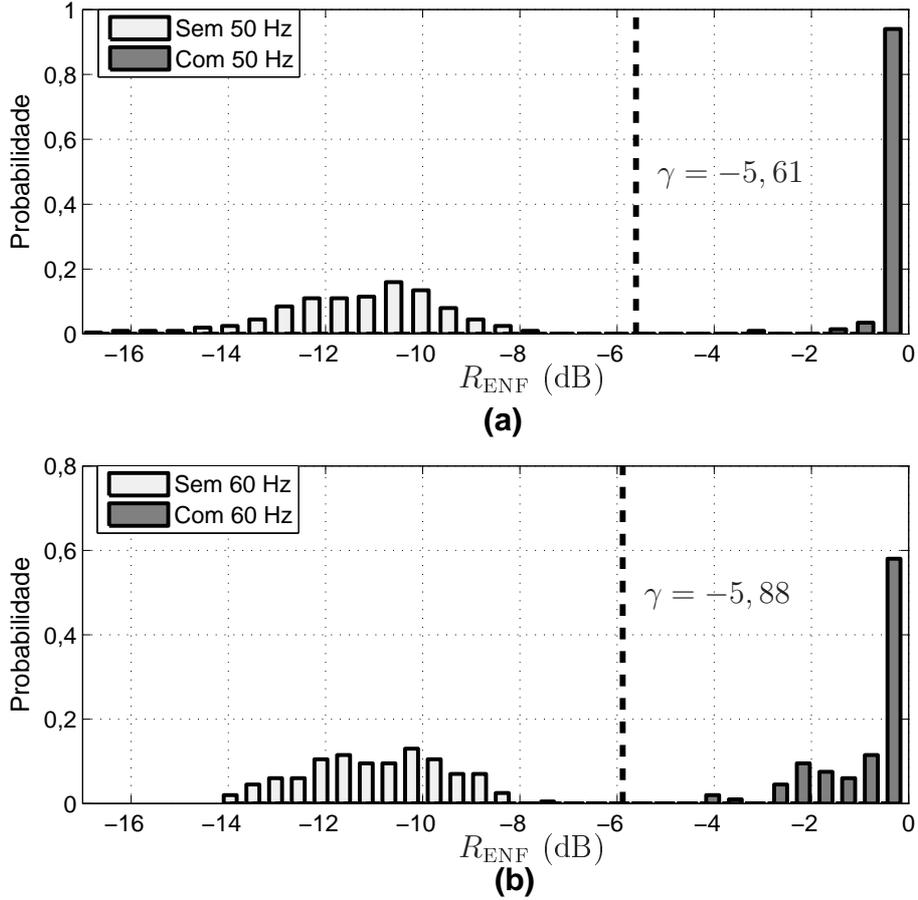


FIG. 5.2: Histograma normalizado da relação  $R_{\text{ENF}}$ . (a) Para 50,0 Hz: sinais com ENF da BASE ESPANHOLA e sinais sem ENF da BASE CARIOCA 2. (b) Para 60,0 Hz: sinais com ENF da BASE CARIOCA 2 e os sinais sem ENF da BASE ESPANHOLA.

da BASE CARIOCA 1 e da BASE CARIOCA 2, respectivamente.

A FIG. 5.3 apresenta um caso de eliminação do fragmento do sinal. É importante considerar que o sinal original possui uma fase estimada com comportamento retilíneo e sem mudanças abruptas; e, que o sinal editado apresenta uma mudança abrupta de fase no ponto de edição P1 (temos, neste caso, uma mudança de fase positiva). As variáveis usadas neste caso foram: frequência de sub-amostragem  $f_{\text{downsample}} = 1000$  Hz; valor nominal da ENF de 50,0 Hz; largura de banda do filtro de 0,8 Hz; tamanho da janela de  $n_1 = 10$  ciclos da ENF com superposição de  $n_2 = 9$  ciclos; e,  $N_{\text{DFT}}$  com 2000 pontos tanto para o método de estimação de fase da DFT como da DFT<sup>1</sup>.

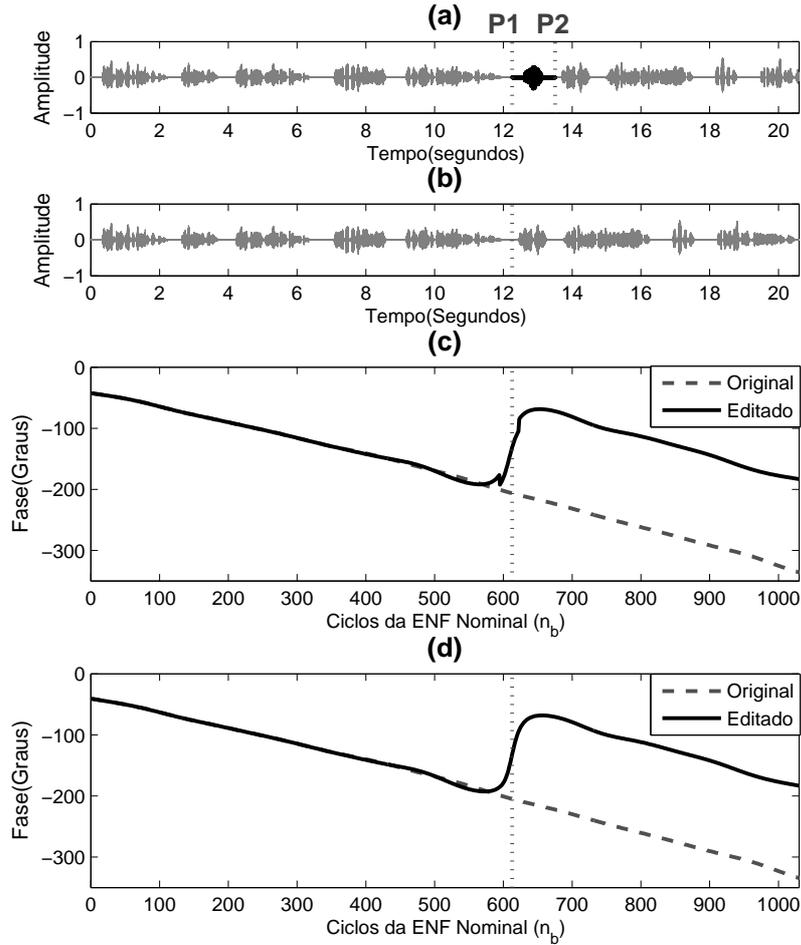


FIG. 5.3: Visualização de uma eliminação de fragmento (Sinal da BASE ESPANHOLA): P1 e P2 são os limites do fragmento do sinal original (P1 é o ponto de edição do sinal editado). O valor nominal de ENF é de 50,0 Hz e a largura de banda do filtro é de 0,8 Hz. O tamanho da janela é de 10 ciclos da ENF com superposição de 9 ciclos.  $N_{DFT} = 2000$  pontos. (a) Sinal Original; (b) Sinal Editado; (c) Estimação de fase usando DFT; (d) Estimação de fase usando DFT<sup>1</sup>.

A FIG. 5.4 apresenta um caso de inserção do fragmento do sinal. Neste caso, existem dois pontos de edição: P1 e P2. Consequentemente, duas mudanças de fase podem ser observadas no caso do sinal editado. A primeira mudança é negativa e a segunda, positiva. Os parâmetros utilizados neste caso foram as mesmas utilizadas para o caso da FIG. 5.3.

A FIG. 5.5 apresenta um outro caso de eliminação de fragmento do sinal. Neste caso, o comportamento da fase estimada da ENF apresenta leves variações. Estas são provocadas pela própria variação na frequência do sinal da rede elétrica, ou seja, o sinal foi gravado

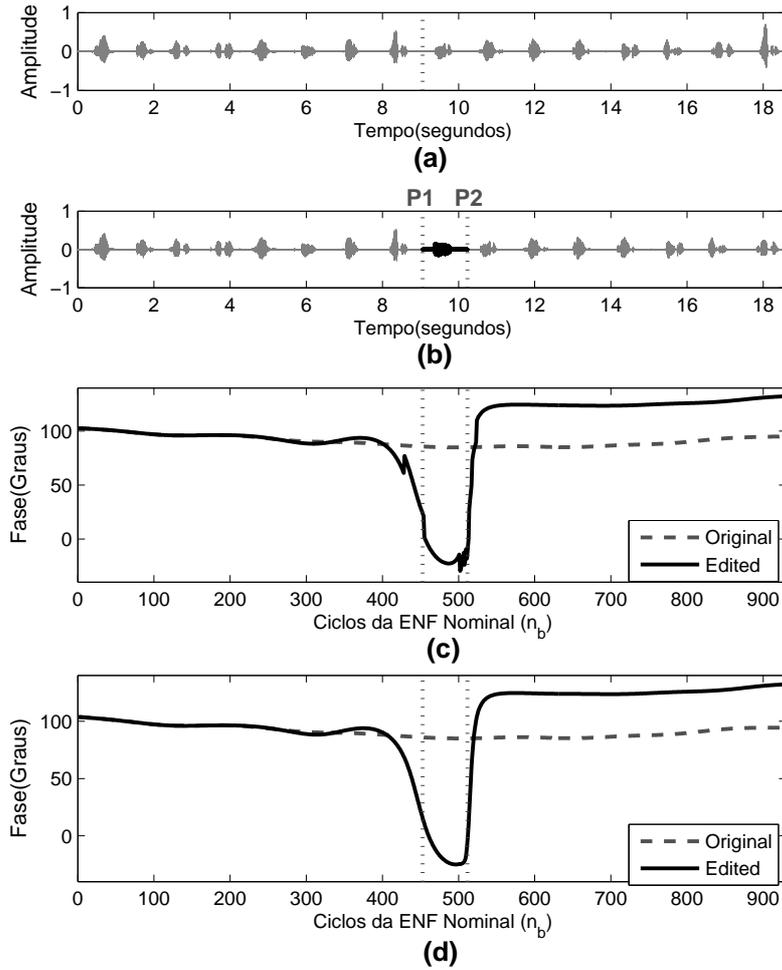


FIG. 5.4: Visualização de uma inserção de fragmento (Sinal da BASE ESPANHOLA): P1 representa o ponto de inserção do sinal original (P1 e P2 são os pontos de edição do sinal editado). O valor nominal de ENF é de 50,0 Hz e a largura de banda do filtro é de 0,8 Hz. O tamanho da janela é de 10 ciclos da ENF com superposição de 9 ciclos.  $N_{\text{DFT}} = 2000$  pontos. (a) Sinal Original; (b) Sinal Editado; (c) Estimação de fase usando DFT; (d) Estimação de fase usando DFT<sup>1</sup>.

num instante quando este fenômeno estava ocorrendo (considerando que este sinal não é sintético). Apesar disso, com a mudança de fase provocada pelo corte ou eliminação do fragmento de sinal, ocorre no sinal editado uma variação mais significativa, e com isso pode-se determinar o ponto de edição do sinal. Os parâmetros usadas neste caso foram: frequência de sub-amostragem  $f_{\text{downsample}} = 1200$  Hz; valor nominal da ENF de 60,0 Hz; e os parâmetros restantes iguais às que foram usadas para os casos da FIG. 5.3 e da

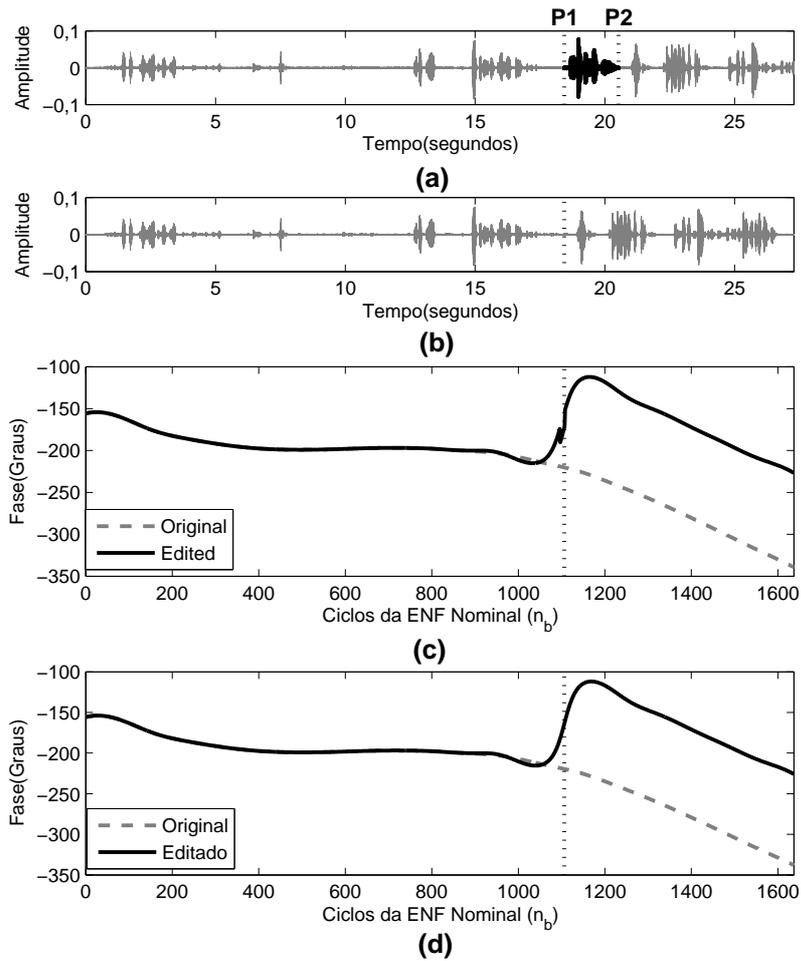


FIG. 5.5: Visualização de uma eliminação de fragmento (Sinal da BASE CARICA 1): P1 e P2 são limites do fragmento do sinal original (P1 é o ponto de edição do sinal editado). O valor nominal de ENF é de 60,0 Hz e a largura de banda do filtro é de 0,8 Hz. O tamanho da janela é de 10 ciclos da ENF com superposição de 9 ciclos.  $N_{DFT} = 2000$  pontos. (a) Sinal Original; (b) Sinal Editado; (c) Estimação de fase usando DFT; (d) Estimação de fase usando  $DFT^1$ .

FIG. 5.4.

Nos casos apresentados até o momento, podemos observar uma melhor resolução nas estimativas de fase quando é utilizado o método da  $DFT^1$ , comparado com o método da DFT (especialmente nas regiões onde ocorrem as transições de fase).

Finalmente, a FIG. 5.6 apresenta um outro caso de inserção de fragmento do sinal. O sinal usado neste exemplo possui não linearidades devidas à ocorrência de saturações. Como consequência destas saturações, existem variações significativas na fase estimada

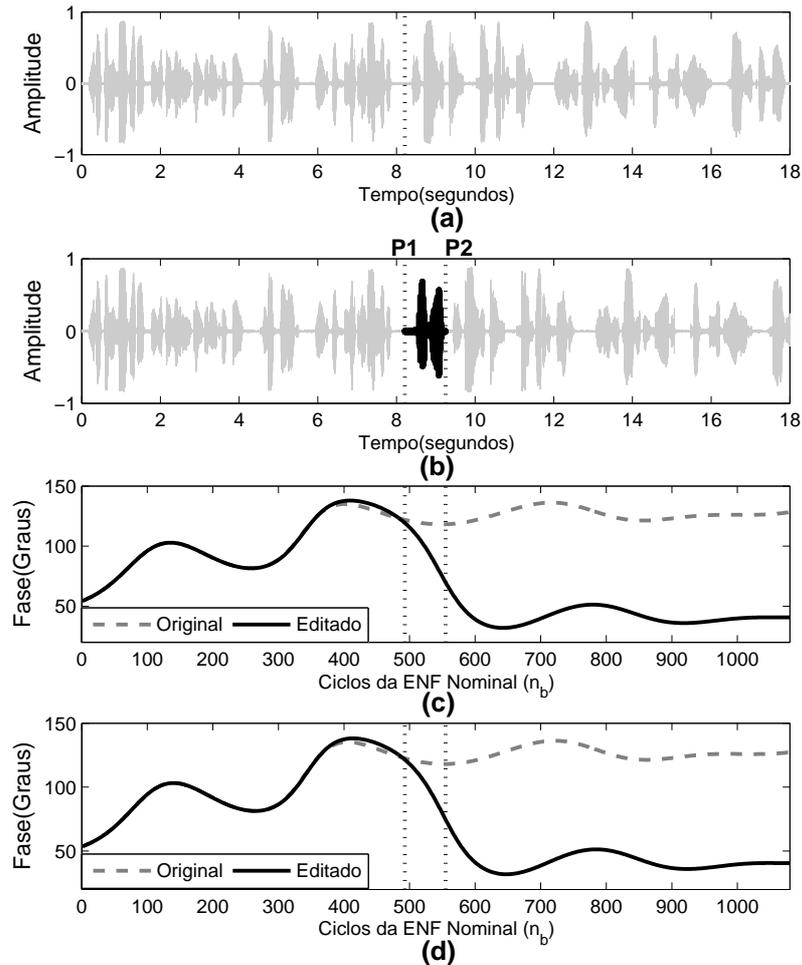


FIG. 5.6: Visualização de uma inserção de fragmento (Sinal da BASE CARIOCA 2): P1 é o ponto de inserção do sinal original (P1 e P2 são os pontos de edição do sinal editado). O valor nominal de ENF é de 60,0 Hz e a largura de banda do filtro é de 0,6 Hz. O tamanho da janela foi de 10 ciclos da ENF com superposição de 9 ciclos.  $N_{\text{DFT}} = 2000$  pontos. (a) Sinal Original; (b) Sinal Editado; (c) Estimação de fase usando DFT; (d) Estimação de fase usando DFT<sup>1</sup>.

da ENF no sinal original. Apesar da edição ter, como nos outros exemplos, provocado mudanças de fase, neste caso torna-se difícil determinar os pontos de edição. Alguns pontos em que ocorre mudança de fase não provocada por edição podem ser confundidos com pontos de edição. Os parâmetros usadas neste caso foram similares ao caso da FIG. 5.5, mas com um filtro passa-banda com largura de faixa de 0,6 Hz.

Adicionalmente, este método visual pode auxiliar a determinar o tipo de edição: recorte (uma única mudança abrupta de fase) e inserção (duas mudanças abruptas de

fase).

### 5.3.3 AVALIAÇÃO DO MECANISMO AUTOMÁTICO

Na Seção 4.3.3 foi apresentado um mecanismo para discriminação automática entre sinais originais e editados mediante a relação de decisão expressa na EQ. 4.6. Sob essa premissa, vamos definir os parâmetros usadas nos bancos de dados para a avaliação do mecanismo automático para detecção de edição.

Para a BASE ESPANHOLA:

- A frequência de amostragem adotada foi  $f_{\text{downsample}} = 1000$  Hz.
- A largura de faixa do filtro passa-banda centrado no valor nominal da ENF (50 Hz) é de 0,8 Hz.
- O sinal filtrado foi segmentado em blocos de durações  $n_1$  iguais a 3, 5 ou 10 ciclos da ENF nominal com sobreposição de  $n_2 = n_1 - 1$  ciclos com respeito ao bloco seguinte.
- Números de pontos da DFT,  $N_{\text{DFT}}$ , iguais a 200, 2000 ou 20000 para os dois métodos de estimação de fase.

Para a BASE CARIOCA 1:

- A frequência de amostragem adotada foi  $f_{\text{downsample}} = 1200$  Hz.
- A largura de faixa do filtro passa-banda centrado no valor nominal da ENF (60 Hz) foi de 0,8 Hz.
- Os parâmetros restantes foram as mesmas usadas para o caso da BASE ESPANHOLA.

Para a BASE CARIOCA 2, os parâmetros usadas foram similares ao caso da BASE CARIOCA 1, mas com um filtro passa-banda com largura de faixa de 0,6 Hz.

A TAB. 5.1 apresenta um resumo dos resultados obtidos quando o método de discriminação automática é utilizado nos sinais da BASE ESPANHOLA de acordo com a relação de decisão expressa na EQ. 4.6. Na TAB. 5.1, podemos observar que tanto os resultados relativos ao método de estimação de fase pela DFT quanto pela DFT<sup>1</sup> apresentam razoáveis valores de erro (neste caso considerado como o ponto de EER na curva DET)

TAB. 5.1: Avaliação do mecanismo automático de autenticação de áudio para BASE ESPANHOLA (100 originais e 100 editados);  $n_1$  representa o número de ciclos da ENF nominal (tamanho da janela). ENF nominal de 50,0 Hz.  $n_2$  considera  $(n_1 - 1)$  ciclos da ENF nominal como superposição entre os blocos para todos os casos. EER é o valor correspondente a  $P_M = P_F$  (Equal Error Rate).

<b>Método</b>	$n_1$	$N_{\text{DFT}}$	<b>EER(%)</b>
DFT	3 (60 amostras)	200	6%
		2000	6%
		20000	7%
	5 (100 amostras)	200	6%
		2000	6%
		20000	8%
	10 (200 amostras)	200	6%
		2000	6%
		20000	9%
DFT <sup>1</sup>	3 (60 amostras)	200	6%
		2000	6%
		20000	6%
	5 (100 amostras)	200	6%
		2000	6%
		20000	6%
	10 (200 amostras)	200	8%
		2000	6%
		20000	6%

para os diferentes valores de tamanho da janela ( $n_1$ ) e número de pontos da DFT ( $N_{\text{DFT}}$ ). Estes valores de erro variam entre 6% e 9%. É importante destacar que todos os sinais do banco de dados utilizados nesta avaliação apresentam condições favoráveis (baixo ruído de fundo e ausência de saturações).

A TAB. 5.2 apresenta um resumo dos resultados obtidos quando o método de discriminação automática é utilizado nos sinais da BASE CARIOCA 1 de acordo com a relação de decisão expressa na EQ. 4.6. Como no caso anterior, os erros obtidos na avaliação do método apresentam valores moderados, entre 7% e 8%, sobre sinais com SNR alta e ausência de saturações.

A TAB. 5.3 apresenta um resumo dos resultados obtidos quando o método de discriminação automática é utilizado nos sinais da BASE CARIOCA 2. Os erros obtidos nesta avaliação variam entre 15% e 16% para os diferentes valores de tamanho da janela

TAB. 5.2: Avaliação do mecanismo automático de autenticação de áudio para BASE CARIOCA 1 (100 originais e 100 editados);  $n_1$  representa o número de ciclos da ENF nominal (tamanho da janela). ENF nominal de 60,0 Hz.  $n_2$  considera  $(n_1 - 1)$  ciclos da ENF nominal como superposição entre os blocos para todos os casos. EER é o valor correspondente a  $P_M = P_F$  (Equal Error Rate).

Método	$n_1$	$N_{\text{DFT}}$	EER(%)
DFT	3 (60 amostras)	200	7%
		2000	7%
		20000	7%
	5 (100 amostras)	200	7%
		2000	7%
		20000	8%
	10 (200 amostras)	200	7%
		2000	7%
		20000	7%
DFT <sup>1</sup>	3 (60 amostras)	200	8%
		2000	8%
		20000	8%
	5 (100 amostras)	200	7%
		2000	7%
		20000	7%
	10 (200 amostras)	200	8%
		2000	7%
		20000	7%

( $n_1$ ) e número de pontos da DFT ( $N_{\text{DFT}}$ ). Neste caso a presença de sinais com condições desfavoráveis (baixo ruído de fundo e saturações no conteúdo do sinal) aumentaram o erro na detecção automática.

Considerando os dados utilizados nos experimentos correspondentes a cada uma das TABs. 5.1, 5.2 e 5.3, o caso particular em que utilizamos o método de estimação de fase da DFT<sup>1</sup> com  $N_{\text{DFT}} = 2000$  e com blocos de duração de  $n_1 = 10$  ciclos da ENF nominal com superposição de  $n_2 = 9$  ciclos (para os quais o erro obtido na discriminação automática foi menor), que inclui os resultados das FIGs. 5.3, 5.4, 5.5 e 5.6, apresentamos na FIG. 5.7 os gráficos dos histogramas da medida característica (*feature*)  $F$ , assim como a localização do limiar de decisão  $\gamma$  para as três bases. Idealmente, nestes histogramas deve ser possível verificar a separação entre as distribuições de  $F$  para o caso dos sinais originais e editados. Claramente, no caso da BASE ESPANHOLA (FIG. 5.7(a)) e no

TAB. 5.3: Avaliação do mecanismo automático de autenticação de áudio para BASE CARIOCA 2 (100 originais e 100 editados);  $n_1$  representa o número de ciclos da ENF nominal (tamanho da janela). ENF nominal de 60,0 Hz.  $n_2$  considera  $(n_1 - 1)$  ciclos da ENF nominal como superposição entre os blocos para todos os casos. EER é o valor correspondente a  $P_M = P_F$  (Equal Error Rate).

Método	$n_1$	$N_{\text{DFT}}$	EER(%)
DFT	3 (60 amostras)	200	15%
		2000	15%
		20000	16%
	5 (100 amostras)	200	15%
		2000	15%
		20000	16%
	10 (200 amostras)	200	15%
		2000	15%
		20000	15%
DFT <sup>1</sup>	3 (60 amostras)	200	15%
		2000	15%
		20000	15%
	5 (100 amostras)	200	15%
		2000	15%
		20000	15%
	10 (200 amostras)	200	16%
		2000	15%
		20000	15%

caso da BASE CARIOCA 1 (FIG. 5.7(b)), pode-se verificar uma separação mais clara entre as distribuições mencionadas. Para o caso da BASE CARIOCA 2 (FIG. 5.7(c)), os histogramas de  $F$  para sinais originais e editados estão mais superpostos devido à presença de sinais desfavoráveis, corrompidos com saturações e ruído, que dificultam a detecção.

O erro de detecção é determinado pelo ponto correspondente à EER (*Equal Error Rate*), ou seja, o valor correspondente a  $P_F = P_M$  na curva DET (*Detection Error Trade-off*). As curvas DET ( $P_M$  em função de  $P_F$ ), assim como a localização do ponto EER para os três bancos de dados considerados, são apresentadas na FIG. 5.8. Os valores dos erros obtidos foram de 6%, 7% e 15% para BASE ESPANHOLA, BASE CARIOCA 1 e BASE CARIOCA 2, respectivamente. Os valores do limiar  $\gamma$  obtidos para o processo de decisão automática determinado pela relação de decisão dada na EQ. 4.6 foram de -201, -188 e -165,38 para BASE ESPANHOLA, BASE CARIOCA 1 e BASE CARIOCA 2,

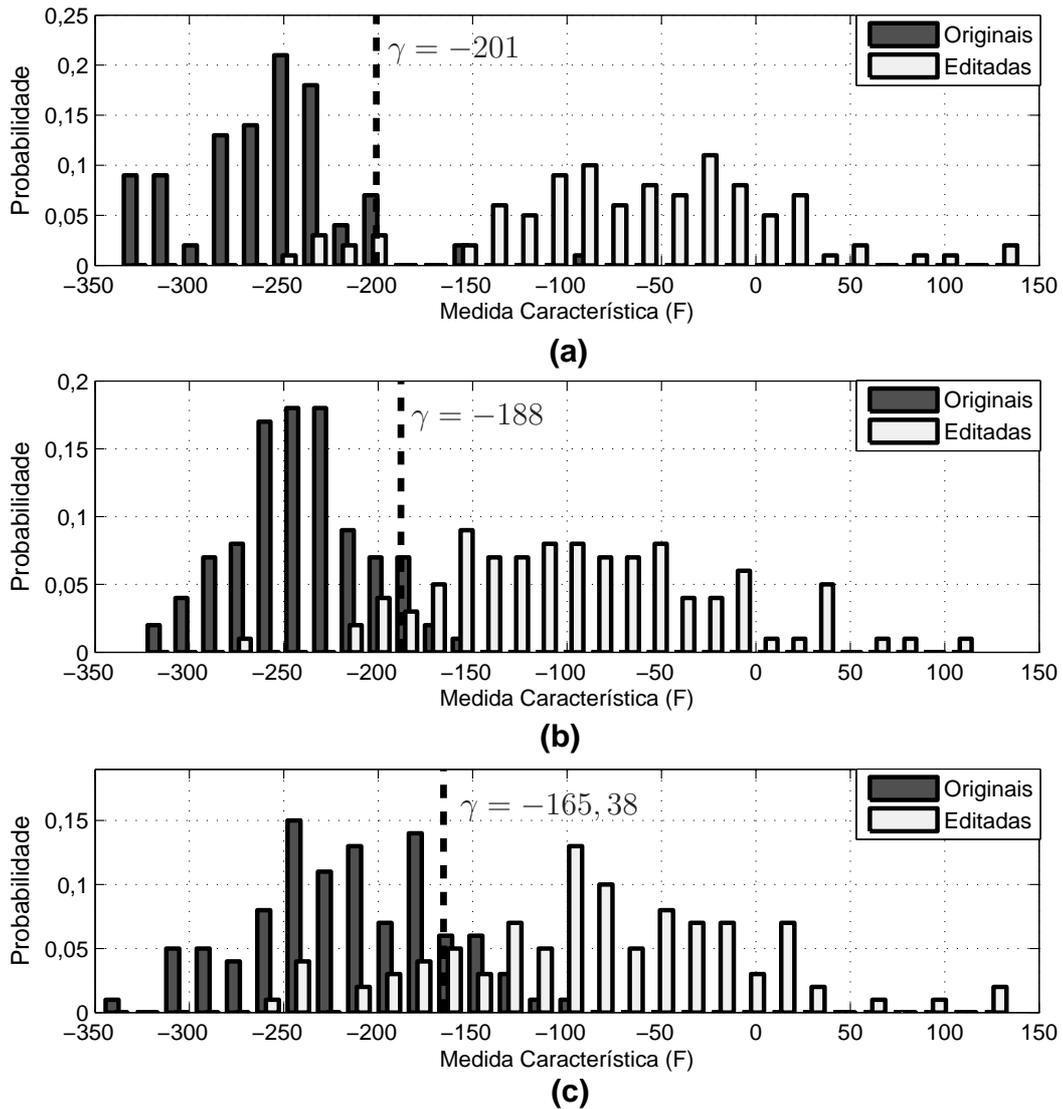


FIG. 5.7: Histograma normalizado da distribuição da medida característica,  $F$ . Foi usado o método de estimação de fase da DFT<sup>1</sup> com  $N_{\text{DFT}} = 2000$  e com blocos de duração de  $n_1 = 10$  ciclos da ENF nominal com superposição de  $n_2 = 9$  ciclos. (a) BASE ESPANHOLA (filtro passa-banda centrado em 50 Hz com largura de faixa de 0,8 Hz); (b) BASE CARIOCA 1 (filtro passa-banda centrado em 60,0 Hz com largura de faixa de 0,8 Hz); (c) BASE CARIOCA 2 (filtro passa-banda centrado em 60,0 Hz com largura de faixa de 0,6 Hz).

respectivamente.

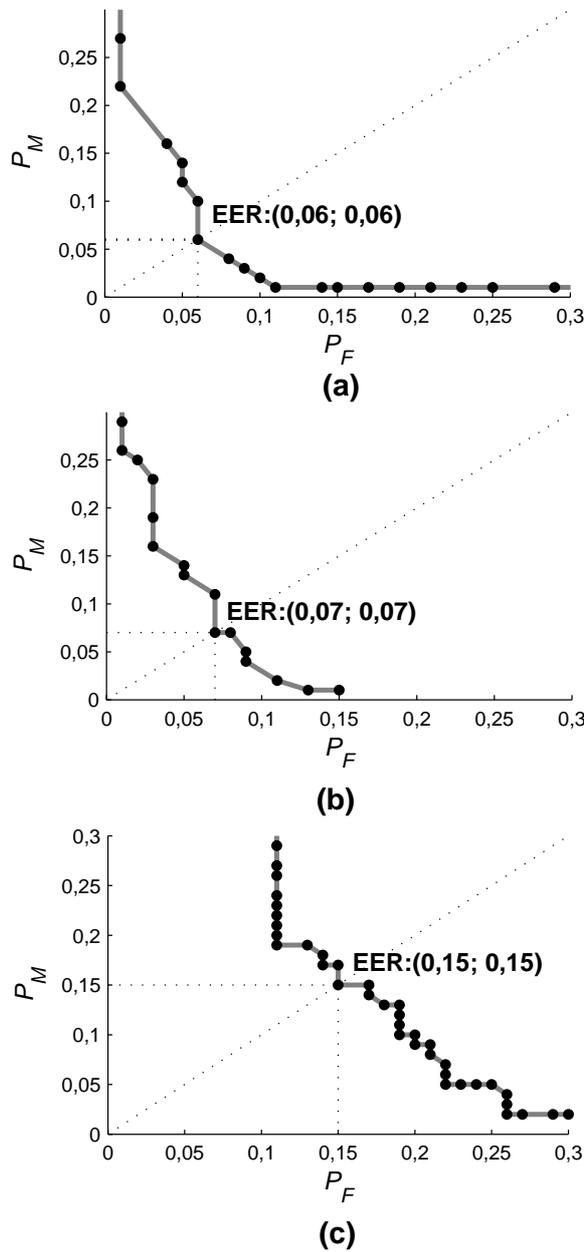


FIG. 5.8: Curva DET:  $P_M$  em função de  $P_F$ . O ponto indicado corresponde à *Equal Error Rate* (EER), ou seja, ao limiar em que  $P_F = P_M$ . Foi usado o método de estimação de fase da DFT<sup>1</sup> com  $N_{\text{DFT}} = 2000$  e com blocos de duração de  $n_1 = 10$  ciclos da ENF nominal com superposição de  $n_2 = 9$  ciclos. (a) BASE ESPANHOLA (filtro passa-banda centrado em 50 Hz com largura de faixa de 0,8 Hz); (b) BASE CARIOCA 1 (filtro passa-banda centrado em 60,0 Hz com largura de faixa de 0,8 Hz); (c) BASE CARIOCA 2 (filtro passa-banda centrado em 60,0 Hz com largura de faixa de 0,6 Hz).

## 5.4 EFEITOS DE DEGRADAÇÕES NA DETECÇÃO

O menores valores de EER obtidos, respectivamente, na avaliação da BASE ESPANHOLA e da BASE CARIOCA 1 estão dentro de limites razoáveis, mas estas bases de dados apresentam condições favoráveis (ruído de fundo reduzido e ausência de saturações). Em situações reais, especialmente em casos periciais, as gravações podem apresentar condições desfavoráveis tais como: saturação, elevado ruído de fundo e variações significativas da própria ENF original (NICOLALDE, 2009c). Estas condições negativas podem ser a causa do erro de 15% obtido na avaliação da BASE CARIOCA 2. Nesta seção, serão analisados estes três efeitos individualmente.

### 5.4.1 EFEITO DE RUÍDO DE FUNDO

Para fazer a análise do efeito do ruído de fundo na detecção, usou-se a BASE ESPANHOLA como referência.

Aplicou-se um algoritmo VAD (*Voice Activity Detector*) (BENYASSINE, 1997) a todos os sinais da base, para separar as regiões de voz ativa das regiões de ruído de fundo. Um exemplo de sua utilização é apresentado na FIG. 5.9.

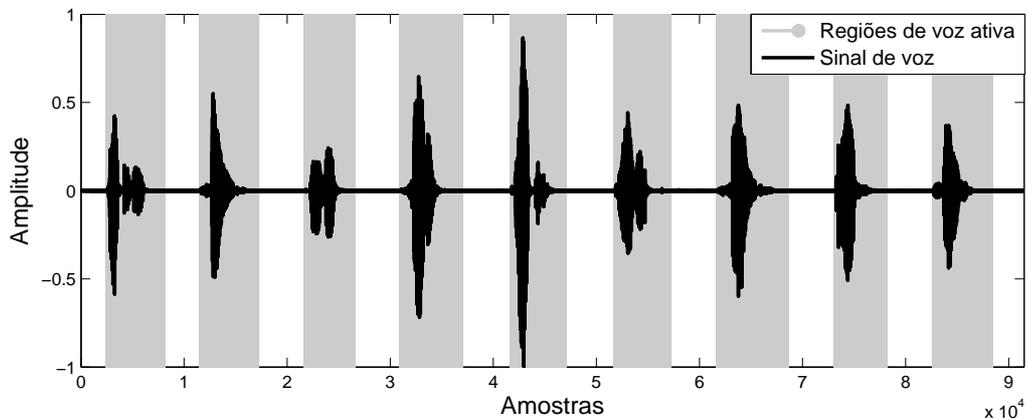


FIG. 5.9: Exemplo de utilização do algoritmo de VAD (BENYASSINE, 1997).

Posteriormente, definimos  $s(n)$  como o sinal de voz puro e  $n_{\text{background}}(n)$  como o ruído de fundo. Os dois sinais são supostos mutuamente decorrelacionados.

A potência média do ruído de fundo (considerada constante) nas regiões em que só

havia ruído é:

$$P_{\text{background}} = E [(n_{\text{background}})^2]. \quad (5.1)$$

A potência média total nas regiões de voz ativa é:

$$P_{\text{voice}} = E [(s + n_{\text{background}})^2] = E [s^2] + E [(n_{\text{background}})^2]. \quad (5.2)$$

A potência do sinal puro é:

$$E [s^2] = P_{\text{voice}} - E [(n_{\text{background}})^2]. \quad (5.3)$$

A razão sinal-ruído original é:

$$\text{SNR}_{\text{original}} = \frac{E [s^2]}{E [(n_{\text{background}})^2]}. \quad (5.4)$$

A idéia é inserir ruído com valor médio zero,  $n_{\text{add}}(n)$ , para modificar a razão sinal-ruído:

$$\text{SNR} = \frac{E [s^2]}{E [(n_{\text{background}})^2] + E [(n_{\text{add}})^2]}. \quad (5.5)$$

Modificando o valor de  $E [(n_{\text{add}})^2]$  e aplicando a técnica de autenticação de áudio proposta (baseada no método de estimação de fase da DFT<sup>1</sup> com os parâmetros usadas nos exemplos das FIGs. 5.3 e 5.4) para a BASE ESPANHOLA, obtemos a curva característica de taxa de erro de detecção em função da SNR.

Foram avaliados três tipos de ruídos, denominados para posterior referência textual como RUÍDO 1, RUÍDO 2 e RUÍDO 3. As características destes ruídos são:

- RUÍDO 1: Ruído branco (isto é, com densidade espectral de potência constante em toda a faixa de frequências) gaussiano.
- RUÍDO 2: Ruído colorido com componentes significativas em frequências baixas. Para obter este ruído, processou-se ruído branco gaussiano com um filtro com função de transferência

$$H(z) = \frac{1}{1 - 0.9z^{-1}}. \quad (5.6)$$

A magnitude da resposta em frequência deste filtro, considerando uma frequência de amostragem de 8000 Hz, é apresentada na FIG. 5.10 em linha contínua.

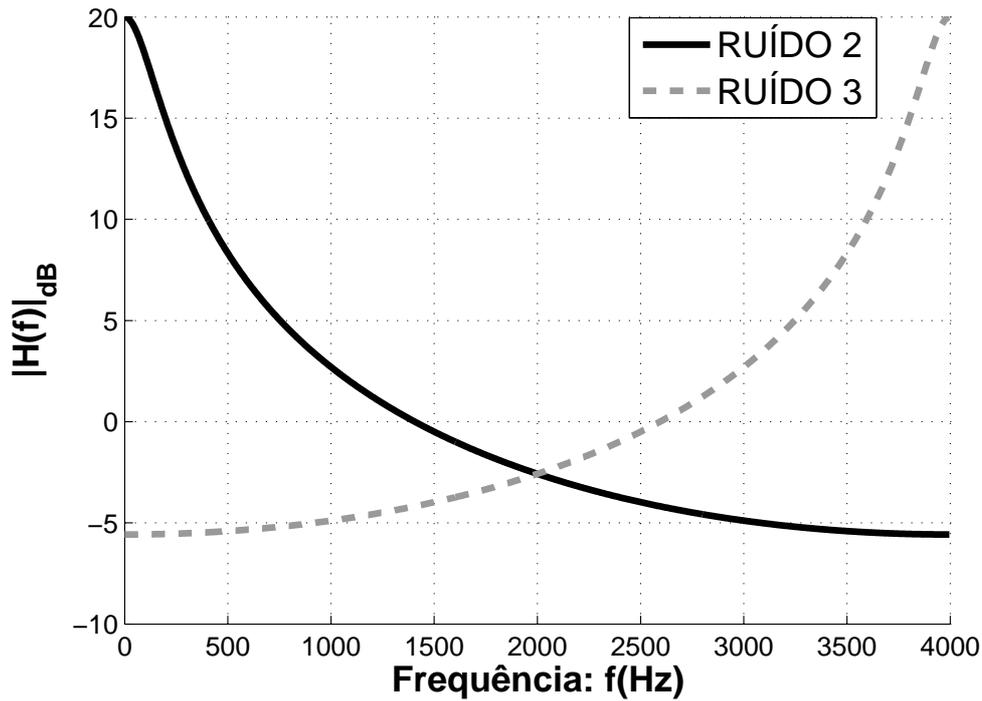


FIG. 5.10: Magnitude da Resposta em frequência da função de transferência  $H(z)$  para os casos de RUÍDO 2 e RUÍDO 3. A frequência de amostragem usada foi de 8000 Hz.

- RUÍDO 3: Ruído colorido com componentes significativas em frequências altas. Para obter este ruído, processou-se ruído branco gaussiano por um filtro com função de transferência

$$H(z) = \frac{1}{1 + 0.9z^{-1}}. \quad (5.7)$$

A magnitude da resposta em frequência deste filtro, considerando uma frequência de amostragem de 8000 Hz, é apresentada na FIG. 5.10 em linha tracejada.

Como consideração adicional para a aplicação do RUÍDO 2 e do RUÍDO 3 nos sinais, é necessário garantir que sua frequência de amostragem é de 8000 Hz, realizando-se sua sub-amostragem, se necessário.

A FIG. 5.11 apresenta as curvas de taxa de erro de detecção em função da SNR quando o método de autenticação de áudio é avaliado com RUÍDO 1, RUÍDO 2 e RUÍDO 3. Observa-se que quando é aplicado ruído com componentes significativas em frequências

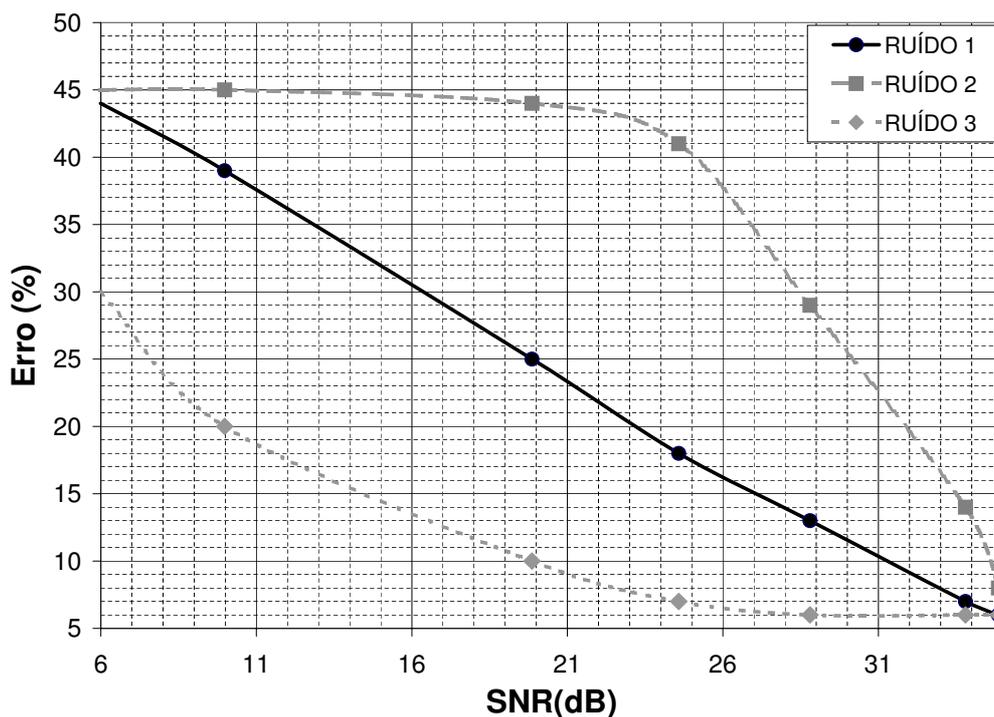


FIG. 5.11: Efeito de ruído: Erro (EER) de detecção na autenticação de áudio em função da SNR. O erro do sistema sem inserção de ruído é de 6%. A razão sinal-ruído inicial é 35 dB.

baixas (com ênfase nas componentes que afetam diretamente a ENF, 50 Hz neste caso), o efeito de ruído é mais sensível. Assim, RUÍDO 2 é o que mais afeta a detecção, seguido de RUÍDO 1 e RUÍDO 3. Nesse contexto, o comportamento da curva de erro de detecção em função da SNR aumenta com o ruído de fundo (menor SNR em dB): de maneira aproximadamente linear para o caso do RUÍDO 1, de maneira próxima à logarítmica para o caso do RUÍDO 2 e de maneira próxima à exponencial para o caso do RUÍDO 3.

Como dados adicionais, temos: a  $SNR_{original}$  (razão sinal-ruído inicial) da BASE ESPANHOLA é de 35 dB e a  $SNR_{original}$  da BASE CARIOCA 2 é de 29,77 dB.

#### 5.4.2 EFEITO DE SATURAÇÕES

Para fazer a análise do efeito da presença de saturações nos sinais na detecção, usou-se a BASE ESPANHOLA como referência. Primeiramente, aplicou-se um algoritmo VAD (BENYASSINE, 1997) a todos os sinais da base, para separar as regiões de voz

ativa das regiões de ruído de fundo. Naturalmente, foram provocadas saturações somente nas regiões de voz ativa. Optou-se por utilizar um limiar fixo para limitar a amplitude das amostras do sinal. O nível desse limiar define o percentual de amostras saturadas nas regiões de voz ativa. A FIG. 5.12 apresenta o caso de um sinal saturado em 3% de suas amostras.

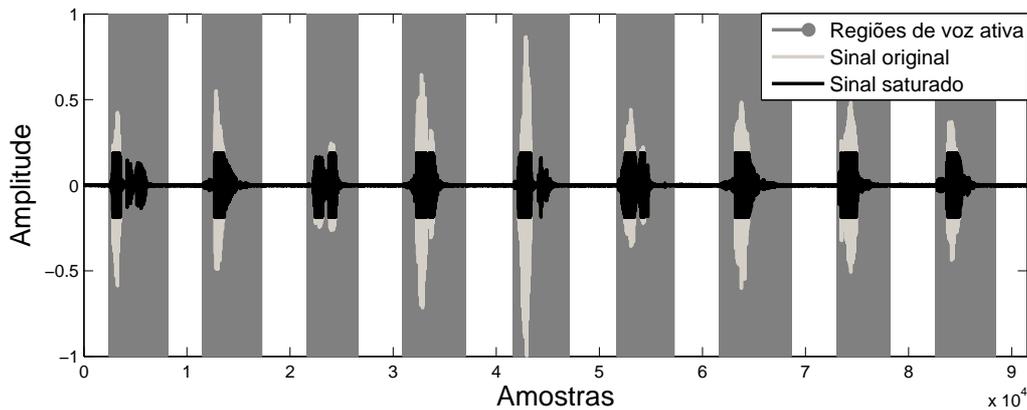


FIG. 5.12: Exemplo de um sinal com 3% das amostras (correspondentes à região ativa de voz) saturadas.

Modificando a percentagem de saturação e aplicando a técnica de autenticação de áudio (baseada no método de estimação de fase da DFT<sup>1</sup> com os parâmetros usadas nos exemplos das FIGs. 5.3 e 5.4) proposta sobre a BASE ESPANHOLA, obtemos a curva característica da taxa de erro de detecção em função da percentagem de saturação. A FIG. 5.13 mostra essa curva. Como se pode observar, níveis de saturação superiores a 0,5% afetam consideravelmente o método de autenticação proposto. Assim sendo, mesmo considerando que as percentagens de saturação encontradas tipicamente em sinais de áudio são baixas, seu efeito não pode ser desconsiderado. Como dado adicional, ressalta-se que a BASE CARIOCA 2 possui 21 sinais saturados de um total de 100 sinais; tais sinais saturados foram obtidas de gravações com volume alto (e não saturadas artificialmente).

### 5.4.3 EFEITO DA VARIAÇÃO DA ENF

Para fazer a análise do efeito da variação da ENF (presente nos sinais) sobre a detecção de edição, foram usados os sinais não editados da BASE ESPANHOLA, da BASE

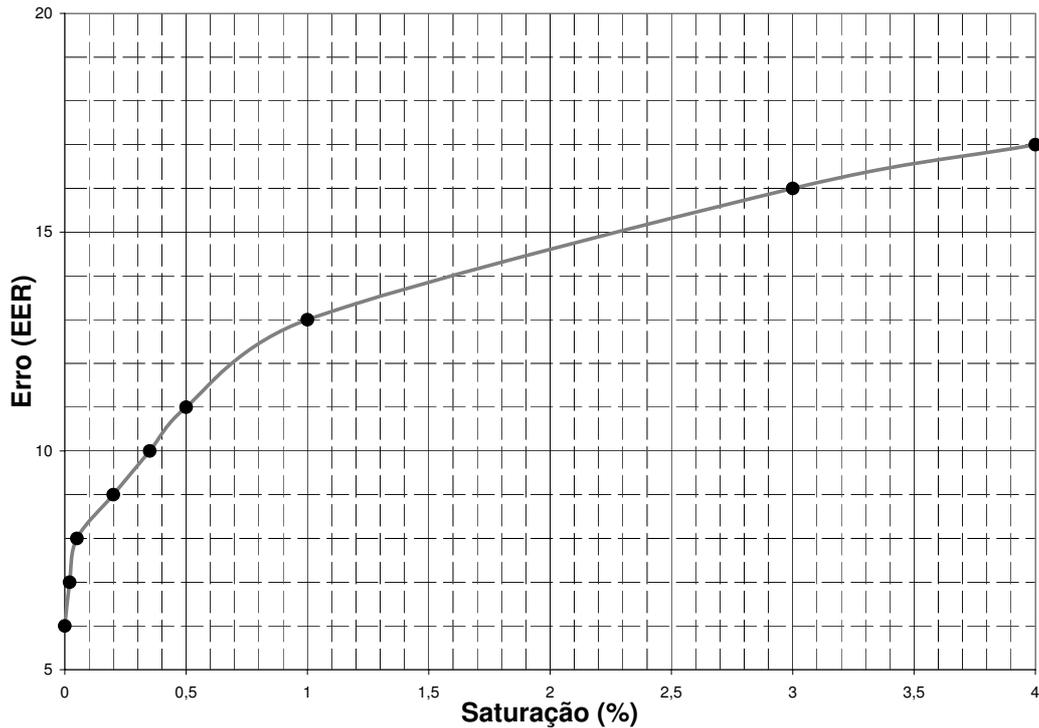


FIG. 5.13: Efeito de saturação: Erro (EER) de detecção na autenticação de áudio em função de percentagem de saturação. O erro do sistema sobre os sinais sem saturação é de 6%.

CARIOCA 1 e da BASE CARIOCA 2 como referências. Utilizou-se a técnica de estimação de frequência da DFT<sup>1</sup> (Seção 3.3) para analisar o comportamento da ENF. As estimativas de frequência foram realizadas em blocos de duração de 10 ciclos da ENF com superposição de 9 ciclos, após a sub-amostragem ( $f_{\text{downsample}} = 1000$  Hz para os casos de sinais que possuem ENF nominal de 50,0 Hz e  $f_{\text{downsample}} = 1200$  Hz para os casos de sinais que possuem ENF nominal de 60,0 Hz) e filtragem (filtro passa-banda com largura de faixa de 0,6 Hz) do sinal original.

A FIG. 5.14 apresenta os histogramas normalizados da distribuição das estimativas da ENF dos blocos de segmentação dos sinais originais da BASE ESPANHOLA, da BASE CARIOCA 1 e da BASE CARIOCA 2. Estes histogramas indicam as variações da ENF de cada uma das três bases (aparentemente similares) com os sinais tomados em conjunto. Contudo, estamos aqui mais interessados nas variações temporais (por sinal individual); desta forma mediu-se o desvio-padrão das estimativas da ENF dos blocos em cada sinal do

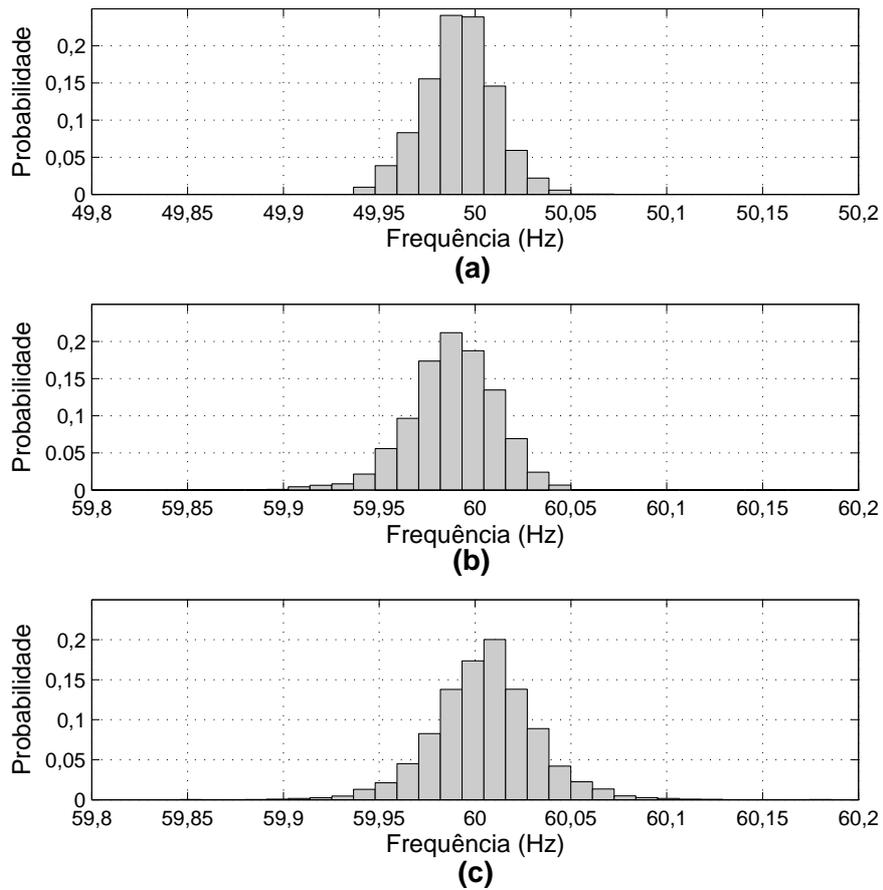


FIG. 5.14: Histograma normalizado da distribuição das estimativas da ENF (sinais não editados). A largura de banda do filtro passa-banda usado foi de 0,6 Hz para os três casos. Foi usado o método de estimação de frequência da DFT<sup>1</sup>. (a) BASE ESPANHOLA; (b) BASE CARIOCA 1; (c) BASE CARIOCA 2.

banco de dados (interesse em verificar como varia a ENF em intervalos curtos de tempo). Os histogramas normalizados correspondentemente são apresentados na FIG. 5.15, em que se verifica que na BASE ESPANHOLA a ENF tem um comportamento temporal mais estável que na BASE CARIOCA 1 e na BASE CARIOCA 2.

Mesmo que alguns sinais da BASE CARIOCA 1 apresentem maiores variações temporais da medida estimada de ENF por sinal do que na BASE ESPANHOLA, as taxas de erro (entre 7% e 8%) obtidas na avaliação estão dentro de limites razoáveis. Isto acontece porque estas variações temporais na BASE CARIOCA 1, com base na FIG. 5.15, são suaves, de tal forma que, quando acontecem mudanças de fase produzidas por edições,

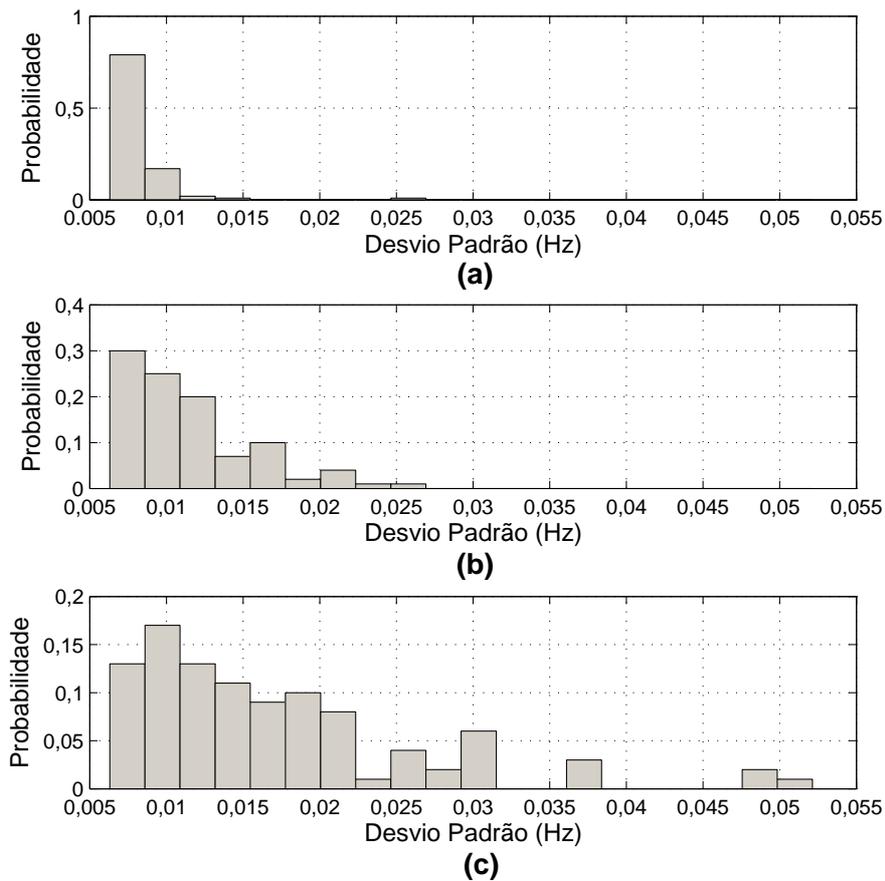


FIG. 5.15: Histograma normalizado da distribuição do desvio-padrão das estimativas da ENF em cada sinal do banco de dados (sinais não editados). A largura de banda do filtro passa-banda usado foi de 0,6 Hz para os três casos. (a) BASE ESPANHOLA; (b) BASE CARIOCA 1; (c) BASE CARIOCA 2.

elas são bastante evidentes. Um exemplo deste fenômeno pode ser visto na FIG. 5.5.

Finalmente, a BASE CARIOCA 2, que apresentou maiores variações temporais da medida estimada de ENF por sinal do que as outras duas bases, gerou maiores taxas de erro na detecção. Porém, neste caso, as saturações e ruído de fundo elevado certamente afetaram mais o resultado (considerando que afetam a medida de estimação de fase da ENF no sinal) que a possível variação temporal da ENF. Espera-se que o efeito de variação da ENF seja mais crítico em regiões onde a energia elétrica seja proveniente de centrais com menor porte.

## 5.5 RESUMO

Neste capítulo foi feita uma avaliação do sistema de autenticação de áudio apresentado no Capítulo 4. Foi apresentado o cálculo de limiares necessários para detectar automaticamente a presença da ENF (50,0 ou 60,0 Hz) nos sinais. Foram também apresentados alguns exemplos para o mecanismo visual de detecção de edição de áudio baseado na observação das mudanças abruptas de fase da ENF provocadas por edições. Adicionalmente, foi feita uma avaliação do mecanismo automático de detecção de edição de áudio para três bancos de dados, dois deles com condições favoráveis (BASE ESPANHOLA e BASE CARIOCA 1) e um deles em piores condições (BASE CARIOCA 2). As melhores taxas de erro (EER) obtidas para estes bancos de dados foram de 6% para a BASE ESPANHOLA, 7% para a BASE CARIOCA 1 e 15% para a BASE CARIOCA 2.

Finalmente, foram analisados alguns efeitos que podem interferir no método de detecção de edição de áudio proposto: ruído de fundo, saturações e variações temporais da ENF no sinal. As duas primeiras situações mostraram-se mais influentes no mecanismo de detecção, ao menos considerando-se variações suaves da ENF. Para o caso de variações da ENF bruscas no sinal, a detecção de edição certamente ficará com um pior desempenho. Mas também é importante destacar que as presenças tanto de ruído de fundo elevado quanto de saturações no sinal podem interferir na medida estimada da ENF (como no caso da BASE CARIOCA 2).

No próximo capítulo são apresentadas as considerações finais sobre todo o trabalho desenvolvido até aqui, que incluem as conclusões e propostas para trabalhos futuros.

## 6 CONSIDERAÇÕES FINAIS

O presente trabalho teve como objetivo a criação de uma ferramenta para autenticação de áudio digital com aplicação em fonética forense. Esta ferramenta é baseada na presença do sinal da rede elétrica nas gravações de áudio efetuadas por equipamentos a ela conectados. Seu fundamento é determinar as mudanças abruptas de ENF provocadas pelas edições feitas nos sinais de áudio (inserções e recortes de fragmentos no/do sinal original).

Os principais tópicos abordados neste trabalho foram:

- Estudo de duas técnicas de estimação de frequência e fase de um tom cossenoidal simples, uma baseada na Transformada Discreta de Fourier de tempo curto (método da DFT) e outra, numa análise de Fourier de alta precisão (método da DFT<sup>1</sup>) que também utiliza a DFT da primeira derivada do sinal. Como na literatura referente à DFT<sup>1</sup> (DESAINTE-CATHERINE, 2000) só é descrito o método para estimação de frequência, foi proposto aqui um método para estimação de fase para o caso de um tom simples (somente uma componente de frequência).
- Desenvolvimento de um procedimento para detecção automática da presença da ENF (50,0 ou 60,0 Hz) nos sinais de áudio.
- Desenvolvimento de um procedimento para permitir localizar visualmente as mudanças de fase da ENF provocadas pelas edições feitas no sinal de áudio.
- Desenvolvimento de um procedimento automático para discriminação entre sinais originais e editados mediante uma relação de decisão associada a uma *feature* que quantifica as mudanças de fase encontradas.
- Avaliação das técnicas propostas com sinais provenientes de três bases de dados: duas bases com sinais gravadas sob condições favoráveis (baixo ruído de fundo e ausência de saturações), denominadas BASE ESPANHOLA (com ENF característica de 50,0 Hz) e BASE CARIOCA 1 (com ENF característica de 60,0 Hz); e uma outra com sinais com saturações e com ruído de fundo elevado, denominada BASE CARIOCA 2 (com ENF característica de 60,0 Hz).

- Estudo dos efeitos de condições que podem influir na detecção, como presença de ruído de fundo, ocorrência de saturações e variações da ENF.

## 6.1 CONCLUSÕES

As principais conclusões relativas a este trabalho são:

- Com o método da DFT<sup>1</sup>, obtêm-se estimativas mais exatas tanto de fase como de frequência do que com o método da DFT, para o mesmo tamanho de janela e o mesmo número de amostras frequenciais. Ao se incrementar o tamanho da janela, a tendência é de melhora nas estimativas de frequência para a DFT e a DFT<sup>1</sup>.
- A técnica de estimação de fase proposta com base no método da DFT<sup>1</sup> mostra resultados satisfatórios. Constatou-se visualmente que praticamente todas as estimativas de fase obtidas eram similares aos dados de fase do sinal original. Sob uma análise estatística, os resultados de erro se mostraram suficientemente pequenos para atestar a eficácia do método.
- A técnica de detecção automática da ENF consegue distinguir a presença ou ausência do sinal da rede elétrica nas gravações digitais. Nos resultados obtidos, o limiar aplicado na relação de decisão para discriminar entre sinais com e sem a ENF permite distinguir as duas situações apropriadamente, tanto para 50,0 Hz quanto para 60,0 Hz.
- O procedimento visual de detecção permite localizar os pontos e inferir os tipos das edições (recortes ou inserções do fragmento do sinal) efetuadas na gravação digital mediante a observação das mudanças de fase da ENF por elas provocadas.
- O mecanismo automático de discriminação entre sinais originais e editados mostra resultados satisfatórios quando se trabalha com sinais gravados sob condições controladas: ruído de fundo baixo (alta SNR), ausência de saturações e pouca variação da ENF. Assim, para o caso da BASE ESPANHOLA e da BASE CARIOCA 1, os menores valores de EER obtidos são de 6% e 7%, respectivamente. Estes pequenos valores de erro correspondem à probabilidade de a pessoa que editou o sinal ter realizado (de modo não intencional) tal edição de forma a provocar uma mudança de fase muito pequena ou nula.

- Para a aplicação específica deste trabalho, detecção de edição com base na mudança de fase do sinal da rede elétrica, o uso dos métodos da DFT e da DFT<sup>1</sup> apresentam resultados similares na discriminação automática entre sinais originais e editados. Neste caso, a exatidão garantida pelo método da DFT<sup>1</sup> na estimação de fase não tem impacto relevante. Por outro lado, com o uso do método da DFT<sup>1</sup> obtém-se melhor resolução com o procedimento visual nas estimativas de fase, especialmente nas regiões onde ocorrem as mudanças abruptas de fase provocadas pelas edições. Além disso, para uma aplicação qualquer onde o tempo de processamento seja importante, o método aqui proposto (estimação de fase usando a DFT<sup>1</sup>) apresenta menor complexidade computacional do que o método baseado na DFT (já que este precisa de um número maior de pontos para o cálculo da DFT).
- Quando existe ruído de fundo elevado no sinal analisado, a possibilidade de erro na detecção aumenta, especialmente quando o ruído possui componentes significativas de potência em frequências baixas (com ênfase nas componentes que afetam diretamente a ENF).
- Quando saturações estão presentes no sinal, a medida estimada da fase da ENF também sofre variações significativas. Neste caso, algumas mudanças de fase provocadas pelas saturações podem ser confundidas com mudanças provocadas por edições. Como no caso anterior, a possibilidade de erro na detecção de edição aumenta.
- Finalmente, para o caso quando as variações da ENF são lentas, as mudanças abruptas de fases ainda apresentam variações mais significativas, e com isso a tarefa de detecção de edição fica factível. Por outro lado, apesar de não testado por não dispormos de sinais de outras regiões, é razoável supor-se que no caso de ocorrerem variações bruscas da ENF embutida no sinal, a detecção de edição irá se complicar. É importante considerar, contudo, que tanto o ruído de fundo elevado como as saturações podem interferir na medida estimada da ENF (como foi o caso da BASE CARIOCA 2).

## 6.2 PROPOSTAS PARA TRABALHOS FUTUROS

Algumas propostas para trabalhos futuros são descritas a seguir:

- A localização dos pontos de edição no sinal neste trabalho é feita somente de maneira visual; como trabalho futuro, pode-se propor um método automático para a detecção dos pontos de edição.
- O mecanismo automático de detecção de edição baseia-se numa relação de decisão expressa na EQ. 4.6. Em seu lugar, poderia-se usar uma decisão baseada em razão de verossimilhança obtida a partir do modelo para as densidades de probabilidade da medida característica  $F$  (ou outra) sob as hipóteses de sinal editado e de sinal original. Em (VAN TREES, 1968), encontra-se um estudo detalhado da técnica de decisão com base na razão de verossimilhança.
- (NICOLALDE, 2009a) discute o uso de distâncias espectrais obtidas mediante o uso de filtros adaptativos empregados como preditores lineares, com o objetivo de obter informação, em cada instante de tempo, sobre mudanças abruptas no espectro, sintomáticas de possíveis edições de áudio. Considerando que o sinal de voz sofre mudanças abruptas no espectro (provocadas, entre outras coisas, pelas transições entre diferentes fonemas) e que as edições geralmente são feitas nas regiões onde não existe sinal de voz, esta análise seria feita preferencialmente nas regiões de silêncio, requerendo um algoritmo VAD (*Voice Active Detection*) para separar as regiões de voz ativa das regiões de silêncio. Pode-se aprofundar este estudo no sentido de se obter uma medida característica baseada nas distâncias espectrais e com ela fazer uma decisão automática.
- Combinar as duas técnicas baseadas na mudança de fase da rede elétrica e no uso de distâncias espectrais. Neste caso, é possível usar as medidas características das duas técnicas e determinar uma decisão bidimensional (fusão de dados). O limiar de decisão seria determinado por uma curva que deve separar as regiões de sinais originais e sinais editados no plano bidimensional das duas medidas características. Esta curva pode ser obtida, por exemplo, mediante o uso de redes neurais.
- Avaliar o método proposto com sinais proveniente de diferentes regiões, tais como Nordeste, Norte e Sul, incluindo localidades em que a ENF não possui um controle muito sofisticado (por exemplo, em regiões alimentadas por usinas de pequeno porte).

## 7 REFERÊNCIAS BIBLIOGRÁFICAS

- BENYASSINE, A., SHLOMOT, E. and SU, H. L. **ITU-T recommendation G.729 annex b: A silence compression scheme for use with g.729 optimized for v.70 digital simultaneous voice and data applications.** *IEEE Communications Magazine*, vol. 35, pp. 64–73, September 1997.
- BISCAINHO, L. W. and ESQUEF, P. A. *Spectral-Based Analysis and Synthesis of Audio Signals. In: Advances in Audio and Speech Signal Processing: Technologies and Applications.* Idea Group, Hershey, 2007. ISBN 0471899550.
- BRAID, A. C. M. *Fonética Forense.* Millenium, Campinas, Brasil, 2ª edição, 2003. ISBN 2147483647.
- BRIXEM, E. B. Enf: Quantification of the magnetic field. In *Proc. of the AES 33rd. International Conference: Audio Forensics, Theory and Practice*, pp. 1–4, Denver, USA, June 2008. AES.
- BUTCHER, A. Forensic phonetics: issues in speaker identification evidence. In *Proc. of the International Conference of the Institute of Forensic Studies*, pp. 1–16, Prato, Italy, July 2002. Institute of Forensic Studies.
- COOPER, A. J. *Detection of Copies of Digital Audio Recordings for Forensic Purposes.* PhD Thesis, Faculty of Technology, The Open University, United Kingdom, 2006.
- COOPER, A. J. The electric network frequency (ENF) as an aid to authenticating forensic digital audio recordings - an automated approach. In *Proc. of the AES 33rd. International Conference: Audio Forensics, Theory and Practice*, pp. 1–10, Denver, USA, June 2008. AES.
- DEAN, D. J. The relevance of replay transients in the forensic examination of analogue tape recordings. Publication 16/91, Police Scientific Development Branch, Home Office Science and Technology Group, 1991.
- DELP, E., MEMON, N. and WU, M. Digital forensics. In *IEEE Signal Processing Magazine*, vol. 26, pp. 14–15. IEEE, March 2009.
- DESAINTE-CATHERINE, M. and MARCHAND, S. **High-precision Fourier analysis of sounds using signal derivatives.** *Journal of the Audio Engineering Society*, vol. 48, pp. 654–667, July 2000.
- GRIGORAS, C. **Digital audio recording analysis: the electric network frequency (ENF) criterion.** *Journal of Speech, Language and the Law*, vol. 12, pp. 63–76, 2005.

- KOENING, B. E. **Authentication of forensic audio recordings.** *Journal of the Audio Engineering Society*, vol. 38, no. 1/2, pp. 3–33, February 1990.
- KOENING, B. E. and LACEY, D. S. **Forensic authentication of digital audio recordings.** *Journal of the Audio Engineering Society*, vol. 57, no. 9, pp. 662–695, September 2009.
- MAHER, R. C. Audio forensic examination. In *IEEE Signal Processing Magazine*, vol. 26, pp. 84–94. IEEE, March 2009.
- MARTIN, A., DODDINGTON, G., KAMM, T., ORDOWSKI, M. and PRZYBOCKI, M. The DET curve in assessment of detection task performance. In *Proc. of the European Conference on Speech Communication and Technology*, pp. 1–4, Rhodes, Greece, September 1997. ISCA.
- NICOLALDE, D. P. e APOLINÁRIO JR., J. A. Evaluating digital audio authenticity with spectral distances and ENF phase change. In *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 1417–1420, Taipei, Taiwan, April 2009a. IEEE.
- NICOLALDE, D. P., APOLINÁRIO JR., J. A. e BISCAINHO, L. W. Autenticação de áudio digital com base na mudança de fase da frequência da rede elétrica. Em *XXVII Simpósio Brasileiro de Telecomunicações–SBrT 2009*, Blumenau, SC, Brasil, Setembro–Outubro 2009b. Sociedade Brasileira de Telecomunicações.
- NICOLALDE, D. P., APOLINÁRIO JR., J. A. e BISCAINHO, L. W. Uma ferramenta para a autenticação de áudio com aplicação em fonética forense. Em *XX Congresso Nacional de Criminalística – III Congresso Internacional de Perícia Criminal*, João Pessoa, PB, Brasil, Outubro 2009c. Associação Brasileira de Criminalística – ABC.
- ORTEGA-GARCIA, J., GONZÁLEZ-RODRÍGUEZ, J. and MARRERO-AGUIAR, V. Ahumada, a large speech corpus in Spanish for speaker characterization and identification. In *Elsevier Speech Communication*, vol. 31, pp. 255–264, June 2000.
- RUMSEY, F. **Forensic audio analysis.** *Journal of the Audio Engineering Society*, vol. 56, pp. 211–217, March 2008.
- SANDERS, R. W. Digital authenticity using the electric network frequency. In *Proc. of the AES 33rd. International Conference: Audio Forensics, Theory and Practice*, pp. 1–11, Denver, USA, June 2008. AES.
- EdiTracker - Dedicated Software Module, User Manual.** Speech Technology Center, St. Petersburg, Russia, 2005.
- VAN TREES, H. L. **Detection, Estimation and Modulation Theory.** Wiley, New York, USA, 1968. ISBN 0471899550.