Autenticação de áudio digital com base na mudança de fase da frequência da rede elétrica

Daniel Patricio Nicolalde, José Antonio Apolinário Jr. e Luiz Wagner Pereira Biscainho.

***Resumo*— Este artigo propõe um método para autenticação de áudio, que consiste em verificar se um sinal de áudio gravado digitalmente foi ou não adulterado. O método baseia-se na verificação da mudança de fase associada à frequência da rede elétrica, quase sempre embutida nas gravações. Ele fornece uma ferramenta visual que permite localizar as mudanças abruptas de fase da frequência da rede indicativas de pontos de edição, e uma medida característica que permite discriminar automaticamente, por uma razão de verossimilhança, sinais originais de editados. Apresentam-se os fundamentos teóricos e questões práticas de implementação da técnica, aferindo-se seu desempenho sobre uma base de sinais reais digitalmente editados.**

***Palavras-Chave*— Autenticação de áudio digital, frequência da rede elétrica, mudança abrupta de fase da frequência da rede.**

***Abstract*— This paper presents a method for audio authentication, which consists in verifying whether or not a digitally recorded audio signal has been adulterated. The method is based on tracking phase changes associated to the electric network frequency, usually embedded in audio recordings. It provides a visual aid to locate abrupt network frequency phase changes, indicatives of edition points, and a feature measure that allows automatic discrimination between real and edited signals via a likelihood ratio. Theoretical bases of the proposed technique are presented along with practical implementation issues; its performance is assessed over a database of real signals which have been digitally edited.**

***Keywords*— Digital audio authenticity, electric network frequency, abrupt phase changes of the network frequency.**

I. Introdução

No campo da criminalística, as provas físicas servem como evidência para reconstituição dos fatos ocorridos em um determinado caso, podendo auxiliar na definição da inocência ou culpa dos implicados. As fitas de áudio são usadas como provas criminalísticas devido ao seu conteúdo, que pode consistir de conversações gravadas diretamente de um microfone ou oriundas de ligações telefônicas.

Com o uso da tecnologia digital, copiar, alterar ou trocar o conteúdo dos sinais de áudio é, hoje em dia, uma atividade simples. Portanto, uma das tarefas da perícia fonética é determinar a autenticidade das gravações de áudio para aceitá-las como provas em procedimentos legais [1], [2]. Existem metodologias para autenticação de áudio proveniente de gravações analógicas [1], [3]; neste caso, a análise é baseada principalmente em:

Daniel Patricio Nicolalde, José Antonio Apolinário Jr. (Programa de Pós-graduação em Engenharia Elétrica, Instituto Militar de Engenharia, Rio de Janeiro, RJ, Brasil, E-mails: danielnicolalde@hotmail.com, apolin@ime.eb.br) e Luiz Wagner Pereira Biscainho (PEE/COPPE, Universidade Federal de Rio de Janeiro, Rio de Janeiro, RJ, Brasil, E-mail: wagner@lps.ufrj.br).

- exame crítico do perito depois de ouvir a gravação;
- inspeção física;
- exames instrumentais no domínio do tempo (forma de onda), da frequência (espectro), e de ambos (espectrograma);
- comparação com exemplos similares;
- testemunhos.

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 [4] que avalia a autenticidade de áudio digital, a literatura técnica associada, principalmente sob a perspectiva de processamento digital de sinais, é pobre. Considerando a presença da frequência da rede elétrica (conhecida pela sigla ENF, de *Electric Network Frequency*) na maioria das gravações [5], este artigo apresenta uma técnica para a autenticação de áudio digital baseada no comportamento da fase da ENF.

A organização deste artigo é a seguinte. A Seção II apresenta conceitos fundamentais da frequência da rede elétrica, assim como o seu comportamento nas gravações. O método usado para a autenticação de áudio digital e os resultados experimentais obtidos são expostos nas Seções III e IV, respectivamente. Finalmente, as conclusões são apresentadas na Seção V.

II. Sobre a frequência da rede elétrica

Os sistemas de transmissão de energia elétrica trabalham conectados a estações e subestações de energia numa rede de alta voltagem para, posteriormente, mediante transformadores de distribuição, reduzir a tensão e 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 [6].

O padrão utilizado para o valor nominal da ENF é de 50 Hz ou 60 Hz. Como exemplos de utilização de 50 Hz, pode-se citar 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 energía elétrica da Inglaterra e do País de Gales, administrado pela "*National Grid Company*" (UCTE), onde a ENF, cujo valor nominal é de 50 Hz, apresenta variações dentro de $\pm$0,2 Hz sobre este valor nominal [6].

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. Por isto, a ENF constitui uma importante ferramenta para a autenticação de áudio.

Em [5] e [6], encontram-se métodos para autenticação de áudio baseados no rastreamento da ENF ao longo do tempo. Estes métodos 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 onde a gravação foi realizada. A técnica proposta neste trabalho para a autenticação de um sinal de áudio digitalizado é baseada no comportamento da fase da ENF e é explicada na Seção III.

III. Método Implementado

Considerando a razoável estabilidade da ENF, o método proposto baseia-se em achar mudanças abruptas na fase desta frequência como sintoma relevante de que um sinal de áudio foi editado.

A. Estimando a fase da ENF

O procedimento de estimação de fase é baseado em [7]. Para começar o procedimento do rastreamento da fase da ENF, realiza-se uma sub-amostragem do sinal de áudio para 1000 Hz ou 1200 Hz, dependendo de o valor nominal da ENF ser 50 Hz ou 60 Hz, respectivamente. O objetivo desta sub-amostragem é diminuir a carga computacional no processamento, assim como trabalhar com um número exato de amostras por ciclo do valor nominal da ENF—neste caso, 20.

Posteriormente, aplica-se ao sinal sub-amostrado um filtro passa-banda, muito estreito e de fase linear, com o objetivo de considerar somente os componentes da banda de interesse. O filtro é centrado no valor nominal da ENF e pode possuir uma largura de faixa entre 0,6 e 1,4 Hz, dependendo da tolerância permitida nesta frequência. Nas simulações realizadas neste trabalho, o filtro foi projetado como um FIR de 10000 coeficientes, sendo a filtragem realizada de modo a não introduzir retardo em relação ao sinal original (usando-se a função *filtfilt* de Matlab®).

Subsequentemente, com o objetivo de obter informação da fase da ENF ao longo do tempo, o sinal filtrado é segmentado em blocos de duração de 3 ciclos de ENF nominal com sobreposição de 2 ciclos com respeito ao bloco seguinte. Isto significa que se deseja procurar as mudanças de fase no sinal

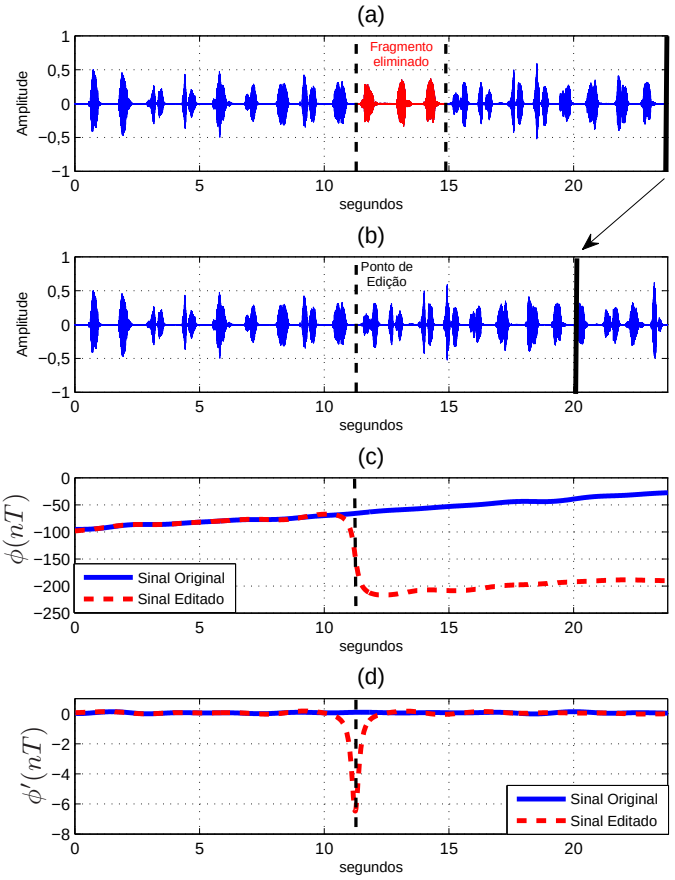

Fig. 1. Edição de áudio: eliminação de um fragmento do sinal provocando uma mudança considerável de fase. O valor nominal da ENF é de 50 Hz, a largura de faixa do filtro usado é de 0,8 Hz e $T = \frac{1}{50}$ segundo (equivalente a um ciclo da ENF nominal). (a) Sinal original; (b) Sinal editado; (c) Curva de estimação de fase na ENF; (d) Curva da derivada da fase estimada.

filtrado a cada intervalo de tempo correspondente a um ciclo da ENF nominal. Em cada bloco do sinal é aplicada uma janela de Hanning e a estimação da fase da ENF é determinada com ajuda da Transformada Discreta de Fourier (DFT) do bloco janelado. O número de pontos considerados para a DFT deve ser tal que o valor específico de um dos seus pontos corresponda exatamente ao valor nominal da ENF (50 ou 60 Hz). Finalmente, o valor em graus do ângulo da DFT nesse ponto é o que corresponde à fase da ENF.

O resultado visual esperado para a curva de estimação da ENF no sinal de áudio digital é teoricamente uma linha reta com valor constante de fase ao longo do tempo para o caso de um sinal que não foi editado e cuja frequência é igual à sua frequência nominal. Mas, devido às variações da ENF com respeito ao seu valor nominal, pode-se obter uma curva da ENF com aspecto aproximadamente retilíneo, com derivada positiva ou negativa conforme o valor real da ENF seja, respectivamente, maior ou menor que o valor nominal.

A Fig. 1 apresenta um caso de edição de áudio em que um fragmento do sinal original foi eliminado com o objetivo de alterar o significado do conteúdo. Como produto deste recorte, nota-se uma mudança de fase neste ponto de corte. A mudança

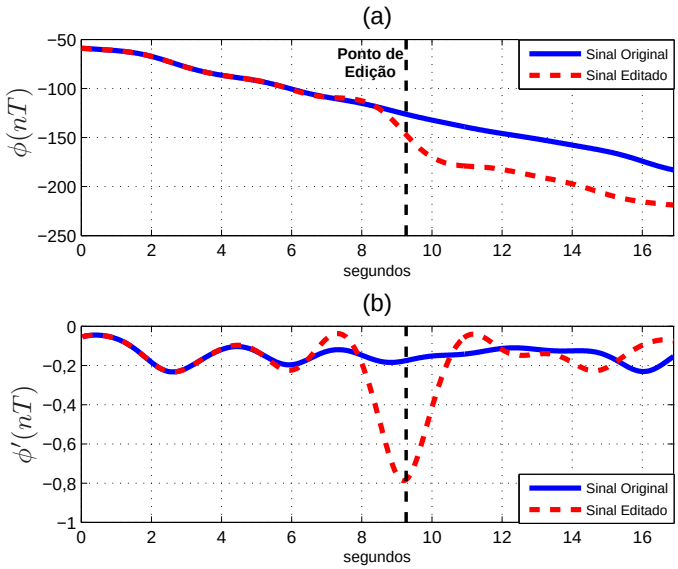

Fig. 2. Edição de áudio: eliminação de um fragmento do sinal provocando uma mudança de fase moderada. O valor nominal da ENF é de 50 Hz, a largura de faixa do filtro usado é de 0,8 Hz e $T = \frac{1}{50}$ segundo (equivalente a um ciclo da ENF nominal). (a) Curva de estimação da fase na ENF; (b) Curva da derivada da fase estimada.

de fase provocada no sinal editado é consideravelmente grande com respeito ao sinal original. A curva de estimação de fase ao longo do tempo, apresentada Fig. 1(c), ilustra este fenômeno.

A Fig. 2 apresenta outro caso de eliminação de um fragmento do sinal de áudio em que a mudança de fase é menor do que a do primeiro caso. Como resultado, a curva da mudança de fase, apresentada na Fig. 2(a), tem um desvio menor com respeito à curva do sinal original. É importante mencionar que quanto menor é a mudança de fase, mais difícil se torna a tarefa de decidir se o sinal foi editado. Por outro lado, é muito pouco provável que uma pessoa, mesmo com conhecimento técnico mas sem conhecimento prévio deste assunto específico, possa evitar este efeito (na opinião dos autores, somente um trabalho profissional poderia obter um sinal editado sem nenhum rastro destas mudanças de fase).

A Fig. 3 apresenta o caso onde foi feita a inserção de um fragmento do sinal. Como produto desta edição, veem-se dois pontos de mudança de fase. Neste caso, a primeira mudança de fase foi negativa e a segunda mudança de fase foi positiva. O resultado destas mudanças é apresentado na Fig. 3(c).

Com a utilização da curva de estimação de fase da ENF, tem-se uma ferramenta visual para auxiliar o processo de autenticação de áudio. Adicionalmente, é importante obter uma medida característica (*feature*) da curva da fase da ENF para poder automatizar o processo da autenticação de áudio.

B. Medindo a mudança de fase da ENF

Ponderando que as mudanças de fase são as que determinam se o sinal de áudio digital foi editado, é importante obter uma medida que caracterize essas mudanças. Primeiramente, considera-se $\phi(n)$ como a medida da estimação de fase da ENF no n-ésimo bloco do sinal e que o sinal de áudio foi

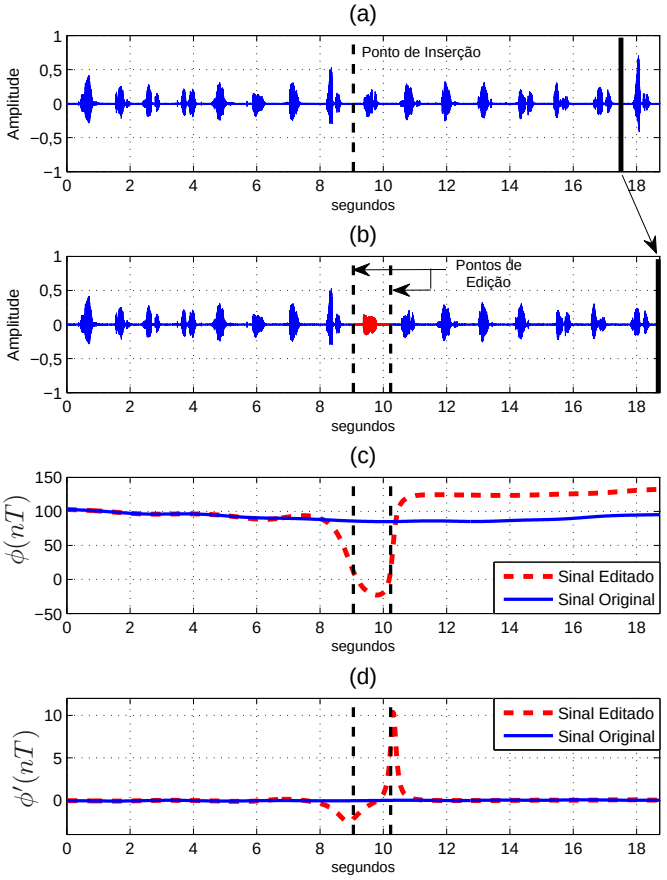

Fig. 3. Edição de áudio: inserção de um fragmento do sinal provocando duas mudanças de fase. O valor nominal da ENF é de 50 Hz, a largura de faixa do filtro usado é de 0,8 Hz e $T = \frac{1}{50}$ segundo (equivalente a um ciclo da ENF nominal). (a) Sinal original; (b) Sinal editado; (c) Curva de estimação de fase na ENF; (d) Curva da derivada da fase estimada.

segmentado num número total de N blocos. Posteriormente, define-se $\phi'(n)$ como a derivada de $\phi(n)$, cuja expressão é dada por:

$$\phi'(n) = \phi(n) - \phi(n-1), \quad 2 \leq n \leq N. \tag{1}$$

O objetivo de computar-se $\phi'(n)$ é verificar as variações temporais no comportamento da curva de estimativa da fase da ENF. Então, quando um sinal não foi editado, espera-se que se assemelhe a uma linha reta com valor constante ao longo do tempo. O valor constante vai depender da inclinação da curva de estimação de $\phi(n)$. Quando o sinal foi editado, tem-se a inclusão 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. As Figs. 1(d), 2(b) e 3(d) apresentam as curvas da derivada da estimação de fase da ENF, assim como as localizações aproximadas dos pontos de edição nos picos destas curvas para os exemplos reais anteriormente

expostos.

O valor médio de $\phi'(n)$ é definido por:

$$m_{\phi'} = \frac{1}{N-1}\sum_{n=2}^{N}\phi'(n). \quad (2)$$

Posteriormente, a medida característica para avaliar a edição de áudio baseada na fase da ENF é definida como:

$$M = 100\log\left\{\frac{1}{N-1}\sum_{n=2}^{N}|\phi'(n) - m_{\phi'}|\right\}. \quad (3)$$

O objetivo desta medida é fornecer um valor que permita discriminar os sinais editados dos não editados. A finalidade de subtrair o valor médio de $\phi'(n)$ é não considerar o valor médio da inclinação da curva de estimação de fase da ENF, provocada pelas variações da ENF em sinais não editadas. O valor logarítmico é utilizado para espalhar os valores de M na decisão entre os sinais editados e os originais.

Para o processo de detecção, definimos o conjunto de hipóteses H, $H \in \{H_O, H_E\}$, onde H_O e H_E representam as hipóteses do que o sinal de áudio digital seja original e editado, respectivamente. Posteriormente, para o procedimento de decisão, a razão de verossimilhança é dada por:

$$M \underset{H_O}{\overset{H_E}{\gtrless}} \gamma, \quad (4)$$

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

$$P_D = P\left(\hat{H} = H_E|H_E\right) = P\left(M > \gamma|H_E\right), \quad (5)$$

$$P_F = P\left(\hat{H} = H_E|H_O\right) = P\left(M > \gamma|H_O\right), \quad (6)$$

$$P_M = P\left(\hat{H} = H_O|H_E\right) = P\left(M < \gamma|H_E\right). \quad (7)$$

Adicionalmente, tem-se que:

$$P_D = 1 - P_M. \quad (8)$$

Para uma detecção ótima, o objetivo é achar o valor de γ que obtenha o maior valor de P_D (diminuindo o valor de P_F e P_M). Para estabelecer este limiar, é necessária a preparação de um banco de dados com sinais editados de áudio digital e comparar com os sinais originais. Posteriormente, com base neste banco de dados, realiza-se a avaliação do método, variando γ e chegando a uma curva de P_M em função de P_F conhecida como curva DET (*Detection Error Tradeoff*) [8]. Nesta curva, o ponto onde $P_M = P_F$ é conhecido como EER (*Equal Error Rate*). O valor de γ correspondente ao EER é usado como o ponto de operação quando se considera que ambos os erros têm igual importância. De qualquer maneira, mesmo que numa dada aplicação o ponto de operação seja escolhido diferente do EER, este valor oferece uma idéia do desempenho global do método.

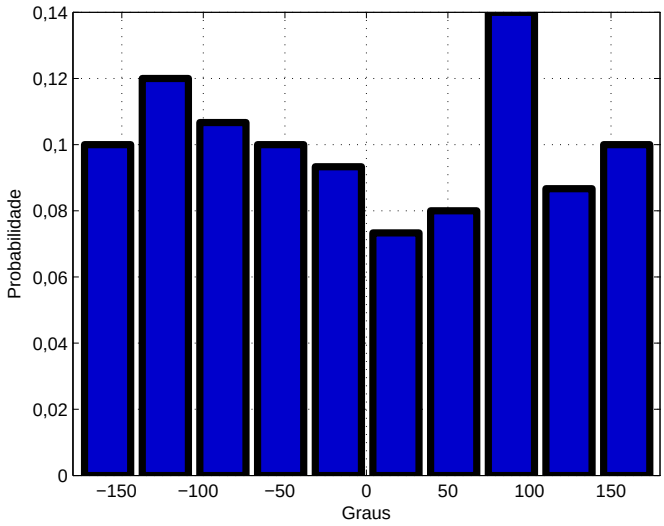

Fig. 4. Histograma normalizado da distribuição das mudanças de fase nos sinais editados.

IV. Resultados Experimentais

Para avaliar o método proposto neste artigo, aplicou-se a técnica a um banco de dados criado com sinais editados. Os sinais originais deste banco foram provenientes de duas bases públicas em castelhano, AHUMADA e GAUDI, obtidas via http://atvs.ii.uam.es/databases.jsp [9]. Todos os sinais utilizados da base estão 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 com frases gravadas. Cada sinal tem duração entre 15 e 30 segundos. Dispõe-se, portanto, de 100 sinais de áudio.

Posteriormente, os dois sinais por cada locutor foram editados. Do primeiro sinal foi eliminado um fragmento, provocando uma mudança de fase (um ponto de edição) como foi feito nos exemplos das Figs. 1 e 2. No segundo sinal, foi inserido um fragmento provocando duas mudanças de fase (dois pontos de edição) como foi feito no exemplo da Fig. 3. 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 mudanças 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 é apresentado na Fig. 4. Ele indica uma distribuição aproximadamente uniforme entre $-180°$ e $+180°$. Ao final, dispõe-se de um banco de dados com 100 sinais originais e 100 sinais editados. Os sinais

usados na preparação das Figs. 1, 2 e 3 fazem parte desse *corpus*.

No processo de estimação da fase da ENF dos sinais da base, utilizou-se uma largura de faixa de 0,8 Hz para o filtro passa-banda centrado no valor nominal da ENF. Os histogramas normalizados da distribuição da medida característica M para os sinais originais e editados da base são mostrados na Fig. 5. A curva DET (P_M em função de P_F) e a localização

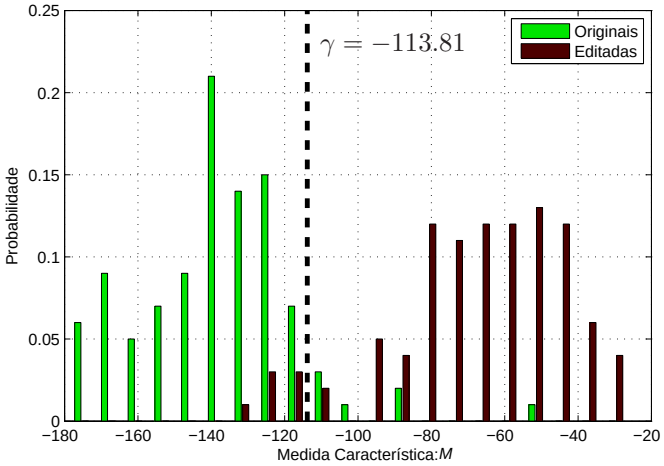

Fig. 5. Histograma normalizado da distribuição da medida característica M, para a banco de dados usado.

do ponto EER são mostrados na Fig. 6. Como se vê, o valor do erro do sistema implementado é de 7%. O valor do limiar de decisão, γ, necessário para obter este valor de EER é de -113.81 e encontra-se indicado sobre a Fig. 5.

V. Conclusões

A técnica proposta para autenticação de áudio digital baseada na mudança de fase da ENF mostrou-se satisfatória para a detecção de edições (recortes e/ou inserções). Esta técnica fornece informação visual que permite que, pela observação das mudanças de fase, localizem-se os pontos em que o sinal foi editado e se infira qual foi o tipo da edição. O cálculo da medida característica M, definida neste artigo, possibilita que, mediante uma razão de verossimilhança, discrimine-se de forma automática entre sinais editados e originais. Tal automatização, a qual permite a quantificação do desempenho do método, é a principal contribuição deste trabalho.

A taxa de acerto de 93% obtida na avaliação do método para o banco de dados usado nos experimentos, considerando uma distribuição de mudanças de fase aproximadamente uniforme, é uma boa confirmação da eficácia do sistema.

Agradecimentos

Os autores agradecem ao CNPq, à FAPERJ, e à CAPES pelo apoio financeiro aos projetos de pesquisa.

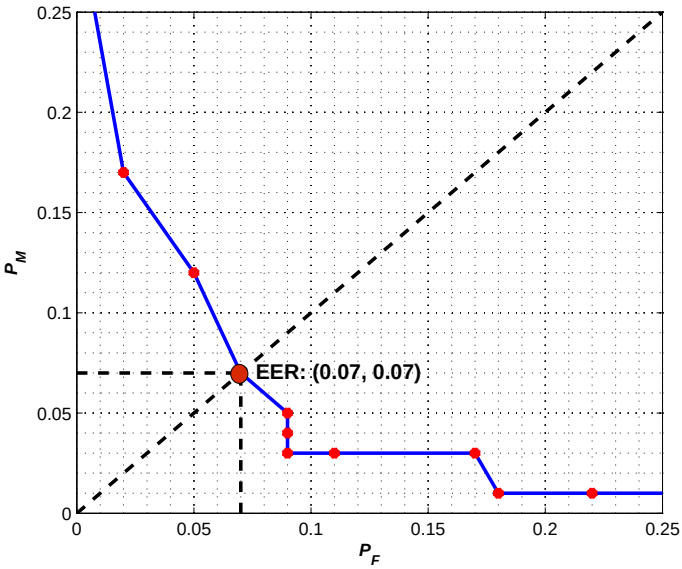

Fig. 6. Curva DET: P_M en função de P_F para o banco de dados usado. O ponto indicado corresponde ao *Equal Error Rate*, ou seja, o limiar onde $P_F = P_M$.

Referências

[1] B. E. Koening, "Authentication of forensic audio recordings," *Journal of the Audio Engineering Society*, vol. 38, pp. 3–33, January/February 1990.

[2] E. B. Brixen, "ENF: quantification of the magnetic field," *AES* 33^{rd} *International Conference: Audio Forensic, Theory and Practice*, Denver, CO, USA, June 2008.

[3] D. J. Dean, "The relevance of replay transients in the forensic examination of analogue tape recordings," *Police Scientific Development Branch, Home Office, Science and Technology Group*, Sandridge, UK, 1991.

[4] *Edit Track, User Manual*. Speech Technology Center, St. Petersburg, Russia, 2005.

[5] R. W. Sanders, "Digital audio authenticity using the electric network frequency," *AES* 33^{rd} *International Conference: Audio Forensic, Theory and Practice*, Denver, CO, USA, June 2008.

[6] A. J. Cooper, "The Electric Network Frequency (ENF) as an aid to authenticating forensic digital audio recordings – an automated approach," *AES* 33^{rd} *International Conference: Audio Forensic, Theory and Practice*, Denver, CO, USA, June 2008.

[7] D. Nicolalde and J. A. Apolinário Jr., "Evaluating digital audio authenticity with spectral distances and ENF phase change," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Taipei, Taiwan, April 2009.

[8] A. Martin, G. Doddington, T. Kamm, M. Ordowski, M. Przybocki, "The DET curve in assessment of detection task performance," *European Conference on Speech Communication and Technology*, Rhodes, Greece, September 1997.

[9] J. Ortega-García, J. González-Rodríguez, and V. Marrero-Aguiar, "AHUMADA, A large speech corpus in spanish for speaker characterization and identification," *Elsevier Speech Communication*, vol. 31, pp. 255–264, June 2000.