ANÁLISE DE ESPECTRO ATRAVÉS DA DETECÇÃO DE EVENTOS ACÚSTICOS ELEMENTARES NO PLANO TEMPO-FREQÜÊNCIA

Christiane Raulino*, Dami Duarte*, Jugurta Montalvão*

* *Universidade Federal de Sergipe (UFS)*
São Cristóvão, Sergipe, Brazil

Emails: chrisraulino@gprufs.org, dami_doria@hotmail.com, jmontalvao@ufs.br

Abstract— An event-based method is presented as an alternative for spectral analysis of acoustic signals. The targeted event is upward level-crossings, from which three essential information are taken: instantaneous frequency between consecutive events, corresponding maximum amplitude and instant of occurrence of each event. This approach is similar to EIH and ZCPA, in this conception, but innovative in terms of spectral estimation, since a Monte Carlo is used instead of histograms. Some illustrations of properties are discussed, such as the intrinsic silence suppression and nonlinear scaling of time, along with experimental illustration through isolated word detection.

Keywords— Level-crossings, Monte Carlo Integral, ZCPA, EIH, MFCC.

Resumo— Um método para análise espectral de sinais acústicos baseada em eventos é apresentado. O evento alvo, nesse método, é o cruzamento ascendente de nível, do qual são extraídas três informações essenciais: a frequência instantânea entre eventos consecutivos, a amplitude correspondentes, e o instante de ocorrência de cada evento. Esse método é similar ao EIH e ao ZCPA, na sua concepção, mas é inovativo na forma como as informações são consolidadas na estimação do espectrograma, onde uma integral de Monte Carlo é usada em lugar de histogramas. Algumas ilustrações de propriedades são discutidas, tais como a supressão intrínseca dos silêncios e a distorção não-linear da escala temporal, juntamente com uma ilustração experimental baseada em detecção de palavras isoladas.

Palavras-chave— Cruzamento de limiares, Integral de Monte Carlo, ZCPA, EIH, MFCC.

1 Introdução

A análise espectral de um sinal é uma ferramenta antiga e importante. Em experiências feitas com prismas, Newton criou, em 1704, o conceito de *spectrum* ao decompor a luz branca em suas cores com diferentes comprimentos de onda. A reboque desse novo conceito, surgiram trabalhos a fim de estudar materiais e sinais através de suas componentes espectrais (Robinson, 1982). No entanto, somente após o trabalho de Jean Baptiste Joseph Fourier, ficou concretizada a teoria matemática de que formas de onda quaisquer podem ser representadas pela uma soma de funções harmônicas ponderadas.

Essa nova forma de representação do sinal proporcionou o surgimento de técnicas para tratamento desses dados no domínio da frequência. Desde então, formas de onda como voz humana, eletrocardiogramas, vibrações mecânicas, entre outras, são analisadas e processadas através de seus espectros. Particularmente, para sinais de voz, é interessante analisar visualmente a evolução do espectro ao longo do tempo, através do espectrograma. Por exemplo, no campo da fonoaudiologia, o espectrograma é utilizado como uma fotografia da fala, que auxilia na correção de sua pronuncia. O método computacional clássico para o cálculo e análise deste gráfico é feito através da Transformada de Fourier em pequenos intervalos fixos de tempo (*Short-Time Fourier transform - STFT*).

Mais especificamente no domínio de sinais acústicos, como a voz humana, há diferentes métodos de análise baseadas nas características do sistema auditivo humano. Tais métodos, ao simular tarefas executadas pelo aparelho auditivo humano, fornecem bom desempenho, por exemplo, em termos de robustez em tarefas de reconhecimento de fala.

Em (Davis and Mermelstein, 1980), *Mel-Frequency Cepstral Coefficients* (MFCC) são apresentados como características de baixa dimensão, representando pequenos trechos de fala. Nesse trabalho, emprega-se o conceito de banda crítica (Fletcher, 1940), aplicando filtros triangulares no espectro do trecho do sinal. A percepção humana no domínio da frequência também é simulada com o uso da escala mel no posicionamento dos filtros. Desta forma, o MFCC apresenta bons resultados em sinais livres de ruído e ganhou popularidade entre os trabalhos de processamento de sinais.

O *Ensemble Interval Histogram* (EIH) (Ghitza, 1994), desenvolvido por Oded Ghitza, nasce a partir de uma análise minuciosa de como o sinal sonoro percorre cada parte do sistema auditivo, e desta análise se destacam dois processos importantes: o modelo do movimento mecânico da membrana basilar da cóclea, modelado como um banco de filtros não-lineares; e o acionamento dos nervos das células ciliadas internas da cóclea, modelados como detectores de cruzamento por limiar. Ghitza aplicou um modelo de filtro condizente com a estrutura fisiológica do ouvido interno. Dessa forma, este modelo age como um banco de filtros passa-faixa, sintonizados em

190 canais diferentes, distribuídos entre 200 Hz e 7000 Hz, resultando em 190 sinais de banda estreita, o que os prepara para a aplicação de detectores de cruzamento por limiar. Sendo assim, a distância entre dois cruzamentos ascendentes fornece a informação de período instantâneo de uma hipotética componente harmônica daquela banda. Para se obter informação complementar de amplitude dessa harmônica, são utilizados cinco níveis diferentes de limiares. Finalmente, elabora-se um histograma (EIH) a partir das informações de período (logo, frequência) e amplitude estimados, que equivale a uma espécie de espectro aproximado (e truncado) do sinal analisado.

O *Zero-Crossings with Peak Amplitudes* (ZCPA) (Kim et al., 1999) é um extrator de características que possui uma abordagem semelhante ao EIH, tendo como estrutura um banco de filtros cocleares seguido de detectores de cruzamentos de nível, que fornecem informações relevantes à composição de um histograma. Sua principal diferença do EIH se dá pelo fato de que o ZCPA utiliza apenas um detector de cruzamento por zero e, para determinar a informação de amplitude, detecta o valor máximo do sinal entre dois cruzamentos por zero. O autor mostra ainda, em (Kim et al., 1999), que o ZCPA obtém melhores taxas de reconhecimento contra outros extratores de características, como o EIH e o MFCC, para sinais coletados em diversos ambientes ruidosos.

De fato, utilizar a informação de cruzamentos por limiar gera boas representações do sinal. Contudo, a ideia de que grande parte da informação está contida nos instantes destes cruzamentos antecede estes trabalhos (Kim et al., 1999) (Ghitza, 1994), como observado em (Kedem, 1986). Uma boa evidência disso, conhecida desde antes de 1948 (Licklider et al., 2012), é que sinais de fala podem ser claramente compreendidos mesmo quando são alterados pela distorção de *clipping infinito*, que codifica esses sinais como seqüências de ± 1, ou seja, sinais nos quais apenas as informações de cruzamento por zero são preservadas.

Nesse contexto, este artigo propõe uma abordagem probabilista para elaboração de espectrogramas de sinais acústicos, e conseqüente extração de suas características. Os resultados deste trabalho são apresentados neste texto como segue: na segunda seção, é descrita o conceito de Eventos Acústicos Elementares (EAE), juntamente com um método simples para sua obtenção; na terceira seção, estes eventos são analisados como ocorrências de uma variável aleatória, da qual estimamos o espectro como uma função de densidade de probabilidade relacionada aos EAE que representam sinal de entrada, através da integral de Monte Carlo. Também nessa seção, é sugerido um novo espaço característico de baixa dimensão para representação de trechos de voz com base nos eventos acústicos. Na quarta seção, é computado e analisado o espectrograma guiado pelos EAE. A quinta seção apresenta a base utilizada para os experimentos realizados, o pré-processamento dos sinais e os detalhes da extração de características. Finalmente, a sexta sessão é dedicada aos resultados do experimentos.

2 Eventos acústicos elementares

O método aqui proposto tem como objetivo obter informações de frequência, ao longo do tempo, de um dado sinal, a partir da coleta seletiva de eventos de detectores de cruzamentos por nível. Todo processo é baseado em modelos simplificados do sistema auditivo humano, descritos consideravelmente em detalhes em (Ghitza, 1994) e (Kim et al., 1999).

Assim como o EIH e o ZCPA, a ideia de desenvolver o método baseado em EAE partiu da análise do sistema auditivo humano, mas usa uma versão muito simplificada dos modelos conhecidos, a saber: Primeiramente é utilizado um filtro de pré-ênfase (função de transferência: $H(z) = 1 - 0,97z^{-1}$). O banco de filtros é montado com 17 FIR (*Finite Impulse Response*) passa-faixas de ordem 100, centrados em 150, 250, 350, 450, 570, 700, 840, 1000, 1170, 1370, 1600, 1850, 2150, 2500, 2900, 3400 e 4000 Hz, respectivamente (i.e. igualmente espaçados na escala Bark) e com bandas iguais à metade de suas respectivas frequências centrais. Filtros FIR foram escolhidos neste método porque tiramos proveito da análise feita em (Kim et al., 1999), para uso no ZCPA, onde se mostra que filtros FIR mais simples, em detrimento dos filtros TW propostos, apresentaram melhores desempenhos. Para diminuir o efeitos de lóbulos laterais no espectro dos filtros, típicos do janelamento retangular, é utilizada a janela de Blackman.

Em seguida, a saída de cada filtro é analisada utilizando detectores de cruzamento por nível ($\lambda_d = \frac{1}{100} \cdot \sum_i h[i]^2$, onde h representa a resposta ao impulso do filtro da banda analisada). O intervalo de tempo, T_n, entre dois instantes, t_n e t_{n+1}, de cruzamentos ascendentes por limiar, λ_d, é linearmente interpolado para reduzir o efeito de truncamento dos instantes de amostragem. De modo semelhante ao método ZCPA, determina-se também a amplitude máxima (em módulo), A_n, associada ao n-ésimo segmento de sinal entre cruzamentos ascendentes. Finalmente, o n-ésimo Evento Acústico Elementar é codificado como um vetor numérico, contendo três medidas, a saber:

$$EAE_n = [t_n \ T_n \ A_n]$$

A obtenção dos EAE e a representação deles no plano tempo-frequência ($t_n \times (1/T_n)$) são ilustrados nas figuras 1 e 2.

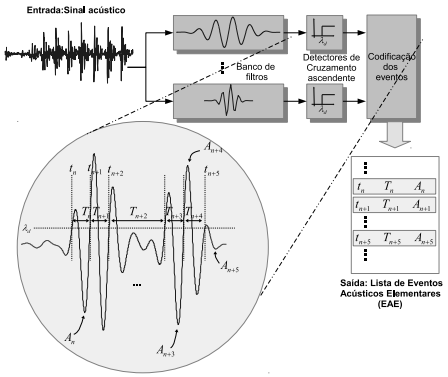

Figura 1: Ilustração do processamento para extração de Eventos Acústicos Elementares (EAE).

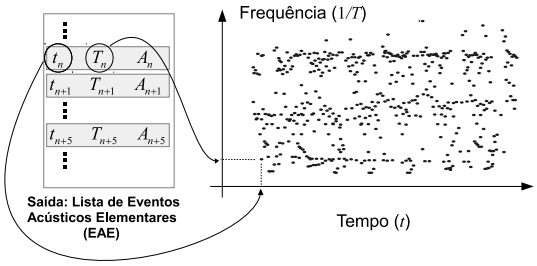

Figura 2: Ilustração do mapeamento de Eventos Acústicos Elementares(EAE) no plano tempo-frequência.

3 Obtenção do contorno espectral

A simples visualização dos eventos como pontos no plano tempo-frequência, sem mesmo considerar a informação de amplitude A_n, já fornece uma espécie de espectrograma (ver figura 3). No entanto, o contorno espectral do sinal, ao longo do tempo, continua desconhecido. Felizmente, esse contorno corresponde à função densidade de probabilidade (PDF) da variável aleatória, $X(t)$, subjacente à geração de EAE no instante t. É sabido que o valor esperado de uma função arbitrária de uma variável aleatória X, $g(X)$, em relação à função densidade de probabilidade $f_X(x)$, é dado por:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx = \langle g, f_X \rangle$$

Usando o método de Monte Carlo, podemos encontrar uma aproximação desta integral, para um número N de instâncias da variável aleatória X, como $\langle g, f_X \rangle = \frac{1}{N} \sum_{i=1}^{N} g(x_i)$, o que pode ser entendido como uma média ponderada das N instâncias disponíveis. Mas vale notar que a escolha da função $g(\cdot)$ é arbitrária, e pode até gerar ponderações negativas nesse cálculo de "média" ponderada. Como veremos mais adiante, a escolha de funções cossenoidais para $g(\cdot)$ é fundamental para o método proposto.

Assim, definindo $x_n^{Hz} = 1/T_n$ como um estimador da frequência instantânea, em Hertz, associada ao n-ésimo EAE, e $x_n^{Mel} = 2595 \log_{10}(1 + x_n^{Hz}/700)$ como o equivalente na escala Mel, pode-se determinar o contorno espectral de um conjunto de EAEs pela decomposição em funções ortogonais de f_X, pelo método de Monte Carlo, seguida de sua recomposição vetorial através dos coeficientes correspondentes. Escolhemos usar como base de decomposição as 12 primeiras funções cossenoidais usadas na Transformada Cosseno Discreta (DCT), onde $g_k(x_n^{Mel}) = \cos(k(2x_n^{Mel} + 1)\pi/(2f_{Max}))$ corresponde à k-ésima função de base, com $k = 1, 2, \ldots, 12$, e $f_{Max} = 2146$ Mels (equivalentes a 4000 Hz).

Visando os contornos suavizados do espectro, f_X,assim como no caso dos coeficientes cepstrais, usamos apenas as 12 primeiras funções de base, o que corresponde a uma limitação na dimensão do espaço de representação do espectro.

Finalmente, o algoritmo correspondente ao método de análise espectral proposto é o seguinte:

- Pré-enfatizar o sinal;
- Utilizar o banco de filtros e, para cada saída, determinar os EAE;
- Segmentar sequencialmente os EAE em blocos de $N = 500$ eventos, com avanços de 100 em 100;
- Utilizar a integral de Monte-Carlo para obter os 12 coeficientes DCT que representam o espectro (aqui termina o algoritmo para extração de vetores de características com dimensão 12);
- Recompor a PDF com os coeficientes DCT calculados para cada bloco de N EAE, e organizar as PDFs estimadas como colunas sequenciais da matriz que representa o espectrograma.

4 Espectrograma guiado por EAE

A figura 3 apresenta uma comparação visual entre espetrograma convencional e o conjunto de EAE (eventos elementares), representados por pontos no plano tempo (t_n, tempo de ocorrência do pacote) *versus* frequência estimada (inverso da duração T_n). Assim como no espectrograma, também na "nuvem" de pontos formada pelos EAE podemos perceber, na forma de variações de densidades de pontos, a evolução das formantes vocálicas correspondentes às três palavras pronunciadas em inglês.

Na figura 4, através da comparação do espectrograma e do equivalente obtido por integração de Monte-Carlo de blocos consecutivos de $N = 500$ EAE ao longo do tempo, fica evidente o efeito de supressão de silêncios provocado pela estratégia aqui proposta. De fato, nos intervalos de silêncio quase não há detecção de eventos acústicos, logo não há produção de EAE (detecção de eventos ou cruzamentos de limiares), ao passo que, na presença de sons ricos em detalhes, sobretudo aqueles que contém componentes em altas frequências, a detecção de eventos atinge taxas altas. Em outras palavras, uma consequência natural do método proposto, baseado em detecção

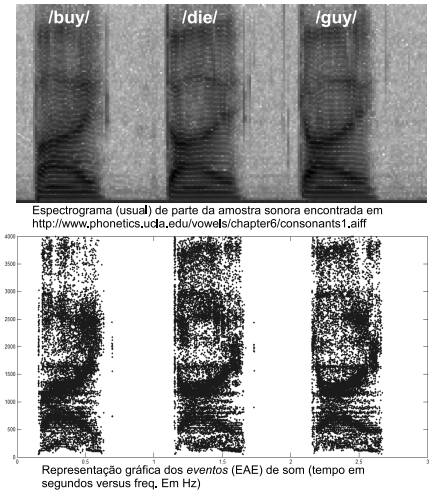

Figura 3: Comparação visual entre espetrograma convencional e o conjunto de EAE, representados por pontos no plano tempo × frequência.

de eventos, é a de que a escala temporal de representação dos sons se deforma não-linearmente, de forma similar ao que seria obtido através de *Dynamic Time Warping* (DTW), mas com a diferença de que, aqui, não há uma referência de alinhamento.

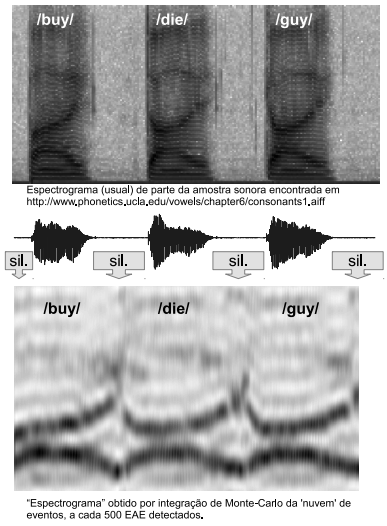

Figura 4: Ilustração do efeito de supressão dos intervalos de silêncio através da deformação não-linear da escala temporal - maior taxa de eventos acústicos detectados implica maior resolução temporal.

5 Base de dados, pré-processamento e extração de características

Os teste relatados aqui foram realizados com uma base de 5 palavras curtas, pronunciadas 10 vezes por cada um dos 8 oradores voluntários (6 homens e 2 mulheres)[1]. As amostras foram coletadas em ambientes não controlados, como domicílios e salas de aulas, numa taxa de 8000 amostras por segundo, e quantização de 16 bits por amostra. A aquisição das amostras foi feita com dispositivos móveis (como *smartphones*), usando seus respectivos microfones embutidos. As palavras pronunciadas por cada orador são os comandos, em português: *avance*, *direita*, *esquerda*, *pare* e *recue*.

Cada sinal foi sistematicamente pré-enfatizado (filtro: $H(z) = 1 - 0.97z^{-}1$). Na obtenção dos MFCC, cada sinal foi segmentado em blocos de 256 amostras, com avanço de 80 entre blocos consecutivos (superposição de $\approx 70\%$). Cada bloco foi então atenuado através de uma janela de Hamming e, finalmente, mapeado em 13 coeficientes cepstrais (13 MFCC). Isto é, cada sinal x, com N_x amostras, foi mapeado em uma matriz com 13 linhas e, aproximadamente, $(N_x - 256)/80$ colunas.

Ainda no caso dos MFCC, o vetor correspondente à primeira linha da matriz de coeficientes é o coeficiente de ordem zero, que expressa a energia acústica de cada bloco de sinal, e que foi sistematicamente descartado da montagem da matriz de características, X_{MFCC}. De fato, o perfil de energia foi apenas usado no processo de pré-alinhamento temporal, explicado em (Montalvão et al., 2012).

Quanto ao ZCPA, usamos 16 filtros FIR igualmente espaçados na escala Bark, entre 200 Hz e 4000 Hz, gerando histogramas com 40 bins também igualmente espaçados na escala Bark, que são transformados (DCT) em 12 coeficientes, organizados como colunas da matriz de características X_{ZCPA}. A rigor, isso corresponde ao ZCPAC (Kim et al., 1999), que empiricamente notamos fornecer melhores resultados que o ZCPA de base.

No caso do método alternativo aqui estudado, cada sinal foi processado gerando um conjunto de EAE. A cada 500 eventos detectados (com avanço de 100 eventos entre blocos), uma integral de Monte Carlo foi usada para gerar 12 coeficientes que representam o contorno espectral suave da janela de sinal correspondente aos 500 EAE. Em seguida, esses coeficientes foram agregados na forma de colunas da matriz de características X_{EAE}.

Sublinhamos também que todos os métodos foram cuidadosamente ajustados para fornecerem os melhores resultados empíricos nas tarefas relatadas aqui.

Em cada caso, as matrizes de características, X_{MFCC}, X_{ZCPA} e X_{EAE}, juntamente com os perfis de energia correspondentes, foram entregues, duas-a-duas, como entradas ao processo de alinhamento temporal, sendo que uma das matrizes foi assumida como sinal de exemplo, ou uma referência da classe de comandos, enquanto que a outra foi tomada como um sinal desconhecido, a

[1]Base disponível para *download* em http://www.biochaves.com/en/download.htm.

ser detectado como sendo ou não da mesma classe do sinal de referência. Vale notar que não se trata de tarefa de classificação, mas de detecção, o que nos permitiu representar os resultados na forma compacta de *Equal Error Rate* EER, que mede o desempenho aproximado de cada detector – quanto menor o EER, melhor o detector. Também é importante ressaltar que o nosso objetivo não é o reconhecimento de fala, para o qual há uma extensa bibliografia que propositalmente não foi considerada aqui. De fato, reduzimos o complexo problema do reconhecimento de fala em um simples problema de detecção, com base em um único exemplo de sinal por vez, no intuito de ressaltar os defeitos e qualidades dos três métodos de extração de características comparados.

6 Resultados Experimentais

Para simular uma situação em que um orador fornece apenas uma amostra de cada comando (restrição extrema que escolhemos impor ao reconhecedor de comandos), uma única amostra de voz foi selecionada aleatoriamente, por vez, e separada como sendo 'a referência de treinamento', juntamente com os rótulos representando o comando pronunciado. Em seguida, as demais amostras da base foram escolhidas aleatoriamente e comparadas, uma-a-uma, à referência, gerando medidas de similaridades registradas e testadas contra os limiares de detecção.

Para uma apresentação sucinta dos desempenhos comparados, optamos por uma busca exaustiva dos limiares de detecção até que as taxas de falsos positivos e falsas rejeições se igualassem, em cada sessão de testes, sendo essa medida tomada como o EER estimado em cada experimento. A figura 5 ilustra esse processo.

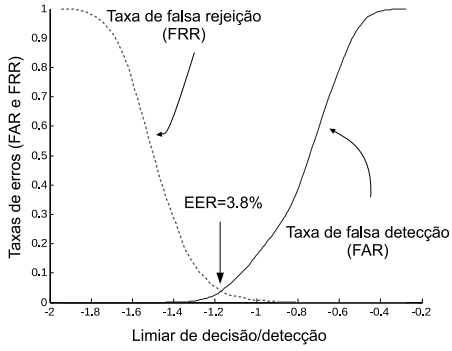

Figura 5: Representação visual das taxas de erros da detecção baseada em Eventos Acústicos Elementares(EAE), para sinais limpos.

Também, para abreviar a apresentação de resultados, destacamos apenas as tarefas de detecção mais difíceis. Isto é, apenas amostras do usuário de referência foram sorteadas em cada simulação, de forma a considerar apenas os casos mais difíceis para o detector de comandos, forçando-o a trabalhar sempre com o mesmo timbre de voz (mesmo orador) para cada par de sons comparados.

Adicionalmente, para cada par de amostra, utilizamos o método de alinhamento de referência (DTW) com as restrições propostas por Itakura, em 1975 (Itakura, 1975), precedido de um pré-alinhamento baseado em perfil de energia, cujos detalhes e motivações podem ser encontrados em (Montalvão et al., 2012). A tabela 1 apresenta os resultados médios obtidos.

Tabela 1: Resultados (EER) no reconhecimento de comandos.

Método	Limpo	SNR=6 dB
MFCC	7.3%	19.8%
ZCPA	8.2%	16.6%
EAE	3.8%	14.3%

Uma segunda bateria de experimentos foi realizada para investigar o efeito da escolha do limiar de detecção de eventos, λ_d, sobre a robustez das características extraídas frente ao ruído aditivo. Na tabela 2, são apresentados os EER para três valores de λ_d, para uma relação sinal/ruído (SNR) de 6 dB. É importante notar que, na geração da referência a partir do sinal limpo, usamos sempre o mesmo λ_d ajustado para a Tabela 1, definido na seção 2, e que apenas no processamento do sinal de teste (impregnado com ruído) é que elevamos o limiar de detecção de eventos para $5\lambda_d$ ou $10\lambda_d$. Esses experimentos simples nos permitem notar que a robustez do método pode ser sensivelmente incrementada pelo ajuste desse limiar ao nível de ruído.

Tabela 2: Variação do EER com o limiar de detecção de cruzamento (SNR=6 dB).

—	λ_d	$5\lambda_d$	$10\lambda_d$
EER →	14.3%	14.1%	14.5%

7 Conclusão e discussões

O método proposto e estudado neste artigo herda do EIH seu aspecto fundamental de detecção de eventos relevantes pelo cruzamento por limiar (limiar diferente de zero). Paralelamente, do ZCPA – que pode ser visto como uma simplificação computacional do EIH –, a nossa proposta herda a valorização de cada par de cruzamento ascendente por limiar como uma peça de informação relevante na extração da característica.

Por consequência, o método estudado neste trabalho também herda o benefício principal de

ambos, EIH e ZCPA, se comparados ao MFCC: a robustez a ruídos. Esse resultado já era esperado, e não representa a maior contribuição deste trabalho. De fato, o nosso foco está voltado para a diferença entre o nosso método e os métodos EIH e ZCPA, a saber: nós não usamos histogramas como estimadores espectrais. Isto é, na nossa abordagem, o contorno suave do comportamento estocástico dos eventos detectados, no plano tempo-frequência, é extraído através de uma integral de Monte Carlo.

Nesse processo de integração probabilística, obtivemos vetores de características com dimensões e conteúdo "equivalentes" aos coeficientes cepstrais, além disso, no intuito de mostrar que esses coeficientes realmente portam as informações relevantes do espectrograma de sinais de voz, re-projetamos os coeficientes no espaço original (como na figura 4), onde se pode perceber visualmente as evoluções das formantes da voz.

Outro diferencial importante entre o método proposto e o ZCPA é a supressão (nos experimentos) da informação de amplitude. Isto é, assumimos a hipótese de partida de que apenas a posição do evento detectado no plano tempo-frequência já contém as informações mais relevantes do sinal sonoro, e que o descarte da informação de amplitude associada a cada evento detectado poderia ser inclusive um fator positivo no incremento da robustez do processamento. Essa expectativa foi confirmada nos resultados apresentados na tabela 1, onde o EAE claramente superou em robustez o ZCPA.

Os experimentos com detecção de comando confirmaram a superioridade do MFCC sobre o ZCPA, no caso de sinais limpos (já esperado), e também notamos uma inversão desse quadro quando ruído foi adicionado ao sinal limpo (novamente esperado). Paralelamente, é notável a superioridade dos resultados obtidos pelos EAE em todos os casos. Adicionalmente, também observamos algo novo em nossos estudos: a robustez do método pode ser incrementada pelo ajuste do limiar de detecção de EAE, λ_d, [2] ao nível de ruído, como evidenciado na Tabela 2. Isso nos remete ao EIH, que utiliza vários limiares ao mesmo tempo, o que finalmente evidencia a relevância indireta do uso da amplitude como informação extra na detecção/seleção dos eventos.

Assim, esses resultados parecem indicar que a seleção dos eventos (EAE) mais relevantes pode ser um aspecto chave para a extração ainda mais robusta de características. Mas, nesse caso, a questão que se impõe é: como atribuir níveis de importância diferenciados aos eventos? O ZCPA "responde" essa questão com a inclusão explícita da informação de amplitude... mas essa solução é questionável, como evidenciano nos nosso resultados experimentais, pois torna o método muito sensível à distribuição de amplitudes do ruído, além de não levar em conta informações temporais sobre a evolução dos eventos detectados (e.g. ritmo ou regularidade temporal).

Na continuação deste trabalho, já iniciada, estudaremos justamente estratégias de seleção dos eventos mais relevantes que sejam alternativas ou complementares ao simples uso da informação de amplitude, levando em conta o aspecto estocástico (dependência temporal/memória) da detecção dos EAE.

[2] Não confundir o limiar de detecção dos EAE com os limiares usados na comparação de sinais e obtenção do EER.

Referências

Davis, S. B. and Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences, *IEEE Transactions on Acoustics, Speech, and Signal Processing* **28**.

Fletcher, H. (1940). Auditory patterns, *Reviews of Modern Physics* **12**.

Ghitza, O. (1994). Auditory models and human performance in tasks related to speech coding and speech recognition., *IEEE Transactions on Speech and Audio Processing* **2**.

Kedem, B. (1986). Spectral analysis and discrimination by zero-crossings, *Proc. IEEE* **74**.

Kim, D.-S., Lee, S.-Y. and Kil, R. M. (1999). Auditory processing of speech signals for robust speech recognition in real-world noisy environments., *IEEE Transactions on Speech and Audio Processing* **7**.

Robinson, E. A. (1982). A Historical Perspective of Spectrum Estimation, *Proceedings of the IEEE* **70**.

Montalvão, J., M. V. P. Montalvão and C. Raulino. (2012). Detecção de orador e palavras em televigilância médica com treinamento mínimo: uma amostra por palavra, in: Proceedings of the XIX Congresso Brasileiro de Automática (CBA), 2012, Campina Grande-PB, Brasil, 2244–2250.

Itakura, F. (1975). Minimum prediction residual applied to speech recognition, in: IEEE Trans. Acoustics, Speech, 1975, ASSP-23, No. 1. 67–72.

Licklider, J. C. R. and I. Pollack. (1948). Effects of Differentiation, Integration, and Infinite Peak Clipping upon the Intelligibility of Speech, in: J. Acoust. Soc. Am. Volume 20, Issue 1, pp. 42-51 (1948);