

Uma Avaliação da Análise Cepstral Generalizada Aplicada ao Espectro de Vogais

Ramiro Barreira e Fábio Violaro

Resumo- A análise cepstral generalizada é uma abordagem unificada na qual a atribuição de valor a um parâmetro γ pode produzir a análise LP ($\gamma=-1$) e a análise cepstral ($\gamma=0$), bem como análises com perfis intermediários ($-1<\gamma<0$). Aplicando a análise cepstral generalizada a sinais de vogais distintas, identificamos condições espectrais críticas para o modelo e avaliamos seu desempenho sob tais condições.

Palavras-Chave- Análise Cepstral Generalizada

Abstract- Generalized Cepstral Analysis is a unified approach in which the assignment of a γ parameter value can produce LP analysis ($\gamma=-1$) and cepstral analysis ($\gamma=0$), as well as analysis with intermediate shapes ($-1<\gamma<0$). By applying generalized cepstral analysis to distinct vowel speech signals, we identified critical spectral conditions for the model and evaluated its performance under such conditions.

Keywords- Generalized Cepstral Analysis

I. INTRODUÇÃO

As análises LP (*Linear Prediction* [1]) e cepstral (*Unbiased Estimation of Log Spectrum* [8][11][2]) são extensamente utilizadas na área de fala para propósitos diversos tais como: codificação, análise e síntese, sendo a análise cepstral implementada na escala mel de frequências (análise mel-cepstral[12]) de uso central em conversão texto-fala via HMM. A abordagem unificada denominada análise cepstral generalizada também tem sido usada com sucesso nas áreas citadas, principalmente em sua implementação na escala mel de frequências.

Em geral deseja-se obter de tais análises uma curva espectral que represente bem as características do trato vocal com o mínimo de coeficientes, de maneira que sejam removidas as interferências produzidas pela fonte (glote) no espectro de fala.

A análise cepstral é classicamente tratada com base na teoria de processamento homomórfico de sinais [15], onde a suavização do espectro é obtida aplicando-se um filtro passa-baixas no domínio cepstral (*lifragem em quefrência*), definido pelos coeficientes de Fourier do espectro logarítmico do sinal em análise. Um refinamento de tal método, proposto em [8] e [11] e denominado UELS (*Unbiased Estimation of Log Spectrum*) pode ser concebido como um método de filtragem inversa [10], onde os coeficientes cepstrais são determinados por um critério de minimização da média quadrática da saída do filtro inverso (erro de predição). Por ter base no espectro

logarítmico, o modelo corresponde a um filtro definido pela exponencial de um filtro FIR com N coeficientes cepstrais, onde N é a ordem da análise. A filtragem inversa, de uso comum na obtenção de coeficientes LP, também pode ser aplicada com base no espectro logarítmico generalizado, o que conduz o modelo a um filtro exponencial generalizado, função cuja expressão é determinada pelo valor atribuído a um parâmetro γ . Para $\gamma=-1$, o filtro exponencial generalizado se reduz ao filtro LP, e para $\gamma=0$, o filtro exponencial padrão é obtido (análise cepstral). O parâmetro γ pode variar continuamente entre -1 e 0 , produzindo análises com perfis intermediários aos das análises LP e cepstral.

Quanto às suas características, as envoltórias espectrais obtidas variando-se o parâmetro γ diferem umas das outras pela sua capacidade de ajuste a picos e vales. A curva da análise LP ($\gamma=-1$) é classicamente conhecida por se ajustar bem aos picos espectrais com o porém de ser deficiente na modelagem de vales acentuados. Isso beneficia a modelagem de formantes às custas de prejudicar anti-formantes (presentes em caso de nasalização). Para o caso cepstral ($\gamma=0$), o modelo produz ajuste equivalente para picos e vales, propiciando picos mais largos e vales mais estreitos do que o caso LP, melhorando assim a modelagem de anti-formantes. Para os modelos espectrais obtidos com valores de γ entre -1 e 0 , observam-se características intermediárias às mencionadas para os casos LP e cepstral.

Este trabalho apresenta a formulação do modelo para a análise cepstral generalizada, o critério para a obtenção dos coeficientes cepstrais generalizados e os perfis espectrais associados aos valores de γ . Também são identificadas as características espectrais (de vogais) críticas para o modelo, sendo o modelo avaliado em exemplos quanto ao seu desempenho sob tais condições.

II. ANÁLISE CEPSTRAL GENERALIZADA

A. Modelo

A função logarítmica generalizada é definida por

$$\ln_{\gamma}(a) = \begin{cases} \frac{a^{\gamma} - 1}{\gamma} & 0 < |\gamma| \leq 1 \\ \ln(a) & \gamma = 0 \end{cases} \quad (1)$$

onde $\ln(a)$ é obtido através de $\lim_{\gamma \rightarrow 0} (a^{\gamma} - 1)/\gamma$. A função inversa do logaritmo generalizado é dada pela exponencial generalizada

$$\exp_{\gamma}(a) = \begin{cases} (1 + \gamma \cdot a)^{1/\gamma} & 0 < |\gamma| \leq 1 \\ e^a & \gamma = 0 \end{cases} \quad (2)$$

sendo e^a obtido através de $\lim_{\gamma \rightarrow 0} (1 + \gamma \cdot a)^{1/\gamma}$. O cepstrum generalizado de um espectro $X(e^{j\omega})$ é definido por

$$\bar{c}_{\gamma}(n) = \int_{-\pi}^{\pi} \ln_{\gamma}(X(e^{j\omega})) e^{j\omega n} d\omega \quad (3)$$

portanto

$$X(e^{j\omega}) = \exp_{\gamma} \left(\sum_{n=-\infty}^{\infty} \bar{c}_{\gamma}(n) e^{-j\omega n} \right) \quad (4)$$

O modelo para a análise cepstral generalizada de ordem N é obtido através de $N+1$ coeficientes cepstrais generalizados na forma

$$H_{\gamma}(z) = \exp_{\gamma} \left(\sum_{n=0}^N c_{\gamma}(n) z^{-n} \right) \quad (5)$$

para o qual assumiremos $-1 \leq \gamma \leq 0$. É conveniente representar $H_{\gamma}(z)$ através de

$$H_{\gamma}(z) = K_{\gamma} \cdot D_{\gamma}(z) \quad (6)$$

onde K_{γ} é a constante de ganho dada por

$$K_{\gamma} = \exp_{\gamma}(c_{\gamma}(0)) \quad (7)$$

e $D_{\gamma}(z)$ é o filtro normalizado dado pela expressão

$$D_{\gamma}(z) = \exp_{\gamma} \left(\sum_{n=1}^N c_{\gamma}'(n) z^{-n} \right) \quad (8)$$

onde os coeficientes $c_{\gamma}'(n)$ estão relacionados com $c_{\gamma}(n)$ através de

$$c_{\gamma}'(n) = \frac{c_{\gamma}(n)}{1 + \gamma c_{\gamma}(0)} \quad (9)$$

Para $\gamma = -1$, o filtro $H_{\gamma}(z)$ toma a seguinte forma:

$$H_{-1}(z) = \frac{K_{-1}}{1 - \sum_{n=1}^N c'_{-1}(n) z^{-n}} \quad (10)$$

que é o modelo para a análise LP. Para $\gamma = 0$, a exponencial generalizada se reduz à exponencial padrão fornecendo o modelo para a análise cepstral:

$$H_0(z) = K_0 \cdot e^{\sum_{n=1}^N c_0(n) z^{-n}} = e^{\sum_{n=0}^N c_0(n) z^{-n}} \quad (11)$$

Como se observa, a análise cepstral generalizada engloba as análises LP e cepstral em uma única representação. O critério para a obtenção dos coeficientes será abordado a seguir.

B. Critério Espectral

Para a obtenção dos coeficientes, utiliza-se o critério de minimização da média quadrática da saída do filtro inverso $D_{\gamma}^{-1}(z)$, cuja entrada é o sinal de fala sob análise. O tratamento para os sinais envolvidos é dado no âmbito de suas densidades espectrais de potência (DEP).

A média quadrática da saída do filtro inverso (erro quadrático médio) é obtida a partir da DEP da entrada $x(n)$ e do módulo quadrático da resposta em frequência do filtro inverso ($|D_{\gamma}^{-1}(e^{j\omega})|^2$). A DEP da entrada pode ser obtida através do periodograma modificado:

$$I_N(\omega) = \left| \sum_{n=0}^{M-1} w(n) x(n) e^{-j\omega n} \right|^2 / \sum_{n=0}^{M-1} w^2(n) \quad (12)$$

onde $x(n)$ é o sinal de entrada com M amostras e $w(n)$ é uma janela à escolha (ex: janela de Blackman).

O produto $I_N(\omega) \cdot |D_{\gamma}^{-1}(e^{j\omega})|^2$ corresponde à DEP da saída do filtro inverso. O que se busca é o conjunto de coeficientes que minimizem o erro quadrático médio, dado por:

$$\varepsilon = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D_{\gamma}(e^{j\omega})|^2} d\omega \quad (13)$$

Para $-1 \leq \gamma \leq 0$, a Hessiana de ε é positiva com respeito aos coeficientes $\bar{c} = [c_{\gamma}'(1), c_{\gamma}'(2), \dots, c_{\gamma}'(N)]^T$, portanto

$$\nabla_{\bar{c}} \varepsilon = \frac{\partial \varepsilon}{\partial \bar{c}} = 0 \quad (14)$$

fornece o mínimo global [9][3]. A constante de ganho é obtida através de

$$K_{\gamma} = \sqrt{\varepsilon_{min}} \quad (15)$$

onde ε_{min} é o valor mínimo de ε obtido resolvendo-se (14). A equação (7) fornece o coeficiente $c_{\gamma}(0)$.

As referências [10] e [14] contém o método para a solução de (14).

III. PERFIL DO MODELO ESPECTRAL

A magnitude espectral em dB obtida através da análise cepstral generalizada é dada por

$$10 \log_{10} (|D_{\gamma}(e^{j\omega})|^2) = 10 \log_{10} [\exp_{\gamma}(f_{\gamma}(\omega))] \quad (16)$$

onde $f_\gamma(\omega)$ é um polinômio trigonométrico dado por

$$f_\gamma(\omega) = d_\gamma(0) + 2 \sum_{k=1}^N d_\gamma(k) \cos(\omega k) \quad (17)$$

onde $d_\gamma(k)$, por sua vez, está relacionado aos coeficientes $c_\gamma'(n)$ através de

$$d_\gamma(k) = \begin{cases} \gamma \sum_{l=1}^{N-k} c_\gamma'(l) c_\gamma'(l+k) & k=1 \\ c_\gamma'(k) + \gamma \sum_{l=1}^{N-k} c_\gamma'(l) c_\gamma'(l+k) & 1 < k < N \\ c_\gamma'(k) & k=N \end{cases} \quad (18)$$

Quando $\gamma = 0$, a exponencial generalizada em (16) é invertida pela função logarítmica e os coeficientes de $f_\gamma(\omega)$ assumem os valores $d_0(0) = 0$ e $d_0(k) = c_0'(k)$ (para $1 \leq k \leq N$), o que resulta em:

$$10 \log_{10} (|D_0(e^{j\omega})|^2) = 10 \log_{10}(e) \cdot f_0(\omega) \quad (19)$$

com

$$f_0(\omega) = 2 \sum_{k=1}^N c_0'(k) \cos(\omega k) \quad (20)$$

A figura 1 ilustra o comportamento da função $10 \log_{10}(\exp_\gamma(x))$ para três valores de γ . Para $\gamma = 0$, a curva é linear, e conforme γ é reduzido ($\gamma = -1/2$ e -1), as curvas ganham caráter não-linear e as variações em x resultam em variações progressivamente mais lentas pela função no intervalo $-1 < x < 0$ e mais rápidas no intervalo $0 < x < 1$.

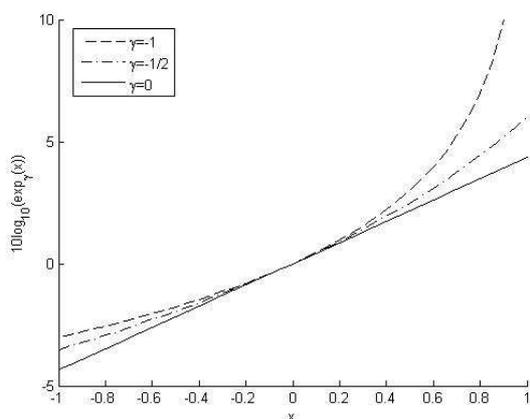


Fig. 1: Curvas para $\gamma = -1$, $\gamma = -1/2$ e $\gamma = 0$ da função $10 \log_{10}(\exp_\gamma(x))$. Para $\gamma = 0$ a função se torna linear na forma $10 \log_{10}(e) \cdot x$.

Se adotarmos $x = f_\gamma(\omega)$, a função assume a forma da expressão (16). Como exemplo, assumindo que $f_\gamma(\omega)$ possui somente o Harmônio $k=5$ e coeficientes fixos (à escolha), a equação (16) resulta em curvas semelhantes às da figura 2.

Quanto maior a ordem N , mais flexibilidade a curva de análise adquire para se moldar à curva do espectro original, seja qual for o valor de γ . Como N em geral é baixo, a curva de análise tende a adquirir um perfil característico do valor de γ . Assim o uso de $f_\gamma(\omega)$ com somente um Harmônico no exemplo da figura 2 permite visualizar aspectos característicos de diferentes valores de γ para o modelo.

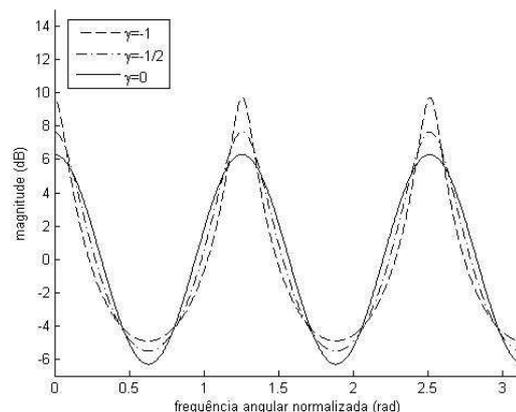


Fig. 2: Curvas para $\gamma = -1$, $\gamma = -1/2$ e $\gamma = 0$ da função $10 \log_{10}[\exp_\gamma(f_\gamma(\omega))]$ utilizando o harmônico $k=5$ de $f_\gamma(\omega)$.

Na figura 2, a curva para $\gamma = -1$ (caso LP), tem picos estreitos e vales notavelmente suaves. Com o aumento progressivo de γ , segue-se a tendência ao alargamento dos picos e à redução na suavidade dos vales, o que pode ser observado na curva para $\gamma = -1/2$ e que culmina na curva para $\gamma = 0$ (caso cepstral). Nesse caso, picos e vales têm suavidades equivalentes, o que é esperado, uma vez que a curva é uma senoide, proveniente da expressão (19) somente com o termo $k=5$ do somatório.

A característica de estreitamento dos picos e suavização dos vales quando γ se aproxima de -1 faz com que a análise para $\gamma = -1$ (LP) tenha as melhores condições para a modelagem de formantes, mas seja a mais problemática para a modelagem de vales acentuados no espectro (que muitas vezes associam-se a anti-formantes). Por outro lado, quando γ se aproxima de 0 (caso cepstral), a tendência de abertura dos picos e estreitamento dos vales beneficia fortemente a modelagem de vales acentuados, às custas de prejudicar a modelagem de formantes, podendo ocorrer ambigüidade na representação daqueles próximos entre si.

IV. CONDIÇÕES ESPECTRAIS CRÍTICAS PARA A ANÁLISE

A fim de se avaliar as características e o desempenho da análise cepstral generalizada, vogais do Português brasileiro foram alvo de análises. As análises foram

efetuadas com diversas ordens, a partir de uma janela de Hamming de 40ms aplicada às vozes de dois locutores adultos do sexo masculino e dois do sexo feminino, amostradas a 22050Hz. Para cada vogal, foram feitas análises para os treze valores consecutivos de γ distanciados entre si de 1/12 no intervalo fechado $[-1,0]$, e gerados os gráficos das envoltórias espectrais e da curva do erro quadrático médio em função de γ . A avaliação dos gráficos permitiu a identificação das características espectrais críticas para a modelagem. Tais características são apontadas, exemplificadas e discutidas a seguir.

O termo *espectro neutro* é utilizado neste trabalho para designar um espectro de fala no qual os formantes ocorrem aproximadamente equidistantes no eixo de frequências e sem a presença de vales acentuados.

Conforme se percorre γ de -1 até 0 , verificam-se, para as condições espectrais especificadas a seguir, as seguintes tendências:

- i. Espectro neutro: nesse caso, o erro tende a crescer uniformemente com γ , já que as boas condições para a modelagem de formantes associadas à análise LP ($\gamma = -1$) são reduzidas com o aumento de γ .
- ii. Espectro com presença de vales acentuados: Nesse caso, iniciando-se a análise em $\gamma = -1$ (onde se tem um vale com modelagem deficiente), conforme se aumenta γ , atinge-se um intervalo crítico de γ (que eventualmente pode cobrir todo o domínio $-1 \leq \gamma \leq 0$) onde a curva do erro tende a decrescer em função do progressivo ajuste do modelo ao vale. O final do intervalo crítico é determinado quando, enfim, a curva do modelo é ajustada ao vale.
- iii. Espectro com presença de formantes próximos entre si: Conforme se aumenta γ atinge-se um intervalo crítico (geralmente pequeno) com ponto de partida determinado pela condição em que o alargamento dos picos modelados desencadeia o início de uma “fusão” entre aqueles próximos entre si. A partir daí, a curva de erro cresce com γ conforme o conjunto de picos se funde. O crescimento é interrompido quando o conjunto forma um platô onde não mais se distinguem os formantes, condição em que deteriorações subseqüentes associadas ao platô (como a formação de um pico único) passam a ser redundantes, tendo pouca influência na curva do erro.

Os espectros associados às tendências **ii** e **iii** podem ser pensados como derivações do *espectro neutro*. A tendência **i** da curva do erro quadrático médio provém de uma condição espectral global (*espectro neutro*) e portanto se manifesta em todo o domínio γ , sofrendo interferências das tendências **ii** e **iii**, provenientes de condições espectrais locais e tendo portanto manifestações em intervalos menores do domínio γ . A tendência **ii** ocorre em intervalos mais extensos do que a tendência **iii**, podendo eventualmente se manifestar em todo o domínio γ . As tendências locais descritas em **ii** e **iii** aparecem sempre superpostas à tendência global **i** e, muitas vezes, superpostas entre si. Um exemplo é o caso em que dois conjuntos isolados de picos próximos entre si, sofrem “fusões” manifestando a tendência **iii** em intervalos

interseccionados de γ . A mesma idéia se aplica à tendência **ii** quando se tem a presença de mais de um vale acentuado no espectro.

As figuras de 3 a 6, referentes à análise de ordem 28 de uma das vozes masculinas, mostram as curvas do erro quadrático médio em função de γ para as vogais “é”, “e”, “o” e “u”. Em cada figura, no gráfico do erro estão marcados três pontos associados a valores de γ relevantes para as discussões. Os demais gráficos representam suas correspondentes envoltórias espectrais sobrepostas ao espectro do sinal de fala que as originou.

Na figura 3, vemos o espectro da vogal “é” onde, se não houvesse um vale acentuado em 4800Hz, teríamos algo próximo do espectro neutro. O progressivo ajuste ao vale provoca na curva do erro, o comportamento de queda descrito em **ii**, tendo como intervalo crítico todo o domínio $-1 \leq \gamma \leq 0$.

A análise efetuada para a vogal “e”, mostrada na figura 4, resultou em uma curva de erro que inicia em queda pela presença de um vale acentuado, como descrito em **ii**, e têm essa tendência invertida quando o modelo se ajusta ao vale, dando lugar à tendência de crescimento descrita em **i**.

Para a vogal “o” na figura 5, observa-se duas manifestações isoladas da tendência **iii**, uma no intervalo que vai de $\gamma = 0$ a $\gamma = -9/12$ (figura 5(c)), devido à deterioração do segundo par de formantes no eixo de frequências, e outra no intervalo que vai de $\gamma = -1/2$ a $\gamma = -3/12$ (figura 5(d)), em função da deterioração do primeiro par de formantes. As tendências se manifestam superpostas à tendência de queda descrita em **ii**, consequência do vale presente em torno da frequência de 4500Hz.

Para a vogal “u”, na figura 6, a curva do erro é perturbada no intervalo $-7/12 \leq \gamma \leq -3/12$ com o efeito conjunto da deterioração dos dois grupos isolados de pares de formantes próximos entre si contidos no espectro, efeito descrito pela superposição de duas manifestações da tendência **iii**.

As análises exemplificadas nas figuras foram aplicadas somente a vogais orais, em que vales espectrais acentuados, quando presentes, não são tão proeminentes quanto nas vogais nasais. No caso das nasais, vales acentuados ocorrem como manifestação de anti-formantes (anti-ressonâncias no trato vocal), de maneira que estão sempre presentes e são proeminentes, fazendo com que a curva do erro seja dominada pela tendência de queda descrita em **ii**.

Para um mesmo locutor, análises efetuadas a diversas ordens e aplicadas a vogais emitidas em diversas frequências fundamentais permitiram concluir que o conhecimento das características espectrais que influenciam na curva do erro não nos possibilita sua previsão a partir somente do conhecimento da vogal em análise. As curvas flutuam em função de mudanças na frequência fundamental, na ordem da análise, além de flutuarem por diferenças entre locutores. Apesar de a estrutura formântica ser característica de uma vogal, quando se trocam os locutores, a posição dos formantes muda em função das diferenças entre seus tratos vocais.

Por exemplo: formantes próximos para um locutor podem ocorrer ainda mais próximos para o outro, o que provoca o deslocamento do intervalo crítico da tendência **iii** para a esquerda. Já o aumento da frequência fundamental leva a um aumento no “período” das ondulações que ocorrem na curva do espectro (provenientes dos pulsos glotais), o que

pode fazer com que a curva da análise passe a acompanhar tais características, prejudicando a envoltória espectral e interferindo na curva do erro. O aumento na ordem das análises reduz o erro como um todo, e desloca de maneira irregular as tendências de sua curva para a direita. Assim, a ordem é mais um fator de flutuações na curva do erro.

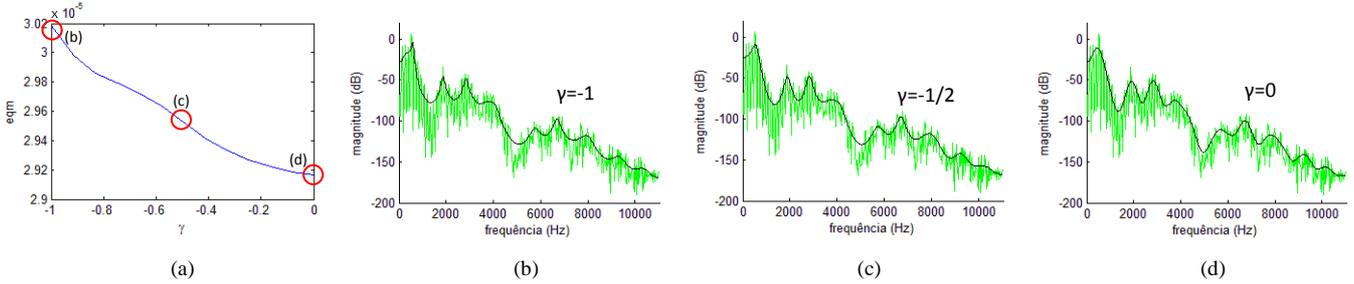


Fig. 3: vogal “é”: (a) erro quadrático médio em função de γ . (b) - (d): envoltórias espectrais para $\gamma=-1$, $\gamma=-1/2$ e $\gamma=0$ sobrepostas ao espectro do sinal de fala

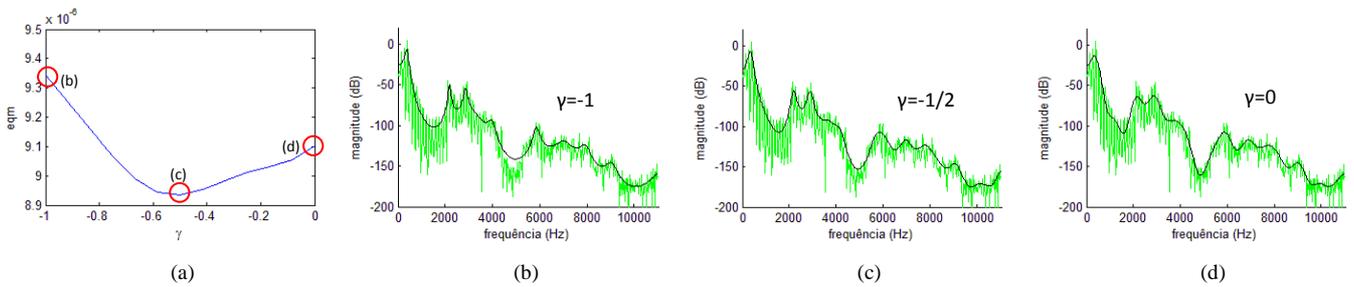


Fig. 4: vogal “e”: (a) erro quadrático médio em função de γ . (b) - (d): envoltórias espectrais para $\gamma=-1$, $\gamma=-1/2$ e $\gamma=0$ sobrepostas ao espectro do sinal de fala.

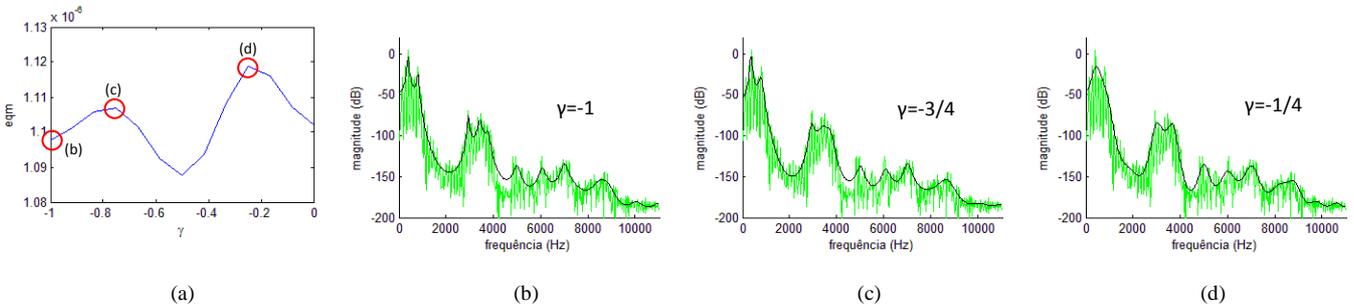


Fig. 5: vogal “o”: (a) erro quadrático médio em função de γ . (b) - (d): envoltórias espectrais para $\gamma=-1$, $\gamma=-3/4$ e $\gamma=-1/4$ sobrepostas ao espectro do sinal de fala.

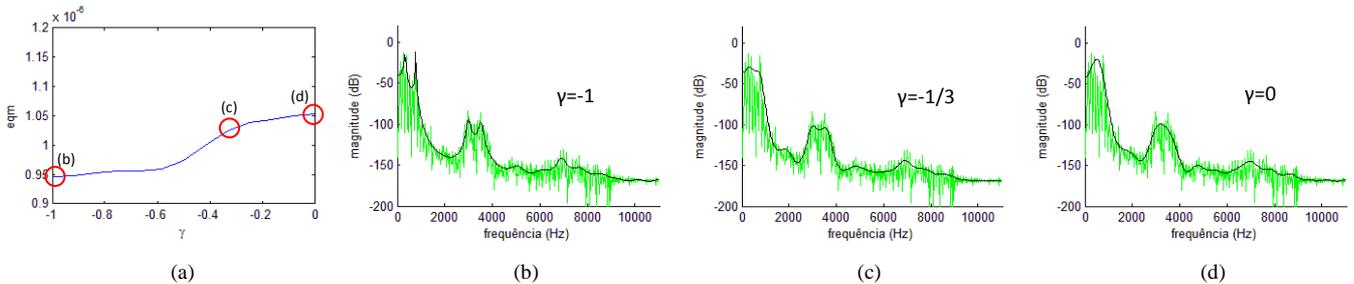


Fig. 6: vogal “u”: (a) erro quadrático médio em função de γ . (b) - (d): envoltórias espectrais para $\gamma=-1$, $\gamma=-1/3$ e $\gamma=0$ sobrepostas ao espectro do sinal de fala.

V. DISCUSSÃO

Neste trabalho foram apontadas as características do espectro que influenciam na qualidade da modelagem da envoltória espectral, por análises efetuadas com diversos valores de γ no intervalo $-1 \leq \gamma < 0$, com base no erro quadrático médio (de predição). Seria interessante a definição de um valor ótimo de γ associado a cada vogal presente na língua, o que não parece ser objetivamente possível, uma vez que a curva de erro não é robusta a fatores como: frequência fundamental, locutor e ordem da análise. A avaliação efetuada nos permite concluir que as curvas de erro das vogais anteriores “a”, “é” e “e” tendem a ser menos acidentadas porque tais vogais não apresentam formantes muito próximos entre si, o que evita os efeitos locais associados à tendência **iii**. Tais efeitos ocorrem para as vogais posteriores “ó”, “o” e “u” pois estas têm formantes próximos entre si. A proximidade do primeiro formante da vogal anterior “i” com seu simétrico na região de frequências negativas também a torna a curva de erro susceptível a tais efeitos.

Uma tentativa de busca por um γ ótimo para cada quadro extraído do sinal de fala exigiria que fossem feitas diversas análises, uma para cada valor do parâmetro γ dentre um conjunto deles, o que envolveria alto custo computacional [3]. Tentativas de otimização de γ são desfavorecidas pelo comportamento irregular da curva do erro. A alternativa é a escolha de um único valor de γ para todas as análises. Em [10], análises efetuadas com ordem 15, apontam $\gamma = -1$ como sendo superior para a análise de vogais orais. Já as nasais são favorecidas com γ próximo de 0, o que se espera, uma vez que os vales acentuados provenientes dos anti-formantes são fortes caracterizadores dos sons nasais. Uma curva média do erro quadrático de predição em função de γ extraída de um minuto de fala (em japonês) aponta $\gamma = -0.4$ como valor ótimo para a fala natural (valor que depende da ordem de análise). Experimentos de análise e ressíntese com ordem 25 apontam que os valores $\gamma = -1/2$ e $\gamma = -1/3$ geram falas sintetizadas com qualidade superior àquelas obtidas com $\gamma = -1$ ou 0. Vale destacar que valores de γ dados por frações negativas de numerador unitário favorecem a implementação do filtro de síntese a eles associados.

A análise cepstral generalizada, quando efetuada na escala mel (escala não-linear de frequências que acompanha a alta resolução do ouvido humano nas baixas frequências), é denominada análise mel-cepstral generalizada. Os coeficientes mel-cepstrais generalizados mostraram bons resultados na área de codificação de fala [6], em experimentos com reconhecimento de palavras isoladas via HMM [7] e

síntese de fala via HMM [5], além de apresentarem, quando utilizados na forma LSP (MGC-LSP), superioridade em características de quantização e interpolação em relação aos coeficientes LSP e mel-cepstrais [4]. O tratamento para a análise mel-cepstral generalizada é mais complexo do que para a análise cepstral generalizada e será abordado nos próximos trabalhos.

REFERÊNCIAS

- [1] B. S. Atal and S. L. Hanauer: “Speech analysis and synthesis by linear prediction of the speech wave”, in *J. Acoust. Soc. America*, vol. 50, no. 2, pp.637–655, Mar. 1971.
- [2] Keiichi Tokuda, Takao Kobayashi and Satochi Imai: “Adaptive Cepstral Analysis of Speech”, *IEEE Trans on Speech and Audio Processing*, vol. 3, no. 6, pp.481–489, Nov. 1995.
- [3] K. Tokuda, T. Kobayashi, T. Chiba and S. Imai: “Spectral estimation of speech by mel-generalized cepstral analysis” translation: *Electronics and Communications in Japan (Part 3)*, vol. 76, no. 2, pp.30–43, July 1993.
- [4] Kazuhito Koishida, Keiichi Tokuda, Takao Kobayashi and Satochi Imai: “Spectral Representation of Speech Based on Mel-Generalized Cepstral Coefficients and Its Properties”, translation: *Electronics and Communications in Japan (Part 3)*, Vol. 83, No. 3, pp.50–59, 2000.
- [5] H. Zen, T. Toda, and K. Tokuda: “The Nitech-NAIST HMM-based speech synthesis system for the Blizzard Challenge 2006,” in *Proc. Blizzard Challenge 2006*, Sept. 2006.
- [6] Kazuhito Koishida, Gou Hirabayashi, Keiichi Tokuda, Takao Kobayashi: “A 16kb/s Wideband CELP-Based Speech Coder Using Mel-Generalized Cepstral Analysis”, *IEICE Trans. Inf. & Syst.*, Vol. E83-D, no. 4, April 2000.
- [7] K. Tokuda, T. Kobayashi, T. Masuko, and S. Imai: “Mel-generalized cepstral analysis - a unified approach to speech spectral estimation”, in *Proc. of ICASSP*, pp.1043–1046, 1994.
- [8] S. Imai and C. Furuichi: “Unbiased estimation of log spectrum”, *Trans. IECE*, vol. J70-A, pp.461–480, Mar. 1987.
- [9] K. Tokuda, T. Kobayashi, R. Yamamoto and S. Imai, “Spectral estimation of speech based on generalized cepstral representation” *Trans. IEICE*, vol. J72-A, pp.457–465, Mar. 1989.
- [10] K. Tokuda, T. Kobayashi, and S. Imai: “Generalized cepstral Analysis of speech – Unified approach to LPC and cepstral method” in *Proc. ICLSP 90*, pp.37–40, 1990.
- [11] S. Imai and C. Furuichi: “Unbiased estimator of log spectrum and its application to speech signal processing”, in *Proc. 1988 EURASIP*, pp.203–206, Sep. 1988.
- [12] T. Fukada, K. Tokuda, T. Kobayashi, and S. Imai: “An adaptive algorithm for mel-cepstral analysis of speech”, in *Proc. ICASSP-92*, pp.1-137-I-140, 1992.
- [13] K. Tokuda, T. Kobayashi, R. Yamamoto and S. Imai: “Spectral estimation of speech based on generalized cepstral representation”, *Trans. IEICE*, vol. J72-A, pp.457–465, mar. 1989.
- [14] R. R. A. Barreira e F. Violaro: “Avaliação da Análise Cepstral Generalizada Aplicada à Modelagem de Vogais”, 7º Congresso de Engenharia de Áudio da AES, maio, 2009.
- [15] A. V. Oppenheim and R. W. Schaffer: “Digital Signal Processing”, 1975.