

Codificador de Voz Baseado em Soma de Gaussianas e no Algoritmo *Matching Pursuit*

Ícaro de Vasconcelos Brito* e Raissa Bezerra Rocha* †

*Universidade Federal de Sergipe – UFS, São Cristóvão, Brasil

†Instituto de Estudos Avançados em Comunicações – Iecom

E-mails: icarovb@hotmail.com e raissa@iecom.org.br

Resumo— Este trabalho investiga o desempenho de um codificador de sinal de voz baseado em soma de gaussianas e no algoritmo *matching pursuit*. A proposta parte do ajuste da forma de onda do sinal de voz por uma combinação de gaussianas. A aproximação é, então, incrementada por meio da aplicação do *matching pursuit*, no qual as formas de onda melhor ajustadas ao sinal são utilizadas como átomos, ponderados por uma matriz esparsa de coeficientes. Avaliações subjetivas do algoritmo demonstram que o codificador apresenta capacidade de compressão similar à de codificadores padronizados, como o ADPCM (G.726) e o LD-CELP (G.728).

Palavras-chave— Codificador, *matching pursuit*, sinal de voz, soma de gaussianas.

I. INTRODUÇÃO

Em um sinal de voz, a memória requerida para gravá-lo, a largura de banda necessária para transmiti-lo e a relação sinal-ruído são diretamente proporcionais ao número de *bits* por amostra [1]. Os codificadores buscam explorar característica da fala, como a redundância existente entre amostras adjacentes ou o estudo de parâmetros prosódicos, entre outros, de forma a reduzir o número de *bits* necessário para representar adequadamente os sinais de voz [2]. A codificação de voz é empregada em diversas aplicações, como *voice mail*, telefone pela *internet*, telefone celular digital, voz criptografada, entre outros.

Neste trabalho, é apresentado um codificador de voz que representa a fala por meio de soma de gaussianas (CSG). O conceito da codificação utilizada é baseado na observação de que a forma de onda do sinal de voz assemelha-se a gaussianas sobrepostas, levando à hipótese de que ela pode ser reconstruída a partir de um vetor formado por informações de médias, desvios e amplitudes de gaussianas.

Os parâmetros gaussianos são, então, combinados com a técnica de *matching pursuit* (MP) que mostra-se a ferramenta ideal para promover tanto o aperfeiçoamento na aproximação do sinal como a redução de parâmetros para codificá-lo, a um custo computacional aceitável.

Além desta seção introdutória, este artigo está dividido em mais três seções. A Seção II contempla o processo de desenvolvimento do codificador baseado na soma de gaussianas. É explorado também a aplicação do algoritmo *matching pursuit* ao codificador pesquisado, analisando os benefícios dessa técnica na codificação de voz. A Seção III apresenta os resultados obtidos bem como a discussão sobre eles. Por fim, as conclusões da pesquisa são descritas na Seção IV.

II. CODIFICAÇÃO DE VOZ POR SOMA DE GAUSSIANAS

A. Descrição do Codificador por Soma de Gaussianas (CSG)

O CSG é, a princípio, um ajuste de curva por soma de gaussianas. No caso do presente trabalho, a curva a ser aproximada é um sinal digital de voz a 8000 amostras/s, com 16 *bits* por amostra e seccionado em janelas de N amostras. A busca pelo número mínimo de gaussianas e seus respectivos parâmetros para representar cada janela é baseada no trabalho de [3], no qual são utilizadas as técnicas de análise de espaço-escala de Witkin [4] para ajustar curvas por uma combinação de gaussianas.

B. Análise de Espaço-Escala

Dada uma sequência de N amostras uniformemente espaçadas $s(x) : \mathfrak{R} \rightarrow \mathfrak{R}$, busca-se aproximá-la por uma soma de gaussianas na forma:

$$y(x) = \sum_{i=1}^n A_i e^{\frac{1}{2} \left(\frac{x - \mu_i}{\sigma_i} \right)^2}, \quad (1)$$

em que $y(x) : \mathfrak{R} \rightarrow \mathfrak{R}$ é uma curva única com parâmetros (amplitude A_i , média μ_i e desvio σ_i) de n gaussianas e cujos valores associados a x são ajustados a $s(x)$ para uma dada precisão.

Na análise de espaço-escala (EE), o sinal é filtrado por núcleos gaussianos em diferentes escalas, gerando o que Witkin chamou de imagem em espaço-escala, como ilustrado na Figura 1. Nessa representação é possível rastrear os pontos críticos da curva, estimando-se com boa precisão o número de gaussianas formantes e os valores iniciais para as respectivas médias e desvios.

O próximo passo é, então, utilizar as combinações de gaussianas provenientes da análise EE para criar um grupo de formas de onda que funcionem como um dicionário especial para o algoritmo *matching pursuit*.

C. Matching Pursuit Aplicado ao CSG

Matching pursuit é um algoritmo no qual um sinal é decomposto em uma expansão linear de formas de onda pertencentes a um dicionário redundante de funções [5]. Nessa técnica, o sinal é iterativamente projetado sobre as funções do dicionário de modo que, após P iterações, a forma de onda melhor adaptada para aproximar o sinal é selecionada.

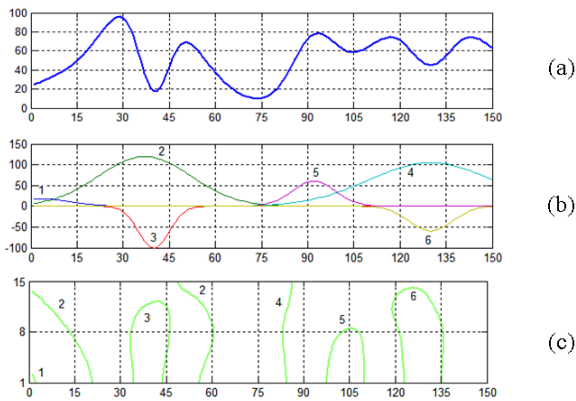


Fig. 1: Exemplo de representação de um sinal em espaço-escala. (a) Sinal original composto por 6 gaussianas aleatórias somadas; (b) Gaussianas formantes; (c) Contorno espaço-escala resultante do sinal (a). Fonte: o autor (2015)

Os dicionários são, em geral, paramétricos ou adaptativos [5]. Os paramétricos, como o dicionário de Gabor, apresentam a vantagem do alto nível de compressão, enquanto que os adaptativos, como as somas de gaussianas da análise EE, destacam-se pelo baixo esforço computacional requerido [6]. Como neste trabalho busca-se equilibrar qualidade acústica e taxa de compressão, os dois modelos são empregados, aplicando-se o tipo de dicionário mais adequado às características dos fragmentos de sinal de voz a serem codificados.

Nessa modificação do CSG, doravante chamado CSG-MP, cada janela do sinal de voz é previamente analisada com as janelas posteriores, a fim de se verificar a similaridade entre elas. Os fragmentos similares do sinal codificado de voz são agrupados e as gaussianas com mais altos índices de correlação com as janelas originais são selecionadas. Essas gaussianas formam, então, M átomos ótimos para aproximar todo o agrupamento. Os fragmentos sem correlação não são analisados pelo EE, sendo codificados de forma independente com M átomos de Gabor, reduzindo a necessidade de armazenar ou transmitir um número maior de dicionários.

O codificador estabelece, também, um limite para a seleção dos coeficientes do MP, de maneira que aqueles cujos módulos são inferiores a um dado limiar são anulados. Quanto maior o valor do limiar, menos coeficientes são utilizados, aumentando a compressão, mas reduzindo a qualidade da aproximação. Dessa maneira, deve-se buscar um número reduzido, mas suficiente, de coeficientes não-nulos e dicionários a serem empregados na reconstrução do sinal aproximado.

III. RESULTADOS E DISCUSSÃO

O CSG-MP foi implementado computacionalmente com uma interface gráfica para que diferentes configurações pudessem ser avaliadas. Para aferir a qualidade do áudio processado, foram realizados testes de escutas, em que 63 participantes classificaram subjetivamente a qualidade do sinal processado segundo uma escala predeterminada de acordo com [7]. A Tabela I apresenta a comparação dos resultados obtidos para o codificador proposto com alguns dos principais

codificadores difundidos da literatura. Os resultados apresentados para o CSG são alcançados com 3 ou 9 átomos, com 50 iterações e com a eliminação de coeficientes de módulo inferior a 0,05.

De acordo com os resultados, o codificador CSG apresenta resultados competitivos com os demais codificadores, oferecendo metade da taxa de *bits* e qualidade semelhantes à do codificador ADPCM (G.726), bem como apresenta características semelhantes ao codificador LD-CELP (G.728).

TABELA I: Comparação de Desempenho dos Codificadores de Voz Estudados e Proposto [8].

Codificador	Taxa de bits (bits/s)	MOS
PCM (G.711)	64k	4,10
ADPCM (G.726)	32k	3,85
LPC-10 (FS 1015)	2,4k	2,40
LD-CELP (G.728)	16k	3,61
CS-ACELP (G.728)	8k	3,92
AMR (G.722)	12,2k	4,10
CSG-MP (M = 3)	12,4k	2,89
CSG-MP (M = 9)	16,5k	3,31

IV. CONCLUSÕES E TRABALHOS FUTUROS

Neste trabalho, é apresentada uma técnica de codificação inovadora, resgatando a análise da forma de onda do sinal de voz, mas não se limitando a ela. Diferentemente dos codificadores de forma de onda convencional, o codificador por soma de gaussianas vale-se também de conceitos das técnicas paramétricas e da aplicação do promissor algoritmo *matching pursuit*. Da análise do sinal de voz filtrado por núcleos gaussianos em diferentes escalas, comprovou-se ser possível extrair gaussianas formantes, cuja soma se aproxima razoavelmente do sinal original. Para se atingir um desempenho melhor, a técnica MP foi adaptada ao codificador.

A partir dos resultados, constata-se que o codificador apresenta potencial, com resultados próximos aos codificadores da literatura. Como trabalhos futuros, pretende-se melhorar a qualidade geral do codificador CSG por meio da incorporação de características perceptuais e técnicas de eliminação de ruídos para elevar consideravelmente o desempenho do CSG-MP. Além disso, a compressão do sinal é um campo a ser mais explorado, podendo revelar meios de se atingir taxas de *bits* ainda mais baixas que as alcançadas.

REFERÊNCIAS

- [1] S. V. Vasegui, *Advanced Digital Signal Processing and Noise Reduction*. Londres: Wiley, 2009.
- [2] M. S. Alencar, *Telefonia Celular Digital*. São Paulo: Editora Érica, 2012.
- [3] A. Goshtasby and O. D. O'Neill, "Curve fitting by a sum of Gaussians." *CVGIP: Graphical Models and Image Processing*, vol. 56, no. 4, pp. 281–288, 1994.
- [4] A. P. Witkin, "Scale space filtering: A new approach to multi-scale description." *Proc. Image Understanding*, 1984.
- [5] S. Mallat and Z. Zhang, "Matching Pursuits with Time-Frequency Dictionaries." *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.
- [6] R. Mazhar, "Optimized Dictionary Design and Classification Using the Matching Pursuit's Dissimilarity Measure." Tese de doutorado, University of Florida, 2009.
- [7] ITU-T, "Methods for Subjective Determination of Transmission Quality." 1996.
- [8] Cisco Systems, Inc. Understanding Codecs: Complexity, Hardware Support, MOS, and Negotiation. <http://www.cisco.com/c/en/us/support/docs/voice/h323/14069-codec-complexity.html>. Visitado em 05 de abril de 2015.