Reconhecimento de Fala Aplicado aos Sinais Extraídos de Misturas de Fontes no Desenvolvimento de SAPDA

Bruno do Amaral[1], José H. Saito[1,2]

[1]Programa de Pós-Graduação em Ciência da Computação - FACCAMP
Campo Limpo Paulista-SP, Brasil

[2] Universidade Federal de São Carlos - UFSCar
São Carlos-SP, Brasil

brunodoamaralifsp@gmail.com, saito@cc.faccamp.br

Abstract. *This article consists of the development of trials concerning to computational processing of an assistance system for people with hearing disabilities (ASPHD) based on a Coktail Party of problem scenario, the blind separation method of sources through independent component analysis and automatic speech recognition. Two experiments were carried consisting in testing the efficacy of FastICA algorithm based on the system performance in terms of the amount of separation and recognition errors, after mixing of the signals.*

Resumo. *Este artigo consiste no desenvolvimento de ensaios relativos a processamentos computacionais de um sistema de assistência às pessoas com deficiência auditiva (SAPDA) baseado em um cenário do problema de Coktail Party, pelo método de separação cega das fontes por meio da análise de componentes independentes e o reconhecimento automático de fala. Foram desenvolvidos dois experimentos que consistem em testar a eficácia do algoritmo FastICA com base no desempenho do sistema em termos de quantidade de erros de separação e reconhecimento, após a mistura dos sinais.*

1. Introdução

O número de pesquisas e projetos na área de tecnologia assistiva vem se tornando cada vez maior, buscando desenvolver sistemas que promovam melhorias na qualidade de vida, possibilitando independência e inclusão social para pessoas portadoras de deficiência (Danesi, 2007). Com o avanço da tecnologia outros campos de pesquisa tiveram avanços significativos até o presente, como em processamento de sinais, onde os sinais elétricos resultantes de falas podem ser processados com diversas finalidades, dentre as quais, a filtragem de fala de um indivíduo e separação de falas individuais, numa mistura de sinais, conhecido como Separação Cega de Fontes (em inglês, *Blind Source Separation-BSS*) (Comon e Jutten, 2010). Outro campo de pesquisa com avanço significativo é o de Reconhecimento Automático de Fala (em inglês, *Automatic Speech Recognition-ASR*) (Benesty, Sondhi e Huang, 2008).

O objetivo deste trabalho é contribuir para o desenvolvimento futuro de um sistema para assistência às pessoas com deficiência auditiva (SAPDA), com base nas conclusões dos

experimentos realizados em simulações computacionais. Os sistemas SAPDA compõem-se de sensores atuadores para a captação de sinais; sistemas computacionais localizados, ou remotos, como em computação em nuvem (Vecchiola, Pandey e Buyya, 2009), para o processamento de sinais captados pelos sensores; e dispositivos atuadores que sinalizam as pessoas, com o resultado do processamento de sinais.

Foram realizados testes preliminares com algoritmos de separação cega das fontes, usando a técnica de Análise de Componentes Independentes (em inglês, *Independent Component Analysis*-ICA) (Stone, 2004), de pequenas frases e os sinais de áudio extraídos das misturas de falas, imitando os sinais captados pelos sensores de um sistema SAPDA. Após a separação de fontes, os sinais extraídos foram submetidos ao reconhecimento de voz, usando uma técnica de ASR. A intenção desta pesquisa é obter o embasamento teórico sobre o processamento de sinais, considerando um cenário de aplicação de um sistema SAPDA.

O trabalho é constituído das seguintes seções. Na Seção 2 são apresentados os fundamentos e metodologia. Na Seção 3 são descritos os experimentos realizados e os resultados obtidos. Finalmente, na Seção 4 são apresentadas as conclusões e os trabalhos futuros.

2. Referencial Teórico e Metodológico

O termo Separação Cega das Fontes se deve ao desconhecimento das fontes dos sinais, tendo como objetivo estimar os sinais de origem a partir das misturas captadas pelos sensores. Um dos métodos mais difundidos para BSS é o uso do ICA. Com esse método se extraem os componentes estatisticamente independentes com distribuições não gaussianas das misturas observadas.

Considerando o cenário com uma pessoa surda distraída, tendo no ambiente um aparelho de rádio ligado numa estação com um locutor apresentando um programa qualquer. Nesse momento ocorre algum problema com uma outra pessoa presente neste ambiente e o mesmo grite chamando o deficiente em um pedido de socorro. Considerando que não há nenhuma outra pessoa na casa, como o deficiente será avisado para realizar o socorro? Nessa casa estão instalados dois microfones, em locais distintos, que fornecem dois sinais para gravação, $x_1(t)\ e\ x_2(t)$. Cada um desses sinais corresponde à mistura entre os dois sinais de voz captados no tempo *t*, um do locutor de rádio, e o outro, da pessoa pedindo socorro. As vozes serão denominadas por $s_1(t)\ e\ s_2(t)$, dadas pelas equações 2.1 e 2.2:

$$x_1(t) = a_{11} * s_1(t) + a_{12} * s_2(t) \quad (2.1)$$

$$x_2(t) = a_{21} * s_1(t) + a_{22} * s_2(t) \quad (2.2)$$

Visto que a_{11}, a_{12}, a_{21} e a_{22} são a soma dos parâmetros relativos às distâncias dos microfones para as fontes das vozes. Este problema caracteriza o que foi denominado de *Cocktail Party* por Hyvärinen e Oja (1997). A mistura simplificada para o problema pode ser dada por $\boldsymbol{x} = \boldsymbol{A} * \boldsymbol{s}$. Ao estimar a matriz inversa de $\boldsymbol{A}$, denominada $\boldsymbol{W}$, $\boldsymbol{W} = \boldsymbol{A^{-1}}$, encontramos a solução do problema, obtendo as fontes originais, conforme equação 2.3.

$$s = W * x \quad (2.3)$$

Uma das formas de se obter a matriz $\boldsymbol{W}$ é calcular a matriz que maximiza a não gaussianidade do vetor $\boldsymbol{s}$, por meio do Teorema do Limite Central. Este Teorema diz

que para uma soma de variáveis aleatórias independentes tendendo ao infinito, a função de densidade de probabilidade dessa soma tenderá a uma distribuição gaussiana; e que quanto menos gaussiana for a distribuição, menor será a mistura existente nos componentes independentes estimados (Hyvärinen, Karhunen e Oja, 2001).

Com base na definição de ICA, Hyvärinen e Oja (1997) propuseram o algoritmo FastICA baseado na interação do ponto fixo para maximização da não-gaussianidade, pelo método da negentropia ou entropia diferencial. Inicialmente, o algoritmo faz uma avaliação dos sinais de mistura, x, se os dados possuem média zero. Caso isso não ocorra, deve ser realizado o processo de transformação para a média zero (*demeaning*). Em seguida, no passo 2 é realizado o processo de branqueamento dos dados que é um pré-processamento para a aplicação do método de ICA. O branqueamento é possível, aplicando-se o algoritmo de PCA (*Principal Component Analysis*) (Hyvärinen e Oja, 1997). O ajuste de uma das linhas da matriz W é descrito como $w^T x$, onde $\boldsymbol{w}$ é um vetor de coeficientes de separação, escolhido aleatoriamente no passo 3 do algoritmo. Durante uma iteração do algoritmo, um novo valor da matriz W^+, é obtido, até que haja a convergência. Na iteração é aplicado um processo de não-linearidade, g, conforme descrito no passo 4 do algoritmo. No passo 5, o valor de W^+ é normalizado. E caso não haja convergência repete-se a iteração, passo 6.

Algoritmo 1: FastICA por negentropia

1. Fazer com que os dados de entrada x possua média zero.
2. Fazer o branqueamento dos dados.
3. Escolher um vetor de peso $\boldsymbol{w}$ inicial (aleatório).
4. $W^+ = E\{xg(w^T x)\} - E\{g'(w^T x)\}w$
5. Normalizar, dividindo W^+ por sua norma.

$$W_{normalizado} = \frac{W^+}{\|W^+\|}$$

6. Se não convergir, voltar ao passo 4.

O algoritmo de reconhecimento automático de fala (ASR) tem como objetivo reconhecer uma determinada sentença falada. Segundo Cuadros (2007), Yu e Deng (2015), os problemas de reconhecimento de fala por máquinas estão relacionados à complexidade da voz humana, formada por fatores como características vocais, entonação, estado emocional do individuo, velocidade da voz, interferência de ruídos dentre outras fatores. Os sistemas atuais de ASR baseiam-se principalmente nos princípios de reconhecimento estatístico de padrões, nos quais o sinais acústicos são transformados, formando uma sequência de simbolos para serem analisados em unidades de sub-palavras (Rabiner e Juang, 2008). Esta metodologia caracteriza uma menor perda de informações e reduz o processamento computacional.

3. Experimentos Realizados

Foram realizados dois experimentos que consistiram na verificação da funcionalidade e desempenho do Algoritmo FastICA, com um número significativo de palavras nos dados utilizados. Para o primeiro experimento foram gravadas três frases com vozes distintas, com o conteúdo conforme Tabela 1.

TABELA 1. Frases com vozes distintas usadas no Experimento I.

F1	"O verdadeiro sentido da existência humana não é simplesmente nascer, viver e morrer, mas sim, deixar um pouco de si em cada momento em que se vive."
F2	"As pessoas costumam dizer que a motivação não dura sempre. Bem, nem o efeito do banho, por isso recomenda-se diariamente."
F3	"Escolha uma ideia. Faça dessa ideia a sua vida. Pense nela, sonhe com ela, viva pensando nela. Deixe cérebro, músculos, nervos, todas as partes do seu corpo serem preenchidas com essa ideia. Esse é o caminho para o sucesso."

A partir dos sons obtidos das gravações, no Matlab foram extraídos os sinais digitais, a uma taxa de 300.000 amostras por segundo. Na Figura 1, ilustra-se na linha superior, três sinais fonte; na linha inferior são mostrados os sinais resultantes de misturas.

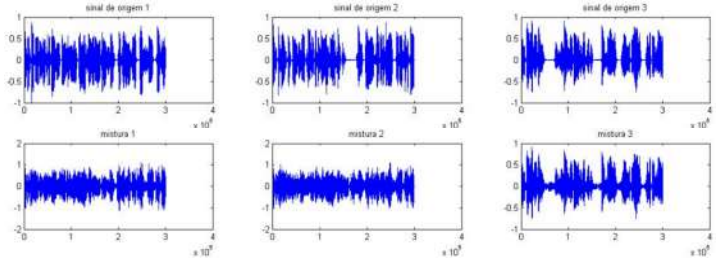

Figura 1. Sinais originais (linha superior); misturas (linha inferior).

As três misturas obtidas a partir da transformação linear entre a matriz de mistura de valores aleatórios, foram gravados em sinais de audio (*wav*) pelo Matlab. Os sons das misturas foram colocados um aúdio por vez no aplicativo de plataforma livre *Dictanote* para o reconhecimento de fala, e como era previsto, não foi possível o reconhecimento. Aplicando o Algoritmo 1, o primeiro componente independente foi calculado após 15 iterações, enquanto o segundo foi calculado após 4 iterações e o último componente após 2 iterações. De posse dos sinais fonte estimados, foi utilizado o sistema ASR. O reconhecimento perfeito das falas, comprovaram a eficácia do Algoritmo 1.
Na Figura 2 são apresentados os componentes extraídos.

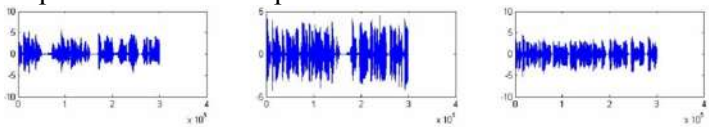

Figura 2. Sinais estimados pelo algoritmo FastICA.

Para o segundo experimento foram considerados três textos literários, conforme Tabela 2. Na aplicação do Algoritmo 1, os sinais extraídos apresentaram 14, 17 e 21 erros, respectivamente, quanto ao texto original, na contagem de palavras erroneamente extraídas, conforme representado na tabela 2. A Figura 3 ilustra o desempenho obtido no Experimento II.

TABELA 2 . Experimento II.

	Qtide total de palavras p/ cada texto	Qtide de erros no sinal extraído pelo Algoritmo-1	Porcentagem de erros (%)	Porcentagem de acertos (%)
Texto1 (Huxley, 1979)	212	14	6,60	93,40
Texto2 (Assis, 1899)	215	17	7,91	92,09
Texto3 (Assis, 1891)	206	21	10,19	89,81

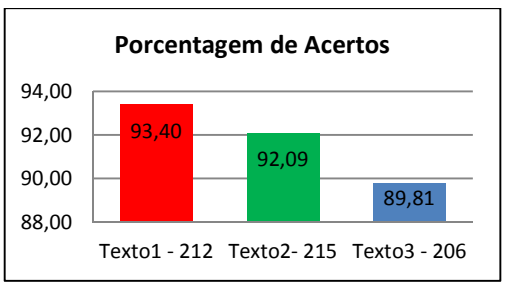

Figura. 3. Gráfico da porcentagem de acertos do Experimento II.

4. Conclusão

Foi descrito o desenvolvimento de ensaios relativos a um sistema de assistência às pessoas com deficiência auditiva (SAPDA) baseado em um cenário do problema de *Coktail Party*. Os sinais de áudio extraídos das misturas pelo algoritmo FastICA para os dois experimentos, foram submetidos ao reconhecimento automático de fala, obtendo o texto escrito confirmando a eficácia do algoritmo com taxas satisfatórias de reconhecimento quanto ao texto original. A intenção desta pesquisa é contribuir para o embasamento teórico sobre os mecanismos necessários para um sistema SAPDA. Apesar desses sistemas serem complexos, os estudos das ferramentas computacionais, como os apresentados, devem permitir o desenvolvimento satisfatório de um SAPDA. Nesse sentido são previstos para os próximos passos, desenvolvimentos de atuadores que permitam às pessoas assistidas receber as informações úteis obtidas pelo sistema.

Referências

Assis, M. (1891)."Quincas Borba". B.L.Garnier, Livreiro-Editor, Rio de Janeiro-RJ.

Assis, M. (1899)."Dom Casmurro". B.L.Garnier, Livreiro-Editor, Rio de Janeiro-RJ.

Benesty, J. Sondhi, M.M. Huang, Y. (2008)."Springer Handbook of Speech Processing". Springer-Verlag, Heidelberg, Alemanha.

Danesi, M.C. (2007). "O admirável mundo dos surdos: novos olhares do fonoaudiólogo sobre a surdez". EDIPUCRS, 2ª ed., Porto Alegre.

Dictanote. Disponivel em https://dictanote.com/, acessado em julho de 2016.

Fast-ICA. Disponível em http://research.ics.aalto.fi/ica/fastica/code/dlcode.shtml, acessado em junho de 2016.

Huxley, A. (1979). "Admirável mundo novo". Trad. V.Oliveira e L.Vallandro, Globo, Porto Alegre-RS.

Hyvärinen A., Oja, E. (1997). "A fast fixed-point algorithm for independent component analysis." Neural computation,v.9, p.1483-1482.

Hyvärinen, A. Karhunen, J. Oja, E. (2001). Independent Component Analysis. John Wiley& Sons, New York.

Jutten, C. Comon, P. (2010)."Handbook of Blind Source Separation: Independent Compont Analysis and Applications". Academic Press, Burlington, MA, USA.

Pedersen, M.S.(2006)."Source Separation for Hearing Aid Applications", Tese de Doutorado, Technical Un. of Denmark, Informatics and Mathematical Modeling, Denmark.

Quadros, C.D.R. (2007). "Reconhecimento de voz e de locutor em ambientes ruidosos: comparacão das técnicas MFCC e ZCPA", Dissertação de Mestrado - Universidade Federal Fluminense, Niterói-RJ.

Rabiner, L. Juang, H. (2008)."Historical Perspective of the Field of ASR/NLU, Handbook of Speech Recognition", (J. Benesty, M.M. Sondhi, Y. Huang editors), Springer-Verlag, Heidelberg, Alemanha, pp. 521-537.

Stone, J.V. (2004). "Independent Component Analysis: A Tutorial Introduction". Bradford Book, Cambridge, MA, USA.

Vecchiola, C. Pandey, S. Buyya, R. (2009)."High-Performance Cloud Computing: A View of Scientific Applications", Proceedings of the 10th International Symposium on Pervasive Systems, Algorithms and Networks, Kaohsiung, Taiwan, December 14-16.

Yu, D. Deng, L. (2015). "Automatic Speech Recognition: A Deep Learning Approach". Springer-Verlag, London.