

RECUPERAÇÃO DE FONTES E ANÁLISES POR COMPONENTES INDEPENDENTES: CONCEITOS, FUNDAMENTOS E APLICAÇÕES EM LINGUÍSTICA

Guilherme de Alencar Barreto*, Charles Casimiro Cavalcante*

Resumo

Este trabalho apresenta uma introdução à técnica de análise por componentes independentes (ICA, Independent Component Analysis), tipicamente utilizado em problemas de recuperação cega em sistemas de informação e sua aplicabilidade em problemas nas áreas de processamento da fala e linguística. São descritos e discutidos os aspectos de modelagem matemática do problema de separação de fontes genérico bem como suas estratégias mais comuns para solução dos problemas supracitados. A ilustração de uma aplicação na separação de sinais de voz combinados linearmente e coletados em vários sensores é mostrada para demonstrar a potencialidade da técnica. Além disso, é realizada uma breve descrição de algumas aplicações de ICA em problemas das áreas de processamento de voz e linguística.

Palavras-Chave: *Análise por componentes independentes, aprendizado estatístico, processamento não-supervisionado, linguística, extração de características, classificação de padrões.*

Abstract

This work brings an introduction on the independent component analysis technique (ICA), usually applied to blind recovering in information systems, and its applicability to problems related to the fields of speech processing and linguistics. The aspects of mathematical modeling are described as well as the most common strategies for the problem solution. The illustration of an application to speech signals separation, linearly combined and acquired in several sensors is presented in order to show the potential

of the technique. Besides that, it is briefly reviewed some recent applications of ICA to the fields of speech processing and linguistics.

Keywords: *Independent component analysis, statistical learning, unsupervised learning, linguistic, feature extraction, pattern classification.*

INTRODUÇÃO

Uma tarefa de extrema importância em sistemas de informação, tais como comunicação de dados, processamento de voz/fala, classificação de padrões, econometria, processamento de imagens e linguística, é a recuperação da informação processada de modo confiável. Tornar esta tarefa automática de tal forma que um mecanismo computacional possa realizá-la não é simples, principalmente quando não se conhece a informação transmitida e quando várias informações compartilham o mesmo meio de transmissão.

Este tipo de problema é conhecido na literatura como *separação cega de fontes* (BSS, sigla em inglês). O nome cego é relacionado ao fato de não se “enxergar” (conhecer) o conteúdo da informação das fontes. Neste caso, são conhecidas, no máximo, algumas de suas características estatísticas elementares. Tradicionalmente, uma das mais usadas suposições, que condiz bastante com os processos físicos e biológicos que tratamos, é a hipótese de independência estatística entre os sinais-fonte (ou simplesmente, fontes). Isto leva a um processamento voltado à geração de estimativas destas fontes que sejam também independentes. Daí o nome *análise por componentes independentes* (ICA, sigla em inglês).

* Departamento de Engenharia de Teleinformática (DETI), Universidade Federal do Ceará (UFC), Emails: guilherme@deti.ufc.br e charles@deti.ufc.br

O problema de separação cega de fontes, ou uma aproximação dele, é bastante comum no nosso dia-a-dia. No caso de uma reunião ou festa, vários tipos de fontes, algumas delas inclusive correlacionadas, estão presentes e perturbam a compreensão e identificação de uma determinada fonte de informação (sinal de voz, por exemplo). Situações como estas são apresentadas ao cérebro humano quase que diariamente e não se nota nenhuma dificuldade para compreender um determinado interlocutor, salvo em meio a ruído excessivo, graças à grande capacidade de processamento do cérebro humano.

As ferramentas e modelos desenvolvidos para este fim têm sua história remontando à década de 80. Desde então, muito esforço para desenvolvimento e compreensão de métodos que permitam a separação das fontes envolvidas tem sido empreendido. E é nesse contexto que se insere este trabalho. Se tomarmos uma construção gramatical (frase, por exemplo) podemos entender como esta é formada por diversas fontes (palavras) que possuem características (classes) diferentes (adjetivo, substantivo, etc.). E a meta, em várias aplicações, é a caracterização e classificação de cada uma das fontes, neste caso, as palavras. Desta forma, um problema de lingüística pode ser solucionado de forma equivalente pelas técnicas existentes para separação de fontes.

Este trabalho se dedica então a mostrar os conceitos e fundamentos de ICA, uma das principais técnicas de separação cega de fontes, bem como fazer uma revisão do estado da arte sobre a aplicação de tais métodos em processamento da fala/voz e lingüística. É mostrado um exemplo em separação de sinais de voz, ilustrando a aplicação da técnica e métodos em problemas relacionados a lingüística e processamento da fala.

O restante deste capítulo se organiza da seguinte forma. Na Seção 1 é descrita a modelagem e os conceitos do problema de recuperação de fontes. A Seção 2 é dedicada à técnica de análise por componentes independentes e suas principais medidas de independência. Uma aplicação em processamento de voz usando é apresentada na Seção 3, mostrando o potencial de utilização de ferramentas de processamento inteligente de sinais. A Seção 4 traz um breve apanhado de trabalhos recentes que utilizam ICA como uma ferramenta útil a problemas de reconhecimento da fala e análise lingüística. O capítulo é concluído na Seção 5.

1 RECUPERAÇÃO DE FONTES: CONCEITOS E MODELAGEM MATEMÁTICA

Seja um mapeamento de um número K de fontes (sinais) e V sinais de ruído, dado por:

$$\mathbf{x}(n) = \mathbf{F}(\mathbf{a}(n), \mathbf{v}(n), n), \quad (1)$$

em que o índice n indica dependência temporal do mapeamento e os diversos sinais são representados pelos vetores dados por:

$$\begin{aligned} \mathbf{a}(n) &= [a_1(n) \quad a_2(n) \quad \dots \quad a_K(n)]^T \\ \mathbf{v}(n) &= [v_1(n) \quad v_2(n) \quad \dots \quad v_V(n)]^T \end{aligned} \quad (2)$$

em que $\mathbf{a}(n)$ e $\mathbf{v}(n)$ são os vetores das fontes e das amostras de ruído, respectivamente.

Então, o sistema $\mathbf{F}(\bullet, \circ, \square)$ fornece sinais de saída, chamados de misturas, cuja representação vetorial é dada

$$\text{por } \mathbf{x}(n) = [x_1(n) \quad x_2(n) \quad \dots \quad x_M(n)]^T.$$

A formulação mostrada na Eq. (1) permite representar o problema da maneira mais genérica possível, isto é:

- o mapeamento pode ser não-linear;
- pode haver dependência temporal;
- as fontes e sinais de ruído podem ser inseridos tanto de forma aditiva quanto multiplicativa.

Entretanto, algumas suposições são geralmente feitas para tornar o problema mais “tratável” e também mais próximo da maioria dos problemas práticos. É bastante usual a consideração de que o sistema $\mathbf{F}(\bullet, \circ, \square)$ é linear e invariante com o tempo. Além disto, as fontes são geralmente consideradas mutuamente independentes e independentes do ruído. Outras considerações que, eventualmente, simplificam a formulação são feitas de acordo com o problema tratado.

Com isto, devido em grande parte à sua simplicidade mas também à sua aplicabilidade, o modelo de um sistema linear invariante no tempo com fontes estacionárias, tem sido o grande foco de estudo da grande maioria dos trabalhos dedicados à BSS. Enfim, considerando-se a linearidade do sistema de mistura e estacionariedade das fontes, pode-se escrever o modelo resultante da seguinte forma:

$$\mathbf{x}(n) = \mathbf{H}\mathbf{a}(n) + \mathbf{v}(n), \quad (3)$$

em que \mathbf{H} é uma *matriz de mistura convolutiva* de ordem $M \times K$ e $V = M$, o que significa que o ruído é inserido somente nos sensores e de forma aditiva. As outras hipóteses consideradas corriqueiramente são as seguintes:

- H1. Os sinais (fontes) $v_i(n)$ e $v_j(n)$ são estacionários e de média nula.
- H2. As fontes $a_i(n)$ são estatisticamente independentes.
- H3. Os sinais de ruído $v_j(n)$ são estatisticamente independentes e independentes das fontes $a_i(n)$.
- H4. O número de sensores é maior ou igual ao número de fontes: $M \geq K$.

Pode-se então esboçar o esquema de separação cega de fontes conforme a Figura 1.

Assim, a meta da separação linear é obter uma matriz de separação \mathbf{W} tal que

$$\mathbf{y}(n) = \mathbf{W}\mathbf{x}(n) = \hat{\mathbf{a}}(n)$$

seja uma boa estimativa dos sinais das fontes.

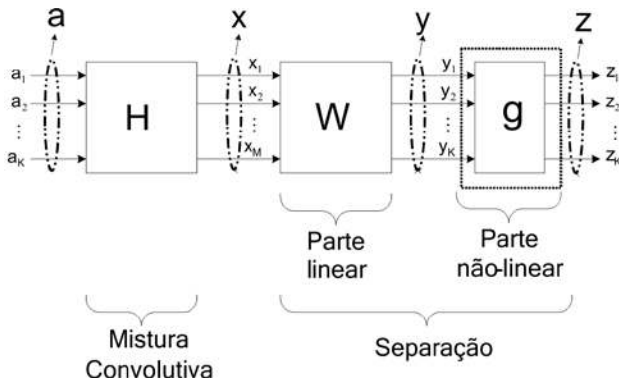


Figura 1 - Esquema do processamento de separação cega de fontes.

De acordo com a hipótese H4 nota-se que, uma vez que o sistema não é sub-determinado, e se a matriz \mathbf{H} é identificada, pode-se ter uma inversão perfeita $\mathbf{W} = \mathbf{H}^{-1}$ na ausência de ruído. Quando o ruído não tem potência baixa o suficiente para ser considerado desprezível, não é possível inverter o sistema perfeitamente, e recuperar exatamente as fontes não é factível. Uma maneira de considerar o ruído de tal forma que o sistema tenha a mesma representação do caso sem ruído é escrever a Eq. (3) da seguinte forma:

$$\mathbf{x}(n) = \mathbf{H}^\dagger \mathbf{a}^\dagger(n) \quad (5)$$

em que

$$\mathbf{H}^\dagger = [\mathbf{H} \quad \mathbf{I}_M]$$

$$\mathbf{a}^\dagger(n) = [\mathbf{a}^T(n) \quad \mathbf{v}^T(n)] \quad (6)$$

em que \mathbf{I}_M é a matriz identidade de ordem M .

Esta maneira de reescrever o modelo facilita visualizar que, no caso com ruído, não é possível inverter perfeitamente de forma linear o sistema (matriz de mistura convolutiva), já que a matriz \mathbf{H}^\dagger tem ordem $M \times (M + K)$. Diante da impossibilidade natural de disponibilizar um aumento na dimensionalidade (ordem) da matriz de mistura convolutiva, o que permitiria uma separação de fontes por um hiperplano, se faz necessária a inserção de um dispositivo não-linear (\mathbf{g}) na parte responsável pela separação. Este requisito deve-se ao fato que técnicas não-lineares são capazes de realizar a separação de sistemas sub-determinados, caso da Eq. (5) (Hyvärinen et al., 2001; Cichocki e Amari, 2002). Vale a pena ressaltar que a real localização do dispositivo não-linear pode não ser exatamente aquela representada na Figura 1, podendo o mesmo estar localizado antes da parte linear da separação (Kofidis, 2001; Hyvärinen e Oja, 2000).

Um ponto de extrema importância na resolução do problema de BSS diz respeito à indeterminação quanto ao escalonamento e à ordem das fontes na saída da matriz de separação \mathbf{W} (Haykin, 2000). Esta indeterminação ocorre sempre que não se possui informação adicional sobre o tipo de sistema (caso comum em BSS). Isto se deve ao fato de que a solução da Eq. (4) não será alterada se uma fonte sofrer uma multiplicação por um escalar e a coluna correspondente da matriz \mathbf{H} for dividida pelo mesmo escalar. O mesmo acontece se houver uma permutação das fontes e uma correspondente permutação das colunas da matriz \mathbf{H} . Assim, o melhor que se pode conseguir é uma solução escalonada e permutada das fontes a_i de tal forma que (Hyvärinen e Oja, 2000):

$$\mathbf{y}(n) = \mathbf{P}\mathbf{D}\mathbf{a}(n) \quad (7)$$

em que \mathbf{P} é uma matriz de permutação de ordem $K \times K$ e \mathbf{D} é uma matriz diagonal e inversível de ordem $K \times K$.

Em problemas de natureza prática, a questão do escalonamento não impõe um severo empecilho já que se pode sempre realizar, de maneira bem simples, um ajuste no ganho dos sinais até atingir-se o nível original através de um controle automático de ganho (CAG). Já no caso da indeterminação da ordem, na maioria das aplicações isto não é relevante, e nos casos em que a identificação da fonte necessita ser realizada, como por exemplo, em problemas de detecção multiusuário em comunicações, codificações especiais são utilizadas para permitir uma classificação direta entre as estimativas e as fontes.

1.1 Estatísticas necessárias ao processamento

Seja o caso sem ruído e com $K = M$, a recuperação das fontes é feita uma vez que a matriz de mistura convolutiva é identificada. Assim sendo, resta saber quais informações estatísticas são apropriadas para que a identificação mencionada seja alcançada.

Seja a matriz de autocorrelação do vetor de observações $\mathbf{x}(n)$ dada por:

$$\mathbf{R}_x = E \left\{ \mathbf{x}(\#)\mathbf{x}^T(n) \right\} = \mathbf{H}\mathbf{R}_a\mathbf{H}^T + \mathbf{H}\mathbf{H}^T, \quad (8)$$

em que o índice n é suprimido devido à estacionariedade da matriz de autocorrelação; $\mathbf{R}_a = \mathbf{I}$ de acordo com as hipóteses (H1-H4) consideradas; e supondo as fontes com potência unitária.

A Eq. (8) sugere que a matriz de mistura convolutiva pode ser identificada através da matriz de correlação dos dados. Também é fácil verificar que uma multiplicação matricial $\mathbf{H}\mathbf{Q}^T$, em que \mathbf{Q} é uma matriz ortogonal, também soluciona a Eq. (8). Isto significa que o uso de \mathbf{R}_x permite \mathbf{H} ser identificado a menos de um fator ortogonal.

Observando-se novamente a Eq. (8), pode-se intuir que a matriz de mistura convolutiva pode ser identificada como uma raiz quadrada de \mathbf{R}_x , ou seja $\mathbf{H} = \mathbf{R}_x^{\frac{1}{2}}$. A extração da raiz quadrada de uma matriz é realizada através da decomposição da matriz em função de seus autovalores (SVD, *Singular Value Decomposition*). De fato, escrevendo-se a matriz \mathbf{H} como

$$\mathbf{H} = \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{V}^T, \quad (9)$$

em que \mathbf{U} e \mathbf{V} são matrizes retangulares de ordem $K \times M$, tais que $\mathbf{U}\mathbf{U}^T = \mathbf{V}\mathbf{V}^T = \mathbf{I}_M$ e $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}_K$, compostas dos autovetores esquerdos e direitos, respectivamente. $\mathbf{\Lambda}$ é uma matriz diagonal contendo os K autovalores não-nulos da matriz $\mathbf{H}\mathbf{H}^T$ (Golub e Loan, 1996; Lacoume et al., 1997).

Então, extrair a raiz quadrada de uma matriz é extrair a raiz quadrada de cada um dos seus autovalores da sua representação em SVD. Reescrevendo a matriz de autocorrelação utilizando-se a Eq. (9) tem-se:

$$\begin{aligned} \mathbf{R}_x &= \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{V}\mathbf{V}^T\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}^T, \\ &= \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \end{aligned} \quad (10)$$

Da equação acima, verifica-se então que a matriz de autocorrelação permite estimar as matrizes \mathbf{U} e $\mathbf{\Lambda}$. Como a matriz \mathbf{V} não aparece na Eq. (10), esta matriz e, por consequência, a matriz \mathbf{H} , não são identificáveis através de estatísticas de segunda ordem (SOS, *Second Order Statistics*), ou seja, utilizando somente a matriz de autocorrelação (Lacoume et al., 1997).

Desta forma, o máximo que se consegue realizar através do uso da matriz de correlação é encontrar uma matriz de separação que é a inversa da matriz de mistura convolutiva identificada, ou seja $\mathbf{T} = \mathbf{R}_x^{\frac{1}{2}}$, as observações na saída do sistema de mistura são representadas por:

$$\bar{\mathbf{x}}(n) = \mathbf{T}\mathbf{x}(n), \quad (11)$$

de tal forma que

$$\begin{aligned} \mathbf{R}_{\bar{\mathbf{x}}} &= \mathbf{T}\mathbf{R}_x\mathbf{T}^T \\ &= \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}^T\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T\mathbf{U}^T\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}, \\ &= \mathbf{I}_K \end{aligned}$$

o que corresponde a *branquear* os dados, já que sinais de média zero são considerados brancos se a sua matriz de autocorrelação é a matriz identidade (Papoulis, 1991; Hyvärinen et al., 2001; Kofidis, 2001; Haykin, 2000; Lacoume et al., 1997).

O processamento realizado pela Eq. (11) corresponde à projeção dos dados $\mathbf{x}(n)$ sobre as principais direções orientadas por \mathbf{R}_x , que são determinadas pelos autovetores

associados aos autovalores mais importantes da SVD da matriz de autocorrelação dos dados (não-branqueados), o que corresponde à conhecida técnica estatística de *análise de componentes principais* (PCA, sigla em inglês) (Coppi & Bolasco, 1989). De uma forma mais direta, os dados são projetados nas direções determinadas pelos principais autovetores de \mathbf{R}_x . O vetor $\bar{\mathbf{x}}$, cujos elementos são os componentes principais de \mathbf{x} , é frequentemente referenciado como padronizado ou esferatizado (Lacoume et al., 1997; Haykin, 2000; Kofidis, 2001).

Como a matriz de autocorrelação é definida positiva, a matriz de branqueamento, também conhecida por *transformação de Mahalanobis* pode ser calculada por (Cichocki e Amari, 2002):

$$\mathbf{T} = \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{U}^T. \quad (13)$$

O processamento de esferatização (*sphering*) reduz o problema para um no qual a matriz de mistura é ortogonal:

$$\bar{\mathbf{x}}(n) = \mathbf{T}\mathbf{H}\mathbf{a}(n) = \mathbf{Q}\mathbf{a}(n) \quad (14)$$

em que $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$. Este tipo de processamento apresenta uma complexidade muito menor para sua solução, sendo assim um importante aspecto do ponto de vista computacional.

É possível mostrar que se consegue encontrar as informações restantes sobre a matriz \mathbf{H} (matriz \mathbf{V}) através da imposição de diferentes correlações para as diferentes fontes associadas a diferentes atrasos temporais. Entretanto, isto implica que ao utilizar somente SOS podemos apenas separar (ou classificar) fontes com diferentes formas espectrais, o que se mostra uma restrição bastante forte ao processamento e aplicabilidade do método (Cavalcante, 2004).

Este ponto ilustra o limite do método através da utilização de PCA, que emprega somente estatísticas de segunda ordem. Neste caso, o uso de SOS somente permite a separação cega quando as fontes são gaussianas, pois neste caso os sinais podem ser completamente descritos através de suas médias e variâncias. De fato, nas considerações anteriores, apenas a condição de decorrelação foi usada e isto confirma o uso para sinais gaussianos, nos quais a condição de independência e decorrelação coincidem (Picinbono, 1993; Papoulis, 1991). Além do mais, a maioria dos sinais encontrados em aplicações práticas, tais como voz, música, dados e imagens, são não-gaussianos.

Desta forma, é necessário utilizar algum critério que leve em conta mais estatísticas dos sinais das fontes que apenas os de segunda ordem. Esta é a principal característica da *análise por componentes independentes*, assunto explorado na próxima seção.

2 ANÁLISE POR COMPONENTES INDEPENDENTES

A técnica ICA (do inglês *Independent Component Analysis*) tem como premissa a hipótese de que as fontes são mutuamente independentes. Informalmente, isto equivale a dizer que os sinais-fonte não dependem um do outro do ponto de vista estatístico. Procura-se então obter um conjunto de fontes também independentes na saída dos filtros de separação mesmo que seus espectros não sejam distinguíveis, premissa básica da técnica PCA (Comon, 1994; Kofidis, 2001; Cao e Liu, 1996; Hyvärinen e Oja, 2000). Para isso, a única restrição imposta sobre a função de densidade de probabilidade (FDP) das fontes é de que as mesmas não sejam gaussianas, ou então que, no máximo uma das fontes seja gaussiana (Kofidis, 2001; Hyvärinen & Oja, 2000).

A imposição de não-gaussianidade se deve ao fato de que uma soma de variáveis aleatórias gaussianas fornece uma distribuição de variáveis aleatórias conjuntamente gaussianas, o que impossibilita inferir qualquer informação sobre as fontes somente a partir das observações (Kofidis, 2001; Hyvärinen & Oja, 2000).

De forma bastante intuitiva, pode-se afirmar que o ponto chave da ICA é a não-gaussianidade das fontes. Então, de maneira muito natural, é esperado que os critérios utilizados para obter a separação cega levem em consideração esta característica dos sinais e busquem explorá-la.

O problema que surge a partir de agora se reflete na seguinte pergunta:

Como medir a não-gaussianidade das fontes e utilizar o fato de que as mesmas são estatisticamente independentes para separá-las/classificá-las?

Para responder a tal questão, são colocados, a seguir, os fundamentos matemáticos e estatísticos da técnica de análise por componentes independentes que se baseiam na não-gaussianidade e na independência estatística das fontes.

2.1 Independência estatística

Supor que os sinais das fontes são independentes significa, estatisticamente, que a fdp conjunta do vetor $\mathbf{a}(n)$ é igual ao produto das fdps marginais dos componentes do vetor K -dimensional. Ou seja, deve-se ter:

$$p_A(\mathbf{a}) \square \prod_{i=1}^K p_{A_i}(a_i) \quad (15)$$

No caso das estimativas das fontes, os sinais de saída oriundos da aplicação de uma transformação sobre sinais contaminados visando a separação das fontes também devem apresentar a mesma relação entre as fdps conjunta e marginais para que a separação seja assegurada, ou seja:

$$p_Y(\mathbf{y}) \square \prod_{i=1}^K p_{Y_i}(y_i) \quad (16)$$

Em termos mais precisos, a meta de ICA é, sob a hipótese de independência das fontes, encontrar uma matriz de separação cujas saídas sejam também fontes mutuamente independentes.

Em termos de estatísticas dos sinais, se as Eqs. (15) e (16) são válidas, tem-se, entre outros resultados:

$$E\{y_1 \cdot y_2 \cdots y_K\} = E\{y_1\} \cdot E\{y_2\} \cdots E\{y_K\} \quad (17)$$

De um ponto de vista geométrico, pode-se caracterizar a diferença entre ICA e PCA da seguinte maneira: PCA procura reduzir a dimensão (para efetuar a separação) dos dados projetando-os em uma dimensão de maneira que a variância dos dados seja máxima (direções determinadas pelos maiores autovalores) e ICA procura reduzir a dimensão dos dados de tal forma que a estrutura dos dados seja preservada ao máximo. A Figura 2 ilustra este conceito para dados bidimensionais. Neste caso, as projeções estão representadas por eixos ortogonais, o que não é regra geral.

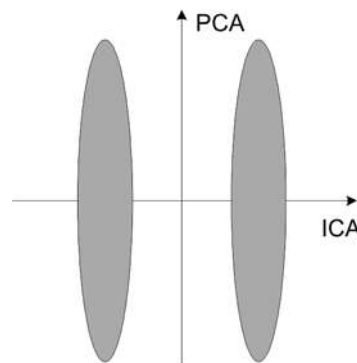


Figura 2 - Visão geométrica da diferença entre ICA e PCA.

Com isso é necessário então discutir algumas medidas de independência para permitir a criação de critérios de separação/classificação, as quais são discutidas na sequência.

2.2 Entropia e informação mútua

Sabendo que a entropia de uma variável aleatória (V.A.), segundo a teoria de Shannon (Shannon, 1948), fornece uma medida da quantidade média de incerteza contida na V.A., pode-se definir, para um vetor de V.A.'s contínuas \mathbf{x} , sua entropia $H(\mathbf{x})$ da seguinte forma:

$$H(\mathbf{x}) \square -E\{\ln[p_{\mathbf{x}}(\mathbf{x})]\} = \int_{-\infty}^{\infty} p_{\mathbf{x}}(\mathbf{x}) \ln[p_{\mathbf{x}}(\mathbf{x})] d\mathbf{x} \quad (18)$$

Seja agora outro vetor de V.A.'s y , e $p_{X|Y}(\mathbf{x}, \mathbf{y})$ a fdp de \mathbf{x} condicionada a \mathbf{y} . Uma medida da incerteza remanescente em \mathbf{x} após a observação de \mathbf{y} é dada pela entropia condicional, representada por:

$$H(\mathbf{x}|\mathbf{y}) = -E\{\ln[p_{X|Y}(\mathbf{x}|\mathbf{y})]\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{X,Y}(\mathbf{x}, \mathbf{y}) \ln[p_{X|Y}(\mathbf{x}|\mathbf{y})] d\mathbf{x} d\mathbf{y}. \quad (19)$$

em que $p_{X,Y}(\mathbf{x}, \mathbf{y}) = p_{X|Y}(\mathbf{x}|\mathbf{y})p_Y(\mathbf{y})$ é a fdp conjunta de \mathbf{x} e \mathbf{y} . Então, a conhecida relação (Gibson, 1993; Haykin, 1998)

$$I(\mathbf{x}, \mathbf{y}) = H(\mathbf{x}) - H(\mathbf{x}|\mathbf{y}). \quad (20)$$

representa a informação contida em \mathbf{x} após a observação de \mathbf{y} . Esta grandeza é chamada de *informação mútua* entre \mathbf{x} e \mathbf{y} .

A informação mútua será de grande importância porque ela é sempre não-negativa e assume o valor zero se e somente se \mathbf{x} e \mathbf{y} são independentes. De fato, isso é esperado já que, para variáveis independentes, a observação de uma delas não fornece informação alguma sobre a outra. Isto é fácil de verificar através da Eq. (20) que, para $I(\mathbf{x}, \mathbf{y}) = 0$, tem-se $H(\mathbf{x}) = H(\mathbf{x}|\mathbf{y})$. Logo, a informação mútua apresenta-se como uma referência confiável para independência estatística.

2.3 Divergência de Kullback-Leibler

Um problema freqüentemente encontrado em inferência estatística é a definição da medida de similaridade entre funções. A Divergência de Kullback-Leibler (KLD, *Kullback-Leibler Divergence*) é uma medida de similaridade entre duas funções estritamente positivas (Kullback e Leibler, 1951). É também chamada de entropia diferencial ou ainda entropia cruzada (Deco e Obradovic, 1996).

Matematicamente, a definição da KLD pode ser explicitada como:

$$D(p_X(\mathbf{x}) \| g_X(\mathbf{x})) = \int_{-\infty}^{\infty} p_X(\mathbf{x}) \cdot \ln \left[\frac{p_X(\mathbf{x})}{g_X(\mathbf{x})} \right] d\mathbf{x}. \quad (21)$$

em que $g_X(\square)$ e $p_X(\square)$ são duas funções estritamente positivas.

É bastante usual a utilização da KLD para a comparação entre duas funções de densidade de probabilidade. Neste caso, a fdp $g_X(\square)$ realiza o papel de uma função de referência. Deve-se observar que $D(p_X(\mathbf{x}) \| g_X(\mathbf{x})) \neq D(g_X(\mathbf{x}) \| p_X(\mathbf{x}))$, significando que $D(\bullet \| \circ)$ não é uma medida de distância, embora seja comum na literatura ser referenciada como distância de Kullback-Leibler. Encontra-se ainda a denominação de divergência assimétrica por ser, matematicamente, uma semi-distância (Kofidis, 2001).

A KLD tem algumas propriedades importantes (Haykin, 1998):

P1. é sempre de valor positivo ou zero; KLD é zero para o caso específico de $p_X(\mathbf{x}) = g_X(\mathbf{x})$.

P2. é invariante com relação às seguintes mudanças nos componentes do vetor \mathbf{x} ;

- permutação de ordem
- escalonamento de amplitude
- transformação montônica não-linear

Usando a KLD, pode-se ainda escrever a informação mútua como:

$$I(\mathbf{x}, \mathbf{y}) = \int \int p_{X,Y}(\mathbf{x}, \mathbf{y}) \ln \left[\frac{p_{X,Y}(\mathbf{x}, \mathbf{y})}{p_X(\mathbf{x}) \cdot p_Y(\mathbf{y})} \right] d\mathbf{x} d\mathbf{y} \quad (22)$$

$$= D(p_{X,Y}(\mathbf{x}, \mathbf{y}) \| p_X(\mathbf{x}) p_Y(\mathbf{y}))$$

o que indica que a informação mútua é uma medida da dependência estatística. Além disso, pode-se ainda definir de forma mais elegante a informação mútua entre os componentes de um vetor K -dimensional (qualquer dimensão) \mathbf{y} como sendo a KLD entre $p_Y(\mathbf{y})$ e $\tilde{p}_Y(\mathbf{y})$ dada por (Kofidis, 2001):

$$I(\mathbf{y}) = D(p_Y(\mathbf{y}) \| \tilde{p}_Y(\mathbf{y}))$$

$$= \int \int \dots \int p_Y(y_1, y_2, \dots, y_K) \ln \left[\frac{p_Y(y_1, y_2, \dots, y_K)}{\prod_{i=1}^K p_Y(y_i)} \right] dy_1 dy_2 \dots dy_K \quad (23)$$

que assume valor igual à zero se e somente se todos os componentes de \mathbf{y} são mutuamente independentes. Através da equação anterior, pode-se ainda escrever que (Haykin, 1998):

$$I(\mathbf{y}) = H(\mathbf{y}) - \sum_{i=1}^K H(y_i) \quad (24)$$

o que significa que minimizar a informação mútua entre os componentes do vetor \mathbf{y} representa tornar a entropia de \mathbf{y} o mais próximo possível da soma de suas entropias marginais.

2.4 Negentropia

A *negentropia* é uma medida de “não-gaussianidade” baseada na medida de entropia diferencial (Touzni, 1998; Hyvärinen e Oja, 2000; Comon, 1994).

Matematicamente, a negentropia é a diferença entre a entropia da V.A. multidimensional \mathbf{y} e a entropia de uma V.A. \mathbf{y}^G de distribuição gaussiana e com os mesmos momentos de ordem um e dois (média e variância) de \mathbf{y} . Assim, temos que $p_{\mathbf{y}^G}(\mathbf{y}) = N(\mu_{\mathbf{y}}, \sigma_{\mathbf{y}}^2)$, em que $\mu_{\mathbf{y}}$ e $\sigma_{\mathbf{y}}^2$ são respectivamente a média e a variância da variável \mathbf{y} , sendo sua negentropia dada por:

$$N_G(\mathbf{y}) \square H(\mathbf{y}^G) - H(\mathbf{y}). \quad (25)$$

Desta forma, a negentropia é uma grandeza não-negativa pois uma variável gaussiana possui sempre a maior entropia para uma variância finita (Picinbono e Barret, 1990). Através da divergência de Kullback-Leibler, a definição de negentropia também pode ser dada por:

$$N_G(\mathbf{y}) \square D(p_Y(\mathbf{y}) \| p_{Y^G}(\mathbf{y})). \quad (26)$$

O uso desta medida torna o processamento de separação das fontes mais robusto, uma vez que a medida de negentropia é um estimador ótimo de não-gaussianidade (Hyvärinen e Oja, 2000). Além disso, a construção de um funcional que maximize a negentropia e, por consequência, torne as fontes estimadas na saída do sistema de separação o menos gaussianas possível, é de fácil dedução (Hyvärinen et al., 2001).

Entretanto, a desvantagem de tal método reside no fato de que a maximização de um funcional que leva em conta a negentropia pressupõe o conhecimento ou a estimativa da fdp das fontes na saída do sistema de separação. Como geralmente não se dispõe do conhecimento a priori das fontes, esta estimativa possivelmente deve ser realizada de forma não-paramétrica.

2.5 Funções de contraste

As funções de contraste foram formalmente definidas por Comon (1994) e se referem a um conjunto de funções utilizadas como critérios de otimização, as quais atingem seu mínimo somente quando a total separação das fontes é alcançada. De uma certa maneira, as funções de contraste já eram utilizadas no processamento autodidata de sinais, mas não havia um critério específico para determinar quais funções poderiam ser utilizadas como contraste.

Uma função $\Psi(\square)$, no espaço de K fdps (distintas ou não) é dita ser um contraste se respeitadas as seguintes condições (Comon, 1994; Kofidis, 2001):

C1. $\Psi(p_Y)$ é invariante a permutações:

$$\Psi(p_{\mathbf{P}Y}) = \Psi(p_Y) \quad \text{para qualquer matriz de permutação } \mathbf{P}$$

em que $\Psi(p_Y)$ representa uma função definida a partir das variáveis da fdp p_Y e $\mathbf{P}Y$ representa uma variável permutada de \mathbf{y} .

C2. $\Psi(p_Y)$ é invariante a mudanças de escala:

$$\Psi(p_{\mathbf{D}Y}) = \Psi(p_Y) \quad \text{para qualquer matriz diagonal } \mathbf{D}$$

em que $\mathbf{D}Y$ representa uma variável escalonada de \mathbf{y} .

C3. Se \mathbf{y} possui componentes independentes, então

$$\Psi(p_{\mathbf{W}Y}) \leq \Psi(p_Y) \quad \text{para qualquer matriz inversível } \mathbf{W}$$

A igualdade em C3 deve ser respeitada se e somente se a matriz \mathbf{W} é uma matriz de permutação generalizada, ou seja, $\mathbf{W} = \mathbf{P}\mathbf{D}$, em que \mathbf{P} é uma matriz de permutação e \mathbf{D} é uma matriz diagonal inversível.

Utilizando-se a definição de informação mútua através da KLD da Eq. (22), e a propriedade P2 da KLD, mostra-se de forma direta que o negativo da informação mútua é uma função contraste para o caso de análise por componentes independentes (Comon, 1994). Ou seja, pode-se escrever

$$\Psi_{ICA}(p_Y) = -I(\mathbf{y}). \quad (27)$$

Esta equação é conhecida como contraste da informação mútua negativa (NMI, *Negative Mutual Information*). Pode-se, ainda, simplificar a expressão do contraste para a NMI de forma a explicitar a dependência com a matriz de separação. Assim, temos (Picinbono, 1993):

$$\Psi_{ICA}(\mathbf{W}) = \ln[\det(\mathbf{W})] - E \left\{ \ln \left[\prod_{i=1}^K p_{Y_i}(y_i) \right] \right\}. \quad (28)$$

2.6 Comentários

Há uma grande gama de outras medidas estatísticas que podem ser aplicadas ao processamento para garantir uma separação/classificação dos sinais de diferentes fontes. Entretanto, para a maioria dos problemas de interesse prático em processamento da fala e lingüística, os modelos lineares e as medidas descritas até aqui se adequam bastante bem para prover a solução de interesse.

Para efeito de uma maior descrição de métodos bem como uma maior revisão das técnicas existentes em ICA o leitor pode verificar as referências Hyvärinen et al. (2001) e Cichocki e Amari (2002).

A seguir, é mostrada uma aplicação em um sistema de mistura de sinais de voz. Problema recorrente tanto em aplicações de engenharia quanto de lingüística.

3 APLICAÇÃO EM RECUPERAÇÃO DE SINAIS DE VOZ

Uma das aplicações clássicas em separação de fontes é derivada diretamente do problema *cocktail party*. A detecção e separação de sinais de voz em um ambiente qualquer apresenta-se como um cenário típico para a aplicação de técnicas de ICA. A Tabela 1 mostra os parâmetros do sistema de mistura/separação.

A potência do ruído em cada sensor é dada através da relação sinal-ruído (SNR, *Signal-to-Noise Ratio*), definida como:

$$\text{SNR (dB)} = 10 \log_{10} \left(\frac{\sigma_a^2}{\sigma_v^2} \right), \quad (29)$$

em que σ_v^2 é a potência (variância) do ruído e σ_a^2 é a potência do sinal de interesse. A Figura 3 ilustra as fontes originais dos sinais de voz. Neste caso, estamos supondo que não há nenhum conhecimento sobre as fontes, nem mesmo sobre suas características espectrais, o que permitiria a utilização de técnicas de separação mais simples. Neste cenário, faz-se necessário o uso de técnicas de ICA.

Tabela 1: Parâmetros do sistema de mistura/separação de processamento de sinais de voz.

Número de fontes	4
Número de sensores	4
Ruído	gaussiano com SNR = 20 dB em cada sensor
Matriz de mistura	mistura instantânea com coeficientes reais aleatórios
Algoritmo de separação	KLD para utilizar estatísticas de ordem dois e de ordem superior

Um detalhe que merece destaque aqui é a consideração de ruído, uma vez que a descrição dos métodos utilizou-se do modelo sem perturbação que não fosse inserida pelas próprias fontes. No caso do exemplo desta seção, ruído significa alguma imperfeição das medidas realizadas no receptor. Isto tem como meta tornar a simulação o mais próximo possível do que se encontra em um sistema de aquisição de dados real, no qual a imperfeição do receptor é um fato que dificilmente se consegue evitar.

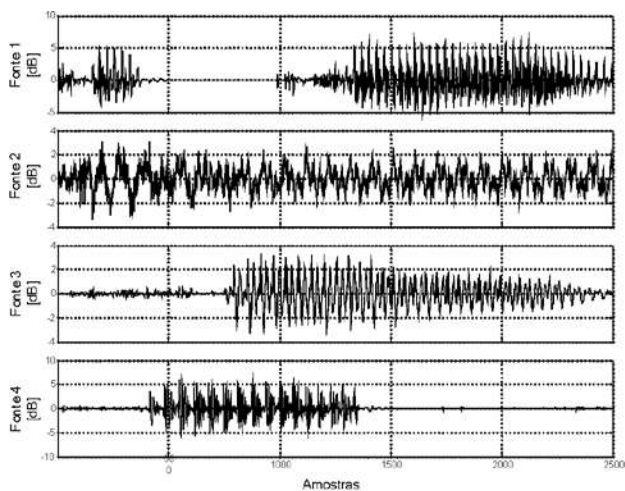


Figura 3 - Fontes dos sinais de voz.

Na Figura 4 observam-se as misturas captadas nos sensores do sistema de separação. Como pode ser facilmente verificado, o conjunto de sinais disponíveis no conjunto de receptores é então uma combinação linear das amostras instantâneas das diversas fontes. Como não se dispõe do sistema de mistura, não é possível inferir informação alguma *a priori* sobre as fontes desconhecidas.

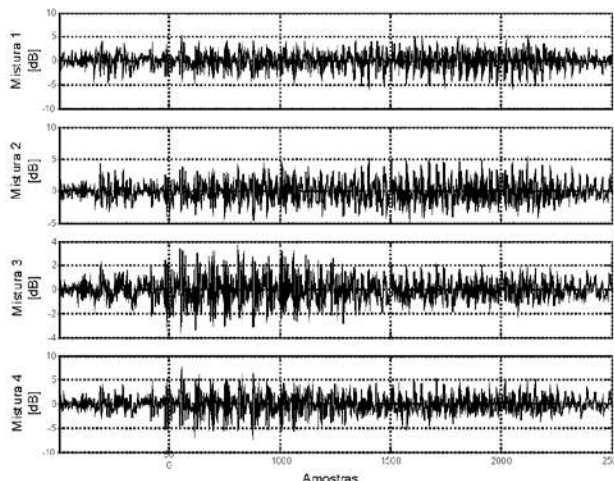


Figura 4 - Misturas disponíveis nos sensores provenientes de sinais de voz.

Neste exemplo, aplicou-se um algoritmo que utiliza funções contraste definidas como função dos cumulantes e das matrizes de correlação dos dados para diferentes atrasos. Desta forma, este algoritmo combina as características de HOS e SOS para extração dos diversos sinais. Pode-se observar, na Figura 5, como o algoritmo é capaz de fornecer boas estimativas das fontes mesmo com a presença de um ruído de potência igual a 20 dB. Claramente, há uma permutação da ordem e escalonamento das estimativa das fontes, conforme previsto pelo modelo teórico.

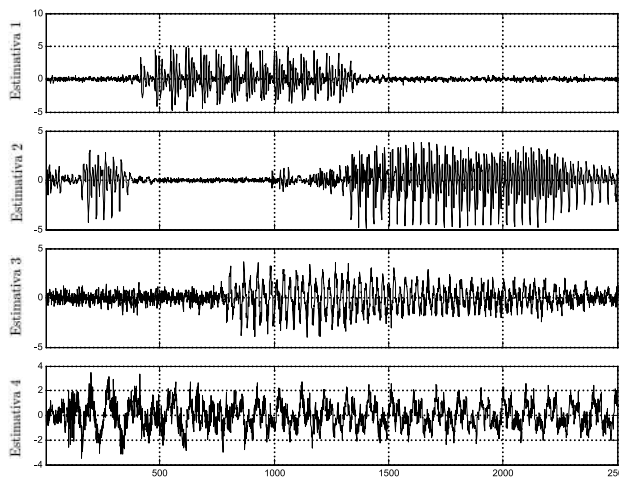


Figura 5: Estimativas das fontes de voz.

Uma forma de medir objetivamente a eficiência da separação é através da matriz global de separação \mathbf{G} , que corresponde à resposta conjunta do sistema, dada por $\mathbf{G} = \mathbf{H}\mathbf{W}$. Uma vez que esta deveria ser, idealmente, a matriz identidade ou uma permutação de suas colunas, a matriz obtida após a convergência ilustra quão próximo da separação perfeita o sistema se encontra. No exemplo em

questão, o módulo dos coeficientes da matriz de separação global é dado por

$$\mathbf{G} = \begin{bmatrix} 0.0094 & 0.0336 & 0.1571 & 1 \\ 1 & 0.0195 & 0.0472 & 0.0079 \\ 0.0048 & 0.0315 & 1 & 0.0124 \\ 0.0095 & 1 & 0.0266 & 0.0132 \end{bmatrix}$$

Vale ressaltar que esta determinação da matriz de separação global só é possível em termos de simulação para se verificar que a separação foi atingida, uma vez que na prática não se dispõe do sistema de mistura para o cálculo da matriz $\mathbf{G} = \mathbf{HW}$.

4. ICA EM PROCESSAMENTO DA FALA E LINGÜÍSTICA

Esta seção traz uma breve descrição de trabalhos recentes nas áreas de processamento de voz/fala e lingüística que utilizam técnicas de ICA, seja como ferramenta principal de análise ou simplesmente como uma técnica de pré-processamento dos dados. É importante destacar que algoritmos de ICA, neste contexto, são basicamente utilizados com o intuito de realizar **extração de características** (*feature extraction*), no jargão da área de reconhecimento de padrões.

Por exemplo, extração de características é uma etapa de crucial importância em problemas de reconhecimento automático da fala (ASR, *automatic speech recognition*). Nesta etapa, o sistema de ASR converte a forma de onda da fala em uma série de vetores de baixa dimensionalidade, em que cada vetor representa um pequeno segmento da fala, a fim de minimizar a carga computacional sobre o classificador subsequente (em geral, baseado em modelos ocultos de Markov). Assim, podemos resumir os papéis principais desempenhados por algoritmos de ICA em tarefas de processamento da fala e lingüística:

(i) extrair informação sobre a estrutura estatística dos sinais de interesse e codificá-la nos vetores ou matrizes (e.g. matriz \mathbf{W}) do modelo matemático adotado para descrever o problema. Estes vetores ou matrizes, por sua vez, podem ser entendidos como detentores de uma versão condensada da informação útil à solução do problema, informação esta que pode ser utilizada, por exemplo, para fins de classificação e reconhecimento da fala ou para inferir a estrutura sintático-semântica por trás de um conjunto de sentenças sendo analisado.

(ii) reduzir o custo computacional envolvido no processamento dos sinais disponíveis, pois em vez de trabalhar diretamente com os sinais propriamente ditos, que são em geral bem ruidosos, pode-se trabalhar com os vetores/matrizes estimados.

4.1 ICA em reconhecimento da fala

Potamitis et al. (2000) utilizam ICA como uma ferramenta estatística útil na obtenção de uma projeção adequada para o vetor de atributos construídos a partir dos coeficientes log-espectrais e cepstrais. O objetivo deste trabalho é reconhecimento automático da fala. Tomando como base o estreito elo entre mecanismos independentes responsáveis pela variabilidade da fala, tais como pulso glotal, função de transferência do trato vocal e radiação do som ao sair pela boca, e o conceito de independência estatística, esses autores propõem uma nova técnica de extração de característica baseada em ICA. O resultado da aplicação desta técnica é uma melhora efetiva no desempenho do sistema de reconhecimento, quando comparado com o desempenho de métodos convencionais baseados apenas em estatísticas de segunda-ordem, tais como PCA, DCT (*discrete cosine transform*) e DLA (*linear discriminant analysis*).

Métodos baseados em estatísticas de segunda ordem, tais como PCA e LDA, vêm sendo utilizados com razoável sucesso no aumento da robustez do processo de reconhecimento automático da fala. Zhao et al. (2003), por exemplo, propõem um novo método de extração de características baseado em ICA, que se mostra ainda mais robusto que alguns anteriormente citados. O método proposto pode extrair efetivamente as componentes de frequência dominante na faixa de 1 a 16Hz a partir do espectro de modulação do sinais de fala. Os autores empreendem uma análise comparativa detalhada entre o método proposto e os tradicionalmente utilizados para o mesmo fim. Resultados preliminares atestam que o método de extração de características baseado em ICA conduz a um desempenho de reconhecimento bem superior aos métodos PCA e LDA em um ambiente ruidoso.

Hoya et al. (2003) apresentam uma nova abordagem para fins de extração de sinais de voz misturados (com ou sem atraso) aplicando ICA em subbandas do sinal de entrada e redes neurais para identificação de locutor. Na abordagem proposta, os sinais de banda passante larga são decompostos em um número fixo de sinais de banda passante estreita. ICA é então aplicada aos sinais decompostos, sendo que um conjunto de redes neurais probabilísticas são usadas para identificar primeiramente o locutor e, em seguida, para compensar “efeitos colaterais”, i.e. efeitos de escala e permutação, ambos particularmente problemáticos para ICA em subbandas. Simulações atestam que o esquema proposto pode efetivamente extrair o sinal de interesse de misturas instantâneas ou com atraso melhor que abordagens tradicionais.

Chien & Chen (2006) apresentam uma nova função-objetivo não-paramétrica para ICA baseada na razão de verossimilhança (*likelihood ratio*). Esta função é obtida a partir de teste de hipótese estatísticos de observações ale-

atórias. Uma função razão de verossimilhança é desenvolvida para medir o grau de confiança na independência entre observações aleatórias. A matriz de separação é estimada através da maximização desta nova função-objetivo e então é aplicada para transformar os dados para um espaço de componentes independentes. Usualmente, o teste de independência é realizado assumindo uma distribuição gaussiana para os dados, o que é inadequado ao uso de ICA. A fim de evitar esta suposição, estes autores propõem uma abordagem não-paramétrica em que as fdp's das V.A.'s envolvidas são determinadas através de métodos baseados em *kernel*. Uma nova técnica de ICA é então realizada por meio desta função objetivo não-paramétrica. Um fato interessante que merece destaque é que os autores aplicaram a nova técnica no aprendizado não-supervisionado de variações desconhecidas de pronúncia. Modelos ocultos de Markov para diferentes categorias (*clusters*) de pronúncia são estimados a fim de caracterizar múltiplas pronúncias, resultando em uma técnica robusta de reconhecimento da fala. A técnica proposta é também aplicada na separação de misturas lineares de sinais de voz e áudio, obtendo resultados melhores que técnicas paramétricas e não-paramétricas baseadas na minimização da informação mútua.

4.2 ICA em análise linguística

A análise de texto simbólico usando métodos estatísticos requer a extração de uma representação numérica ou estatística. Informação contextual tem sido amplamente usada na análise estatística de corpora de linguagem natural. Uma representação útil para palavras pode ser contruída levando-se em consideração o contexto em que elas ocorrem. Por contexto entende-se palavras ao redor da palavra sendo analisada. Isto se baseia no fato empírico e intuitivo de que palavras que ocorrem em contextos similares tendem a ser similares. O tamanho do contexto varia de algumas palavras vizinhas, toda a sentença ou todo o texto.

Seguindo a linha de raciocínio apontada no parágrafo anterior, Honkela & Hyvärinen (2004) usam ICA para encontrar relações sintáticas e semânticas para fins de recuperação da informação através da análise de corpora. Segundo estes autores, ICA possui vantagens claras sobre dois outros métodos comumente utilizados: *análise semântica latente* (LSA, sigla em inglês) e *mapas auto-organizáveis* (SOM, sigla em inglês). LSA é um método simples para geração automática de conceitos que sejam úteis, por exemplo, em tarefas de recuperação da informação. Contudo, estes conceitos não são facilmente interpretáveis por humanos. SOM é uma técnica conexionista que pode ser usada para gerar um diagrama explícito que caracteriza as relações entre as palavras, refletindo categorias sintáticas em nível global e categorias semânticas em nível local (Honkela et al., 1995). A técnica baseada em SOM, contudo,

não provê categorias distintas para as palavras em si. Já ICA, quando aplicada a dados formados por palavras e seu contexto (palavras em torno de uma principal) é capaz de gerar atributos distintos que refletem tanto categorias sintáticas quanto semânticas, e de um modo que seja facilmente interpretável por humanos. Isto é obtido sem qualquer supervisão humana ou uso de informação prévia sobre o corpora, seja esta de natureza morfológica, sintática ou semântica.

Em um trabalho mais recente, Honkela et al. (2004) mostram que ICA aplicada ao contexto de palavras produz características que, ao refletir explicitamente as categorias sintáticas e semânticas das palavras, são facilmente interpretadas por humanos. Este resultado é obtido sem qualquer supervisão humana ou uso de corpora etiquetados (*tagged*), que viesse a ter alguma informação prévia, seja morfológica, sintática ou semântica. Os resultados incluem tanto a emergência de categorias ou características claramente distintivas, quanto uma representação distribuída das mesmas. Isto se baseia no fato de que uma palavra pode pertencer simultaneamente a várias categorias em intensidades variadas. Esses autores argumentam que a abordagem proposta pode prover insights acerca dos potenciais mecanismos cognitivos envolvidos no aprendizado e entendimento da linguagem natural, uma vez que eles demonstram que uma grande parte do conhecimento lingüístico é de natureza emergente e baseada em mecanismos específicos de aprendizagem.

Väyrynen et al. (2004), ao estudarem a linguagem escrita como se ela fosse um sinal multidimensional em vez de uma seqüência de símbolos, mostram que a aplicação de ICA a palavras e seus contextos leva ao surgimento atributos lingüísticos úteis. O grau de similaridade entre os atributos obtidos via ICA e as categorias lingüísticas convencionais das palavras é examinado em detalhes neste estudo. Os resultados obtidos levam à conclusão de que ICA tem melhor desempenho que métodos clássicos, tal como PCA.

Lagus et al. (2004) analisam o uso de morfemas ao analisar em um grande corpus textual da língua finlandesa por meio de ICA. Um morfema é a menor unidade gramatical que se pode identificar. Os autores observam a emergência de representações lingüísticas para os morfemas. Em um nível mais amplo, as principais categorias semânticas são observadas. Em um nível mais detalhado, as componentes independentes obtidas indicam potenciais papéis temáticos que os morfemas podem estar representando. Um ponto sugerido como pesquisa futura por estes autores é a utilização das componentes independentes descobertas em aplicações de processamento da linguagem natural (e.g. reconhecimento automático de locutor).

Kumaran et al. (2005) propõem um novo modelo estatístico de linguagem para capturar dependências temporais longas, largamente presentes na linguagem natural,

e com as quais os humanos são inerentemente capazes de lidar. Por exemplo, dado um conjunto de palavras relacionadas entre si (por exemplo, em uma sentença), um humano pode facilmente identificar o contexto em que tal conjunto está ocorrendo, mesmo que este contexto envolva um longo espaçamento entre as palavras. Esses autores desenvolvem um modelo de linguagem baseado em ICA que captura este tipo de dependência melhor que abordagens convencionais. Além disso, eles mostram que ICA provê mecanismos de decomposição das relações entre palavras que podem ser facilmente interpretados por humanos, quando comparado com outros modelos. A abordagem proposta é então combinada com o modelo N-grama convencional para produzir um modelo de linguagem. Os resultados obtidos mostram a viabilidade do modelo proposto para fins de reconhecimento da fala.

Väyrynen & Lindh-Knuutila (2006) reportam os primeiros resultados da extração de representações semânticas de palavras a partir de corpora paralelos multilíngües. ICA é usada para extrair um certo número de atributos a partir de estatísticas calculadas a partir de palavras em seus respectivos contextos. Os atributos extraídos possuem significado e são multilíngües, enquanto palavras são representadas como conceitos associados. O espaço de atributos criado também é multilíngüe. É interessante observar que, neste espaço, palavras relacionadas em diferentes línguas aparecem próximas umas às outras, tornando possível encontrar traduções para palavras entre as línguas.

Finalmente, Brock (2006) investiga a plausibilidade de se realizar, de modo computacionalmente eficiente, a classificação de dados acústicos de alta dimensionalidade usando coeficientes mel-cepstrais e ICA para extração de características. O autor desenvolve um método para calcular os coeficientes mel-cepstrais de categorias de dados acústicos agrupados por gênero musical ou linguagem falada. Em seguida, ICA é empregada em cada categoria para extrair características (*features*) temporais dos coeficientes em um nível de resolução mais alto. Estes conjuntos de componentes independentes únicas representam os padrões temporais subjacentes a uma determinada classe de dados acústicos. Os resultados obtidos mostram claramente que a unicidade das componentes independentes para cada classe de informação acústica provê um mecanismo legítimo para separação e classificação dos dados.

5 CONCLUSÕES

Este trabalho apresentou uma introdução à técnica de ICA, tipicamente utilizada em problemas de recuperação cega em sistemas de informação e sua aplicabilidade em problemas nas áreas de processamento da fala e linguística. Àqueles interessados em se aprofundar nestes assuntos recomenda-se, além das referências já citados, a leitura

adicional de Torkkola (1999) e Mansour et al. (2000), em que conceitos, idéias, aplicações e técnicas de ICA são reportadas em detalhes.

REFERÊNCIAS BIBLIOGRÁFICAS

- Brock, J. L. *Acoustic Classification Using Independent Component Analysis*. Dissertação de Mestrado, Rochester Institute of Technology, 2006.
- Cao, X.-R.; Liu, R.W. General Approach to Blind Source Separation. *IEEE Transactions on Signal Processing*, 44(3):562-571, 1996.
- Cavalcante, C. C. *Sobre Separação Cega de Fontes: Proposições e Análise de Estratégias para Processamento Multi-Usuário*, Tese de Doutorado, UNICAMP, 2004.
- Chien, J.-T.; Chen, B.-C. A New Independent Component Analysis for Speech Recognition and Separation, *IEEE Transactions on Audio, Speech, and Language Processing*, 14(4):1245-1254.
- Cichocki, A.; Amari, S. *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. John Wiley & Sons, 2002.
- Comon, P. Independent Component Analysis: A New Concept? *Signal Processing*, 36(3):287-314, 1994.
- Coppi, R., & Bolasco, S. (Eds.). *Multway Data Analysis*. Netherlands: Elsevier Science Publishers B.V. (North-Holland), 1989.
- Deco, G.; Obradovic, D. *An Information-Theoretic Approach to Neural Computing*. Springer-Verlag, 1996.
- Gibson, J. D. *Principles of Digital and Analog Communications* (2nd ed.). New Jersey, USA: Prentice-Hall, 1993.
- Golub, G. H.; Loan, C. F. V. *Matrix Computations* (3rd ed.). Baltimore, USA: The Johns Hopkins University Press, 1996.
- Haykin, S. *Neural Networks: A Comprehensive Foundation* (2nd ed.). Prentice Hall, 1998.
- Haykin, S. (Ed.). *Unsupervised Adaptive Filtering (Vol. I: Source Separation)*. John Wiley & Sons, 2000.
- Honkela, T.; Hyvärinen, A. Linguistic Feature Extraction using Independent Component Analysis. *Proceedings of the 2004 IEEE International Joint Conference on Neural Networks (IJCNN'2004)*, vol. 1, p. 279-284.
- Honkela, T.; Hyvärinen, A.; Väyrynen, J. *Emergence of Linguistic Features: Independent Component Analysis of Contexts*. *Proceedings of the Ninth Neural Computation and Psychology Workshop (NCPW9)*, Plymouth, England, p. 1-10, 2004.

- Honkela, T.; Pulkki, V.; Kohonen, T. Contextual Relations of Words in Grimm Tales Analyzed by Self-Organizing Map. *Proceedings of the International Conference on Artificial Neural Networks (ICANN'95)*, vol. II, p. 3-7, 1995.
- Hoya, T.; Barros, A. K.; Rutkowski, T.; Cichocki, A. Speech Extraction based upon a combined subband independent component analysis and neural memory. *Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA'2003)*, pp. 355-360, 2003.
- Hyvärinen, A.; Oja, E. Independent Component Analysis: Algorithms and Applications. *Neural Networks*, 13(4-5):411-430, 2000.
- Hyvärinen, A.; Oja, E.; Karhunen, J.. *Independent Component Analysis*. John Wiley & Sons, 2001.
- Jutten, C., and Karhunen, J., Advances in Blind Source Separation (BSS) and Independent Component Analysis (ICA) for Nonlinear Mixtures. *International Journal of Neural Systems*, 14(5):267-292, 2004.
- Kofidis, E. Blind Source Separation: Fundamentals and Recent Advances. Mini-curso no XIX *Simpósio Brasileiro de Telecomunicações (SBrT 2001)*, Fortaleza-CE, Setembro, 2001.
- Kullback, S.; Leibler, R. A. On Information and Sufficiency. *Annals of Mathematical Statistics*, 22:79-86, 1951.
- Kumaran, R. S.; Narayanan, K.; Gowdy, J.N. Language Modeling using independent component Analysis for Automatic Speech Recognition. *Proceedings of the 13th European Signal Processing Conference (EUSIPCO'2005)*, Antalya, Turkey, September 4-8, p. 1-4, 2005.
- Lacoume, J.-L. ; Amblard, P.-O. ; Comon, P. *Statistiques d'Ordre Supérieur pour le Traitement du Signal*. Paris, Masson, 1997.
- Lagus, K.; Creutz, M.; Virpioja, S. *Latent Linguistic Codes for Morphemes using Independent Component Analysis*. *Proceedings of the Ninth Neural Computation and Psychology Workshop (NCPW9)*, Plymouth, England, p. 1-5, 2004.
- Mansour, A.; Barros, A. K.; Ohnishi, N. Blind separation of sources: Methods, assumptions and applications. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E83-A, no. 8, pp. 1498-1512, 2000.
- Papoulis, A. *Probability, Random Variables and Stochastic Processes* (3rd ed.). McGraw-Hill International, 1991.
- Picinbono, B. *Random Signals and Systems*. Englewood Cliffs, New Jersey: Prentice Hall International Editions, 1993.
- Picinbono, B.; Barret, M. Nouvelle Présentation de la Méthode du Maximum d'Entropie. *Traitement du Signal*, 7(2) :153-158, 1990.
- Potamitis, L.; Fakotamis, N.; Kokkinakis, G. Independent Component Analysis Applied to Feature Extraction for Robust Automatic Speech Recognition, *Electronics Letters*, 36(23):1977-1979, 2000.
- Shannon, C. E. A Mathematical Theory of Communication. *Bell Systems Technical Journal*, Vol. 27, 379-423, 623-656, 1948.
- Torkkola, K. Blind separation for audio signals – are we there yet. *Proceedings of the International Symposium on Independent Component Analysis and Blind Signal Separation (ICA'99)*, Aussois, France, pp. 239–244, January 1999.
- Touzni, A. *Performance et Robustesse en Égalisation Aveugle Spatio-Temporelle*. Tese de doutorado, Université de Cergy-Pontoise, France, 1998.
- Väyrynen, J. J.; Lindh-Knuutila, T. Emergence of multilingual representations by independent component analysis using parallel corpora. *Proceedings of the Ninth Scandinavian Conference on Artificial Intelligence (SCAI 2006)*, p. 101-105, 2006.
- Väyrynen, J.; Honkela, T.; Hyvärinen, A. Independent Component Analysis of Word Contexts and Comparison with Traditional Categories. *Proceedings of the 6th Nordic Signal Processing Symposium (NORSIG'2004)*, Espoo, Finland, p. 300-303, 2004.
- Zhao, J. ; Kuang, J.; Xie, X. Data-driven temporal processing using independent component analysis for robust speech recognition, *Proceedings of the 2003 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT'2003)*, p. 729-732.