

UNIVERSIDADE FEDERAL DO MARANHÃO
CENTRO TECNOLÓGICO
CURSO DE PÓS-GRADUAÇÃO EM ENGENHARIA DE ELETRICIDADE

**EXTRAÇÃO DE SINAIS DE VOZ EM
AMBIENTES RUIDOSOS POR
DECOMPOSIÇÃO EM FUNÇÕES BASES
ESTATISTICAMENTE INDEPENDENTES**

NATÁLIA COSTA LEITE ABREU

São Luís, Ma

2003

EXTRAÇÃO DE SINAIS DE VOZ EM AMBIENTES RUIDOSOS POR DECOMPOSIÇÃO EM FUNÇÕES BASES ESTATISTICAMENTE INDEPENDENTES

Dissertação de Mestrado submetida à Coordenação do Curso de Pós Graduação em Engenharia de Eletricidade da Universidade Federal do Maranhão – UFMA como parte dos requisitos necessários para obtenção do grau de Mestre.

Por

NATÁLIA COSTA LEITE ABREU

Dezembro, 2003

EXTRAÇÃO DE SINAIS DE VOZ EM AMBIENTES RUIDOSOS POR DECOMPOSIÇÃO EM FUNÇÕES BASES ESTATISTICAMENTE INDEPENDENTES

NATÁLIA COSTA LEITE ABREU

Dissertação aprovada em 11 de dezembro de 2003.

Prof. Dr. Allan Kardec Duailibe Barros Filho - UFMA
(Orientador)

Prof. Dr. Henio Henrique Aragão Rego - CEFET-MA
(Membro da Banca Examinadora)

Prof. Dr. Zair Abdelouahab - UFMA
(Membro da Banca Examinadora)

EXTRAÇÃO DE SINAIS DE VOZ EM AMBIENTES RUIDOSOS POR DECOMPOSIÇÃO EM FUNÇÕES BASES ESTATISTICAMENTE INDEPENDENTES

MESTRADO

Área de Concentração: CIÊNCIA DA COMPUTAÇÃO

NATÁLIA COSTA LEITE ABREU

Orientador: Prof. Dr. Allan Kardec Duailibe Barros Filho

**Curso de Pós-Graduação
Em Engenharia de Eletricidade da
Universidade Federal do Maranhão**

Este trabalho é dedicado a toda minha família.
Dedico-o ainda ao meu querido
Henry Felipe.

AGRADECIMENTOS

A Deus.

A minha mãe Zélia Abreu pela insistência comigo fazendo-me voltar atrás em minhas decisões erradas.

A minha mãe “Zizi” (*in memoriam*) e meu pai Francisco Xavier por terem compreendido minha ausência nos finais de semana.

Ao meu orientador, Profº Dr. Allan Kardec Duailibe Barros Filho pela enorme paciência, tolerância e constante estímulo à independência na condução desta pesquisa.

Ao Profº. Nelson Camelo pelos conselhos, apoio e conversas amigas.

A todos os meus amigos do PIB, em especial a Márcio Santos, Fausto Lucena, Maxwell Borges e Ricardo Robson por disponibilizarem tempo para ajudar-me.

Aos amigos Flávia Cristina, Ivanilde Freire, Edil James e Claudionice dos Santos pelos incentivos, longas conversas e orações.

Ao amigo Ronaldo Mourão, pelas providenciais caronas e descontraídas conversas no início do curso.

Ao Instituto Pindorama, meu local de trabalho, pelas horas em que fui liberada para ir a UFMA.

A Coordenação do Curso de Pós-Graduação em especial Alcides Neto.

Enfim, não é fácil listar todos que, holisticamente falando, contribuíram de algum modo para a realização deste trabalho. Assim, antecipo minhas desculpas e agradecimentos aos que não me ocorreram agora e que não figuram na lista acima.

*“Eu vi ainda debaixo do sol
que a corrida não é para os mais ligeiros,
nem a batalha para os mais fortes,
nem o pão para os mais sábios,
nem as riquezas para os mais inteligentes,
mas tudo depende do tempo e do acaso.”*

Eclesiastes 11

RESUMO

A constante busca para aperfeiçoar e estreitar o relacionamento entre homens e máquinas, tornando-o mais natural, não é nenhuma novidade. Conseqüentemente, o reconhecimento da voz possibilitará uma manipulação mais fácil e prática de equipamentos dotados com a capacidade de compreender a fala humana. Neste sentido e utilizando-se dos conhecimentos disponíveis na literatura de como o cérebro humano processa informações, alguns métodos propostos procuram simular computacionalmente essa habilidade, voltados principalmente à extração de um sinal de voz de uma mistura de sons, na tentativa de, por exemplo, aumentar a taxa de reconhecimento e inteligibilidade. A extração da voz pode ser obtida usando medidas de um único ou múltiplos canais. Para extrair uma voz em um único canal, propomos usar as características da voz pelo conceito de codificação eficiente, que procura imitar o modo como o córtex auditivo trata as informações, utilizando-se da técnica de Análise de Componentes Independentes (ICA), obtendo as funções bases dos sinais de entrada e recuperando o sinal estimado, mesmo quando são adicionadas interferências. Através de simulações comprovamos também a eficiência da técnica usada, primeiro, na recuperação de um sinal de voz com a utilização das funções bases de outro sinal e, segundo, frente a efeitos de reverberação. Esta técnica pode ser usada para extrair uma única fala eficazmente, como também prenuncia um modo novo de chegar ao problema de reconhecimento da fala/orador.

Palavras-chaves: Simple Canal de Voz, Cocktail Party, Análise de Componente Independente, Reconhecimento da Fala.

ABSTRACT

The constant search for the improvement and strengthening of the relationship between humans and machines turning it more natural is common place. Consequently, the recognition of speech will turn, easier and practical the handling of equipments supplied with the capacity to understand the human speech. In this sense and with the use of the available knowledge information in the literature as how the human brain processes informations, some suggested methods try to simulate this ability in the computer, especially devoted to the extraction of a speech signal of mixed sounds, attempting, for example to increase the recognition and comprehension rate. The extraction of speech can be obtained by measures of a single-channel or multiple the channels. In order to extract the speech in a single channel, it is proposed here to use the speech characteristics introducing the concept of efficient codification, that tries to imitate the way the auditory cortex gets information using the method of *Independent Component Analysis* (ICA), getting the basis functions of the input signals and retrieving the estimated signal even when we add interferences to it. Our simulations also prove the efficiency of our method against reverberation effects and the recovery of speech signal by the handling of basis function of other speech signals. This technique can be used efficiently both to extract a single speech, as well as highlighting new ways of approaching the speech/speaker recognition problem.

Keywords: Single Channel Speech, Cocktail Party, Independent Component Analysis, Speech Recognition.

SUMÁRIO

| | | |
|-----------|---|-----------|
| 1. | INTRODUÇÃO | 15 |
| 1.1 | OBJETIVO | 16 |
| 1.2 | ESTADO DA ARTE | 16 |
| 1.3 | ORGANIZAÇÃO DO TRABALHO | 19 |
| 2. | O SISTEMA NERVOSO | 21 |
| 2.1 | ORGANIZAÇÃO DO SISTEMA NERVOSO | 21 |
| 2.1.1 | <i>Bilhões de Neurônios</i> | 22 |
| 2.1.2 | <i>Sistema Nervoso Central</i> | 24 |
| 2.1.2.1 | <i>O Cérebro</i> | 24 |
| 2.1.2.2 | <i>Áreas Associativas do Córtex</i> | 26 |
| 2.1.2.3 | <i>Plasticidade Neuronal</i> | 27 |
| 2.2 | NEURÔNIO ARTIFICIAL | 27 |
| 2.2.1 | <i>Redes Neurais Artificiais</i> | 28 |
| 2.3 | SISTEMA SENSORIAL | 31 |
| 2.3.1 | <i>Os Sentidos</i> | 31 |
| 2.3.1.1 | <i>Audição</i> | 32 |
| 2.3.1.2 | <i>Órgãos da Audição</i> | 32 |
| 2.4 | CODIFICAÇÃO EFICIENTE DE SONS NATURAIS | 36 |
| 3. | ANÁLISE DE CENÁRIO AUDITIVO | 42 |
| 3.1 | REVERBERAÇÃO | 43 |
| 3.1.1 | <i>Filtro de Resposta Finita - FIR</i> | 45 |
| 4. | SEPARAÇÃO CEGA DE FONTE (BSS) | 47 |
| 4.1 | FORMULAÇÃO DO PROBLEMA DA BSS | 47 |
| 4.1.1 | <i>Hipóteses Básicas</i> | 48 |
| 4.2 | PCA / ICA | 49 |
| 4.2.1 | <i>Análise de Componentes Principais</i> | 50 |
| 4.2.1.1 | <i>Entendendo PCA</i> | 52 |
| 4.2.2 | <i>Análise de Componentes Independentes</i> | 54 |
| 4.2.2.1 | <i>Definições da Análise de Componentes Independentes</i> | 55 |
| 4.2.2.2 | <i>Branqueamento</i> | 57 |
| 4.2.2.3 | <i>Funções Objetivas</i> | 59 |
| 4.2.2.4 | <i>Algoritmos para ICA</i> | 60 |
| 4.2.2.5 | <i>Aplicações de ICA</i> | 62 |

| | |
|--------------------------------|------------|
| 5. MODELO PROPOSTO..... | .64 |
| 5.1 INTRODUÇÃO | 64 |
| 5.2 O MÉTODO | 65 |
| 6. RESULTADOS..... | .70 |
| 6.1 INTRODUÇÃO | 70 |
| 6.21 DISCUSSÃO | 83 |
| 7. CONCLUSÃO | .85 |
| BIBLIOGRAFIA..... | .86 |

LISTA DE FIGURAS

| | |
|--|----|
| FIGURA 1: NEURÔNIO BIOLÓGICO | 22 |
| FIGURA 2: COMPOSIÇÃO DE UM NEURÔNIO..... | 23 |
| FIGURA 3: CÉREBRO HUMANO..... | 25 |
| FIGURA 4: ÁREAS CEREBRAIS COM SUAS FUNÇÕES ESPECÍFICAS | 26 |
| FIGURA 5: ESTRUTURA DE NEURÔNIO ARTIFICIAL..... | 28 |
| FIGURA 6: ARQUITETURA DE RNA COM DUAS SAÍDAS | 29 |
| FIGURA 7: ÓRGÃOS DA AUDIÇÃO..... | 33 |
| FIGURA 8: ANÁLISE DE TEMPO-FREQÜÊNCIA | 38 |
| FIGURA 9: ANÁLISE DE TEMPO E FREQÜÊNCIA PARA VOCALIZAÇÃO DE ANIMAIS | 39 |
| FIGURA 10: ANÁLISE DE TEMPO E FREQÜÊNCIA PARA SONS AMBIENTAIS..... | 40 |
| FIGURA 11 ANÁLISE DE TEMPO E FREQÜÊNCIA PARA VOZ HUMANA. | 40 |
| FIGURA 12: REPRESENTAÇÃO MONAURAL EM BLOCO | 43 |
| FIGURA 13: REPRESENTAÇÃO DO DECAIMENTO EXPONENCIAL. | 44 |
| FIGURA 14: DIAGRAMA EM BLOCOS DA FUNÇÃO DE TRANSFERÊNCIA..... | 46 |
| FIGURA 15: REPRESENTAÇÃO DE SINAIS NO PROCESSO DE MISTURA E SEPARAÇÃO | 48 |
| FIGURA 16: DIREÇÃO DA EXPANSÃO DO FLUXO DE DADOS..... | 51 |
| FIGURA 17: DIAGRAMA EM BLOCOS PARA ICA | 56 |
| FIGURA 18: DIAGRAMA EM BLOCOS DO SISTEMA..... | 64 |
| FIGURA 19: SINAIS ORIGINAIS E RESPECTIVAS FUNÇÕES BASES OBTIDAS..... | 71 |
| FIGURA 20: ERRO VERSUS NÚMERO DE FUNÇÕES BASES / SEM RUÍDO | 72 |
| FIGURA 21: ERRO VERSUS NÚMERO DE FUNÇÕES BASES / MUSICAL | 73 |
| FIGURA 22: ERRO VERSUS NÚMERO DE FUNÇÕES BASES / VOC. ANIMAL..... | 74 |
| FIGURA 23: ERRO ENTRE SINAL DESEJADO E SINAL RECUPERADO COM AS BASES DE OUTROS SINAIS, SEM RUÍDO ADICIONADO | 75 |
| FIGURA 24: RESPOSTA AO IMPULSO - FILTRO COM 50 ATRASOS..... | 76 |
| FIGURA 25: RESPOSTA AO IMPULSO - FILTRO COM 100 ATRASOS..... | 76 |
| FIGURA 26: RESPOSTA AO IMPULSO - FILTRO COM 150 ATRASOS..... | 77 |
| FIGURA 27: RESPOSTA AO IMPULSO - FILTRO COM 200 ATRASOS..... | 77 |
| FIGURA 28: RESPOSTA AO IMPULSO - FILTRO COM 250 ATRASOS..... | 78 |
| FIGURA 29: SINAL DE VOZ ORIGINAL E COM EFEITO DA REVERBERAÇÃO PARA FILTRO COM 50 ATRASOS | 79 |
| FIGURA 30: SINAL DE VOZ ORIGINAL E COM EFEITO DA REVERBERAÇÃO PARA FILTRO COM 100 ATRASOS | 79 |

| | |
|---|----|
| FIGURA 31: SINAL DE VOZ ORIGINAL E COM EFEITO DA REVERBERAÇÃO PARA FILTRO COM 150 ATRASOS | 80 |
| FIGURA 32: SINAL DE VOZ ORIGINAL E COM EFEITO DA REVERBERAÇÃO PARA FILTRO COM 200 ATRASOS | 80 |
| FIGURA 33: SINAL DE VOZ ORIGINAL E COM EFEITO DA REVERBERAÇÃO PARA FILTRO COM 250 ATRASOS | 81 |
| FIGURA 34: GRÁFICO COMPARATIVO DAS MÉDIAS DOS ERROS SEM E COM REVERBERAÇÃO .. | 82 |
| FIGURA 35: MÉDIA DOS ERROS PARA FILTRO COM 250 ATRASOS..... | 82 |

ABREVIATURAS E SÍMBOLOS

| | |
|------|---------------------------------------|
| AI | Index Articulation |
| AM | Modulação em Amplitude |
| ASA | Auditory Scene Analysis |
| ASR | Automatic Speech Recognition |
| BSS | Blind Source Separation |
| CASA | Computational Auditory Scene Analysis |
| DB | Decibel |
| DFT | Discrete Fourier Transform |
| EEG | Eletroencefalograma |
| F0 | Frequência Fundamental |
| FIR | Finite Impulse Response |
| ICA | Independent Component Analysis |
| IS | Itakura-Saito Distortion |
| K-L | Kullback-Lieber |
| MEG | Magnetoencefalograma |
| MSE | Mean Square Error |
| PCA | Principal Component Analysis |
| RNA | Redes Neurais Artificiais |
| SNC | Sistema Nervoso Central |
| SNP | Sistema Nervoso Periférico |
| SNR | Signal to Noise Ratio |
| s | Fonte de Sinal Original |
| A | Matriz de Mistura |

| | |
|----------------------|---------------------------------|
| x | Sinais Observados / Misturados |
| W | Inversa da Matriz de Mistura |
| u | Sinais Separados |
| λ_p | Autovalor |
| e_i | Autovetor |
| Cx | Matriz de Covariância ou R |
| E | Esperança Matemática |
| I | Matriz Identidade |
| B | Matriz Ortogonal |
| k | Kurtosis |
| \prod | Produtório |
| d_k | Sinal Desejado ou de referência |
| y_k | Sinal Estimado ou Saída |
| ε_k | Erro Instantâneo |
| ε^2_k | Erro Quadrático Instantâneo |
| ε_{\min} | Mínimo Erro Médio Quadrático |
| P | Vetor Coluna |
| ∇ | Gradiente |
| w^* | Vetor Peso Ótimo |

CAPÍTULO 1

INTRODUÇÃO

Os crescentes avanços tecnológicos fazem com que o reconhecimento da fala seja um campo de estudos fascinante e ao mesmo tempo desafiador; uma vez que é bastante grande a gama de aplicações onde o uso da voz tem o papel de agilizar e facilitar a realização de tarefas cotidianas, buscando extrair da fala informações relevantes para a realização do reconhecimento. [1]

Um antigo desejo do homem foi sempre poder controlar suas máquinas por meio da fala. As pesquisas relacionadas ao reconhecimento iniciaram sem a utilização do computador, por meio de experimentos mecânicos que simulavam a produção sonora humana. [2]

Com a evolução da tecnologia, as máquinas predominam em quase todos os cenários. Sendo assim, nada melhor do que dotar tais equipamentos com a capacidade de percepção e compreensão da voz humana- que é a forma mais simples, natural e eficaz do ser humano expressar seus pensamentos e sentimentos-, desta forma humanizando o relacionamento homem-máquina.

Nos ambientes reais, a fala normalmente acontece em meio a interferências acústicas, mas o homem possui uma notável habilidade para separar a voz de um simples orador de outros sons, o que não acontece da mesma forma com os equipamentos, pois possuem baixa robustez a ruído. A extração de voz envolvendo um único canal de entrada (monaural) em um ambiente como o *Cocktail Party* [3] permanece como um desafio para os sistemas computacionais e vários são os métodos propostos que procuram solucionar esse problema, na tentativa de, por exemplo, aumentar a taxa de reconhecimento de voz.

A extração da voz, necessária para o seu reconhecimento, pode ser obtida usando uma ou múltiplas medidas de canais. Alguns métodos envolveram as características

harmônicas da voz humana [4], e outros fazem uso da redundância entre os canais, usualmente por ICA [5]. Para extrair uma voz com um único canal, propomos usar as características da voz pelo conceito de codificação eficiente, que imita o modo de informação do córtex auditivo. Simulações mostram que esta técnica, ICA, pode ser usada para extrair com eficácia uma única voz, como também prenuncia um novo modo de tratar o problema de reconhecimento de voz.

1.1 Objetivo

Fazer a redução de ruído e recuperação de um sinal de voz em um ambiente ruidoso conhecido como *Cocktail Party*. O sinal será decomposto em funções bases obtidas por Análise de Componentes Independentes, ICA, conduzindo este estudo ao problema de reconhecimento de voz/orador.

1.2 Estado da Arte

Atualmente, os estudos pertinentes à construção de sistemas computacionais para a análise de cenário auditivo (CASA) monaural estão em constante desenvolvimento [6]. A seguir são descritos alguns estudos para melhorar a qualidade da fala e inteligibilidade, com soluções envolvendo métodos de subtração de espectros, ou com o uso de estimações da estrutura harmônica da voz humana, por sua frequência fundamental [4], ou ainda por Análise de Componentes Independentes, ICA:

Hu e Wang [7] observaram a deficiência da maioria dos sistemas CASA em lidar com frequências altas e propuseram um sistema para separação de fala envolvendo baixas e altas frequências.

Para sinais de frequências baixas, o sistema gerou segmentos baseados na

continuidade temporal e correlação cruzada entre os canais que estão próximos. Estes segmentos foram agrupados comparando a periodicidade destas respostas com o *pitch*¹ calculado da fala desejada. Por outro lado, canais de alta-freqüência geram segmentos baseados em modulação de amplitude (AM) e continuidade temporal. Estes segmentos vão se agrupando comparando taxas de repetição AM com a freqüência fundamental (F0) calculada da fala desejada. Foi utilizado no modelo um método novo para a localização de *pitch*. O modelo de estimativa de *pitch* foi desenvolvido através da segregação inicial baseado em *pitch* global, os *pitch*'s calculados foram corrigidos através de propriedades psicoacústicas, e como resultado foi possível obter contornos de *pitch* próximos da fala sem ruído. De posse do *pitch* as operações para baixas e altas freqüências sinalizavam de forma diferente. O sistema foi testado e avaliado com vários tipos de misturas e rendeu melhores resultados quando comparado com o sistema de Wang-Brown [6] que segue características similares, pois gerou relações menores de perda de energia, obtendo relações semelhantes ao outro quanto ao ruído residual e aumentou a relação sinal ruído (SNR) para todas as intrusões.

Também foram desenvolvidos algoritmos para melhorar um simples canal usando a transformada discreta de Fourier (DFT), tal como algoritmos que usam ICA [5] ou algoritmos do tipo subtrativo. [8]

Virag [8] propôs um algoritmo baseado na família dos algoritmos subtrativos, explorando as propriedades de mascaramento do sistema auditivo humano, embora introduzissem um ruído residual com uma estrutura não natural no melhoramento da fala.

Testes foram realizados com diferentes algoritmos subtrativos para verificar a qualidade do sistema e a melhor redução de ruído foi obtida no caso do ruído gaussiano branco quando da Relação Sinal Ruído (SNR), e pelo fato dos resultados obtidos na saída dos sistemas possuírem os SNR's similares, Virag chegou à conclusão de que o segmento SNR não era um indicador suficiente de qualidade da fala. Então, utilizando-se das medidas objetivas de qualidade (IS) e medida objetiva de inteligibilidade (AI) e SNR=0 foram realizados novos testes com diferentes tipos de ruídos, onde a mesma concluiu que o método

¹ Freqüência de maior potência ou freqüência fundamental.

proposto conduzia a melhores resultados para um ouvinte humano se comparado aos outros algoritmos do tipo subtrativo. O algoritmo também foi testado por um reconhecedor baseado no modelo de Markov de reconhecimento de fala e obteve bons resultados com relação ao sinal ruído em todos os casos, exceto para ruídos que se assemelhavam a fala com SNR's maior ou igual a 3dB. Os resultados mostraram que o ruído de fundo foi reduzido e que o ruído residual foi menos estruturado do que com métodos clássicos, enquanto a distorção permaneceu aceitável.

O melhoramento da fala em um simples canal possui duas grandes desvantagens: (1) introduz ruído residual com uma estrutura não natural na fala melhorada. (2) um sistema reconhecedor de fala não pode corretamente identificar um sinal de fala quando a SNR é muito baixa.

Virag [8] achou o melhor balanço entre a redução de ruído, a distorção da fala e o nível de ruído residual baseado no fenômeno do mascaramento, mas não superou a segunda desvantagem citada.

Martin [9] propôs um novo algoritmo para localizar a variação de SNR para vários ruídos ambientais, embora esse algoritmo tenha introduzido ruído musical aborrecedor nos segmentos silenciosos da fala.

Combinando as vantagens entre os algoritmos de Virag [8] e Martin [9], Xiaoping, et al. em [10] propuseram um algoritmo baseado na estatística mínima para localizar a variação de SNR, explorando as propriedades do mascaramento do sistema auditivo humano para superar as limitações do melhoramento do canal tipo subtrativo em ruído de fundo adicionado a muito baixa SNR. Isso permitiu achar a melhor troca entre a redução de ruído, a distorção da fala e o nível de ruído residual.

Lewicki [11] usou ICA para derivar códigos eficientes de diferentes classes de sons naturais, incluindo vocalizações animais, sons ambientais e fala humana, para comprovar a teoria da codificação eficiente que tem explicado as propriedades dos campos receptivos no

córtex visual primário pela derivação eficiente dos códigos visuais das estatísticas das imagens naturais. Mostrou que vocalizações de animais assemelham-se a uma transformada de Fourier, sons ambientais não-biológicos a uma transformada Wavelet e a voz humana, por sua vez, parecia ser codificada como uma mistura dessas duas classes.

Nossa proposta é semelhante a uma rede neural que aprende as funções bases dos sinais de entrada por ICA, para recuperar a fala desejada, conduzindo este estudo ao problema de reconhecimento, pois a constante busca para aperfeiçoar e estreitar o relacionamento entre homens e máquinas, tornando-o mais natural, não é nenhuma novidade. Conseqüentemente, o reconhecimento de voz possibilitará uma manipulação mais fácil e prática de equipamentos providos com a capacidade de compreender a fala humana.

1.3 Organização do Trabalho

Neste capítulo apresentamos a proposta do trabalho, abordando alguns trabalhos correlatos.

O segundo capítulo consta de assuntos relevantes ao entendimento dos capítulos seguintes com noções básicas sobre o funcionamento do sistema nervoso central, sistema auditivo e redes neurais artificiais.

No terceiro capítulo tratamos do cenário auditivo deste trabalho destacando o ambiente denominado de *Cocktail Party*.

No quarto capítulo abordamos as técnicas ICA (Análise de Componentes Independentes) e PCA (Análise de Componentes Principais) que usamos para a separação de fontes cegas de sinais, incluindo aplicações e algoritmos de ICA.

O quinto capítulo é o foco do trabalho. Nele propomos uma metodologia para

recuperação de sinal, que consiste de duas fases: uma de aprendizagem de parâmetros de todos os sinais de entrada e outra de estimação do sinal desejado.

Para validar a metodologia proposta, fizemos simulações cujos resultados são analisados e discutidos no sexto capítulo.

No sétimo capítulo concluímos o trabalho realizado, incluindo sugestões para continuidade desta linha de pesquisa.

CAPÍTULO 2

O SISTEMA NERVOSO

Neste capítulo apresentamos ao leitor uma visão geral relativa ao conhecimento que serve de suporte para o entendimento dos capítulos seguintes.

2.1 Introdução

Iniciamos com algumas noções sobre o Sistema Nervoso Central e RNA (Redes Neurais Artificiais), enfatizando as características de como o cérebro codifica informações auditivas. Descrevemos, também, o mecanismo de sua simulação através de RNA. Não temos a intenção de que tais preliminares seja um texto para leitura independente dos objetivos expostos nesta dissertação. Aos leitores interessados em maiores detalhes nos temas envolvidos, recomendamos a leitura em outras fontes.

2.2 Organização do Sistema Nervoso

O sistema nervoso humano é uma maravilha de engenharia e comunicação que guarda dados, analisa e utiliza informações geradas pelos estímulos sensoriais (calor, luz, movimento, som). Tudo que acontece no nosso organismo é controlado pelo sistema nervoso, uma complexa rede que inclui nervos, medula espinhal e cérebro.

Cada órgão do corpo humano tem funções específicas e alguns deles trabalham de forma involuntária. Os nervos estão em toda parte do corpo. É através deles que o cérebro recebe e transmite mensagens.

Existe apenas um sistema nervoso. Ele pode ser dividido em duas partes: o sistema nervoso central (SNC), formado pelo encéfalo e pela medula espinhal; e o sistema nervoso

periférico (SNP), composto de nervos cranianos, que nascem do cérebro; de nervos espinhais, que nascem da medula, e de glânglios a eles associados. Os dois sistemas trabalham integrados.

Todo ser vivo dotado de um sistema nervoso é capaz de modificar o seu comportamento em função de experiências passadas. Essa modificação comportamental é chamada de aprendizado e ocorre no sistema nervoso através da propriedade chamada plasticidade cerebral. [12]

As células estão presentes em todas as partes do corpo, inclusive no sistema nervoso, onde são chamadas de neurônios, e são responsáveis pela transmissão dos impulsos para o cérebro e do cérebro para outros órgãos.

2.1.1 Bilhões de Neurônios

As células neuronais são consideradas unidades anatomo-fisiológicas do cérebro humano. Estima-se que existam aproximadamente 15 bilhões destas células, responsáveis por todas as funções do sistema. Tais células possuem um papel essencial na determinação e raciocínio humano. A Figura 1 representa um neurônio biológico.

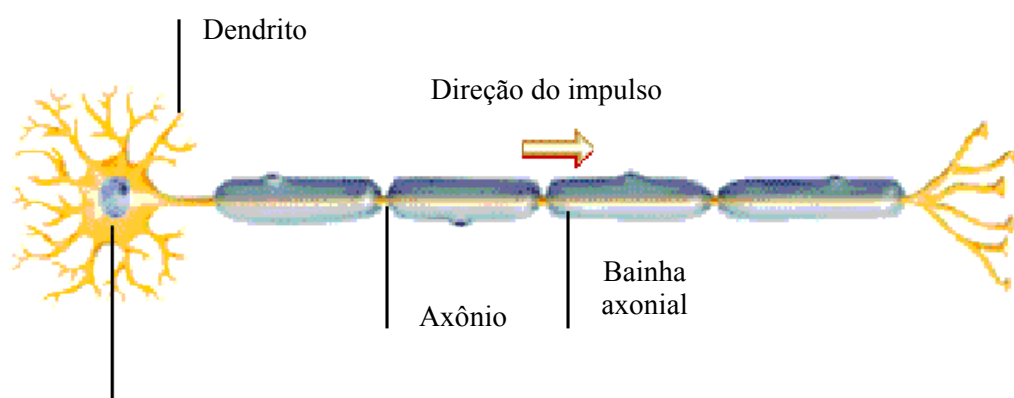


Figura 1 – Neurônio biológico
– Fonte: [13]

Existem diversos tipos de neurônios, com diferentes funções dependendo da sua localização e estrutura morfológica, mas em geral constituem-se dos mesmos componentes básicos, conforme mostrado na Figura 2:

- Corpo celular (soma) – região geralmente estrelada, constituída pela membrana, organelas celulares e núcleo;
- Dendritos – prolongamentos curtos, ramificados e geralmente numerosos que captam os impulsos nervosos e os conduzem ao corpo celular.
- Axônio – prolongamento único, cilíndrico e com ramificações em sua extremidade livre. Ele é o responsável pela transmissão dos impulsos nervosos, podendo ser revestido ou não por mielina (bainha axonal).

Entre os neurônios existe uma região especializada, responsável pela transmissão do impulso nervoso de um neurônio para o outro. Essa região denomina-se sinapse. O neurônio é uma célula altamente especializada capaz de receber impulsos elétricos e processá-los dentro de seu corpo celular (ou soma) e finalmente transmiti-lo através de seu axônio para os dendritos de outros neurônios. [14, 15, 16]

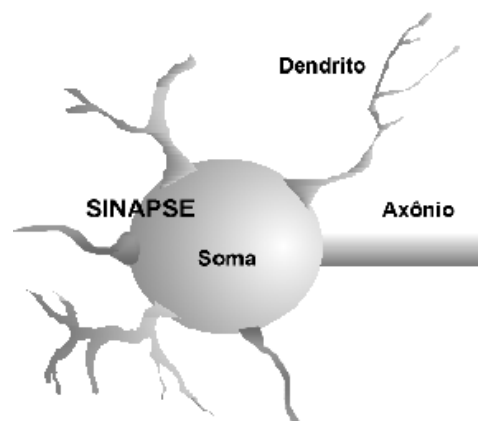


Figura 2 – Composição de um neurônio
– Fonte: [13]

Em uma sinapse os neurônios não se tocam, permanecendo um espaço entre eles denominado fenda sináptica, onde um neurônio pré-sináptico liga-se a um outro denominado neurônio pós-sináptico. O sinal nervoso (impulso), que vem através do axônio da célula pré-sináptica chega em sua extremidade e provoca na fenda a liberação de neurotransmissores depositados em bolsas chamadas de vesículas sinápticas, dando continuidade à propagação do sinal.

Um neurônio pode receber ou enviar entre 1.000 a 100.000 conexões sinápticas em relação a outros neurônios, dependendo de seu tipo e localização no sistema nervoso. O número e a qualidade das sinapses em um neurônio pode variar, entre outros fatores, pela experiência e aprendizagem, demonstrando a capacidade plástica do sistema nervoso.

2.1.2 Sistema Nervoso Central

O sistema nervoso central é constituído pelo encéfalo e pela medula espinhal, é o principal controlador das atividades do organismo. O encéfalo localiza-se dentro da caixa craniana e é formado por três órgãos: cérebro, cerebelo e tronco do encéfalo. A medula situa-se no canal medular, dentro da coluna vertebral.

Todos esses órgãos são formados por dois tipos de substâncias: a cinzenta e a branca. A cinzenta, constituída pela concentração dos corpos celulares dos neurônios, é responsável pelo recebimento dos impulsos nervosos. A branca, formada pela reunião dos axônios das células nervosas, tem como função a transmissão dos impulsos até os órgãos e vice-versa.

2.1.2.1 O Cérebro

É o maior, mais importante e complexo órgão do sistema nervoso, pois é o centro de todo o organismo. É responsável por manter o equilíbrio do corpo e coordenar os movimentos. A Figura 3 representa um cérebro humano.

O cérebro possui três tipos de funções:

- Receber os impulsos sensitivos que chegam dos órgãos dos sentidos: visão, olfato, audição, etc;
- Enviar para os músculos impulsos que determinarão movimentos voluntários: correr, andar, sentar, etc;
- Ser o centro da inteligência, o que permite atos como recordar, raciocinar e sentir emoções.

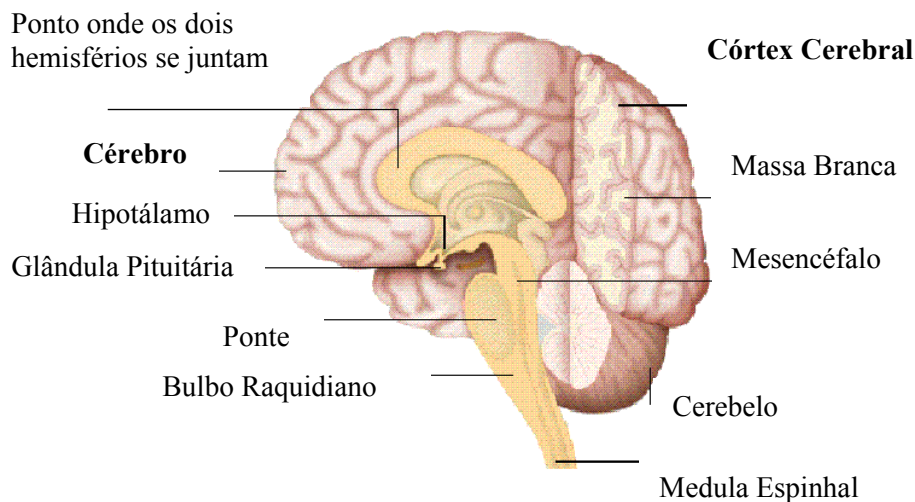


Figura 3 – Cérebro humano

– Fonte: [13]

O cérebro possui duas partes, chamadas de hemisfério direito e hemisfério esquerdo. Cada uma delas possui lóbulos (ou lobos) responsáveis por trabalhos diferentes. Os lóbulos da frente relacionam-se com nosso comportamento e alguns movimentos voluntários; os lóbulos temporais (em cada um dos lados da cabeça) cuidam da memória e da fala; os de trás analisam o que vemos; os da parte superior têm a ver com o tato e outras sensações. Os hemisférios cerebrais são muito parecidos, cada metade do cérebro controla o lado oposto do organismo. O hemisfério esquerdo é o responsável pelo controle das funções sensoriais do lado direito do corpo, enquanto que o direito controla as mesmas atividades, só que do lado esquerdo.

Quando se divide frontalmente o cérebro, verificamos uma camada externa de cerca de 8 cm de espessura, o córtex cerebral. Ela é formada por substância cinzenta (agrupamento de corpos celulares), sob a qual se encontra uma camada de substância branca (axônio dos neurônios). O córtex é enrugado e coberto de saliências chamadas circunvoluções, que triplicam essa área e permitem que caibam milhões de neurônios no cérebro. É no córtex cerebral que os impulsos nervosos são recebidos e analisados.

2.1.2.2 Áreas Associativas do Córtex

O córtex cerebral é organizado em áreas funcionais que assumem tarefas receptoras, integrativas ou motoras no comportamento, Figura 4. São responsáveis por todos os nossos atos conscientes, nossos pensamentos e pela capacidade de respondermos a qualquer estímulo ambiental de forma voluntária. Existe um verdadeiro mapa cortical com divisões precisas a nível anatomo-funcional, mas todo ele está praticamente sempre mais ou menos ativado dependendo da atividade que o cérebro desempenha, visto a interdependência e a necessidade de integração constante de suas informações frente aos mais simples comportamentos.

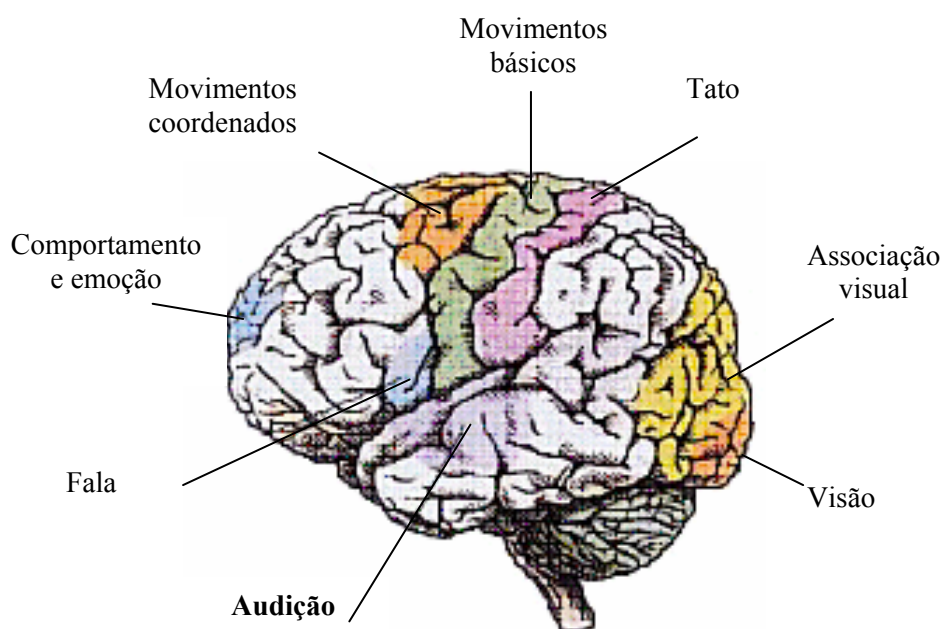


Figura 4 – Áreas cerebrais com suas funções específicas.

– Fonte: [17]

2.1.2.3 Plasticidade Neuronal

A plasticidade neural é a propriedade do sistema nervoso que permite o desenvolvimento de alterações estruturais em resposta à experiência, e como adaptação a condições mutantes e a estímulos repetidos.

Pesquisas em Neurobiologia têm comprovado que a plasticidade do sistema nervoso é uma característica única em relação a todos os outros sistemas orgânicos.

Este fato é melhor compreendido através do conhecimento morfológico-estrutural do neurônio, da natureza das suas conexões sinápticas e da organização das áreas associativas cerebrais. Sem dúvida nenhuma a aprendizagem pode levar a alterações estruturais no cérebro [12]. A cada nova experiência do indivíduo, portanto, redes de neurônios são rearranjadas, outras tantas sinapses são reforçadas e múltiplas possibilidades de respostas ao ambiente tornam-se possíveis. Portanto, o mapa cortical de um adulto está sujeito a constantes modificações com base no uso ou atividade de seus caminhos sensoriais periféricos.

2.2 Neurônio Artificial

O neurônio artificial é uma estrutura lógico-matemática que procura simular a forma, o comportamento e as funções de um neurônio biológico. Assim sendo, os dendritos foram substituídos por entradas, cujas ligações com o corpo celular artificial são realizadas através de elementos chamados de peso (simulando as sinapses). Os estímulos captados pelas entradas são processados pela função de soma, e o limiar de disparo do neurônio biológico foi substituído pela função de transferência, conforme mostrado na Figura 5. Em resumo, a operação de cada unidade de processamento pode ser descrita da seguinte forma: os sinais apresentados à entrada são associados cada qual a um determinado peso, que determina sua influência na saída da unidade, executa-se a soma ponderada destes sinais que produz um nível de atividade, se este nível de atividade exceder um certo limite à unidade produz uma determinada resposta de saída.

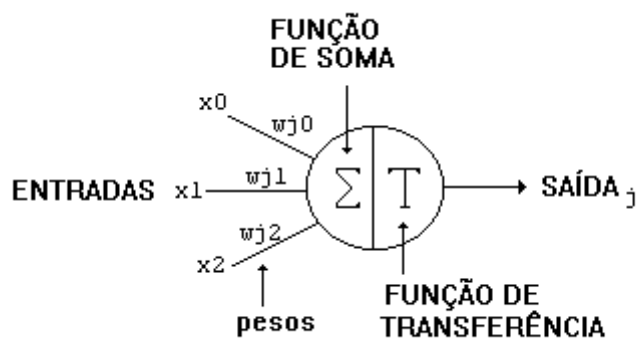


Figura 5 – Estrutura de neurônio artificial

– Fonte: [18]

Combinando diversos neurônios artificiais podemos formar o que é chamado de rede neural artificial. As entradas, simulando uma área de captação de estímulos, podem ser conectadas em muitos neurônios, resultando, assim, em uma série de saídas, onde cada neurônio representa uma saída. Essas conexões, em comparação com o sistema biológico, representam o contato dos dendritos com outros neurônios, formando assim as sinapses. A função da conexão em si é tornar o sinal de saída de um neurônio em um sinal de entrada de outro, ou ainda, orientar o sinal de saída para o mundo externo (mundo real). As diferentes possibilidades de conexões entre as camadas de neurônios podem gerar n números de estruturas diferentes. [19]

2.2.1 Redes Neurais Artificiais

As Redes Neurais Artificiais (RNA's) ou modelos conexionistas ou ainda sistemas de processamento paralelo e distribuído como também são conhecidas [14, 15] foram desenvolvidas, originalmente, na década de 40, pelo neurofisiologista Warren McCulloch e pelo matemático Walter Pitts, que dentro do espírito cibernético, fizeram uma analogia entre células nervosas vivas e o processo eletrônico num trabalho publicado sobre "neurônios formais". O trabalho consistia num modelo de resistores variáveis e amplificadores representando conexões sinápticas de um neurônio biológico.

Desde então, mais enfaticamente a partir da década 80, diversos modelos de redes neurais artificiais têm surgido com o propósito de aperfeiçoar e aplicar esta tecnologia. Algumas destas propostas tendem a aperfeiçoar mecanismos internos da rede neural para aplicação na indústria e negócios, outras procuram aproximá-las ainda mais dos modelos biológicos originais. A Figura 6 apresenta a arquitetura de uma RNA.

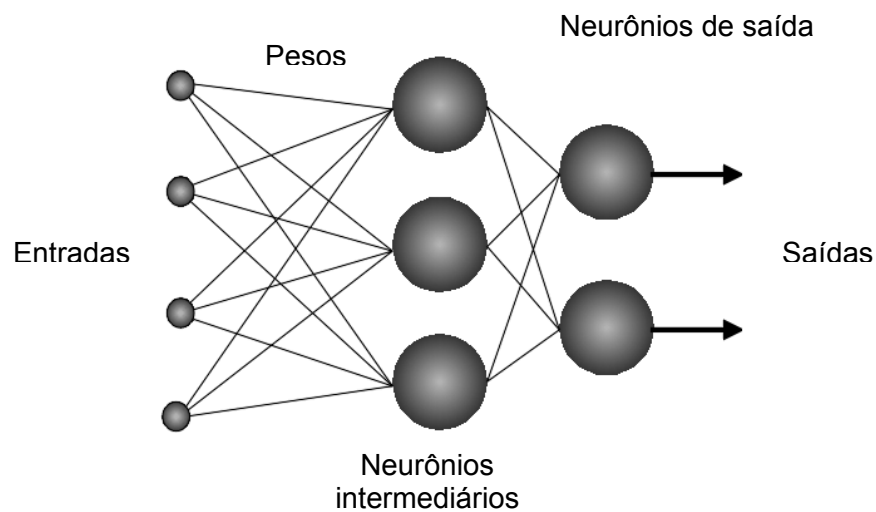


Figura 6 – Arquitetura de RNA com duas saídas.

– Fonte: [18]

Estas redes têm dado boas contribuições para um grande número de problemas, incluindo o reconhecimento da fala, devido às suas características intrínsecas. Elas podem implementar facilmente um grau enorme de computação paralela, porque as redes neurais formam uma estrutura altamente paralela de elementos computacionais simples e funcionalidades idênticas. Elas possuem uma grande tolerância a falhas. Os pesos das conexões da rede não precisam ser planejados para serem fixos, eles podem se adaptar em tempo real para oferecer um desempenho melhor. Essa é a base do conceito de aprendizado adaptativo, que é inerente à estrutura das redes neurais. Uma grande rede neural artificial pode ter centenas ou milhares de unidades de processamento; já o cérebro humano tem aproximadamente uma centena de bilhões de neurônios. [14, 16]

As variantes de uma rede neural são muitas, e combinando-as, podemos mudar a arquitetura conforme a necessidade da aplicação, ou ainda, conforme o gosto do projetista. Basicamente, os itens que compõem uma rede neural e, portanto, sujeito a modificações, são os seguintes:

- conexões entre camadas
- camadas intermediárias
- quantidade de neurônios
- função de transferência
- algoritmo de aprendizado

Todas as RNA's possuem alguma regra de treinamento, onde os pesos de suas conexões são ajustados de acordo com os padrões apresentados, de tal forma que podemos dizer que uma importante propriedade é a capacidade de aprender, por exemplo, e fazer interpolações e extrapolações do que aprendem. Isso é feito através de um processo iterativo de ajustes aplicados aos pesos das conexões ao qual denomina-se treinamento. O aprendizado ocorre quando a RNA atinge uma solução generalizada para uma classe de problemas.

Um conjunto de regras bem definidas para que a RNA possa aprender a solução de um problema é chamado de algoritmo de aprendizado. Existem muitos algoritmos de aprendizado específicos para determinados modelos de redes neurais, estes algoritmos diferem entre si principalmente pelo modo como os pesos são modificados. Outro ponto importante é a categorização das situações de aprendizado das RNA's. Neste contexto podemos citar as seguintes classes de aprendizado: [14, 15]

- *Aprendizado Supervisionado*, quando é utilizado um agente externo que indica à rede a resposta desejada para o padrão de entrada;
- *Aprendizado Não Supervisionado*, quando não existe um agente externo indicando a resposta desejada para os padrões de entrada.

É certo que os neurônios artificiais já trazem grandes semelhanças com os neurônios biológicos, e é certo também que essas semelhanças já produzem resultados significativos quando aplicados de forma adequada. Entretanto, muitos ajustes ainda são necessários para assemelhar as redes neurais artificiais ainda mais às redes neurais biológicas. Desse modo, aproximando esses dois sistemas, acreditamos que ambientes neurais artificiais possam ser úteis para elucidar mecanismos neurais complexos e ainda desconhecidos.

É certo também que processos cognitivos artificiais farão cada vez mais parte de nossas vidas, seja através de sistemas computacionais específicos ou mesmo para a pesquisa da neurociência. O uso de redes neurais na fabricação de microchips já revela uma das tendências da indústria da informática, e empresas multinacionais não estão apostando nestas novas tecnologias apenas por curiosidade. Sobre redes neurais, artigos científicos próprios das áreas biológicas também têm revelado outra tendência deste paradigma.

2.3 Sistema Sensorial

Por meio dos sentidos nos relacionamos com o ambiente que nos cerca. Os órgãos são sensíveis aos estímulos, isto é, aos vários tipos de interações com a energia do ambiente, como a luz, o som, etc. Esses órgãos captam as informações e as transformam na “linguagem” corporal dos sinais nervosos, que chegam ao cérebro pelos nervos sensoriais. No cérebro, eles são classificados, processados, analisados e interpretados. A análise desses sinais se dá no córtex cerebral, onde existe um centro especializado para essa tarefa e essa interpretação cerebral denomina-se sensação.

2.3.1 Os Sentidos

A idéia de que há cinco sentidos vem da época de Aristóteles (384-322 aC), filósofo da Grécia Antiga, que os enumerou há mais de 2300 anos. As primeiras crenças a respeito de certos órgãos do sentido parecem muito estranhas hoje em dia. Algumas concepções errôneas

surgiram em parte pela falta de compreensão sobre a natureza dos raios de luz, dos odores, sabores e das ondas sonoras; o que pode ser claramente observado no pensamento dos egípcios antigos, que acreditavam que o corpo ouvia e também respirava pelas orelhas.

Hoje dizemos que o sistema sensorial humano é capaz de reconhecer cinco sentidos fundamentais: tato, olfação, gustação, visão e audição. Cada um deles é percebido num órgão específico. Assim, por meio da interpretação desses sentidos em nosso cérebro é que podemos produzir respostas favoráveis à nossa sobrevivência.

2.3.1.1 Audição

É o sentido que nos permite a percepção dos sons. Grande parte do prazer em nossas vidas depende da nossa capacidade de ouvir [20]. Por meio desse sentido, também tomamos conhecimento do ambiente em que vivemos e somos capazes de nos comunicar.

Nossas orelhas são órgãos surpreendentes. Permitem que ouçamos do mais baixo sussurro ao mais alto ruído. Mesmo quando dormimos, as orelhas permanecem em prontidão. São capazes de mascarar os ruídos ao redor. Quando estamos acordados, um dos mais importantes trabalhos das orelhas é ajudar na comunicação. Elas trabalham com o cérebro e o sistema nervoso: captam um som e o transformam numa mensagem que podemos receber e reconhecer.

As orelhas ajudam no aprendizado da fala, participam do aprendizado na leitura, ajudando a relacionar as letras à sons. Outra atribuição é auxiliar no equilíbrio e graças a elas caminhamos sem cair.

2.3.1.2 Órgãos da Audição

Os nossos órgãos da audição, as orelhas (antigamente chamadas de ouvidos), são divididas em três regiões: externa, média e interna conforme pode ser observado na Figura 7.

Além de possuírem receptores para a percepção de sons, estes órgãos estão relacionados ao equilíbrio (orelha interna), pois fornecem ao cérebro informações sobre a posição do nosso corpo.

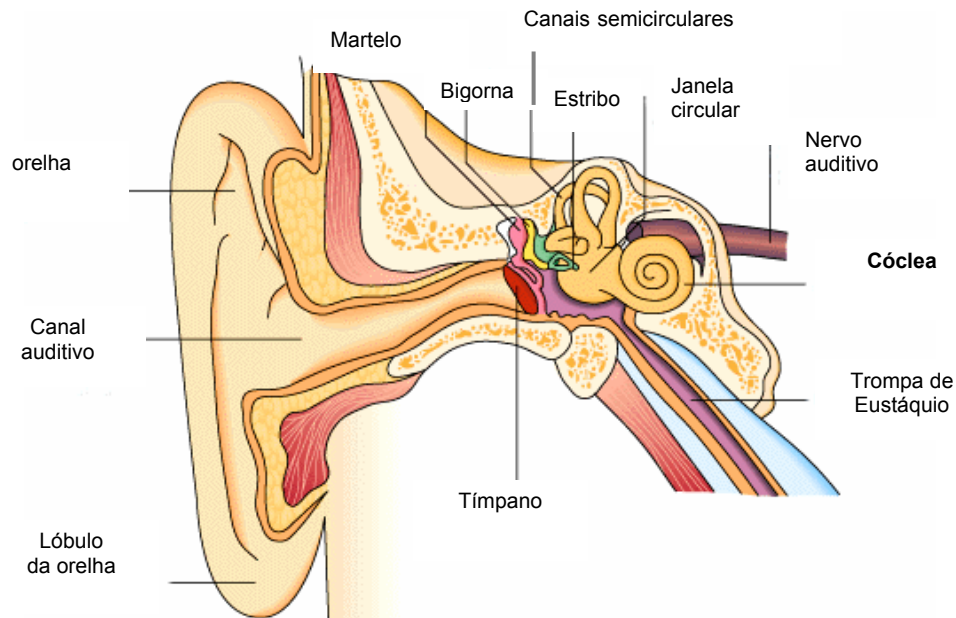


Figura 7 – Órgãos da audição

– Fonte: [21]

Para compreendermos melhor como as orelhas tornam possível ouvir os sons, é preciso entender o que são e como são produzidos esses sons.

O som é o resultado da vibração dos corpos e só se processa nos meios materiais.

Nossa orelha só é capaz de distinguir sons cujos números de vibrações estão compreendidos entre 20 e 20.000 por segundo. As vibrações dos limites inferiores produzem sons graves e a dos limites superiores, sons agudos.

A diferença de tempo entre a chegada de uma onda a uma das orelhas e depois a

outra é de milésimos de segundos, mas as orelhas e o cérebro detectam essa diferença e assim podemos descobrir a direção do som.

a) Orelha Externa

É constituída pelo pavilhão auricular (ou simplesmente orelha), formação cartilaginosa, em forma de concha, coberta pela pele e cheia de saliências e reentrâncias, exceto na parte inferior, chamada lóbulo cuja função está ligada à captação e direcionamento das ondas sonoras; e pelo meato acústico externo (conduto auditivo), canal com aproximadamente 2,5 cm de profundidade que se estende da aurícula até a membrana do tímpano.

A orelha externa capta o som, processo no qual o canal auditivo atuando como ressonador reproduz o dobro ou o triplo as ondas sonoras.

b) Orelha Média

Transmite o som. É constituída pela caixa do tímpano, cavidade contendo a cadeia de ossículos (martelo, bigorna e estribo) que transmitem as vibrações do tímpano à janela oval. A caixa do tímpano comunica-se, atrás, com as cavidades mastóideas, que funcionam como caixa de ressonância, e na frente, com a faringe, pela Trompa de Eustáquio, cuja abertura, durante a deglutição, restabelece uma pressão interna igual à pressão atmosférica.

Os ossículos, no interior da orelha média atuam como um sistema de alavancas, podendo duplicar ou triplicar as vibrações do tímpano. Os músculos da orelha média modificam o desempenho desse sistema de alavancas, atuando como uma unidade amplificadora, além de atuarem como um controlador de volume, pois protegem a orelha contra sons excessivamente altos diminuindo as vibrações que chegam do tímpano. Embora esses pequenos músculos, os menores no corpo humano, no fim do canal ossicular, possam ser movidos voluntariamente, suas contrações são, normalmente, reflexos desencadeados quando o som excede determinado nível.

b.1) A Janela Oval

É a comunicação entre a orelha média e a orelha interna. O estribo passa as vibrações para a janela oval, que é uma membrana que cobre uma abertura na "caixa de ossos" (*bony case*) da cóclea.

O tamanho da janela oval (15 a 30 vezes menor do que o tímpano), faz com que se produza à amplificação crítica necessária para se igualar a impedância (propriedade que os corpos têm de se oporem ao movimento) entre as ondas de som no ar e o fluido coclear. Além da amplificação do sistema de alavancas da orelha média, esta concentração de forças amplifica as vibrações do som que chegam cerca de 15 a 30 vezes. O resultado desses três mecanismos (o tubo de ressonância do canal auditivo, que pode aumentar a pressão do ar em até 10 vezes; o sistema de alavancas, que pode triplicá-lo; e a localização estratégica do tímpano e da janela oval) podem amplificar a onda sonora em até mais de 800 vezes antes que o líquido da orelha interna se ponha em movimento.

c) Orelha Interna

Está localizada mais profundamente no crânio, numa escavação do osso temporal. A orelha interna, também conhecida como labirinto, divide-se em três partes: canais semicirculares, vestíbulo e cóclea (caracol). Os canais semicirculares e o vestíbulo relacionam-se ao equilíbrio e a cóclea, à audição.

c.1) A Cóclea

É um tubo espiralado semelhante à concha de um caracol, cheio de um fluido denominado líquido coclear. [20]

A força mecânica amplificada transmitida da orelha média para a orelha interna pelos ossículos, é imediatamente transformada, na cóclea, em pressão hidráulica. Essa pressão

hidráulica move-se para o canal coclear e para os receptores da audição que formam o órgão espiral (órgão de Corti). A cóclea é responsável pela transformação das ondas sonoras em mensagens compreensíveis ao cérebro. Quando as ondas transmitidas pelo estribo vibram de encontro à janela oval, o fluido da cóclea também vibra. Esses movimentos vibratórios são, então, transformados em impulsos nervosos, os quais viajam ao cérebro pelo nervo vestibulo coclear (auditivo). Assim ocorre a percepção do som.

c.2) Sistema de Transmissão da Cóclea para o Cérebro

Este sistema contém 30.000 fibras nervosas que saem do Órgão de Corti para formarem o nervo auditivo. Estas fibras são agrupadas de acordo com a frequência do sinal sonoro que elas carregam, o número de fibras que um som requer, dará ao cérebro a indicação da sua intensidade.

Vindo do cérebro, fibras nervosas descendentes podem transportar instruções da parte de trás do cérebro até a orelha, para filtrar, e com isso eliminar alguns sinais que o cérebro determina como sendo sem importância, e se concentrar em outros. Depois de passar pelo núcleo coclear, alguns dos nervos descendentes do cérebro vão para a orelha média, onde elas controlam os músculos usados na defesa contra os sons muito altos e prejudiciais ao sentido da audição.

As fibras nervosas levam sinais sonoros a diferentes partes do córtex auditivo dependendo da frequência que eles conduzem. O córtex auditivo situa-se em um profundo sulco chamado Fossa de Sylvian. Os sons agudos terminam na parte mais interna da Fossa de Sylvian, e os sons mais graves, mais perto da superfície.

2.4 Codificação Eficiente de Sons Naturais

Já sabemos que o cérebro humano codifica informações sensoriais, mas o que tem

sido alvo de estudo é o desenvolvimento que ocorre no mesmo quanto ao uso de diferentes estratégias de codificação.

Em meio a diferentes codificações, temos a codificação dos sons, que para os humanos é uma tarefa menos complicada do que para os reconhedores de fala artificiais, mesmo sendo de nosso conhecimento que a audição humana não é perfeita, pois além das limitações físicas do ouvido, o som tem que viajar através dos nervos até o córtex auditivo do cérebro onde ele é transformado em diferentes percepções das quais tomamos consciência.

Duas características de como o cérebro humano possivelmente processa informações são o mascaramento e codificação eficiente. Enquanto o primeiro acontece na cóclea, supõe-se que o posterior acontece no córtex auditivo. Nossa proposta é baseada nesta última.

Entende-se por mascaramento auditivo como a "audibilidade diminuída de um som devido à presença de outro". O mascaramento auditivo consiste de mascaramento em frequência e mascaramento temporal. Se um som é mascarado depois de um som mais forte é chamado pós-mascaramento, e se é mascarado antes em tempo é chamado pré-mascaramento. O pré-mascaramento existe só por um curto momento (20 ms). O pós-mascaramento tem efeito até 200 ms.

No estudo intitulado como Codificação Eficiente de Sons Naturais, Lewicki [11] abordou o problema de quais estratégias particulares o cérebro utiliza para distinguir os sons e codificá-los de formas diferentes.

O som que chega às nossas orelhas é formado por diferentes frequências e o sistema auditivo codifica esses sons através da decomposição do sinal de amplitude que chega. A idéia que se tem é que várias formas de filtros codificam os sons (banco de filtros lineares) e que o código dependerá da classe do som em questão.

Lewicki avaliou três classes de sons como representantes de um ambiente auditivo natural: sons ambientais, vocalização de animais e a fala humana. Como exemplo de sons

ambientais, há o esmigalhamento de folhas e estalamento de gravetos, pois requer uma rápida e apurada localização auditiva. Esses sons são não-harmônicos, de banda larga e de pequena duração. Vocalizações de animais são tipicamente harmônicas, de relativa banda larga e longa duração. E a voz humana possui propriedades tanto de um quanto de outro som, harmônico e não-harmônicos.

A codificação eficiente tem explicado as propriedades dos campos receptivos no córtex visual primário pela derivação eficiente dos códigos visuais das estatísticas das imagens naturais. [22]

Para explicar esta teoria no sistema auditivo, Lewicki usou ICA para derivar códigos eficientes para as classes de sons descritas anteriormente. Ele mostrou que sons naturais são codificados dependendo da classe, “se assemelhando a uma transformada de Fourier e Wavelet” que podem ser representadas conforme a Figura 8.

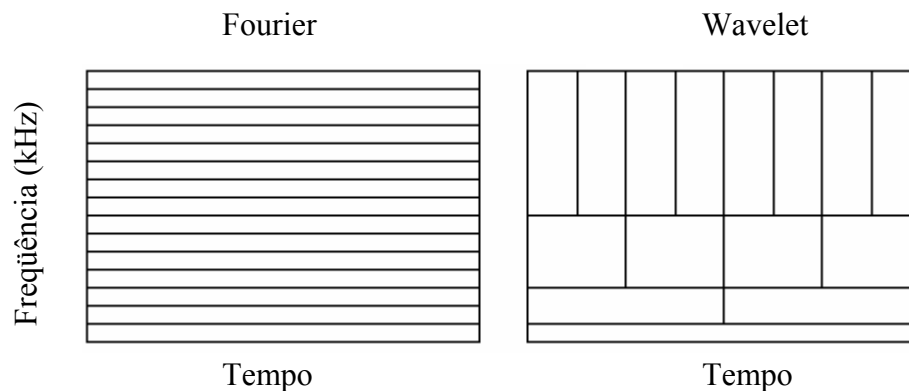


Figura 8 – Análise de tempo-freqüência.

– Fonte: [11]

A transformada de Fourier é usada para transportar uma função $f(t)$ no domínio do tempo para outra função $F(w)$ no domínio da freqüência. Seja $F(w)$ a Transformada de Fourier de um sinal $f(t)$. A formulação da transformada é:

$$F(w) = \sum_{k=-\infty}^{\infty} f(t)e^{-jwk} \quad (1)$$

A equação (1) acima usa como função de análise a exponencial complexa. Esta função se estende no domínio do tempo de $[-\infty \dots \infty]$.

Na transformada Wavelet à medida que as frequências vão aumentando, a largura da função da Wavelet-mãe diminui no tempo, enquanto que na série de Fourier a largura não muda no tempo.

As formas dos filtros ICA para vocalizações de animais são essencialmente representações Fourier, com muitos filtros tendo oscilações senoidais e pequena modulação em amplitude. A transformada de Fourier representa um sinal por uma superposição linear de senoides. Deste modo, os filtros estão localizados na frequência, e não no tempo conforme mostrado na Figura 9.

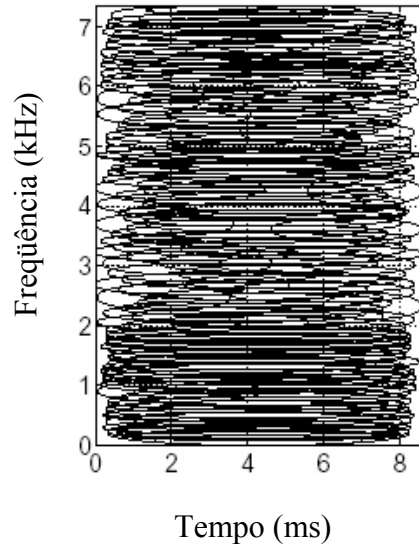


Figura 9 – Análise de tempo e frequência para vocalização de animais
– Fonte: [11]

Os sons ambientais não biológicos assemelham-se a uma transformada Wavelet, que é composta de filtros que estão localizados tanto no tempo quanto na frequência. Uma representação Wavelet faz um balanço entre a resolução na frequência e no tempo conforme a Figura 10.

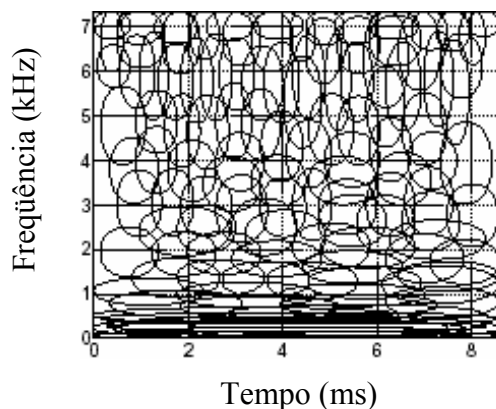


Figura 10 – Análise de tempo e frequência para sons ambientais.

– Fonte: [11]

A representação eficiente da voz humana é quase idêntica a dos sons naturais combinados com vocalizações de animais como pode ser observado na Figura 11.

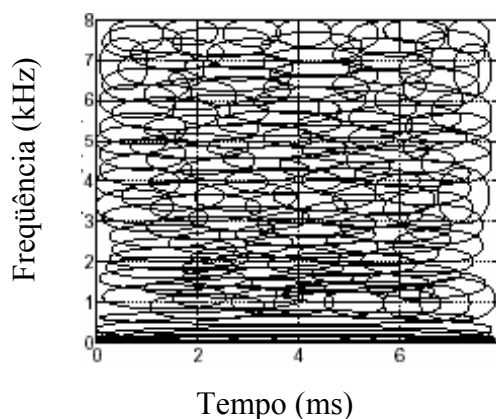


Figura 11 – Análise de tempo e frequência para voz humana.

– Fonte: [11]

A estrutura estatística dos sinais determina como as formas dos filtros ICA derivados da codificação eficiente de um conjunto de dados diferentes são distribuídos no espaço tempo-freqüência.

Cada elipse representada nas Figuras 9, 10 e 11 é um esquema da extensão de um filtro ICA no espaço tempo-freqüência.

Tendo estabelecido as bases de funcionamento do cérebro, principalmente com relação ao sistema auditivo, torna-se necessário abordar o cenário auditivo em que se desenvolveu o trabalho - o ambiente *Cocktail Party* -, com destaque para um de seus mais freqüentes efeitos, a reverberação.

CAPÍTULO 3

ANÁLISE DE CENÁRIO AUDITIVO

Em ambiente real, a fala normalmente acontece em meio a interferências acústicas, mas o homem tem uma notável habilidade para separar a voz de um simples orador de outros sons. As máquinas propostas com esta finalidade não apresentam este mesmo desempenho, dada sua baixa robustez a ruídos. Pois, em um ambiente como o *Cocktail Party* todos os sons, exceto o desejado, podem ser tratados como ruído. A expressão análise de cenário auditivo (ASA) tem sido introduzida para descrever esse processo. [23]

3.1 Introdução

Entre os problemas em ASA, talvez o mais conhecido seja o *Cocktail Party*. O *Cocktail Party* pode ser entendido da seguinte maneira: é um ambiente onde os componentes de frequência das diversas fontes de sinais existentes (vozes, música, ar, ruído, entre outras) chegam à orelha de uma pessoa, completamente misturados e com reverberação. A análise desta cena envolve questões perceptivas como direção e localização das fontes (sinais originais).

O problema no *Cocktail Party* é separar um ou mais desses sinais originais ou aumentar sua inteligibilidade. Tal situação pode ser analisada através de duas perspectivas [2]: pela contribuição de múltiplos canais de entrada (vários microfones) ou de um simples canal de entrada (um simples microfone – caso monaural). Isto tem sido alvo de muitos pesquisadores que há algumas décadas vêm tentando melhorar os resultados conseguidos através de simulação computacional. Este estudo iniciou-se em 1953 com Cherry, tendo ele descoberto que capacidade de separação de uma das vozes existentes em uma sala está ligada ao gênero da pessoa, a intensidade da voz e a localização dela. [24]

Analisando a situação do ponto de vista de um ouvinte, esta tarefa é bastante simples e intuitiva. De uma perspectiva psicológica ou fisiológica, há uma série de evidências que foram compostas para explicar o efeito; tais como interações entre o sinal, o sistema auditivo e o sistema nervoso central. Do ponto de vista acústico, o problema é semelhante ao de se fazer uma separação da voz de uma pessoa em um espectrograma² contendo sinais de voz de várias pessoas em uma situação de conversa e outros ruídos. Neste caso, com o espectrograma - produto de uma máquina -, essa tarefa é difícil até mesmo para um especialista.

Na nossa proposta utilizamos um simples canal, o que torna essa tarefa mais difícil, tendo em vista que neste caso, temos tanto o(s) sinal (is) desejado (s) quanto o ruído juntos, conforme a seguinte representação.

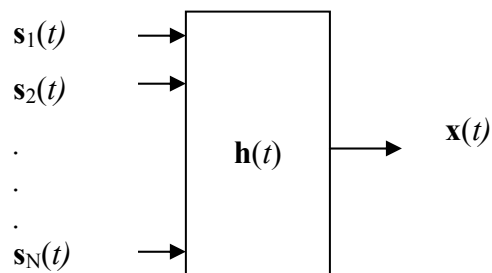


Figura 12 – Representação monaural em bloco.

3.1 Reverberação

Reverberação é a sensação desagradável de retorno da própria voz - eco - dentro de um banheiro, numa sala totalmente vazia, ou em qualquer outro ambiente, como por exemplo em um meio de telecomunicação. ocorre por causa da reverberação própria do ambiente. É um fenômeno em que os sons permanecem por algum tempo num ambiente depois de emitidos, devido à reflexão das ondas sonoras nas paredes, nos objetos ou nos materiais.

² Gráfico de domínio da frequência.

Ocorrem tantas reflexões quantas necessárias para que toda a energia do som se dissipe no ambiente, daí a sua permanência. [25]

Uma maneira de entender o conceito de reverberação é através do eco: em um lugar amplo, as reflexões de um som ocorrem muito espaçadamente e a orelha distingue cada reflexão particular, pois bem, o eco é apenas uma reverberação lenta. O eco acontece quando uma onda sonora bate em uma barreira e volta para o local de origem. São essas ondas que propagam o som, fazendo vibrar as moléculas que compõe o ar refletindo.

A Figura 13 descreve o som recebido por um único ouvinte em função do tempo em consequência de um impulso a uma determinada distância. O som direto recebido é seguido por sons refletidos distintos e então por um conjunto de muitos sons refletidos que se misturam e se sobrepõem no que é chamado reverberação. O atraso entre o som direto e o primeiro som refletido é uma característica significativa para um ambiente, tão importante quanto o tempo total da reverberação.

Se o impulso de um som constante cessa, o som reverberante em um ambiente deteriora em uma forma aproximadamente exponencial.

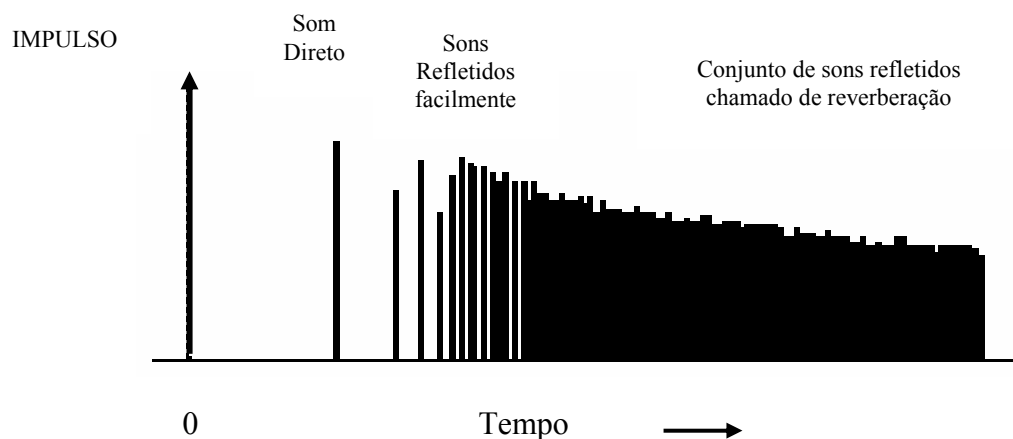


Figura 13 – Representação do decaimento exponencial.

– Fonte: [25]

3.1.1 Filtro de Resposta Finita ao Impulso - FIR

Nestes filtros as saídas do sistema dependem somente da entrada presente e de um número finito de entradas passadas, então o filtro tem uma resposta impulsiva finita. Os valores passados não têm influência no cálculo dos valores de saída presentes.

Em se tratando de filtros e resposta ao impulso, isso conduz naturalmente para a noção de função de transferência.

Seja $s(n)$ um sinal digitalizado, $h(n)$ a resposta ao impulso de uma sala para uma posição particular de um ouvinte, e $y(n)$ a potência do som em uma posição particular, podemos escrever o produto da convolução.

$$y(n) = \sum_{k=0}^{\infty} s(k) * h(n - k) \quad (2)$$

Tomando a transformada Z de cada membro da equação (2), temos a expressão:

$$Y(z) = S(z) * H(z) \quad (3)$$

Onde $Y(z)$, $S(z)$ e $H(z)$ são definidas por:

$$Y(z) = \sum_{k=0}^{\infty} y(k)z^{-k} \quad (4)$$

$$S(z) = \sum_{k=0}^{\infty} s(k)z^{-k} \quad (5)$$

$$H(z) = \sum_{k=0}^{\infty} h(k)z^{-k} \quad (6)$$

A função $H(z)$ é a representação matemática da resposta ao impulso do filtro. Ou seja, a função de transferência do sistema “sala”.

Esse sistema pode ser representado como mostra a Figura 14:

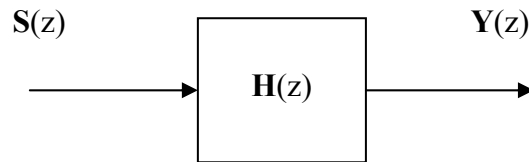


Figura 14 – Diagrama em blocos da função de transferência.

O problema do *Cocktail Party* consiste em separar fontes. Esse processo, mais apropriadamente chamado de separação cega de fontes é tratado em detalhes no capítulo seguinte.

CAPÍTULO 4

SEPARAÇÃO CEGA DE FONTES (BSS)

A Separação Cega de Fontes (BSS - *Blind Source Separation*) consiste em recuperar dentre uma mistura de sons, sinais não observados, denominados fontes. A solução inicial para este problema foi proposta por Herault and Jutten [26]. Hoje esse tópico é bastante conhecido e novos algoritmos continuam desenvolvidos com o mesmo propósito.

4.1 Introdução

A separação cega de fontes (BSS) é uma aplicação clássica do modelo de Análise de componentes independentes [26]. Vale ressaltar que desde que as pesquisas utilizando ICA para essa aplicação iniciaram-se, muitos autores não usam o termo ICA, e sim, a notação BSS.

ICA é uma metodologia usada para BSS, ou ainda para a modelagem de dados em diferentes aplicações. BSS é um problema para o qual, além da própria ICA, aplicam-se outras metodologias que fazem uso, por exemplo, da informação da frequência ou das propriedades dos espectros. [27, 28]

4.2 Formulação do problema da BSS

O propósito da separação cega de fonte é recuperar n fontes de sinais de um processo de m misturas. Levando em consideração que o número de misturas é bem maior que as fontes originais, isto é, $m \geq n$ [29] e que a mistura é linear e instantânea. Então a BSS procura achar uma matriz de mistura usando somente os sinais observados.

4.1.1 Hipóteses Básicas

Embora o processo de separação seja dito como “cego”, nós fazemos algumas hipóteses sobre as fontes de sinais [30]:

- Hipótese 1: As fontes são estatisticamente independentes.
- Hipótese 2: As misturas são convolutivas;
- Hipótese 3: Somente uma das fontes tem distribuição gaussiana;
- Hipótese 4: A matriz de mistura tem *rank* cheio.

A seguir é ilustrado na Figura 15 (a) o processo de mistura e separação de um conjunto de n fontes de sinais $s_1(t), \dots, s_n(t)$, misturadas linearmente por uma matriz \mathbf{A} , gerando a matriz $x(t) = x_1(t), \dots, x_n(t)$.

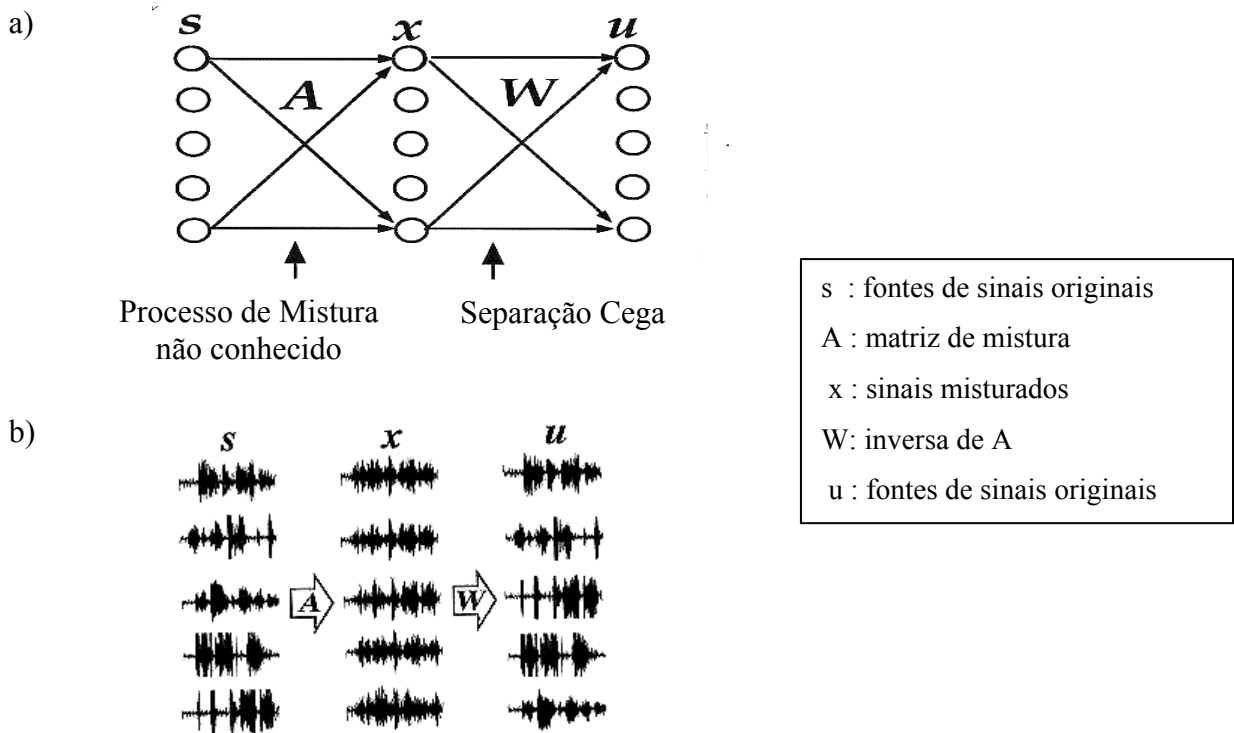


Figura 15 – Representação de Sinais no Processo de Mistura e Separação

– Fonte: [31]

Partindo do princípio de que não conhecemos nada sobre as fontes originais e o processo de mistura, a tarefa é obter os sinais originais a partir do sinal misturado $\mathbf{x}(t)$.

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \quad (7)$$

Para a separação cega das fontes e baseado na Hipótese 1 citada anteriormente, o problema seria solucionado com uma estimativa da inversa da matriz de mistura.

Sendo:

$$\mathbf{u}(t) = \mathbf{W}\mathbf{x}(t) \quad (8)$$

e substituindo a equação (7) na equação (8) temos:

$$\mathbf{u}(t) = \mathbf{W}\mathbf{A}\mathbf{s}(t) \quad (9)$$

se $\mathbf{W} = \mathbf{A}^{-1}$, teremos uma matriz identidade na equação (9), que resultará na solução do problema de encontrar as fontes originais, que pode ser reescrito como mostra as equações (10) e (11).

$$\mathbf{s}(t) = \mathbf{A}^{-1}\mathbf{x}(t) \quad (10)$$

$$\mathbf{s}(t) = \mathbf{W}\mathbf{x}(t) \quad (11)$$

4.2 PCA / ICA

Um problema enfrentado na análise de dados, estatística, processamento de sinal e nas pesquisas de redes neurais é o de se encontrar uma representação apropriada dos dados, por meio de uma transformação linear dos dados originais. É importante para a análise

subseqüente, seja ela qual for, que os dados sejam representados de uma maneira que facilite o processo de entendimento.

Os métodos de transformações lineares conhecidos e comumente usados para representação dos dados para este exemplo são a Análise de Componentes Principais (PCA), a Análise de Fator e Busca da Projeção. Em seguida, vamos descrever a PCA e um método de transformação linear recentemente desenvolvido: a Análise de Componentes Independentes (ICA), em que a representação desejada é de minimizar a dependência estatística dos componentes. Tal representação parece capturar a estrutura essencial dos dados em muitas aplicações.

Usar transformações lineares faz o problema da representação de variáveis aleatórias multidimensionais, computacionalmente e teoricamente mais simples, facilitando a interpretação dos resultados.

4. 2.1 Análise de Componentes Principais - PCA

Análise de Componentes Principais é um método que se baseia em estatística de segunda ordem. Como exemplo clássico de método de segunda ordem além do PCA temos também a Análise de Fator. [32, 33]

A PCA é uma transformação linear ortogonal de um espaço p -dimensional para um espaço m -dimensional, $m \leq p$, onde um número (possivelmente) de variáveis correlacionadas serão transformadas em um número menor de variáveis não correlacionadas chamadas *componentes principais*, ou seja, as coordenadas dos dados no novo espaço são não correlacionadas e a maior quantidade de variância dos dados originais é preservada usando-se somente algumas poucas coordenadas [34]. E o primeiro componente é o que contribui com maior energia.

Em outras palavras, PCA é uma técnica para encontrar a direção em que o fluxo de

dados está mais alongado, Figura 16. Essas direções representam a maioria das informações, sendo assim é importante conhecê-las, pois através delas será permitido o armazenamento dos dados de forma comprimida para posteriormente reconstruí-los com a mínima quantidade de distorção.

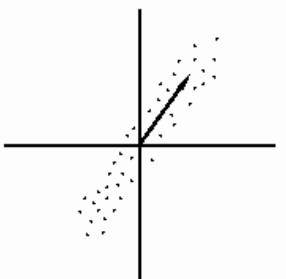


Figura 16 – Direção da extensão do fluxo de dados.

– Fonte: [35]

PCA é usado em estatística para extrair as principais relações entre dados de dimensionalidades elevadas. Uma maneira comum de encontrar as componentes principais de um conjunto de dados é calculando os autovetores da matriz de correlação. Esses vetores informam as direções nas quais há um alongamento maior dos dados. A projeção dos dados sobre os autovetores são as componentes principais. Os autovalores nos dão uma indicação da quantidade de informação que a respectiva componente principal representa.

Dessa forma o objetivo da PCA pode ser resumido no seguinte:

- dadas p variáveis x_1, x_2, \dots, x_p , encontrar combinações lineares dessas para produzir índices z_1, z_2, \dots, z_p que sejam não correlacionados. Onde z_i componentes principais e os índices z são ordenados de forma que $\text{var}(z_1) \geq \text{var}(z_2) \geq \dots \geq \text{var}(z_p)$ e $\text{var}(z_i)$ representa a variância de z_i .

A redução de dimensionalidade se resume no seguinte: se a maioria dos índices

apresentarem variâncias tão pequenas a ponto de serem ignoradas, a variação no conjunto de dados pode ser apropriadamente descrita pelos poucos índices z que retêm as maiores variâncias.

O procedimento para uma Análise de Componentes Principais resume-se em encontrar os autovalores e os correspondentes autovetores da matriz de covariância do conjunto de dados considerado, assumindo que os autovalores estão ordenados por valores decrescentes, $\lambda_1 \geq \lambda_2 \geq \dots \lambda_p \geq 0$, então λ_i corresponde à i -ésima componente principal. Ou seja:

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p \quad (12)$$

e em particular, $\text{var}(Z_i) = \lambda_i$ e $a_{i1}, a_{i2}, \dots, a_{ip}$ são os elementos do autovetor correspondente.

4.2.1.1 Entendendo PCA

PCA é um método estatístico clássico. Essa transformação linear tem sido usada em análise e compressão de dados. Baseia-se na representação estatística de variáveis aleatórias. Suponha que nós tenhamos um vetor aleatório \mathbf{x} , onde:

$$\mathbf{x} = (x_1, \dots, x_m)^T \quad (13)$$

sendo a média do vetor \mathbf{x} denotada por:

$$\boldsymbol{\mu}_x = E[\mathbf{x}] \quad (14)$$

a matriz de covariância do mesmo conjunto de dados é:

$$\mathbf{C}_x = E[(\mathbf{x} - \boldsymbol{\mu}_x)(\mathbf{x} - \boldsymbol{\mu}_x)^T] \quad (15)$$

A componente de \mathbf{C}_x , denotada por c_{ij} , representa a covariância entre os componentes de variáveis aleatórias x_i e x_j . O componente c_{ii} é a variância do componente x_i . A variância de um componente indica a distribuição dos valores dos componentes ao redor do valor significativo. Se duas componentes x_i e x_j dos dados são não correlacionados, sua variância é zero ($c_{ij} = c_{ji} = 0$). A matriz de covariância é por definição sempre simétrica.

De uma amostra de vetores x_1, \dots, x_m , nós podemos calcular uma amostra significativa da matriz de covariância como uma estimação da média e da matriz de covariância.

De uma matriz simétrica tal como a matriz de covariância, nós calculamos uma base ortogonal para achar os autovalores e autovetores. O autovetor e_i e o correspondente autovalor λ_i são as soluções da equação.

$$\mathbf{C}_x e_i = \lambda_i e_i \quad (16)$$

para $i = 1, \dots, n$;

Para simplificar nós supomos que os λ_i são distintos. Esses valores podem ser estabelecidos, por exemplo, achando as soluções da equação característica.

$$|\mathbf{C}_x - \lambda \mathbf{I}| = 0 \quad (17)$$

onde o \mathbf{I} é a matriz identidade tendo a mesma ordem que \mathbf{C}_x e a $|\cdot|$ denota o determinante da matriz. Se um dado vetor tem n componentes, a equação característica torna-se de ordem n . Isso é relativamente fácil de resolver, se n for pequeno. A solução dos autovalores e correspondentes autovetores é uma tarefa não trivial, e existem muitos métodos.

Ordenando os autovetores em uma ordem de autovalores descendentes (*largest first*),

podemos criar uma base ortonormal com o primeiro autovetor tendo a direção da maior variância de dados. Por este caminho, nós podemos encontrar direções em cada conjunto de dados com a quantia mais significativa de energia.

Os componentes são extraídos na ordem do mais explicativo “*de maior energia*” para o menos explicativo. Teoricamente o número de componentes é sempre igual ao número de variáveis. Entretanto, alguns poucos componentes são responsáveis por grande parte da explicação total. A desvantagem desta estratégia é que gira o eixo na direção dos sinais independentes só no caso onde esses sinais têm uma distribuição Gaussiana. ICA apareceu como uma solução para este problema.

A análise de componentes principais seja talvez a mais antiga técnica conhecida na análise de dados multivariados. Foi primeiramente introduzida por Pearson (1901), que usou PCA num contexto biológico para análise de regressão linear. Foi utilizado por Hotelling (1933) num trabalho feito em psicometria. E apareceu novamente e quase independentemente no contexto de teoria probabilística, como considerado por Karhunen (1947); e foi generalizado em seguida por Loève (1963). Desta forma, em algumas áreas de aplicação, pode ser chamada também de discreta transformação Karhunen-Loève, ou transformação Hotelling. [35, 36]

Algumas aplicações de controle de processo podem ser executadas utilizando-se redes neurais simples, cujo mapeamento entrada/saída tornam-se lineares depois da aprendizagem. Vários algoritmos de aprendizagem são realizações neurais amplamente utilizadas em técnicas estatísticas de análise de componentes principais.

4.2.2 Análise de Componentes Independentes – ICA

Análise de Componentes Independentes é um método de análise proposto por Jutten e Héroult a partir de 1983, é uma extensão da PCA que vem sendo estudada no contexto de separação cega de fonte (BSS) de suas misturas lineares, usando a independência estatística

como critério de separação de fontes sem o conhecimento dos coeficientes de mistura, nem a provável distribuição das fontes do sinal. [37, 38, 39]

Este método tem recebido bastante atenção devido ao seu potencial em aplicações de processamento de sinais como telecomunicações, processamento de sinais médicos e sistemas de reconhecimento de voz.

Diferentemente da PCA é um método que utiliza estatística de alta ordem, e neste contexto temos como exemplo também a busca da projeção, redução de redundância e a deconvolução cega.

4.2.2.1 Definições da Análise de Componentes Independentes

Na literatura há pelo menos três definições básicas para ICA linear. As diferenças entre as definições não são enfatizadas completamente. Provavelmente por ICA ser um tópico de pesquisa novo. E sendo assim a maioria das pesquisas se concentrou na mais simples dessas definições. [40]

Nas seguintes definições um vetor aleatório de variáveis observáveis é assim determinado:

$$\mathbf{x} = (x_1, \dots, x_m)^T \quad (13)$$

- Definição 1: (definição geral) – ICA do vetor aleatório \mathbf{x} consiste em usar uma transformação linear $\mathbf{s}=\mathbf{W}\mathbf{x}$ de modo que os componentes s_i sejam tão independentes um do outro quanto possível no sentido de maximizar/minimizar alguma função $F(s_1, \dots, s_m)$. Esta definição é a mais geral no sentido de que não é feita nenhuma suposição com relação aos dados. Naturalmente, esta definição é também completamente vaga porque deve-se definir uma medida de independência para s_i .

- Definição 2: (modelo ruidoso) – ICA do vetor aleatório \mathbf{x} , consiste em estimar o seguinte modelo para os dados:

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n} \quad (18)$$

onde as variáveis latentes (componentes) s_i no vetor $\mathbf{s} = (s_1, \dots, s_n)^T$ são supostamente independentes. A matriz \mathbf{A} é uma matriz de mistura $m \times n$, e \mathbf{n} é o vetor aleatório dimensional do ruído. Esta definição reduz o problema de ICA à estimação de um modelo normal de variáveis latentes. Entretanto esse problema de estimação não é muito simples, e conseqüentemente a maioria das pesquisas de ICA concentrou-se na seguinte definição simplificada.

- Definição 3: (modelo silencioso) – ICA do vetor aleatório \mathbf{x} , consiste em estimar o seguinte modelo para os dados:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (19)$$

onde \mathbf{A} e \mathbf{s} são como na definição 2. E o vetor do ruído foi omitido. Este foi o modelo introduzido por Jutten e Héroult em seu artigo [26]. E mais cedo, por Jutten em sua tese de Doutorado.

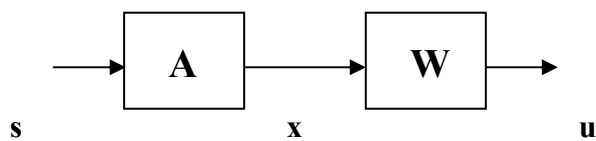


Figura 17 – Diagrama em blocos para ICA.

A identificação do modelo de ICA, conforme representação do diagrama da Figura 17 foi tratada em [41]. Impondo as seguintes limitações fundamentais (além da suposição básica da independência estatística), a identificação do modelo pode assegurar que:

- Os elementos de s_i são mutualmente independentes (estatisticamente independentes) fontes de sinais; e x_i sinais misturados;
- Todos os componentes s_i , com a exceção para apenas um componente, devem ser não gaussianos;
- O número das misturas lineares observadas m deve ser tão grande quanto o número dos componentes independentes n , isto é, $m \geq n$;
- A matriz \mathbf{A} deve ser de *rank* completo da coluna;
- Não conhecemos os componentes s_i nem \mathbf{A} .

4.2.2.2 Branqueamento

Conforme o diagrama em blocos da Figura 18, nós temos m variáveis escalares aleatórias x_1, x_2, \dots, x_m as quais são assumidas como uma combinação linear de n desconhecidas s_1, s_2, \dots, s_n componentes independentes que são mutuamente estatisticamente independentes, e possuem média zero. Além de que, consideramos $n \leq m$. Organizando as variáveis observadas x_i no vetor $\mathbf{x} = (x_1, \dots, x_m)^T$ e as variáveis s_i no vetor \mathbf{s} , respectivamente; podemos então escrever a relação linear como:

$$\mathbf{x} = \mathbf{A} \mathbf{s} \quad (19)$$

Onde, \mathbf{A} é uma matriz desconhecida $m \times n$ de *rank* completo, chamada de matriz de mistura.

O problema básico de ICA é então estimar a componente original s_i da mistura x_j ou, equivalentemente, estimar a matriz de mistura \mathbf{A} . A restrição fundamental do modelo de ICA é que podemos estimar componentes independentes não-Gaussianas (com exceção se somente

uma das componentes independente é Gaussiana). Além disso, nem as energias nem os sinais das componentes independentes podem ser estimados, porque nenhuma constante multiplicando uma componente independente na equação (19) pode ser cancelada por dividir a coluna correspondente da matriz de mistura \mathbf{A} pela mesma constante.

Por conveniência matemática, definimos aqui que a componente independente s_i tem variância unitária. Isto faz as (não-Gaussiana) componentes independentes únicas. Embora nenhuma ordem seja garantida entre as componentes independentes.

Na separação cega da fonte, os valores observados de \mathbf{x} correspondem a uma realização de um m -dimensional sinal em tempo discreto $\mathbf{x}(t)$, $t = 1, 2, \dots$. As componentes $s_i(t)$ são chamadas sinais fonte, são usualmente fontes de sinais originais não corrompidos ou fontes de ruídos.

O problema de estimar a matriz \mathbf{A} na equação (20) pode ser simplificada executando um pré-branqueamento do dado \mathbf{x} , [41]. Temos por pré-branqueamento a divisão dos sinais processados pelos seus respectivos desvios padrões, normalizando.

Dessa forma o vetor observado \mathbf{x} é linearmente transformado para o vetor $\mathbf{u} = \mathbf{W}\mathbf{x}$ com seus elementos u_i mutuamente decorrelacionados e todos com variância unitária. Assim a matriz correlação de \mathbf{u} é igual a identidade: $E[\mathbf{u}\mathbf{u}^T] = \mathbf{I}$. Esta transformação é sempre possível e pode ser realizada pela Análise de Componente Principal (PCA). Ao mesmo tempo, a dimensionalidade dos dados seriam reduzidas de forma que a dimensão da transformação do vetor de dado \mathbf{u} igualaria a n , o número de componentes independentes. Com o conseqüente efeito de reduzir ruído. Depois da transformação teremos:

$$\mathbf{u} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{A}\mathbf{s} = \mathbf{B}\mathbf{s} \quad (20)$$

Onde $\mathbf{B} = \mathbf{W}\mathbf{A}$ é uma matriz ortogonal devido as suposições sobre as componentes s_i . Fazendo-se $E[\mathbf{u}\mathbf{u}^T] = \mathbf{B}E[\mathbf{s}\mathbf{s}^T]\mathbf{B}^T = \mathbf{B}\mathbf{B}^T = \mathbf{I}$. Assim nós reduzimos o problema de encontrar uma

arbitrária matriz \mathbf{A} de rank completo para o problema de encontrar uma matriz ortogonal \mathbf{B} , nesse caso temos $\mathbf{s}=\mathbf{B}^T\mathbf{u}$. Se a i -ésima coluna de B é b_i , então a i -ésima componente independente pode ser calculada pela observação de \mathbf{u} como $s_i=(\mathbf{b}_i)^T\mathbf{u}$.

A estimação dos modelos dos dados da análise de componentes independentes é executada geralmente formulando uma função objetiva e um algoritmo de otimização. [40, 41]

$$ICA = \text{função objetiva} + \text{algoritmo de otimização}$$

Onde as propriedades estatísticas (por exemplo, consistência, variância assintótica e robustez) do método de ICA dependem da escolha da função objetiva, e as propriedades algorítmicas (isto é, velocidade de convergência, capacidade de memória e estabilidade) dependem do algoritmo de otimização.

4.2.2.3 Funções Objetivas

ICA pode ser entendida no contexto de estimação de distribuição de probabilidade. Vamos primeiro lembrar que a distribuição em comum de um conjunto de observações independentes, $x = (x_1, \dots, x_n)^T$ tomadas de uma mesma distribuição $p(x, \theta)$ é determinado pelo produto das distribuições marginais quando os elementos são mutuamente independentes, isto é.,

$$p(x, \theta) = \prod_i^N p(x_i, \theta) \quad (21)$$

Dado a modificação dos parâmetros θ , algoritmos de ICA deveriam achar uma estimação da verdadeira densidade $p(x, \theta)$, onde θ é um parâmetro verdadeiro. Um modo para achar a distribuição atual é por meio da divergência Kullback-Liebert (K-L), dada por,

$$l(p, \tilde{p}) = \int p(x, \theta) \ln \frac{p(x, \theta)}{\tilde{p}(x, \theta)} dx \quad (22)$$

ou, através de medidas para quantificar a similaridade de uma distribuição qualquer com relação à gaussiana, cumulantes de alta-ordem como Kurtosis [42]

$$k = E[x_i^4] - 3E[x_i^2]^2 \quad (23)$$

Sinais que apresentam Kurtosis com sinal negativo são comumente chamado de subgaussianos e o contrário, sinais que apresentam Kurtosis positivas são chamados de supergaussianos. Os primeiros normalmente apresentam uma distribuição com maior concentração nas regiões marginais enquanto que os últimos apresentam uma distribuição com um pico mais pontiagudo e uma cauda mais comprida. Quanto maior o valor do cumulante, mais longe estará da gaussiana. Em outras palavras, quando o cumulante é máximo garante independência.

Após escolher um princípio de estimação para ICA, necessitamos de um método prático para sua execução. E muitos são os algoritmos para ICA com características distintas, baseadas em estabilidade, velocidade de convergência etc. Em seguida, comentamos alguns desses algoritmos.

4. 2.2.4 Algoritmos para ICA

a) HERÁULT-JUTTEN (HJ) – O pioneirismo na ICA é associado ao trabalho de Heráult-Jutten (1983) que propuseram um algoritmo neural de separação cega. Tal algoritmo é simples, porém falha na separação de mais de duas fontes. [26]

b) ALGORITMO DE COMON – Em 1994 é introduzido o conceito de uma função de contraste independente de permutação de mudanças de amplitude que deve ser maximizada para se alcançar o critério de independência. Comon [41] mostrou que a medida de informação mútua pode ser usada como uma função de contraste. E que o processo de separação constitui-se em duas partes distintas: o pré-branqueamento, que pode ser obtido com o uso de PCA e que garante a decorrelação estatística; e a obtenção de uma

transformação ortogonal que maximize a entropia das densidades dos sinais obtidos. A transformação tem que ser ortogonal para manter a condição de não correlação obtida após o branqueamento.

c) FastICA – Este algoritmo para ICA desenvolvido por Hyvarinen e Oja [43] em 1997 pode ser usado para a separação cega de fonte e extração da característica. Mostra que regras de aprendizagem de redes neurais podem ser transformadas em um ponto fixo de interação, proporcionando um algoritmo simples, que não depende de nenhum parâmetro usualmente definido, e é rápido para convergir a mais rápida solução concedida por um dado. O algoritmo encontra em tempo todos os componentes independentes não-gaussiano, sem se importar com sua probabilidade de distribuição. O cálculo pode ser realizado em modo batch ou semi-adaptativo. A convergência de velocidade desse algoritmo é cúbica.

d) JADE – Em Cardoso [44], 1999 é feita uma análise de métodos algébricos baseados em cumulantes mostrando como correlações de alta ordem podem ser eficientemente exploradas para encontrar componentes independentes. Este algoritmo é chamado de Algoritmo de Jacobi devido ao fato de se procurar maximizar as medidas de independência através do método de diagonalização de Jacobi. As medidas de independência usadas são baseadas em cumulantes de quarta ordem. A vantagem em relação a outros algoritmos é que se podem modificar os parâmetros de treinamento em passos grandes sem ter problemas de convergência. As grandes vantagens desse método quando comparado aos métodos baseados em gradientes é que não há parâmetros a serem sintonizados na implementação básica e não são necessários algoritmos complexos de aprendizagem. Normalmente nos métodos baseados em gradientes é utilizada uma aprendizagem baseada em alguma heurística. O problema desta técnica é que falha em algumas distribuições.

e) RICA – O algoritmo “Robust ICA” foi proposto por Cichocki e Barros [45] em 1999 para ser aplicado em casos onde os sinais aparecem contaminados por sinais de ruído como eletroencefalograma e magnetoencefalograma (EEG / MEG), por exemplo. Trabalha com estatísticas de segunda ordem e ao contrário de outros algoritmos permite trabalhar apenas com estatística de segunda ordem e ao contrário dos demais algoritmos permite a

separação de sinais com baixa medida de kurtosis, ou seja, gaussianos. A abordagem é a de desenvolver um algoritmo que permita extrair apenas as fontes potencialmente interessantes. A proposta do RICA é a de extrair as fontes sequencialmente, uma de cada vez. Para cada etapa de execução do algoritmo um sinal é extraído e excluído da mistura. Assim na próxima etapa um outro sinal será obtido.

Os autores confirmam, com testes, a validade do método proposto para aplicação na separação de sinais com Kurtosis próximas de zero tendo obtido a separação em situações nas quais outros algoritmos falharam.

4.2.2.5 Aplicações de ICA

Uma das aplicações clássicas do modelo de ICA é na separação cega de fonte (BSS), e como exemplo temos o problema do *Cocktail Party* explicado anteriormente.

Vale ressaltar que desde que as pesquisas utilizando ICA com aplicação para a separação cega de fonte muitos autores que tratam o problema de ICA não usam o termo ICA, e usam simplesmente a notação de separação cega de fonte. É bom deixar claro a distinção entre esses termos. ICA é um problema teórico ou ainda modelagem de dados para diferentes aplicações, enquanto que a BSS é uma aplicação que pode ser resolvida usando diferentes aproximações teóricas, incluindo, mas não se limitando a somente ICA. Em outras palavras, o problema da separação cega de fonte pode ser resolvido usando muitos métodos diferentes de ICA como, por exemplo, a informação da frequência, ou as propriedades dos espectros. [27, 28, 46]

Em [47, 48, 49] temos resultados da aplicação de ICA para separar dados de um EEG e MEG.

Uma outra aplicação é no setor econômico, o que pode ser observado em [50]. Recentes aplicações na área de telecomunicações foram publicadas [51]. Podemos ver também em [52] aplicações de ICA para a extração da característica; que é motivada pela

teoria da redução de redundância. E uma aplicação menos direta dos métodos de ICA pode ser mostrada na deconvolução cega.

Os conhecimentos estabelecidos neste capítulo servem de base para o desenvolvimento da proposta de novo modelo, a seguir.

CAPÍTULO 5

MODELO PROPOSTO

O modelo proposto e desenvolvido neste trabalho, Figura 18, aplicado à BSS não é muito diferente de uma rede neural artificial (RNA) e está dividido em duas fases: uma de aprendizagem e uma de estimação do erro médio quadrático, ou fase de funcionamento, também denominada de teste. A diferença entre o modelo e uma RNA está no fato de que não usamos a não-linearidade do modelo neural.

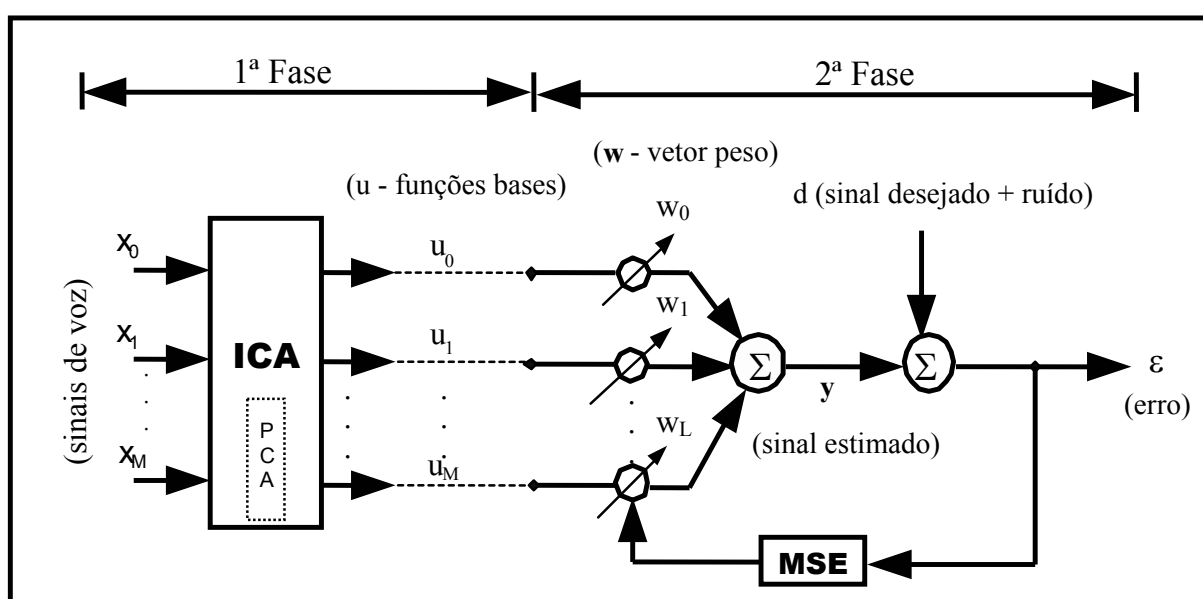


Figura 18 - Diagrama em blocos do sistema: 1ª Fase (Aprendizagem) e 2ª Fase (Teste).

5.1 Introdução

A fase de aprendizagem, à esquerda no diagrama em blocos é onde o sistema aprende as funções bases dos sinais de voz de entrada pelo algoritmo de ICA e na segunda fase, de teste, calcula o sinal desejado por uma estimação simples de erro médio quadrático (MSE). Nesta proposta a recuperação de um sinal de voz é feita através da minimização do erro médio

quadrático em que o sinal estimado será descrito como uma soma ponderada das funções bases, sendo estas obtidas através de aprendizado por ICA.

5.2 O Método

Neste trabalho consideramos que o sinal de voz observado $\mathbf{x}(t)$ é codificado em um conjunto de \mathbf{M} respostas $\mathbf{a}(t) = [a_1(t) \dots a_M(t)]^T$. O Objetivo do código eficiente é derivar um vetor de filtros $\mathbf{u}(t) = [u_1(t) \dots u_M(t)]^T$, ou funções bases que minimizem a dependência estatística mútua entre as respostas. O sinal observado pode ser escrito como:

$$\mathbf{x} = \mathbf{a}^T \mathbf{u} \quad (24)$$

Uma estimação de \mathbf{u} e \mathbf{a} pode ser considerada tanto pela análise de componentes principais, PCA, ou análise de componentes independentes, ICA, por eles suporem que qualquer elemento de \mathbf{u} são mutuamente não correlacionados ou independentes, respectivamente. Na teoria, um código ideal transforma a entrada dos sinais para que as saídas sejam estatisticamente independentes, removendo o ruído.

O sinal observado, conforme mostra a equação:

$$\mathbf{x}_k = [x_{0k}, x_{1k}, \dots, x_{Mk}]^T \quad (25)$$

compõe a entrada do algoritmo de ICA no modelo. Nesta notação, T representa a matriz transposta, k subscrito é usado como indicador de tempo e x_k é um vetor coluna que irá conter amostras de um mesmo som de comprimento 200, tomadas aleatoriamente para o algoritmo de ICA fazer o cálculo das funções bases dos sinais, ou seja, o objetivo é derivar um vetor de filtros, ou funções bases que minimizem a dependência estatística mútua entre as respostas; representados na figura do modelo por :

$$\mathbf{u} = [u_{1k}, u_{2k}, \dots, u_{Mk}]^T \quad (26)$$

São muitos algoritmos propostos para ICA, utilizamos o algoritmo proposto por Aapo Hyvarinen [43], na primeira fase do nosso modelo, fase esta chamada de aprendizagem e onde as funções bases e respostas são estimadas com altos Kurtosis, aqueles que apresentam uma distância maior da gaussiana.

A segunda fase é a estimação do erro médio quadrático considerando que um sinal desejado, corrompido por ruído, é adicionado ao modelo.

O sinal desejado d (desejado+ruído), faz parte dos sinais da fase de aprendizagem e que o comprimento da contribuição da aprendizagem, caso contrário a saída seria nula e tem que ser pequeno o bastante de forma que por aquele tempo em particular, o sinal é estacionário.

Nesta segunda fase do modelo temos que o vetor peso é representado na equação (27) por:

$$\mathbf{w}_k = [w_{0k}, w_{1k}, \dots, w_{Lk}]^T \quad (27)$$

e que o sinal estimado y_k será descrito como uma soma ponderada das funções bases. Conforme mostrado abaixo:

$$\mathbf{y}_k = \mathbf{w}_k^T \mathbf{u}_k^T \quad (28)$$

Em seguida é feita uma comparação da saída y_k com o sinal desejado \mathbf{d}_k para a obtenção de um sinal de erro ε_k , instantâneo ajustando ou otimizando o vetor peso para minimizar este sinal, ou seja, através do algoritmo de minimização do Erro Médio Quadrático (MSE), pois quanto menor o erro obtido, mais próximo o sinal estimado estará do sinal desejado. O sinal erro instantâneo com índice de tempo k é dado por:

$$\varepsilon_k = d_k - y_k \quad (29)$$

Substituindo a equação (28) na equação (29) temos uma nova expressão:

$$\varepsilon_k = \mathbf{d}_k - \mathbf{u}_k^T \mathbf{w} = \mathbf{d}_k - \mathbf{w}^T \mathbf{u}_k \quad (30)$$

O vetor peso, para simplificar a demonstração da função desempenho do sistema adaptativo, não possui o índice k , pois nesta demonstração não necessitamos ajustar o vetor peso. Fazendo o quadrado da equação (30), iremos obter o erro quadrático instantâneo.

$$\varepsilon_k^2 = \mathbf{d}_k^2 - \mathbf{w}^T \mathbf{u}_k \mathbf{u}_k^T \mathbf{w} - 2\mathbf{d}_k - \mathbf{u}_k^T \mathbf{w} \quad (31)$$

As variáveis ε_k , \mathbf{d}_k e \mathbf{u}_k são estatisticamente estacionárias e fazendo o valor esperado da equação (31), obteremos o erro médio quadrático, expresso pela equação (32).

$$E[\varepsilon_k^2] = E[\mathbf{d}_k^2] + \mathbf{w}^T E[\mathbf{u}_k \mathbf{u}_k^T] \mathbf{w} - 2E[\mathbf{d}_k \mathbf{u}_k^T] \mathbf{w} \quad (32)$$

Devemos observar que o valor esperado da soma é a soma do valor esperado, mas o valor esperado do produto é o produto do valor esperado, somente se as variáveis forem estatisticamente independentes. Sendo que \mathbf{u}_k e \mathbf{d}_k não são geralmente independentes, o erro médio quadrático pode ser representado por conveniência através da matriz covariância de \mathbf{u}_k , designada pela letra \mathbf{R} e pelo vetor coluna, designado por \mathbf{P} .

$$\mathbf{R} = E[\mathbf{u}_k \mathbf{u}_k^T] = E \begin{bmatrix} \mathbf{u}_{0k}^2 & \mathbf{u}_{0k} \mathbf{u}_{1k} & \cdots & \mathbf{u}_{0k} \mathbf{u}_{lk} \\ \mathbf{u}_{1k} \mathbf{u}_{0k} & \mathbf{u}_{1k}^2 & \cdots & \mathbf{u}_{1k} \mathbf{u}_{lk} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{u}_{lk} \mathbf{u}_{0k} & \mathbf{u}_{lk} \mathbf{u}_{1k} & \cdots & \mathbf{u}_{lk}^2 \end{bmatrix} \quad (33)$$

$$\mathbf{P} = E[\mathbf{d}_k \mathbf{u}_k] = E[\mathbf{d}_k \mathbf{u}_{0k} \mathbf{u}_{1k} \dots \mathbf{d}_k \mathbf{u}_{1k}]^T \quad (34)$$

Os elementos de \mathbf{R} e \mathbf{P} são todos estatisticamente constantes de segunda ordem, quando \mathbf{u}_k e \mathbf{d}_k são estacionários. Agora temos o erro médio quadrático em termos da equação (33) e (34), podendo ser reescrito da seguinte forma:

$$\text{MSE} = \xi = E[\varepsilon_k^2] = E[d_k^2] + \mathbf{w}^T \mathbf{R} \mathbf{w} - 2 \mathbf{P}^T \mathbf{w} \quad (35)$$

A superfície de desempenho do sistema adaptativo é geralmente uma parábola ou uma hipérbole, caso exista mais de dois pesos. Ela deve ser côncava para cima, caso contrário poderia existir pesos que tornariam o erro médio quadrático negativo, que por definição não existe.

A superfície é projetada sobre o plano do vetor peso, com um ponto representando o vetor peso ótimo, o qual é representado por \mathbf{w}^* , que implica no mínimo erro médio quadrático obtido com o método.

O erro médio quadrático é uma função quadrada das componentes do vetor peso, quando o sinal de entrada e a resposta desejada são variáveis estocásticas estacionárias, então existe um único ponto global ótimo, no mínimo local existente.

Em muitos processos adaptativos o ajuste do vetor peso na procura do mínimo erro médio quadrático é feito pelo método do gradiente. O gradiente para o erro mínimo médio quadrático é designado por $\nabla(\xi)$ ou simplesmente pelo operador ∇ . Pode ser determinado através da equação (36), obtendo um vetor coluna.

$$\nabla = \frac{\partial \xi}{\partial \mathbf{w}} = \left[\frac{\partial \xi}{\partial w_0} \quad \frac{\partial \xi}{\partial w_1} \quad \dots \quad \frac{\partial \xi}{\partial w_L} \right]^T$$

$$\nabla = 2\mathbf{R}\mathbf{w} - 2\mathbf{P} \quad (36)$$

Sendo que \mathbf{R} e \mathbf{P} já foram definidos pelas equações (33) e (34) respectivamente. Esta expressão é obtida expandindo a equação (35) derivando-a em relação a cada componente do vetor peso.

Para o mínimo erro médio quadrático, teremos que encontrar o vetor peso ótimo, dada a superfície de desempenho do sistema adaptativo, que é geralmente uma parábola ou uma hipérbole, caso existam mais de dois pesos. A superfície é projetada sobre o plano do vetor peso, com um ponto representando o vetor peso ótimo, o qual torna o gradiente igual a zero. Podemos representar este vetor peso por \mathbf{w}^* .

$$\nabla = 0 = 2\mathbf{R}\mathbf{w}^* - 2\mathbf{P}$$

$$\mathbf{w}^* = \mathbf{R}^{-1}\mathbf{P} \quad (37)$$

O mínimo erro médio quadrático é agora obtido, por substituição de \mathbf{w} na equação (35) por \mathbf{w}^* . A vantagem do erro médio quadrático é que ele só tem um ponto mínimo, um ponto ótimo \mathbf{w}^* que irá nos dar um erro mínimo, que é a solução do problema.

$$\begin{aligned} \xi_{\min} &= E[\varepsilon_k^2] = E[\mathbf{d}_k^2] + \mathbf{w}^{*T}\mathbf{R}\mathbf{w}^* - 2\mathbf{P}^T\mathbf{w}^* \\ &= E[\mathbf{d}_k^2] + [\mathbf{R}^{-1}\mathbf{P}]^T \mathbf{R}\mathbf{R}^{-1}\mathbf{P} - 2\mathbf{P}^T\mathbf{R}^{-1}\mathbf{P} \\ &= E[\mathbf{d}_k^2] + \mathbf{P}^T\mathbf{R}^{-1}\mathbf{P} - 2\mathbf{P}^T\mathbf{R}^{-1}\mathbf{P} \\ &= E[\mathbf{d}_k^2] - \mathbf{P}^T\mathbf{R}^{-1}\mathbf{P} \\ \xi_{\min} &= E[\mathbf{d}_k^2] - \mathbf{P}^T\mathbf{w}^* \end{aligned} \quad (38)$$

Para validar o modelo proposto foram feitas simulações, cujos resultados são mostrados e discutidos.

CAPÍTULO 6

RESULTADOS

Neste capítulo mostramos as simulações que foram realizadas para a validação do modelo proposto.

6.1 Introdução

As simulações consistiram em usar o software Matlab, versão 6.5 e um conjunto de sinais de locução de números proferidos de zero até nove, em português (zero, um, dois, três, quatro, cinco, seis, sete, oito, nove) por um orador masculino obtido da Base de dados da UFMG através do Professor Hani Yehia, a uma frequência de amostragem de 8 kHz. O orador repetiu cada número dez vezes. Foram removidas manualmente partes de cada expressão vocal - subtração espectral -, preservando somente o que continha a fala atual.

Simulação 1:

a) Nós testamos a validade da teoria extraindo diferentes números de funções bases. Isto foi feito escolhendo-se um número e tomando aleatoriamente de qualquer parte das dez expressões vocais um vetor de treinamento de comprimento 200. Treinamos o algoritmo de ICA que tem como introduzir esses vetores de treinamento e tomamos as funções bases com o kurtosis mais alto. A Figura 19 mostra exemplos de sinais originais e a obtenção das funções bases.

b) Nós extraímos 20, 40, 60, 80, 100, 120 e 140 funções bases dos sinais de voz de entrada onde um orador masculino proferiu os números de zero até nove em português e calculamos a estimação do erro pelo algoritmo MSE, para compor outra entrada aleatoriamente escolhida, como descrito no diagrama de blocos do sistema. Nós realizamos 100 iterações do algoritmo e calculamos a média dos erros tanto para PCA quanto para ICA conforme mostrado na Figura 20, de acordo com o número, de funções bases.

Na Figura 19 mostramos amostras de dois sinais originais e suas respectivas funções bases obtidas através do algoritmo de ICA.

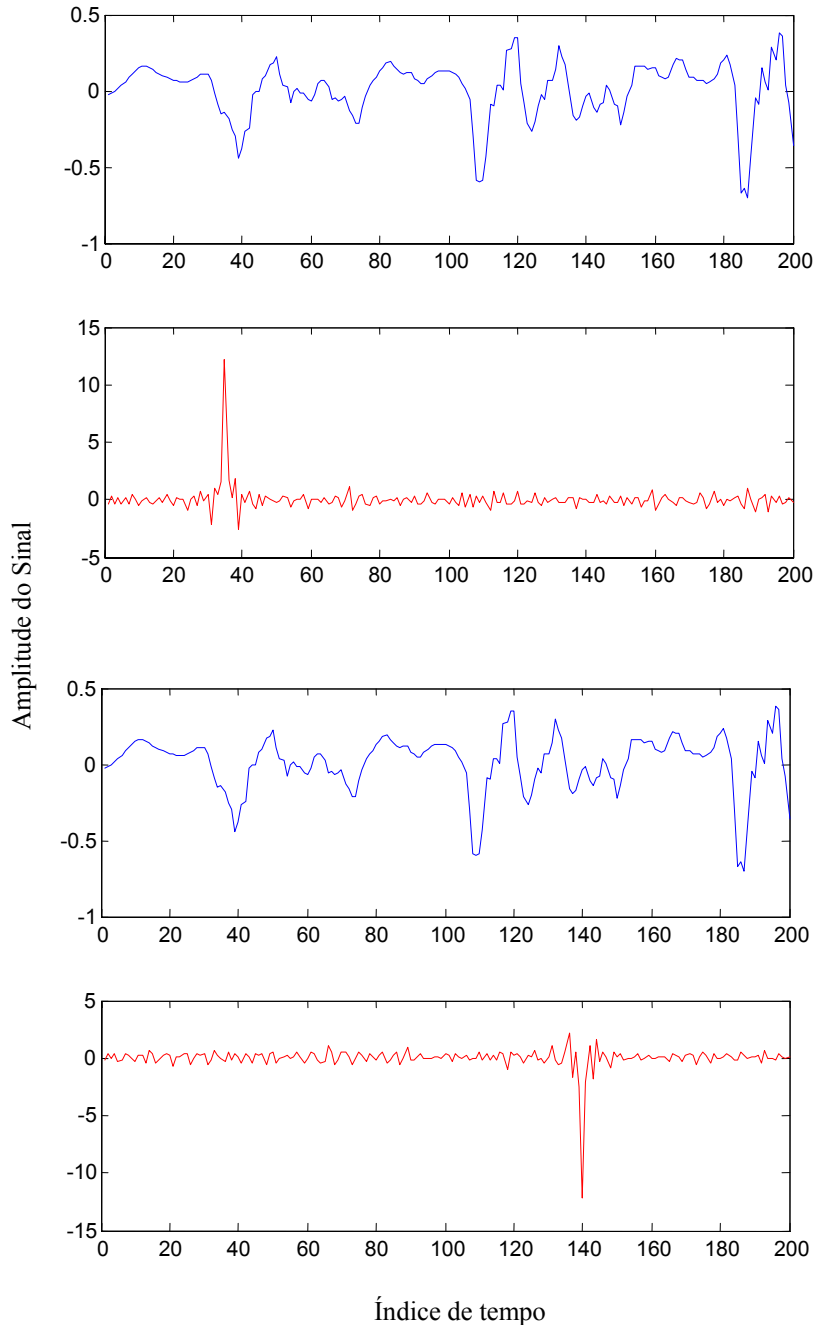


Figura 19 – Sinais originais (azul) e respectivas funções bases obtidas (vermelho) – Simulação 1 (a).

Na Figura 20 mostramos a média dos erros para as diferentes palavras, onde um orador masculino proferiu os números de zero até nove em português, respectivamente para PCA e ICA sem adicionar ruído ao sistema.

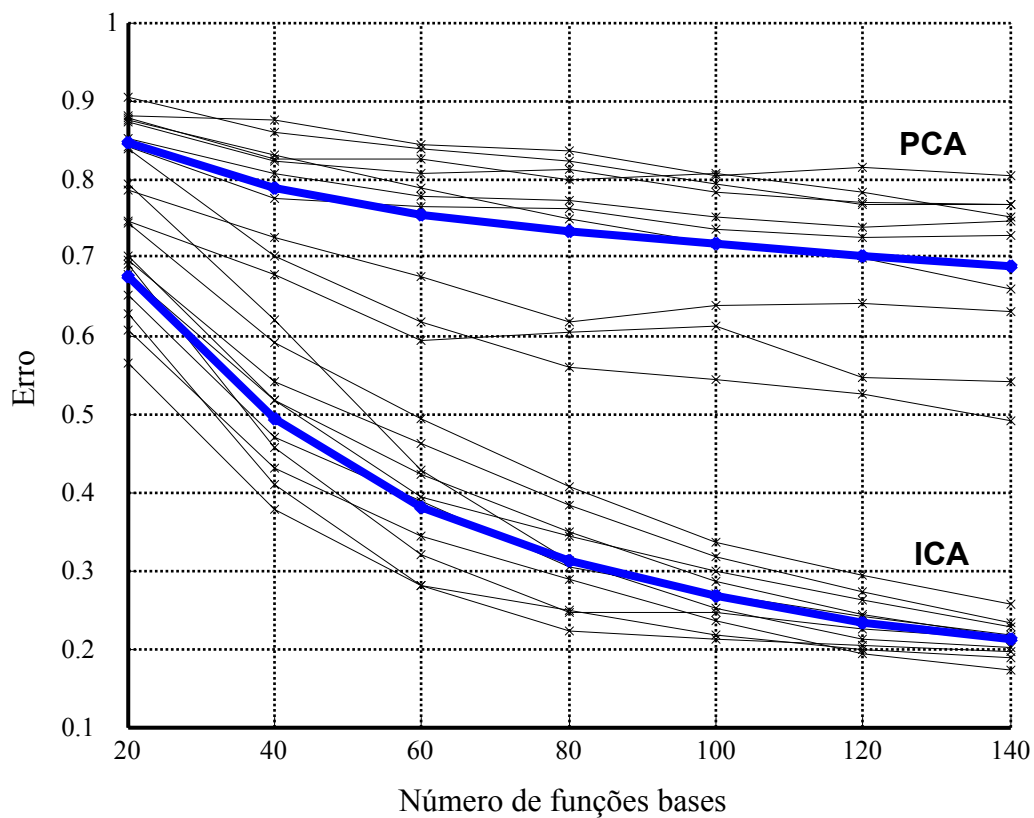


Figura 20 - Erro versus número de funções bases sem ruído adicionado – Simulação 1 (b).

Simulação 2:

a) Foram feitas simulações com diferentes tipos de ruído, para testar a robustez do método frente a sinais indesejáveis. Segue exemplo com ruído do tipo música adicionado para as diferentes palavras onde um orador masculino proferiu os números de zero até nove em português, com 100 iterações do algoritmo e cálculo da média dos erros para PCA e ICA respectivamente, representada pela linha azul conforme a Figura 21.

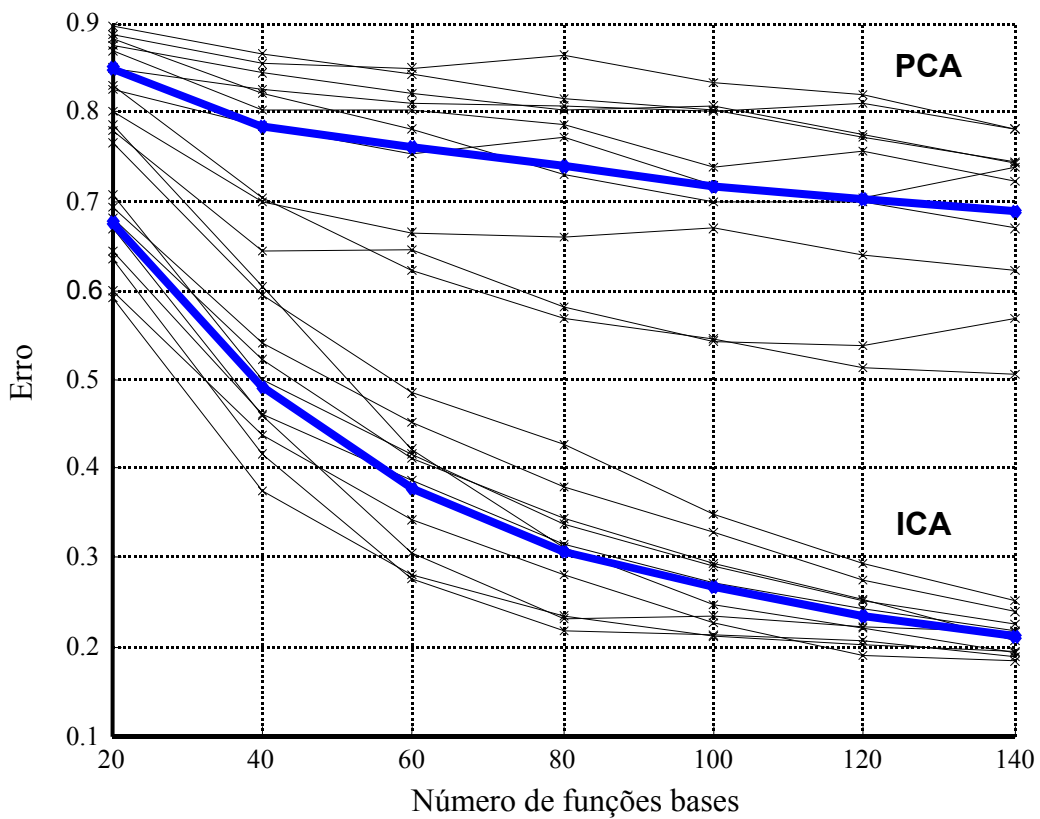


Figura 21 - Erro versus número de funções bases com ruído do tipo música adicionado - Simulação 2 (a).

b) Representação gráfica com ruído do tipo vocalização de animal adicionado para as diferentes palavras onde um orador masculino proferiu os números de zero até nove em português, com 100 iterações do algoritmo e cálculo da média dos erros para PCA e ICA respectivamente, indicado pela linha azul conforme mostra a Figura 22.

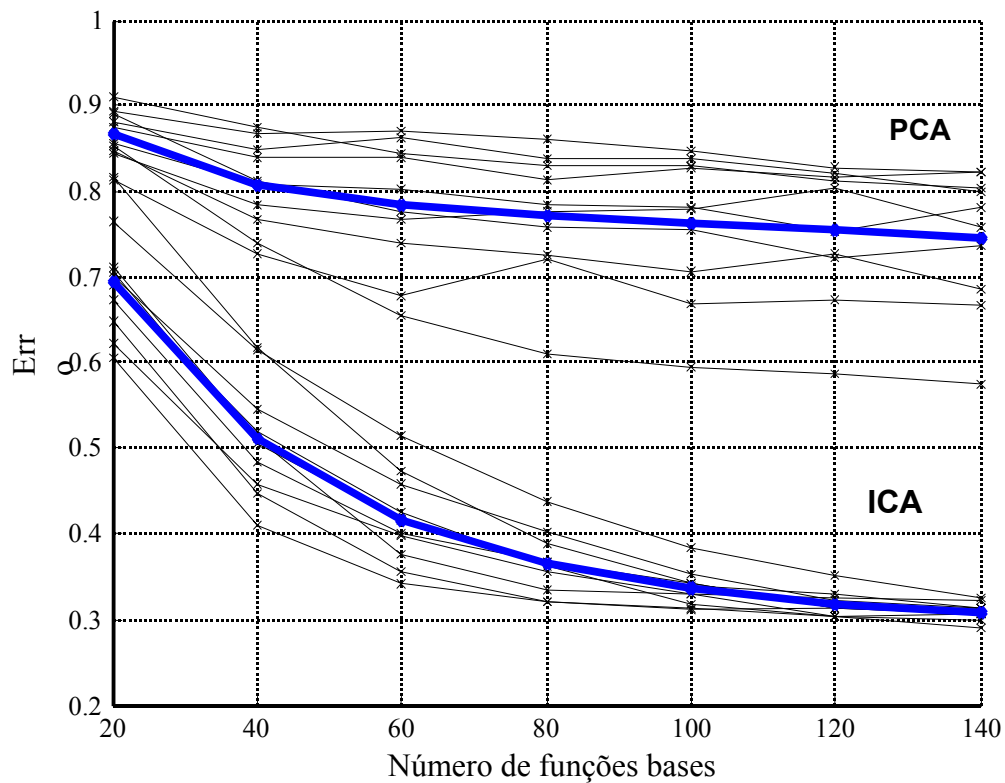


Figura 22 – Erro versus número de funções bases com ruído de vocalização de animal adicionado Simulação 2 (b).

Simulação 3: Também recuperamos o sinal de voz de locução de um dado número através da utilização das funções bases de outro número, sem ruído adicionado. Em outras palavras, as funções bases foram extraídas de um número proferido diferente daquele do sinal desejado. Testando o algoritmo para 100 iterações, o erro resultante foi pequeno, aproximadamente 0,28, conforme pode ser observado na Figura 23.

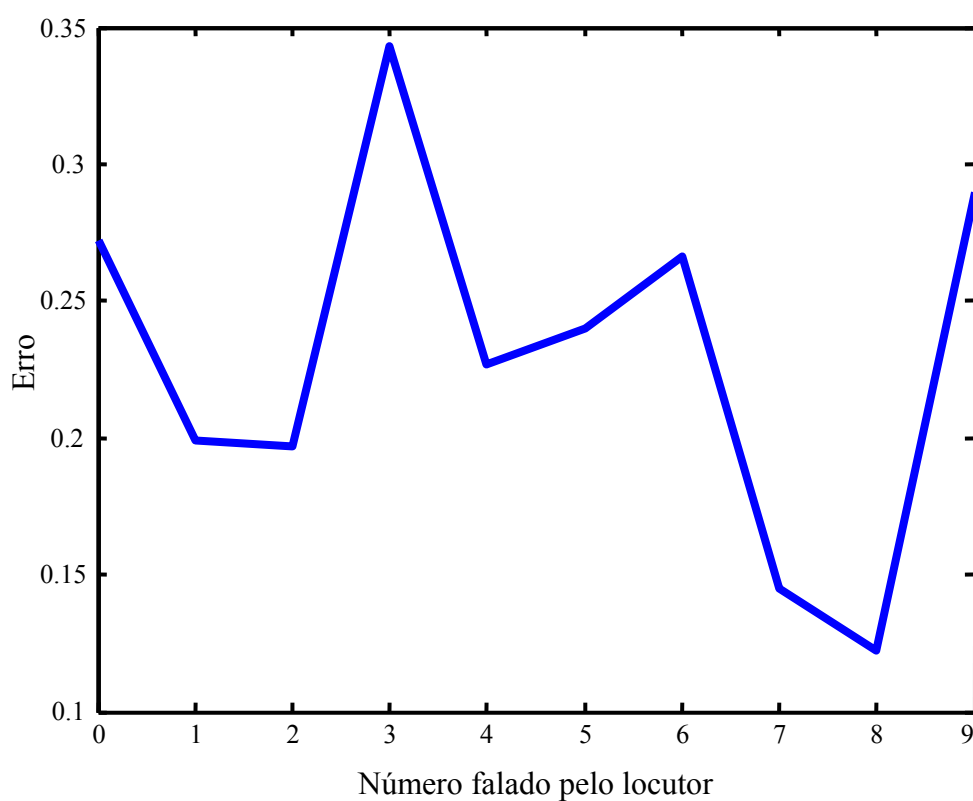


Figura 23 – Erro entre o sinal desejado e o sinal recuperado com as bases de outros sinais, sem ruído adicionado.

Simulação 4:

Nesta simulação testamos a sensibilidade do método quanto à reverberação. Geramos várias respostas ao impulso com decaimento tipo exponencial, por ser uma forma semelhante à resposta ao impulso de um ambiente real. As Figuras 24, 25, 26, 27 e 28 mostram as respostas ao impulso geradas artificialmente. Criamos um filtro FIR com resposta finita e aplicamos esse sinal ao sistema, fixamos o número de funções bases e calculamos a média dos erros neste ambiente com reverberação para 100 iterações. Variamos o comprimento do filtro para 50, 100, 150, 150, 200 e 250 atrasos.

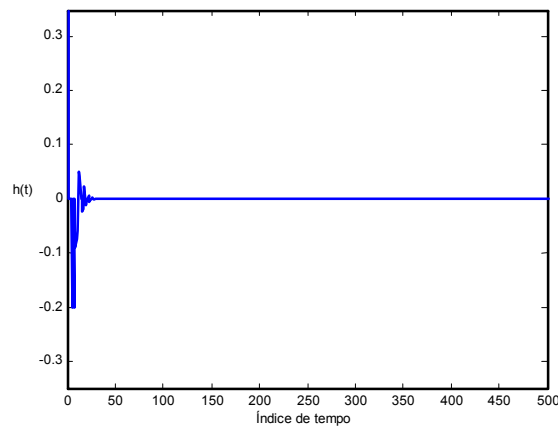


Figura 24 - Representação de resposta ao impulso $h(t)$ gerada artificialmente para filtro com 50 atrasos.

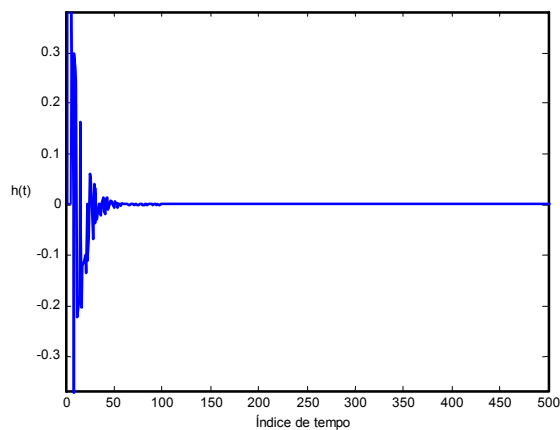


Figura 25 - Representação de resposta ao impulso $h(t)$ gerada artificialmente para filtro com 100 atrasos.

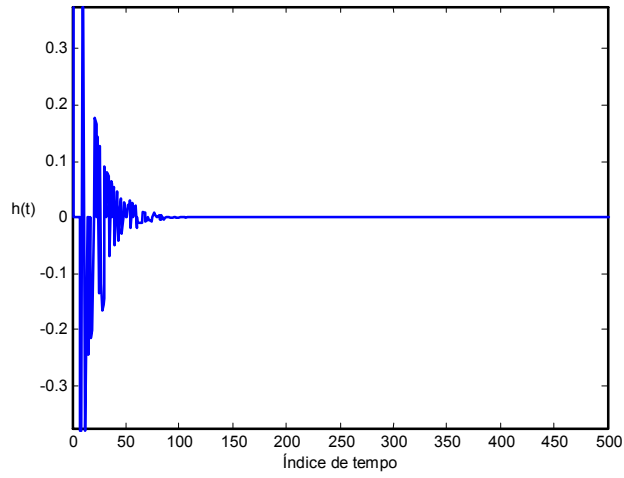


Figura 26 - Representação de resposta ao impulso $h(t)$ gerada artificialmente para filtro com 150 atrasos.

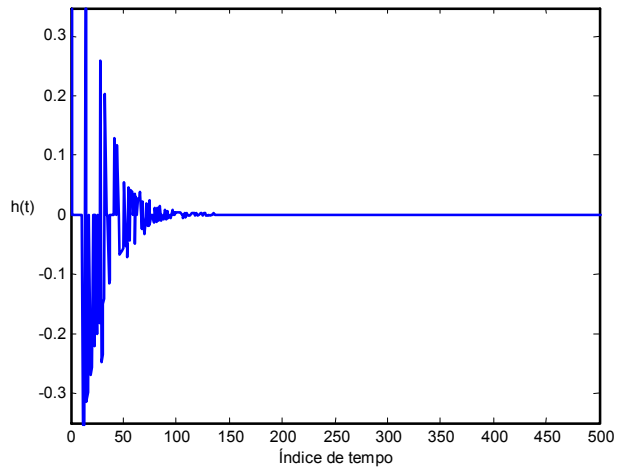


Figura 27 - Representação de resposta ao impulso $h(t)$ gerada artificialmente para filtro com 200 atrasos

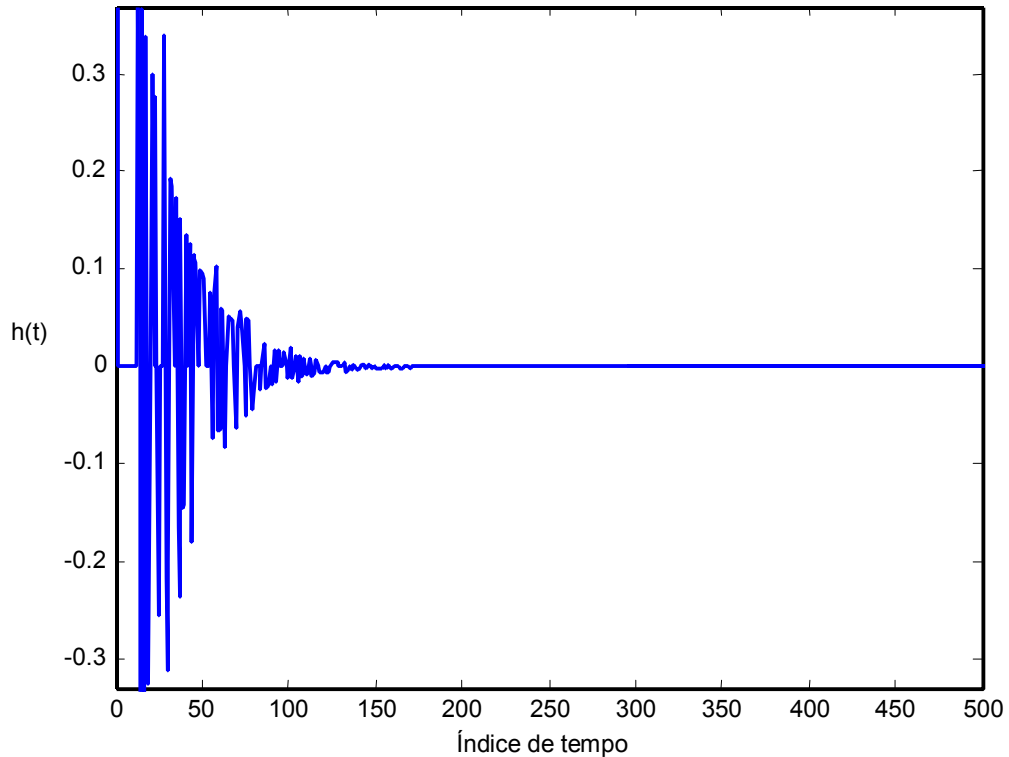


Figura 28 - Representação de resposta ao impulso $h(t)$ gerada artificialmente para filtro com 250 atrasos.

uma frequência de amostragem de 8 kHz, o comprimento de 250 atrasos corresponde a um atraso temporal de 31,25 ms, entendido como o comprimento aproximado da resposta ao impulso da reverberação em uma sala.

As Figuras 29, 30, 31, 32 e 33 apresentam uma amostra de tamanho 200 de um sinal de voz original (linha em azul) e a mesma amostra com efeito da reverberação (linha em vermelho), aplicado ao ambiente – “sala” -, para os atrasos de 50, 100, 150, 200 e 250 respectivamente.

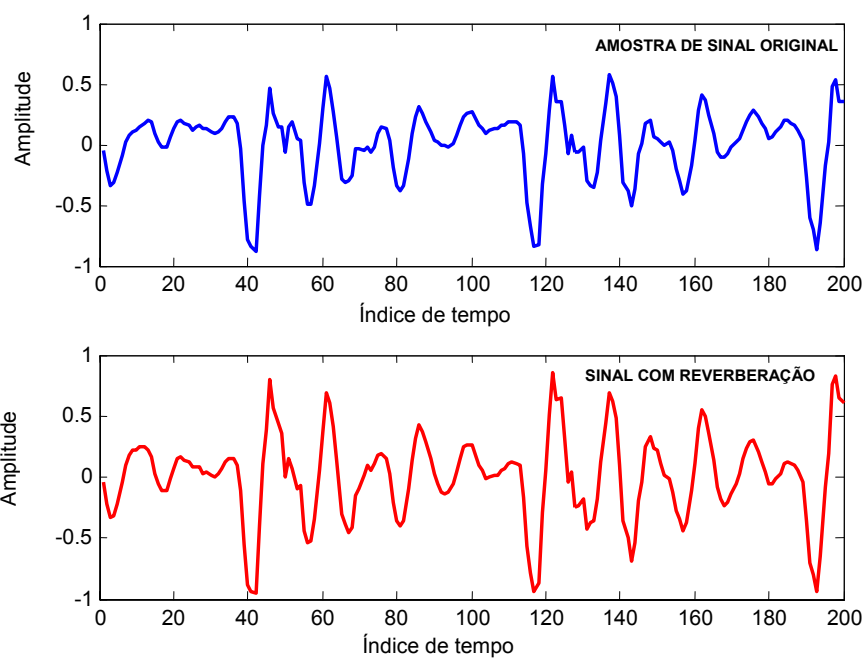


Figura 29 – Sinal de voz original e com efeito da reverberação para filtro com 50 atrasos.

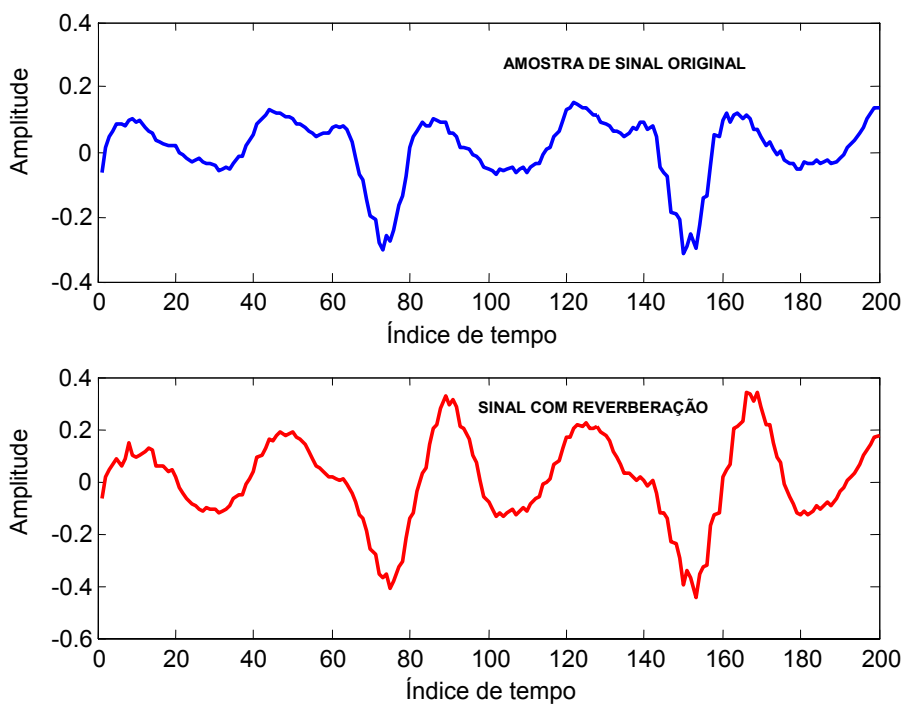


Figura 30 – Sinal de voz original e com efeito da reverberação para filtro com 100 atrasos.

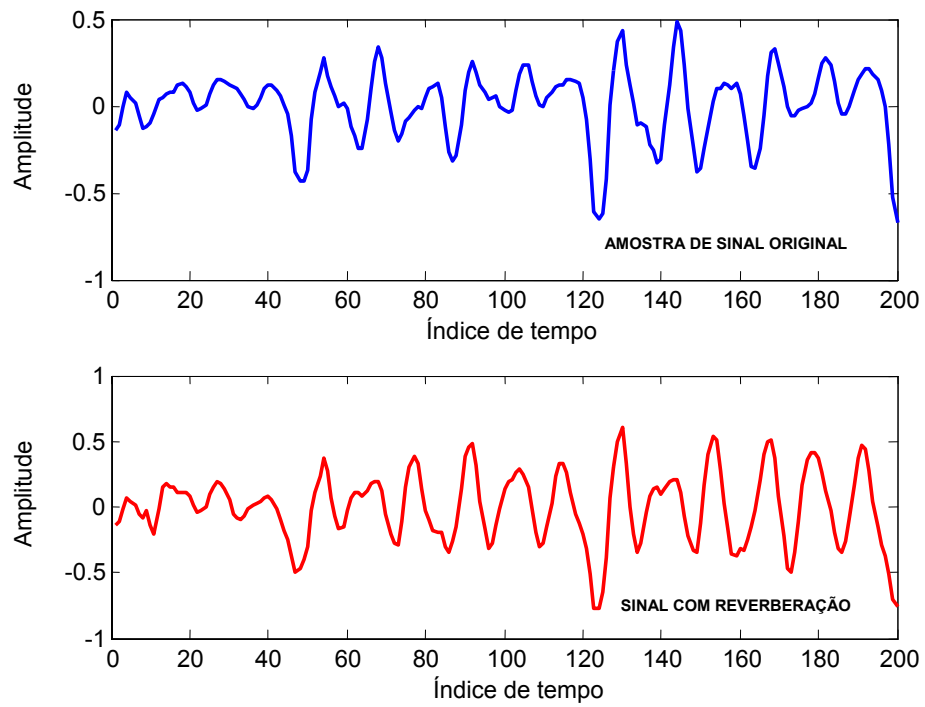


Figura 31 – Sinal de voz original e com efeito da reverberação para um filtro com 150 atrasos.

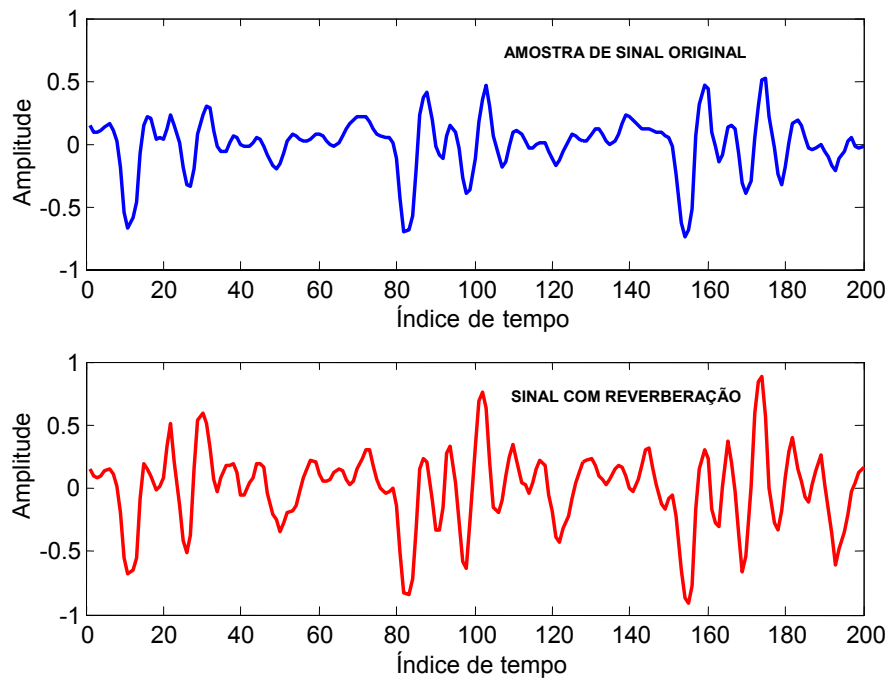


Figura 32 – Sinal de voz original e com efeito da reverberação para um filtro com 200 atrasos.

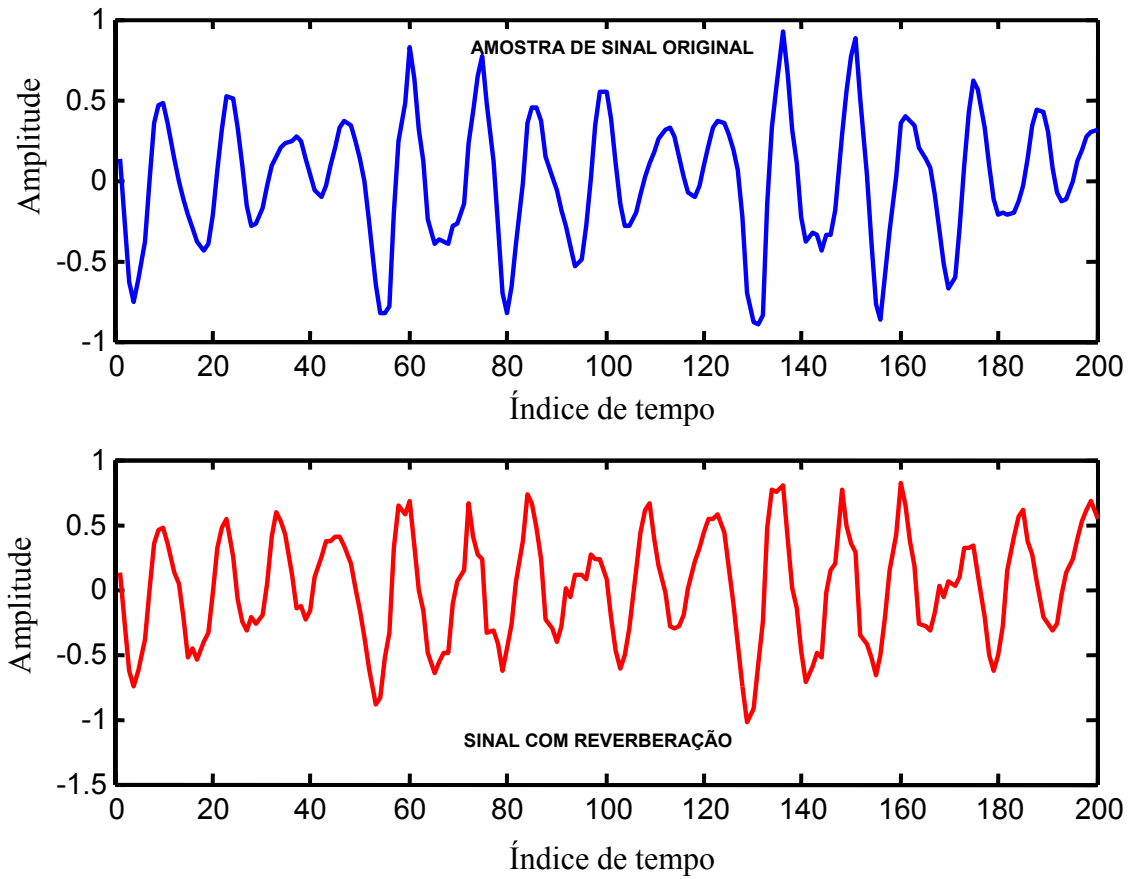


Figura 33 – Sinal de voz original e com efeito da reverberação para um filtro com 250 atrasos.

Em seguida temos o gráfico comparativo dos resultados das médias dos erros dos sinais de voz sem (em vermelho) e com o filtro (em azul), mostrado na Figura 34. A Figura 35 representa a média dos erros para um filtro com 250 atrasos.

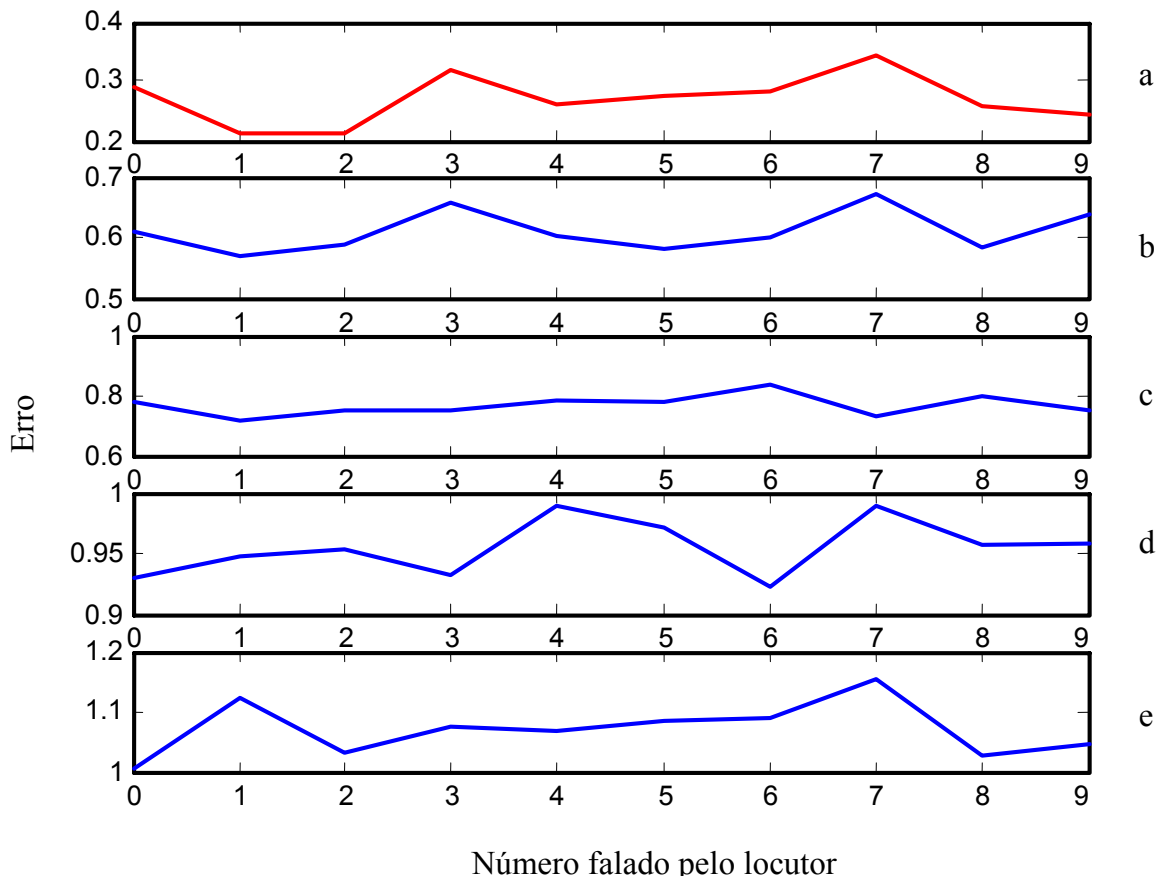


Figura 34 – a) Média dos erros para sinal sem reverberação, Média dos erros com reverberação: b) para filtro com 50 atrasos, c) 100 atrasos, d) 150 atrasos, e) 250 atrasos.

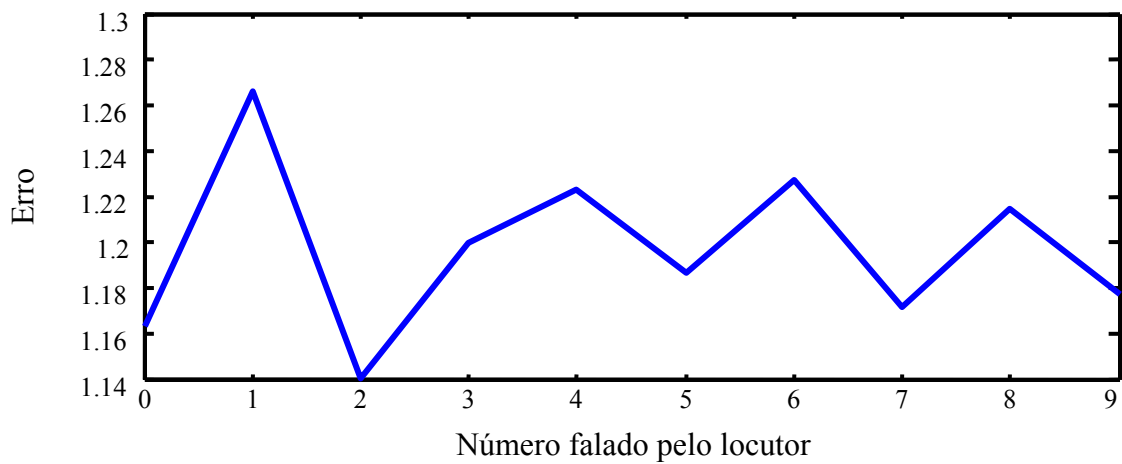


Figura 35 - Média dos erros para filtro com 250 atrasos.

6.2 Discussão

Fazer a extração de sinais de voz de um ambiente como o *Cocktail Party* utilizando-se apenas de um canal de entrada, o que dificulta ainda mais a recuperação do sinal de voz, não é uma tarefa muito fácil para as máquinas, diferentemente dos seres humanos, pois os sinais apresentam uma série de ruídos que comprometem a inteligibilidade daquele que se deseja extrair. Com o objetivo de eliminar tais ruídos recorreremos a Análise de Componentes Independentes (ICA) para derivar filtros ou funções bases que junto ao algoritmo de erro médio quadrático (MSE) funcionam como um filtro recursivo em que os pesos são atualizados e realimentados, conforme as iterações do código eficiente.

Utilizando-se de sinais provenientes do sub-espço de sinais de entrada como sinal desejado e adicionando a este diferentes tipos de ruídos nós checamos a validade desse método quanto a redução de ruído pela minimização do erro e conseqüente recuperação do sinal.

Pelo cálculo das médias dos erros dos sinais, representado pela linha em azul, respectivamente para ICA e PCA obtivemos na Simulação 1 que PCA soma pouco benefício na recuperação da voz. Através da Figura 20 podemos observar que a recuperação por PCA não é muito afetada pelo aumento do número de funções bases. Pelo contrário, para alguns números o erro é realmente maior com o número do aumento das bases. Por outro lado, ICA parece ser o mais apropriado, pois podemos ver facilmente que há uma queda exponencial quando o número de funções bases é aumentado, bem como para diferentes tipos de ruídos adicionados ao sinal desejado conforme Figuras 21 e 22, onde a média para ICA mostra-se melhor a cada iteração do processo, ou seja, com o aumento do número de funções bases há um melhor aprendizado e conseqüentemente a diminuição do erro e redução de ruído, o que comprova a eficiência do código frente a interferências indesejáveis (diferentes tipos de ruídos), no caso aqui musical e vocalização de animal.

Através da simulação 3 recuperamos um sinal de voz através da utilização das funções bases de outro sinal. O que deixa claro que o sistema utiliza-se das características que compõe a voz do orador (componentes independentes) e dentro do sub-espço de palavras proferidas será possível à recuperação de qualquer sinal de voz com as funções bases de outro, pois o erro mostrou-se pequeno, Figura 23.

Na simulação 4 verificamos a sensibilidade do método quanto a reverberação e foi possível observar que o erro aumenta com o comprimento do filtro.

CAPÍTULO 7

CONCLUSÃO

Há alguns pontos interessantes que podem ser destacados neste trabalho. Tem uma base biológica forte que é a codificação eficiente baseada na forma como o córtex auditivo codifica as informações, levando em consideração a capacidade que as máquinas têm de aprender através de processos repetitivos e melhorar seu desempenho. Essa é uma vantagem refletida em nossas simulações. Segundo, possui uma base estatística comprovada através da literatura da análise de componentes independentes, ICA. Mostramos um sistema baseado em codificação eficiente que codifica e faz a redução de ruído para a recuperação de um sinal desejado, evidenciando tanto a robustez do método a ruídos aditivos quanto a efeitos de reverberação que ocorrem usualmente em ambiente real.

Para continuidade e aperfeiçoamento desta linha de pesquisa sugerimos:

- Testar a sensibilidade do sistema para outros tipos de ruídos, como, por exemplo, ruído gaussiano para verificar o comportamento do algoritmo de ICA;
- Realizar experimentos em ambientes diversos com produção de diferentes interferências e avaliar o desempenho de resultados.
- Aplicar este modelo para reconhecimento de voz;
- Usar este trabalho como preliminar, para aplicações em sub-áreas de Processamento de Voz, utilizando as funções bases obtidas por ICA.

BIBLIOGRAFIA

- [1] J. F. Valiati, Reconhecimento de voz para comandos de direcionamento por meio de redes neurais, Porto Alegre: CPGCC da UFRGS, 1996.
- [2] I. H. Witten, Principles of computer speech, New York, Academic Press, 1982.
- [3] B. Arons, A review of the cocktail party, Cambridge, MA: MIT Laboratory, 1990.
- [4] T. W. Parsons, Separation of speech from interfering speech by means of harmonic selection, Journal of the Acoustical Society of America, 60, 911-918, 1976.
- [5] A. K. Barros, T. Rutkowski, F. Itakura, N. Ohnishi, Estimation of speech embedded in a reverberant and noisy environment by independent component analysis and wavelets, IEEE Trans. on Neural Networks, Vol. 13, No. 4, 888-893, 2002.
- [6] D. L. Wang and G. J. Brown, Separation of speech from interfering sounds based on oscillatory correlation, IEEE Trans. Neural Network, Vol. 10, 684-697, 1999.
- [7] G. Hu and D. Wang, Monaural speech segregation based on pitch tracking and amplitude modulation, IEEE Trans. on Signal Processing, 2002.
- [8] N. Virag, Single channel speech enhancement based on masking properties of the human auditory system, IEEE Trans. on Signal Processing, Vol 7, Nº. 2, 126-137, 1999.
- [9] R. Martin, Noise power spectral density estimation based on optimal smoothing and minimum statistics, IEEE Trans. Speech and Audio Processing, Vol 9, 504-512, 2001.
- [10] J. Xiaoping, F. Hua and Y. Tianren, A single channel speech enhancement method based on masking properties and minimum statistics, IEEE Trans. 2002.
- [11] M. S. Lewicki, Efficient coding of natural sounds, Nature Neuroscience 5(4):356-363, 2002.
- [12] E. R. Kandel, J. H. Schwartz, T. M. Jessel, Principles of neural science, 3. ed. USA: Prentice-Hall, 1991.
- [13] <http://www.epub.org.br/cm/n05/tecnologia/nervoso.htm>. Como funciona o sistema nervoso, Acessado em 17/08/03.

- [14] A. P. Braga, T. B. Ludemir, and A.C.P.L.F. Carvalho, *Redes neurais artificiais teoria e aplicações*, LTC, 2000.
- [15] B. Kröse and P. Smagt, *An introduction to neural networks*, University of Amsterdam, 1996.
- [16] G. M. Rahim, *Artificial neural network for speech analysis / synthesis*, Cambridge University, Cambridge, UK, 1994.
- [17] http://www.corpohumano.hpg.ig.com.br/sistema_nervoso/cerebro/cerebro_2.html. Corpo humano. Acessado em 14/11/03.
- [18] <http://www.epub.org.br/cm/n05/tecnologia/rna.htm>. O que são as redes neurais artificiais, Acessado em 17/08/03.
- [19] <http://www.furb.rct-sc.br/malcon>. Redes neurais artificiais: aprendizado e plasticidade, Acessado em 07/03/03.
- [20] B. C. J. Moore, *An introduction to the psychology of hearing*, Fourth edition, 2002.
- [21] <http://geocities.yahoo.com.br/saladefisica5/leituras/ouvido.htm>. Ouvido. Acessado em 14/11/03.
- [22] B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive-field properties by learning a sparse code for natural images, *Nature* 381, 607 – 609, 1996.
- [23] A. S. Bregman, *Auditory scene analysis*, Cambridge, MA: MIT press, 1990.
- [24] E. C. Cherry, Some experiments on the recognition of speech, with one and two ears, *Journal of the Acoustical Society of America*, 25, 975-979. 1953.
- [25] <http://www.universitario.net/ilevm/reverber.htm>, Reverberação, Acessado em 02/02/03.
- [26] C. Jutten and J. Herault, Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture, *Signal Processing*, 24:1-10, 1991.
- [27] V. Capdevielle, Ch. Serviere, and J. Lacoume, Blind separation of wide-band sources in the frequency domain, In *Proc. ICASSP-95*, volume 3, pages 2080-2083, Detroit, Michigan, USA, May 9-12 1995.
- [28] L. Tong, R. -W. Liu, V.C. Soon, and Y.-F. Huang, Indeterminacy and identifiability of blind identification, *IEEE Trans. on Circuits and Systems*, 38, 1991.
- [29] J. F. Cardoso, *Blind signal separation: statistical principles*, *Proceedings of the IEEE*, vol.83, 1998.

- [30] A. Mansour, A. K. Barros, and N. Ohnishi, Blind separation of sources: methods, assumptions and applications, IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences, vol. E83-A, 1498-1511, 2000.
- [31] A. J. Bell, and T.J. Sejnowski, An information maximization approach to blind separation and blind deconvolution, Neural Comp. 7, 1129-1159, 1995.
- [32] H. H. Harman. Modern factor analysis, University of Chicago Press, 2nd edition, 1967.
- [33] M. Kendall. Multivariate analysis, Charles Griffin & Co., 1975.
- [34] <http://www.din.uem.br/ia/pca/resumo.htm>; Análise de Componentes Principais (PCA), Acessado em 18/03/03.
- [35] <http://www-karc.crl.go.jp/avis/peper/pca.html>, What is Principal Component Analysis, Acessado em 2/02/2003.
- [36] S. Haykin, Redes neurais: princípios e prática, 2^a edição, 2001.
- [37] S. Amari and A. Cichocki, A new learning algorithm for blind signal separation, Lab. For Information Representation FRP, 1997.
- [38] A. J. Bell and T. J. Sejnowski, An information-maximization approach to blind source separation and blind deconvolution, Computational Neurobiology Laboratory, The Salk Institute, 1996.
- [39] J. Karhunen, Neural approaches to independent component analysis and source separation, Helsinki University of Technology, Laboratory of Computer and Information Science, 1996.
- [40] A. Hyvarinen, and E. Oja, Survey on independent component analysis, 1999.
- [41] P. Comon, Independent component analysis – a new concept? Signal Processing, 36:287-314, 1994.
- [42] N. Delfosse and P. Loubaton, adaptive blind separation of independent sources: a deflation approach, Signal Processing, vol. 45, 59 - 83, 1995.
- [43] A. Hyvarinen and E. Oja, A fast fixed-point algorithm for independent component analysis, Neural Computation (9), 1483 - 1492, 1997.
- [44] J. F. Cardoso and A. Souloumiac, Blind beamforming for non gaussian signals, IEEE Proceedings F, 140(6): 362-370, 1993

- [45] A. K. Barros and A. Cichocki, extraction of specific signals with temporal structure, *Neural Computation*, Vol. 13, No. 9, 1995-2004, 2001.
- [46] A. Belouchrani, K. Abeb Meraim, J.-F. Cardoso, and E. Moulines, A blind source separation technique based on second order statistics, *IEEE Trans. on S.P.*, 45(2): 434-44, 1997.
- [47] R. Vigário, Extraction of ocular artifacts from eeg using independent component analysis, *Electroenceph. Clin. Neurophysiol.*, 103 (3): 395-404, 1997.
- [48] R. Vigário. V. Jousmäki, M. Hämäläinen, R. Hari, and E. Oja, Independent component analysis for identification of artifacts in magnetoencephalographic recordings, In *Advances in Neural Information Processing 10 (Proc. NIPS'97)*, 229-235, Cambridge, MA, MIT Press, 1998.
- [49] S. Makeig, A. J. Bell, T.-P. Jung, and T.-J. Sejnowski, Independent component analysis of electroencephalographic data, In *Advances in Neural Information Processing Systems 8*, 145-151. MIT Press, 1996.
- [50] K. Kiviluoto and E. Oja, Independent component analysis for parallel financial time series, In *Proc. ICONIP'98*, Vol. 2, 895-898, Tokyo, Japan, 1998.
- [51] T. Ristaniemi and J. Joutsensalo, On the performance of blind source separation in cdma downlink, In *Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, 437-441, Aussois, France, 1999.
- [52] A. J. Bell and T.J. Sejnowski. The independent component of natural scenes are edge filters, *Vision Research*, 37:3327-338, 1997.
- [53] C. Loesch, e S. Sari,. *Redes neurais artificiais: fundamentos e controle*, Blumenau: FURB, 1996.
- [54] C. Schauf, D. Moffet e S. Moffet, *Fisiologia humana*, Rio de Janeiro: Guanabara, 1993.
- [55] H. B. Barlow and W. Rosenblith, Possible principles underlying the transformations of sensory messages, *Sensory Communication*, 217-234. MIT Press, Cambridge, MA, 1961.
- [56] J. Karhunen, R. Vigário, J. Hurri. And E. Oja, *Applications of neural blind separation to signal and image processing*, Helsinki University of Tecnology, Laboratory of Computer and Information Science, 1997.

- [57] M. Aoki, M. Okamoto, S. Aoki, H. Matsui, T. Sakurai and Y. Kaneda, Sound source segregation based on estimating incident angle of each frequency component of input acquired by multiple microphones, 2000.
- [58] M. Tafner, M. Xerez e I. Rodrigues, Redes neurais artificiais: introdução e princípios de neurocomputação, Blumenau: EKO, 1995.
- [59] N. Paraga and J-P Nadal, Blind source separation with time dependent mixtures, 1996.
- [60] P. Foster. T. Schalk, Speech recognition the complete practical reference guide, New York, Telecom Library,1993.
- [61] S. Amari, A. Cichocki and H.H.Yang, A new learning algorithm for blind signal separation, Advances in Neural Information Processing Systems 8, MIT press, 1996
- [62] S. F. Boll, Suppression of acoustic noise in speech using spectral subtraction, IEEE Trans. Acoustic. Speech, Signal Process. ASSP-27, 113-120, 1979.
- [63] T. W. Lee, Independent component analysis: theory and applications, Kluwer Academic Publishers. 1998.
- [64] T. W. Lee, Introduction to independent component analysis, Computational Neurobiology Laboratory, The Salk Institute, 1998.
- [65] www.hyperphysics.phy-astr.gsu.edu/hbase/acoustic/reverb.html. Acessado em 10/10/02.
- [66] S. Handel, Listening: An introduction to the perception of auditory events, MIT Press, 1989.
- [67] I.T. Jolliffe, Principal component analysis springer verlag, ISBN 0-387-96269-7 or ISBN 3-540-96269-7,1986.