

ANÁLISE DA RESPOSTA DE REDES NEURAIS DO TIPO MLP PARA RECONHECIMENTO DE LOCUTOR CONSIDERANDO DISTINTAS FERRAMENTAS DE PRÉ-PROCESSAMENTO

RODRIGO T. CAROPRESO¹, LUISA B. LIBONI², RICARDO A. S. FERNANDES³, IVAN N. SILVA⁴

¹*Departamento de Desenvolvimento Eletrônico - Smar Equipamentos Industriais
Rua Dr. Antônio Furlan Jr, 1028 – Centro - Sertãozinho – CEP 14170-480.*

²*Instituto Federal de Educação, Ciência e Tecnologia de São Paulo
Rua Américo Ambrósio, 269 – Jardim Canaã – Sertãozinho – SP – CEP 14169-263.*

³*Departamento de Engenharia Elétrica - CCET - Universidade Federal de São Carlos
Rod. Washington Luís, km 235 - SP-310, CEP 13565-905, São Carlos – SP*

⁴*Departamento de Engenharia Elétrica - EESC - Universidade de São Paulo
Av. Trabalhador São-carlense, 400 - CEP 13566-590, São Carlos – SP*

E-mails: rcaropreso@smar.com.br, luisa.liboni@gmail.com, ricardo.asf@ufscar.br, insilva@sc.usp.br

Abstract— Pattern classification is a very well-known and researched problem through many perspectives. Usually, the speaker recognition still stands as a field which a great potential for optimization problem. Based on this premise, this paper proposes the use of signal processing tools altogether with a Multi Layer Perceptron on the speaker recognition task. The results have shown relevant results, assuring the proposed method as a very good approach on this particular problem.

Keywords— Artificial Neural Networks, Speaker Recognition, Cepstrum, Perceptron.

Resumo— O problema de classificação de padrões é bastante estudado e abordado de diversas formas. Particularmente, o reconhecimento de fala mostra-se como uma área onde há ainda muito a ser explorado. Desta forma, este artigo propõe a utilização de ferramentas destinadas ao pré-processamento de sinais junto a Redes Neurais Artificiais do tipo MLP para o reconhecimento de locutor. Com base nos treinamentos e validações das redes neurais, foi possível notar o satisfatório desempenho alcançado pela metodologia proposta.

Palavras-chave— Redes neurais artificiais, identificação de locutor, pré-processamento de sinais.

1 - Introdução

O reconhecimento da voz humana e de um interlocutor tem sido objeto de estudo cada vez mais frequente na atualidade e, desta forma, o Reconhecimento Automático de Locutores (RAL) tem sido vastamente pesquisado pela comunidade científica (Hafen e Henry, 2012), (Sahidullah e Saha, 2012), (Sadiç e Bilginer, 2011), (Ajmera e Holambre, 2010).

Partindo-se do resultado destas pesquisas, tornou-se possível a elaboração de procedimentos e algoritmos capazes de processar e executar o RAL. Entretanto, apesar dos avanços obtidos pelas pesquisas, ainda não há uma metodologia capaz de compreender toda a fala de um ser humano.

De forma sumária, pode-se definir o RAL como sendo a capacidade de reconhecer uma determinada pessoa por meio de sua voz, sendo esta uma variação do Reconhecimento de Voz.

As técnicas mais conhecidas na literatura para o problema do RAL encontram-se divididas em: técnicas estatísticas e técnicas determinísticas.

Assim, nas técnicas estatísticas, o reconhecimento se dá pela medida da probabilidade condicional ao se comparar os padrões durante a observação do

modelo, destacando-se os Modelos Ocultos de Markov (Gajšek et al, 2012) e os Modelos de Mistura Gaussiana (Sadiç e Bilginer, 2011).

Já nas técnicas determinísticas, é feito um processo de alinhamento para se calcular a distância entre o sinal e o modelo, onde assume-se que o padrão a ser reconhecimento é uma réplica do modelo adotado. Dentre as técnicas determinísticas pode-se destacar a Quantização Vetorial (Kumar e Rao, 2011) e os Classificadores Polinomiais (Ajmera e Holambre, 2010).

Por outro lado, diversas pesquisas têm agregado sistemas inteligentes ao RAL, (Cumani e Laface, 2012, Casagrande, 1997, Khaled, 2011, Swamy et al, 2012, Ajmera e Homalbre, 2010, Jian, Lin e Xiaouying, 2010)

Dentre as pesquisas baseadas em sistemas inteligentes, as realizadas por (Casagrande, 1997), (Khaled, 2011), e (Swamy et al, 2012) empregam Redes Neurais Artificiais (RNA) para a realização do RAL.

Nestes métodos, o papel da RNA consiste em fazer a classificação dos conjuntos de características de cada locutor, as quais são extraídas durante uma fase de pré-processamento do sinal, permitindo assim, sua identificação na fase de operação do sistema.

Seguindo o contexto supracitado, este trabalho propõe uma análise de desempenho de RNAs no processo do RAL, onde combinou-se um pré-processamento de sinal baseado na extração dos Coeficientes Mel-Cepstrais a uma RNA do tipo *Perceptron* de Múltiplas Camadas como classificador.

Deste modo, na Seção 2 será apresentada uma descrição do modelo do sistema fonador adotado na literatura correlata. Na Seção 3, será descrito o pré-processamento do sinal de voz utilizado. A Seção 4 apresenta a arquitetura do RAL baseado em RNA. Na Seção 5, são apresentados os resultados obtidos pela metodologia proposta. Por fim, na Seção 6, são feitas as considerações finais e conclusões do trabalho.

2 – Modelagem do Sistema Fonador

É importante destacar que o problema de reconhecimento de fala é de difícil tratabilidade, pois, há uma grande variabilidade acústica entre distintos locutores. Além disso, também há uma notória diferença de resposta entre os dispositivos transdutores.

Assim, em (Rabiner, 1978) foi proposta uma função de transferência que pudesse modelar um sistema fonador. Esta função de transferência relaciona uma entrada que representa a excitação das cordas vocais com uma saída que representa a produção de voz:

$$F(z) = \frac{G}{(1 - \sum_{k=1}^N \alpha_k \cdot z^{-k})} \quad (1)$$

onde α_k representa os parâmetros dependentes das características e as dimensões do trato vocal; e G representa o ganho do sistema.

O modelo do trato vocal representado pela Equação 1 consiste numa simplificação, pois, somente a informação de amplitude da densidade espectral de potência é utilizada. Entretanto, para grande parte das aplicações esta modelagem é considerada aceitável.

Por ser um sinal de variação temporal lenta, quando examinado em períodos de tempo pequenos (entre 5 ms e 100 ms, suas características podem ser consideradas estacionárias. Destaca-se que nas aplicações práticas é comum considerar intervalos inferiores a 10 ms).

3 – O pré-processamento

Destaca-se que a base de dados de voz utilizada neste trabalho foi a Speaker Recognition v1.0, desenvolvida pelo *Center of Spoken Language Understanding (CSLU)* do *Oregon Graduate Institute (OGI)*, a qual consiste na gravação de amostras de voz de 90 participantes ao telefone. Cada um dos participantes gravou sua voz em doze sessões ao longo de um período de vários meses.

Os sinais de voz foram amostrados com frequência de amostragem de 8 kHz e codificados em PCM 8 bits.

Partindo-se desta base de dados, desenvolveu a etapa de pré-processamento dos sinais, a qual pode ser melhor compreendida por meio da Figura 1.

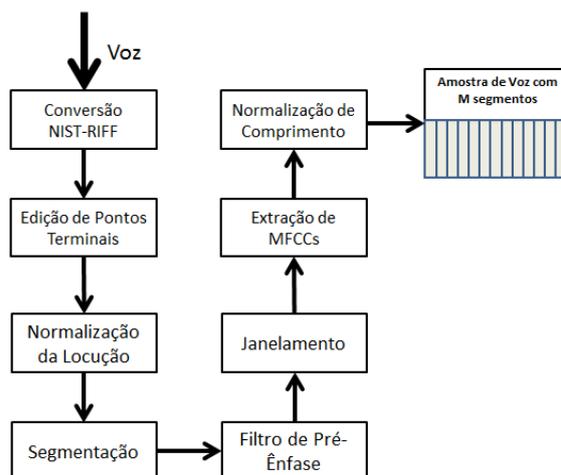


Figura 1 – Etapa de pré-processamento dos sinais.

Inicialmente, os sinais de voz foram convertidos do formato NIST para RIFF. Na sequência, houve a edição dos pontos terminais com o intuito de minimizar trechos de silêncio e ruído excessivo no início e final de cada locução.

A partir deste ponto, empregou-se uma normalização aos dados com o objetivo de minimizar a possibilidade de interferências no sistema de reconhecimento, as quais poderiam ser causadas por grandes diferenças de energia entre as amostras.

Desta forma, as locuções foram segmentadas por meio de conjuntos de 512 pontos (duração de 64 ms) com 50% de sobreposição ao passo que a janela deslizava pelas amostras. Após esta segmentação, as locuções foram submetidas a um filtro de pré-ênfase cuja função de transferência é dada pela Equação (2):

$$H(z) = 1 - 0.95z^{-1} \quad (2)$$

Assim, realizou-se um janelamento dos segmentos, onde cada um deles passa por uma janela de Hamming, dada pela Equação (3):

$$W_{HAMMING}(n) = 0.54 - 0.46\cos(2\pi \frac{n}{LJ}) \quad (3)$$

onde LJ é a largura da janela utilizada (512 pontos).

A partir das janelas de *Hamming*, são obtidos os coeficientes Mel-Cepstrais, os quais foram empregados devido aos bons resultados observados por meio da literatura correlata (Sahidullah e Goutam, 2012), (Nakagawa et al, 2012) e que são devidamente apresentados na Subseção 3.1.

3.1 – Cálculo dos Coeficientes Mel-Cepstrais

Os coeficientes Mel-Cepstrais têm sido os mais utilizados na tarefa do RAL. A escala Mel foi apresentada pela primeira vez nos trabalhos de (Stevens, 1940) e pode ser visualizada por meio da Equação (4):

$$Mel = 2595 \log\left(\frac{f}{700}\right) \quad (4)$$

Por meio da Figura 2, é possível observar o processo de obtenção dos coeficientes Mel-Cepstrais.

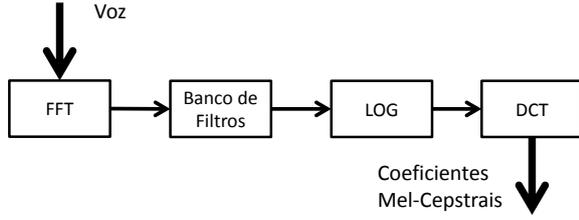


Figura 2 – Cálculo dos coeficientes mel-cepstrais.

Cabe ressaltar que os coeficientes Mel-Cepstrais são definidos como sendo a Transformada Discreta do Cosseno para a saída logarítmica de um banco de filtros triangulares distribuídos de acordo com a escala Mel. Deste modo, a equação (5) define tal conceito:

$$C_p(j) = \frac{2}{K} \sum_{k=1}^K \log(X_k(j)) \cdot \cos\left(p\left(k - 0.5\right) \frac{\pi}{K}\right) \quad (5)$$

onde $C_p(j)$ é o p -ésimo coeficiente Mel-Cepstral da janela j , K é o número de filtros e $X_k(j)$ é a energia do k -ésimo filtro calculada para a janela j .

É importante comentar que a escolha do número de filtros segue o trabalho de (Magni, 1998), onde são empregados 80 filtros com 24 Coeficientes Mel-Cepstrais. No entanto, o coeficiente Mel-Cepstral(0) foi desprezado, uma vez que sua informação (energia do segmento de amostras) influencia de maneira indesejável o sistema de reconhecimento.

A última etapa do pré-processamento consiste na normalização do comprimento das locuções, a fim de evitar amostras excessivamente longas para um dado locutor.

Como referência, foi utilizado o truncamento simples, que consiste em eliminar o excesso de termos do vetor de locuções até que atinja o valor desejado para ser usado como entrada da RNA.

A este procedimento de referência, foi realizada uma comparação com duas técnicas que são descritas na sequência.

3.2 – Trace Segmentation (TS)

O algoritmo conhecido como *Trace Segmentation* (TS) baseia-se na hipótese de que, apesar de diferenças temporais, para sinais de voz de uma mesma categoria, as flutuações no espectro de frequências ao longo do tempo irão ocorrer na mesma sequência, porém, em intervalos de tempo de comprimentos distintos.

Assim, considerando um sinal de voz u representado pela sequência de m vetores N -dimensionais ($v_1^u, v_2^u, \dots, v_m^u$), obtidos em intervalos de tempo uniformes, os quais podem ser vistos como pontos não uniformemente espaçados ao longo de uma curva em um espaço N -dimensional.

O processo de TS envolve a divisão do comprimento espacial total da trajetória L_t , em n_p segmentos de

igual comprimento, cuja distância euclidiana entre eles seja mantida aproximadamente uniforme. Assim, concatenação destes vetores pode ser usada como uma representação da trajetória original.

Para o cálculo dos vetores, inicialmente, deve-se calcular o comprimento espacial total da trajetória, L_t , através da soma das distâncias euclidianas dos m vetores sucessivos que compõe a sequência original, conforme a Equação (6):

$$L_t = \sum_{i=1}^{m-1} d(\vec{v}_i, \vec{v}_{i+1}) \quad (6)$$

onde,

$$d(\vec{v}_i, \vec{v}_{i+1}) = \left(\sum_{j=1}^N (v_{i,j}^u - v_{i+1,j}^u)^2 \right)^{1/2} \quad (7)$$

Para uma dada quantidade de *frames* n_p no vetor normalizado, deve-se calcular o comprimento de cada segmento espacial, conforme a Equação (8):

$$L_s = \frac{L_t}{n_p} \quad (8)$$

Desta maneira, para cada *frame* i do vetor de saída, deve-se localizar os dois *frames* da sequência original que recaem em cada lado do i -ésimo segmento. Isto pode ser obtido calculando-se o valor de j por meio das Inequações (9) e (10):

$$\sum_{k=1}^j d(v_k^u, v_{k+1}^u) > i \cdot L_s \quad (9)$$

$$\sum_{k=1}^{j-1} d(v_k^u, v_{k+1}^u) < i \cdot L_s \quad (10)$$

No caso de se utilizar a interpolação linear para definir o *frame* do vetor de saída, deve-se calcular o valor de t_i^u , conforme as Equações (11) e (12):

$$t_i^u = v_j^u + (v_{j+1}^u - v_j^u) \cdot \alpha \quad (11)$$

$$\alpha = \frac{i \cdot L_s - \sum_{k=1}^{j-1} d(v_k^u, v_{k+1}^u)}{d(v_j^u, v_{j+1}^u)} \quad (12)$$

3.3 – Linear Time Warping

Os algoritmos de alinhamento linear (*Linear Time Warping*) estão entre os algoritmos de implementação mais simples e podem ser usados tanto para expansão quanto compressão do vetor de padrões que representa a fala do locutor.

Assim, o método consiste em separar as linhas do vetor que compõe a locução em seções e calcular a média de cada seção de forma a produzir um único vetor de atributos por seção. Desta forma, o vetor final será composto sempre pelo mesmo número de linhas, porém a quantidade de vetores pertencentes a uma mesma seção será variado.

A quantidade de vetores contidos numa seção será dada, então, por:

$$Q = \frac{L_{locucao}}{L_{norm}} \quad (13)$$

onde $L_{locucao}$ é a quantidade de vetores de atributos que compõem a locução e L_{norm} é o tamanho da locução normalizada, ou seja, a quantidade de vetores de atributos que irá compor a locução “reduzida”.

A menos que a quantidade de linhas da locução seja um múltiplo de L_{norm} , o valor de Q não será um valor inteiro. Devido a isso, nem todas as linhas serão aproveitadas no *Linear Time Warping*.

3.4 – Minimal Temporal Information

Devido às locuções poderem apresentar comprimentos distintos, torna-se comum a normalização do tamanho das locuções, principalmente, quando se deseja reconhecer palavras. Entretanto, em problemas do tipo RAL, isto não é necessário conforme demonstrado por (Tmoszczuk, 1998), cuja proposta baseia-se no agrupamento dos segmentos das locuções em *Minimal Temporal Information* (MTI).

As MTIs são conjuntos de segmentos consecutivos extraídos a partir da sequência de segmentos que representam uma locução (com sobreposição ou não) conforme pode ser visualizado por meio da Figura 3.

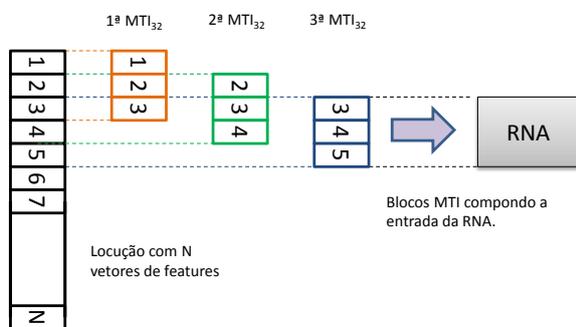


Figura 3 – Construção de MTIs.

Neste trabalho, empregou-se a MTI31 sem perda de generalidade, uma vez que o objetivo proposto é avaliar o desempenho da RNA na tarefa do RAL e o ônus computacional é menor (devido a um menor nível de interpolação entre os segmentos).

4 – Arquitetura do RAL Baseado em Redes Neurais Artificiais

O sistema computacional que realiza o RAL pode ser representado pelo seguinte diagrama de blocos.

No sistema da Figura 4, após a etapa de pré-processamento, faz-se necessário o uso de um mecanismo capaz de realizar a tarefa de classificação, a qual é representada pelos blocos “Armazenamento dos Padrões” e “Comparação de Padrões”. Estes blocos são representados pela Rede Neural Artificial (RNA).

Cabe comentar que as Redes Neurais Artificiais são modelos computacionais inspirados no cérebro humano capazes de adquirir e manter o conhecimento a

partir de exemplos de um determinado processo ou problema (Haykin, 2001).

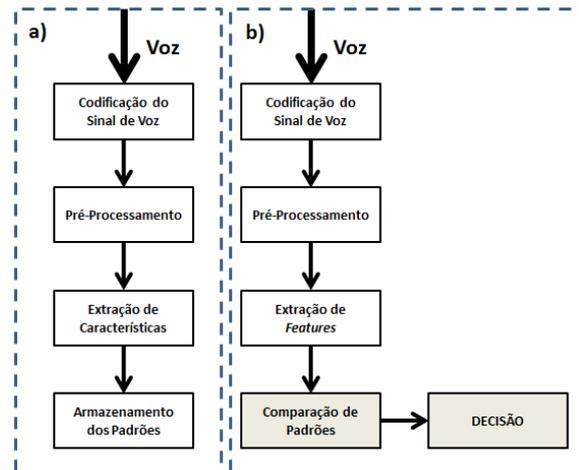


Figura 4 – Sistema de RAL: a) Fase de Preparação; b) Fase de Operação.

Neste trabalho foi utilizada uma rede do tipo Perceptron de Múltiplas Camadas, treinada com o algoritmo *Backpropagation*. Esta RNA recebe como entradas os conjuntos de MTIs gerados pela etapa de pré-processamento previamente descrita. A RNA também possui N saídas, onde cada uma representa um locutor (conforme Figura 5).

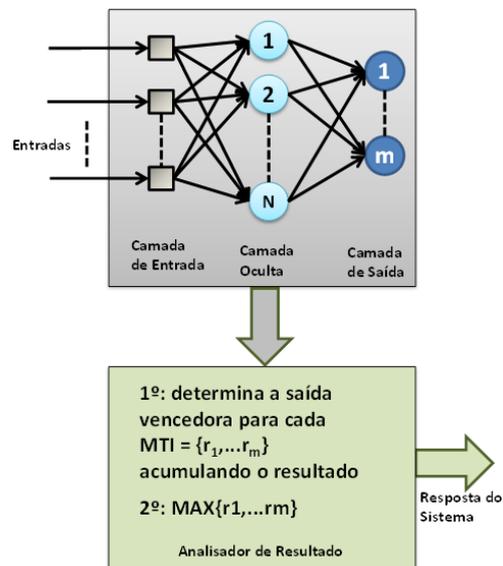


Figura 5 – Reconhecedor de locutor baseado em RNA.

Durante a etapa de validação da RNA, verifica-se qual das saídas apresentou o maior valor e, portanto, o locutor reconhecimento será o que representa esta saída. Destaca-se que esta abordagem é conhecida como *winner takes all*.

Como consequência do uso das MTIs, surge a necessidade de se criar um critério de decisão adicional, pois, para cada locução apresentada, será gerado um vetor com M saídas (uma saída para cada MTI). A fim de tornar o sistema mais flexível e permitir a identificação de falsos positivos, algumas informações devem ser acrescentadas ao 2º critério de decisão. Assim, idealizou-se a utilização de um *threshold* τ ou parâmetro de vigilância, de forma a

permitir o ajuste fino dos critérios de decisão do sistema.

Desta forma, caso o resultado obtido pela RNA apontar um conjunto significativo de MTIs para um dado Locutor N (ou seja, uma porcentagem em relação ao total de MTIs da locução superior a τ) e este conjunto for único, então a saída do sistema irá considerar o reconhecimento bem sucedido e o locutor identificado será o indivíduo N . Caso contrário, a saída do sistema será “Locutor Indefinido”.

5 – Resultados Obtidos

O sistema foi construído de forma a analisar 10 locutores escolhidos da base de dados descrita anteriormente, organizados da seguinte forma:

- 10 locuções (por locutor) foram separadas para o “treinamento” da RNA;
- 6 locuções (por locutor) foram separadas para validar o desempenho da RNA.

Em uma primeira análise, a RNA foi treinada considerando-se distintos valores para a quantidade de neurônios da camada oculta (35, 45, 55, 65 e 75 neurônios). Já a taxa de aprendizado foi mantida constante em $\eta=0.015$ para todos os treinamentos.

Cabe comentar que cada topologia de RNA previamente citada foi avaliada também em função da variação do *threshold* τ , na faixa de 0,1 a 1,0. No entanto, notou-se que valores baixos de τ podem conduzir a decisões errôneas (falsos positivos), uma vez que apenas uma pequena parte do sinal de voz acaba sendo associado a um locutor, conferindo baixa confiabilidade à RNA.

Por outro lado, valores excessivamente altos (próximos de 1,0) não necessariamente conduzem a uma taxa de acerto maior, pois, significa que a RNA deve classificar quase a totalidade da locução para um específico locutor.

Desta forma, verificou-se que valores adequados para τ estão entre 0,7 e 0,85, onde uma parcela suficientemente significativa da frase (mais de 70%) foi classificada pela RNA como sendo pertencente a apenas um locutor.

Obtidos os resultados para cada topologia de RNA, verificou-se que a melhor quantidade de neurônios na camada oculta é 55. Assim, por meio da Figura 6, são mostrados os resultados desta topologia de RNA em relação à variação do *threshold*.

Observa-se que se for empregado um $\tau = 0,7$, a taxa de acerto do sistema encontra-se em torno de 86%, ou seja, garante um resultado bastante robusto para o RAL.

Considerando-se a topologia da RNA com 55 neurônios na camada oculta, foram realizados treinamentos adicionais para se avaliar a melhor ferramenta de pré-processamento a ser empregada:

Trace Segmentation (TS), Truncamento Simples ou *Linear Time Warping* (LTW).

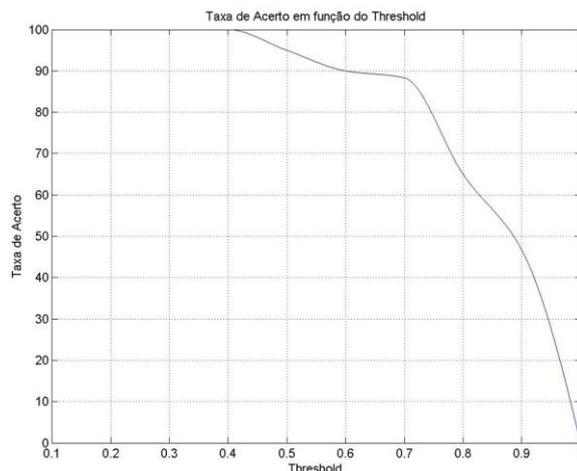


Figura 6 – Taxa de acerto da RNA com 55 neurônios na camada oculta.

Na sequência, verificou-se a resposta da RNA para cada ferramenta de pré-processamento e também para a variação do *threshold*. Assim, por meio das Figuras 7, 8 e 9, pode-se observar nitidamente que as técnicas de TS e LTW melhoram sensivelmente a eficiência do RAL, cuja taxa de acerto para fica acima de 90% com uma resposta mais plana para uma faixa maior, principalmente, para a técnica de TS.

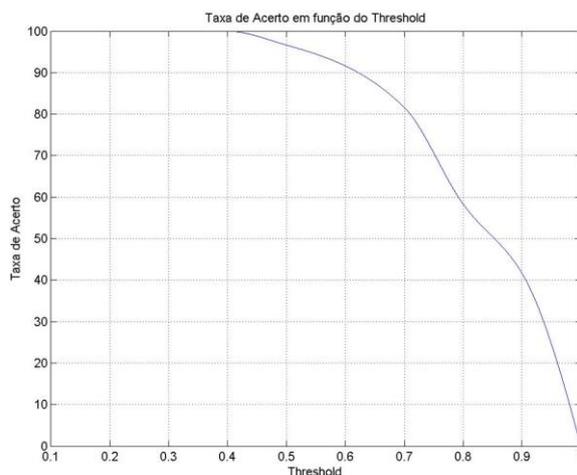


Figura 7 – Taxa de acerto utilizando Truncamento.

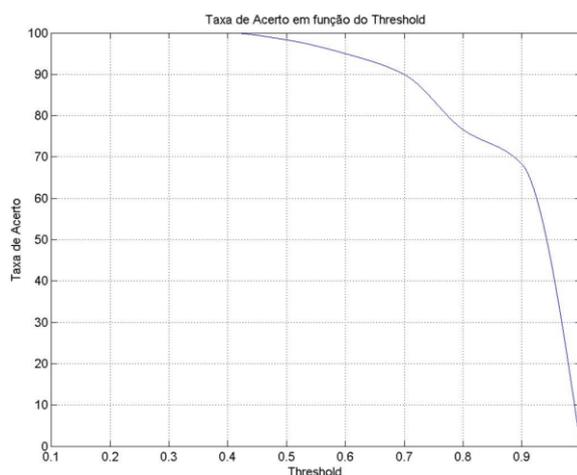


Figura 8 – Taxa de acerto utilizando Trace Segmentation.

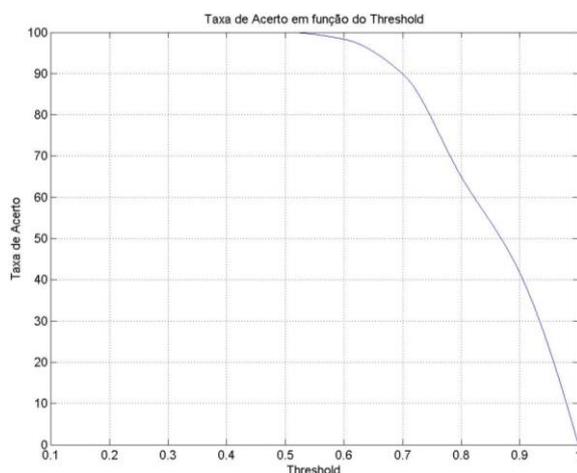


Figura 9 – Taxa de acerto utilizando *Linear Time Warping*.

6 – Conclusão

A partir dos testes realizados, fica clara a influência do pré-processamento no desempenho do RAL, com melhores resultados para *Trace Segmentation*, cuja taxa de acerto ficou superior à técnica de *Linear Time Warping*. A utilização do *threshold* de decisão τ facilitou as análises de resultados, uma vez que permitiu a visualização do comportamento de cada RNA ao longo de uma faixa de resultados, além de permitir um ajuste adicional para a taxa de acerto do sistema, sem que haja a necessidade de se recorrer a um novo treinamento. Como trabalhos futuros pode-se sugerir a análise de outras arquiteturas de RNA e outras ferramentas de pré-processamento.

Referências Bibliográficas

- AJMERA, P.K.; HOLAMBE, R.,S. **Speaker Recognition Using Auditory Features and Polynomial Classifier**. International Journal of Computer Applications, 2010, Vol.1(14), p.86
- CASAGRANDE, R. Redes neurais artificiais com retardos temporais aplicadas ao reconhecimento automático do locutor. São Paulo, 1997, 80p. Dissertação de Mestrado, EPUSP.
- CUMANI, S.; LAFACE, P. **Analysis of Large-Scale SVM Training Algorithms for Language and Speaker Recognition**. IEEE Transactions on Audio, Speech, and Language Processing, July 2012, Vol.20(5), pp.1585-1596
- GAJŠEK, R., MIHELIČ, F., DOBRIŠEK, S. **Speaker state recognition using an HMM-based feature extraction method**. Computer Speech & Language, 2012, Vol.27(1), pp.135-150
- HAFEN, R.; HENRY, M. **Speech information retrieval: a review**. Multimedia Systems, 2012, Vol.18(6), pp.499-518
- HAYKIN, S. **Redes Neurais – Princípios e prática**. 2. Ed. Porto Alegre, RS: Bookman, 2001
- JIAN, C.;LIN,L.; XIAOYING, S. Fuzzy kernel vector quantization with entropy and sectional set for speaker recognition under limited data. Computer Mechatronics, Control and Electronic Engineering (CMCE), 2010 International Conference Vol 3., pp.376-379.
- KHALED, D. **Wavelet entropy and neural network for text-independent speaker identification**. Engineering Applications of Artificial Intelligence, 2011, Vol.24(5), pp.796-802
- KUMAR, S., RAO, M. **Design Of An Automatic Speaker Recognition System Using MFCC, Vector Quantization and LBG Algorithm**. International Journal on Computer Science and Engineering, 2011, Vol.3(8), p.2942
- MAGNI, A.B. Reconhecimento automático do locutor com coeficientes Mel-Cepstrais e redes neurais artificiais. São Paulo, 138p., Dissertação de Mestrado, EPUSP, 1998.
- NAKAGAWA, S, WANG, L. OHTSUKA, S. **Speaker Identification and Verification by Combining MFCC and Phase Information**. IEEE Transactions on Audio, Speech, and Language Processing, May 2012, Vol.20(4), pp.1085-1095
- RABINER, L.R.; SCHAFER, R.W. Digital processing of speech signals. Englewood Cliffs, Prentice Hall, 1978.
- SADIÇ, S.; BILGINER, G. M. **Common vector approach and its combination with GMM for text-independent speaker recognition**. Expert systems with applications [0957-4174], ano:2011 vol:38 fasc:9 pág:11394 -11400
- SAHIDULLAH, M.; SAHA, G. **Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition**. Speech Communication, May, 2012, Vol.54(4), p.543(23)
- STEVENS, S.S.; VOLKMAN, J. The relation of pitch to frequency. American Journal of Psychology, vol.53, p.329, 1940. citado em DELL93.
- SWAMY, S., SHALINI, T., NAGABHUSHAN, S. P., NAWAZ, S., & RAMAKRISHNAN, K. V. (2012). **Text Dependent Speaker Identification and Speech Recognition Using Artificial Neural Network**. In Global Trends in Computing and Communication Systems (pp. 160-168). Springer Berlin Heidelberg
- TMOSZCZUK, A.P. Reconhecimento automático do locutor com redes neurais artificiais do tipo radial basis function (RBF) e Minimal Temporal Information (MTI). São Paulo, Dissertação de Mestrado, 135p. EPUSP, 1998.