ROBUSTEZ AO RUÍDO: ANÁLISE DE DESEMPENHO ENTRE PARÂMETROS RASTA E OUTROS PARÂMETROS USUAIS EM SISTEMAS DE RECONHECIMENTO AUTOMÁTICO DA FALA

Rubem Dutra Ribeiro Fagundes[1] e César Boesche[2]

***Resumo*— A robustez ao ruído em sistemas de reconhecimento automático de fala (SRAF) tem sido objeto de estudo de inúmeros trabalhos científicos, pois representa um grande desafio tornar estes sistemas utilizáveis nos ambientes onde irão operar. A maior parte destes trabalhos aborda técnicas que tornam o sistema robusto ao ruído através da adaptação de modelos estatísticos a um novo ambiente [20], que não aquele no qual o sistema foi treinado. Já o presente trabalho, aborda técnicas que atuam diretamente na etapa inicial do sistema, no *front-end*. Esta etapa é responsável por extrair parâmetros do sinal de voz, que posteriormente são submetidos à etapa de classificação de padrões. As técnicas PLP (Perceptual Linear Prediction) [1] e RASTA (RelAtive SpecTrAl) [9, 10], objetos de estudo deste trabalho, tentam imitar certas características desejáveis do sistema auditivo humano, buscando com isso, a extração de parâmetros que dão mais robustez ao sistema.**

***Palavras-Chave*—Reconhecimento Automático de fala com ruído, Robustez ao Ruído, Processamento de fala, RASTA, PLP, Processamento Digital de Sinais, Interface Vocal Homem-Máquina.**

Abstract— **The robustness in SRAF has been the subject of several scientific works, because it represents a big obstacle to the use of this kind of systems, considering the environments where they will work . Most of these scientific works rely mainly on stochastic methods, which try to adapt the statistical models to another environment [20], different from that where the system was trained. This research will show some techniques that work directly on the SRAF front-end. This part of the system is responsible for extracting parameters from the speech signal, which will be the inputs to the next part, the pattern classification process. The PLP (Perceptual Linear Prediction) [1] and RASTA (RelAtive SpecTrAl) [9, 10] techniques, subjects of this work, try to mimic some desirable characteristics of the human auditory system, in order to get parameters that give more robustness to the system. The advantages of these techniques over the others become evident after the analysis of the comparative results.**

***Keywords*—Automatic Speech Recognition System with Noise, ASR, Noise Robustness, Speech processing, RASTA, PLP, Digital Signal Processing, DSP, Vocal Man-Machine Interface.**

I. INTRODUÇÃO

O tema central do presente trabalho é a imunidade ao ruído em SRAF. É sabido que um SRAF sofre uma degradação considerável em seu desempenho na presença de ruído. Surgiu, então, há algumas décadas atrás, uma área de pesquisa dedicada exclusivamente em lidar com esta falta de robustez ao ruído. Várias propostas de melhorias foram apresentadas desde então, dentre elas as que modelam a audição humana, as quais mostraram-se bastante promissoras nesta tarefa, e hoje são adotadas quase como um padrão na etapa de extração de parâmetros dos SRAF.

II. SISTEMAS DE RECONHECIMENTO AUTOMÁTICO DE FALA (SRAF)

Um SRAF tem como objetivo converter um sinal de voz em palavras. É composto de etapas bem distintas, cada uma delas executando tarefas específicas, tais como: análise do sinal de voz (extração de parâmetros), classificação de padrões e decodificação utilizando modelos lingüísticos. Desde o surgimento destes sistemas, cada uma destas etapas evoluiu de forma a torná-los menos sensíveis ao ruído e às variações na fala. Surgiu, então, o conceito de robustez ao ruído. Um SRAF é dito robusto quando seu desempenho degrada pouco em ambientes com ruído.

III. RECONHECIMENTO DE FALA EM AMBIENTES COM RUÍDO

Os SRAF projetados para operar em ambientes controlados, treinados a partir de amostras de voz livres de ruído, têm seus desempenhos prejudicados quando passam a operar em ambientes com ruído. Por exemplo, um sistema de reconhecimento de palavras isoladas treinado com amostras de voz reais apresenta 100% de precisão com amostras de voz livres de ruído, podendo este desempenho cair para 30% quando este for utilizado no interior de um carro viajando a 90km/h [**5**].

A. Técnicas de compensação de ruído

1) Técnica de análise PLP

A técnica PLP (*Perceptual Linear Prediction*) [**1**, **4**] é bastante utilizada em reconhecimento automático de fala. Nesta técnica utiliza-se um banco de filtros que tenta reproduzir alguns aspectos das respostas cocleares, tal como a assimetria das respostas em relação à freqüência central. Os filtros são definidos com a forma descrita pela Equação 1, sendo esta uma aproximação da curva de mascaramento assimétrica de Schroeder [**16**]. Esta técnica, à semelhança da técnica MFCC, presta-se apenas a uma análise em freqüência e não a uma análise temporal.

[1] Professor Adjunto do Departamento de Engenharia Elétrica, Faculdade de Engenharia da PUCRS. E-mail: rubemdrf@pucrs.br

[2] Mestre em Engenharia Elétrica pela PUCRS. E-mail: cesar.boesche@gmail.com.br

$$\Psi(\Omega)=\begin{cases}0 & para\ \ \Omega<-1.3\\ 10^{2.5(\Omega+0.5)} & para\ -1.3\le\Omega\le-0.5\\ 1 & para\ -0.5\le\Omega\ \le\ 0.5\\ 10^{-1.0(\Omega-0.5)} & para\ \ 0.5\le\Omega\le 2.5\\ 0 & para\ \ \Omega>2.5\end{cases} \quad \textbf{(1)}$$

onde ψ(Ω) é a funcão que simula as bandas-críticas e Ω representa a freqüência na escala *Bark*. As freqüências centrais destes filtros são espaçadas de aproximadamente uma banda crítica (1 Bark), o que corresponde a utilizar 18 filtros para cobrir uma gama de freqüências entre 0 e 16.9 Bark (0 e 5 kHz). Nesta técnica são ainda incorporados dois conceitos psicofísicos que são: a variação da sensação de intensidade sonora subjetiva ("loudness") em função da freqüência e em função da intensidade (ver Figura 1). Da curva de igual intensidade subjetiva a 40dB deriva-se um fator de pré-ênfase associado aos filtros e que tem como efeito a redução das respostas a baixas freqüências (ver Figura 2). À energia dos sinais à saída dos filtros é aplicada uma raiz cúbica para simular a relação não linear entre a intensidade do som e sua percepção subjetiva. Com esta técnica obtém-se uma representação espectral de tempo curto compacta referida como espectro de potência de bandas-críticas.

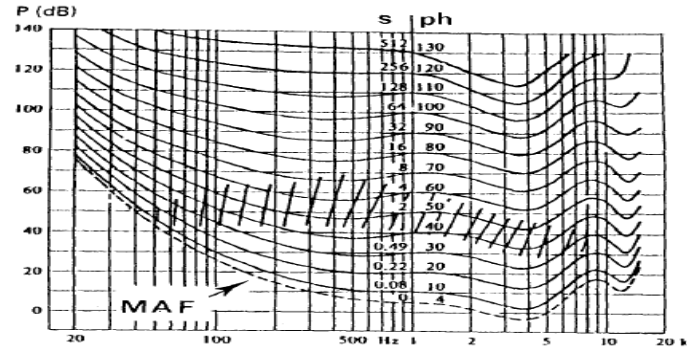

Figura 1 - Curvas de audibilidade humana e de igual intensidade subjetiva (nível de pressão sonora, P, em função da freqüência). [**3**].

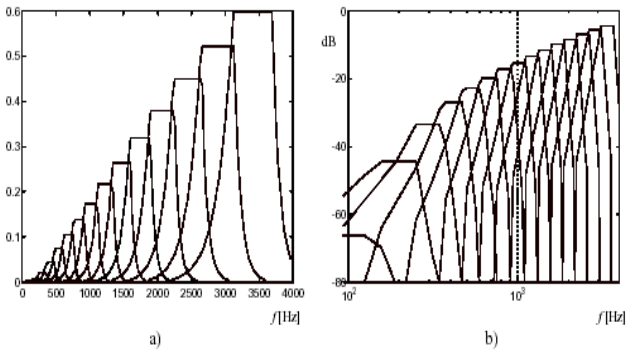

Figura 2 - Resposta em freqüência de 15 filtros referentes à modelação PLP para uma freqüência de amostragem de 8kHz. Os filtros já têm incorporado a função de pré-ênfase de igual "loudness". a) Escala linear. b) Escala logarítmica.

A técnica de análise PLP compreende as seguintes etapas:

1. Inicialmente, o sinal de voz é filtrado por um filtro passa-baixa *anti-aliasing* e logo depois amostrado;
2. O sinal é segmentado em *frames*, que contêm amostras que são ponderadas através da aplicação de uma janela, usualmente do tipo Hamming;
3. A FFT é aplicada em cada *frame*, resultando o espectro de potência de tempo-curto do sinal de voz ;
4. O espectro de potência de tempo-curto é transportado para a escala de freqüência Bark. Esta escala vai de 1 a 24 Barks, correpondendo às primeiras 24 bandas-críticas do ouvido humano;

$$\Omega(\omega)=6\ln\left[\left(\frac{\omega}{1200\pi}\right)+\sqrt{\left(\frac{\omega}{1200\pi}\right)^2+1}\right] \quad \textbf{(2)}$$

5. O espectro de potência de tempo-curto é, então, convoluído com um banco de filtros (Equação 3 e Figura 2) para derivar um espectro similar ao espectro de potência de bandas críticas (Equação 3). São utilizados filtros sobrepostos com respostas trapezoidais em amplitude, igualmente espaçados na escala de freqüência Bark, com larguras de banda e freqüências centrais espaçadas de aproximadamente 1 Bark ;

$$\theta\left(\Omega_i\right)=\sum_{\Omega=-1.3}^{2.5}P\left(\Omega-\Omega_i\right)\cdot\Psi(\Omega) \quad \textbf{(3)}$$

6. O espectro de bandas críticas é pré-enfatizado pela curva de igual intensidade subjetiva (Equação 4), simulando a variação da sensação de intensidade sonora subjetiva (*loudness*) em função da freqüência a um nível de 40 dB (Equação 5). A Equação 4 representa uma função de transferência de um filtro com assíntotas de 12 dB/oct entre 0 e 400 Hz, 0 dB/oct entre 400 e 1200 Hz, 6 dB/oct entre 1200 e 3100 Hz, 0 dB/oct entre 3100 Hz e 5000 Hz e -18 dB/oct entre 5000 Hz e a freqüência de Nyquist.

$$E(\omega)=\frac{\left(\omega^2+56.8\times10^6\right)\cdot\omega^4}{\left(\omega^2+6.3\times10^6\right)^2\cdot\left(\omega^2+0.38\times10^9\right)\cdot\left(\omega^6+9.58\times10^{26}\right)} \quad \textbf{(4)}$$

$$\Xi\left(\Omega(\omega)\right)=E(\omega)\cdot\theta\left(\Omega(\omega)\right) \quad \textbf{(5)}$$

7. É aplicada uma raiz cúbica no espectro de bandas críticas pré-enfatizado (Equação 6), simulando a regra de potência do sistema auditivo humano [**17**], ou seja, a relação não-linear entre a intensidade sonora e sua percepção subjetiva. Esta etapa promove uma compressão da amplitude do espectro, e juntamente com a etapa anterior, reduz a variação da amplitude espectral do espectro de potência de bandas-críticas, de forma que se possa utilizar um modelo auto-regressivo de mais baixa ordem na etapa seguinte;

$$\Phi(\Omega)=\sqrt[3]{\Xi(\Omega)} \quad \textbf{(6)}$$

8. Finalmente, o espectro resultante da etapa anterior, Φ(Ω), é aproximado pelo espectro de um modelo contendo somente pólos, utilizando-se o método de autocorrelação da modelagem espectral de modelos contendo somente pólos [**18**]. Os coeficientes de autocorrelação são usados para a obtenção dos coeficientes cepstrais.

B. RASTA

Esta técnica utiliza processamento temporal, ao contrário das técnicas LPC, MFCC e PLP. A idéia chave por trás da abordagem temporal, que torna um SRAF mais robusto, é que o espectro do sinal de voz varia a uma razão diferente daquela

em que variam as formas potenciais de interferência acústica. Quando isto é verdadeiro, a filtragem das seqüências temporais dos parâmetros espectrais (as trajetórias espectrais) do sinal de voz em um domínio no qual a voz e a interferência são, aproximadamente, aditivas, pode suprimir os efeitos da interferência.

Desta forma é possível suprimir ruídos aditivos, cujos espectros variam mais lentamente ou mais rapidamente do que as porções do sinal de voz que carregam a informação lingüística, através da filtragem das trajetórias espectrais de potência [**7**], porque a voz e os ruídos são aditivos no domínio espectral de potência. Esta idéia é similar à da subtração espectral [**8**], porém não requer um mecanismo de detecção de voz, entre outras vantagens.

Similarmente, as formas espectrais desconhecidas contidas no sinal de voz podem ser suprimidas pela filtragem em escala logarítmica das trajetórias espectrais de potência [**9**, **10**]. Se esta filtragem inclui um componente passa-alta que suprime as variações a taxas abaixo de 1 Hz, esta também suprimirá o espectro médio do sinal de voz, que pode melhorar a independência de locutor em SRAF [**11**]. A presença deste componente tenderá a equalizar o espectro modulado dos parâmetros apresentados para o SRAF [**12**]. Esta equalização pode produzir parâmetros, cujas estatísticas temporais se ajustam melhor aos HMMs do que aqueles não filtrados [**11**]. Um componente passa-baixa que suprime variações à taxas acima de 16 Hz pode também melhorar o desempenho de um SRAF, porque estas variações não carregam informações lingüísticas significativas e porque variações a estas taxas podem não ser caracterizadas precisamente pela etapa inicial de processamento do sinal de voz [**12**, **11**]. Assim, tem-se um filtro passa-banda na faixa de 1 a 16 Hz.

Portanto, a técnica RASTA incorpora à técnica PLP uma filtragem da freqüência de modulação do espectro, para compensar formas espectrais desconhecidas e ruído aditivo.

A técnica de análise RASTA compreende as seguintes etapas:

1. Inicialmente o sinal de voz é filtrado por um filtro passa-baixa *anti-aliasing* e logo após amostrado;
2. O sinal é segmentado em *frames*, que contêm amostras que são ponderadas através da aplicação de uma janela, usualmente do tipo Hamming ;
3. A FFT é aplicada em cada *frame*, resultando o espectro de potência de tempo-curto do sinal de voz;
4. O espectro de potência de tempo-curto é transportado para a escala de freqüência Bark (Equação 2);
5. O espectro de potência de tempo-curto é, então, convoluído com um banco de filtros (Equação 1 e Figura 2) para derivar um espectro similar ao espectro de potência de bandas-críticas (Equação 3);
6. A saída de cada filtro é processada através de uma não linearidade compressiva e desprovida de memória. No processamento L-RASTA (*Logarithmic-RASTA*) *esta* não linearidade equivale a

$$y = \ln(x) \tag{7}$$

Enquanto no processamento J-RASTA equivale a

$$y = \ln(1 + Jx) \tag{8}$$

que é aproximadamente linear para pequenos valores de Jx (J muito menor que 1) e aproximadamente logarítmica para grandes valores de Jx (J muito maior que 1). Nas duas equações acima, y é o coeficiente cepstral e x, a saída de cada filtro.

7. Os coeficientes dos espectros de potência de banda crítica, transformados de forma não linear na etapa anterior, são filtrados através de um filtro passa-banda tipo IIR (Equação 9) com uma banda passante entre 1 e 12 Hz. Esta filtragem enfatiza aquelas partes do sinal que variam à taxas características da voz, enquanto suprime os elementos que variam à taxas mais baixas ou mais altas;

$$H(z) = \frac{z^4}{10}\left(\frac{2 + z^{-1} - z^{-3} - 2z^{-4}}{1 - 0.94z^{-1}}\right) \tag{9}$$

8. Os coeficientes dos espectros de potência filtrados são processados através de uma não linearidade expansiva e desprovida de memória. No processamento L-RASTA esta não linearidade equivale a

$$x = e^{y} \tag{10}$$

enquanto no processamento J-RASTA equivale a

$$x = \frac{e^{y}}{J} \tag{11}$$

onde y é o coeficiente filtrado e x, o coeficiente convertido de volta para a escala linear;

9. O espectro de potência de bandas críticas é pré-enfatizado pela curva de igual intensidade subjetiva (Equação 5);
10. É aplicada uma raiz cúbica no espectro de bandas críticas pré-enfatizado (Equação 6), simulando a regra de potência do sistema auditivo humano, ou seja, a relação não-linear entre a intensidade sonora e sua percepção subjetiva. Esta etapa promove uma compressão da amplitude do espectro, e juntamente com a etapa anterior, reduz a variação da amplitude espectral do espectro de potência de bandas críticas, de forma que se possa utilizar um modelo auto-regressivo de mais baixa ordem;
11. Finalmente, o espectro resultante da etapa anterior é aproximado pelo espectro de um modelo contendo somente pólos, utilizando-se o método de autocorrelação da modelagem espectral de modelos contendo somente pólos [**18**]. Os coeficientes de autocorrelação são usados para a obtenção dos coeficientes cepstrais:

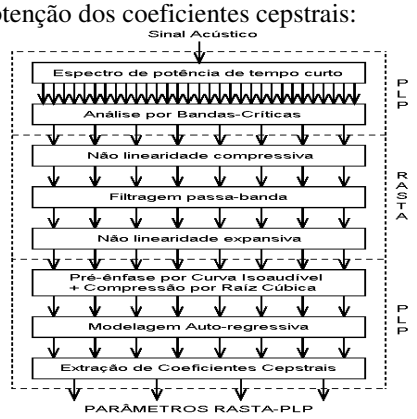

Figura 3 - Identificação das etapas da técnica de análise RASTA.

IV. METODOLOGIA

A. Introdução

O sistema implementa um reconhecedor de palavras isoladas extraídas da base de dados TIMIT [**23**], através de um aplicativo desenvolvido originalmente para esta tarefa, e denominado NIST2RIFF. Este aplicativo integra as ferramentas disponíveis no software HTK (*HMM Tool Kit*) [**19**], juntamente com a aplicação que implementa a técnica RASTA.

B. A base de dados de fala TIMIT

A TIMIT é uma base de dados fonético-acústica de fala contínua em Inglês, que foi desenvolvida para prover amostras de voz para aquisição e aprimoramento do conhecimento fonético-acústico, bem como, para o desenvolvimento e avaliação de SRAF.

A TIMIT contém um total de 6300 sentenças. Cada locutor, de um total de 630 locutores, pronuncia 10 destas sentenças. O presente trabalho faz uso apenas das sentenças do tipo *Compact (SX)* e delas extrai as palavras que irão compor o vocabulário reconhecido pelo sistema.

C. A base de dados de ruídos NOISEX-92

A NOISEX-92 [**21**] é a base de dados da qual foram extraídos os ruídos utilizados na etapa de testes do sistema. Os ruídos foram amostrados com um ADC de 16 bits a uma freqüência de 19.98 kHz e pré-filtrados com um filtro *anti-aliasing*. A duração de cada ruído é de 235 segundos. Segue abaixo uma descrição sucinta de cada um dos ruídos. Os nomes dos ruídos aparecem entre parênteses.

1) Ruído no interior de um veículo (volvo *)*

Ruído presente no interior de um Volvo 340 a uma velocidade de 120 km/h, em 4ª marcha, numa estrada asfaltada, num dia de chuva.

2) Ruído no interior de uma fábrica (factory1 *)*

Ruído produzido no interior de uma fábrica, próximo a um cortador de placas e a um equipamento de solda elétrica.

3) Ruído no interior de uma cantina (babble *)*

Este ruído provém de 100 pessoas conversando em uma cantina. As vozes individuais são ligeiramente audíveis.

D. Procedimento de teste

Numa primeira etapa foram executados dois testes imprescindíveis para validação do sistema e obtenção de percentuais de desempenho referenciais para o mesmo: o teste fechado e o teste aberto sem ruído, descritos abaixo.

- O teste fechado de um SRAF é efetuado na ausência de ruído e tem como objetivo validar o sistema. O desempenho neste caso deve atingir 100% para que o sistema esteja conforme, uma vez que os locutores que testam o sistema são os mesmos que o treinaram.
- O teste aberto de um SRAF, na ausência de ruído, tem o propósito de revelar o melhor desempenho alcançado por este, servindo de referência para os testes com ruído. No teste aberto, os locutores que testam o sistema são outros, que não aqueles que o treinaram.

Numa segunda etapa foram realizados os testes abertos na presença de ruído.

V. RESULTADOS

Os resultados foram obtidos através do aplicativo NIST2RIFF. Este aplicativo automatizou os testes de reconhecimento do sistema, aplicando seqüencialmente no sinal de voz, os tipos de ruídos e SNRs definidos para cada teste.

A. Reconhecimento de fala sem ruído

O reconhecimento de fala sem ruído, conforme já mencionado, compreende dois testes importantes que objetivam habilitar o SRAF para a tarefa designada, sendo eles o teste fechado e o teste aberto. As tabelas a seguir apresentam os resultados dos testes fechado e aberto, na ausência de ruído, para o sistema implementado no presente trabalho.

Tabela V.1 - Desempenho do reconhecedor (em %) para teste fechado sem ruído

Parâmetros / SNR(dB)	LPC_E	MFCC_E	PLP_E	RASTA_E
-	100.00	100.00	100.00	100.00

Tabela V.2 - Desempenho do reconhecedor (em %) para teste aberto sem ruído.

Parâmetros / SNR(dB)	LPC_E	MFCC_E	PLP_E	RASTA_E
-	64.29	71.43	85.71	92.86

1) Análise dos resultados obtidos

Os resultados da Tabela V.1 validam o SRAF, pois apresentam 100 % de desempenho para todos os testes.

Os resultados da Tabela V.2 já dão mostra da superioridade das técnicas de análise perceptuais, PLP e RASTA.

B. Reconhecimento de fala com ruído

As tabelas a seguir apresentam os resultados do reconhecimento de fala do sistema face aos diferentes ruídos adicionados ao sinal de voz.

Tabela V.3 - Desempenho do reconhecedor (em %) face ao ruído no interior de um veículo.

Parâmetros / SNR(dB)	LPC_E	MFCC_E	PLP_E	RASTA_E
-6	14.29	28.57	57.14	64.29
-3	21.43	50.00	78.57	64.29
0	35.71	57.14	78.57	85.71
3	35.71	64.29	85.71	85.71
6	35.71	64.29	85.71	92.86
9	50.00	71.43	78.57	92.86
12	57.14	78.57	78.57	92.86
15	64.29	78.57	78.57	92.86
18	64.29	71.43	78.57	92.86

Tabela V.4 - Desempenho do reconhecedor (em %) face ao ruído no interior de uma fábrica

Parâmetros / SNR(dB)	LPC_E	MFCC_E	PLP_E	RASTA_E
-6	0.00	21.43	21.43	7.14
-3	0.00	35.71	35.71	7.14
0	0.00	35.71	42.86	14.29
3	0.00	28.57	42.86	35.71
6	0.00	42.86	42.86	57.14
9	14.29	42.86	57.14	57.14
12	21.43	50.00	64.29	71.43
15	21.43	50.00	64.29	71.43
18	28.57	64.29	78.57	71.43

Tabela V.5 - Desempenho do reconhecedor (em %) face ao ruído no interior de uma cantina.

Parâmetros / SNR(dB)	LPC_E	MFCC_E	PLP_E	RASTA_E
-6	0.00	7.14	28.57	28.57
-3	0.00	14.29	28.57	21.43
0	0.00	21.43	35.71	35.71
3	0.00	28.57	35.71	42.86
6	0.00	42.86	64.29	50.00
9	14.29	50.00	71.43	50.00
12	21.43	57.14	71.43	78.57
15	21.43	57.14	71.43	85.71
18	21.43	64.29	78.57	92.86

1) Análise dos resultados obtidos

Os resultados apresentados na tabela V.4 e na Tabela V.5 confirmam a conhecida degradação da representação LPC em condições de ruído, devido, principalmente, à subida das médias dos parâmetros nos segmentos do sinal onde antes existia silêncio.

É impressionante o resultado obtido com a modelação RASTA com um número tão reduzido de parâmetros. Isto se deve à normalização cepstral e a suavização espectral derivada da modelação LPC implícita nesta técnica de análise, e também devido à reduzida dimensão do conjunto de treino.

VI. CONCLUSÕES

Os resultados mostraram que existem de fato vantagens em utilizar a representação auditiva, e permitem concluir que as técnicas PLP e RASTA apresentam excelente desempenho em ambientes com ruídos que variam lentamente suas características espectrais (ruído no interior de um carro, por exemplo). Por outro lado, estas técnicas não apresentam desempenho considerável quando o ruído for correlacionado com o sinal de voz (ruído no interior de uma cantina, por exemplo). Deve-se também notar o pequeno número de coeficientes (8 no total) necessários para estes tipos de técnicas, o que propicia uma economia de memória de armazenamento e maior eficiência de banda em um canal de comunicação, nas situações onde for necessária a transmissão do sinal parametrizado.

VII. REFERÊNCIAS

1. H. Hermansky, "Perceptual Linear Predictive (PLP) Analysis of Speech", J. Acoust. Soc. Am. 87(4), pp. 1738-1752, Apr. 1990.
2. Shaugnessy, D. "Speech Communications: Human and Machine". 2ed.IEEE Press.2000.
3. P. Buser, M. Imbert, "Audition", MIT Press, 1992.
4. J-C. Junqua, H. Wakita, H. Hermansy, "Evaluation and Optimization of Perceptually-Based ASR Front-End", IEEE trans. Speech and Audio Proc., Vol. 1, No. 1, Jan 1993.
5. P. Lockwood and J. Boudy, "Experiments with a Non-linear Spectral Subtractor, HMMs and the projection, for robust speech recognition in cars", Eurospeech '91, Vol. 11, Nos. 2-3, pp. 215-228, 1992.
6. Jean-Claude Lunqua, "The Lombard Reflex and its role on human listeners and automatic speech recognisers", JASA 93(1) Jan 1993 p510-524.
7. H. G. Hirsch, P. Meyer, and H. W. Ruehl. Improved speech recognition using high-pass filtering of subband envelopes. In *Proceedings of Eurospeech 1991*, pages 413-146, 1991.
8. Steven F. Boll. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-27(2):113-120, April 1979.
9. Hynek Hermansky, Nelson Morgan, Aruna Bayya, and Phil Kohn. Compensation for the effect of the communication channel in auditory-like analysis of speech (RASTA-PLP). In *Proceedings of Eurospeech 1991*, pages 1367-1370, 1991.
10. Hynek Hermansky and Nelson Morgan. RASTA processing of speech. *IEEE Transactions on Speech and Audio Processing*, 2(4):578-589, October 1994.
11. Climent Nadeu, Pau Pachès-Leal, and Biing-Hwang Juang. Filtering the time sequences of spectral parameters for speech recognition. *Speech Communication*, 22(4):315-332, September 1997.
12. Climent Nadeu and Biing-Hwang Juang. Filtering of spectral parameters for speech recognition. In ICSLP 94. *Proceedings of the 1994 International Conference on Spoken Language Processing*, pages 1927-1930, 1994.
13. Nelson Morgan and Hynek Hermansky. RASTA extensions: Robustness to additive and convolutional noise. In *Proceedings of the ESCA Workshop on Speech Processing in Adverse Environments*, pages 115-118, 1992.
14. S.Furui, "Speaker-Independent Isolated Word Recognition Using Dynamic Features of Speech Spectrum", IEEE, ASSP-34, No.1, Feb.86.
15. H. Hermansky, N. Morgan, A. Bayya, P. Kohn, "RASTA-PLP Speech Analysis Technique", Proc. ICASSP-92, pp. I.121-I.124, 1992.
16. Schroeder, M.R. : "Recognition of Complex Acoustic Signals", in Life Sciences Research Report 5, T.H. Bullock, Ed., p. 324, Abakon Verlag, Berlin.
17. Stevens, S.S. : "On the psychophysical law", Psychological Review, 64, pp. 153-181.
18. Makhoul, J. : "Spectral linear prediction: properties and applications", IEEE Trans. ASSP-23, pp. 283-296.
19. S. Young, G. Evermann, D. Kershaw, G. Moore, J. Odell, D. Ollason, V. Valtchev, P. Woodland, "The HTK Book - (for HTK Version 3.1)", Cambridge University, Dec. 2001.
20. Sanches, I. : "Improved speech recognition through the use of noise-compensated hidden Markov models", University of London, Nov. 1994.
21. A. P. Varga et al., "The NOISEX-92 study on the effect of additive noise on automatic speech recognition", in *Technical Report, DRA Speech Research Unit*, 1992.
22. Furui, S. "Digital Speech Processing, Synthesis and Recognition". Tokyo: Tokai University Press, 1985.
23. TIMIT. "TI-Texas Instruments & MIT", CD1-1.1, 1990.