MELHORAMENTO DE SINAIS DE VOZ ATRAVÉS DE MÉTODO DE SUPRESSÃO DE RUÍDOS POR BALANCEAMENTO ESPECTRAL ADAPTATIVO

Benedito G. Aguiar Neto

Laboratório de Sinais, Imagens e Computação Gráfica
Departamento de Engenharia Elétrica - UFPB
58.100 - Campina Grande - PB.

Este trabalho apresenta um estudo e avaliação de um método de supressão de ruídos aplicado a sinais de voz degradados, baseado em um filtro Wiener-Kolmogoroff e na estimação espectral em curtos intervalos de tempo. A filtragem do sinal de voz degradado é levada a efeito através de um balanceamento do espectro do sinal degradado, em função do espectro do sinal degradante estimado nos intervalos de pausas.

São apresentados os resultados obtidos para uma degradação dos sinais de voz por ruído de automóvel e ruído branco. Os resultados mostraram um considerável aumento na Relação Sinal-Ruído do sinal de voz, o que levou a um significativo melhoramento na qualidade e intelegibilidade do sinal degradado.

1. INTRODUÇÃO

Em sistemas digitais de transmissão de voz, é comum a presença de componentes indesejáveis, na forma de ruídos acústicos, que são sobrepostos ao sinal de voz, levando a uma redução da qualidade e intelegibilidade da conversação entre usuários. Os ruídos acústicos são produzidos por fontes externas localizadas no meio ambiente dos usuários, sendo freqüentes, por exemplo, em sistemas móveis de comunicações ou sistemas de comunicações a céu aberto. Exemplos de tais ruídos ambientais são: ruídos de trânsito, ruídos de motores e ruídos em zonas industriais.

A degradação dos sinais de voz por ruídos acústicos apresenta-se de forma intermitente, ou seja, todos os valores das amostras do sinal são atingidos. Desta forma, para redução do ruído, devem ser utilizadas técnicas não seletivas, que permitam uma filtragem contínua do sinal degradado |1|. Estas técnicas incluem os chamados "métodos de supressão de ruídos" |1,2| que se baseiam na teoria dos filtros ótimos e/ou teoria de estimação espectral a curtos intervalos de tempo.

Neste trabalho, é estudado um "método de supressão de ruídos" no qual a redução do ruído é levada a efeito através de um balanceamento adaptativo da amplitude espectral do sinal degradado |1|. O balanceamento espectral consiste em determinar-se uma função de transferência de um filtro Wiener-Kolmogoroff, obtida a partir de estimações do espectro do sinal de voz degradado e do espectro do sinal de ruído, que é utilizada para obter-se uma modificação da amplitude espectral do sinal de voz degradado. Este método foi inicialmente proposto em |2| e estudado em uma estrutura de filtros polifásicos em |3|. Este trabalho apresenta uma nova variante do método, baseada em estimações espectrais a curtos intervalos de tempo, por uso da transformada discreta de Fourier (TDF).

II. PRINCÍPIOS DO SISTEMA ADAPTATIVO DE SUPRESSÃO DE RUÍDOS

A Fig. 1 mostra a estrutura geral de um sistema adaptativo de supressão de ruídos. Estes sistemas utilizam, em geral, algum tipo de filtro adaptativo cuja resposta ao impulso é determinada em função das propriedades estatísticas do sinal a ser melhorado e do ruído. Estas propriedades, em geral, não são disponíveis nestes sistemas. Um pré-conhecimento delas não é também possível, se os sinais são não estacionários. Por isso, as informações necessárias a respeito da estatística dos sinais devem ser obtidas a partir do sinal degradado, e em curtos intervalos de tempo correspondentes a segmentos de igual comprimento $y(n,k)$, onde $n = 1,2,\ldots, n_L$ e $k = 1,2,\ldots, n_S$. n_L e n_S representam o número de amostras de cada segmento e o número de segmentos respectivamente. Os parâmetros do filtro são calculados e atualizados a cada novo segmento. Supõe-se, portanto, que os sinais são suficientemente estacionários nos intervalos de tempo considerados. Para sinais de voz, pode-se considerar que esta estacionariedade existe em intervalos de 16 ms a 32 ms |4|. Uma outra consideração a ser feita é a de que o sinal de voz não seja correlacionado com o ruído.

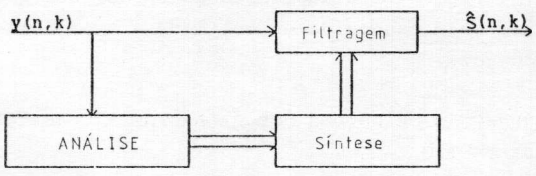

Fig. 1 - Estrutura geral de um sistema adaptativo de supressão de ruídos

O ponto de partida para o método de supressão de ruídos estudado é a equação de Wiener-Hopf

$$R_{ys}(i) = \sum_{j\varepsilon I} h_{j,opt} . R_{yy}(i-j) \qquad i\varepsilon I \qquad (1)$$

obtida para os filtros ótimos denominados por "filtros Wiener-Kolmogoroff (WK)" |5|. Nestes filtros, a resposta ao impulso $h_{j,opt}$, onde $j\varepsilon I$, é obtida segundo o critério do mínimo erro médio quadrático. A eq. 1 mostra que, para o cálculo de $h_{j,opt}$, são necessárias apenas a autocorrelação, $R_{yy}(i)$, do sinal de voz degradado e a correlação $R_{ys}(i)$, entre o sinal degradado, y(n), e o sinal original, s(n). Quaisquer outras informações sobre os sinais são supérfluas.

Para o filtro WK não causal, I = (-∞,∞), tem-se:

$$R_{ys}(i) = h_{i,opt} * R_{yy}(i) = h_{opt} * R_{yy}(i) \qquad (2)$$

Aplicando-se a transformada de Fourier à eq. 2, obtém-se a seguinte função de transferência:

$$H_{opt}(j\Omega) = \frac{S_{ys}(j\Omega)}{S_{yy}(\Omega)} \qquad (3)$$

onde $S_{ys}(j\Omega)$ é a densidade espectral de potência cruzada entre y(n) e s(n) e $S_{yy}(\Omega)$ é a densidade espectral de potência (DEP) de y(n).

Seja o sinal de voz degradado y(n) = s(n) + n(n), onde n(n) é o sinal de ruído. Pode-se mostrar que, para o caso de s(n) e n(n) não serem correlacionados entre si |6|,

$$S_{ys}(j\Omega) = S_{ss}(\Omega) \qquad (4a)$$

e

$$S_{yy}(\Omega) = S_{ss}(\Omega) + S_{nn}(\Omega). \qquad (4b)$$

Substituindo-se as eq. 4a e 4b na eq. 3, obtém-se o seguinte algoritmo de supressão de ruídos:

$$H_{opt}(\Omega) = \begin{cases} 1 - \dfrac{S_{nn}(\Omega)}{S_{yy}(\Omega)} & \text{para } S_{nn}(\Omega) < S_{yy}(\Omega) \\ 0 & \text{para } S_n(\Omega) \geq S_{yy}(\Omega) \end{cases} \qquad (5)$$

Observando-se a eq. 5 vê-se que a função de transferência $H_{opt}(\Omega)$ permite a supressão do ruído, sem que, para tanto, seja conhecida ou estimada a estatística do sinal de voz.

III. SUPRESSÃO DE RUÍDOS POR BALANCEAMENTO ESPECTRAL ADAPTATIVO

No sistema de supressão de ruídos por balanceamento espectral adaptativo, a filtragem do sinal degradado é realizada através de modificações da amplitude espectral deste sinal, segundo a eq. (5). Entretanto, esta equação não pode ser utilizada diretamente, pois os sinais de voz e freqüentemente também os sinais de ruído são não estacionários. Contudo, as densidades espectrais de potência, $S_{yy}(\Omega)$ e $S_{nn}(\Omega)$, podem ser substituídas por suas respectivas estimações $S_{yy}(\Omega,k)$ e $S_{nn}(\Omega,k)$, determinadas em curtos k-ésimos intervalos de tempo, isto é:

$$S_{yy}(\Omega,k) = \hat{E}[|Y(\Omega,k)|^2] \qquad (6)$$

$$S_{nn}(\Omega,k) = \hat{E}[|N(\Omega,k)|^2] \qquad (7)$$

onde $|Y(\Omega,k)|$ e $|N(\Omega,k)|$ são as amplitudes espectrais do sinal y(n,k) e n(n,k), respectivamente. Em segmentos de atividade de voz, a DEP do sinal degradado, devido à sua forte não estacionariedade pode ser aproximada pelo quadrado da amplitude espectral do k-ésimo segmento:

$$S_{yy}(\Omega,k) \cong |Y(\Omega,k)|^2 \qquad (8)$$

A estimação $S_{nn}(\Omega,k)$ da DEP do sinal de ruído é levada a efeito durante os intervalos de pausas através da média quadrática da amplitude espectral $|Y(\Omega,k)| \cong |N(\Omega,k)|$, para k = 1,2, ... n_p, determinada em n_p intervalos de pausas do sinal degradado:

$$S_{nn}(\Omega,k) \cong \frac{1}{n_p} \sum_{k=1}^{n_p} |Y(\Omega,k)|^2 \qquad (9)$$

Com os valores estimados segundo as eq. (8) e (9), obtém-se

$$H_{opt}(\Omega,k) = \begin{cases} 1 - Q(\Omega,k) & \text{para } 0 < Q(\Omega,k) < 1 \\ 0 & \text{para } Q(\Omega,k) \geq 1 \end{cases} \qquad (10)$$

com

$$Q(\Omega,k) = \frac{\hat{E}[|N(\Omega,k)|^2]}{|Y(\Omega,k)|^2} \qquad (11)$$

onde Q(Ω,k) é independente do nível absoluto do sinal degradante e foi denominado em |3| como "fator relativo de degradação.

A estimação do espectro de um segmento do sinal de voz original é obtida através de um balanceamento espectral da amplitude espectral do k-ésimo segmento do sinal degradado:

$$\hat{S}(\Omega,k) = \{H_{opt}(\Omega,k) . |Y(\Omega,k)| \, e^{j\theta(\Omega,k)}\} \qquad (12)$$

Nesta estimação, é considerado que a fase do sinal de voz degradado representa uma aproximação utilizável da fase do sinal de voz original, ou seja:

$$\arg\hat{S}(\Omega,k) = \arg Y(\Omega,k) = \theta(\Omega,k) \qquad (13)$$

Esta aproximação baseia-se no fato de que o ouvido humano é relativamente insensível a degradações de fase

|7,8|.

IV. REALIZAÇÃO POR TRANSFORMADA DISCRETA DE FOURIER

A Fig. 2 mostra a estrutura do método de supressão de ruídos por balanceamento espectral adaptativo. Este método é realizado através da Transformada Discreta de Fourier.

O sinal de voz degradado é dividido em segmentos de L amostras com 50% de superposição entre si, que são multiplicados por uma janela de Hamming. O k-ésimo segmento, a ser filtrado y(n,k), é então transformado em um vetor de coeficientes de Fourier:

$$Y_{i,k} = \mathcal{F}\{y(n,k)\} \qquad i = 0,1,2,\ldots,L-1 \qquad (14)$$

A supressão do ruído é feita por uma modificação do módulo das componentes do vetor $Y_{i,k}$:

$$|\hat{S}_{i,k}| = C_{i,k}\ |Y_{i,k}| \qquad (15)$$

Na transformação do k-ésimo segmento, y(n,k), de L amostras, cada componente do vetor $Y_{i,k}$ tem um espaçamento de freqüência de $2\pi/L$. Os coeficientes de balanceamento espectral $C_{i,k}$ são, portanto:

$$C_{i,k} = H(i2\pi/L,k) \qquad i = 0,1,2,\ldots,L-1 \qquad (16)$$

Ao módulo das componentes do vetor balanceado, $\hat{S}_{i,k}$, é associada a fase de $Y_{i,k}$. Através da transformação inversa do vetor $\hat{S}_{i,k}$, obtém-se então o segmento do sinal de voz melhorado, $\hat{S}(n,k)$. Na região de superposição dos segmentos são então somadas as amostras de transformações subseqüentes.

Durante os intervalos de pausas, o detetor de pausas ativa a atualização da estimação da DEP do sinal degradante. Nos intervalos de atividade de voz são utilizados os valores da DEP do último intervalo de pausa. É considerado que a DEP do sinal degradante permanece, de forma aproximada, inalterada nos intervalos de atividade de voz subseqüentes até uma próxima atualização nos intervalos de pausas.

A adaptação às propriedades estatísticas variantes no tempo dos sinais de voz é melhor à medida que o segmento de análise é menor. Por outro lado, para obter-se uma aceitável resolução do espectro, o comprimento do segmento deve ser escolhido de forma a ter-se, no mínimo, o dobro do período da freqüência fundamental esperada |9|. Como compromisso entre essas duas exigências, foi escolhido um segmento de 256 amostras. O janelamento pela função de Hamming permite uma mudança "suave" nas fronteiras dos segmentos, de modo a serem evitadas descontinuidades na fase.

Resultados preliminares mostraram a presença de um ruído residual produzido quando algumas componentes espectrais do sinal tratado são fortemente atenuadas ou até totalmente suprimidas devido a estimações espectrais errôneas. Este mecanismo produz "buracos" no espectro do sinal tratado, cuja percepção auditiva se apresenta de natureza tonal com freqüência variável. Estes ruídos residuais são comumente denominados na literatura como "ruídos musicais" |10,11| e são freqüentes nos intervalos de baixa energia do sinal. Um considerável melhoramento da qualidade do sinal tratado, pode, portanto, ser obtido se o ruído musical puder ser reduzido ou mascarado. Com este objetivo, foram efetuadas algumas modificações no algoritmo estudado. Uma modificação consistiu em utilizar-se, para a determinação de $Q_{i,k}$ e $H_{i,k}$ nos intervalos de pausas, não a estimação da DEP, $\hat{E}[|N_{i,k}|^2]$, mas sim a atual estimação de cada k-ésimo intervalo, $|Y_{i,k}|^2$. Isto leva a $Q_{i,k} = 1$ nestes intervalos, significando uma total supressão do ruído nos mesmos. Este procedimento leva, entretanto, a indesejáveis supressões de segmentos de baixa energia do sinal de voz, se estes segmentos são interpretados como intervalos de pausas.

Desta forma, podem ser produzidos saltos bruscos dos intervalos de pausas para intervalos de atividade de voz com alta energia, ou o contrário, que são percebidos auditivamente de forma não natural. Isto pôde, entretanto, ser compensado, prolongando-se a região detetada de atividade de voz.

Uma outra modificação consistiu em não permitir-se que

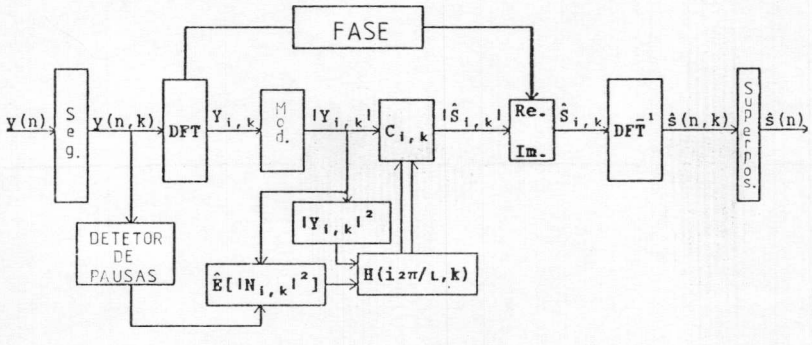

Fig. 2 - Estrutura geral do método de supressão de ruídos por balanceamento espectral adaptativo

os valores espectrais atingissem valores abaixo de um determinado valor após o balanceamento, de forma a evitar-se a impressão de curtas interrupções no espectro. Entretanto, a impressão subjetiva da perturbação nos intervalos de pausas deve corresponder à mesma nos intervalos de atividade de voz. Desta forma, foi também fixado o valor mínimo a ser atingido pelos valores espectrais nos intervalos de pausas nos mesmos valores fixados nos intervalos de atividade de voz.

Com estas modificações, os coeficientes de balanceamento são obtidos da seguinte forma:

$$C_{i,k} = \begin{cases} H(Q_{i,k}) & |\hat{S}_{i,k}| > E[|N_{i,k}|^2] \\ \lambda \hat{E}[|N_{i,k}|^2] & |\hat{S}_{i,k}| \leq E[|N_{i,k}|^2] \end{cases} \quad (17)$$

onde $0 < \lambda << 1$ e

$$Q_{i,k} = \begin{cases} \hat{E}[|N_{i,k}|^2]/|Y_{i,k}|^2 & \text{em atividade de voz} \\ 1 & \text{nas pausas} \end{cases} \quad (18)$$

Uma boa percepção auditiva foi feita para valores de λ situados entre 0,05 e 0,1.

V. CONDIÇÕES EXPERIMENTAIS

Como material de teste do método de supressão de ruídos estudado, foram utilizados os seguintes sinais de prova:

a) Sinal de voz, com uma duração de 7,68 s, correspondente a um texto falado por um interlocutor masculino (SP1)

b) Sinal de voz, com uma duração de 5,632 s, falado por dois interlocutores masculinos (SP2)

c) Ruído acústico de uma Kombi movida à diesel, gravado no seu interior do início da partida até uma velocidade de 80 Km/h.

Os sinais acima listados foram digitalizados com uma freqüência de amostragem de 8 KHz e 14 bits/amostra.

A Fig. 3 mostra a forma do sinal de voz correspondente a um trecho de 704 ms de duração do sinal de prova SP1. Neste gráfico, é destacada a forma pseudo-periódica da vogal /a/ e a forma ruidosa da consoante /s/. Pode-se ainda reconhecer que a energia dos sons surdos (p. ex., /s/) é consideravelmente mais baixa que a dos sons sonoros (p. ex., /a/).

Os sinais de voz degradados foram obtidos adicionando-se os sinais de ruído, de modo a obter-se uma desejada Relação Sinal-Ruído (SNR).

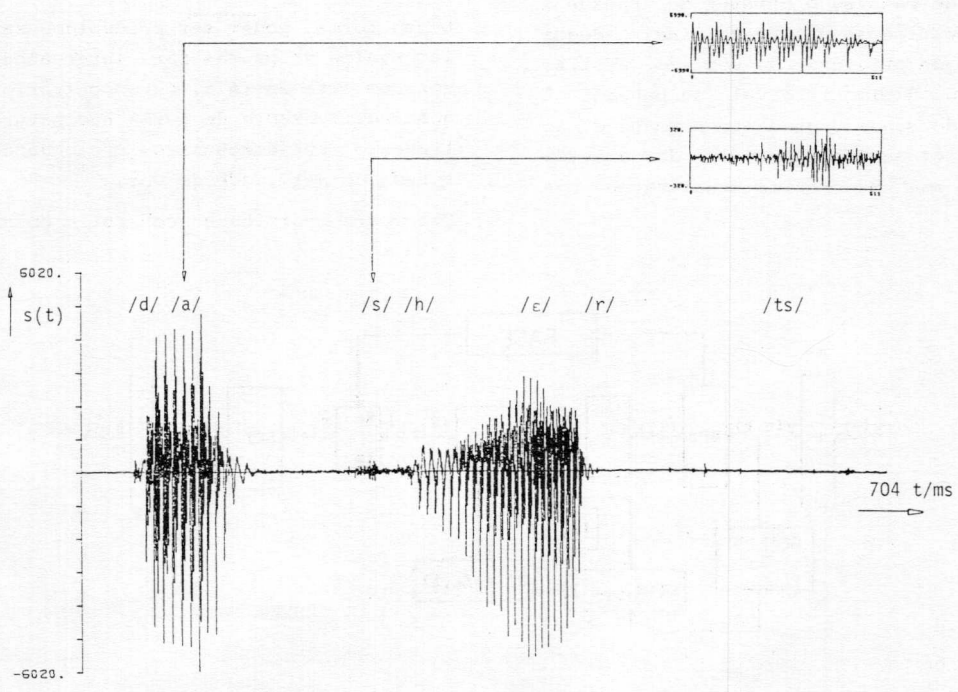

Fig. 3 - Forma do sinal de voz "Das Hertz",/das herts/

VI. RESULTADOS DA SIMULAÇÃO E DISCUSSÃO

A tabela mostra os resultados do método de supressão de ruídos por balanceamento espectral, para uma Relação Sinal-Ruído = 10 dB. São mostrados o ganho da SNR (SNR-G) e o ganho da Relação Sinal-Ruído-Segmental (SegSNR-G) para os sinais de voz SP1 e SP2 com degradação por ruído de automóvel e ruído branco.

Tipo de Ruído	SNR-G dB	SegSNR-G dB	Sinal de Prova
Automóvel	5,18	8,54	SP1
Branco	4,15	7,85	
Automóvel	4,93	15,99	SP2
Branco	6,94	17,26	

Tabela - Resultados do método de supressão de ruídos para os sinais de voz SP1 e SP2 degradados por ruído de automóvel e ruído branco.

Os resultados mostram um considerável ganho na SNR e particularmente da SegSNR, que mostram a eficácia do método estudado, tanto na supressão do ruído de automóvel quanto do ruído gaussiano. Na figura 4, é mostrado, de forma visual o efeito do método de supressão de ruídos através do espectrograma do sinal de voz após o tratamento. Nesta figura, percebe-se, de forma considerável, a eficácia do método na supressão do ruído, principalmente nos intervalos de pausas; isto explica os altos valores obtidos para SegSNR-G, uma vez que o sinal de prova SP2 possui um longo intervalo de pausa. Entretanto, percebe-se também, comparando-se a Fig.4.a como a Fig. 4.c, que segmentos de baixa energia do sinal original (p. ex., /s/) são também parcialmente suprimidos, se estes segmentos são interpretados como pausas.

O êxito de um método de supressão de ruídos é, portanto, bastante dependente de uma detecção de pausas livre de erros, pois a adaptação do algoritmo de supressão de ruídos é realizada somente nos intervalos de pausas, através de uma atualização dos valores expectrais estimados para o sinal degradante. O melhoramento da voz degradada depende, portanto, também da estacionariedade do sinal degradante, pois, se este sinal for fortemente não estacionário, dificilmente os parâmetros do filtro, atualizados dirante os intervalos de pausas, terão o mesmo efeito nos intervalos de atividade de voz, uma vez que, nestes intervalos, não é possível uma atualização do espectro do ruído.

Os critérios de avaliação do método estudado foram, em primeira linha, de caráter objetivo através da medição da SNR e SegSNR. Além disso, foram levados a efeito testes subjetivos, através da reprodução acústica dos sinais de voz após o tratamento, que confirmaram, de forma qualitativa, o melhoramento na qualidade e inteligibilidade dos sinais de voz.

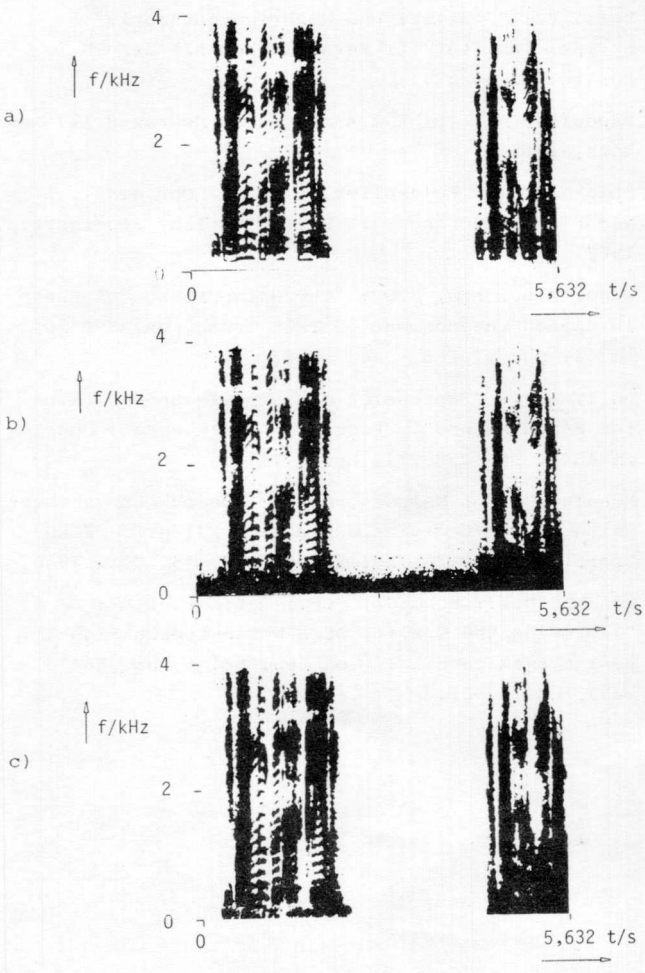

Fig. 4 - Espectrograma para o sinal de prova SP2: a) Sinal de voz original, b) Sinal de voz degradado por ruído de automóvel e c) Sinal de voz tratado.

VII. AGRADECIMENTOS

O autor agradece ao prof. Dr.-Ing. Peter Noll, do "Institut für Fernmeldetechnik" da Universidade Técnica de Berlim, pelo valioso apoio ao desenvolvimento deste trabalho.

VIII. REFERÊNCIAS BIBLIOGRÁFICAS

|1| Aguiar Neto, B.G.:"Signalaufbereitung in digitalen Sprachübertragungssystemen". Dissertation, Technische Universität Berlin, 1987.

|2| Lim, J.S.; Oppenheim, A.V.:"Enhancement and Bandwidth Compression of Noise Speech". Proc. of the IEEE, Vol. 67, No. 12, pp. 1586-1604, Aug. 1979.

|3| Vari, P.: "Verfahren zur Digitalen Verbesserung Gestörter Sprache". TEKADE Tech. Mitteilungen, S. 70-76, 1983.

[4] Fellbaum, K.: "Sprachsignalverarbeitung and Sprachübertragung. Springer-Verlag, Berlin, 1984.

[5] Noll, P.: "Statistische Nachrichtentheorie". Skript, Institut für Fernmeldetechnik der TU Berlin, WS 86/87.

[6] Papoulis, A.: "Signal Analysis". McGraw-Hill, New York, 1985.

[7] Zwicker, E.: Feldkeller, R.: "Das Ohr als Nachrichtenempfänger". Hirzel-Verlag, Stuttgart, 1967.

[8] Wang, D.L.; Lim, J.S.: "The Unimportance of Phase in Speech Enhancement". IEEE Trans. on ASSP-30, No. 4, pp. 679-681, Aug. 1982.

[9] Boll, S.F.: "Supression of Noise in Speech Using the SABER Method". Proc. on the Internat. Conf. on ASSP, pp. 208-211, April 1979.

[10] McAulay, R.J.; Malpass, M.L.: "Speech Enhancement Using a Soft-Decision Supression Filter". IEEE Trans. on ASSP-28, No. 2, pp. 137-145, Oct. 1980.

[11] Sondni, M.M.; Schmidt, C.E.; Rabiner, L.R.: "Improving the Quality of a Noise Speech Signal". Bell System Tech. J., Vol. 60, No. 6, pp. 1847-1859, Oct. 1981.