

9618030

**UNIVERSIDADE ESTADUAL DE CAMPINAS  
FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO  
DEPARTAMENTO DE COMUNICAÇÕES**

***Tese de Doutorado***

**REDUÇÃO DE RUÍDO EM SINAIS DE VOZ NOS SISTEMAS RÁDIO**

**MÓVEIS VEICULARES**

Este exemplar corresponde à redação final da tese  
apresentada por Jozué Vieira Filho  
e aprovada pela Comissão  
 julgadora em 22 / 08 / 96.  
José Geraldo Chiquito  
Orientador

Tese apresentada à Faculdade de Engenharia  
Elétrica e de Computação da Universidade  
Estadual de Campinas como parte dos  
requisitos para a obtenção do título de  
DOUTOR EM ENGENHARIA ELÉTRICA

Autor: Jozué Vieira Filho  
Orientador: Professor Dr. José Geraldo Chiquito

***Campinas, agosto de 1996***

E BC  
 AMADA: UNICAMP  
 73/12  
 Ex. 28622  
 667196  
☐ D ☒  
 DO RB 11,00  
 A 19109196  
 CPD

M-00092061-2

FICHA CATALOGRÁFICA ELABORADA PELA  
 BIBLIOTECA DA ÁREA DE ENGENHARIA - BAE - UNICAMP

V673r Vieira Filho, Jozué  
 Redução de ruído em sinais de voz nos sistemas rádio-  
 móveis veiculares / Jozué Vieira Filho.--Campinas, SP:  
 [s.n.], 1996.

Orientador: José Geraldo Chiquito.  
 Tese (doutorado) - Universidade Estadual de Campinas  
 Faculdade de Engenharia Elétrica e de Computação.

1. Ruído. 2. Voz. 3. Veículos. 4. Sistemas de  
 comunicação móvel. I. Chiquito, José Geraldo. II.  
 Universidade Estadual de Campinas. Faculdade de  
 Engenharia Elétrica e de Computação. III. Título.

Trabalho realizado com o apoio do CNPq, através do programa normal de doutorado no Brasil (6 meses de bolsa), e do programa “Doutorado Sandwich”, dentro do convênio CNPq-CEFI, na França (18 meses).

---

À minha família e a Sônia

---

## *AGRADECIMENTOS*

- Ao departamento TSS/CMC do CENTRE NATIONAL D'ETUDE DES TELECOMMUNICATIONS (CNET), onde desenvolveu-se uma grande parte deste trabalho, dentro do programa de “doutorado sandwich” do CNPq;
- Ao professor Dr. José Geraldo Chiquito pelo apoio a idéia inicial deste trabalho e pela orientação durante o desenvolvimento do mesmo;
- Ao Dr. PASCAL SCALART, engenheiro com quem trabalhei um ano e meio no CNET e sem dúvida um grande responsável pelos resultados obtidos;
- Ao senhor ANDRÉ GILLOIRE, chefe do grupo de acústica do CNET/TSS/CMC, por ter viabilizado minha ida a esse grande centro de pesquisa, pelas discussões técnicas, pela revisão dos artigos, enfim por toda sua contribuição;
- A Faculdade de Engenharia de Ilha Solteira (DEE-FEIS) da Universidade Estadual Paulista (UNESP) que me permitiu um ano e meio de afastamento para ir ao CNET;
- A todos os meus amigos, brasileiros e franceses, com quem convivi durante este processo, em Campinas, Ilha Solteira e Lannion;
- Agradecimentos especiais a Sônia, quem eu amo, pelo incentivo e apoio durante o desenvolvimento deste trabalho, e aos amigos Antônio Padilha Feltrin e Carlos Roberto Minussi;
- Agradecimento especialíssimo ao grande amigo Edson Nagle, pelo seu incentivo e pela excelente revisão deste trabalho.

## ***SIMBOLOGIA E ABREVIACÕES***

- ♦ *MSV* - Melhoramento de Sinais de Voz;
- ♦ *HFTs* - Terminais Telefônico Mãos-Livres (“Hands-Free Telephones”);
- ♦ *PITCH* - Frequência fundamental de um sinal de voz;
- ♦ *STFT* - “Short Time Fourier Transform” ou Transformada de Fourier de curto prazo;
- ♦ *SE* - Subtração Espectral;
- ♦ *SNR* - Relação Sinal/Ruído;
- ♦ *SNR\_prio* - Relação Sinal/Ruído a priori;
- ♦ *SNR\_post* - Relação Sinal/Ruído a posteriori;
- ♦ *LPC* - Predição Linear;
- ♦ *DVS* - Detecção de intervalos Voz-Silêncio;
- ♦ *DCEP* - Distância Cepstral
- ♦  $\sigma_i^2$  - Potência Média
- ♦  $\gamma_{ii}$  - Auto-Espectro de um sinal  $i$ ;
- ♦  $\mathcal{F}[\bullet]$  - Transformada de Fourier;
- ♦  $\rho(\omega)$  - Função de Coerência;
- ♦  $\wedge$  - Indica Estimação

## **RESUMO**

Neste trabalho propõem-se técnicas de redução de ruído para aplicação nos sistemas de comunicação rádio-móveis veiculares. As técnicas propostas são baseadas na estimação da relação sinal/ruído a priori, quando um ou dois microfones são disponíveis para captação do sinal de voz no veículo. Os resultados foram obtidos a partir de um estudo das técnicas clássicas de redução de ruído, como subtração espectral, filtragem de Wiener, etc. Mostra-se que essas técnicas podem ser definidas em função da relação sinal/ruído a posteriori ( $SNR_{post}$ ). Este parâmetro é analisado e identificado como o responsável pelo fenômeno do *ruído musical* nas técnicas clássicas de redução de ruído que usam um único microfone para captar sinal. Para melhorar o desempenho das técnicas clássicas propõe-se o uso da relação sinal/ruído a priori ( $SNR_{prio}$ ). Esta modificação permite a redefinição dos filtros clássicos de redução de ruído, reduzindo-se pela primeira vez nesses métodos o fenômeno do ruído musical. Estende-se o uso da relação sinal/ruído a priori para os sistemas baseados na função de coerência, que apresentam como problema principal a fraca redução de ruído nas frequências baixas e a introdução de reverberação no sinal processado. Mostra-se que a função de coerência pode ser redefinida a partir da  $SNR_{prio}$ , melhorando a redução de ruído nas frequências baixas e eliminando o problema da reverberação. Os resultados finais, avaliados por meio de testes objetivos e subjetivos, mostram que estes novos métodos são superiores aos métodos existentes, satisfazendo as condições necessárias para uso nos sistemas rádio-móveis veiculares, isto é, eficácia na redução do ruído e baixa complexidade matemática.

## **CONTEÚDO**

<b>INTRODUÇÃO.....</b>	<b>01</b>
------------------------	-----------

### ***CAPÍTULO I - Ambiente Veicular e Problemas Associados aos Telefones Mãos-Livres***

<i>I.1- Introdução.....</i>	<i>04</i>
<i>I.2 - Ambiente veicular e o ruído.....</i>	<i>07</i>
<i>1.2.1 - Ruído do motor.....</i>	<i>08</i>
<i>1.2.2 - Ruído de rolamento dos pneus.....</i>	<i>08</i>
<i>1.2.3 - Ruído aerodinâmico.....</i>	<i>09</i>
<i>1.2.4 - Ruído do ambiente.....</i>	<i>09</i>
<i>I.3 - Características frequenciais do ruído obtido em um veículo.....</i>	<i>10</i>
<i>I.4 - Conclusões.....</i>	<i>12</i>

### ***CAPÍTULO II - Técnicas de Redução de Ruído em Sinais de Voz - Uma visão Geral***

<i>II.1 - Introdução.....</i>	<i>14</i>
<i>II.2 - Métodos que usam um único microfone para captar o sinal.....</i>	<i>15</i>
<i>II.2.1 - Subtração Espectral.....</i>	<i>15</i>
<i>II.2.2 - Filtragem de Wiener.....</i>	<i>18</i>
<i>II.2.3 - Filtragem baseada na máxima-verossimilhança.....</i>	<i>20</i>
<i>II.2.4 - Técnica de Ephraim &amp; Malah.....</i>	<i>23</i>
<i>II.2.5 - Filtragem tipo pente.....</i>	<i>27</i>
<i>II.2.6 - Outros métodos.....</i>	<i>29</i>
<i>II.3 - Métodos que usam dois microfones para captar o sinal.....</i>	<i>30</i>



II.3.1 - Filtragem adaptativa.....	31
II.3.2 - Uso da função de coerência - Método 1.....	32
II.3.2.1 - Função de Correção de Fase.....	35
II.3.2.2 - Cálculo da função de coerência.....	36
II.3.3 - Uso da função de coerência - Método 2.....	38
II.4 - Conclusões.....	39

**CAPÍTULO III - O uso da Relação Sinal Ruído no melhoramento de Sinais de Voz**

III.1 - Introdução.....	41
III.2 - Técnicas baseadas na relação sinal/ruído a posteriori.....	42
III.2.1 - Subtração Espectral.....	42
II.2.2 - Filtro de Wiener generalizado.....	44
II.2.3 - Estimação pela máxima-verossimilhança.....	45
III.3 - Análise da relação sinal/ruído (SNR) a posteriori.....	46
III.4 - Análise da relação sinal/ruído (SNR) a priori.....	48
III.5 - Técnicas clássicas de redução de ruído usando SNR a priori.....	50
III.6 - Função de coerência usando a SNR a priori.....	52
III.7 - Introdução de um modelo a dois estados.....	55
III.8 - Conclusões.....	59

**CAPÍTULO IV - Técnicas Auxiliares relacionadas à Redução de Ruído**

IV.1 - Introdução.....	60
IV.2 -Estimação usando a Transformada de Fourier.....	61
IV.3 - Detecção voz-silêncio (DVS).....	62
IV.3.1 - DVS usando a função de coerência.....	63
IV.3.2 - DVS usando um microfone auxiliar isolado do sinal de voz.....	66
IV.4 - Estimação da potência do ruído.....	68
IV.5 - Estimação das SNR a posteriori e a priori.....	69

IV.6 - Conclusões.....71

**CAPÍTULO V - Resultados Obtidos e Avaliações**

V.1 - Introdução.....72

V.2 - Avaliação objetiva usando a distância cepstral e o ganho na SNR temporal.....73

    V.2.1 - Distância cepstral - Teoria básica.....73

    V.2.2 - Ganho na SNR.....75

    V.2.3 - Obtenção dos sinais para as avaliações objetivas.....76

V.3 - Avaliação subjetiva.....77

V.4 - Resultados dos testes objetivos.....78

    V.4.1 - Sinal usado nos testes objetivos.....79

    V.4.2 - Resultados obtidos com a subtração espectral original.....80

    V.4.3 - Resultados obtidos com o método de Ephraim & Malah .....83

    V.4.4 - Resultados obtidos com a função de coerência original.....85

    V.4.5 - Resultados obtidos com a nova subtração espectral.....87

    V.4.6 - Resultados obtidos com a nova função de coerência.....93

V.5 - Avaliação dos testes objetivos.....95

V.6 - Resultados e avaliação dos testes subjetivos.....97

V.7 - Conclusões.....100

**CAPÍTULO VI - Comentários Finais.....102**

**REFERÊNCIAS BIBLIOGRÁFICAS.....105**

**APÊNDICE.....109**

## **INTRODUÇÃO**

Os estudos sobre redução de ruído em sinais de voz em sistemas de telecomunicações foram significativos nos anos 70. O objetivo principal sempre foi a melhoria do sinal nos aspectos físicos (recuperação da forma de onda original) e auditivo (inteligibilidade). A motivação para tais estudos é o vasto campo de aplicações possíveis.

Neste trabalho, a redução de ruído resulta em um aumento da inteligibilidade da voz e é denominada de *melhoramento de sinais de voz (MSV)*. Dentro de um contexto geral, a redução de ruído em sinais de voz pode ser aplicada em várias atividades, dentre as quais pode-se citar:

- *Sistemas telefônicos convencionais;*
- *Postos telefônicos usados em ambientes ruidosos, tais como escritórios;*
- *Restauração de gravações antigas (discos, fitas, etc.);*
- *Codificação de sinais de voz ruidosas;*
- *Reconhecimento de voz em transmissões ruidosas;*
- *Verificação de locutor em transmissões ou gravações ruidosas;*
- *Postos telefônicos móveis, como em veículos, trens, helicópteros, etc.*

Neste estudo, o objetivo principal é a análise e desenvolvimento de técnicas de melhoramento de sinais de voz para serem aplicadas nos sistemas telefônicos celulares instalados em veículos. Estes sistemas, aqui denominados de *Telefones Mãos-Livres (HFTs)*<sup>1</sup>, estão sendo desenvolvidos para evitarem-se restrições no uso da telefonia celular. Vem se generalizando em vários países a proibição do uso de sistemas celulares

---

<sup>1</sup> Abreviatura obtida do termo original em inglês "Hands-Free Telephones" (HFTs) e será usada ao longo deste texto.

convencionais em veículos por seus motoristas, com a justificativa de segurança. O motorista, ao conduzir um automóvel e usar um aparelho portátil para estabelecer uma comunicação, necessita de uma das mãos (e as vezes até das duas, como no caso da discagem) para utilizar o aparelho telefônico, provocando instabilidade na condução do automóvel. Uma primeira solução que pode ser cogitada é o uso de um sistema com fones de ouvido e microfone integrados acoplado à cabeça do motorista (como no caso dos equipamentos usados em aviões e helicópteros). Esta solução não agrada aos usuários por ser incômoda e desajeitada. Por estes motivos, as empresas querem investir em um sistema mais moderno, que é o caso dos *HFTs*. Nestes sistemas, microfone e alto-falante estão distanciados do locutor e mesmo o controle do aparelho (por exemplo, a discagem) é feito por comando de voz. No entanto, para se colocar em operação tal sistema, dois problemas, que afetam diretamente o interlocutor conectado ao telefone veicular, devem ser resolvidos: primeiro, o fato de distanciar-se o microfone do locutor no veículo provoca *uma queda acentuada na relação sinal/ruído (SNR) do sinal transmitido*; segundo, o uso de um alto-falante num ambiente veicular, que é altamente reverberante, provoca o surgimento de *ecos acústicos*. Estes problemas, em conjunto, inviabilizam uma comunicação com qualidade satisfatória. Assim, o desenvolvimento de terminais tipo *HFTs* passa por um estudo sobre *REDUÇÃO DE RUÍDO E ANULAÇÃO DE ECO ACÚSTICO*. No estudo desenvolvido e apresentado aqui, o objetivo principal é a redução do ruído, sendo o eco acústico considerado um problema isolado.

Dentro do estudo de melhoramento de sinais de voz, com aplicação nos sistemas telefônicos mãos-livres, são discutidos e propostos nesta tese métodos de redução de ruído que usam um ou dois microfones na captação do sinal a ser transmitido. O ruído presente no ambiente veicular, originário do próprio veículo e do ambiente onde se trafega, é tratado de forma global, mesmo se tendo ruídos predominantes, com o que é o caso do ruído do motor. A opção da globalização do ruído foi tomada tendo em vista a obtenção de técnicas de redução de ruído mais simples, eficientes e de aplicação imediata em sistemas que exijam operação em tempo real, como no caso dos *HFTs*. Com este objetivo, o texto foi estruturado como segue:

- No Capítulo I é apresentada a problemática do ruído no ambiente veicular. São apresentados alguns pontos críticos relacionados à captação do ruído, bem como os

problemas ligados à colocação de microfone(s) e alto-falante(s) distantes do locutor-ouvinte.

- No Capítulo II faz-se um resumo das principais técnicas de redução de ruído conhecidas, com destaque para aquelas possíveis de serem implementadas nos sistemas *HFTs*. Estas técnicas, baseadas na utilização de um ou dois microfones, têm seus aspectos teóricos destacados.
- No Capítulo III propõe-se um conjunto de técnicas para utilização nos sistemas *HFTs*. Os estudos foram centrados nas técnicas que usam um único microfone para captar o sinal, como subtração espectral, filtro de Wiener, filtros baseados nas estatísticas dos sinais de voz e ruído, etc. Inicialmente, analisa-se o comportamento dos parâmetros principais dessas técnicas, identificados como sendo a *relação sinal/ruído a posteriori* ( $SNR_{post}$ ) e a *relação sinal/ruído a priori* ( $SNR_{prio}$ ). Mostra-se a vantagem de utilizar-se a  $SNR_{prio}$  no lugar da  $SNR_{post}$  e propõem-se técnicas de redução de ruído baseadas na  $SNR_{prio}$ , quando um ou dois microfones são disponíveis para captar o sinal. Também neste capítulo, introduz-se o conceito de incerteza na avaliação do sinal a ser processado, melhorando-se o desempenho dos métodos de redução de ruído.
- No Capítulo IV, são discutidos alguns pontos relacionados ao MSV, onde a base é a *Estimação Espectral de Curto Prazo* através do uso da Transformada Discreta de Fourier<sup>2</sup>. Discute-se também técnicas de *detecção voz-silêncio* (*DVS*) e a estimação dos parâmetros necessários para implementação dos filtros estudados no capítulo III.
- No Capítulo V é feito um estudo comparativo entre as técnicas estudadas e propostas, usando-se medidas de distorção, de ganho na relação sinal/ruído, espectrogramas e análises subjetivas.
- Finalmente, no Capítulo VI são apresentados os comentários finais com relação aos resultados obtidos, discutindo-se novas possibilidades para um avanço nos estudos sobre redução de ruído em sinais de voz.

---

<sup>2</sup>Em inglês, a Transformada Discreta de Fourier quando aplicada na estimação de curto prazo, é denominada "Short Time Fourier Transform" (STFT).

## **CAPÍTULO I**

### ***AMBIENTE VEICULAR E PROBLEMAS ASSOCIADOS AOS TELEFONES MÃOS-LIVRES***

#### ***1.1 - Introdução***

A expansão dos sistemas de telefonia celular propiciou uma flexibilidade indiscutível nas comunicações pessoais. Um dos destaques desses sistemas é sua utilização em veículos. O uso do telefone celular dentro de um veículo abre um dos mercados mais promissores do setor. Se, do ponto de vista técnico, os problemas relacionados ao aparelho celular estão praticamente superados, do ponto de vista de segurança, os aparelhos usados atualmente nos veículos têm causado polêmica. O questionamento está no perigo envolvido quando uma pessoa dirige e usa ao mesmo tempo um telefone. Além de perder a concentração, nos sistemas atuais o condutor é obrigado a usar uma das suas mãos unicamente para o sistema telefônico, colocando em risco a si mesmo e todos que trafegam no mesmo ambiente (ruas, rodovias, etc.). Por razão de segurança, vários países têm proibido o uso de aparelhos celulares por pessoas dirigindo veículos. Para viabilizar o uso de telefones celulares em veículos, as empresas têm concentrado esforços no sentido de desenvolver um novo terminal específico para veículos, denominado *Telefone Mãos-Livres (HFT)*, mas que pode, também,

ser usado em outras circunstâncias (por exemplo, escritórios). Nos sistemas convencionais (fone ao ouvido e microfone à boca), os sinais transmitidos apresentam boa relação sinal/ruído e não existe um caminho de retorno do sinal do fone para o microfone. Já nos *HFTs*, microfone e fone devem ficar distantes do locutor, eliminando assim a necessidade de se utilizar as mãos. Porém, os prejuízos na qualidade do sinal são grandes, com uma queda acentuada da relação sinal/ruído (SNR) e um acoplamento forte entre fone e microfone.

A operação de um sistema *HFT* deve garantir, a princípio, a captação do sinal a ser transmitido com boa relação sinal/ruído. Além disso, do ponto de vista de recepção (no veículo), é preciso garantir-se uma difusão acústica do sinal recebido de forma a proporcionar-se uma boa escuta. Com microfone e alto-falante no mesmo ambiente, é impossível obter-se um sistema sem que haja acoplamento entre transmissão e recepção. Na figura I.1 tem-se um diagrama esquemático representando o ambiente veicular, onde observa-se alguns dos possíveis caminhos entre locutor-ouvinte, microfone e alto-falante. A posição típica do microfone é o pára-sol do condutor, sendo que outros lugares podem ser utilizados (por exemplo o painel). Com esta configuração, o sinal proveniente do alto-falante terá um forte acoplamento com o microfone, gerando eco para quem está do outro lado da linha. O sinal proveniente do locutor sofrerá um perda de amplitude e será refletido no interior do veículo, chegando no microfone com um pequeno índice de reverberação e com uma SNR baixa. O resultado é um sinal bem degradado e que será praticamente ininteligível para o comunicador conectado ao veículo.

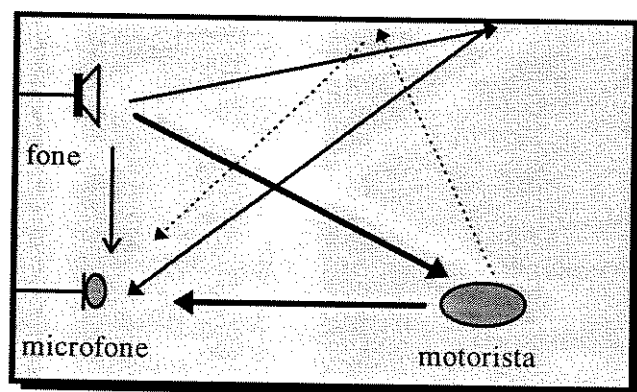
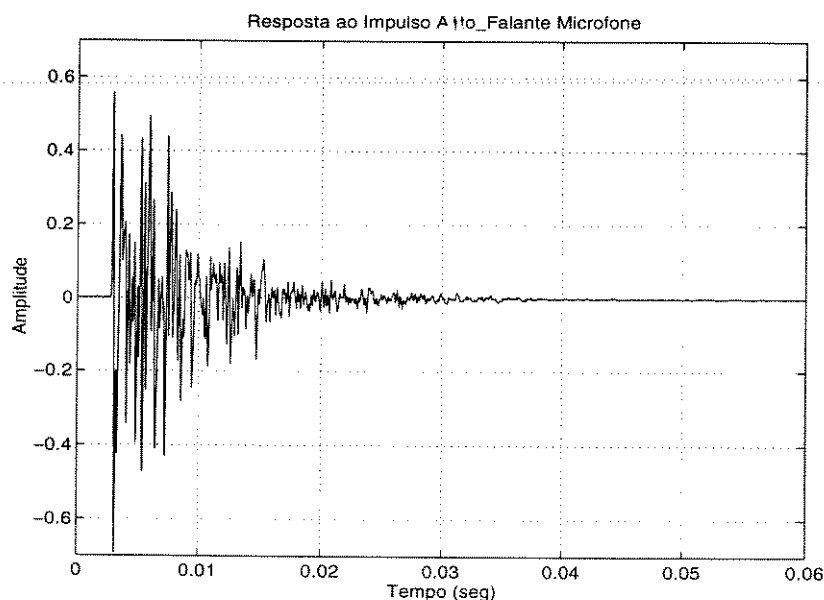


Figura I.1 - Ambiente veicular típico

Para ilustrar alguns dos problemas mencionados anteriormente, na figura I.2 apresenta-se a resposta ao impulso entre alto-falante e microfone. Observa-se uma resposta longa e com amplitude importante. Estas características permitem que o sinal proveniente do alto-falante chegue ao microfone praticamente com a mesma amplitude que chega ao motorista, obtendo-se assim um forte eco que pode inviabilizar uma comunicação. Na figura I.3 tem-se a resposta ao impulso medida entre locutor e microfone (colocado no painel). O principal interesse é verificar o atraso na captação do sinal pelo microfone. Quanto maior o atraso, maior será a queda da SNR do sinal captado pelo microfone. No caso da reverberação, a baixa amplitude da resposta ao impulso indica que ela pode ser desprezada.

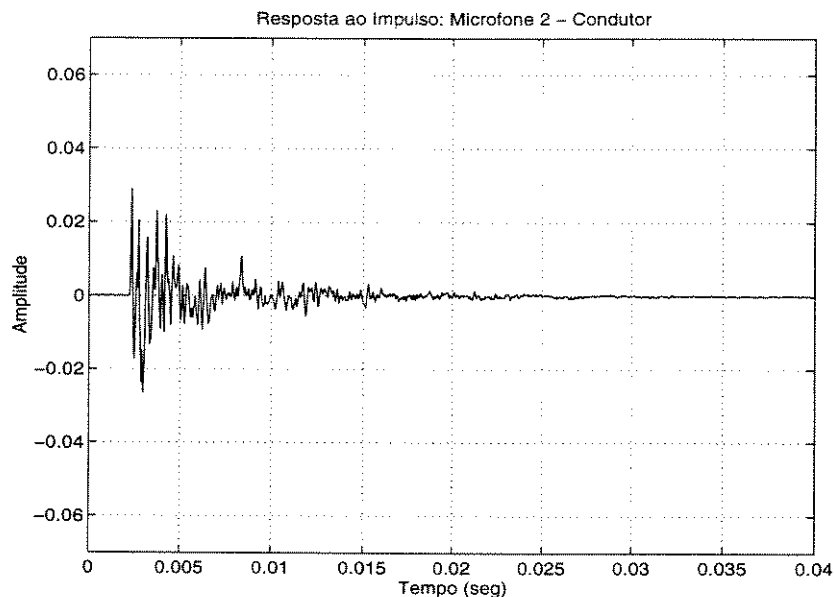
As medidas e as análises realizadas na obtenção das respostas ao impulso foram feitas no CNET em estudos anteriores [SCALART, 1994]. A técnica usada é baseada na análise da correlação-cruzada entre uma sequência pseudo-aleatória transmitida e sua equivalente recebida pelo microfone. Nestas respostas estão incluídos todos os efeitos inerentes aos dispositivos utilizados (amplificadores, filtros, microfones, etc.). Nas medições o motorista foi simulado por um boneco especial com um pequeno alto-falante posicionado na "boca". Este boneco representa bem os efeitos da cabeça e tórax de uma pessoa<sup>1</sup>.



*Figura I.2 - Resposta ao impulso entre alto-falante e microfone*

<sup>1</sup> O CNET possui um banco de dados completo sobre tais problemas e estas figuras foram gentilmente cedidas, já que elas fazem parte de relatórios internos não-publicáveis.





*Figura I.3 - Resposta em ao impulso entre Condutor e Microfone*

## ***I.2 - O Ambiente veicular e o ruído (LESCO, 1992)***

Três componentes básicas são responsáveis pela formação do ruído no ambiente veicular: O ruído do motor, o ruído do atrito dos pneus (ou ruído de rolamento) e o ruído do vento. Também, considerando o fraco isolamento acústico dos veículos, tem-se o ruído do ambiente onde se trafega. Estes ruídos têm características predominantemente estacionárias, o que facilita a redução dos seus efeitos. No entanto, deve-se prever uma parcela de ruído não-estacionária, proveniente do cruzamento entre veículos. A eliminação deste tipo de ruído é

extremamente difícil e normalmente ele deve ser considerado como sinal útil<sup>2</sup> para evitar maiores distorções no sinal melhorado.

### ***1.2.1 - Ruído do motor***

Característico de cada tipo de veículo, o ruído do motor contribui fortemente na composição final do ruído. Uma parte significativa é resultante da rotação do motor, contribuindo de acordo com a frequência fundamental e suas harmônicas, que estão associadas ao número de cilindros. Quando o motor entra em operação e o veículo parte em movimento existirá um ruído adicional proveniente das seguintes fontes: caixa de marchas, sistema de transmissão, escapamento e sistema de refrigeração. O ruído resultante tem características estacionárias, é predominante nas baixas frequências, na faixa de 0 a 1500 Hz, e sua potência dependerá da velocidade do veículo. Para uma velocidade média de 100 km/h, por exemplo, seu nível sonoro e sua potência são importantes.

### ***1.2.2 - Ruído de rolamento dos pneus***

O contato dos pneus com o solo gera um ruído tipicamente estacionário e sua potência dependerá de vários fatores. Inicialmente, deve-se dizer que a fabricação de um pneu é baseada em diferentes requisitos que no final buscam a segurança. Assim, é preciso definir-se o seu coeficiente de atrito, sua largura, sua rugosidade, etc. Estas definições dependerão também do tipo de rodovia, das condições climáticas, etc. O ruído resultante dependerá da

---

<sup>2</sup> O cálculo da potência média do ruído é normalmente feito considerando-se várias janelas de sinal onde não se tem voz. Se um ruído não-estacionário for incluído nesse cálculo, ele permanecerá influenciando a média durante um certo tempo, mesmo não existindo mais. Tomando-o como sinal útil, ele é eliminado da média e processado como voz. Nesse caso, apenas as componentes estacionárias serão eliminadas.

rugosidade do pneu, do tipo de rodovia e da velocidade do veículo. A potência obtida é significativa e o ruído é predominante na faixa de 0 a 3000 Hz.

### ***1.2.3 - Ruído aerodinâmico***

A penetração do veículo dentro da atmosfera resulta num forte atrito entre o ar e a carroceria. O ruído resultante deste atrito dependerá da velocidade do veículo e da força e direção do vento. Porém, considerando a dificuldade de obter-se uma carroceria hermética, um outro ruído proveniente deste atrito pode aparecer sob a forma de "assobio". Tendo que o ruído aerodinâmico é tipicamente estacionário, o "assobio" contribuirá para o fortalecimento de uma faixa estreita do espectro. De uma forma geral, o ruído aerodinâmico apresenta uma potência relativamente fraca em relação às outras componentes de ruído (considerando um veículo com portas e janelas bem fechadas).

### ***1.2.4 - Ruído do ambiente***

O lugar físico por onde trafega o veículo (ruas, avenidas, rodovias, etc.) pode ser definido como uma fonte de ruído variável, com características espectrais as mais diversas. Estas variações dependem de vários fatores, podendo-se citar como exemplos: quantidade de veículos, largura da rodovia, tipo de construções, etc. O problema mais sério vem do cruzamento de veículos (ruas e avenidas de mão-dupla) e de ultrapassagens (ruas e avenidas de mão-única). O efeito sobre o sinal captado pelo microfone é a incorporação de um ruído não-estacionário, que é mais acentuado quanto se trafega em ruas estreitas com prédios altos. Como citado anteriormente, este tipo ruído é geralmente processado como sinal útil, reduzindo-se apenas o ruído de fundo estacionário.

### ***1.3 -Características freqüenciais do ruído obtido em um veículo***

A caracterização espectral individual dos ruídos gerados no ambiente veicular móvel é algo que exige alguns cuidados (por exemplo, desligar o motor a uma certa velocidade para gravar o ruído proveniente dos pneus). Considerando que as informações dadas no item anterior (que foram obtidas de relatórios internos do CNET) são suficientes para dar uma idéia geral de cada ruído isoladamente, neste item é mostrado como o espectro do ruído varia com a velocidade e o tipo de revestimento de uma rodovia.

Na figura I.4 têm-se os espectros de dois ruídos obtidos em um mesmo veículo, numa mesma rodovia e em duas velocidades. Observa-se que o efeito do aumento de velocidade é um aumento da potência do ruído praticamente em toda a faixa de freqüência. Isto é esperado visto que o aumento de velocidade afeta todas as componentes geradoras do ruído (ruído de rolamento dos pneus, aerodinâmico, do motor, etc.).

Na figura I.5, têm-se os espectros de outros dois ruídos gravados com a mesma velocidade mas em rodovias diferentes. No primeiro caso, a rodovia apresenta um asfalto liso e no segundo caso, o asfalto é rugoso. Verifica-se que o ruído resultante será tanto mais forte nas freqüências baixas quanto maior for a rugosidade do asfalto. Isto significa que o ruído nas freqüências baixas tem uma forte influência do ruído proveniente rolamento dos pneus. O fortalecimento de uma faixa de freqüência indica uma ressonância entre o ruído do rolamento dos pneus e a carroceria do veículo. A pequena variação do ruído nas freqüências altas é facilmente explicada, visto que a manutenção de uma mesma velocidade implica em ruídos aerodinâmicos e de motor constantes.

As figuras I.4 e I.5 mostram que o ruído resultante em um veículo em movimento tem uma forte potência nas baixas freqüências (0 - 1500) Hz, com uma queda acentuada nas altas<sup>3</sup>. O grande problema é que o sinal de voz também tem sua composição mais importante

---

<sup>3</sup> Os espectros de ruído apresentados foram obtidos usando-se um sinal amostrado a 16 kHz, FFT de 512 pontos e janelamento de Hanning com sobreposição de 50%. Usou-se um trecho de ruído com 140 ms de duração, valor típico usado para estimar o ruído em sistemas práticos.

nas frequências baixas (500 a 2500) Hz. Além do mais, o ruído mais incômodo vem das baixas frequências. O resultado é um sinal difícil de ser melhorado sem que se introduzam distorções na voz original. Assim, um dos compromissos que deve ser considerado no melhoramento do sinal de voz é o nível de redução de ruído com relação às distorções introduzidas pelas técnicas utilizadas.

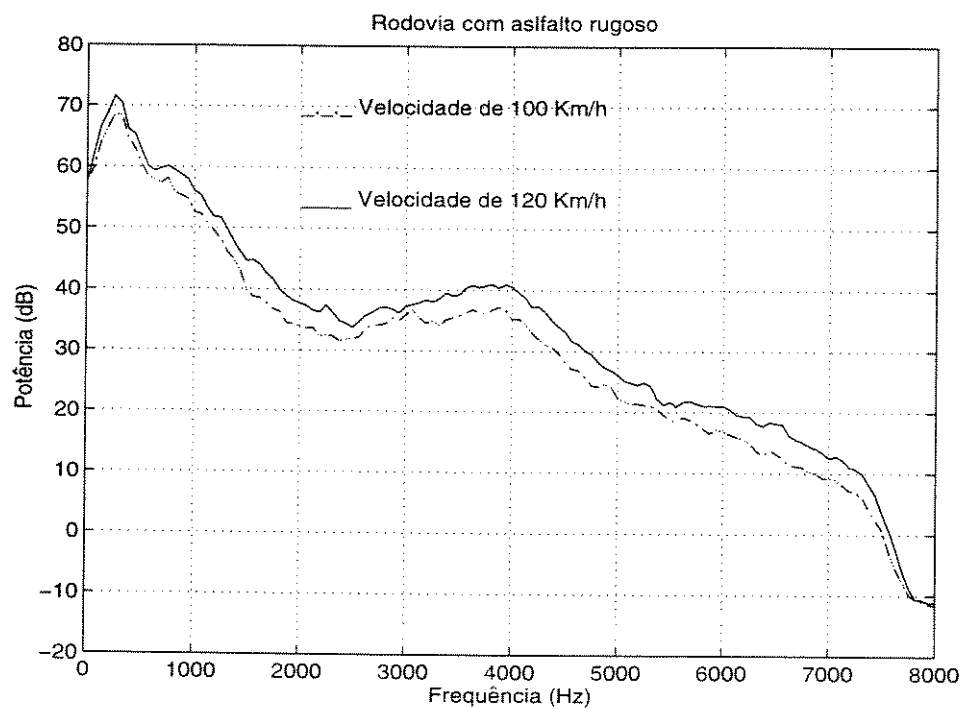


Figura I.4 - Espectro do Ruído - mesma rodovia, velocidades diferentes

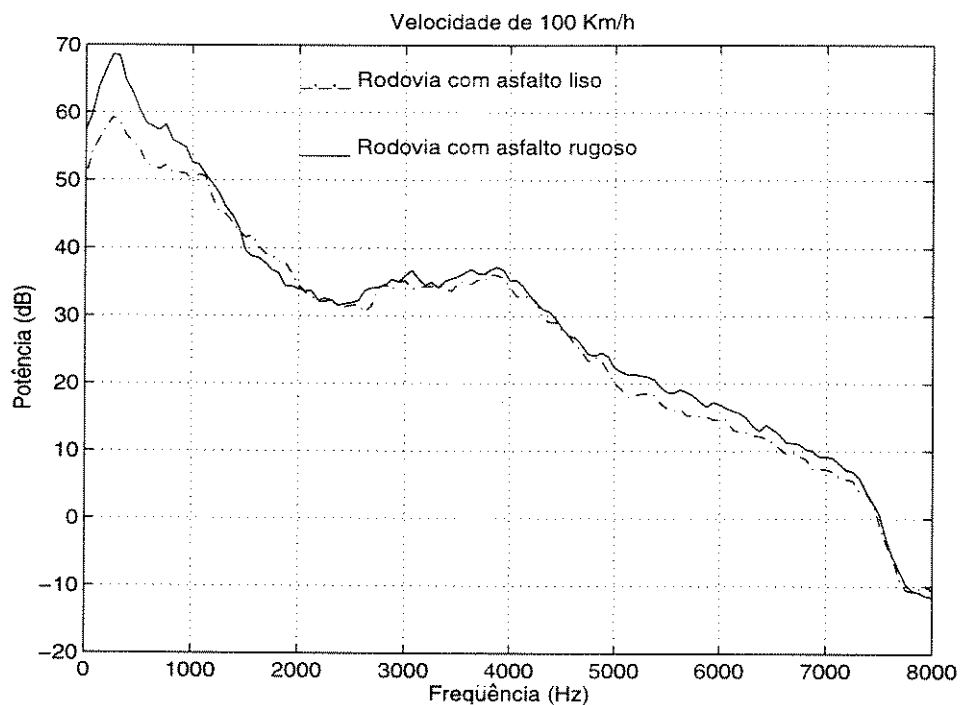


Figura I.5 - Espectro do Ruído - mesma velocidade (100 km/h), rodovias diferentes

#### I.4 - Conclusões

Apresentou-se neste capítulo uma visão geral dos problemas envolvidos no ambiente veicular. Mostrou-se que o uso de um terminal telefônico especial para veículos, os HFTs, apresenta muitas dificuldades, sendo que os dois principais problemas a serem atacados são o *eco acústico* e o *ruído*. Como o objetivo deste trabalho se restringe ao estudo de técnicas para a redução do ruído, mostrou-se que, nesse sentido, o problema maior é que o ruído obtido em um ambiente veicular apresenta forte concentração de energia nas frequências baixas, como ocorre também com os sinais de voz. Mostrou-se que a distribuição do espectro do ruído

depende muito do tipo de piso da rodovia (influenciando as frequências baixas) e que o aumento de velocidade do veículo provoca um aumento na sua potência. Enfim, tem-se como resultado um sinal ruidoso difícil de ser melhorado, tendo em vista que uma forte redução de ruído pode implicar numa forte distorção no sinal de voz processado. É com estas informações que estudaram-se e desenvolveram-se técnicas para serem implementadas nos sistemas HFT, e que serão apresentadas ao longo deste trabalho.

## **CAPÍTULO II**

### **TÉCNICAS DE REDUÇÃO DE RUÍDO EM SINAIS DE VOZ**

#### **- Uma Visão Geral -**

##### **II.1 - Introdução**

A redução de ruído em sinais de voz tem um vasto campo de aplicação e começou a ser explorada de maneira mais intensa na década de 70. No entanto, já nos anos 40, EGAN & WIENER (1946) publicaram trabalhos científicos enfocando o problema da inteligibilidade dos sinais de voz. Na época, com aplicações voltadas para as telecomunicações, eles usavam o efeito de mascaramento do ruído. Com a continuação dos estudos, as técnicas de melhoramento de sinais de voz (MSV) evoluíram consideravelmente e hoje pode-se dividi-las em três grupos básicos:

- 1. MSV usando estimação da amplitude espectral de curto prazo;*
- 2. MSV explorando a característica de periodicidade dos sinais de voz;*
- 3. MSV baseada em um modelo de produção para os sinais de voz.*



O MSV, para ser aplicado nos sistemas rádio-móveis veiculares, tem que satisfazer dois critérios básicos: simplicidade matemática e eficiência. Uma rápida análise nos grupos apresentados anteriormente mostra que as técnicas com base nos grupos 2 e 3, que envolvem respectivamente detecção de período de “pitch” (frequência fundamental do sinal de voz) e análise LPC (Linear Predictive Coding) na presença de ruído, são muito complexas. Para as técnicas baseadas no grupo 1, a principal vantagem é a facilidade de implementação, tornando-as atrativas para uso nos HFTs (YANG, 1993). Nessas técnicas, a ferramenta básica é a Transformada de Fourier de curto prazo (STFT)<sup>1</sup>. Neste capítulo apresenta-se uma visão geral das técnicas de MSV, usando-se um ou dois microfones para captação do sinal a ser processado.

## ***II.2 - Métodos que usam um único microfone para captar o sinal***

### ***II.2.1 - Subtração espectral***

Considerando os sinais de voz e ruído como processos aleatórios estacionários e independentes, a técnica baseada na subtração espectral utilizada por BOLL (1979) foi uma das primeiras a oferecer possibilidades reais de implementação prática. Seu principal atrativo é a simplicidade matemática.

Seja, por exemplo, um sinal de voz puro,  $v(t)$ , degradado por um ruído aditivo  $r(t)$ , formando um sinal ruidoso  $y(t)$  como apresentado na equação II.1:

$$y(t) = v(t) + r(t) \quad (\text{II.1})$$

No domínio da frequência tem-se

---

<sup>1</sup> A STFT é uma ferramenta básica na implementação de sistemas de redução de ruído e seu uso é apresentado no capítulo IV.

$$Y(\omega) = V(\omega) + R(\omega) \quad (\text{II.2})$$

Calculando-se o quadrado na equação II.2 obtém-se:

$$|Y(\omega)|^2 = |V(\omega)|^2 + V(\omega)^* \cdot R(\omega) + V(\omega) \cdot R(\omega)^* + |R(\omega)|^2 \quad (\text{II.3})$$

Usando a hipótese de que o ruído é aditivo e descorrelacionado com o sinal de voz, da equação II.3 tem-se:

$$|V(\omega)|^2 = |Y(\omega)|^2 - |R(\omega)|^2 \quad (\text{II.4})$$

Das equações II.1 e II.4, verifica-se que a aplicação da técnica baseada na subtração espectral (SE) passa pela estimação do espectro do ruído. Além disso, observa-se que não existe uma recuperação da fase do sinal. Na verdade, não se conhece procedimentos para a recuperação da fase do sinal de voz original e utiliza-se nesses casos a fase do sinal ruidoso com bons resultados<sup>2</sup>. Assim, o sinal de voz estimado será dada por<sup>3</sup>:

$$\hat{V}(\omega) = |A(\omega)| \cdot \frac{Y(\omega)}{|Y(\omega)|} \quad (\text{II.5})$$

onde:

$$|A(\omega)| = \begin{cases} \sqrt{|Y(\omega)|^2 - |\hat{R}(\omega)|^2} & \text{se } |Y(\omega)|^2 \geq |\hat{R}(\omega)|^2 \\ 0 & \text{caso contrário} \end{cases} \quad (\text{II.6})$$

com  $|\hat{R}(\omega)|^2$  sendo a potência média estimada do ruído.

<sup>2</sup> Estudos relacionados à inteligibilidade (FLANAGAN, 1972) mostram que o ouvido não discerne bem as distorções de fase.

<sup>3</sup> O filtro obtido foi baseado nas potências dos sinais envolvidos. Pode-se também obter um filtro diretamente das amplitudes (equação II.2) (VIEIRA-FILHO et alii, 1995).

Da equação II.6, verifica-se que o filtro obtido pelo método da SE só é realizável se a potência estimada do ruído for menor ou igual à potência do sinal ruidoso. No entanto, considerando o fato de se ter sinais com fases aleatórias, isto não é garantido. Esta aleatoriedade da fase pode, no processo de adição dado na equação II.2, resultar na formação de um sinal ruidoso de potência inferior a do ruído. A “retificação de meia-onda” representada pela equação II.6 é uma das “soluções” para este problema. Poder-se-ia da mesma maneira tomar o valor absoluto na equação II.4. Qualquer que seja a solução adotada, o resultado será uma alteração aleatória das amplitudes nas frequências onde isto ocorre, acarretando a geração de tons indesejáveis no sinal processado. Esses tons são denominados na literatura de *RUÍDO MUSICAL* e é um dos maiores inconvenientes na aplicação da subtração espectral e de outras técnicas (CAPPÉ, 1994 e EPHRAIM, 1992).

A figura II.1 resume o uso da subtração espectral aplicada ao MSV, destacando-se a reutilização da fase do sinal ruidoso (equação II.5) na reconstrução do sinal processado, como citado anteriormente.

Vários trabalhos têm explorado este método (AGUIAR, 1989; VASEGHI & FRAYLING, 1992 e HÄNDEL, 1995) e o objetivo é obter um processamento sem o inconveniente do ruído musical, o que tornaria esta técnica muito atrativa, devido à baixíssima complexidade computacional envolvida. Para se obter uma maior redução do ruído, utiliza-se muito uma superestimação do ruído de fundo (LIM & OPPENHEIM, 1979). O objetivo nesses casos é reduzir ao máximo a potência média do ruído. Entretanto, este processo pode aumentar ainda mais o problema<sup>4</sup> do ruído musical.

---

<sup>4</sup> Quanto maior a potência do ruído maior será a possibilidade de se ter na equação II.4 um resultado menor que ZERO. Assim, maior será a frequência nas retificações necessárias para viabilizar o filtro.

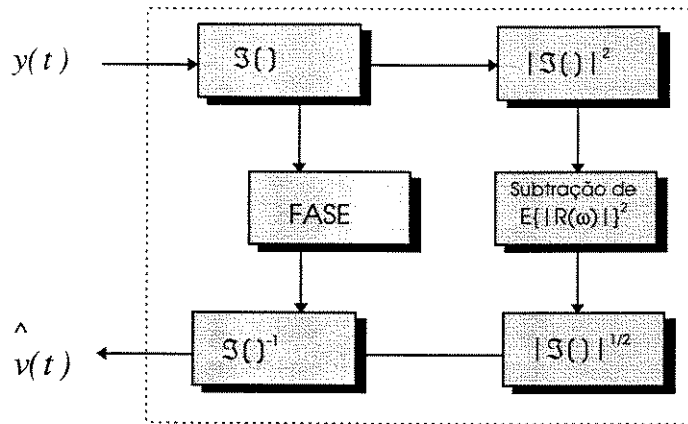


Figura II.1 - Esquema Básico da Subtração Espectral

## II.2.2 - Filtragem de WIENER

Um método alternativo para a redução do ruído, que generaliza o método da subtração espectral, é o Filtro de Wiener. Muito utilizado em diferentes aplicações de processamento de sinais, o filtro de Wiener é baseado no princípio da minimização do erro quadrático médio e nesse sentido realiza uma filtragem ótima. Sua aplicação parte do princípio de que tem-se disponível o espectro do sinal original e do ruído.

Da equação II.1, e usando as mesmas notações do item anterior, o sinal de voz estimado é resultante do sinal ruidoso passado por um filtro ótimo, isto é,

$$\hat{V}(\omega) = H_w(\omega) \cdot Y(\omega) \quad (\text{II.7})$$

Considerando os espectros de potência (estimados) dos sinais envolvidos, como na equação II.4, o filtro de Wiener, não causal, é dado por:

$$H_w(\omega) = \frac{\phi_{vv}(\omega)}{\phi_{vv}(\omega) + \phi_{rr}(\omega)} \quad (\text{II.8})$$

onde  $\phi_{vv}(\omega)$  e  $\phi_{rr}(\omega)$  são as densidades espectrais de potência do sinal de voz puro e do ruído (tidos como processos aleatórios estacionários e independentes).

A obtenção da resposta do filtro de Wiener, dado na equação II.8, pode ser realizada de diferentes formas (DELLER et alii, 1993)]. Uma solução padrão, também explorada na subtração espectral, é estimar a potência do ruído nos intervalos de silêncio (ausência de voz) do sinal ruidoso. Para resolver a não-causalidade do filtro, duas soluções básicas podem ser utilizadas: numa primeira, o cálculo do filtro pode ser baseado no resultado do espectro de voz estimado na janela de análise precedente<sup>5</sup>. Uma segunda possibilidade é obter-se uma estimação do sinal de voz de acordo com a equação II.5 (subtração espectral). Independente do método adotado, o filtro de Wiener pode ser reescrito de uma maneira mais geral, considerando a estimação das potências dos sinais de voz e ruído, dada por:

$$H_w(\omega) = \left\{ \frac{|\hat{V}(\omega)|^2}{|\hat{V}(\omega)|^2 + \alpha \cdot |\hat{R}(\omega)|^2} \right\}^\beta \quad (\text{II.9})$$

Considerando o uso da subtração espectral para resolver o problema da não-causalidade do filtro, deve-se ter  $|\hat{V}(\omega)|^2 = |\hat{Y}(\omega)|^2 - |\hat{R}(\omega)|^2$ .

Na equação II.9, o fator  $\alpha$  é usado para permitir um controle do nível de redução de ruído, considerando que a potência do ruído estimada pode estar abaixo do seu valor real. O fator  $\beta$  permite a obtenção de uma variedade de sub-filtros de Wiener. Normalmente, tem-se  $0 < \beta \leq 1$  e  $\alpha \geq 1$ .

Da equação II.9, fazendo-se  $\beta = 1/2$  e  $\alpha = 1$ , obtém-se o filtro baseado na subtração espectral. O fator  $\alpha$  também pode ser aplicado diretamente na SE.

Para se ter uma idéia do potencial de atenuação desta técnica, na figura II.2 são apresentadas as curvas teóricas obtidas da equação II.9 para SE ( $\beta=1/2$  e  $\alpha=1$ ) e Wiener

---

<sup>5</sup> O janelamento é típico na estimação espectral usando STFT.

Padrão ( $\beta=\alpha=1$ ). Na mesma figura apresenta-se também a curva de atenuação para a SE baseada diretamente na equação II.2, o que é denominado SE em AMPLITUDE (VIEIRA-FILHO et alii, 1995). Observa-se que o filtro de Wiener apresenta uma atenuação intermediária em relação à SE em potência e a SE em amplitude. A escolha de um ou de outro filtro deve ser feita considerando-se o compromisso entre redução de ruído efetiva e distorções resultantes do processo de redução do ruído. Uma forte atenuação pode gerar fortes distorções no sinal processado. Nesse sentido, a SE em amplitude, que permite uma forte atenuação, pode não ser interessante.

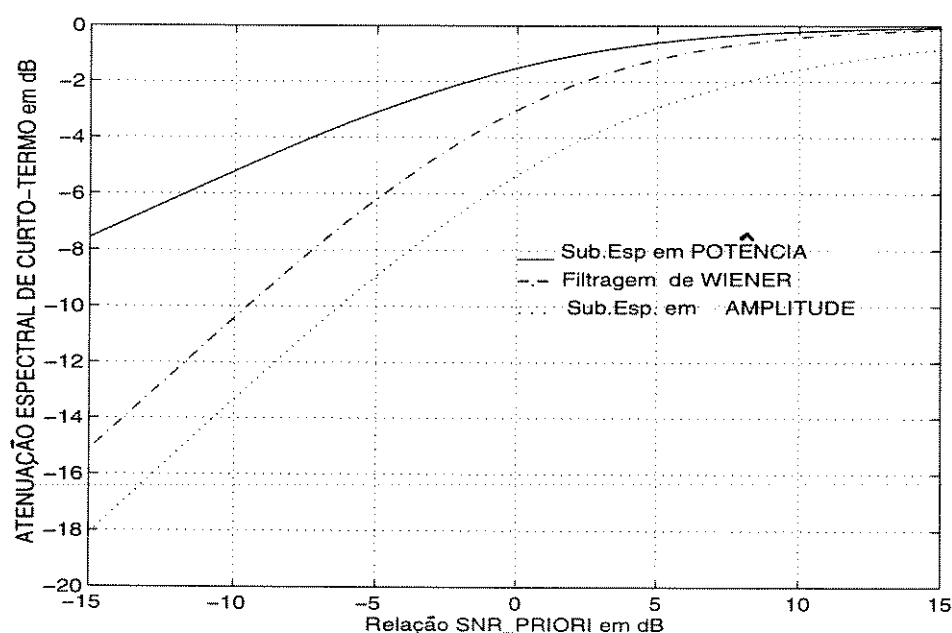


Figura II.2 - Curvas teóricas de redução de ruído para a SE e filtro de Wiener

### II.2.3 - Filtragem baseada na máxima-verossimilhança

Nesta técnica, os sinais de voz e ruído são modelados como processos aleatórios gaussianos e independentes. Aplica-se então o critério da máxima-verossimilhança<sup>6</sup>.

<sup>6</sup> Maximização da função densidade de probabilidade que descreve estatisticamente um determinado evento.

Para um sinal de voz puro, pode-se usar a seguinte representação:

$$V(\omega) = |V(\omega)| \cdot e^{j\alpha(\omega)} \quad (\text{II.10})$$

onde  $|V(\omega)|$  determina a envoltória do sinal e  $\alpha(\omega)$  sua fase.

Para o ruído aditivo, a função densidade de probabilidade é dada por:

$$p[Y(\omega)|V(\omega), \alpha(\omega)] = \frac{1}{\pi \cdot \sigma_R^2(\omega)} \cdot \exp\left[-\frac{|R(\omega)|^2}{\sigma_R^2(\omega)}\right] \quad (\text{II.11})$$

Considerando a formação do sinal ruidoso como dada na equação II.2 tem-se  $|R(\omega)|^2 = |Y(\omega)|^2 - 2 \cdot |V(\omega)| \cdot \Re[e^{-j\alpha(\omega)} \cdot Y(\omega)] + |V(\omega)|^2$ . A variável  $\sigma_R^2(\omega)$  é a potência média do ruído em cada frequência analisada.

A substituição de  $|R(\omega)|^2$  na equação II.11 mostra a existência de um efeito da fase, que pode ser eliminado através de uma maximização do valor médio da função de verossimilhança, dado por:

$$\overline{p[Y(\omega)|V(\omega)]} = \int_0^{2\pi} p[Y(\omega)|V(\omega), \alpha(\omega)] \cdot p[\alpha(\omega)] \cdot d\alpha \quad (\text{II.12})$$

onde  $p[\theta(\omega)]$  representa a função densidade de probabilidade da fase. Considerando esta distribuição como sendo uniforme entre 0 e  $2\pi$ , tem-se que o valor médio da equação II.12 é dado por:

$$\overline{p[Y(\omega)|V(\omega)]} = \frac{1}{\pi \sigma_R^2(\omega)} \exp\left[\frac{|Y(\omega)|^2 - |V(\omega)|^2}{\sigma_R^2(\omega)}\right] \cdot \frac{1}{2\pi} \int_0^{2\pi} \exp\left[\frac{2 \cdot |V(\omega)| \cdot \Re[e^{-j\alpha(\omega)} \cdot Y(\omega)]}{\sigma_R^2(\omega)}\right] \cdot d\alpha \quad (\text{II.13})$$

A integral na equação anterior é conhecida como função de Bessel modificada do primeiro tipo e de ordem ZERO, e é dada por:

$$Io(|X(\omega)|) = \frac{1}{2\pi} \cdot \int_0^{2\pi} \exp\left\{\Re\left[e^{-j\alpha(\omega)} \cdot X(\omega)\right]\right\} \cdot d\alpha \quad (\text{II.14})$$

onde  $X(\omega)$  é dado por:

$$X(\omega) = \frac{2 \cdot |V(\omega)| \cdot Y(\omega)}{\sigma_R^2(\omega)} \quad (\text{II.15})$$

Para grandes valores de  $|X(\omega)|$  (maiores do que 3), a função de Bessel modificada pode ser aproximada por

$$Io(|X|) \equiv \frac{1}{\sqrt{2\pi|X|}} \exp(|X|) \quad (\text{II.16})$$

Assim, substituindo a equação II.16 em II.13 e maximizando esta última, chega-se ao seguinte estimador de máxima-verossimilhança para a envoltória do sinal de voz:

$$\left|\hat{V}(\omega)\right| = \frac{1}{2} \cdot \left[|Y(\omega)| + \sqrt{|Y(\omega)|^2 - \sigma_R^2(\omega)}\right] \quad (\text{II.17})$$

Na equação II.17, dividindo tudo por  $|Y(\omega)|$ , observa-se que o filtro resultante terá uma atenuação de 0,5 mais os valores obtidos na estimação. Assim, a máxima atenuação possível com este método é 6 dB<sup>7</sup>. Este valor é muito fraco quando comparado com a SE ou com o filtro de Wiener., indicando uma baixa performance na redução do ruído.

---

<sup>7</sup> Observando-se a equação II.2 e II.17, verifica-se que o método da máxima-verossimilhança tem a mesma curva de atenuação da subtração espectral em potência deslocada de 6 dB acima (figura II.2)



Como nos métodos baseados na subtração espectral, a fase do sinal ruidoso é usada para recuperar o sinal processado e, para a realização do filtro, deve-se ter  $|Y(\omega)|^2 - \sigma_R^2(\omega) \geq 0$ .

A técnica aqui apresentada foi aplicada por McAULAY & MALPASS (1980), que também introduziram o conceito de *incerteza na avaliação do sinal de voz*, melhorando seu desempenho. Essa modificação será discutida no capítulo III.

#### **II.2.4 - Técnica de Ephraim & Malah**

Nesta técnica, desenvolvida por EPHRAIM & MALAH (1984), os sinais de voz e ruído também são modelados estatisticamente como processos aleatórios estacionários e independentes. Eles consideram que os coeficientes da expansão de Fourier são variáveis aleatórias gaussianas estatisticamente independentes. A base matemática da modelagem é o teorema do limite central, considerando que cada coeficiente de Fourier é, no final, uma soma ponderada (ou integral) de variáveis aleatórias resultantes de um grande número de amostras do processo. Assume-se que os processos possuem média igual a ZERO.

De acordo com a equação II.2, os sinais de voz puro e ruidoso podem ser escritos como segue:

$$V(\omega) = |V(\omega)| \cdot e^{j\alpha(\omega)} \quad e \quad Y(\omega) = |Y(\omega)| \cdot e^{j\theta(\omega)} \quad (\text{II.18})$$

O objetivo é estimar o espectro de amplitude do sinal de voz  $|V(\omega)|$  a partir do sinal ruidoso  $y(t)$ , dentro do intervalo de observação  $[0 - T]^8$ .

---

<sup>8</sup> *Intrinsecamente relacionado ao janelamento do sinal ruidoso*

Considerando que as componentes espectrais são estatisticamente independentes, obtém-se então o estimador MMSE diretamente do sinal ruidoso, ou seja,

$$\left| \hat{V}(\omega) \right| = E\{V(\omega) | y(t), 0 \leq t \leq T\} = E\{V(\omega) | Y_0, Y_1, Y_2, \dots\} = E\{V(\omega) | Y_k\} \quad (\text{II.19})$$

onde

$Y_k = \text{amostra de } |Y(\omega)|$  e  $k$  representa a frequência específica analisada dentro do intervalo observado.

Da equação II.19 tem-se :

$$\left| \hat{V}(\omega) \right| = \frac{\int_0^\infty \int_0^{2\pi} v(\omega) \cdot p[Y(\omega)/v(\omega), \alpha(\omega)] \cdot p[v(\omega), \alpha(\omega)] \cdot d\alpha(\omega), dv(\omega)}{\int_0^\infty \int_0^{2\pi} p[Y(\omega)/v(\omega), \alpha(\omega)] \cdot p[v(\omega), \alpha(\omega)] \cdot d\alpha(\omega), dv(\omega)} \quad (\text{II.20})$$

onde  $p(\cdot)$  representa a função densidade de probabilidade e  $v(\omega)$  representa a variável aleatória do espectro de amplitude  $V(\omega)$  do sinal de voz.

Assumindo o modelo gaussiano tem-se:

$$p[Y(\omega) | v(\omega), \alpha(\omega)] = \frac{1}{\pi \sigma_R^2(\omega)} \cdot \exp \left\{ -\frac{1}{\sigma_R^2(\omega)} \left| Y(\omega) - v(\omega) \cdot e^{j\alpha(\omega)} \right|^2 \right\} \quad (\text{II.21A})$$

e

$$p[v(\omega), \alpha(\omega)] = \frac{v(\omega)}{\pi \sigma_v^2(\omega)} \cdot \exp \left\{ -\frac{v^2(\omega)}{\sigma_v^2(\omega)} \right\} \quad (\text{II.21B})$$

Nas equações II.21A e II.21B,  $\sigma_v^2(\omega) = E\{|V(\omega)|^2\}$  e  $\sigma_R^2(\omega) = E\{|R(\omega)|^2\}$ , representam as variâncias do sinal de voz e do ruído, respectivamente.

Substituindo-se as equações II.21A e II.21B na equação II.20, chega-se ao seguinte estimador:

$$\left| \hat{V}(\omega) \right| = \Gamma(1,5) \cdot \frac{\sqrt{\eta(\omega)}}{\gamma(\omega)} \cdot M(-0,5; 1,0; -\eta(\omega)) \cdot |D(\omega)| \quad (\text{II.22})$$

onde  $\Gamma(\cdot)$  representa a função gama e  $M(a; c; x)$  representa a função hipergeométrica (MIDDELTON, 1960).

Tem-se também que:

$$\eta(\omega) = \frac{\xi(\omega)}{1 + \xi(\omega)} \cdot \gamma(\omega), \text{ com } \xi(\omega) = \frac{\sigma_v^2(\omega)}{\sigma_R^2(\omega)} \text{ e } \gamma(\omega) = \frac{|Y(\omega)|^2}{\sigma_R^2(\omega)} \quad (\text{II.23})$$

A equação II.22 fornece os dois termos mais importantes do estimador desenvolvido por EPHRAIM & MALAH (1984), isto é,  $\xi(\omega)$  e  $\gamma(\omega)$ , que representam as relações sinal/ruído a *PRIORI* e a *POSTERIORI*, respectivamente. Vale lembrar que estes termos foram introduzidos inicialmente por McAULAY & MALPASS (1980).

A equação II.23 pode ainda ser desenvolvida, pois a função hipergeométrica pode ser escrita em termos de funções de Bessel<sup>9</sup>. Neste caso tem-se:

$$\left| \hat{V}(\omega) \right| = \Gamma(1,5) \cdot \frac{\sqrt{\eta(\omega)}}{\gamma(\omega)} \cdot \exp\left[\frac{-\eta(\omega)}{2}\right] \cdot \left\{ [1 + \eta(\omega)] I_0\left[\frac{-\eta(\omega)}{2}\right] + \eta(\omega) I_1\left[\frac{-\eta(\omega)}{2}\right] \right\} \cdot |Y(\omega)| \quad (\text{II.24})$$

---

<sup>9</sup> A demonstração da resposta do filtro dado na equação II.24 é apresentada no apêndice A.

Na equação II.24,  $I_0[\cdot]$  e  $I_1[\cdot]$  representam as funções de Bessel modificada do tipo um e de ordens ZERO e UM, respectivamente.

O parâmetro chave deste estimador é a relação sinal/ruído a priori, que deve ser cuidadosamente calculada, como será visto posteriormente. No entanto, pode-se antecipar que (em se tratando de uma estimação visto que o sinal disponível já incorpora o ruído) uma relação sinal/ruído a priori superestimada poderá causar distorções no sinal de voz processado. E o contrário, uma subestimação deixará um ruído residual forte.

Na figura II.4 têm-se as curvas de atenuação teóricas para este método. Sendo a resposta do filtro uma função das relações sinal/ruído a priori e a posteriori, obtiveram-se as curvas para quatro valores da SNR a posteriori. Observa-se que uma baixa SNR a priori implica numa maior atenuação (mais acentuada quanto mais forte for a SNR a posteriori), significando uma forte presença de ruído e, possivelmente, uma ausência de voz. Ao contrário, para uma SNR a priori alta, praticamente não se tem atenuação ( $\approx 0$  dB), indicando neste caso uma forte presença do sinal de voz, não importando o valor da SNR a posteriori.

EPHRAIM & MALAH, baseados no trabalho de McAULAY & MALPASS, desenvolveram um termo adicional considerando a probabilidade de presença ou ausência de voz nos sinais analisados, o que também será discutido no capítulo III.

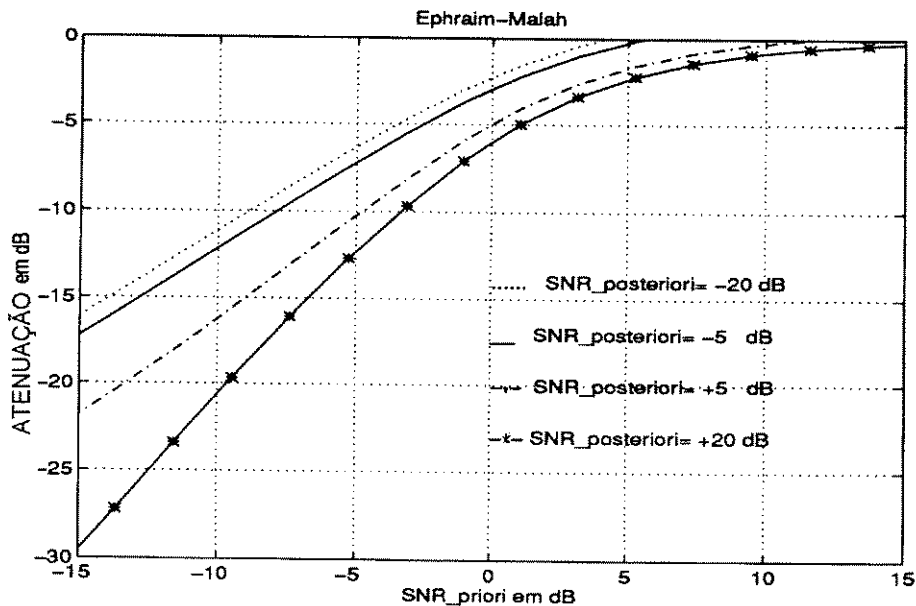


Figura II.4 - Atenuações obtidas com o método de Ephraim & Malah

**II.2.5 - Filtragem tipo pente**

Uma das principais características do sinal de voz é a sua periodicidade, identificada por meio da *freqüência fundamental* ou “*pitch*” (DELLER, 1993 e FLANAGAN, 1972). Isto implica numa forte concentração de energia em faixas de freqüência bem definidas. Assim, supondo um ruído aditivo de faixa larga (por exemplo, do tipo ruído branco), o uso de filtros passa-faixa estreitos, com freqüência central múltipla da freqüência fundamental, pode permitir uma boa estimação do sinal de voz original.

Na figura II.5 tem-se um diagrama que resume a operação de um sistema de redução de ruído baseado em um filtro adaptativo tipo pente.

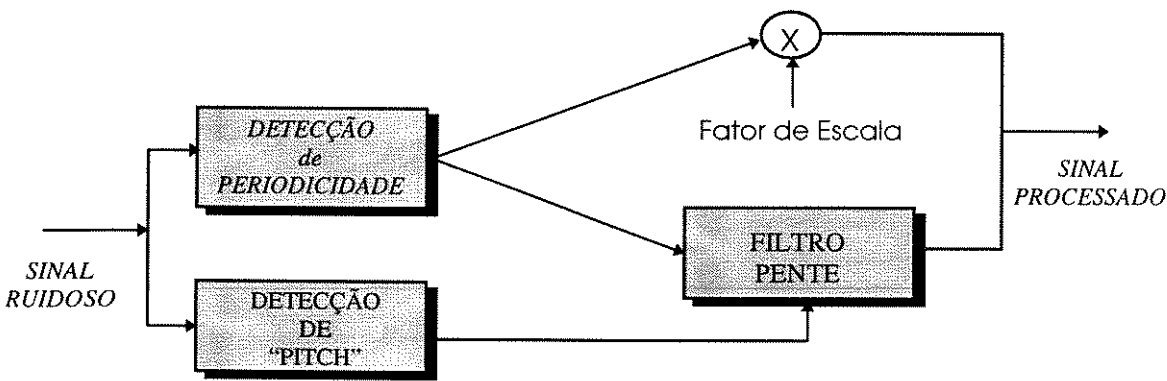


Figura II.5 - MSV usando filtro pente adaptativo

A detecção de periodicidade exige, na verdade, a implementação de um sistema que permita segmentar o sinal de voz em termos de sons surdos e sonoros. O objetivo é utilizar esta informação para melhorar o desempenho do sistema nas partes do sinal de voz onde predominam os sons surdos.

O sinal de voz estimado na saída do filtro pente será dado por:

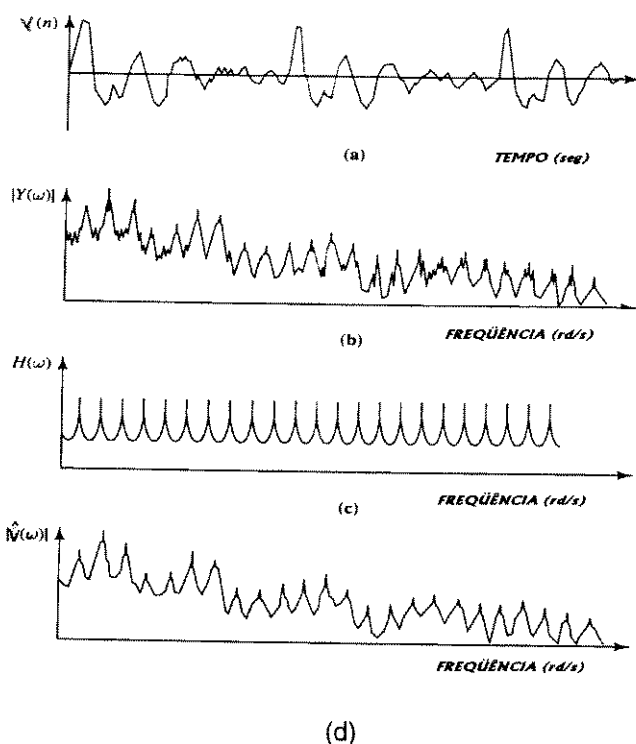
$$\hat{v}(n) = \sum_{i=-L}^L c(i) \cdot y(n - iT_o) \quad (\text{II.25})$$

onde  $c(i)$  são os coeficientes do filtro, que tem comprimento  $2L+1$ , e  $T_o$  é o período fundamental (*período de "pitch"*) do sinal de voz em número de amostras. A constante  $L$ , que define o tamanho do filtro, indica, na verdade, o número de períodos a ser considerado na estimação do sinal de voz. Isto significa também o atraso que ocorrerá no processamento, já que o sinal de voz estimado será uma versão "ponderada" e atrasada do sinal ruidoso. Normalmente,  $L$  varia de 1 a 6. Os filtros usados geralmente são baseados em janelas, como as de Hanning, Hamming, Blackman, etc. Por exemplo, no trabalho desenvolvido por LIM et alii (1979), onde avaliou-se um filtro adaptativo tipo pente, foi utilizada uma janela de Hamming que define a seguinte equação para o cálculo dos coeficientes:

$$c(i) = \frac{0,54 + 0,46 \cos(2\pi \cdot i / (2L+1))}{\sum_{i=-L}^L 0,54 + 0,46 \cos(2\pi \cdot i / (2L+1))} \quad (\text{II.26})$$

Um exemplo de um filtro pente (fixo) é apresentado na figura II.6. O princípio básico é reforçar o sinal em todas as frequências harmônicas da fundamental.

Quando o sinal analisado não apresenta sons periódicos identificáveis, o sinal processado será simplesmente uma versão atenuada do sinal ruidoso. Isto introduz fortes discontinuidades entre as janelas processadas e o resultado segundo LIM et alii. (1979) é uma forte distorção. Também, a própria detecção do "pitch" pode ser fonte de erros e conseqüentemente de distorções.



*Figura II.6 - Etapas típicas de uma filtragem tipo pente*

## **II.2.6 - Outros Métodos**

O estudo do MSV é vasto e existem vários outros métodos baseados em um único microfone. Por exemplo, LIM & OPPENHEIM (1978), enfocam o MSV baseado na estimação dos parâmetros LPC (Linear Predictive Coding) usando um filtro somente com pólos para modelar a voz. A idéia básica é estimar os parâmetros LPC do sinal ruidoso e aplicá-los no cálculo de um filtro de Wiener não-causal, que é então usado para filtrar de forma direta o sinal ruidoso. Esse filtro é o mesmo dado na equação II.8, sendo que a potência do sinal pode ser estimada como segue:

$$|\hat{V}(\omega)|^2 = \frac{g^2}{\left| 1 - \sum_{k=1}^p a_k \cdot e^{-j\omega k} \right|^2} \quad (\text{II.27})$$

onde  $p$  e  $g$  são, respectivamente, a ordem e o ganho do modelo. A potência do ruído, como na subtração espectral, é estimada durante os intervalos de ausência de voz.

Técnicas ainda mais complexas, como as que usam redes neurais, também têm surgido (BARBIER, 1990). No entanto, além do grande número de cálculos envolvidos, os métodos baseados em redes neurais ainda deixam fortes distorções no sinal estimado.

Uma outra linha de estudo, que tem despertado interesse, baseia-se nas características do ouvido humano (TSOUKALAS et alii, 1993; SCHROEDER et alii, 1979 e CHENG & O'SHAUGHNESSY, 1991). O objetivo neste caso é o de reduzir o ruído presente no sinal usando como base funções que modelam o sistema auditivo.

### ***II.3 - Métodos que usam dois microfones para captar o sinal***

O grande problema existente nas técnicas apresentadas anteriormente é a necessidade de estimação do ruído nos intervalos de ausência de voz. Se um segundo microfone for disponível, outras técnicas podem ser desenvolvidas. Para aproveitar este fato, existem duas possibilidades. Na primeira, um microfone é usado diretamente para captar o sinal de voz ruidoso e um segundo microfone é usado para captar apenas o ruído. Isto sugere um *filtro adaptativo*. Na segunda, os dois microfones captam ao mesmo tempo sinal e ruído e sugere um método baseado na *função de coerência*.



II.3.1 - Filtragem adaptativa

Considerando dois microfones, um captando sinal e ruído e um outro captando somente ruído, pode-se pensar de forma direta na filtragem adaptativa. Na figura II.7 tem-se um diagrama resumindo o melhoramento de sinais voz por meio de uma filtragem adaptativa quando dois microfones estão disponíveis.

Os ruídos presentes em M1 e M2 são considerados decorrelacionados do sinal de voz. No entanto, supõe-se que o ruído presente em M1 é uma transformação linear do ruído presente em M2. O objetivo é estimar o ruído  $r_1(n)$  e subtraí-lo do sinal do ruído.

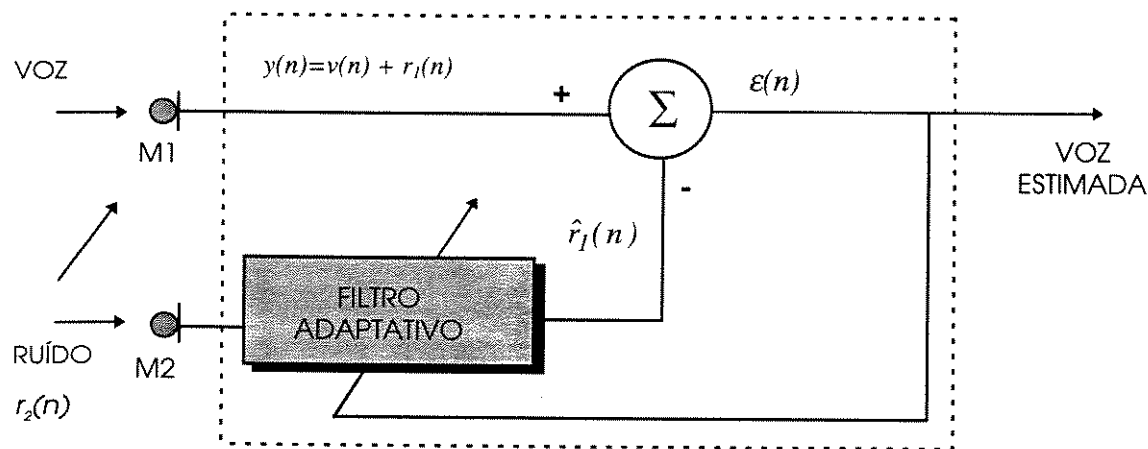


Figura II.7 - MSV usando filtragem adaptativa

Na figura II.8 tem-se uma estrutura típica para o filtro apresentado na figura II.6.

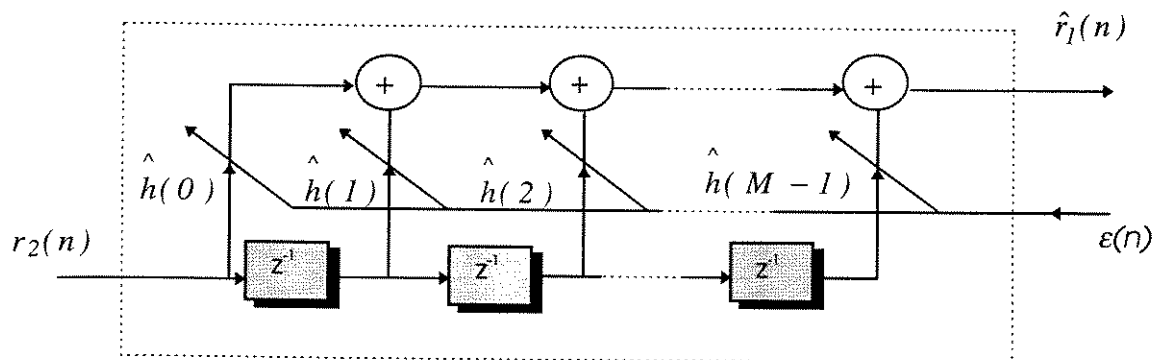


Figura II.8- Estrutura possível para o Filtro Adaptativo

O ruído a ser estimado é dado por:

$$\hat{r}_1(n) = \sum_{i=0}^{M-1} \hat{h}(i) \cdot r_2(n-i) \quad (\text{II.28})$$

O erro para a adaptação dos coeficientes do filtro é:

$$\varepsilon(n) = r_1(n) - \hat{r}_1(n) \quad (\text{II.29})$$

Usando o princípio da ortogonalidade (HAYKIN, 1991), o filtro a ser estimado deve satisfazer à equação:

$$\mathfrak{R}_{r_2 r_2} \cdot H = \mathfrak{R}_{y r_2}, \quad (\text{II.30})$$

onde

$\mathfrak{R}_{r_2 r_2}$  é a matriz de autocorrelação do ruído no microfone M2,  $\mathfrak{R}_{y r_2}$  é correlação cruzada do sinal ruidoso e do ruído no microfone M2 e  $H$  é uma matriz coluna dos coeficientes do filtro a ser estimado.

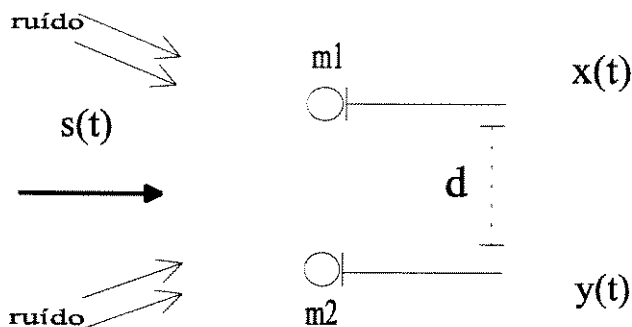
A obtenção do filtro a partir da equação II.30 pode definir diferentes sistemas adaptativos (WIDROW et alii, 1975 e KIM & UN, 1986). Uma possibilidade é o uso direto do algoritmo LMS (DELLER, 1993).

Além da grande quantidade de cálculos, um dos principais problemas para o uso de filtros adaptativos é o de colocar em fase sinal ruidoso e ruído. Como discutido anteriormente, não é tarefa fácil obter-se um ruído isolado e correlacionado com o ruído que deseja-se eliminar.

### **II.3.2 - Uso da função de coerência - método 1**

O uso de um filtro adaptativo parte do pressuposto de que um dos microfones capta apenas ruído. Realisticamente, pode-se dizer que tal situação não é comum. Assim, uma outra possibilidade é considerar que os dois microfones captem uma mistura de voz e ruído.

Seja  $s(t)$  uma fonte de sinal captada por dois microfones,  $m1$  e  $m2$ , separados de uma distância  $d$ , como indicado na figura II.8.



*Figura II.8 - Sistema com dois microfones não isolados*

Considerando que os ruídos que chegam aos microfones são descorrelacionados entre si e que o sinal de voz consegue atingir os dois microfones com o máximo de correlação, ALLEN et alii (1977) propuseram um sistema de redução da reverberação, adaptado aqui à redução do ruído sem modificações, conforme diagrama apresentado na figura II.9. O ganho  $G(\omega)$  do filtro na figura II.9 será próximo de um quando a coerência entre os dois sinais captados for muito forte (o que se espera para o sinal de voz). De forma contrária, o ganho será próximo de zero quando a coerência entre os sinais for fraca (o que se espera para o ruído). Ainda na figura II.9,  $A(\omega)$  representa uma função de correção de fase, o que permite a obtenção de um único sinal para ser processado.

Uma análise direta da figura II.9 mostra que:

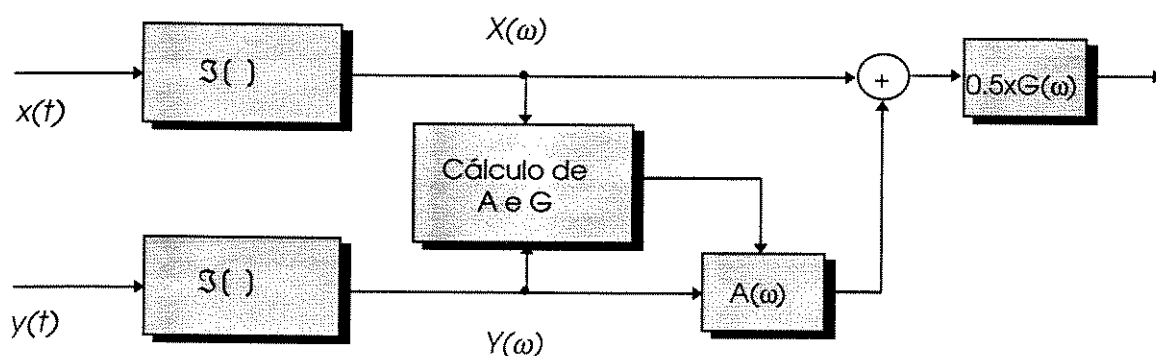
**A** - Se  $x$  e  $y$  são idênticos, então o sinal estimado será  $\hat{s} = x = y$ ;

**B** - Se  $x$  e  $y$  são atrasados, um em relação ao outro, de um tempo  $T$  então  $\hat{s}(t) = x(t - T) = y(t)$ ;

*C* - Se o atraso  $T > L$ , onde  $L$  é tamanho da janela de sinal analisada, então  $x$  e  $y$  tendem a ser sinais descorrelacionados. Neste caso  $\hat{s} \rightarrow 0$ ;

*D* - Finalmente, se  $x$  e  $y$  são totalmente descorrelacionados  $\hat{s} = 0$ .

A idéia básica é "eliminar" todas as frequências onde os sinais  $x$  e  $y$  são descorrelacionados e deixar passar somente as frequências correlacionadas. Neste ponto a distância entre os dois microfones passa a ser um dos parâmetros mais importantes. Normalmente  $d$  não pode ser muito pequena pois fornecerá dois sinais completamente correlacionados. Por outro lado  $d$  não pode ser muito grande, de forma a obterem-se dois sinais completamente descorrelacionados. O ideal é encontrar uma distância que permita uma forte correlação entre os sinais de voz captados e uma forte descorrelação entre os ruídos.



*Figura II.9 - Método de Allen-Berkley*

Do ponto de vista teórico, a distância  $d$  pode ser escolhida a partir de uma análise da função de coerência (MARRO, 1995).

Do ponto de vista prático, após várias análises de funções de coerência obtidas a partir de dois microfones separados por diferentes distâncias, um bom compromisso entre redução de ruído e recuperação do sinal de voz original pode ser obtido escolhendo-se uma distância entre 15 cm e 35 cm (valores obtidos através de simulações).

Para o uso da função de coerência duas funções básicas devem ser implementadas: a função de correção de fase e a função de coerência propriamente dita.

### II.3.2.1 - Função de correção de fase

Esta função permite eliminar o atraso entre os dois sinais para possibilitar a soma das duas entradas. Com esta correção, denominada de *formação de via única*, obtém-se um único sinal para ser filtrado. Esse sinal apresenta uma relação sinal/ruído superior à dos sinais originais, o que sugere o uso de antenas acústicas (SLYH & MOSES, 1993) na captação de sinais em ambientes ruidosos.

O cálculo da função de correção de fase  $A(\omega)$  é feito por:

$$A(\omega) = \frac{X(\omega)Y^*(\omega)}{|X(\omega)||Y(\omega)|} \quad (\text{II.31})$$

onde \* indica o conjugado. De acordo com a figura II.9, o sinal resultante após a soma das duas entradas será:

$$S(\omega) = Y(\omega) \cdot A(\omega) + X(\omega) = Y(\omega) \cdot \frac{X(\omega)Y^*(\omega)}{|X(\omega)||Y(\omega)|} + X(\omega) \quad (\text{II.32})$$

Da equação II.32 observa-se que o produto  $Y(\omega) A(\omega)$  fornece um sinal com módulo de  $Y(\omega)$  e fase de  $X(\omega)$ . Assim, o sinal resultante  $S(\omega)$  terá a contribuição das amplitudes dos dois sinais, mas usando a fase de uma das entradas. No caso, da entrada  $x(t)$ .

### II.3.2.2 - Cálculo da função de coerência

A função de coerência é definida por:

$$\rho(\omega) = \frac{\gamma_{xy}(\omega)}{\sqrt{\gamma_{xx}(\omega) \gamma_{yy}(\omega)}} \quad (\text{II.33})$$

onde  $\gamma_{xy}(\omega)$  é denominado *inter-espectro* e  $\gamma_{xx}(\omega)$  e  $\gamma_{yy}(\omega)$  são os *auto-espectros* das entradas  $x$  e  $y$ , respectivamente. Na figura II.9, o ganho  $G(\omega)=|\rho(\omega)|$ , o que significa que o *módulo da função de coerência (MSC)* é o próprio *ganho do filtro* redutor de ruído.

Quando os sinais presentes nos microfones 1 e 2, em uma determinada frequência  $\omega$ , forem bastante correlacionados, nessa frequência o módulo da função de coerência tenderá para a UNIDADE. Isto deveria ocorrer quando os sinais de voz estivessem presentes nas duas entradas. De forma contrária, se os sinais presentes nas duas entradas forem descorrelacionados, o MSC tenderá para ZERO, o que deveria acontecer quando o ruído fosse predominante nas duas entradas. No entanto, existe uma faixa de frequência onde voz e ruído são ambos correlacionados. Nessa faixa, o MSC não é nulo e depende da distância entre os microfones.

Para se ter uma idéia, na figura II.10 apresenta-se um exemplo do módulo da função de coerência para um sinal de voz ruidoso gravado dentro de um veículo e captado por dois microfones distanciados de 30 cm. Na primeira parte tem-se o MSC para os intervalos de silêncio (figura II.10a) e na segunda tem-se o MSC para os instantes de voz (figura II.10b). Nota-se que o ruído apresenta uma boa coerência nas frequências mais baixas, o mesmo ocorrendo com o sinal de voz.

Sendo o módulo da função de coerência o próprio filtro redutor de ruído, observa-se na figura II.10b que o sinal processado poderá apresentar alguma redução de ruído na faixa de 0 a 1250 kHz, mas continuará com um ruído significativo na faixa de 1500 a 2750 kHz. Isto pode

ser explicado observando-se o fato de que na formação dos sinais de voz existe uma contribuição de sons periódicos (conhecidos como sonoros) e sons com características ruidosas (conhecidos como surdos). Assim, uma redução de ruído nessa faixa resultaria na introdução de grandes distorções no sinal processado, o que seria contrário à idéia básica do uso da função de coerência (redução de ruído com o mínimo de distorção no sinal processado). Globalmente, para os exemplos apresentados na figura II.10, a aplicação da função de coerência na redução do ruído não indica um bom processamento devido à boa coerência do ruído.

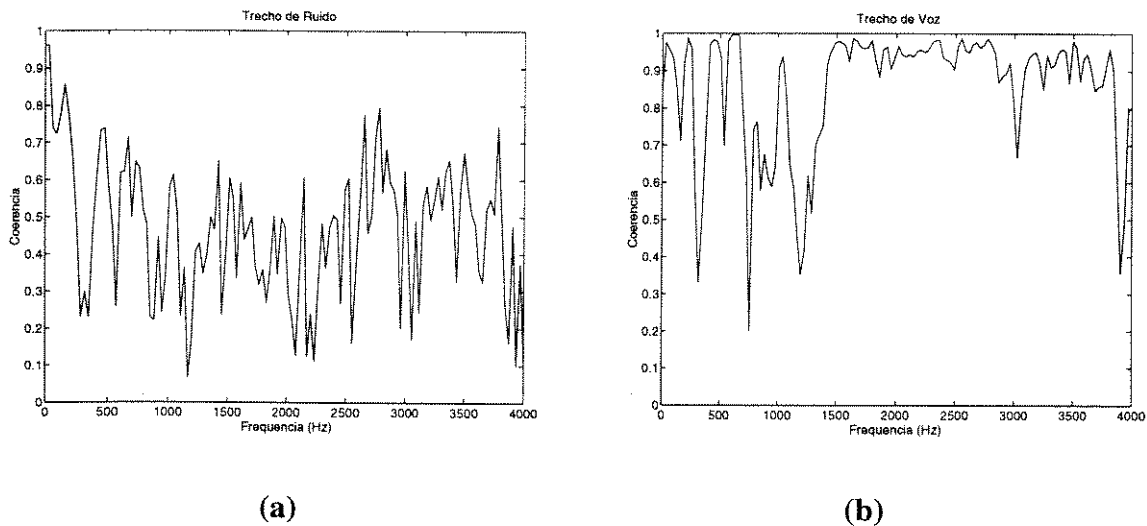


Figura II.10 - (a) Coerência do ruído (b) Coerência da voz

Para se utilizar a função de coerência, é necessária uma estimação dos inter-espectros e dos auto-espectros. A metodologia normalmente utilizada é baseada numa média ponderada e recursiva dos espectros de amplitudes.

Definindo:

$$\gamma_{xx}(\omega) = \frac{X(\omega) \cdot X(\omega)^*}{\dots}$$

- Auto-Espectro da entrada m1 (II.34)

$$\gamma_{yy}(\omega) = \overline{Y(\omega) \cdot Y(\omega)^*} \quad \text{- Auto-Espectro da entrada m2} \quad (\text{II.35})$$

$$\gamma_{xy}(\omega) = \overline{X(\omega) \cdot Y(\omega)^*} \quad \text{- Inter-Espectro entre m1 e m2} \quad (\text{II.36})$$

uma forma direta de obterem-se os auto-espectros e o inter-espectro é:

$$\gamma_{ij}(\omega)_T = \alpha \cdot \gamma_{ij}(\omega)_{T-1} + (1-\alpha) \cdot Z(\omega)_T \cdot W^*(\omega)_T, \quad (\text{II.37})$$

onde  $Z(\omega)$  e  $W(\omega)$  podem ser  $X(\omega)$  e  $Y(\omega)$ . A constante  $T$  indica o intervalo atual de análise e  $\alpha$  é o fator de esquecimento. O valor de  $\alpha$  é escolhido de acordo com a constante de tempo estabelecida para o cálculo da média e dependerá da frequência de amostragem dos sinais e do intervalo de tempo tomado para análise. Vale salientar que uma constante de tempo grande (por exemplo, três vezes maior do que o intervalo de análise) provocará um índice de reverberação significativo no sinal processado, mas com uma boa redução de ruído. O contrário poderá minimizar a reverberação mas eliminará pouco ruído.

Os ruídos presentes em  $m1$  e  $m2$  são considerados descorrelacionados do sinal de voz. No entanto, supõe-se que o ruído presente em  $m1$  é uma transformação linear do ruído presente em  $m2$ . O objetivo é estimar o ruído  $r_1(n)$  e subtraí-lo do sinal do ruído.

### **II.3.3 - Uso da função de coerência - método 2**

Na figura II.11 tem-se uma versão modificada do algoritmo proposto inicialmente por ALLEN et alii (1977), que foi estudado por LE BOUQUIN (1990). O princípio é o mesmo, com o ganho final sendo baseado na função de coerência. No entanto, não há formação de um



sinal único a partir das duas entradas. O processamento final é feito tomando-se uma das entradas e multiplicando-a pelo ganho obtido através da coerência entre os dois sinais. Em termos práticos elimina-se a necessidade de cálculo da função de correção de fase, o que garante uma vantagem inicial em relação à complexidade computacional. Entretanto, espera-se um ruído residual mais significativo, pois a formação de via única garante a princípio um ganho na relação sinal/ruído.

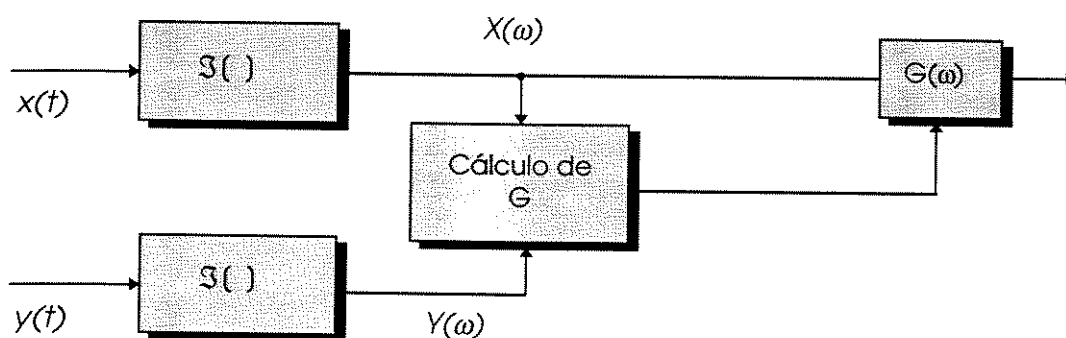


Figura II.11 - Algoritmo de ALLEN et alii (1977) modificado

## II.4- Conclusões

Tentou-se neste capítulo resumir as principais técnicas de redução de ruído com possibilidade de aplicação direta em MSV. Nota-se um vasto número de técnicas existentes, sendo necessário, portanto, restringir a base de estudo. Neste trabalho, serão privilegiadas as técnicas que possam ser implementadas em sistemas de operação em tempo real, como é o caso dos sistemas *HFTs*. Como orientação de escolha, usou-se a complexidade de implementação (ou computacional). Alguns comentários sobre as técnicas apresentadas neste capítulo podem ser antecipados.

De uma maneira geral, os métodos baseados na subtração espectral e generalizados pelo filtro de Wiener, têm como maior atrativo o número reduzido de cálculos. Como maior inconveniente, tem-se um tipo desagradável de ruído residual resultante do processamento, denominado de ruído musical. Este problema também aparece no método baseado na estimação de máxima-verossimilhança. No entanto, o método baseado na minimização do erro quadrático médio desenvolvido por EPHRAIM & MALAH (1984) não apresenta o fenômeno do ruído musical e tem sido bem explorado (CAPPÉ, 1994, e HAKKINEN, 1993). O inconveniente neste caso é o número elevado de cálculos, que inviabiliza, no momento, sua aplicação em tempo real. Os métodos que usam filtragem tipo pente e modelagem LPC para o sinal ruidoso são baseados no período de “*pitch*”. No entanto, a estimação do período de “*pitch*” na presença de ruído constitui um problema significativo, ainda em estudo, o que inviabiliza a aplicação dessas técnicas.

Nos métodos baseados em dois microfones, o maior problema para aplicação de um sistema adaptativo é a necessidade de uso de um microfone isolado que capte um ruído coerente com o ruído presente no sinal de voz (ao mesmo tempo) e não capte voz. Também, neste caso, um fator complicador passa a ser a obtenção de sinais em fase. Com o uso de dois microfones mais próximos, explorando a coerência dos sinais envolvidos, a maior vantagem é a simplicidade. No entanto, o sinal processado por esta técnica apresenta um efeito de reverberação que depende do nível de redução de ruído que se deseja atingir. Além disso, as componentes do ruído em baixa frequência são fracamente reduzidas.

O estudo realizado neste capítulo direcionou o trabalho para explorar as técnicas baseadas na subtração espectral, filtragem de Wiener, máxima-verossimilhança e a técnica de Ephraim & Malah, para o uso com um único microfone. No caso de dois microfones, o estudo foi centrado na função de coerência. O resultado final é um conjunto de novas técnicas de MSV, apresentadas e avaliadas nos capítulos seguintes.

## **CAPÍTULO III**

# **O USO DA RELAÇÃO SINAL/RUÍDO NO MELHORAMENTO DE SINAIS DE VOZ**

### **III.1 - Introdução**

O estudo apresentado no capítulo II orientou a seleção de um conjunto de técnicas que podem ser implementadas nos terminais telefônicos mãos-livres (*HFTs*), usados nos sistemas rádio-móveis veiculares. Um estudo mais aprofundado detectou pontos importantes relacionados a essas técnicas. Por exemplo, foi confirmado o aparecimento do ruído musical nas técnicas como subtração espectral, filtragem de Wiener e máxima-verossimilhança, mas com boa redução de ruído e simplicidade computacional. Também confirmou-se o bom desempenho da técnica desenvolvida por Ephraim & Malah, com um número muito elevado de cálculos<sup>1</sup>. Definiu-se assim um conjunto de técnicas baseadas em um único microfone para serem exploradas. Para os métodos baseados em dois microfones, considerou-se mais interessante o uso de técnicas baseadas na função de coerência. Esta decisão foi tomada com base nas dificuldades práticas envolvidas no uso

---

<sup>1</sup> A implementação do filtro de Ephraim & Malah envolve o cálculo de duas funções de Bessel, de ordem zero e um, e uma exponencial, a mais que os outros filtros. É importante lembrar também que o tempo disponível para o cálculo de um filtro de redução de ruído em tempo real depende da frequência de amostragem do sinal.

de filtros adaptativos, que exige um microfone isolado do outro do ponto de vista de captação de voz.

Com estas constatações, neste capítulo discutem-se inicialmente os problemas relacionados com a subtração espectral, que é generalizada pelo filtro de Wiener, e a estimação usando a máxima-verossimilhança. Uma análise destes métodos mostrou que a *relação sinal-ruído a posteriori* ( $SNR_{post}$ ) é usada como parâmetro principal. Estuda-se também a técnica proposta por EPHRAIM & MALAH (1984), mostrando que o bom desempenho desta técnica não vem exatamente do uso de um modelo estatístico, mas sim do fato de que o filtro calculado é função da *relação sinal/ruído a priori* ( $SNR_{prio}$ ).

O estudo dos parâmetros utilizados por estas técnicas permite que se proponham novos métodos, baseados em um ou dois microfones, que usam como parâmetro principal a  $SNR_{prio}$ . O resultado são técnicas que não apresentam o fenômeno do ruído musical, para o caso das que usam um único microfone, e um novo método de cálculo para a função de coerência, que elimina o problema da reverberação.

Introduz-se também neste capítulo o modelo a dois estados para a análise de presença e ausência do sinal de voz, que foi proposto inicialmente por McAULAY & MALPASS (1980). Dois estimadores de incerteza para o sinal de voz são apresentados e propõe-se o controle de alguns parâmetros para usufruir-se dos seus potenciais.

## **III.2 - Técnicas de MSV baseadas na relação sinal/ruído a posteriori**

### **III.2.1 - Subtração espectral (SE)**

Retoma-se neste item os pontos principais da subtração espectral, já discutida no capítulo II. Assim, seja um sinal de voz puro  $v(t)$ , degradado por um ruído aditivo  $r(t)$  e que resulta no sinal ruidoso  $y(t)$ . Sendo voz e ruído considerados processos aleatórios estacionários e independentes, no domínio da frequência tem-se:

$$|Y(\omega)|^2 = |V(\omega)|^2 + |R(\omega)|^2 \quad (\text{III.1})$$

Seja  $H(\omega)$  um filtro redutor de ruído, com módulo dado por:

$$|H(\omega)| = \frac{|V(\omega)|}{|Y(\omega)|} \quad (\text{III.2})$$

Utilizando as relações da subtração espectral, tem-se o seguinte estimador de  $H(\omega)$ :

$$|\hat{H}(\omega)|^2 = \frac{|Y(\omega)|^2 - |\hat{R}(\omega)|^2}{|Y(\omega)|^2} \quad (\text{III.3})$$

Define-se, após McAulay & Malpass, a *relação sinal/ruído a posteriori* estimada como sendo:

$$SNR_{post}(\omega) = \frac{|Y(\omega)|^2}{|\hat{R}(\omega)|^2} \quad (\text{III.4})$$

O filtro redutor de ruído dado na equação III.3 pode então ser reescrito como segue:

$$|\hat{H}(\omega)| = \sqrt{1 - \frac{1}{SNR_{post}(\omega)}} \quad (\text{III.5})$$

O filtro dado na equação III.5 será definido apenas se  $SNR_{post}(\omega) \geq 1$ , o que implica numa possível “retificação” para os casos onde a potência do sinal ruidoso esteja abaixo da potência do ruído estimada. Mesmo assim, o fato de se escrever o filtro em função da  $SNR_{post}$  já pode resultar em um melhor desempenho do filtro, tendo em vista que a retificação será feita na relação sinal/ruído a posteriori, e não diretamente como apresentado no capítulo II. Isto significa que, dependendo do processo de estimação da  $SNR_{post}$ , pode-se reduzir o efeito do ruído musical.

### III.2.2 - Filtro de Wiener generalizado

O filtro de Wiener a ser estimado, já discutido no capítulo II, é dado por:

$$|\hat{H}_w(\omega)| = \frac{|\hat{V}(\omega)|^2}{|\hat{V}(\omega)|^2 + |\hat{R}(\omega)|^2} \quad (\text{III.6})$$

Aplicando-se a análise realizada no item anterior e resolvendo o problema da não-causalidade pela aplicação do estimador da subtração espectral, tem-se o filtro de Wiener generalizado dado por:

$$|\hat{H}_w(\omega)| = \left\{ \frac{|Y(\omega)|^2 - |\hat{R}(\omega)|^2}{|Y(\omega)|^2 - |\hat{R}(\omega)|^2 + \alpha \cdot |\hat{R}(\omega)|^2} \right\}^\beta \quad (\text{III.7})$$

Desenvolvendo a equação III.7, obtém-se:

$$|\hat{H}_w(\omega)| = \left\{ \frac{|Y(\omega)|^2}{|Y(\omega)|^2 + (\alpha - 1) \cdot |\hat{R}(\omega)|^2} - \frac{|\hat{R}(\omega)|^2}{|Y(\omega)|^2 + (\alpha - 1) \cdot |\hat{R}(\omega)|^2} \right\}^\beta \quad (\text{III.8})$$

Usando a definição dada na equação III.4, a equação III.7 pode ser novamente reescrita como segue:

$$|\hat{H}_w(\omega)| = \left\{ \frac{1}{1 + \frac{(\alpha - 1)}{\hat{SNR}_{post}(\omega)}} - \frac{1}{\hat{SNR}_{post}(\omega) + (\alpha - 1)} \right\}^\beta \quad (\text{III.9})$$

Finalmente, o filtro de Wiener generalizado, em função da relação sinal/ruído a posteriori, é dado por:

$$|\hat{H}_w(\omega)| = \left\{ \frac{SNR_{post}(\omega) - 1}{SNR_{post}(\omega) + (\alpha - 1)} \right\}^{\beta} \quad (III.10)$$

O valores de  $\alpha$  e  $\beta$  definem o filtro, sendo que, normalmente,  $\alpha \geq 1$  e  $\beta$  varia entre 0 e 1. Fazendo-se  $\alpha = 1$ , a potência do ruído usada no cálculo do filtro será exatamente a potência estimada. Se  $\alpha < 1$ , tem-se uma subestimação da potência do ruído, o que significa uma menor redução do ruído. Finalmente, se  $\alpha > 1$ , tem-se uma superestimação da potência do ruído, significando uma melhor redução do ruído, mas aumentando-se o fenômeno do ruído musical. A atenuação do filtro será tanto maior quanto mais próximo de 1 for o valor de  $\beta$ . A escolha destes parâmetros, normalmente feita através de simulações, define no final os níveis de distorção, de ruído musical e de redução de ruído do sinal processado.

### III.2.3 - Estimação pela máxima-verossimilhança

No capítulo II viu-se que o sinal de voz estimado pelo método da máxima-verossimilhança era dado por:

$$|\hat{V}(\omega)| = \frac{1}{2} \left\{ |Y(\omega)| + \sqrt{|Y(\omega)|^2 - |\hat{R}(\omega)|^2} \right\} \quad (III.11)$$

Usando a definição da equação III.4 pode-se descrever a resposta do filtro redutor de ruído como segue:

$$|\hat{H}(\omega)| = \frac{1}{2} \left[ 1 + \sqrt{1 - \frac{1}{SNR_{post}(\omega)}} \right] \quad (III.12)$$

Como na subtração espectral, é preciso ter  $SNR_{post}(\omega) \geq 1$ .

### III.3 -Análise da relação sinal/ruído a posteriori

Notou-se no estudo dos filtros baseados na relação sinal/ruído a posteriori que existe uma função comum a todos os filtros, dada por:

$$\hat{Q}(\omega) = \frac{\hat{SNR}_{post}(\omega) - 1}{\hat{SNR}_{post}(\omega)} \quad (III.13)$$

Para a realização desses filtros é preciso que  $\hat{Q}(\omega)$  seja maior ou igual a zero, o que significa que  $\hat{SNR}_{post}(\omega) \geq 1$ . No entanto, sendo voz e ruído sinais com fases aleatórias, a formação do sinal ruidoso não garante  $SNR_{post} \geq 1$ . Na figura III.1 mostra-se o mecanismo de formação do sinal ruidoso. Se for, por exemplo, fixada uma fase para o sinal de voz, vê-se que a aleatoriedade da fase do ruído e de sua potência podem fazer com que o sinal ruidoso possua uma potência instantânea inferior à potência do ruído, significando, assim, uma  $SNR_{post} < 1$ . Esta situação exigirá o uso da retificação para garantir-se a realizabilidade dos filtros.

Para avaliar-se o comportamento da  $SNR_{post}$ , tomou-se uma senóide de frequência igual a 1 kHz e adicionou-se um ruído gaussiano branco. Foram observados dois sinais ruidosos com relação sinal/ruído de 20 dB e 4 dB, respectivamente. Os histogramas para os dois casos são apresentados nas figuras III.2.A e III.2.B. Quando a SNR é de 20 dB (figura III.2.A), observa-se um baixo desvio padrão e um bom comportamento da  $SNR_{post}$ . Já na figura III.2.B, onde SNR é de 4 dB, observa-se a existência de muitos valores de  $SNR_{post}$  abaixo de 0 dB e um desvio padrão relativamente alto (lembrando que a SNR real é de 4 dB). Esta irregularidade em  $SNR_{post}$  é responsável por um desempenho insatisfatório dos filtros quando a SNR é baixa. No entanto, para altas SNR, o sinal ruidoso é praticamente confundido com o sinal puro e o desempenho da  $SNR_{post}$  é satisfatório.



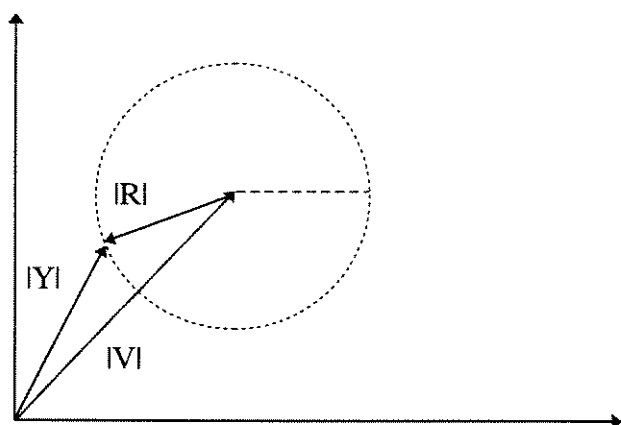


Figura III.1 - Mecanismo de formação do sinal ruidoso

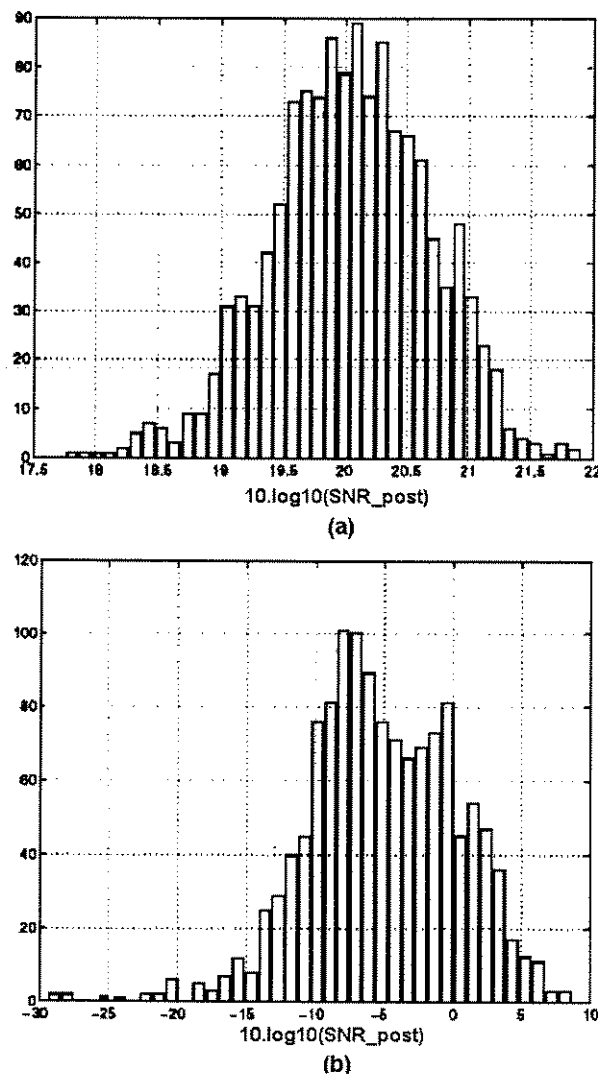


Figura III.2 - Histograma da SNR\_post para (a) SNR = 20 dB e (b) SNR = 4dB

### III.4 - Análise da relação sinal/ruído a priori

Os problemas apresentados pelas técnicas clássicas baseadas na  $SNR_{post}$  incentivam o estudo de técnicas alternativas. Neste sentido, Ephraim & Malah desenvolveram uma técnica onde a principal vantagem é a ausência do ruído musical (CAPPÉ, 1994). Esse filtro, apresentado no capítulo II, pode ser escrito na seguinte forma:

$$|\hat{H}(\omega)| = \Gamma(1,5) \cdot \frac{\sqrt{\eta(\omega)}}{\gamma(\omega)} \cdot \exp\left[\frac{-\eta(\omega)}{2}\right] \cdot \left\{ [1 + \eta(\omega)] I_0\left[\frac{-\eta(\omega)}{2}\right] + \eta(\omega) I_1\left[\frac{-\eta(\omega)}{2}\right] \right\} \quad (\text{III.14})$$

onde

$$\eta(\omega) = \frac{S\hat{N}R_{prio}(\omega)}{1 + S\hat{N}R_{prio}(\omega)} \cdot S\hat{N}R_{post}(\omega) \quad (\text{III.15})$$

e

$$S\hat{N}R_{prio}(\omega) = \frac{|\hat{V}(\omega)|^2}{|\hat{R}(\omega)|^2} \quad (\text{III.16})$$

A equação III.14 foi obtida considerando que os coeficientes da transformada de Fourier do sinal de voz apresentam uma distribuição gaussiana. No entanto, o modelo gaussiano pode não ser o mais apropriado para o sinal de voz, visto que outros modelos têm sido propostos (PORTER & BOLL, 1994 e ZELINSKI & NOLL, 1984). Com base nessa “controvérsia”, resolveu-se estudar esse filtro de maneira mais detalhada. Um dos primeiros pontos que foi investigado foi o próprio modelo gaussiano. Algumas simulações, realizadas na faixa de frequência de 800 Hz e 3 kHz, mostraram que o modelo mais adequado para os coeficientes de Fourier do sinal de voz seria baseado numa distribuição log-normal (SCALART et alii, 1996). No entanto, a tentativa de se obter um filtro teórico usando tal distribuição esbarra em equações que ainda não possuem soluções analíticas conhecidas. O uso de uma aproximação numérica já foi investigado por

PORTER & BOLL (1984), mas do ponto de vista prático não é viável. Os resultados obtidos serviram de indicativo para que se verificasse com mais atenção o filtro proposto por Ephraim & Malah.

Nas equações III.14 e III.15, verifica-se que o parâmetro principal do filtro redutor de ruído obtido por Ephraim & Malah é a *relação sinal/ruído a priori*, sendo a  $SNR_{post}$  um parâmetro secundário. Assim, um dos primeiros pontos observados foi o comportamento da  $SNR_{prio}$ . Com o mesmo procedimento adotado para o estudo da  $SNR_{post}$ , analisou-se a  $SNR_{prio}$ . Na figura III.3 tem-se o histograma para a  $SNR_{prio}$  (senóide de 1 kHz e SNR de 20dB e 4dB, respectivamente).

Observa-se que o comportamento de  $(SNR_{prio} + 1)$ , para uma SNR de 20 dB, é praticamente o mesmo obtido com a  $SNR_{post}$  dado na figura III.2.A. De fato, uma análise rápida da figura III.1 mostra que uma alta relação sinal/ruído faz com que  $|Y(\omega)|$  se aproxime de  $|V(\omega)|$ . Já no caso de uma SNR baixa (4 dB na figura III.3.B), o termo  $(SNR_{prio} + 1)$  apresenta uma média bem abaixo de 4 dB. Uma comparação com a figura III.3.B mostra, entretanto, que o termo  $(SNR_{prio} + 1)$  apresenta um desvio padrão muito menor do que o obtido com a  $SNR_{post}$ . Estas avaliações mostram que a relação sinal/ruído a priori apresenta um comportamento superior ao obtido com a  $SNR_{post}$ , sendo assim um indicativo do bom desempenho do filtro proposto por Ephraim-Malah. Para se comprovar a importância da  $SNR_{prio}$ , introduziram-se pequenas modificações na estimação da  $SNR_{prio}$ <sup>2</sup>. O resultado foi um sinal de voz estimado que apresentava o fenômeno do ruído musical, além de distorções superiores às obtidas com filtros clássicos, como a subtração espectral, por exemplo.

Com os testes realizados, que mostraram a superioridade da  $SNR_{prio}$  em relação a  $SNR_{post}$ , e as controvérsias em relação ao modelo gaussiano para o sinal de voz, é possível que o sucesso do filtro proposto por Ephraim & Malah não seja devido à exploração de um modelo estatístico para o sinal de voz. Esta dúvida incentivou o estudo dos filtros clássicos de redução de ruído, onde percebeu-se que era possível redefini-los usando a  $SNR_{prio}$ .

---

<sup>2</sup> A estimação da  $SNR_{prio}$  é apresentada no capítulo IV. Ela incorpora uma parcela da  $SNR_{post}$ . A modificação feita aumentava a contribuição da  $SNR_{post}$ .

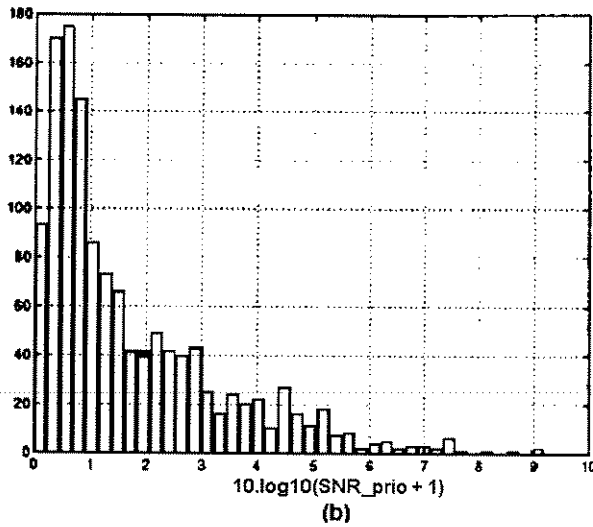
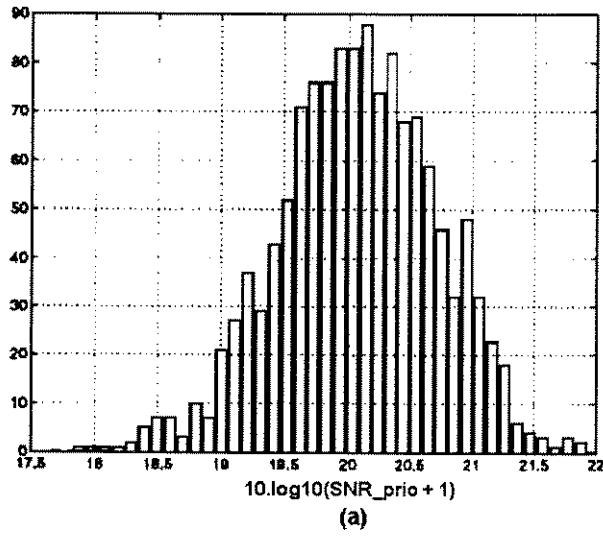


Figura III.3 - Histograma da  $SNR_{prio}$  para (a)  $SNR = 20$  dB e (b)  $SNR = 4$  dB

### III.5 - Técnicas clássicas de redução de ruído usando a $SNR_{prio}$

Os estudos apresentados anteriormente mostraram a importância da utilização da  $SNR_{prio}$ , que pode ser uma solução na eliminação do ruído musical. Tendo em vista os problemas discutidos com relação à utilização da  $SNR_{post}$ , torna-se importante explorar o uso da  $SNR_{prio}$  nas técnicas clássicas.

É possível mostrar-se que a  $SNR_{post}$  pode ser escrita como uma função da  $SNR_{prio}$ . Assim, retomando-se à equação III.1 e dividindo por  $|R(\omega)|^2$ , obtém-se:

$$\frac{|Y(\omega)|^2}{|R(\omega)|^2} = \frac{|V(\omega)|^2}{|R(\omega)|^2} + 1 \quad (\text{III.17})$$

De acordo com as definições das equações III.4 e III.16, tem-se que:

$$SNR_{post}(\omega) = 1 + SNR_{prio}(\omega) \quad (\text{III.18})$$

A estimação da  $SNR_{post}$  com base na  $SNR_{prio}$  elimina definitivamente a necessidade de se fazer uma retificação. Uma consequência imediata é a redefinição dos filtros clássicos de redução de ruído.

Considerando a equação III.18, os filtros definidos nas equações III.5, III.10 e III.12 podem ser escritos em função da  $SNR_{prio}$  como segue:

### ***Subtração Espectral***

$$|\hat{H}(\omega)| = \sqrt{\frac{\hat{SNR}_{prio}(\omega)}{1 + \hat{SNR}_{prio}(\omega)}} \quad (\text{III.19})$$

### ***Filtro de Wiener generalizado***

$$|\hat{H}_w(\omega)| = \left\{ \frac{\hat{SNR}_{prio}(\omega)}{\alpha + \hat{SNR}_{prio}(\omega)} \right\}^\beta \quad (\text{III.20})$$

### ***Máxima-Verossimilhança***

$$|\hat{H}(\omega)| = \frac{1}{2} \left[ 1 + \sqrt{\frac{\hat{SNR}_{prio}(\omega)}{1 + \hat{SNR}_{prio}(\omega)}} \right] \quad (\text{III.21})$$

A nova versão dos filtros não altera as curvas de atenuação apresentadas no capítulo II. Os tons aleatórios que são obtidos quando se trabalha diretamente com as potências dos sinais ou com a  $SNR_{post}$  desaparecem com o uso da  $SNR_{prio}$ . O resultado deverá ser uma diminuição significativa do problema do ruído musical ou mesmo sua eliminação, com o mesmo nível de redução de ruído. Como será visto posteriormente, o desempenho dos filtros depende diretamente do método (ou ajuste) usado na estimação da  $SNR_{prio}$ .

### **III.6 - Função de coerência usando a $SNR_{prio}$**

O cálculo da função de coerência, como desenvolvido no capítulo II, apresenta como ponto crítico o uso de uma recursividade na estimação dos espectros dos sinais. Isto tem como principal inconveniência a introdução de um efeito de reverberação no sinal estimado.

A experiência dos trabalhos baseados em um único microfone com o uso de estimadores baseados na relação sinal/ruído a priori permitiu a proposição de um novo método de cálculo para a função de coerência. Com este novo método, elimina-se a recursividade na estimação dos espectros dos sinais de voz recebidos nos 2 microfones.

De acordo com a figura II.9 (capítulo II), os sinais recebidos podem ser representados como:

$$\begin{aligned} x(t) &= v_1(t) + r_1(t) \\ y(t) &= v_2(t) + r_2(t) \end{aligned} \quad e \quad (III.22)$$

onde  $v_1(t)$  e  $v_2(t)$  são os sinais de voz puros e  $r_1(t)$  e  $r_2(t)$  são os ruídos captados nos microfones.

Considerando que os ruídos podem ser coerentes entre si mas que eles são descorrelacionados dos sinais de voz, pode-se rescrever os auto-espectros como segue:

$$\gamma_{xx}(\omega) = \gamma_{v_I v_I}(\omega) + \gamma_{r_I r_I}(\omega) \quad (\text{III.23})$$

$$\gamma_{yy}(\omega) = \gamma_{v_2 v_2}(\omega) + \gamma_{r_2 r_2}(\omega) \quad (\text{III.24})$$

O novo inter-espectro será dado por:

$$\gamma_{xy}(\omega) = [V_I(\omega) + R_I(\omega)] \cdot [V_2(\omega) + R_2(\omega)]^* \quad (\text{III.25})$$

Desenvolvendo a equação III.25 tem-se:

$$\gamma_{xy}(\omega) = V_I(\omega) \cdot V_2(\omega)^* + R_I(\omega) \cdot R_2(\omega)^* + V_I(\omega) \cdot R_2(\omega)^* + R_I(\omega) \cdot V_2(\omega)^* \quad (\text{III.26})$$

Considerando as hipóteses de independência entre sinais e ruídos, obtém-se então:

$$\gamma_{xy}(\omega) = V_I(\omega) \cdot V_2(\omega)^* + R_I(\omega) \cdot R_2(\omega)^* \quad (\text{III.27})$$

A nova função de coerência é finalmente dada por:

$$\rho(\omega) = \frac{\gamma_{v_I v_2}(\omega) + \gamma_{r_I r_2}(\omega)}{\sqrt{[\gamma_{v_I v_I}(\omega) + \gamma_{r_I r_I}(\omega)] \cdot [\gamma_{v_2 v_2}(\omega) + \gamma_{r_2 r_2}(\omega)]}} \quad (\text{III.28})$$

A equação III.28 pode ainda ser escrita como segue:

$$\rho(\omega) = \frac{\gamma_{r_I r_2}(\omega) \left[ 1 + \frac{\gamma_{v_I v_2}(\omega)}{\gamma_{r_I r_2}(\omega)} \right]}{\sqrt{\gamma_{r_I r_I}(\omega) \left[ 1 + \frac{\gamma_{v_I v_I}(\omega)}{\gamma_{r_I r_I}(\omega)} \right] \cdot \gamma_{r_2 r_2}(\omega) \left[ 1 + \frac{\gamma_{v_2 v_2}(\omega)}{\gamma_{r_2 r_2}(\omega)} \right]}} \quad (\text{III.29})$$

Definindo:

$$\frac{\gamma_{v_1v_1}(\omega)}{\gamma_{r_1r_1}(\omega)} = SNR\_prio1(\omega) \quad \text{e} \quad \frac{\gamma_{v_2v_2}(\omega)}{\gamma_{r_2r_2}(\omega)} = SNR\_prio2(\omega) \quad (\text{III.30})$$

como as *relações sinal/ruído a priori* na saída de cada microfone;

$$\frac{\gamma_{v_1v_2}(\omega)}{\gamma_{r_1r_2}(\omega)} = SNR\_prio\_12(\omega) \quad (\text{III.31})$$

como a *relação sinal/ruído a priori cruzada* entre as saídas dos dois microfones e

$$\rho_{ruído}(\omega) = \frac{\gamma_{r_1r_2}(\omega)}{\sqrt{\gamma_{r_1r_1}(\omega) \cdot \gamma_{r_2r_2}(\omega)}} \quad (\text{III.32})$$

como a função de coerência do ruído, pode-se finalmente rescrever a equação III.29 como:

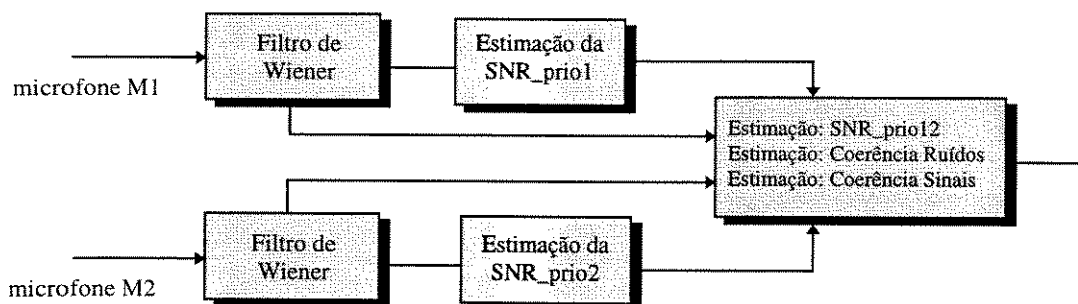
$$\rho(\omega) = \rho_{ruído}(\omega) \frac{[1 + SNR\_prio\_12(\omega)]}{\sqrt{[1 + SNR\_prio1(\omega)] \cdot [1 + SNR\_prio2(\omega)]}} \quad (\text{III.33})$$

Da equação III.33 observa-se que o novo cálculo da função de coerência entre os sinais depende diretamente das relações sinal/ruído definidas, o que elimina o problema de cálculos recursivos nos espectros dos sinais de voz. No entanto, este cálculo continua a ser necessário para a determinação da função de coerência do ruído, mas não afeta a qualidade do processamento final, como apresentado no capítulo V.

Para estimar as relações sinal-ruído pode-se aplicar qualquer método de redução de ruído baseado em um único microfone. Particularmente, neste trabalho utiliza-se a filtragem de Wiener.

Na figura III.4 tem-se um diagrama esquemático resumindo o novo método de cálculo da coerência.





*Figura III.4 - Novo método de cálculo para função de coerência*

A aplicação de métodos baseados em um único microfone na estimação das  $SNR_{prio1}$  e  $SNR_{prio2}$  impõe o uso de um sistema de detecção voz/silêncio (DVS). No capítulo V discute-se o uso da DVS, que é um ponto importante na aplicação de sistemas baseados em um único microfone.

A necessidade de uma DVS no novo método de cálculo da função de coerência apresenta-se com uma desvantagem em relação ao método original. Em termos de complexidade computacional, os dois métodos estão próximos, com o novo método necessitando das mesmas transformações em frequência usadas no método original.

### **III.7 - Introdução de um modelo a dois estados**

Em muitos métodos usam-se sistemas de detecção de intervalos de voz e silêncio, que selecionam as pausas naturais durante uma conversação, considerando total ausência de sinal de voz nesses intervalos. No entanto, é possível que, mesmo nos instantes considerados de pausa, exista uma fraca presença de voz em algumas frequências. De maneira análoga, podem-se ter trechos de conversação onde algumas componentes apresentam fraca energia como, por exemplo, nos instantes onde se tem os sons ditos *sonoros*. Estas considerações sugere um modelo estatístico da probabilidade de presença

ou ausência de voz durante o processamento, denominado de *modelo a dois estados*, que pode levar a um aumento do desempenho do filtro redutor de ruído.

A idéia inicial de usar-se um modelo a dois estados foi introduzida por McAulay & Malpass, sendo mais tarde explorada também por Ephraim & Malah. Esse modelo pode ser estendido sem perdas de generalidade para os filtros clássicos aqui desenvolvidos. Assim, pode-se incorporar aos filtros clássicos um fator de incerteza, de forma a obter-se um novo filtro  $Hr(\omega)$  dado por:

$$Hr(\omega) = H(\omega) \cdot I(\omega, q) \quad (\text{III.34})$$

onde  $I(\omega, q)$  é o fator de incerteza em cada frequência  $\omega$  analisada, dado por

$$I(\omega, q) = \frac{\Lambda[Y(\omega), q]}{1 + \Lambda[Y(\omega), q]} \quad (\text{III.35})$$

Na equação III.35,  $\Lambda(\omega, q)$  é uma relação de máxima-verossimilhança que representa o estimador de incerteza na análise do sinal, e é dado por:

$$\Lambda[Y(\omega), q] = \mu \cdot \frac{p[Y(\omega) | H^1]}{p[Y(\omega) | H^0]} \quad (\text{III.36})$$

Para cada frequência  $\omega$  analisada, tem-se:

$$\mu = [1 - q] / q;$$

$q$  é a probabilidade de ausência de voz ;

$H^1$  é a hipótese de presença do sinal de voz;

$H^0$  é a hipótese de ausência de sinal de voz.

Na aplicação do fator de incerteza, a probabilidade  $q$  de ausência de voz é normalmente a mesma para todas as frequências.

No método desenvolvido por McAulay & Malpas, utiliza-se  $q = 0,5$  e aplica-se o critério de minimização do erro quadrático médio, de forma a se obter:

$$\Lambda(\omega) = \exp[-SNR\_Priori(\omega)] \cdot Io\left[2 \cdot \sqrt{SNR\_priori(\omega) \cdot SNR\_post(\omega)}\right] \quad (III.37)$$

onde  $Io[\cdot]$  representa a função de Bessel modificada do tipo um e de ordem zero.

No desenvolvimento feito por Ehpraim & Malah,  $q$  é uma variável que pode ser controlada. Para a estimação da incerteza foi aplicado o critério da máxima-verossimilhança, obtendo-se:

$$\Lambda[Y(\omega), q] = \mu \cdot \frac{\exp\left[\frac{SNR\_post(\omega) \cdot SNR\_prio(\omega)}{1 + SNR\_prio(\omega)}\right]}{1 + SNR\_prio(\omega)} \quad (III.38)$$

Nas equações III.37 e III.38 observa-se a presença da  $SNR\_prio$ , o que vem confirmar sua importância nos sistemas de redução de ruído.

Para avaliar o comportamento dos dois estimadores de incerteza, na figura III.5 tem-se as curvas para o fator de incerteza,  $I(\omega)$ . Para os dois estimadores, fixou-se a probabilidade de ausência de voz em 0,5. Esta visualização é importante pois mostra as consequências de uma estimação ruim da  $SNR\_prio$ . Por exemplo, na figura III.5.A, onde tem-se  $SNR\_post=0$  dB, o sistema pode introduzir altas distorções no sinal processado se se estima uma  $SNR\_prio$  acima de 0 dB. Observa-se no início que a incerteza está próxima do valor inicial estabelecido, 0,5. No entanto, com o aumento da  $SNR\_prio$ , a tendência é uma rápida queda na probabilidade de presença de voz. A explicação vem do fato de que uma baixa  $SNR\_post$  indica uma baixa probabilidade de presença de voz, já que as análises realizadas não consideram as fases dos sinais. Por exemplo, considere-se uma frequência onde se tem voz, mas que devido à alta potência do ruído, de fase contrária, tem-se uma baixa  $SNR\_post$ . Neste caso, a aplicação do fator de incerteza irá provocar uma forte atenuação nessa frequência. No caso de uma  $SNR\_post$  alta, o fator de incerteza praticamente não interfere na operação do filtro. Nos casos intermediários, o efeito do fator de incerteza dependerá das estimações de  $SNR\_prio$  e  $SNR\_post$ .

Comparando-se os dois métodos, verifica-se que eles apresentam comportamentos parecidos. Observa-se, entretanto, que o método proposto por Ephraim & Malah é menos sensível aos aumentos de  $SNR_{prio}$ , para uma mesma  $SNR_{post}$  (figura III.5, A e B)

Na implementação dos filtros propõe-se um controle de  $I(\omega)$  a partir dos valores de  $SNR_{post}$ , de forma que:

$$I(\omega, q) = \begin{cases} \frac{\Lambda[Y(\omega), q]}{1 + \Lambda[Y(\omega), q]}, & \text{se } SNR_{post}(\omega) > 0 \text{ dB} \\ 1 - q, & \text{se } SNR_{post}(\omega) \leq 0 \text{ dB} \end{cases} \quad (III.39)$$

As restrições impostas pela equação III.39 evitam uma queda descontrolada da potência do sinal processado, mas deixa um nível de ruído residual mais significativo.

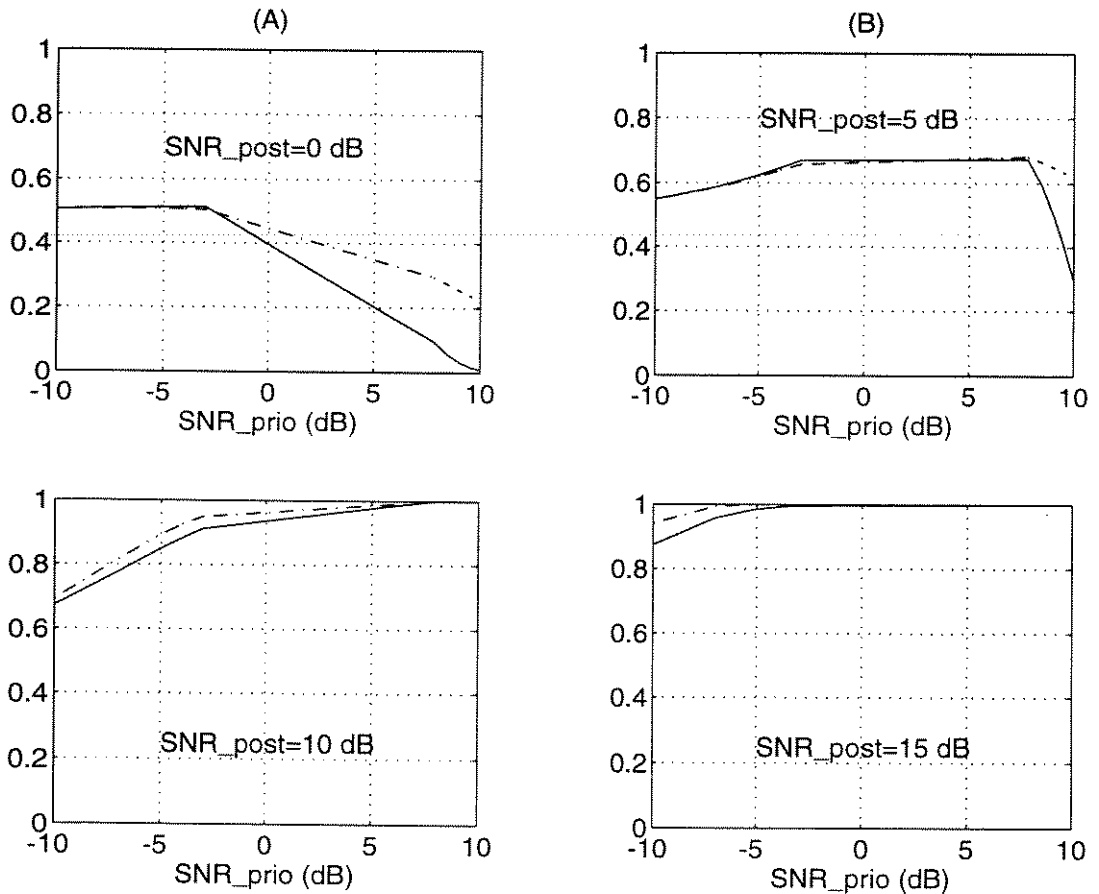


Figura III.4 - Curvas dos fatores de incerteza para o sinal de voz  
( — McAulay & Malpass, - - - Ephraim & Malah )

### III.8 - CONCLUSÕES

Neste capítulo mostrou-se que os filtros clássicos de redução de ruído utilizam como parâmetro principal a relação sinal/ruído a posteriori ( $SNR_{post}$ ). Uma análise detalhada destes filtros permitiu concluir que a presença do ruído musical nos sinais processados pelos os mesmos é resultado do comportamento ruim da  $SNR_{post}$ .

Um estudo mais completo da técnica desenvolvida por Ephraim & Malah, que possibilita um processamento sem introdução do ruído musical, mostrou que o seu sucesso pode ser resultado do uso da relação sinal/ruído a priori ( $SNR_{prio}$ ) como parâmetro básico, tendo em vista que o modelo estatístico adotado é questionável.

Finalmente propôs-se a introdução da  $SNR_{prio}$  nos filtros clássicos de redução de ruído, com um ou dois microfones, o que resultou no desenvolvimento de novos tipos de filtros. Isto significa uma descoberta de novos métodos para serem aplicados em melhoramento de sinais de voz com uma baixíssima complexidade matemática. No caso de um único microfone, estes novos filtros resolvem o problema básico do ruído musical, introduzido pelos filtros clássicos. Este resultado é muito importante pois, desde a apresentação da subtração espectral como método de redução de ruído por BOLL (1979) ou BEROUTI et alii (1979), não se conseguiam resultados com ausência de ruído musical. Já no caso de dois microfones com base na função de coerência, elimina-se o problema da reverberação e aumenta-se o desempenho na redução do ruído nas frequências baixas.

Além disso, observando-se as vantagens do uso de um estimador de incerteza nas análises dos sinais de voz (McAULAY & MALPASS, 1980 e EPHRAIM & MALAH, 1984), propôs-se uma generalização do mesmo, que pode ser aplicado a qualquer filtro redutor de ruído sem perda de generalidade. Mostra-se, no entanto, que o fator de incerteza deve ser controlado de acordo com os parâmetros  $SNR_{post}$  e  $SNR_{prio}$ .

## **CAPÍTULO IV**

### ***TÉCNICAS AUXILIARES RELACIONADAS À REDUÇÃO DE RUÍDO***

#### ***IV.1 - Introdução***

O desenvolvimento teórico de filtros de redução de ruído é apenas uma parte do trabalho no melhoramento de sinais de voz. Considerando a aplicação de técnicas de processamento digital de sinais, mais particularmente a estimação espectral de curto prazo, torna-se necessário que alguns parâmetros importantes, como tamanho da janela de análise, tipo de janela, intervalo de sobreposição entre janelas, etc., sejam ajustados adequadamente. Um dos fatores mais problemáticos na prática é a estimação da potência do ruído quando se usam técnicas baseadas em um único microfone. A dificuldade neste caso vem da necessidade de se detectar os intervalos de voz e silêncio (DVS), tarefa difícil quando se têm baixas relações sinal/ruído.

Neste capítulo apresenta-se inicialmente um modelo generalizado da estimação de curto prazo em sistemas de MSV. Baseado neste modelo discutem-se e propõem-se soluções para sua implementação.

## IV.2 - Estimação usando a transformada de Fourier

A Transformada Discreta de Fourier na estimação de curto prazo (STFT) tem sido largamente utilizada no processamento de sinais de voz (RABINER & SCHAFER, 1988). Sua aplicação viabiliza a estimação de pequenas parcelas do espectro do sinal de voz, consideradas estacionárias quando os intervalos de análise variam entre 20ms e 40ms.

Na figura IV.1 tem-se um diagrama resumindo a aplicação da STFT em sistemas de MSV. O sinal, já digitalizado, passa inicialmente por uma janela temporal, com dois objetivos: o primeiro é garantir um sinal discreto de duração limitada para permitir o uso da transformada discreta de Fourier; o segundo é assegurar trechos do sinal que sejam praticamente estacionários. O tipo de janela a ser adotado não constitui aqui um objeto de estudo, visto que existe uma variedade de janelas utilizáveis. O importante é definir a mais conveniente, o que pode ser feito através de simulações. Associado à janela, deve-se definir também o intervalo de sobreposição das janelas consecutivas. Um valor típico é 50% para uso com a janela de Hanning.

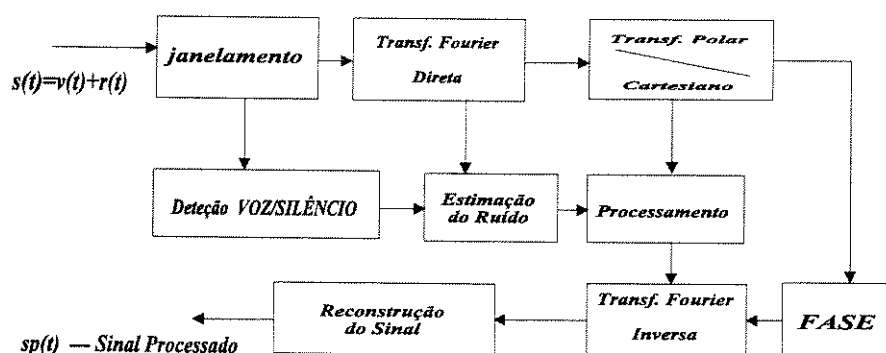


Figura IV.1 - Melhoramento de sinais de voz baseado na STFT

No caso de trabalhar-se com apenas um microfone, é necessário uma detecção de voz/silêncio (DVS). O objetivo é determinar os intervalos de silêncio para se estimar a potência do ruído, que é suposto estacionário.

Na etapa seguinte deve-se escolher o tipo de filtro para o processamento do sinal ruidoso, que pode ser obtido a partir das várias técnicas apresentadas no capítulo III. Estas técnicas não consideram a estimação da fase. Na aplicação da transformada inversa de Fourier usa-se a fase do sinal ruidoso. Finalmente, para reconstruir o sinal estimado usa-se um método de síntese. O método mais adotado em MSV, quando baseado na STFT, é o “*Overlap Addition*” (OLA), podendo também utilizar-se um banco de filtros. Um ponto importante no método OLA é a necessidade de uso do recurso do “*Zero Padding*”<sup>1</sup> no sinal de entrada. O objetivo é evitar que o sinal reconstituído apresente “aliasing”<sup>2</sup>.

Para denotar o uso da STFT nas implementações, usar-se-á a partir de agora uma notação em que se destaca a janela ( $w_i$ ) analisada e a frequência ( $f_k$ ) processada.

### **IV.3 - Detecção voz/silêncio**

A aplicação de sistemas baseados em um único microfone impõe o uso de uma técnica de detecção voz/silêncio (DVS) para segmentar o sinal ruidoso. Esta tarefa pode parecer simples à primeira vista, se for considerado, por exemplo, um sinal de voz puro. No entanto, para sinais com baixas SNR, o problema torna-se complexo.

Nos estudos desenvolvidos neste trabalho, onde o objetivo principal foi a obtenção de filtros simples e eficazes para uso em MSV, considerou-se uma DVS ideal. No entanto, considerando a aplicação aos sistemas *HFTs*, propõem-se dois métodos para DVS usando dois microfones<sup>3</sup>. Um, estudado inicialmente por LE BOUQIN & FAUCON (1993), tem

---

<sup>1</sup> Processo em que uma determinada sequência tem seu comprimento aumentado usando-se ZEROS.

<sup>2</sup> O uso da STFT em processamento de sinais de voz é bem discutido por RABINER & SCHAFER (1988).

<sup>3</sup> Existem vários métodos para DVS como, por exemplo, os propostos por FREEMAN et alii (1989) e LYNCH et alii (1987).



como base a função de coerência. O outro utiliza um microfone auxiliar para a obtenção de uma potência média do ruído.

#### **IV.3.1 - DVS usando a função de coerência**

Nos capítulos II e III viu-se que dois microfones próximos podem fornecer sinais de voz coerentes e ruídos quase incoerentes, mesmo considerando as restrições das baixas frequências. Com esta definição explora-se a coerência na seleção de trechos de voz e silêncio.

Para uma determinada janela de análise ( $w_i$ ), cada frequência ( $f_k$ ), contribuirá com seu grau de coerência. Assim, pode-se definir um valor médio de coerência para cada janela, o qual é dado por:

$$\overline{|\gamma(w_i)|} = \frac{2}{N} \sum_{k=0}^{N/2-1} |\gamma(w_i, f_k)| \quad (\text{IV.1})$$

onde  $N$  é o comprimento da FFT usada na STFT.

Considerando que o sinal de voz apresenta uma coerência aproximadamente uniforme em todas as frequências e que o ruído apresenta um certo grau de coerência apenas nas baixas frequências, obtém-se no final um valor médio que permite a detecção de voz e silêncio.

Nas figuras IV.2 e IV.3 têm-se exemplos do cálculo da equação IV.1 e o equivalente obtido para a DVS. A obtenção pura e simples do valor médio para a função de coerência não define uma DVS. É preciso identificar a janela exata onde tem-se voz e a janela exata onde tem-se somente ruído.

Considerando que existe uma pausa natural entre o momento em que se liga um sistema de transmissão e o momento em que se começa a falar, pode-se fazer um aprendizado inicial do valor médio da coerência para o ruído.

Assumindo  $L$  como sendo o número de janelas iniciais onde somente o ruído está presente, define-se uma média inicial de coerência que é dada por:

$$\phi = \frac{1}{L} \sum_{i=0}^{L-1} \frac{2}{N} \sum_{k=0}^{N/2-1} |\gamma(w_i, f_k)| \quad (\text{IV.2})$$

O valor de  $\phi$  dá um indicativo da coerência inicial do ruído e pode ser usado para uma tomada de decisão. O critério de decisão é dado por:

$$\text{VOZ:} \quad \text{se } \frac{\overline{|\gamma(w_i)|}}{\phi} < \varepsilon$$

*NÃO-VOZ:* caso contrário

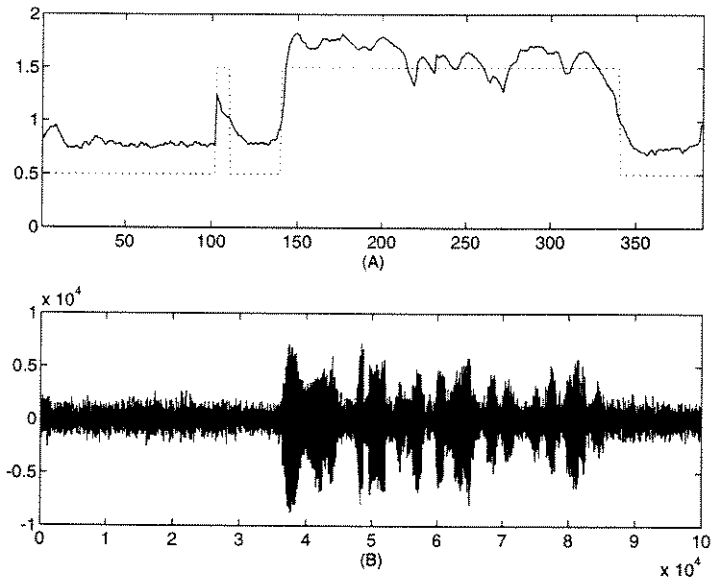
O parâmetro  $\varepsilon$  define o critério e, de acordo com o resultado de várias simulações, varia no intervalo de 1,0 a 1,5.

De acordo com as figuras IV.2 e IV.3, obtidas para  $\varepsilon = 1$ , verifica-se um desempenho altamente satisfatório de DVS. Nestes exemplos, o valor de  $\phi$  foi obtido fazendo-se  $L = 10$ , o que significa um tempo de aprendizado para a coerência de  $312,5\text{ms}^4$ . Ainda com relação às figuras IV.2 e IV.3, as curvas para o valor médio da função de coerência estão normalizadas com relação a  $\phi$ , isto é, tem-se na verdade  $\overline{|\gamma(w_i)|} / \phi$ . Também é possível no trecho de silêncio (figura IV.2), que o ruído tenha sido identificado como sinal de voz, o que aconteceria se houvesse uma forte coerência do ruído nesse intervalo. Esta identificação não perturba o processamento. Na verdade, pode haver nesses casos a detecção de um ruído não-estacionário. Isto é importante, pois o fato de tomar-se essas janelas como sendo de sinal de voz útil ( $\text{DVS}=1,5$ , neste exemplo) evitar-se-á um aumento abrupto da potência do ruído no intervalo analisado. Este procedimento elimina a possibilidade de aparecer distorções no sinal processado devido a uma DVS incorreta. Logo, os trechos de ruído não-estacionários terão apenas uma redução do ruído de fundo, considerado estacionário.

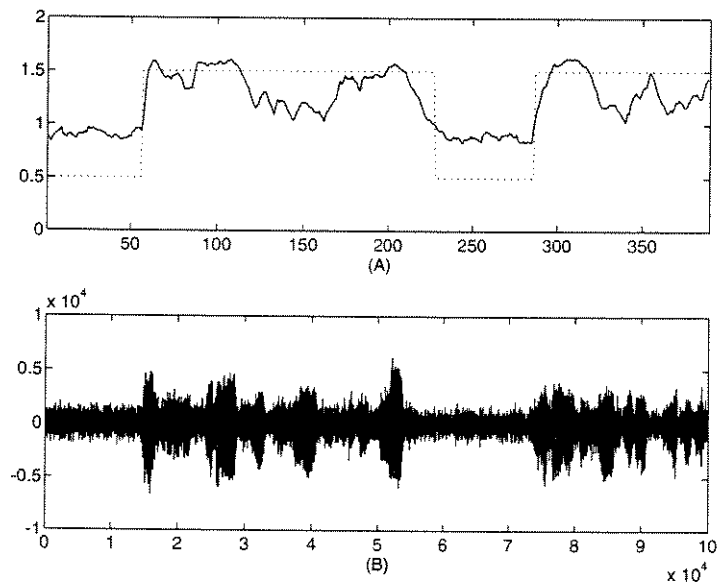
O principal inconveniente no uso da DVS com base na função de coerência é o grande número de cálculos necessários. No entanto, se esta técnica for combinada com um filtro redutor de ruído de baixa complexidade, como por exemplo a subtração espectral, o resultado final pode ser um sistema eficiente e com uma DVS robusta.

---

<sup>4</sup> Trabalhou-se com uma FFT de 512 pontos para um sinal amostrado a 16 kHz.



*Figura IV.2* - (a) - Valor médio da função de coerência e DVS equivalente  
(b) - Sinal ruidoso 1



*Figura IV.3* - (a) - Valor médio da função de coerência e DVS equivalente  
(b) - Sinal ruidoso 2

#### IV.3.2 - DVS usando um microfone auxiliar isolado do sinal de voz

A possibilidade de uso de um sistema adaptativo no MSV com um microfone isolado é questionável devido à dificuldade de se obterem sinais realmente descorrelacionados e não defasados. No entanto, pode-se usar um segundo microfone para estimar uma potência média auxiliar, que servirá na identificação dos intervalos de voz e silêncio. Neste caso, não é preciso ter um ruído completamente descorrelacionado do sinal de voz e não existem problemas de defasagem.

Considerando as características do ambiente veicular, decidiu-se colocar um único microfone para o locutor. O segundo foi colocado no bagageiro do veículo para captar somente ruído. O resultado foi a obtenção de um ruído com potência média muito próxima da potência média do ruído presente no sinal captado pelo microfone do locutor. Pensou-se em estimar de forma direta a potência do ruído mas os testes iniciais mostraram que o espectro em frequência do ruído captado no microfone colocado no bagageiro apresentava alguns valores superiores aos obtidos no ruído estimado diretamente do sinal captado pelo microfone do locutor. Isto poderia introduzir variações importantes na SNR e conseqüentemente distorções no sinal processado.

O sistema proposto e utilizado para DVS usando um microfone auxiliar e isolado é apresentado na figura IV.4. É importante um ajuste na amplitude do ruído para que ele apresente a mesma variação dinâmica do ruído real<sup>5</sup>. Isto é feito aproveitando-se o fato de que, ao se ligar o sistema, apenas o ruído estará presente (como suposto para uma DVS usando a função de coerência). Nas simulações realizadas este ajuste foi feito numericamente antes do processamento. A partir do momento em que os sinais estão ajustados, calcula-se a SNR por janela, que é dada por:

$$SNR(w_i) = 10 \cdot \log \left[ \frac{\sum_{n=0}^{L-1} y^2(n)}{\sum_{n=0}^{L-1} r_2^2(n)} \right] \quad (IV.3)$$

---

<sup>5</sup> Ruído presente no sinal captado pelo microfone do locutor.

onde  $L$  representa o tamanho da janela temporal escolhida e  $w_i$  a janela analisada.

Para evitar que transições abruptas ocorram na SNR entre janelas, suaviza-se o resultado por meio de uma filtragem recursiva de primeira ordem. Este procedimento tem por objetivo facilitar a decisão entre VOZ e SILÊNCIO. A filtragem recursiva é dada por:

$$\overline{\overline{SNR}}(w_i) = \alpha \cdot \overline{\overline{SNR}}(w_{i-1}) + (1 - \alpha) \cdot SNR(w_i) \tag{IV.4}$$

A equação IV.4 introduz uma dependência da SNR atual em relação a SNR calculada nas janelas anteriores. A constante  $\alpha$ , que pode ser interpretada como um fator de esquecimento, controla o grau de dependência e varia normalmente entre 0 e 1.

Na figura IV.5 tem-se um exemplo de DVS usando este método. Usou-se o mesmo sinal da figura IV.3 e fez-se  $\alpha=0,5$ . Tem-se a SNR por janela de acordo com a equação IV.3 (figura IV.5a) e após a filtragem, como na equação IV.4 (figura IV.5b). Verifica-se que a filtragem permite uma melhor definição dos intervalos de voz e silêncio.

Como no método baseado na função de coerência, tem-se um limiar para a decisão entre voz e silêncio. O critério adotado parte do pressuposto que a SNR média nos intervalos de silêncio é limitada a um fator  $\delta$ . Com algumas simulações verificou-se que normalmente  $\delta < 2$ . Para obtenção da figura IV.5 adotou-se  $\delta = 0,1$ .

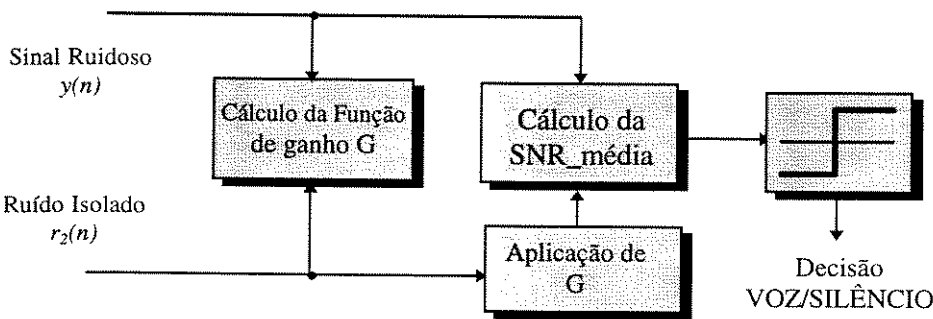


Figura IV.4 - Esquema básico de DVS usando um microfone auxiliar

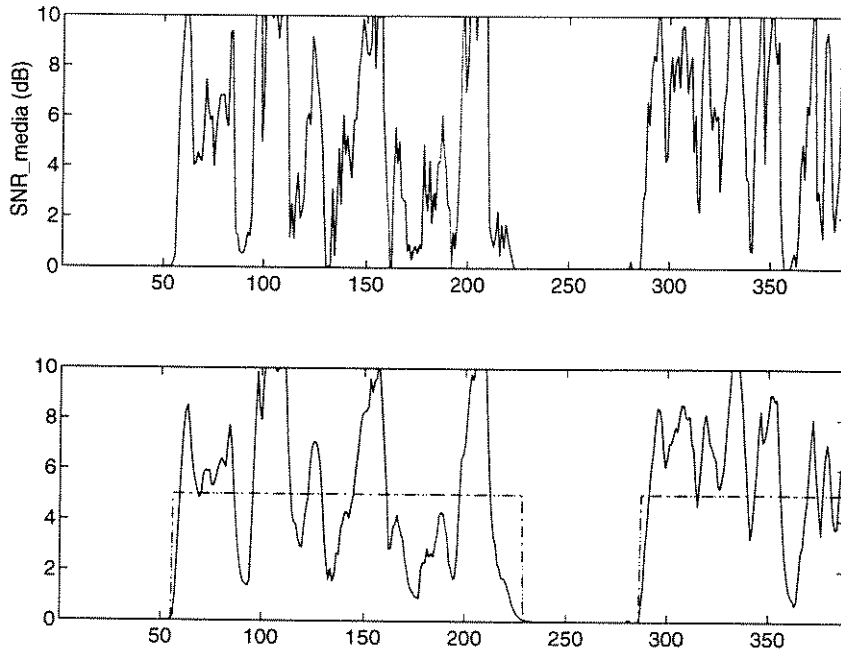


Figura IV.5 - (a) - SNR (b) DVS equivalente e SNR\_média filtrada

#### IV.4 - Estimação da potência do ruído

A aplicação dos filtros redutores de ruído desenvolvidos neste trabalho passa por uma estimação da potência do ruído. Considerando os intervalos de silêncio definidos pela DVS e o ruído estacionário, pode-se estimar a potência do ruído usando uma filtragem recursiva de primeira ordem, como segue:

$$\sigma_r^2(w_i, f_k) = \beta \cdot \sigma_r^2(w_{i-1}, f_k) + (1 - \beta) \cdot |Y(w_i, f_k)|^2 \quad (\text{IV.5})$$

onde  $\beta$  é uma constante que define a dependência do ruído na janela atual de análise em relação às janelas anteriores e  $|Y(w_i, f_k)|^2$  é a potência do sinal ruidoso (que será selecionado como ruído de acordo com a DVS). Esta estimação é fundamental para um

bom desempenho do filtro redutor de ruído. Por exemplo, uma variação brusca de potência de uma janela a outra pode contribuir fortemente para o surgimento do ruído musical. Nesse casos, um valor adequado do fator  $\beta$  passa a ser muito importante.

O cálculo de  $\beta$  é baseado no tamanho da janela de análise, da frequência de amostragem do sinal e da constante de tempo que garanta o esquecimento das informações passadas. Assim, tem-se:

$$\beta = \exp\left[-L / (2 \cdot f_a \cdot T_c)\right] \quad (\text{IV.6})$$

onde  $L$  é o tamanho da janela,  $f_a$  a frequência de amostragem e  $T_c$  o tempo associado ao fator de esquecimento. Um valor típico para  $T_c$  é 140 ms.

Existem alguns algoritmos que exploram a estimação do ruído com um fator de superestimação, principalmente os que são baseados na subtração espectral direta. No entanto, ao utilizarem-se os mesmos algoritmos com base na  $SNR_{prio}$ , pode-se introduzir fortes distorções no sinal de voz, com pouquíssimo ganho em relação à redução do ruído.

#### **IV.5 - Estimação da SNR a posteriori e a priori**

A relação sinal/ruído a posteriori pode ser obtida diretamente do sinal ruidoso captado e da potência do ruído estimada. Assim, considerando que  $|Y(w_i, f_k)|^2$  representa a potência do sinal ruidoso, tem-se:

$$\hat{SNR}_{post}(w_i, f_k) = \frac{|Y(w_i, f_k)|^2}{\sigma_r^2(w_i, f_k)} \quad (\text{IV.7})$$

Verifica-se que a  $SNR\_post$  acompanha de forma direta as variações do sinal ruidoso, o que confirma sua sensibilidade em relação às mudanças bruscas de fase do ruído.

A estimação da relação sinal/ruído a priori depende do próprio sinal de voz estimado, o que caracteriza um sistema não-causal. Para resolver este problema, considera-se que a variação de potência de uma janela para outra não é significativa. Estas considerações justificam o uso do estimador recursivo proposto por Ephraim & Malah, que é dado por:

$$\hat{SNR\_prio}(w_i, f_k) = \mu \cdot \frac{|\hat{V}(w_{i-1}, f_k)|^2}{\sigma_r^2(w_i, f_k)} + (1 - \mu) \cdot T[\hat{SNR\_post}(w_i, f_k) - 1] \quad (IV.8)$$

onde  $T[\cdot]$  indica uma transformação sobre  $[SNR\_post(w_i, f_k) - 1]$  e  $\mu$  representa o grau de dependência da  $SNR\_prio$  com relação a  $SNR\_post$ . Lembrando o capítulo III, onde definiu-se  $SNR\_post(w_i, f_k) = SNR\_prio(w_i, f_k) + 1$ , o uso da  $SNR\_post$  na estimação da  $SNR\_prio$  representa uma contribuição da  $SNR\_prio$  da atual janela de análise, mas que deverá passar por uma transformação, visto que  $SNR\_post$  pode ser menor que 1. Normalmente usa-se a retificação de meia-onda como transformação, como proposto por BOLL (1979). No entanto, isto significa uma potência NULA para a voz naquela frequência, o que não parece ser razoável. Assim, propõe-se nesses casos a utilização de um valor mínimo para  $SNR\_post$ , aqui representado como  $\delta$ , de forma que:

$$Se \ \hat{SNR\_post}(w_i, f_k) - 1 \leq 0 \ \text{então} \ \ SNR\_post(w_i, f_k) = \delta \quad (IV.9)$$

Voltando à equação IV.8, tem-se que a contribuição da  $SNR\_post$  na  $SNR\_prio$  depende do valor de  $\mu$ . Nos estudos realizados, como também no trabalho original de EPHRAIM & MALAH (1984), verificou-se que as fortes variações na  $SNR\_post$  podem afetar o desempenho da  $SNR\_prio$ . Entretanto, se for definido  $\mu=0$ , perde-se o ajuste fino na estimação da  $SNR\_prio$ . Assim, o valor típico adotado para  $\mu$  varia entre 0,9 e 1. Particularmente neste trabalho, observou-se que 0,98 é um valor bastante razoável.



## **IV.6 - Conclusões**

Neste capítulo foram mostrados os pontos principais envolvidos na aplicação de técnicas de MSV, quando se usa a STFT.

Um ponto importante para implementação das técnicas estudadas no capítulo II é a detecção voz/silêncio (DVS). É a partir da segmentação do sinal em intervalos de voz e silêncio que estima-se a potência média do ruído. Este fato é importante pois os parâmetros básicos dos filtros, isto é, as relações sinal/ruído a posteriori e a priori, são baseados na potência média estimada do ruído. Neste estudo, apresentaram-se dois métodos para DVS, baseados em dois microfones, obtendo-se bons resultados.

Com relação a  $SNR_{prio}$ , a estimação é baseada na janela anterior processada, incorporando-se, no entanto, uma parcela da  $SNR_{post}$  da janela atual de análise. Este tipo de estimação é denominado de “decisão dirigida” e não é linear, o que dificulta uma análise teórica do seu comportamento (EPHRAIM & MALAH, 1984).

## **CAPÍTULO V**

### **RESULTADOS OBTIDOS E AVALIAÇÕES**

#### **V.1 - Introdução**

Os métodos estudados e propostos nos capítulos anteriores foram implementados e avaliados considerando-se medidas objetivas e subjetivas. Com relação às medidas objetivas, buscou-se principalmente uma forma de se avaliarem as distorções resultantes do processamento e o nível de redução de ruído. Para a avaliação, diferentes métodos são disponíveis (DELLER, 1993, e GRAY & MARKEL, 1976), mas neste trabalho optou-se pela medida da distância cepstral como forma de avaliar as distorções, e a medida do ganho na SNR média temporal para obter os níveis médios de redução de ruído. A principal vantagem da distância cepstral em relação aos outros métodos segundo, segundo GRAY (1976), é que ela permite a obtenção de curvas de distorção melhores relacionadas aos resultados que se obtêm com testes subjetivos. No caso do uso do ganho na SNR ruído temporal, a vantagem é a simplicidade e o bom indicativo da redução média do ruído. Ainda como forma de visualização da redução de ruído, usou-se o espectrograma, que dá um indicativo freqüencial do nível de redução de ruído. No que diz respeito às avaliações subjetivas, os pontos explorados foram a inteligibilidade e o conforto auditivo do sinal processado, quando comparado ao sinal original. Os testes realizados seguiram um caráter informal, considerando a existência de metodologias sofisticadas para avaliações subjetivas. No entanto, estes testes são suficientes para se avaliarem os resultados, obtendo-se uma boa equivalência com as medidas objetivas (distorção associada à inteligibilidade e redução de ruído ao conforto auditivo).

As avaliações e comparações foram feitas para as seguintes técnicas: Subtração Espectral Original, Ephraim & Malah, Função de Coerência Original, Função de Coerência baseada na SNR\_prio e Subtração Espectral baseada na SNR\_prio (considerando também as incertezas nas medidas). Estas técnicas foram escolhidas porque são suficientes para que se demonstre o que há de melhor e de pior dentre as técnicas apresentadas neste trabalho.

## ***V.2 - Avaliação objetiva usando a distância cepstral e o ganho na SNR temporal***

### ***V.2.1 - Distância cepstral - Teoria básica***

A distância cepstral representa uma medida Euclidiana ponderada entre dois sinais, usando como base seus cepstros. Na verdade, isto representa uma distância direta entre os coeficientes dos seus modelos LPC. O objetivo é mostrar o quanto dois sinais estão distanciados espectralmente.

De um modo geral, sejam

$$\frac{\sigma}{S(z)} \quad e \quad \frac{\sigma'}{S'(z)} \quad (V.1)$$

os modelos espectrais de dois sinais. A diferença entre o logaritmo destes dois modelos fornece uma distância D, que é dada por:

$$D(\theta) = \ln \left( \frac{\sigma}{|S(e^{j\theta})|} \right) - \ln \left( \frac{\sigma'}{|S'(e^{j\theta})|} \right) \quad (V.2)$$

sendo  $\theta$  a frequência normalizada no plano  $z$  e  $\pi$  representa metade da frequência de amostragem.

A distância espectral logarítmica é então definida como:

$$(d_p)^p = \frac{1}{2\pi} \int_{-\pi}^{\pi} |D(\theta)|^p d\theta \quad (V.3)$$

onde  $p$  pode ser definido como um parâmetro de ajuste da medida. Para grandes valores de  $p$ , a distância obtida reflete apenas as grandes distorções.

De acordo com a equação V.1, se  $S(z)$  e  $S'(z)$  são polinômios em  $z^{-1}$  com raízes no interior do círculo unitário e  $S(\infty)=1$ , mostra-se através de uma expansão em série de Taylor (GRAY, 1976) que

$$\ln[S(z)] = - \sum_{k=1}^{\infty} c_k z^{-k} \quad (V.4)$$

onde os  $c_k$  representam os coeficientes cepstrais. Conseqüentemente, e considerando-se a equação V.4, tem-se que

$$\ln[\sigma^2 / |S(e^{j\theta})|^2] = - \sum_{k=-\infty}^{\infty} c_k z^{-j\theta k} \quad (V.5)$$

onde  $c_0 = \ln(\sigma^2)$  e  $c_{-k} = c_k$ .

Das equações V.2 e V.5 tem-se que a distância cepstral quadrática é dada por:

$$(d_2)^2 = \sum_{k=-\infty}^{\infty} (c_k - c'_k)^2 = (c_0 - c'_0)^2 + \sum_{k=1}^{\infty} (c_k - c'_k)^2 \quad (V.6)$$

A implementação desta equação é baseada na limitação do número de coeficientes cepstrais. Apesar de parecer um “truncamento” da medida, o que se faz de fato é suavizar a distância de acordo com o número de coeficientes escolhidos. Assim, tomando-se  $L$  coeficientes cepstrais, a distância cepstral final é dada por:

$$(d_{cep})^2 = (c_0 - c'_0)^2 + \sum_{k=1}^L (c_k - c'_k)^2 \quad (V.7)$$

Naturalmente, quanto maior o valor de  $L$  melhor a definição da distância. Mas, para o caso de se usar um modelo LPC de ordem  $x$  para a obtenção dos coeficientes cepstrais, um valor razoável para  $L$  será  $L=2x$ .

### **V.2.2 - Ganho na SNR**

Existem várias possibilidades para obter-se uma medida da SNR, mesmo querendo-se utilizá-la para uma avaliação das distorções presentes em um determinado sinal (DELLER, 1993)]. No entanto, neste trabalho o objetivo principal é ter um valor aproximado do ganho obtido na SNR para mostrar o nível de redução de ruído que se pode obter com um determinado método.

Seja  $y(n)$  o sinal ruidoso e  $\hat{v}(n)$  o sinal de voz estimado, ambos já discretizados. Uma avaliação temporal de suas SNR pode ser obtida de acordo com as seguintes equações:

$$SNR_y = \frac{\sum_n y^2(n)}{\sum_n r^2(n)} \quad (V.8)$$

$$SNR_{\hat{v}} = \frac{\sum_n \hat{v}^2(n)}{\sum_n \hat{r}^2(n)} \quad (V.9)$$

sendo:

$y(n)$  - sinal ruidoso;

$r(n)$  - o ruído presente no sinal ruidoso;

$\hat{v}(n)$  - o sinal de voz puro passado pelo filtro de redução de ruído;

$\hat{r}(n)$  - ruído passado pelo filtro de redução de ruído.

O valor de  $n$  varia de acordo com o tamanho da janela analisada. As potências dos ruídos são obtidas durante os intervalos de silêncio.

O ganho na relação sinal ruído, em decibéis, é definido pela relação:

$$G_{(dB)} = 10 \cdot \log_{10} \frac{SNR_{\hat{v}}}{SNR_y} \quad (V.10)$$

Quando o processamento do sinal ruidoso é bom, o ganho na equação V.10 é superior a 0 dB nos intervalos de voz e aproximadamente igual a 0 dB nos intervalos de silêncio. O valor obtido dá um indicativo global da redução de ruído, sem mostrar detalhes de variações frequenciais.

### **V.2.3 - Obtenção dos sinais para as avaliações objetivas**

Na aplicação da distância cepstral, o uso direto dos sinais ruidoso e processado levará em consideração as possíveis diferenças de potência entre eles. Naturalmente, estas diferenças podem ser um indicativo de distorção, mas elas não são, geralmente, correlacionadas com o aspecto auditivo. Uma maneira encontrada para privilegiar apenas as distorções que são correlacionadas com o aspecto auditivo é usar as hipóteses iniciais consideradas nos próprios métodos de redução de ruído.

Seja  $v(n)$  o sinal de voz puro,  $r(n)$  o ruído e  $y(n) = v(n) + r(n)$  o sinal ruidoso. Considerando a hipótese de linearidade, a passagem do sinal ruidoso  $y(n)$  pelo filtro redutor de ruído com resposta ao impulso  $h(n)$  é representada, em frequência, por:

$$Y(\omega) \cdot H(\omega) = V(\omega) \cdot H(\omega) + R(\omega) \cdot H(\omega) \quad (V.11)$$

Conhecendo-se os sinais  $v(n)$  e  $r(n)$  a priori<sup>1</sup>, adotou-se a seguinte metodologia para obtenção dos sinais usados nas medidas objetivas:

- 1 - A partir do sinal ruidoso  $y(n)$  calcula-se o filtro redutor de ruído de acordo com o método adotado;
- 2 - Passam-se  $v(n)$  e  $r(n)$  pelo filtro calculado em 1 obtendo-se os sinais  $\hat{v}(n)$  e  $\hat{r}(n)$ ;
- 3 - Aplica-se, finalmente, as equações V.7, V.8, V.9 e V.10 para obter-se a distância cepstral e o ganho na SNR temporal.

---

<sup>1</sup> Hipótese válida apenas para testes de desempenho, quando se tem disponível os sinais de voz e ruído, separadamente.

O diagrama apresentado na figura V.1 resume a obtenção dos sinais usados nos testes objetivos descritos anteriormente.

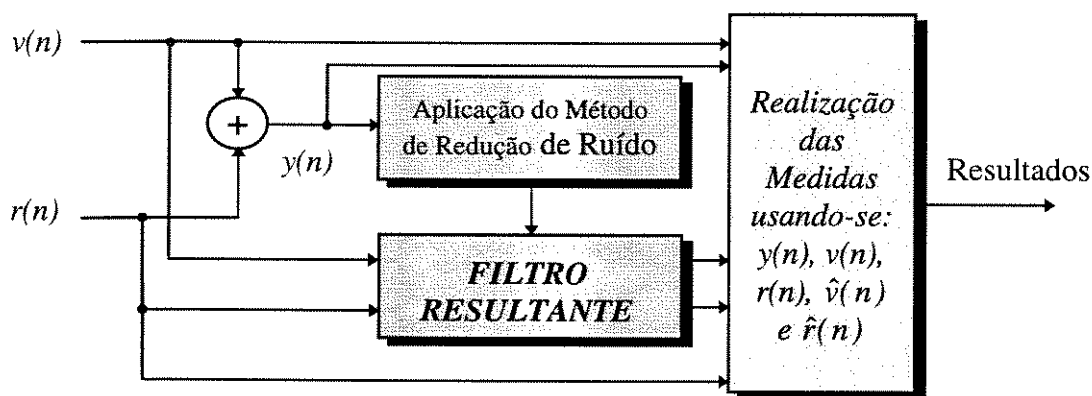


Figura V.1 - Método usado na avaliação objetiva

### V.3 - Avaliação subjetiva

O uso de medidas objetivas na avaliação de sistemas de redução de ruído possibilita uma comparação entre métodos avaliando-se principalmente a melhoria na qualidade física do sinal. Para associar essas comparações com a inteligibilidade, faz-se normalmente uma avaliação subjetiva. A melhoria na inteligibilidade não exige aqui uma recuperação do sinal original mas sim um sinal mais confortável do ponto de vista auditivo<sup>2</sup>.

A realização de testes subjetivos pode parecer simples. No entanto, se for considerada a diversidade de locutores, de ouvintes e ainda das frases, a aplicação de tais testes passa a ser algo difícil de ser organizado. Mesmo assim, desconsiderar os testes subjetivos é fugir da metodologia mais completa que existe na avaliação de sistemas de melhoramento de sinais de voz. Neste trabalho, como já colocado anteriormente, e considerando que as diferenças obtidas entre as técnicas são claras, não se usam métodos considerados científicos nas avaliações, como realizado por exemplo por SILVA (1992).

<sup>2</sup> Esta distinção se faz necessária, visto que normalmente associa-se uma técnica de redução de ruído a uma recuperação física do sinal (forma de onda original). O problema é que não se conhece ainda um sistema capaz de reconhecer melhor o sinal de voz do que o sistema auditivo humano.

No entanto, fazem-se testes que possam indicar claramente quais são os sinais mais confortáveis auditivamente e, conseqüentemente, as melhores técnicas.

Num avaliação subjetiva deve-se, inicialmente, destacar a composição dos sinais de testes. Neste sentido, o banco de dados utilizado é formado por:

**1** - Frases faladas em francês, que fazem parte do banco de dados do CNET dentro do estudo de sistemas *HFTs*.

**2** - Frases em português gravadas originalmente sem ruído e degradadas artificialmente por um ruído de carro a uma velocidade aproximada de 100 km/h.

A metodologia (*informal*) adotada para os testes foi a seguinte:

**TESTE1** - Em uma primeira etapa, deseja-se saber do ouvinte se o sinal processado é melhor, igual ou pior (em termos de conforto auditivo) do que o sinal ruidoso original. O ouvinte responde dizendo IGUAL, SIM (*melhor*) ou NÃO (*pior*).

**TESTE2** - Em um segundo teste escuta-se uma única frase resultante dos vários métodos de processamento. As frases processadas por cada técnica são ouvidas por cada pessoa duas vezes em uma sequência diferente. O objetivo é verificar se o ouvinte não privilegia um método pela sua posição em relação ao sinal de teste (ruidoso). O ouvinte deve apenas informar qual é a melhor frase.

#### **V.4 - Resultado dos testes objetivos**

A aplicação dos testes objetivos resultou em um conjunto de curvas que mostram as distorções introduzidas por uma determinada técnica e os ganhos na relação sinal/ruído temporal obtidos. Ainda como forma de avaliação obtiveram-se os espectrogramas e formas de ondas no tempo do sinal limpo, do sinal ruidoso e dos sinais processados por cada uma das técnicas avaliadas. O espectrograma fornece uma visualização da redução de ruído para pequenas faixas de frequência em um espectro que vai de 0 a 8 kHz. Os sinais



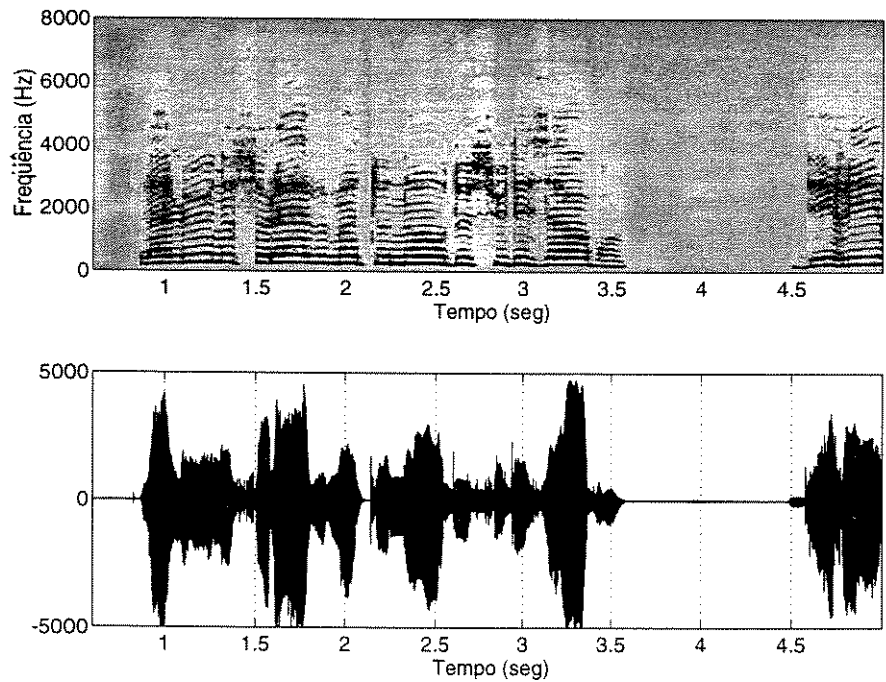
utilizados foram amostrados a um taxa de 16 k amostras/s após filtragem passa-baixa com frequência de corte de 7,2 kHz. O ruído de fundo utilizado é o de um carro pequeno (Renault 25) rodando a 100 km/h. Todas as curvas foram obtidas com o auxílio do software de simulação MATLAB. Todos os sinais foram processados com a mesma configuração para a STFT (Short Time Fourier Transform), isto é, janelas de Hanning com intervalos de sobreposição de 50%, 512 amostras por janela (31,25 ms de sinal) e FFT (Fast Fourier Transform) de 1024, sendo 512 de zeros formando a configuração “ZERO PADDING”. A constante de tempo usada no cálculo da potência média do ruído é de 140 ms. Esta configuração também é adotada nas avaliações subjetivas.

**V.4.1 - Sinal usado nos testes objetivos**

**A - Frase (em francês)**

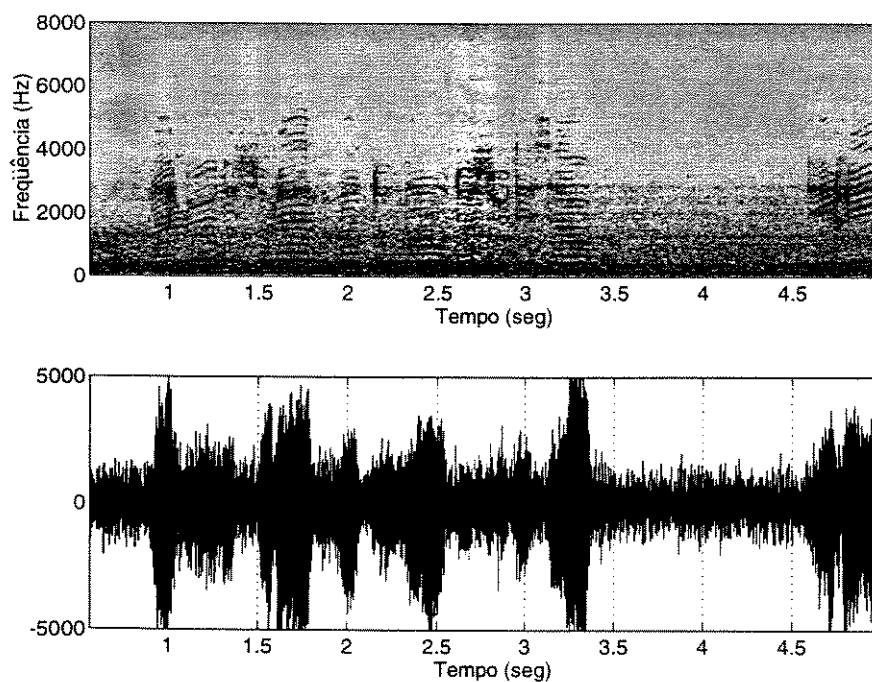
*“LA LUNE SE LÈVE MAINTENANT AU-DESSUS DES ARBRES”*

**B - Espectrograma e Forma de Onda do Sinal sem Ruído**



*Figura V.2 - Espectrograma e forma de onda do sinal puro*

### ***C - Espectrograma e Forma de Onda do Sinal com Ruído***

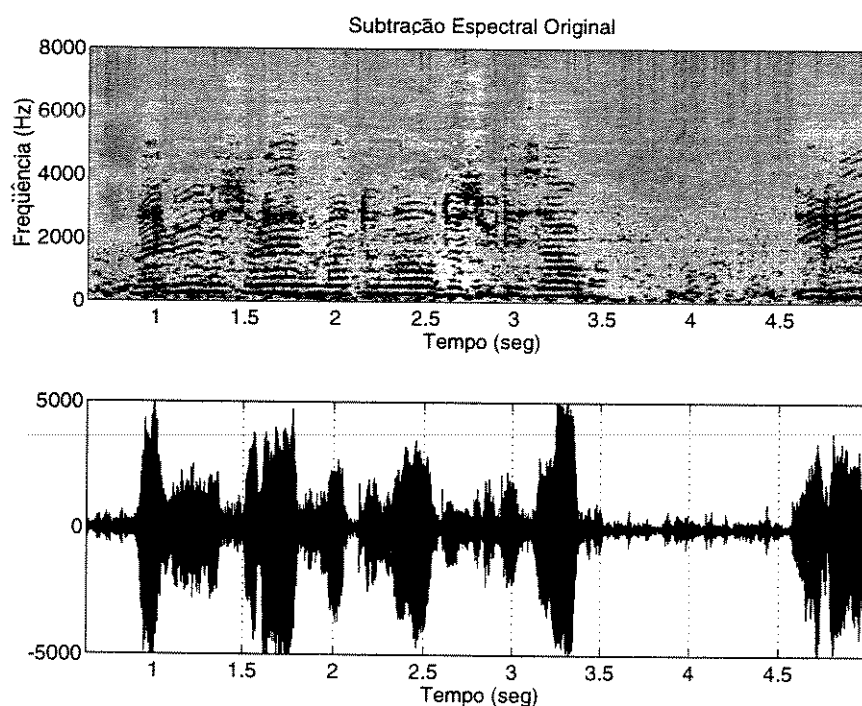


*Figura V.3 - Espectrograma e forma de onda do sinal ruidoso*

#### ***V.4.1 - Resultados obtidos com a SUBTRAÇÃO ESPECTRAL ORIGINAL (SEP)***

***Espectrograma e forma de onda (figura V.4):*** Nos períodos de voz (entre 1s e 3,5s, por exemplo) verifica-se que o sinal processado apresenta uma distribuição espectral próxima a do sinal original. No entanto, nos intervalos de silêncio (entre 3,5s e 5,5s, por exemplo), observam-se pequenos pontos numa faixa significativa de frequência. Esses pontos representam ruído musical (pequenos picos podem ser visualizados na forma de onda no mesmo intervalo) e também estão presentes nos intervalos de voz. O método apresenta uma boa redução de ruído a partir de 1000 Hz. Em termos de forma de onda, observa-se uma proximidade desta com a do sinal original, além de uma boa indicação de redução do ruído.

**Distorções (figura V.5) e Ganho na SNR (figura V.6):** As distorções resultantes são principalmente devido ao ruído musical, conforme mostra o diagrama da distância cepstral. No histograma, verifica-se que os valores da distância cepstral são concentrados entre 0,4 e 0,7., com grande contribuição dos intervalos de silêncio. Em termos de redução de ruído, o ganho apresentado pela subtração espectral é relativamente bom, dando uma média de 4,88 dB.



*Figura V.4 - Espectrograma e forma de onda no tempo*

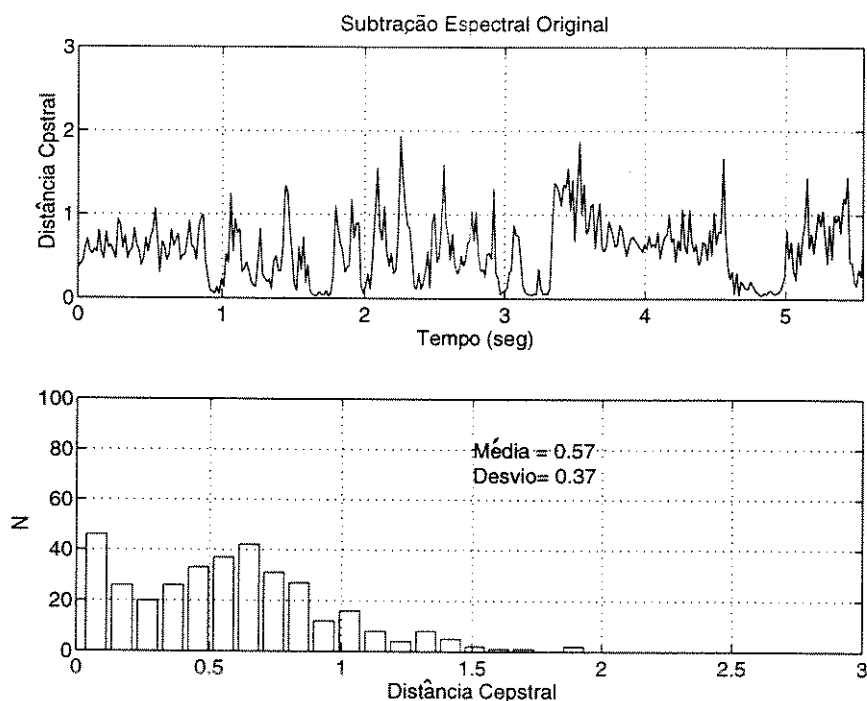


Figura V.5- Distorções

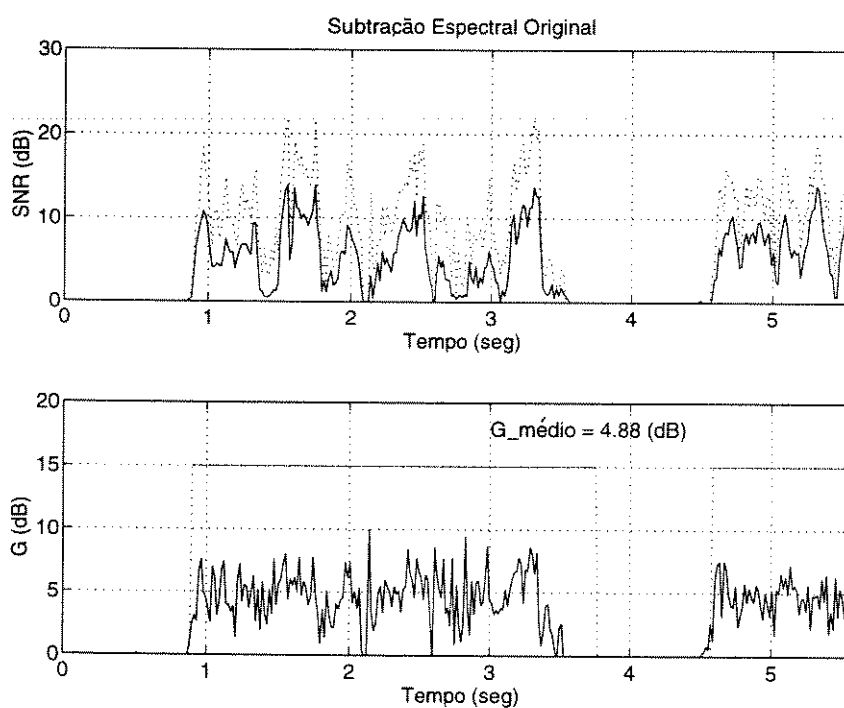


Figura V.6- SNR temporal e Ganho<sup>3</sup>

<sup>3</sup> Em todas as figuras onde apresentam-se a SNR temporal e o ganho (G) equivalente, tem-se que: para as curvas SNR, \_\_\_\_\_ é a SNR do sinal ruidoso e ..... é a SNR do sinal processado; para as curvas de ganho, \_\_\_\_\_ é G e ..... indica uma DVS..

### V.4.3 - Resultados obtidos com o método de EPHRAIM & MALAH (EP\_MA)

**Espectrograma e forma de onda (figura V.7):** Verifica-se neste método uma boa redução de ruído também a partir de 1000 Hz. O espectrograma é muito mais limpo do que o apresentado na figura V.4 e não se constata os pontos característicos do ruído musical. Porém, é fácil perceber algumas distorções, como por exemplo em torno do ponto 3,5s. Nesse ponto (que pode ser visto também na forma de onda) observa-se que um trecho de voz foi praticamente eliminado. Além disto, nos intervalos de 2,5s a 3,2s, observa-se uma forte atenuação na forma de onda temporal. Já nos intervalos de silêncio o ruído é reduzido de forma bem uniforme.

**Distorções (figura V.8) e Ganho na SNR (figura V.9):** As distorções resultantes são claras na curva da distância cepstral. Observa-se que durante os intervalos de silêncio existe menos distorção do que nos intervalos de voz. O pico maior de distorção é justamente no ponto 3,5s. No histograma tem-se que a média é de 0,74, com muitos valores entre 0,3 e 0,8 e com alguns outros chegando a mais de 2. O ganho apresentado é muito bom, chegando a uma média de 6,66 dB.

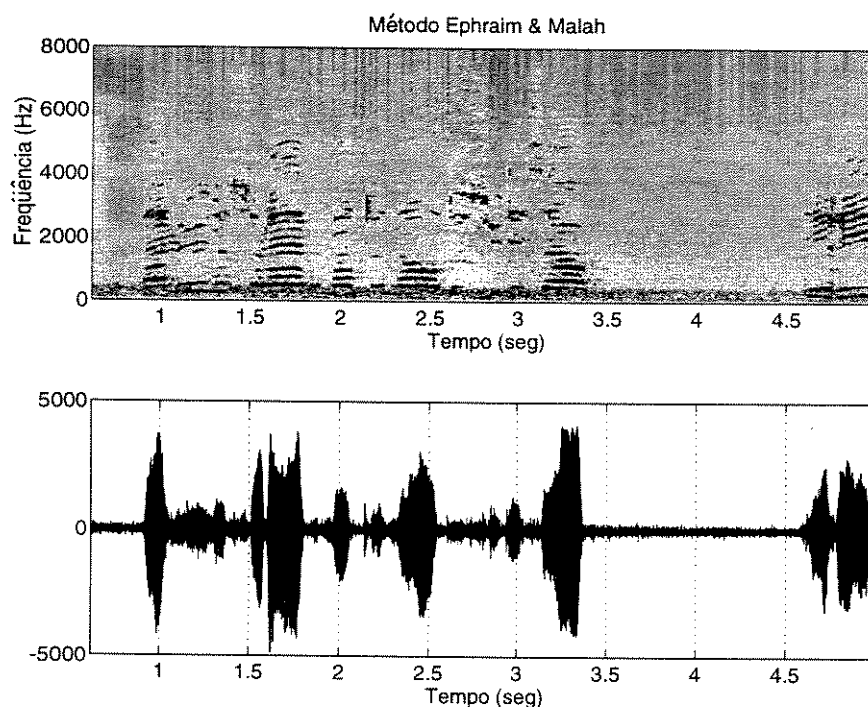


Figura V.7 - Espectrograma e forma de onda

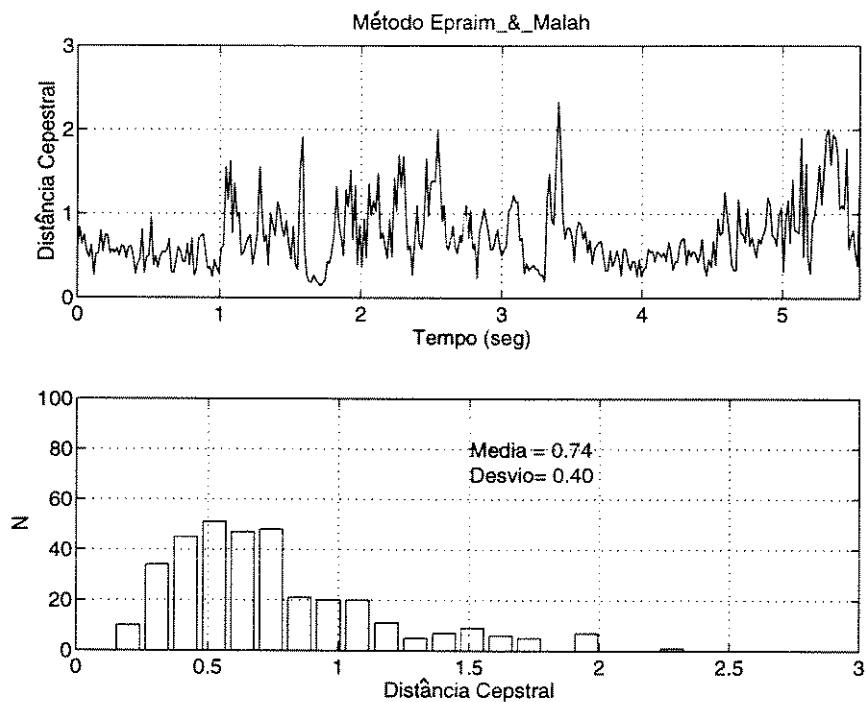


Figura V.8- Distorções

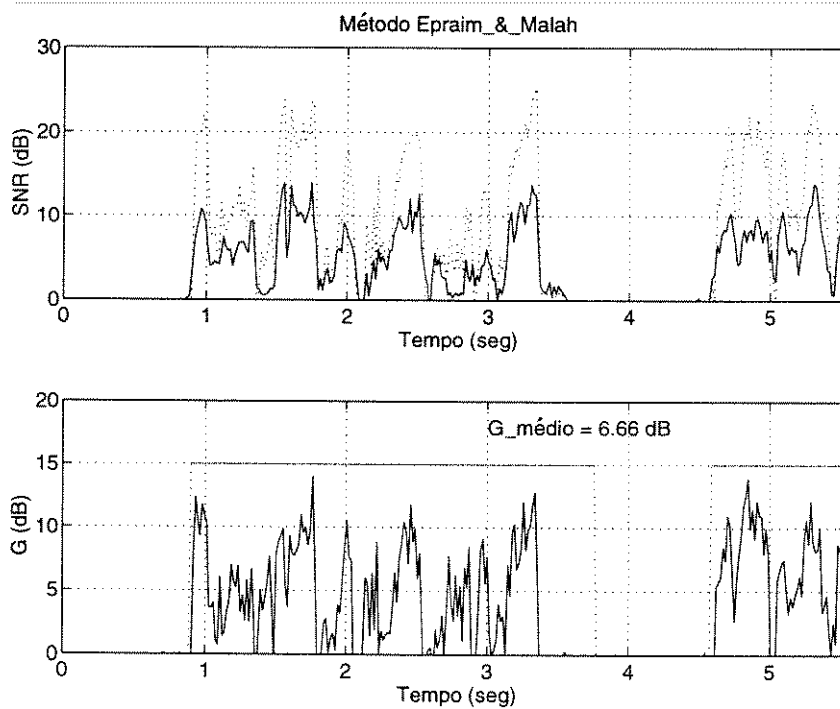
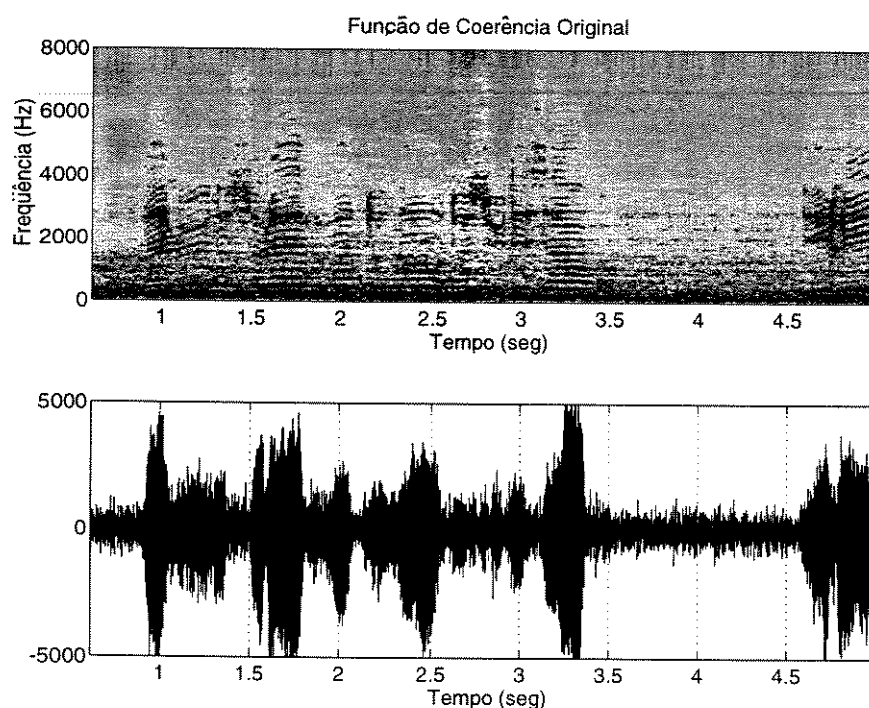


Figura V.9- SNR temporal e Ganho

### ***V.4.3 - Resultados obtidos com a FUNÇÃO DE COERÊNCIA ORIGINAL (FCOER)***

***Espectrograma e forma de onda (figura V.10):*** Observa-se neste método um sinal processado mais próximo do sinal ruidoso do que do sinal de voz puro. Pelo espectrograma verifica-se que a redução do ruído só começa a ser significativa a partir de 3000 Hz. O ruído resultante em baixa frequência, que também é o mais incômodo auditivamente, persiste. Não há indícios de ruído musical.

***Distorções (figura V11) e Ganho na SNR (figura V12):*** Uma das maiores vantagens da função de coerência é a baixa distorção introduzida no sinal processado. A média da distância cepstral apresentada é relativamente baixa (0,37), com valores variando na faixa de 0,1 e 0,5. No entanto, o ganho na SNR obtido é muito inferior aos apresentados pelos outros métodos, com uma média de 1,6 dB.



***Figura V.10 - Espectrograma e forma de onda***

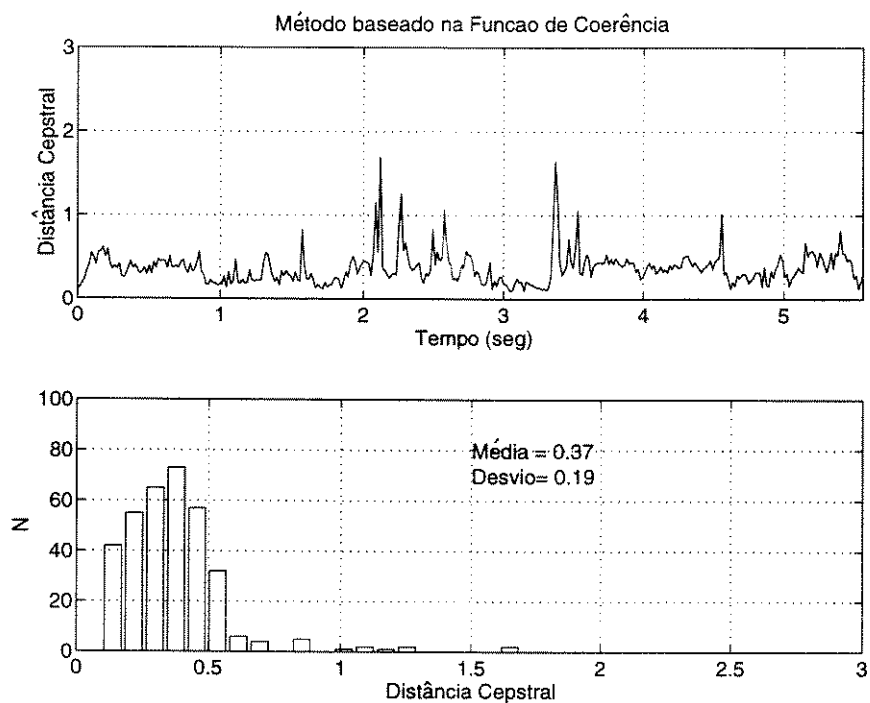


Figura V.11- Distorções

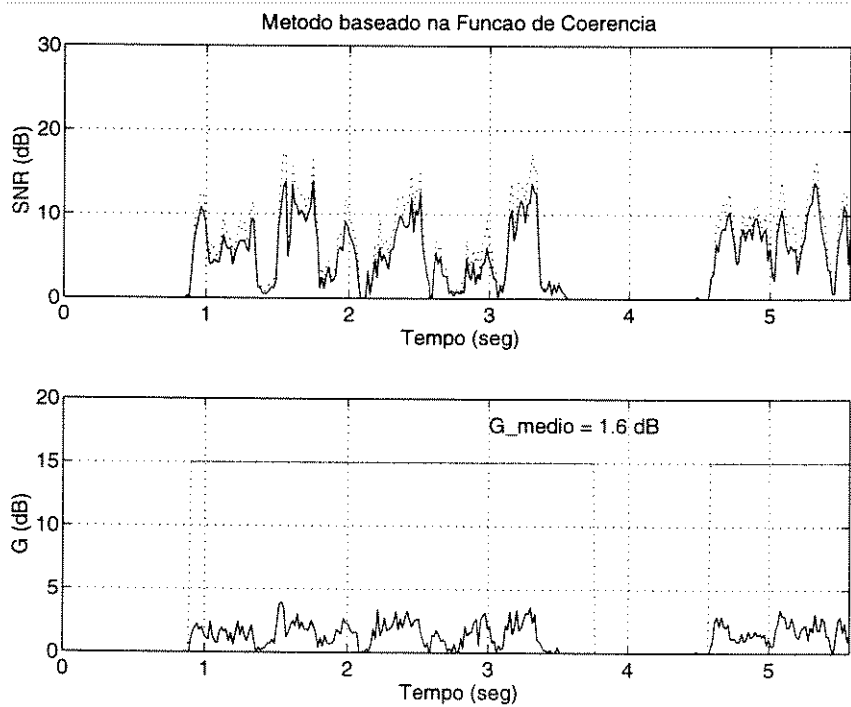


Figura V.12- SNR temporal e Ganho



#### ***V.4.4 - Resultados obtidos com a NOVA SUBTRAÇÃO ESPECTRAL (NSEP)***

##### ***V.4.4.1 - Sem considerar as incertezas (NSEP)***

***Espectrograma e forma de onda (figura V.13):*** Observa-se neste método um espectrograma mais próximo daquele obtido com o sinal original, mas também próximo do mostrado na figura V.7 (método de Ephraim & Malah). Na verdade, tem-se neste método a vantagem de ausência de ruído musical com uma boa redução de ruído. No entanto, têm-se ainda distorções, mas que são inferiores às obtidas na figura V.7. Esta melhora pode ser verificada até mesmo em se observando a forma de onda, onde se têm amplitudes mais próximas daquelas do sinal original.

***Distorções (figura VI4) e Ganho na SNR (figura VI5):*** O valores das distorções são claramente inferiores aos obtidos com o método de Ephraim & Malah. Também, a média global (0,47) é inferior à média obtida com a subtração espectral, mas com as distorções estando mais presentes nos intervalos de voz (entre 1s e 4s, por exemplo). O histograma mostra muitos valores abaixo de 0,1 e uma média razoável entre 0,2 e 0,8. A redução ruído é muito boa, com o ganho na SNR tendo uma média de 5,57 dB. Este valor está entre os valores obtidos com a subtração espectral original e os valores obtidos com o método de Ephraim & Malah.

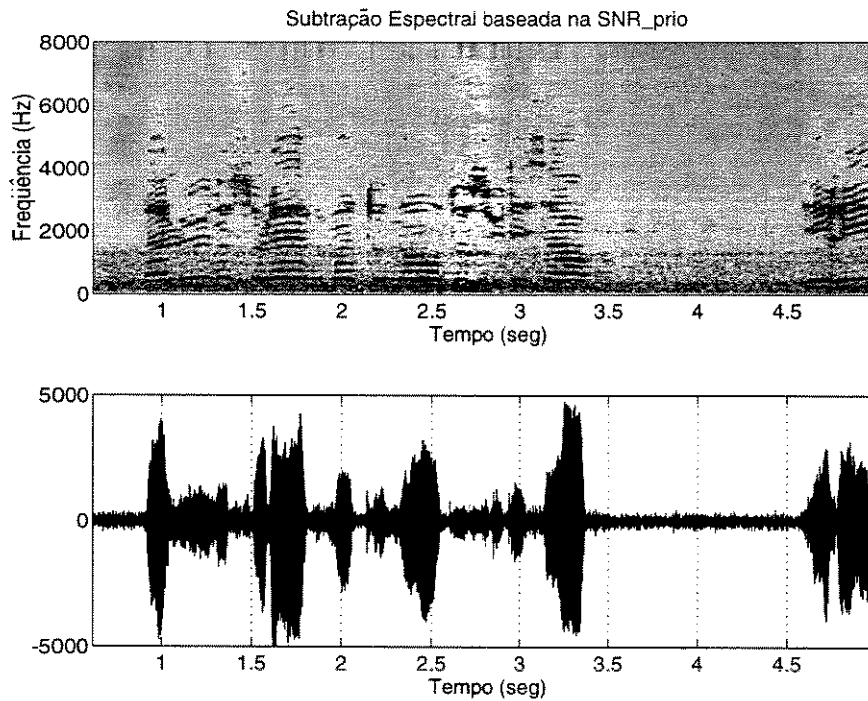


Figura V.13 - Espectrograma e forma de onda

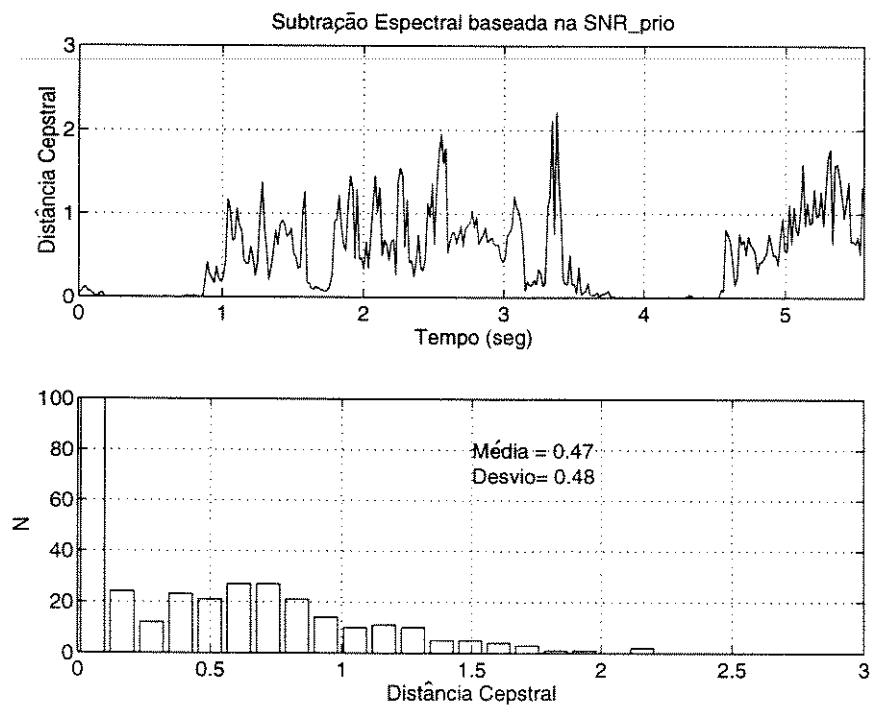
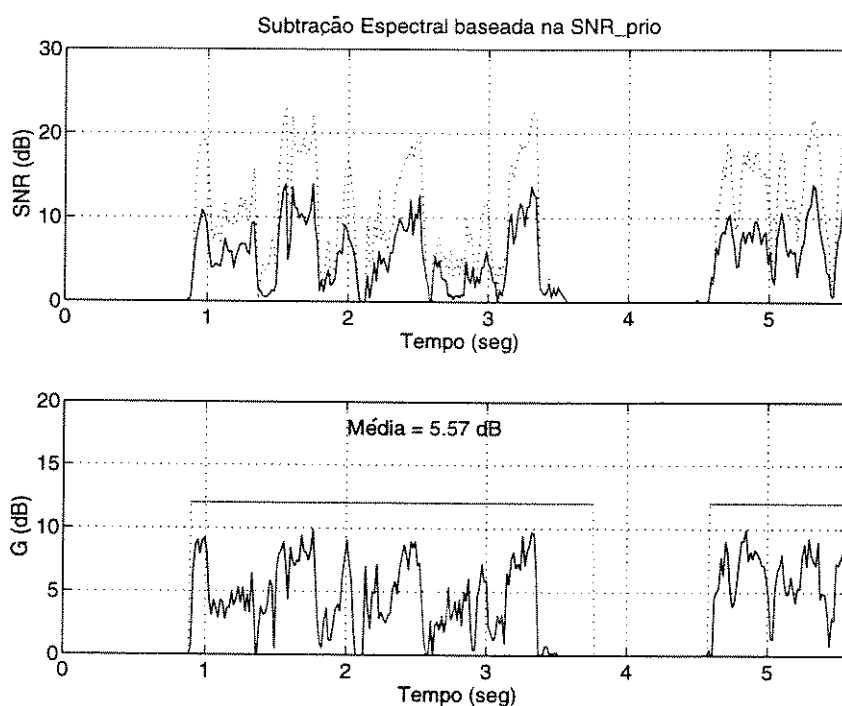


Figura V.14- Distorções



*Figura V.15- SNR temporal e Ganho*

#### ***V.4.4.2 - Considerando as incertezas- Método Ephraim & Malah (NSEPII)***

***Espectrograma e forma de onda (figura V.16):*** Neste método os resultados gerais variam entre o método da nova subtração espectral e o método de Ephraim & Malah. Tem-se boa redução de ruído, com ausência de ruído musical mas com distorções.

***Distorções (figura V.17) e Ganho na SNR (figura V.18):*** Obteve-se neste método uma média de distorção de 0,52, bem abaixo de 0,74 (método Ephraim & Malah) mas superior à nova subtração espectral sem o uso da incerteza (0,47). O valor médio do ganho obtido foi muito bom (6,50 dB), principalmente considerando-se a média de distorção obtida.

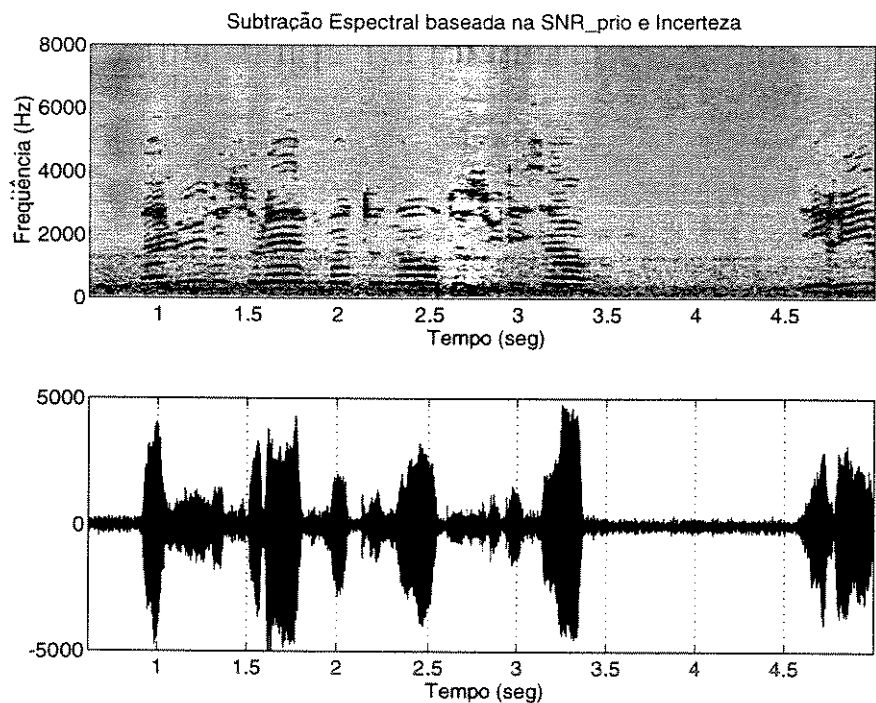


Figura V.16 - Espectrograma e forma de onda

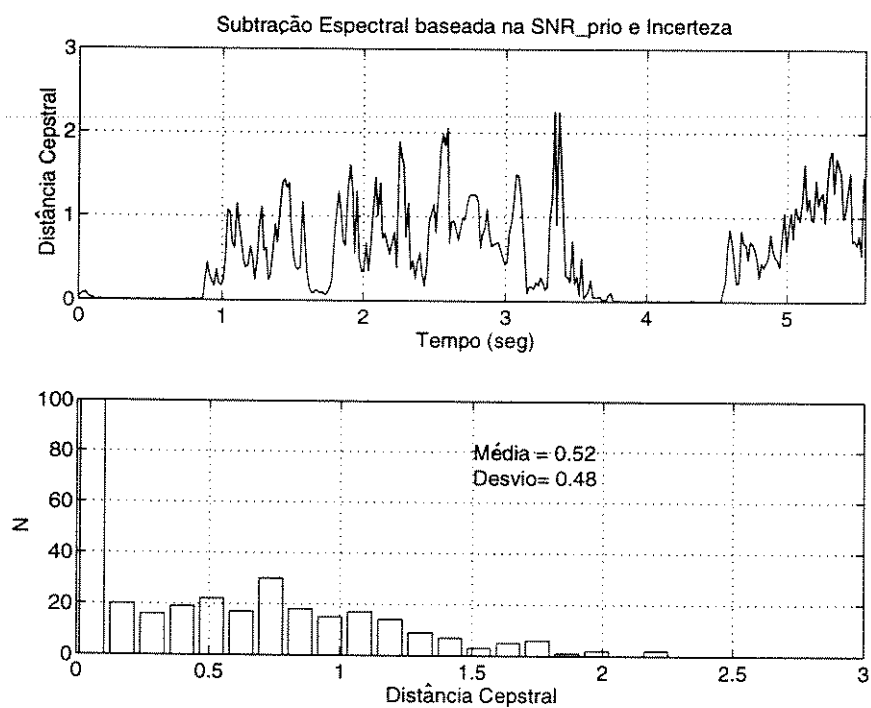


Figura V.17- Distorções

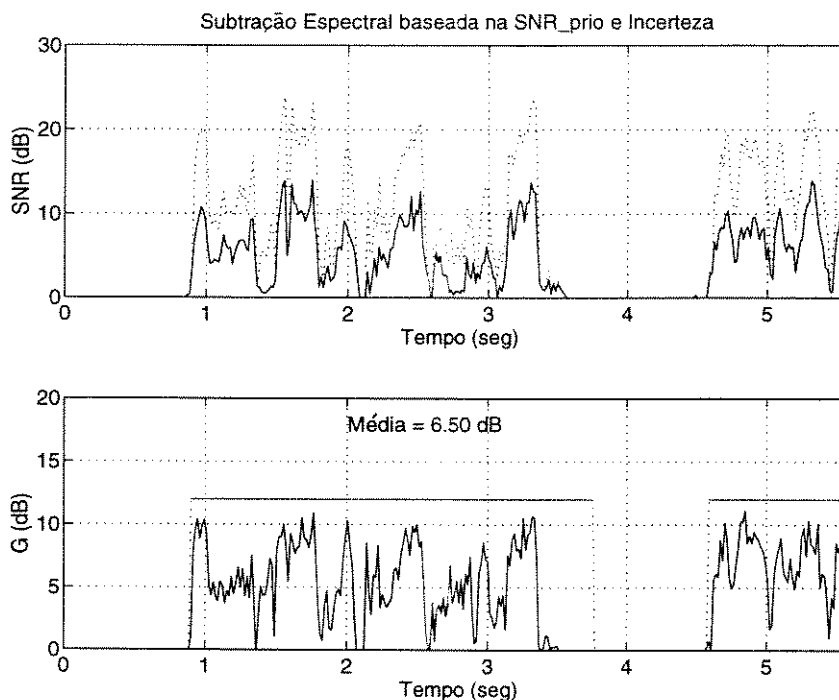
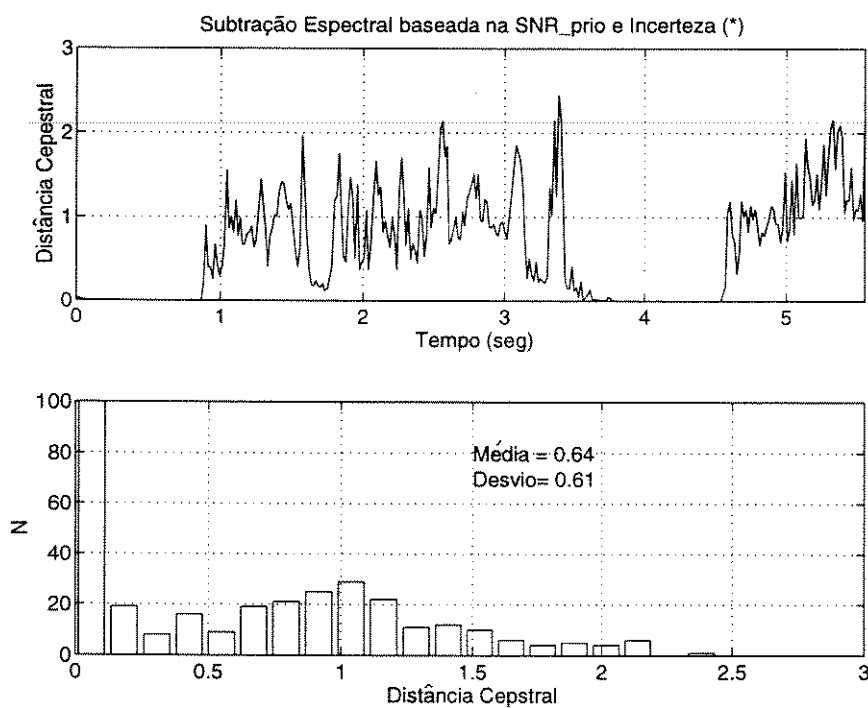
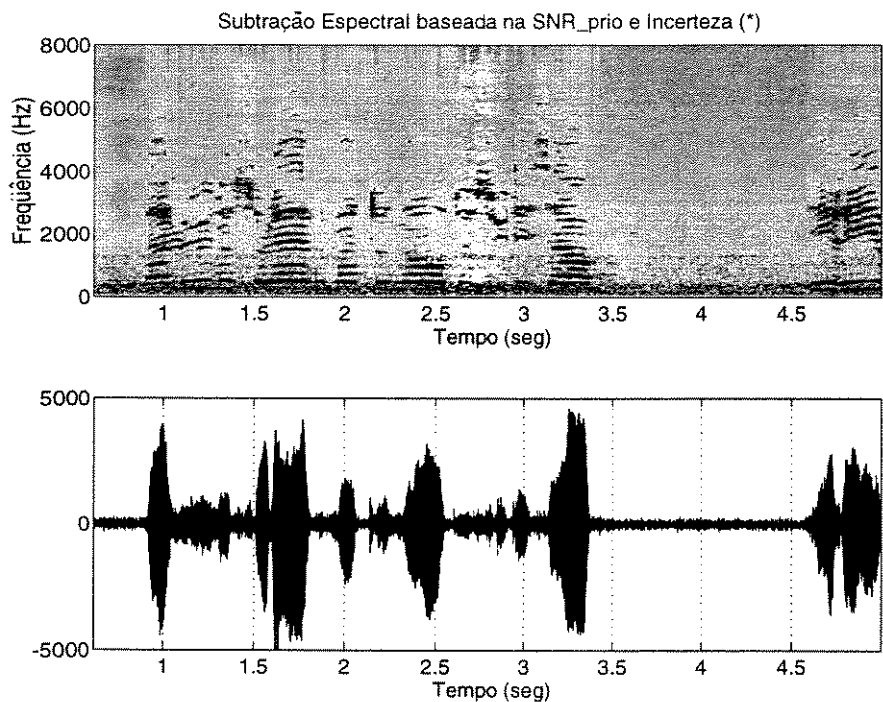


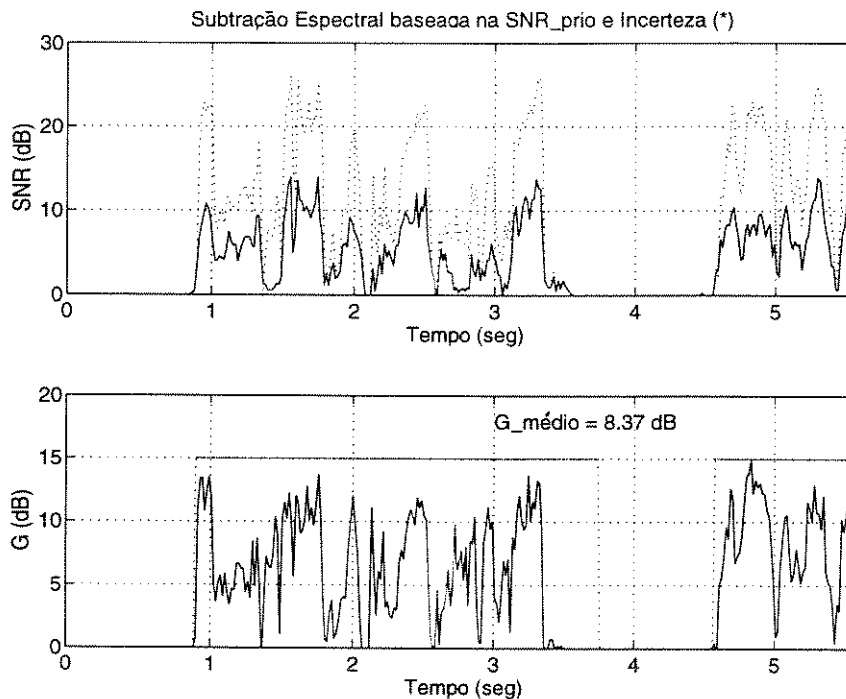
Figura V.18- SNR temporal e Ganho

#### V.4.4.3 - Considerando as incertezas- Método McAulay & Malpass (NSEPI2)

**Espectrograma e forma de onda (figura V.19):** Neste método, os resultados gerais variam entre o método da nova subtração espectral e o método de Ephraim & Malah, com boa redução de ruído, sem ruído musical e com distorções.

**Distorções (figura V.20) e Ganho na SNR (figura V.21):** As distorções obtidas são significativas, com média de 0,64 e muitos valores acima de 1, como indicado no histograma. Este método apresenta-se melhor do que o método de Ephraim & Malah mas é inferior à nova subtração espectral usando a incerteza proposta por Ephraim & Malah. O valor médio do ganho obtido foi o mais alto dentre todos os métodos (8,37 dB).





*Figura V.21- SNR temporal e Ganho*

#### ***V.4.5 - Resultados obtidos com a NOVA FUNÇÃO DE COERÊNCIA (NCOER)***

***Espectrograma e forma de onda (figura V.22):*** Observa-se que este espectrograma é parecido com o mostrado na figura V.10 (função de coerência original), mas apresentando uma melhor definição do sinal de voz (por exemplo, nos intervalos de 1s a 2s).

***Distorções (figura V.23) e Ganho na SNR (figura V.24):*** As distorções obtidas, com uma média de 0,56, são superiores às obtidas com a função de coerência normal. No entanto, tem-se uma redução de ruído muito mais significativa, com média de 3,45 dB.

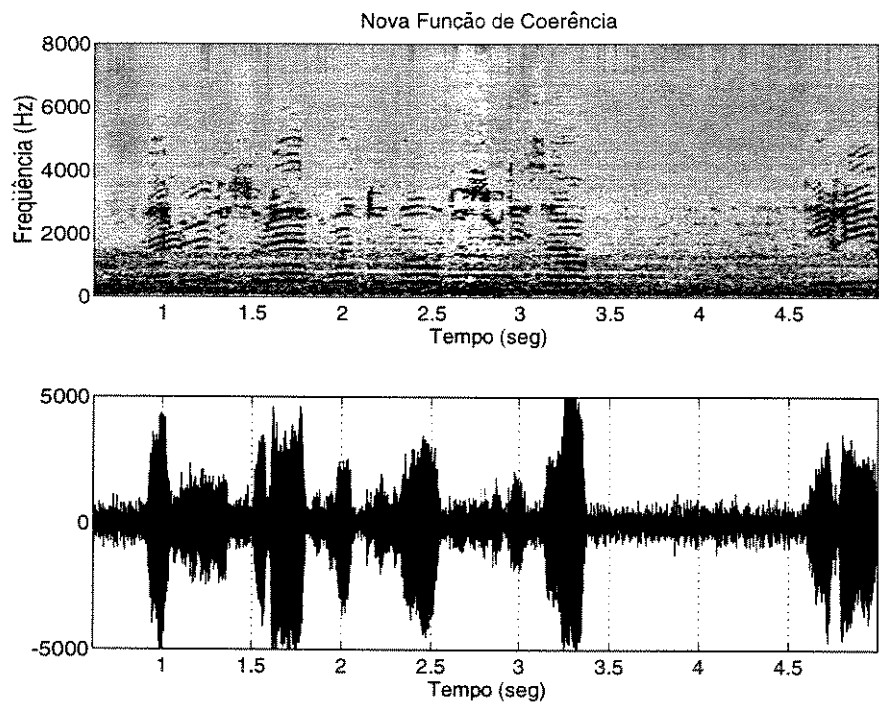


Figura V.22 - Espectrograma e forma de onda

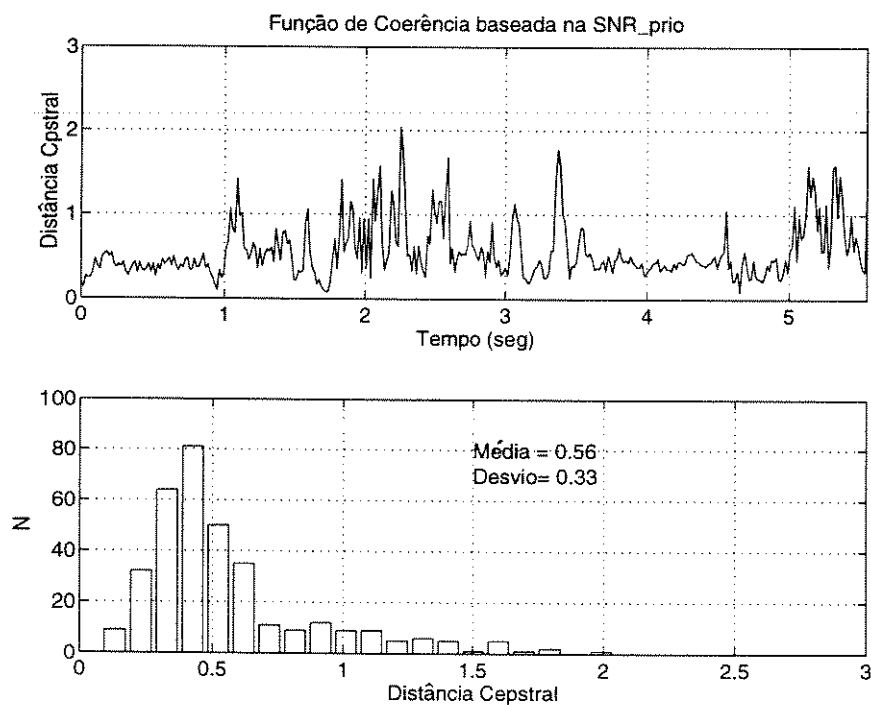


Figura V.23- Distorções



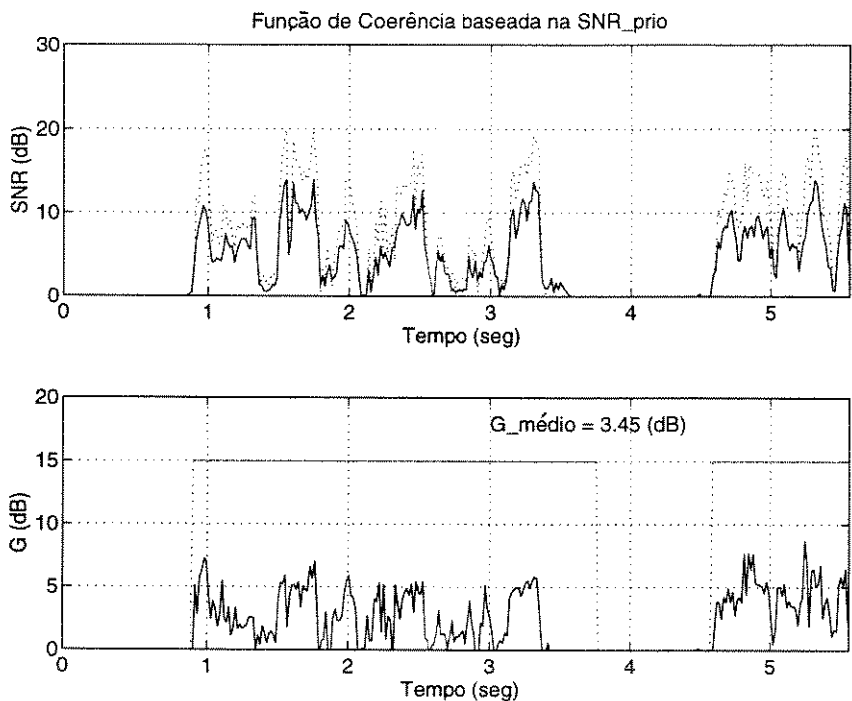


Figura V24 - SNR temporal e Ganho

## V.5 - Avaliação dos testes objetivos

A apresentação das curvas anteriores mostra com detalhes os resultados dos testes objetivos. No entanto, para uma avaliação mais rápida, mostram-se na tabela V.1 os valores médios de distorções e ganhos na SNR. No cálculo da média das distorções foram considerados os trechos de silêncio. Adotou-se este critério visto que distorções dentro dos intervalos de silêncio também são incômodas auditivamente.

	<i>SEP</i>	<i>NSEP</i>	<i>NSEPI1</i>	<i>NSEPI2</i>	<i>EP_MA</i>	<i>FCOER</i>	<i>NFCOER</i>
<i>Dist. Cepstral</i>	M=0,57	M=0,47	M=0,52	M=0,64	M=0,74	M=0,37	M=0,56
	D=0,37	D=0,48	D=0,48	D=0,61	D=0,40	D=0,19	D=0,33
<i>Ganho na SNR (dB)</i>	4,88	5,57	6,50	8,37	6,66	1,6	3,45

Tabela V.1 - Distorção e Ganho na SNR obtidos (M=média; D=desvio padrão)

Concentrando-se inicialmente nas técnicas baseadas em um único microfone, verifica-se que o maior índice médio de distorção foi obtido com a técnica de Ephraim & Malah (EP\_MA). Este fato, já discutido por SCALART (1996), mostra que o uso do modelo Gaussiano pode não ser o mais adequado, apesar do método apresentar um bom valor médio de ganho na SNR. O menor índice médio de distorção é obtido com o novo método de subtração espectral (NSEP), que, no entanto, apresenta um desvio padrão alto. Isto é explicado pelo bom desempenho do método na eliminação do ruído musical, mas introduzindo distorções no sinal de voz. Isto pode ser melhor analisado se forem observados os valores obtidos com a subtração espectral original (SEP). Observa-se que o desvio padrão está abaixo daquele obtido na NSEP, mesmo se a média da distorção é mais elevada. O que acontece é que nos instantes de silêncio, as distorções são mínimas com a NSEP e elevadas com a SEP, como pode ser visto nas figuras V.5 e V.8. No entanto, um balanço geral, incorporando aqui o ganho da SNR, mostra que a NSEP é superior a SEP.

Observando-se agora a NSEP com a introdução do conceito de incerteza nas realizações das medidas, verifica-se que o método baseado na incerteza proposta por Ephraim & Malah (NSEPI1) apresenta menor índice médio de distorção do que o proposto por McAulay & Malpass (NSEPI2). No entanto, o ganho na SNR é maior com a NSEPI2.

Para os métodos baseados na função de coerência, o fraco desempenho na redução do ruído (baixo ganho na SNR) para a função de coerência original (FCOER) é compensado pelo baixo índice médio de distorção. Para o novo método (NFCOER), observa-se um melhor ganho na SNR mas um índice de distorção superior. Na verdade, este é um problema generalizado na redução de ruído, isto é, para se ter uma melhor redução de ruído é preciso sacrificar a qualidade do sinal processado.

Em termos de redução de ruído por faixa de frequência, observando-se os espectrogramas é fácil comprovar que a melhor técnica é a NSEPI2. Pode-se visualizar também o problema do ruído musical na SEP e o baixo índice de redução de ruído nas baixas frequências, na FCOER.

O resultado final dos teste objetivos resumidos na tabela V.1 aponta a NSEP como a técnica mais adequada, dando um bom compromisso entre redução de ruído e distorção. Ela é também a técnica com menor complexidade computacional.

## **V.6 - Resultados e avaliação dos testes subjetivos**

A metodologia adotada nos testes subjetivos permitiu uma boa comparação entre as técnicas. As frases utilizadas nos testes foram:

Frase 1: *“La lune se lève maintenant au- dessus des arbres - Des rires montent de la cour de récréation - Ça et là, la prairie se piquait de fleurs”.*

Duração: 10,63 segundos

Locutor: Masculino

Frase 2: *“Tout s'est animé dès que le soleil s'est levé - Ma voiture est en panne devant ce pont - Ces femmes portent encore une coiffe”*

Duração: 10,63 segundos

Locutor: Feminino

Frase 3: *“A bolsa ficará estável ou sofrerá uma pequena queda”*

Duração: 5,38 segundos

Locutor: Masculino

Frase 4: *“O convênio permite o intercâmbio porque visa a integração entre alunos de culturas diferentes”*

Duração: 7,42 segundos

Locutor: Masculino

### **V.6.1 - Resultados subjetivos para o TESTE1**

Neste teste, onde o objetivo é verificar se o ouvinte prefere o sinal ruidoso ou o sinal processado, utilizaram-se as frases 1 e 3. Para a frase 1 ouviram-se apenas 5 pessoas, devido às dificuldades de se encontrar voluntários dominando a língua francesa. Todos consideraram que o sinal processado era melhor do que o sinal original, não importando a técnica utilizada. Para a frase 3 ouviram-se 12 pessoas e o resultado não foi unânime, com preferências variadas. Infelizmente, não foi possível avaliar as frases em português

usando-se dois microfones, pois a mesma foi obtida em um sistema com um único microfone. Os resultados finais são resumidos na tabela V.2.

AVALIAÇÃO	SEP	NSEP	NSEPII	NSEPI2	EP_MA	FCOER	NFCOER
<i>Frase 1</i>	100%	100%	100%	100%	100%	100%	100%
<i>Frase 3</i>	66,7%	91,6%	83,3%	83,3%	75%	----	----

*Tabela V.2<sup>4</sup> - Preferência pelo sinal processado em relação ao sinal ruidoso original*

O resumo apresentado na tabela V.2 mostra que a melhor técnica para processar o sinal ruidoso é a subtração espectral (SE) baseada na SNR\_prio (NSEP), seguida dos dois outros métodos de SE usando as incertezas nas avaliações (NSEPII e NSEPI2). O pior desempenho fica com a subtração espectral normal (SEP). Uma explicação possível para a diferença nos resultados obtidos com a frase 1, em francês, e a frase 2 em português, é que a frase em francês foi gravada dentro do veículo enquanto que a frase em português foi gravada em um ambiente acusticamente isolado e depois somada a um ruído de carro. Naturalmente, a frase gravada no veículo incorporou também o problema da reverberação, o que não aconteceu com o sinal de voz gravado em português. Um outro ponto importante é que a frase em francês é mais longa que a frase em português. Isto faz com que o ouvinte seja mais incomodado pelo ruído. Mesmo assim, de maneira geral, pode-se dizer que o sinal processado por qualquer método é mais agradável do que o sinal original.

**V.6.2 - Resultados subjetivos para o TESTE2<sup>5</sup>**

Neste teste todas as frases foram utilizadas. O objetivo é saber qual a técnica que melhor satisfaz o ouvinte. Utilizaram-se 05 voluntários em francês e 12 em português, sendo obtidas 10 e 24 respostas para cada frase, respectivamente (usaram-se duas

<sup>4</sup> SEP - Subtração Espectral Original; NSEP - Subtração Espectral baseada na SNR\_prio; NSEPII - NSEP com o uso da incerteza proposto por Ephraim & Malah; NSEPI2 - NSEP com o uso da incerteza proposto por McAulay & Malpass; EP\_MA - Técnica proposta por Ephraim & Malah; COER - Método baseado na função de coerência original; NCOER - Coerência baseada na SNR\_prio.

<sup>5</sup> Neste teste, para as frases em português, além de não haver teste com dois microfones, também não foi usado o método NSEPI2 (muito inferior nos testes objetivos à NSEP e NSEPII).

seqüências diferentes com as mesmas frases). Os resultados foram variados e mais uma vez a técnica baseada em dois microfones só foi avaliada com as frases em francês. O resultado final obtido é apresentado na tabela V.3, sendo o número de respostas colocado para cada frase. Vale salientar que o teste foi realizado apenas para saber qual é o melhor método, não importando em que nível se encontram os outros. O ouvinte pôde escolher mais de uma frase ou optar pela frase original (ruidosa).

AVALIAÇÃO	SEP	NSEP	NSEPII	NSEPI2	EP_MA	COER	NCOER
<i>Frase 1</i>	0	3	1	0	0	3	4
<i>Frase 2</i>	0	4	2	1	1	1	2
<i>Frase 3</i>	1	12	8	----	1	----	----
<i>Frase 4</i>	1	14	11	----	1	----	----

*Tabela V.3 - Resultado dos testes subjetivos*

Os resultados apresentados na tabela V.3 mostram uma grande vantagem para a técnica NSEP (subtração espectral baseada na SNR\_prio). Após cada avaliação interrogaram-se os ouvintes. A explicação para as respostas parecem interessantes e vale a pena citá-las aqui. Por exemplo, nos testes realizados com as frases em francês, a SEP (original) foi considerada muito fraca em relação às outras técnicas. Vê-se também que as técnicas baseadas em dois microfones (COER e NCOER) apresentam um bom índice de aceitação, com uma pequena vantagem para a coerência usando a SNR\_prio. A explicação dos ouvintes é que o ruído residual nas frases processadas usando dois microfones não era incômodo. Na verdade, os métodos baseados na função de coerência não apresentam um bom índice de redução nas baixas frequências mas conseguem ter um bom desempenho nas frequências altas. Assim, é de se esperar que quanto maior a relação sinal/ruído (SNR) melhor será seu desempenho (boa redução nas altas e pouca distorção). Tanto é verdade que na frase 2, que é mais ruidosa do que a frase 1, os índices são menores. No geral, todos os métodos conseguiram ser mencionados, menos a SEP. Para as frases em português, onde foi possível apenas uma avaliação dos métodos baseados em um único microfone, os resultados foram muito bons para a NSEP. Também, a SE original foi citada. A grande vantagem aqui foi o maior número de ouvintes. A desvantagem foi o fato de a voz ter sido

gravada em um ambiente acústico isolado, não reproduzindo os efeitos reverberantes de um veículo. Isto não invalida os testes. Pelo contrário, pode ser um bom indicativo do desempenho na redução de ruído em outros contextos.

## **V.7 - CONCLUSÕES**

O objetivo deste capítulo foi mostrar uma comparação entre os métodos estudados e propostos neste trabalho. Para tal usaram-se métodos de avaliações objetivos e subjetivos. Na realização das avaliações objetivas mostrou-se que as técnicas propostas são superiores às técnicas originais e também às outras técnicas. No contexto de redução de ruído usando-se apenas um microfone, a técnica mais eficiente, apresentando bom compromisso entre redução de ruído e distorção, é a subtração espectral baseada na relação sinal/ruído e com a incorporação do parâmetro de incerteza (NSEPI1). Quanto aos métodos que utilizam dois microfones baseados na função de coerência, mostrou-se que eles são inferiores na redução do ruído mas introduzem bem menos distorção do que os métodos baseados em um único microfone. Entre os dois métodos baseados na função de coerência, as medidas objetivas mostraram que o novo método, usando uma função de coerência baseada na SNR a priori, permite uma melhor redução do ruído. No entanto, apresenta mais distorções.

Em termos de testes subjetivos, os resultados estão próximos aos obtidos nos testes objetivos. Os métodos baseados em dois microfones são mencionados com razoável frequência, sendo que o método baseado na função de coerência modificada tem ligeira vantagem. Se comparados todos em conjunto, a nova subtração espectral é superior. A introdução do parâmetro de incerteza (NSEPI1 e NSEPI2) não produziu o mesmo efeito obtido nos testes objetivos. No entanto, a NSEP e NSEPI1 estão próximas, indicando que a introdução do parâmetro de incerteza sugerido por Ephraim & Malah é melhor do que o sugerido por McAulay & Malpass.

Uma conversa informal com os participantes dos testes subjetivos revelou que a duração das frases usadas nos testes é muito importante. Por exemplo, a frase 3, em

português apresenta apenas 5,63 segundos de duração. Neste curto intervalo de tempo, segundo alguns ouvintes, o sinal ruidoso é suportável. Ao se repetir o mesmo sinal mais vezes (teste informal) o ouvinte acaba sendo incomodado e até o ruído musical, encontrado fortemente na subtração espectral original, passa a ser mais aceitável do que o ruído original.

## **CAPÍTULO VI**

### **COMENTÁRIOS FINAIS**

Neste trabalho estudaram-se e desenvolveram-se novas técnicas de redução de ruído com aplicações voltadas para os sistemas rádio-móveis veiculares, obtendo-se resultados altamente satisfatórios. Foram mostradas as características do ambiente móvel e do ruído nele obtido, dando-se uma idéia geral das dificuldades para se reduzir ruído nesses sistemas. O problema principal é que o ruído obtido no ambiente móvel tem forte concentração de energia nas frequências baixas, como os sinais de voz.

Os estudos foram iniciados com a apresentação de algumas das técnicas clássicas de redução de ruído, como subtração espectral, filtragem de Wiener, método Ephraim & Malah, função de coerência, etc., envolvendo a captação de sinais com um e dois microfones. Nesses estudos constataram-se dois pontos importantes para as técnicas baseadas em um único microfone: primeiro, podem-se ter sinais processados eficientemente mas com técnicas complexas em termos de implementação computacional, como a técnica de Ephraim & Malah; segundo, podem-se ter técnicas simples computacionalmente, mas que não apresentam resultados satisfatórios. É o caso, por exemplo, da subtração espectral clássica, que deixa no sinal processado um ruído residual muito incômodo denominado de *ruído musical*. Com relação aos métodos baseados na função de coerência, que usa dois microfones, o problema maior é que não se consegue um bom nível de redução de ruído nas frequências baixas, além de uma pequena reverberação introduzida no sinal processado.



As novas técnicas obtidas foram resultado de um estudo de dois parâmetros importantes nos métodos de redução de ruído avaliados: a *relação sinal/ruído a posteriori* ( $SNR_{post}$ ) e a *relação sinal/ruído a priori* ( $SNR_{prio}$ ). Mostrou-se que o grande problema das técnicas que apresentam o inconveniente do “ruído musical” é que elas são definidas diretamente como função da  $SNR_{post}$ . Nos estudos realizados verificou-se que este parâmetro não apresenta uma boa performance quando a relação sinal/ruído é baixa (tipicamente menor do que 10 dB). No entanto, verificou-se que as técnicas baseadas na  $SNR_{prio}$  apresentavam bons resultados. Definindo a  $SNR_{post}$  em função da  $SNR_{prio}$ , obtiveram-se soluções para as técnicas clássicas de redução de ruído, que usam um ou dois microfones para captar sinal.

As novas técnicas obtidas, considerando um único microfone para captação do sinal, apresentam vantagens significativas em relação aos métodos clássicos. Uma primeira é a simplicidade computacional, que pode viabilizar aplicações em tempo real. Uma segunda é a eficiência na redução do ruído. Finalmente, as técnicas propostas não apresentam o inconveniente do ruído musical. Para o caso de dois microfones, obteve-se um novo método de cálculo para a função de coerência. A vantagem em relação ao método original é a melhor redução do ruído nas frequências baixas e a ausência de reverberação. Nos dois casos, o parâmetro principal explorado foi a relação  $SNR_{prio}$ .

Para a escolha de métodos baseados em um ou dois microfones têm-se duas considerações importantes: uma é que nos métodos baseados em um único microfone tem-se distorções significativas introduzidas no sinal processado; a outra é que no caso de se terem dois microfones explorando a coerência entre os sinais, não se consegue obter uma boa redução de ruído nas frequências baixas. Com os resultados apresentados neste trabalho, pode-se afirmar que as técnicas baseadas em um único microfone apresentam desempenho satisfatório, e que, de forma global, são superiores aos obtidos com técnicas baseadas na função de coerência.

Ainda dentro do contexto de redução de ruído nos sistemas rádio-móveis veiculares, apresentaram-se propostas para um sistema de detecção de intervalos de voz-silêncio. O método é baseado em dois microfones, sendo que um capta voz e ruído e o outro apenas

ruído. Esta parte é importante pois verificou-se que a maioria dos métodos de redução de ruído depende de uma estimação da potência do ruído de fundo, considerado estacionário.

Alguns pontos podem ser estudados para um melhoramento das técnicas de redução de ruído propostas neste trabalho, como por exemplo:

- Modificar o estimador de relação sinal/ruído a priori;
- Melhorar a captação do sinal no ambiente veicular;
- Incorporar características da fala e do sistema auditivo.

Os pontos destacados anteriormente não são atrelados um ao outro e podem contribuir isoladamente para um melhor desempenho na redução de ruído em ambientes veiculares. Acredita-se, por exemplo, que uma melhora na estimação da relação sinal/ruído a priori possa aumentar significativamente o desempenho das técnicas propostos. Também, um melhor sistema de captação do sinal no ambiente veicular resultará num aumento da relação sinal/ruído do sinal. Pode-se, neste caso, se pensar no uso, por exemplo, de matrizes de microfones. Em termos de características da fala, mesmo sabendo-se que as técnicas com base, por exemplo, na detecção de ‘pitch’, ainda apresentam baixa performance, não há como não considerarem-se aspectos intrínsecos da fala para se obterem técnicas que introduzam menos distorção no sinal processado. Nesse sentido, pode-se também explorar os métodos baseados em um modelo do sistema auditivo humano.

Um último ponto diz respeito ao problema dos codificadores, usados nos sistemas rádio-móveis, quando se tem um sinal de voz com relação sinal/ruído baixa. Os métodos apresentados não eliminam o ruído sem introduzir alguma distorção no sinal de voz. Este problema será passado para o codificador, que deverá apresentar um desempenho menos satisfatório, para sinais com baixa relação sinal/ruído. O fato é que a maioria das técnicas atacam o problema da inteligibilidade, enquanto que, para se aumentar as performances dos codificadores, é preciso uma melhora na qualidade física do sinal de voz. Pode-se assim, por exemplo, se pensar no melhoramento do sinal de voz do ponto de vista auditivo após a decodificação do sinal.

## **REFERÊNCIAS BIBLIOGRÁFICAS**

- AGUIAR, B.G. "Melhoramento de Voz Degradada por Método Baseado em Subtração Espectral Adaptativa", *Anais do 7º Simpósio Brasileiro de Telecomunicações*, Florianópolis-SC, p.54-59, setembro-1989.
- ALLEN, J.B et alii. "Multimicrophone Signal Processing Technique to remove reverberation from Speech Signals" *J. Acoust. Soc. Amer.*, vol 62, No. 4, 1977.
- BARBIER, R.L et alii., "Trainable Noise Subtraction Filter for Speech Enhancement in Car", *V Europ. Sig. Proc. Conference (EUSIPCO)*, p. 1111-1114, Barcelona-1990.
- BEROUTI, M et alii, "Enhancement of Speech Corrupted by Acoustic Noise", *Proc. IEEE Intern. Conf. on Acoust, Speech, and Signal Processing*, April-1979.
- BOLL, S.F. "Suppression of Acoustic Noise in Speech using Spectral Subtraction", *IEEE Trans. Acoust., Speech, and Signal Processing (ASSP)*, vol. 29, p113-120, April-1979.
- BRANCACCIO, P., "Experiments on noise reduction techniques with robust voice detector in car environment" *Proc. Eurospeech* , p. 1259-1262, 1993.
- CAPPÉ, O. "Elimination of the Musical Noise Phenomenon with the Ephraim and Malah Noise Suppressor", *IEEE Trans. on Acoust., Speech and Signal Processing (TASSP)*, April 1994.
- CHENG, Y.M & O'SHAUGHNESSY, "Speech Enhancement Based Conceptually on Auditory Evidence, *IEEE Trans. on Acoust., Speech and Signal Processing*, vol.39, No.9, sept-1991.
- DELLER, J.L., PROAKIS, J.G & HANSEN, J.H.L, *Discrete-Time Processing of Speech Signals*, Ed. Macmillam, New York, 1993.
- EGAN, J.P. & WIENER, F.M. "On the intelligibility of Bands of Speech in Noise", *Journal of the Acoustical Society of America (JASA)*, Vol.18, No.2, October-1946.
- EPHRAIM, Y, "Statistical-Model-Based Speech Enhancement Systems", in *Proc. of IEEE*, Vol. 80, No. 10, p. 1526-1555, october,1992.

- EPHRAIM, Y & D. MALAH, D., "Speech Enhancement Using Minimum Mean Square Error Short-Time Spectral Amplitude Estimador, *IEEE Trans. on Acoust., Speech and Signal Processing*, vol.32, N° 6 - December-1984
- FLANAGAN, J.L. *Speech Analysis Synthesis and Perception*, Ed. Springer-Verlag, 1972.
- FRAZIER, R.H. et alii, "Enhancemet of Speech by Adaptive Filtering", *Proceedings IEEE Intern. Conf. on Acoust., Speech, and Signal Processing*, April-1976.
- FREEMAN, D.K. et alii, "The voice activity detector of the Pan-European digital cellular mobile telephone services", *IEEE Intern. Conf. on Acoust., Speech, and Signal Processing*, p.369-373, Glasgow, 1989.
- GUELOU, Y, BENAMAR, Abdelkrim and SCALART, Pascal, "Analysis of two Structures for Combined Acoustic Echo Cancellation and Noise Reduction", *IEEE Intern. Conf. on Acoust., Speech, and Signal Processing*, Atlanta, 1996.
- GRAY JR, A.H & MARKEL, J.D, "Distance Measures for Speech Processing", *IEEE Trans. on Acoust. Speech and Signal Processing*, vol.24, No.5, october-1976.
- HÄNDEL, P., "Low Distortion Spectral Subtration for Speech Enhancement", in *Proc.EUROSPREECH, 4<sup>th</sup> European Conference on Speech Communication and Technology*, Madrid, p. 1549-1552, september,1995.
- HAYKIN, S, "*Adaptive Filter Theory*", Ed. Prenctive-Hall, 2<sup>a</sup> edição, 1991.
- HAKKINEN, J. & VANANEN M., "Background noise suppressor for a car hands-free microphone" *4th ICSPAT*, pp. 300-307, Santa-Clara, USA, 1993.
- KIM, J.W and UN, C.K., "Enhancement of Noisy Forward/Backward Adaptive Digital Filtering", ", *IEEE Intern. Conf. on Acoust., Speech, and Signal Processing*, p.89-92, Tokio, 1986.
- KOBATAKE, H. et alii, "Linear Predictive Coding of Speech Signals in a Hih Ambient Noise Enivronmet", *IEEE Intern. Conf. on Acoust., Speech, and Signal Processing* June-1978.
- LE BOUQUIN, R. "On Using Coherence Function for Noise Reduction", *Signal Processing V: Theories and Applications*, Ed. Elsevier Science Publishers, 1990.
- LE BOUQUIN, R & FAUCON, G., "Study of a voice activity detector and its influence on a noise reduction system", *Speech Communication*, Vol. 16, No.1, p.245-254, 1995.
- LESCO, Rapport, "Etude des facteurs dégradant la qualité de la Parole en Radiotéléphonie Mains-Libres dans les véhicules", *Tome I, Rapport Technique INRETS*, Nov.,1992

- LIM, J.S & OPPENHEIM, A.V. "All-Pole Modeling of Degraded Speech", *IEEE Trans. on Acoust., Speech, and Signal Processing*, Vol.ASSP-26, No.3, June-1978.
- LIM, J.S & OPPENHEIM, A.V. "Enhancement and Bandwidth Compression of Noisy Speech", *Proceedings of the IEEE*, Vol.67 No.12, December-1979.
- LIM, J.S, et alii "Evaluation of an Adaptive Comb Filtering Method for Enhancing Speech Degraded by White Noise Addition, *IEEE Trans. on Acoust., Speech and Signal Processing*, vol.26, No. 4, october-1979.
- LIM, J.S,"Evaluation of a Correlation Subctratation Method for Enhancing Speech Degraded by Addictive White Noise", *IEEE, Trans. on Acoust., Speech and Signal Processing*, vol.26, october-1979.
- LYNCH, J.F et alii, "Speech/Silence segmentation for real-time coding via rule based adaptive endpoint detection", *IEEE Int. Conf. on Acoust., Speech and Signal Processing*, p.1348-1351, Dallas, 1987.
- MARRO, C. "Analyse de traitements multicauteurs pour le débruitage et la déréverberation du signal de parole, *Rapport Interne CNET*, 1995.
- McAULAY, J.R & MALPASS, M, "Speech Enhancement Using a Soft-Decision Noise Suppression Filter", *IEEE Trans. on Acoust., Speech and Signal Processing*, vol.28, N° 2 April 1980.
- PORTER, J.E and BOLL, F, "Optimal Estimators for Spectral Restoration of Noisy Speech", ", *IEEE Int. Conf. on Acoust., Speech and Signal Processing*, p. 18A.2.1-18A.2.4, 1984.
- RABINER, L.R & SHAFER, R.W, "Digital Processing of Speech Signals", Ed. Prentice-Hall, 1998.
- ROGERS et alii., "Neural Network Enhacement for a Two Speaker Separation System", *IEEE Int. Conf. on Acoust., Speech and Signal Processing*, p. 357-360, 1989.
- SCALART, P, "Debruitage Pour le Rádio-Téléphone Mains-Libres", *Rapport Interne CNET*, octobre-1994.
- SCALART, P. & VIEIRA-FILHO, J., "Speech Enhancement based on a Priori Signal-to-Noise Ratio Estimation", *IEEE Int. Conf. on Acoust., Speech and Signal Processing*, Atlanta, 1996.
- SCALART, P; VIEIRA-FILHO, J. & CHIQUITO, J.G, "On Speech Enhancement Algorithms based on MMSE Estimation", to be presented in the *VIII Europ. Sig. Proc. Conference (EUSIPCO)*, Trieste- Italy, september,1996.
- SCHROEDER, M.R, ATAL, B.S and HALL, J.,L, "Optimizing Digital Speech Coders by Exploiting Masking properties of the Human Ear", *Journal of Acoustical Soc. of America*, 1979, pp. 1647-1652.

- SILVA, A.J.S, "Quantização Vetorial: Aplicações a um VOCODER LPC", tese de mestrado, UFPb, Campina Grande, 1992.
- SLYH, R.E and MOSES, R.L, "Microphone Array Speech Enhancement in Overdetermined Signal Scenarios", *IEEE Int. Conf. on Acoust., Speech and Signal Processing*, p. II.347 - II.350, Minneapolis, 1993.
- TSOUKALAS, D et alii, "Speech Enhancement Using Psychoacoustic Criteria", *IEEE Int. Conf. on Acoust., Speech and Signal Processing*, p. II.359 - II.362, Minneapolis, 1993.
- VASEGHI, S.V & FRAYLING-CORK, R, "Restoration of Old Gramophone Recordings", *Journal of Audio Eng. Soc (JAEC)*, Vol. 40, No. 10, p. 791-800, October 1992.
- VIEIRA FILHO,J, SACART, P & CHIQUITO, J.G, "Redução de Ruído em Sinais de Voz: Análise e avaliação das técnicas clássicas baseadas na subtração espectral a curto-termos (SECT)", *13º Simpósio Brasileiro de Telecomunicações*, p.685-689, Águas de Lindóia, SP, setembro de 1995.
- WIDROW, B. et alii, "Adaptive Noise Cancelling: Principles and Applications", *Proceedings of the IEEE*, Vol.63, No.12, December-1975.
- YANG, J, "Frequency Domain Noise Suppression Approches in Mobile Telephone Systems, *IEEE Int. Conf. on Acoust., Speech and Signal Processing* , p. II.363-II.366, Minneapolis, 1993.
- ZELINSKI, R & NOLL, P., "Adaptive transform coding of speech signals", *IEEE Trans. Acoust. Speech and Signal Processing*, vol.25, march 1984.

## ***APÊNDICE***

---

## APÊNDICE

### *Obtenção do filtro redutor de ruído proposto por Ephraim & Malah*

Seja o sinal de voz ruidoso  $y(t)$  obtido a partir da soma de um sinal de voz puro  $v(t)$  com um ruído  $r(t)$ , analisado em um intervalo limitado entre 0 e T; sejam:

$$V(\omega) = |V(\omega)| \cdot e^{j\alpha(\omega)}; R(\omega) = |R(\omega)| \cdot e^{j\theta(\omega)} \quad Y(\omega) = |Y(\omega)| \cdot e^{j\theta(\omega)} \quad (A1)$$

suas representações em frequência.

A amplitude estimada do sinal  $v(t)$ , de acordo com as considerações dadas no capítulo II, em cada frequência  $\omega$  analisada, é dada por:

$$\left| \hat{V}(\omega) \right| = \frac{\int_0^{2\pi} \int_0^{2\pi} v(\omega) \cdot p[Y(\omega)/v(\omega), \alpha(\omega)] \cdot p[v(\omega), \alpha(\omega)] \cdot d\alpha(\omega) \cdot dv(\omega)}{\int_0^{2\pi} \int_0^{2\pi} p[Y(\omega)/v(\omega), \alpha(\omega)] \cdot p[v(\omega), \alpha(\omega)] \cdot d\alpha(\omega) \cdot dv(\omega)} \quad (A2)$$

onde

$$p[Y(\omega)/v(\omega), \alpha(\omega)] = \frac{1}{\pi\sigma_r^2(\omega)} \cdot \exp\left\{-\frac{1}{\sigma_r^2(\omega)} |Y(\omega) - v(\omega) \cdot e^{j\alpha(\omega)}|^2\right\} \quad (A3)$$

e

$$p[v(\omega), \alpha(\omega)] = \frac{v(\omega)}{\pi\sigma_v^2(\omega)} \cdot \exp\left\{-\frac{v^2(\omega)}{\sigma_v^2(\omega)}\right\} \quad (A4)$$

Tem-se ainda que  $\sigma_v^2(\omega) = E\{|V(\omega)|^2\}$  e  $\sigma_r^2(\omega) = E\{|R(\omega)|^2\}$ , que representam as variâncias do sinal de voz e do ruído, respectivamente.

Substituindo-se as equações A4 e A3 em A2 e considerando todas as variáveis como função de  $\omega$  (para facilitar a escrita das equações), obtém-se:



$$|\hat{V}(\omega)| = \frac{\int_0^\infty \int_0^{2\pi} v \frac{1}{\pi \sigma_R^2} \cdot \exp\left[-\frac{1}{\sigma_R^2} |Y - v \cdot e^{j\alpha}|^2\right] \cdot v \frac{1}{\pi \sigma_R^2} \cdot \exp\left[-\frac{v^2}{\pi \sigma_R^2}\right] d\alpha dv}{\int_0^\infty \int_0^{2\pi} v \frac{1}{\pi \sigma_R^2} \cdot \exp\left[-\frac{1}{\sigma_R^2} |Y - v \cdot e^{j\alpha}|^2\right] \cdot v \frac{1}{\pi \sigma_R^2} \cdot \exp\left[-\frac{v^2}{\pi \sigma_R^2}\right] d\alpha dv} \quad (A5)$$

Sendo  $|Y - v \cdot e^{j\alpha}|^2 = |Y|^2 - 2 \Re[Y \cdot (v \cdot e^{j\alpha})^*] + v^2$ , a equação A5 pode ser rescrita como:

$$|\hat{V}(\omega)| = \frac{\int_0^\infty \int_0^{2\pi} v^2 \cdot \exp\left[-\frac{v^2}{\sigma_r^2} \left(\frac{1}{\sigma_r^2} + \frac{1}{\sigma_v^2}\right)\right] \cdot \exp\left\{-\frac{1}{\sigma_r^2} [2 \Re(Y \cdot v \cdot e^{-j\alpha})]\right\} d\alpha dv}{\int_0^\infty \int_0^{2\pi} v \cdot \exp\left[-\frac{v^2}{\sigma_r^2} \left(\frac{1}{\sigma_r^2} + \frac{1}{\sigma_v^2}\right)\right] \cdot \exp\left\{-\frac{1}{\sigma_r^2} [2 \Re(Y \cdot v \cdot e^{-j\alpha})]\right\} d\alpha dv} \quad (A6)$$

Tendo que  $Y(\omega) = |Y(\omega)| \cdot e^{j\theta(\omega)}$ , então

$$2 \Re[Y \cdot (v \cdot e^{j\alpha})^*] = 2 \Re[|Y| \cdot e^{j\theta} \cdot v \cdot e^{-j\alpha}] = 2 \cdot |Y| \cdot v \cdot \cos(\theta - \alpha) \quad (A7)$$

Substituindo A7 em A6, obtém-se

$$|\hat{V}(\omega)| = \frac{\int_0^\infty v^2 \cdot \exp\left[-\frac{v^2}{\sigma_r^2} \left(\frac{1}{\sigma_r^2} + \frac{1}{\sigma_v^2}\right)\right] dv \cdot \int_0^{2\pi} \exp\left\{-\frac{1}{\sigma_r^2} [2 \cdot |Y| \cdot v \cdot \cos(\theta - \alpha)]\right\} d\alpha}{\int_0^\infty v \cdot \exp\left[-\frac{v^2}{\sigma_r^2} \left(\frac{1}{\sigma_r^2} + \frac{1}{\sigma_v^2}\right)\right] dv \cdot \int_0^{2\pi} \exp\left\{-\frac{1}{\sigma_r^2} [2 \cdot |Y| \cdot v \cdot \cos(\theta - \alpha)]\right\} d\alpha} \quad (A8)$$

Sabendo-se que (GRADSHTEYN & RYZHIK, 1980)

$$I_n(z) = \frac{1}{2\pi} \int_0^{2\pi} \cos(n\beta) \cdot \exp(2 \cdot \cos \beta) d\beta; \text{ definindo-se } \beta = (\nu - \alpha), z = \frac{1}{\sigma_R^2} 2|Y|v \text{ e } n=0; \text{ e}$$

lembrando-se que  $\cos(\theta - \alpha) = -\cos(\alpha - \theta)$ , da equação A8, obtém-se:

$$|\hat{V}(\omega)| = \frac{\int_0^\infty v^2 \cdot \exp\left[-\frac{v^2}{\sigma_r^2} \left(\frac{1}{\sigma_r^2} + \frac{1}{\sigma_v^2}\right)\right] \cdot I_0\left(\frac{1}{\sigma_r^2} 2|Y|v\right) dv}{\int_0^\infty v \cdot \exp\left[-\frac{v^2}{\sigma_r^2} \left(\frac{1}{\sigma_r^2} + \frac{1}{\sigma_v^2}\right)\right] \cdot I_0\left(\frac{1}{\sigma_r^2} 2|Y|v\right) dv} \quad (A9)$$

Definindo:

$$\frac{1}{\sigma^2} = \left( \frac{1}{\sigma_r^2} + \frac{1}{\sigma_v^2} \right) \quad e \quad \Theta = \frac{SNR\_prio}{1 + SNR\_prio} \cdot SNR\_post \quad (A10)$$

com

$$SNR\_prio = \frac{\sigma_v^2}{\sigma_r^2} \quad e \quad SNR\_post = \frac{|Y|^2}{\sigma_r^2} \quad (A11)$$

sendo, respectivamente, as relações sinal/ruído a priori e a posteriori, obtém-se

$$\left| \hat{V}(\omega) \right| = \frac{\int_0^\infty v^2 \cdot \exp\left[-\frac{v^2}{\sigma^2}\right] \cdot I_0\left(2 \cdot v \sqrt{\frac{\Theta}{\sigma^2}}\right) dv}{\int_0^\infty v \cdot \exp\left[-\frac{v^2}{\sigma^2}\right] \cdot I_0\left(2 \cdot v \sqrt{\frac{\Theta}{\sigma^2}}\right) dv} \quad (A12)$$

A integral presente na equação A12 é resolvida usando-se a função hipergeométrica. Sabe-se que (MIDDELTON, 1960)

$$\int_0^\infty x^\mu \cdot e^{-\alpha \cdot x^2} J_\nu(\beta \cdot x) dx = \frac{\beta^\nu \Gamma(0,5\nu + 0,5\mu + 0,5)}{2^{(\nu+1)} \alpha^{0,5(\mu+\nu+1)} \Gamma(\nu+1)} \cdot F\left(\frac{\nu + \mu + 1}{2}; \nu + 1; \frac{-\beta^2}{4\alpha}\right) \quad (A13)$$

onde  $\Gamma(\bullet)$  é a função gama e  $F(\bullet)$  é a função hipergeométrica.

Tendo que  $J_\nu(x) = I_\nu(ix)$ , da equação A12 e A13, tem-se:

$x = v$ ;  $\alpha = \frac{1}{\sigma^2}$ ; e  $\beta = 2i\sqrt{\frac{\Theta}{\sigma^2}}$ . Com estes valores, a equação A12 pode ser resolvida diretamente de A13, sendo portanto dada por:

$$\left| \hat{V}(\omega) \right| = \alpha^{-\frac{1}{2}} \frac{\Gamma(1,5)}{\Gamma(1,0)} \cdot \frac{F(1,5; 1,0; \Theta)}{F(1,0; 1,0; \Theta)} \quad (A14)$$

Para resolver a equação A14 tem-se que:

$F(\lambda; \gamma; z) = e^z F(\gamma - \lambda; \gamma; -z)$  e  $F(\lambda; \lambda; z) = e^z$ . Assim, de A14 tem-se que:

$$\left| \hat{V}(\omega) \right| = \alpha^{-\frac{1}{2}} \Gamma(1,5) \cdot F(-0,5; 1,0; -\Theta) \quad (\text{A15})$$

De acordo com a tabela apresentada por MIDDELTON (1960) sabe-se que:

$$F(-0,5; 1; -x) = e^{-x/2} \left[ (1+x) I_0(x/2) + x I_1(x/2) \right] \quad (\text{A16})$$

Também, de acordo com as equações A10 e A11, tem-se que

$$\alpha^{-1/2} = \left( \frac{1}{\sigma^2} \right)^{-1/2} = \frac{\sqrt{\Theta}}{SNR_{post}} \cdot |Y(\omega)| \quad (\text{A17})$$

Finalmente tem-se o filtro de Ephraim & Malah, dado por:

$$\left| \hat{V}(\omega) \right| = \Gamma(1,5) \cdot \frac{\sqrt{\Theta(\omega)}}{SNR_{post}(\omega)} \cdot \exp\left[\frac{-\Theta(\omega)}{2}\right] \cdot \left\{ [1+\Theta(\omega)] I_0\left[\frac{-\Theta(\omega)}{2}\right] + \eta(\omega) I_1\left[\frac{-\Theta(\omega)}{2}\right] \right\} \cdot |Y(\omega)|$$

---

## REFERÊNCIAS BIBLIOGRÁFICAS

- GRADSHTEYN, I.S & RYZHIK, I.M, "Table of Integrals, Series, and Products", Ed. Academics, New York, 1980.
- MIDDELTON, D, "Introduction to Statistical Communication Theory", Ed. McGraw-Hill, New York, 1960.