

Henrique Carlos Diniz

**EFEITO DA FASE EM TÉCNICAS DE REDUÇÃO DE
RUÍDOS EM SINAIS DE VOZ BASEADAS NA SNR A
PRIORI**

Ilha Solteira–SP
Setembro, 2013

Henrique Carlos Diniz

**EFEITO DA FASE EM TÉCNICAS DE REDUÇÃO DE
RUÍDOS EM SINAIS DE VOZ BASEADAS NA SNR A
PRIORI**

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Engenharia Elétrica da Faculdade de Engenharia de Ilha Solteira da UNESP, como parte dos requisitos para a obtenção do título de Mestre em Engenharia Elétrica.

Especialidade: Automação.

Jozué Vieira Filho
Orientador

Ilha Solteira–SP
Setembro, 2013

FICHA CATALOGRÁFICA

Elaborada pela Seção Técnica de Aquisição e Tratamento da Informação
Serviço Técnico de Biblioteca e Documentação da UNESP - Ilha Solteira.

D585e	Diniz, Henrique Carlos. Efeito da fase em técnicas de redução de ruídos em sinais de voz baseadas na SNR a priori / Henrique Carlos Diniz. – Ilha Solteira : [s.n.], 2013 67 f. : il. Dissertação (mestrado) - Universidade Estadual Paulista. Faculdade de Engenharia de Ilha Solteira. Área de conhecimento: Automação, 2013 Orientador: Jozué Vieira Filho Inclui bibliografia 1. Ruído. 2. Ruído em sinais de voz. 3. SNR a priori. 4. FFT.
-------	---



UNIVERSIDADE ESTADUAL PAULISTA
CAMPUS DE ILHA SOLTEIRA
FACULDADE DE ENGENHARIA DE ILHA SOLTEIRA

CERTIFICADO DE APROVAÇÃO

TÍTULO: Efeito da fase em técnicas de redução de ruídos em sinais de voz baseadas na SNR a priori

AUTOR: HENRIQUE CARLOS DINIZ

ORIENTADOR: Prof. Dr. JOZUE VIEIRA FILHO

Aprovado como parte das exigências para obtenção do Título de Mestre em Engenharia Elétrica ,
Área: AUTOMAÇÃO, pela Comissão Examinadora:

Prof. Dr. JOZUE VIEIRA FILHO
C Experimental de São João Da Boa Vista / Reitoria

Profa. Dra. SUELY CUNHA AMARO MANTOVANI
Departamento de Engenharia Elétrica / Faculdade de Engenharia de Ilha Solteira

Prof. Dr. MARCO APARECIDO QUEIROZ DUARTE
Departamento de Matemática / Universidade Estadual de Mato Grosso do Sul

Data da realização: 27 de setembro de 2013.

“Talvez o certo para você é o errado para mim. Claro, cada um é cada um, ninguém é igual a ninguém. Não vai ser por isso que vamos nos desentender. Duas pessoas discutirem e não chegarem a uma conclusão igual, é a melhor prova de que cada ser humano tem o seu valor e identidade própria.” - *Raul Seixas*

RESUMO

A redução de ruído em sinais de voz é importante nos mais variados contextos. Dentre as diferentes técnicas encontradas na literatura, a maioria tem como base o processamento no domínio da frequência, com implementações via Transformada de Fourier. A fase do sinal processado é, nesses casos, a própria fase do sinal ruidoso. Neste trabalho, o objetivo é avaliar os efeitos da fase do sinal limpo no processamento. Para tal, foram feitos testes utilizando métodos eficientes de redução de ruído em sinais de voz, tendo como base a Subtração Espectral e a Minimização do Erro Quadrático Médio, implementados usando a relação sinal/ruído a priori. Os resultados mostram que a fase é importante e altera a qualidade do sinal processado, abrindo caminho para novas técnicas de redução de ruído.

Palavras-chaves: Ruído em sinais de voz. SNR a priori. FFT.

ABSTRACT

The noise reduction in speech signals is important in various contexts. Among the various techniques in the literature, most of them are based in the frequency domain, with implementation through Fourier transform. The phase of the processed signal is, in these cases, the actual phase of the noisy signal. In this work, the objective is to evaluate the effects of the phase in the processing. To this end, tests were made using efficient methods of noise reduction in speech signals, based on the Spectral Subtraction and Minimizing the mean square error, implemented using the a priori signal/noise ratio. The results show that phase is important and changes the quality of the processed signal, opening studies to new techniques for noise reduction.

Keywords: Noise in speech. a priori SNR. FFT.

LISTA DE ILUSTRAÇÕES

Figura 1 – Forma de onda do sinal de voz.	18
Figura 2 – Diagrama de Blocos da aplicação da STFT.	24
Figura 3 – Janela de Kaiser.	25
Figura 4 – Formação do sinal ruidoso.	31
Figura 5 – Diagrama do processamento utilizando a fase.	33
Figura 6 – Formação do sinal ruidoso.	35
Figura 7 – Experimento com a soma dos vetores.	36
Figura 8 – Exemplo de ângulo entre os vetores onde ocorre subtração.	36
Figura 9 – Adaptação para avaliação da fase.	37
Figura 10 – Espectrogramas do sinal 1.	40
Figura 11 – Espectrogramas do sinal 2.	41
Figura 12 – Espectrogramas do sinal 3.	42
Figura 13 – Espectrogramas do sinal 4.	43
Figura 14 – Espectrograma do sinal 1 com SNR de 5 dB processados pela SE e estimativa da SNR a priori obtida pela TSNR .	45
Figura 15 – Espectrograma do sinal 1 com SNR de 15 dB processados pela SE e estimativa da SNR a priori obtida pela TSNR .	46
Figura 16 – Espectrograma do sinal 2 com SNR de 5 dB processados pela SE e estimativa da SNR a priori obtida pela TSNR .	47
Figura 17 – Espectrograma do sinal 2 com SNR de 15 dB processados pela SE e estimativa da SNR a priori obtida pela TSNR .	48
Figura 18 – Espectrograma do sinal 3 com SNR de 5 dB processados pela SE e estimativa da SNR a priori obtida pela TSNR .	49
Figura 19 – Espectrograma do sinal 3 com SNR de 15 dB processados pela SE e estimativa da SNR a priori obtida pela TSNR .	50
Figura 20 – Espectrograma do sinal 4 com SNR de 5 dB processados pela SE e estimativa da SNR a priori obtida pela TSNR .	51
Figura 21 – Espectrograma do sinal 4 com SNR de 15 dB processados pela SE e estimativa da SNR a priori obtida pela TSNR .	52
Figura 22 – Espectrograma do sinal 1 com SNR de 15 dB processados pela MMSE e estimativa da SNR a priori obtida pela TSNR .	53
Figura 23 – Espectrograma do sinal 1 com SNR de 5 dB processados pela MMSE e estimativa da SNR a priori obtida pela TSNR .	54
Figura 24 – Espectrograma do sinal 2 com SNR de 15 dB processados pela MMSE e estimativa da SNR a priori obtida pela TSNR .	54
Figura 25 – Espectrograma do sinal 2 com SNR de 5 dB processados pela MMSE e estimativa da SNR a priori obtida pela TSNR .	55

Figura 26 – Espectrograma do sinal 3 com SNR de 15 dB processados pela MMSE e estimativa da SNR a priori obtida pela TSNR.	55
Figura 27 – Espectrograma do sinal 3 com SNR de 5 dB processados pela MMSE e estimativa da SNR a priori obtida pela TSNR.	56
Figura 28 – Espectrograma do sinal 4 com SNR de 15 dB processados pela MMSE e estimativa da SNR a priori obtida pela TSNR.	56
Figura 29 – Espectrograma do sinal 4 com SNR de 5 dB processados pela MMSE e estimativa da SNR a priori obtida pela TSNR.	57

LISTA DE TABELAS

Tabela 1 – Resultados da PESQ para a SE aplicada ao sinal 1 com relação sinal/ruído de 5dB	58
Tabela 2 – Resultados da PESQ para a SE aplicada ao sinal 1 com relação sinal/ruído de 15dB	58
Tabela 3 – Resultados da PESQ para a SE aplicada ao sinal 2 com relação sinal/ruído de 5dB	58
Tabela 4 – Resultados da PESQ para a SE aplicada ao sinal 2 com relação sinal/ruído de 15dB	59
Tabela 5 – Resultados da PESQ para a SE aplicada ao sinal 3 com relação sinal/ruído de 5dB	59
Tabela 6 – Resultados da PESQ para a SE aplicada ao sinal 3 com relação sinal/ruído de 15dB	59
Tabela 7 – Resultados da PESQ para a SE aplicada ao sinal 4 com relação sinal/ruído de 5dB	60
Tabela 8 – Resultados da PESQ para a SE aplicada ao sinal 4 com relação sinal/ruído de 15dB	60
Tabela 9 – Resultados da PESQ para a MMSE aplicada ao sinal 1 com relação sinal/ruído de 5dB	61
Tabela 10 – Resultados da PESQ para a MMSE aplicada ao sinal 1 com relação sinal/ruído de 15dB	61
Tabela 11 – Resultados da PESQ para a MMSE aplicada ao sinal 2 com relação sinal/ruído de 5dB	62
Tabela 12 – Resultados da PESQ para a MMSE aplicada ao sinal 2 com relação sinal/ruído de 15dB	62
Tabela 13 – Resultados da PESQ para a MMSE aplicada ao sinal 3 com relação sinal/ruído de 5dB	62
Tabela 14 – Resultados da PESQ para a MMSE aplicada ao sinal 3 com relação sinal/ruído de 15dB	63
Tabela 15 – Resultados da PESQ para a MMSE aplicada ao sinal 4 com relação sinal/ruído de 5dB	63
Tabela 16 – Resultados da PESQ para a MMSE aplicada ao sinal 4 com relação sinal/ruído de 15dB	63

LIST OF ABBREVIATIONS AND ACRONYMS

TF	Transformada de Fourier
TDF, DFT	Transformada Discreta de Fourier (<i>Discrete Fourier Transform</i>)
FFT	Transformada Rápida de Fourier (<i>Fast Fourier Transform</i>)
STFT	Transformada de Fourier de Curto Prazo (<i>Short Time Fourier Transform</i>)
ISTFT	Transformada de Fourier de Curto Prazo Inversa (<i>Inverse Short Time Fourier Transform</i>)
SE	Subtração Espectral (<i>Spectral Subtraction</i>)
MMSE	Minimização do Erro Quadrático Médio (<i>Minimum Mean Square Error</i>)
SNR _{post}	Relação Sinal/Ruído a Posteriori (<i>a Posteriori Signal-to-Noise Ratio</i>)
SNR _{prio}	Relação Sinal/Ruído a Priori (<i>a Priori Signal-to-Noise Ratio</i>)
DVS	detecção de voz/silêncio
MSV	melhoramento de sinais de voz
OLA	<i>Overlap Addition</i>
PESQ	<i>Perceptual Evaluation of Speech Quality</i>
TSNR	Redução de Ruído em Duas Etapas (<i>Two Step Noise Reduction</i>)
MTSNR	Redução de Ruído em Duas Etapas Modificado (<i>Modified Two Step Noise Reduction</i>)

LIST OF SYMBOLS

$\hat{\bullet}$	indica estimação de um valor
σ_r^2	estimação da potência do ruído
$v(t)$	sinal de voz puro
$r(t)$	ruído aditivo
$y(t)$	sinal de voz contaminado pelo ruído aditivo
$ V(\bullet) ^2$	potência do sinal de voz puro
$ R(\bullet) ^2$	potência do ruído
$ Y(\bullet) ^2$	potência do sinal ruidoso
w_i	i -ésima janela analisada
f_k	frequência processada

SUMÁRIO

1	INTRODUÇÃO	14
2	TÉCNICAS BÁSICAS DE REDUÇÃO DE RUÍDO	17
2.1	SUBTRAÇÃO ESPECTRAL (SE)	17
2.2	ERRO QUADRÁTICO MÉDIO MÍNIMO (MMSE)	20
2.3	APLICAÇÃO DA RELAÇÃO SINAL/RUÍDO NA REDUÇÃO DE RUÍDO EM SINAIS DE VOZ	21
2.4	APLICAÇÃO DA SNR A POSTERIORI AO MSV	22
2.5	TÉCNICAS CLÁSSICAS DE REDUÇÃO DE RUÍDO USANDO A SNR A PRIORI	22
3	APLICAÇÕES BÁSICAS USANDO FFT	24
3.1	APLICAÇÃO DA TRANSFORMADA DE FOURIER DE CURTO PRAZO	24
3.2	ESTIMAÇÃO DA POTÊNCIA DO RUÍDO	25
3.3	ESTIMAÇÃO DA SNR A POSTERIORI	26
4	MÉTODOS PARA ESTIMAÇÃO DA SNR A PRIORI	27
4.1	DECISÃO DIRIGIDA (DD)	27
4.2	TÉCNICA DE REDUÇÃO DE RUÍDO EM DUAS ETAPAS (TSNR)	28
4.3	TÉCNICA DE REDUÇÃO DE RUÍDO EM DUAS ETAPAS MODIFICADO (MTSNR)	29
4.4	DECISÃO DIRIGIDA COM VALOR DE ALFA AUTOAJUSTÁVEL	29
4.5	VALOR DE α ADAPTATIVO USANDO A FUNÇÃO SIGMÓIDE	30
5	IMPLEMENTAÇÕES E AVALIAÇÕES	31
5.1	MECANISMO DE FORMAÇÃO DE UM SINAL RUIDOSO	31
5.2	USO DA SUBSTITUIÇÃO DA FASE NO PROCESSAMENTO	32
5.3	COMPARAÇÃO ENTRE FASE DO RUÍDO E SINAL PURO	34
6	RESULTADOS	38
6.1	SINAIS UTILIZADOS	38
6.2	RESULTADOS OBTIDOS	39
6.2.1	Espectrogramas	39
6.2.1.1	<i>Resultados da Subtração Espectral</i>	<i>44</i>
6.2.1.2	<i>Resultados da Minimização do Erro Quadrático Médio</i>	<i>53</i>
6.2.2	Medição de Qualidade Objetiva	57
7	CONCLUSÕES	65

1 INTRODUÇÃO

As técnicas de redução de ruído têm como propósito melhorar o sinal de voz degradado por um ruído aditivo, reduzindo o ruído de fundo presente no sinal. Um bom algoritmo deve realizar uma boa recuperação da onda original (a recuperação física do sinal), melhorar a inteligibilidade (os aspectos auditivos) do sinal ou reduzir a fadiga do ouvinte (DELLER; HANSEN; PROAKIS, 1993).

Artigos recentes, como os apresentados em Gerkmann, Krawczyk e Rehr (2012), Krawczyk e Gerkmann (2012), Paliwal, Wójcicki e Shannon (2011), têm procurado estudar o efeito da fase no processamento de sinais ruidosos quando o objetivo é melhorar a qualidade do sinal. O processamento da fase tem se mostrado importante e pode fazer com que métodos clássicos de redução de ruído possam ser aprimorados.

Obter uma maior qualidade ou inteligibilidade em sinais de voz ruidosos pode também contribuir para o aperfeiçoamento da performance em outras aplicações, tais como compressão de sinais, reconhecimento de sinais de voz e outros (DELLER; HANSEN; PROAKIS, 1993). Devido à grande possibilidade de aplicações, a discussão neste trabalho sempre tratará de algoritmos de aprimoramento aplicados a ruídos aditivos para falantes e ouvintes com sistemas de fala e auditivos normais.

Neste trabalho, será discutida a redução de ruído através da subtração espectral (SE) (BOLL, 1979b) e do método proposto por Ephraim e Malah (1984), avaliando a importância que a fase possa ter para o aprimoramento de sinais de voz. Para isto, foram feitos dois tipos de testes: o primeiro consiste em gerar um sinal ruidoso a partir de um sinal puro, guardando sua fase original para posterior restauração; já o segundo realiza uma comparação entre as fases do sinal puro e do ruído e, a partir dessa comparação, é possível aplicar um filtro redutor de ruído de maneira mais coerente com a degradação do sinal original. Testes comparativos são realizados para avaliar os diferentes resultados, mantendo como referência os resultados obtidos com os métodos originais. Embora tal situação não se aplique em um caso real, estes testes fornecem uma melhor compreensão dos limites de desempenho que são possíveis de se atingir através da fase e a sua importância.

Um dos principais objetivos neste trabalho é avaliar se os métodos tradicionais são sensíveis à variação de fase dos sinais de voz e ruído, o que pode levar a novas pesquisas na área de redução de ruído em sinais de voz.

Os ruídos de fundo considerados neste trabalho são do ambiente e aditivos. Alguns exemplos são: o ruído da cabine de um avião, ruído de carro, ruído no interior de uma fábrica, etc.

Os estudos sobre redução de ruído em sinais de voz se tornaram mais significativos a partir da década de 70, mas o notável desenvolvimento tecnológico que vem ocorrendo nas últimas décadas fomentou novos estudos e pesquisas.

No trabalho proposto por Ephraim (1992a) foi desenvolvido um estudo sobre estima-

tiva bayesiana para o aprimoramento de sinais degradados por ruídos aditivos estatisticamente independentes do sinal de voz, onde estimadores baseados no método do mínimo erro quadrático (MMSE) e na máxima probabilidade a posteriori (MAP) foram obtidos utilizando modelo oculto de Markov. Já no trabalho desenvolvido por Virag (1999), foi proposto um algoritmo baseado nas propriedades de mascaramento do sistema auditivo humano, permitindo uma adaptação automática de alguns parâmetros em sistemas de redução de ruído.

Mais recentemente, alguns trabalhos já vêm introduzindo discussões a respeito da influência das fases dos sinais de voz e ruído em algoritmos e técnicas de redução de ruído, buscando aumentar a eficiência de técnicas de redução de ruído já consagradas. Fardkhaleghi e Savoji (2010) discutem e avaliam o papel que a fase tem nas técnicas de aprimoramento utilizando a filtragem de Wiener e estatísticas mínimas de Martin. No artigo apresentado por Langarani, Veisi e Sameti (2012), é avaliado o aprimoramento obtido ao aplicar-se a fase no MMSE de amplitude espectral de curto tempo e método MMSE de amplitude log-espectral.

Sem dúvida, o problema da fase está colocado como um desafio para os novos estudos sobre redução de ruído em sinais de voz.

A redução de ruídos em sinais de voz é uma área de pesquisa de processamento de sinais que vem se desenvolvendo ao longo das últimas décadas, focando principalmente na melhoria da qualidade do sinal acústico. No entanto, as aplicações que utilizam sinais de voz têm evoluído e o ruído continua sendo um problema que merece muita atenção. Aplicativos que possuem reconhecimento de voz para execução de tarefas são, geralmente, pensados para facilitar o usuário. No caso de comandos de voz, tais aplicativos devem considerar a captação de ruído durante o envio dos comandos falados, pois geralmente o ambiente é ruidoso e irá gerar um ruído aditivo no sinal a ser processado. O ruído captado pelo microfone degrada a qualidade perceptiva da fala, a inteligibilidade e reduz a eficiência dos algoritmos de reconhecimento automático da fala, prejudicando, assim, o sistema como um todo.

Da mesma forma que as aplicações que utilizam sinais de voz têm evoluído, as técnicas de redução de ruído se tornam cada vez mais necessárias, estimulando pesquisas que buscam tanto o aprimoramento de técnicas clássicas como o desenvolvimento de novas técnicas.

Dentro do estudo de melhoramento de sinais de voz, com aplicação em redução de ruídos, são discutidos nesta dissertação diferentes métodos para estimativa da SNR a priori, sendo comparados com a SNR a priori obtida através dos valores reais de sinal de voz e ruído. É avaliada também a influência que a fase de um sinal pode ter no processamento da magnitude, procurando observar quais os melhores resultados que podem ser obtidos a partir de técnicas clássicas. Com este objetivo, após este capítulo introdutório, o texto está estruturado como segue:

- No capítulo 2 são apresentadas duas técnicas clássicas de redução de ruído: a subtração espectral proposto por Boll (1979b) e o método do erro quadrático médio mínimo proposta por Ephraim e Malah (1984). Adicionalmente, são apresentados e discutidos os parâmetros sinal/ruído a priori e a posteriori.

- No capítulo 3 faz-se um resumo da aplicação da Transformada de Fourier de Curto Prazo em sistemas de redução de ruído em sinais de voz, sendo apresentados os métodos utilizados para a estimativa da potência do ruído e a estimativa da SNR a posteriori.
- No capítulo 4 são apresentados e discutidos diferentes métodos para a estimação da SNR a priori, com testes e comparações que visam determinar a melhor técnica para ser usada nos métodos de redução de ruído estudados.
- No capítulo 5 são discutidos alguns pontos relacionados à influência que a fase do sinal puro e do ruído têm no momento da geração do sinal ruidoso. Além disso, são apresentadas duas propostas para o uso da fase no processamento, como uma forma de avaliar se os métodos tradicionais que não consideram a fase no processamento têm sua eficiência reduzida por desconsiderar essa informação.
- No capítulo 6 são apresentados os sinais que foram utilizados nos testes, os parâmetros usados nas implementações e os resultados obtidos. Para avaliar os resultados, foram usados espectrogramas e a medida objetiva PESQ.
- Por fim, no capítulo 7 são apresentadas as conclusões do trabalho.

2 TÉCNICAS BÁSICAS DE REDUÇÃO DE RUÍDO

A redução de ruídos em sinais de voz (*Speech Enhancement*) é um procedimento que necessita de uma análise específica para cada situação, devida à variedade de tipos de ruídos e os diferentes efeitos que podem causar em um sinal. Outra dificuldade encontrada se refere aos parâmetros a serem utilizados, que podem variar para cada tipo de aplicação.

Existem dois aspectos principais que devem ser considerados: inteligibilidade e qualidade. A inteligibilidade refere-se ao entendimento que o ouvinte tem das palavras presentes no sinal, sendo portanto algo mais objetivo; já a qualidade diz respeito à opinião do ouvinte, sendo assim algo mais subjetivo.

Diversos métodos para redução de ruídos operam no domínio da frequência, onde uma das principais ferramentas é a Transformada de Fourier de Curto Prazo (STFT).

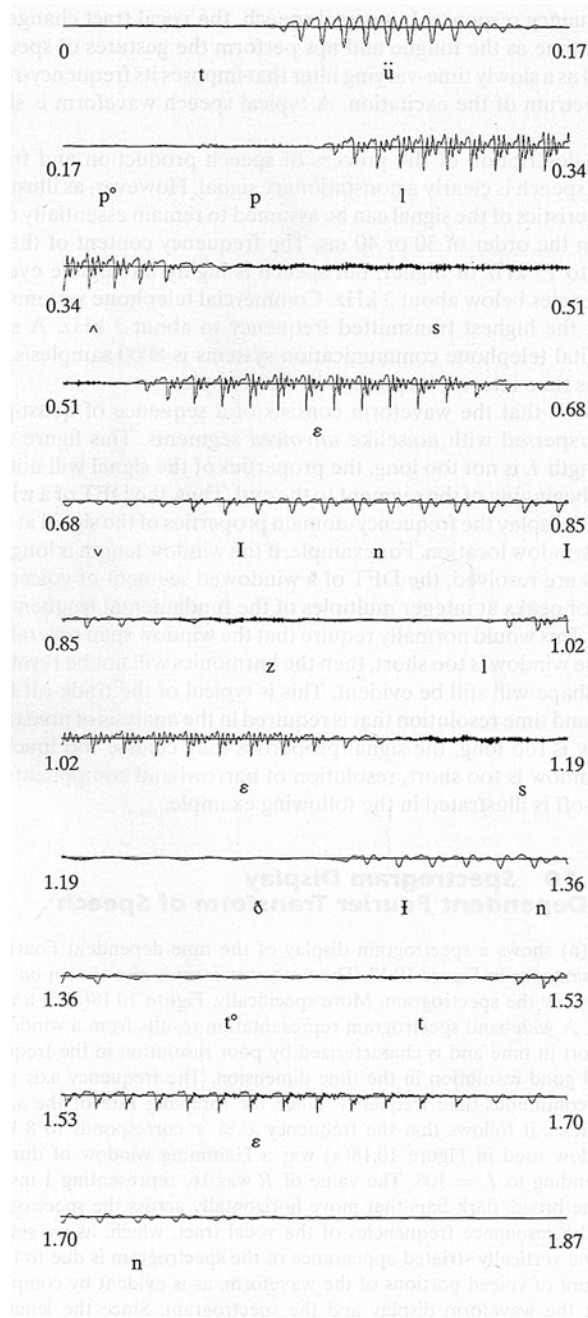
Neste capítulo serão estudadas duas importantes técnicas de redução de ruídos em sinais de voz, a subtração espectral (SE) (BOLL, 1979a; BOLL, 1979b) e o método do erro quadrático médio mínimo (MMSE) (EPHRAIM; MALAH, 1984). Na sequência, serão apresentados os dois principais parâmetros usados na melhoria das técnicas clássicas de redução de ruído: a relação sinal ruído a priori e a relação sinal ruído a posteriori.

2.1 SUBTRAÇÃO ESPECTRAL (SE)

Um ruído de fundo adicionado a um sinal de voz pode prejudicar processos que utilizem este sinal, tal como sistemas de reconhecimento automático de fala (ou voz). Como estes sistemas podem ser utilizados em qualquer ambiente é necessário o uso de métodos que permitam reduzir os possíveis ruídos adicionados ao sinal de voz original (ou puro). Dentre esses métodos tem-se a subtração espectral, uma técnica que realiza a melhoria no sinal de voz ruidoso através de uma subtração entre o espectro do sinal ruidoso e o espectro do ruído que contaminou o sinal. Para isto é necessária uma estimativa do espectro do ruído, que deve ser realizada durante os períodos de silêncio (trechos do sinal sem voz) (BOLL, 1979b).

A subtração espectral pode ser aplicada somente em sinais estacionários. No entanto, como pode ser visto na figura 1, onde cada linha tem duração de 0,17 ms, os sinais de voz são sinais claramente aleatórios e não estacionários. Contudo, como visto na figura 1 é possível assumir que as características do sinal permaneçam essencialmente constantes em intervalos de tempo de 30 a 40 ms. Nesses casos, pode-se considerar o sinal como localmente estacionário, sendo possível utilizar a Transformada de Fourier de Curto Prazo (STFT) para um processamento no domínio da frequência (OPPENHEIM; SCHAFER; BUCK, 1999).

Figura 1 – Forma de onda do sinal de voz “Two Plus seven is less than ten.”.



Source: Oppenheim, Schafer e Buck (1999)

Tomando, como exemplo, um sinal de voz puro $v(t)$ contaminado por um ruído aditivo $r(t)$, obtem-se um sinal ruidoso $y(t)$ da seguinte forma:

$$y(t) = v(t) + r(t) \quad (1)$$

Após realizar o processo de amostragem do sinal representado pela equação (1) tem-se,

na forma discreta:

$$y[n] = v[n] + r[n] \quad (2)$$

No domínio da frequência tem-se:

$$Y(\omega) = V(\omega) + R(\omega) \quad (3)$$

Tomando-se o quadrado na equação (3) e assumindo que o ruído é aditivo e descorrelacionado com o sinal de voz, obtém-se a seguinte equação:

$$|V(\omega)|^2 = |Y(\omega)|^2 - |R(\omega)|^2 \quad (4)$$

A partir da equação (4) é possível estimar o espectro de amplitude do sinal original, usando-se o sinal ruidoso e uma estimativa do ruído adicionado ao sinal.

Na maioria das técnicas de redução de ruído, apenas a magnitude do sinal ruidoso é processada, mantendo-se a fase para recuperação do sinal processado. Tal procedimento foi considerado apropriado por muito tempo, pois todos os estudos indicavam que a fase do sinal ruidoso era suficiente para garantir uma boa qualidade no processamento. Porém estudos recentes, como apresentados em Gerkmann, Krawczyk e Rehr (2012), Krawczyk e Gerkmann (2012), Paliwal, Wójcicki e Shannon (2011), mostram pesquisas com a proposta de avaliar possíveis benefícios relacionados ao processamento da fase.

A redução de ruído pode ser vista de forma simplificada como uma filtragem. No caso da SE básica, de acordo com a equação (4) e considerando que a potência do ruído é estimada e dada por $E\{|R(\omega)|^2\}$, a resposta em frequência do filtro redutor de ruído para processamento da magnitude será:

$$|H(\omega)| = \frac{\sqrt{|Y(\omega)|^2 - E\{|R(\omega)|^2\}}}{|Y(\omega)|} \quad (5)$$

Pela equação (5), observa-se que o método da subtração espectral pode ser aplicado apenas quando a potência estimada do ruído $E\{|R(\omega)|^2\}$ for menor ou igual à potência do sinal ruidoso $|Y(\omega)|^2$, o que não pode ser garantido devido à aleatoriedade da fase.

A aleatoriedade da fase pode, no processo de adição da equação (3), resultar na formação de um sinal ruidoso de potência inferior à do ruído. Uma possível solução é aplicar uma retificação de meia-onda onde, para cada frequência ω com potência de $|Y(\omega)|^2$ menor do que a estimativa da potência do sinal ruidoso $E\{|R(\omega)|^2\}$, a saída resultante será 0, como apresentado na equação (6) (BOLL, 1979a; BOLL, 1979b):

$$|H_1(\omega)| = \begin{cases} \frac{\sqrt{|Y(\omega)|^2 - E\{|R(\omega)|^2\}}}{|Y(\omega)|}, & \text{se } |Y(\omega)|^2 \geq E\{|R(\omega)|^2\} \\ 0 & \text{caso contrário} \end{cases} \quad (6)$$

Outra alternativa seria tomar o valor absoluto da equação (5), obtendo-se assim a seguinte função de transferência:

$$|H_2(\omega)| = \frac{\sqrt{||Y(\omega)|^2 - E\{|R(\omega)|^2\}}}{|Y(\omega)|} \quad (7)$$

No entanto, ambas as técnicas de atenuação apresentadas resultarão na presença de um ruído remanescente, causado pela alteração aleatória das amplitudes nas frequências ω onde isto ocorre. Este ruído é conhecido como ruído musical, sendo uma referência à presença de tons puros no ruído residual (CAPPE, 1994; EPHRAIM, 1992b).

2.2 ERRO QUADRÁTICO MÉDIO MÍNIMO (MMSE)

Na técnica desenvolvida por Ephraim e Malah (1984), os sinais de voz e ruído são modelados estatisticamente, sendo considerados processos aleatórios estacionários e independentes. Nesta técnica, os coeficientes da expansão de Fourier são considerados variáveis aleatórias gaussianas estatisticamente independentes. O modelamento é baseado no teorema do limite central, onde cada coeficiente de Fourier é considerado uma soma ponderada (ou integral) de variáveis aleatórias resultantes das amostras do processo. Assume-se que os processos possuem média igual a zero.

A partir da equação (3), é possível reescrever os sinais de voz puro e ruidoso como seguem:

$$\begin{aligned} V(\omega) &= |V(\omega)| e^{j\alpha(\omega)} \\ Y(\omega) &= |Y(\omega)| e^{j\theta(\omega)} \end{aligned} \quad (8)$$

O objetivo do MMSE é obter uma estimativa de $|V(\omega)|$ a partir do sinal degradado $y(t)$ no intervalo $0 \leq t \leq T$. Considerando as componentes espectrais como estatisticamente independentes, o estimador MMSE pode ser desenvolvido a partir de $|Y(\omega)|$ como definido na seguinte equação:

$$\begin{aligned} |\hat{V}(\omega)| &= E\{|V(\omega)| \mid y(t), 0 \leq t \leq T\} \\ &= E\{|V(\omega)| \mid Y_0, Y_1, Y_2, \dots, Y_k\} \\ &= E\{|V(\omega)| \mid Y_k\} \end{aligned} \quad (9)$$

onde, Y_k é a amostra de $|Y(\omega)|$, k indica a frequência que está sendo analisada no intervalo e $E\{\cdot\}$ representa o operador esperança.

Da equação (9) obtém-se:

$$|\hat{V}(\omega)| = \frac{\int_0^\infty \int_0^{2\pi} v(\omega) p[Y(\omega) \mid v(\omega), \alpha(\omega)] p[v(\omega), \alpha(\omega)] d\alpha(\omega) dv(\omega)}{\int_0^\infty \int_0^{2\pi} p[Y(\omega) \mid v(\omega), \alpha(\omega)] p[v(\omega), \alpha(\omega)] d\alpha(\omega) dv(\omega)} \quad (10)$$

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

Tendo definido a equação (10), o seguinte estimador é obtido:

$$|\hat{V}(\omega)| = \Gamma(1,5) \frac{\sqrt{\eta(\omega)}}{\gamma(\omega)} e^{\left(\frac{-\eta(\omega)}{2}\right)} \left\{ [1 + \eta(\omega)] I_0 \left[\frac{\eta(\omega)}{2} \right] + \eta(\omega) I_1 \left[\frac{\eta(\omega)}{2} \right] \right\} \cdot |Y(\omega)| \quad (11)$$

Da equação (11), tem-se:

$$\eta(\omega) = \frac{\xi(\omega)}{1 + \xi(\omega)} \gamma(\omega), \quad \text{sendo que,} \quad (12)$$

$$\xi(\omega) = \frac{\sigma_V^2(\omega)}{\sigma_R^2(\omega)} \quad \text{e} \quad \gamma(\omega) = \frac{|Y(\omega)|^2}{\sigma_R^2(\omega)}$$

Além disso, $I_0[\cdot]$ e $I_1[\cdot]$ representam as funções de Bessel modificadas de ordens zero e um e $\Gamma(\cdot)$ representa a função gama. Os parâmetros $\xi(\omega)$ e $\gamma(\omega)$ são interpretados como as relações sinal/ruído (SNR) a priori e a posteriori, respectivamente (MCAULAY; MALPASS, 1980).

Tendo definido o estimador na equação (11), o filtro definido em Ephraim e Malah (1984), pode ser escrito como segue:

$$|\hat{H}(\omega)| = \frac{\Lambda(\omega)}{1 + \Lambda(\omega)} G_{MMSE}(\omega) \quad (13)$$

onde, $G_{MMSE}(\omega)$ é definido pelo estimador apresentado na equação (11) da seguinte forma:

$$G_{MMSE}(\omega) = \frac{|\hat{V}(\omega)|}{|Y(\omega)|} \quad (14)$$

e $\Lambda(\omega)$ é a razão de verossimilhança generalizada, definida pela seguinte equação:

$$\Lambda(\omega) = \mu(\omega) \frac{e^{\eta(\omega)}}{1 + \text{SNR}_{\text{prio}}(\omega)} \quad (15)$$

Na equação (15), tem-se $\mu(\omega) = \frac{(1-q)}{q}$ e o valor q representa a probabilidade de ausência do sinal, um valor típico que apresenta bons resultados é 0,2 (EPHRAIM; MALAH, 1984).

2.3 APLICAÇÃO DA RELAÇÃO SINAL/RUÍDO NA REDUÇÃO DE RUÍDO EM SINAIS DE VOZ

No item anterior foi apresentada a técnica de redução de ruído baseada na subtração espectral. Agora serão introduzidos dois novos parâmetros: a relação sinal/ruído a posteriori (SNR_{post}) e a relação sinal/ruído a priori (SNR_{prio}). Estes novos parâmetros podem aprimorar a redução de ruídos através da subtração espectral. A aplicação apenas da SNR_{post} não permite eliminar completamente o ruído musical, abrindo espaço para a aplicação da SNR_{prio} .

Considerando-se os sinais puro, ruidoso e o próprio ruído aditivo, os parâmetros SNR_post e SNR_prio podem ser definidos como seguem nas equações (16) e (17) (MCAULAY; MALPASS, 1980):

$$\text{SNR}_{\text{post}}(\omega) = \frac{|Y(\omega)|^2}{|R(\omega)|^2} \quad (16)$$

$$\text{SNR}_{\text{prio}}(\omega) = \frac{|V(\omega)|^2}{|R(\omega)|^2} \quad (17)$$

2.4 APLICAÇÃO DA SNR A POSTERIORI AO MSV

Na seção 2.1, foi apresentado o equacionamento da função de transferência do filtro redutor de ruído, definida na equação (5). A partir das equações (5) e (16), o filtro pode ser reescrito da seguinte forma:

$$|\hat{H}(\omega)| = \sqrt{1 - \frac{1}{\text{SNR}_{\text{post}}(\omega)}} \quad (18)$$

O filtro apresentado na equação (18) será válido apenas se $\text{SNR}_{\text{post}}(\omega) \geq 1$, o que pode resultar em uma retificação caso a potência do sinal ruidoso esteja abaixo da potência do ruído estimado. No entanto, a aplicação do filtro em função da SNR_post pode resultar numa melhoria de desempenho, pois a retificação será feita na relação sinal/ruído a posteriori e não na forma direta como foi visto na seção 2.1. Assim, dependendo do processo de estimação da SNR_post, podem-se reduzir os efeitos do ruído musical (CAPPE, 1994).

2.5 TÉCNICAS CLÁSSICAS DE REDUÇÃO DE RUÍDO USANDO A SNR A PRIORI

A SNR_prio é uma boa alternativa para a solução do problema do ruído musical resultante da aplicação da SNR_post. Assim é possível reescrever a SNR_post em função da SNR_prio, efetuando-se a divisão da equação (4) por $|R(\omega)|^2$ como segue:

$$\frac{|Y(\omega)|^2}{|R(\omega)|^2} = \frac{|V(\omega)|^2}{|R(\omega)|^2} + 1 \quad (19)$$

A partir das definições das equações (16), (17) e (19) obtém-se a seguinte relação:

$$\text{SNR}_{\text{post}}(\omega) = \text{SNR}_{\text{prio}}(\omega) + 1 \quad (20)$$

A estimação da SNR_post com base na SNR_prio elimina definitivamente a necessidade de se efetuar uma retificação. Redefinindo então o filtro da subtração espectral apresentado na

equação (18), um novo filtro pode ser obtido como segue (SCALART; VIEIRA FILHO, 1996):

$$|\hat{H}(\omega)| = \sqrt{\frac{\text{SNR}_{\hat{\text{prio}}}(\omega)}{1 + \text{SNR}_{\hat{\text{prio}}}(\omega)}}. \quad (21)$$

3 APLICAÇÕES BÁSICAS USANDO FFT

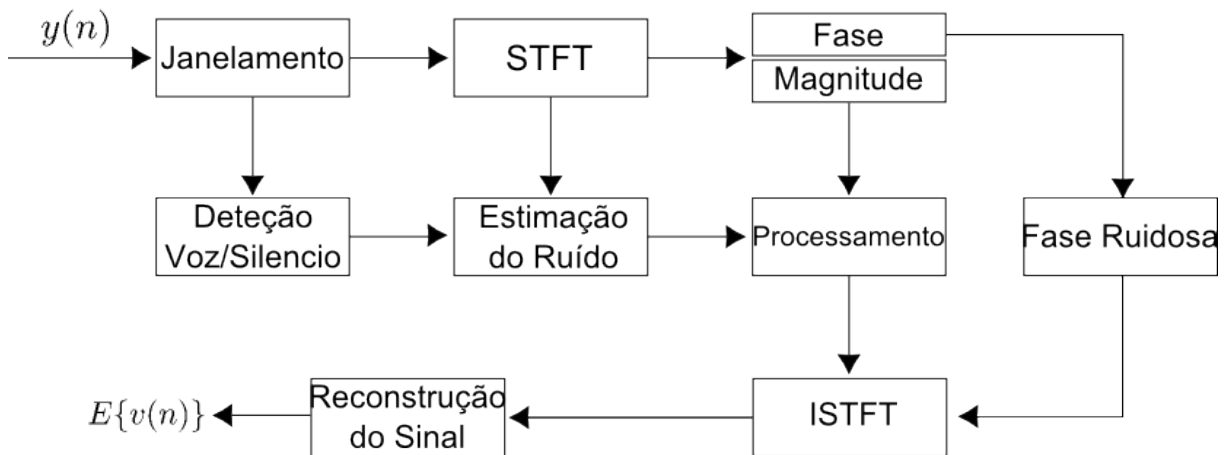
Neste capítulo, é apresentada a técnica que utiliza a Transformada de Fourier de Curto Prazo para a aplicação dos filtros estudados no capítulo anterior, bem como as técnicas de estimação para o ruído e para o parâmetro SNR a posteriori.

3.1 APLICAÇÃO DA TRANSFORMADA DE FOURIER DE CURTO PRAZO

O conceito da Transformada de Fourier de Curto Prazo é, geralmente, mais utilizado em técnicas para obtenção de estimativas espectrais em sinais não estacionários, como é o caso da subtração espectral quando aplicada a sinais de voz que possuem uma forte característica de não-estacionariedade. Um sinal não-estacionário é um sinal cujas propriedades estatísticas variam com o tempo. A Transformada de Fourier de Curto Prazo permite a estimação do espectro do sinal de voz em pequenos trechos, normalmente variando de 20 ms à 40 ms, que podem ser considerados estacionários (OPPENHEIM; SCHAFER; BUCK, 1999).

Na figura 2 apresenta-se o diagrama do sistema baseado na Melhoria de Sinais de Voz (MSV) através da STFT.

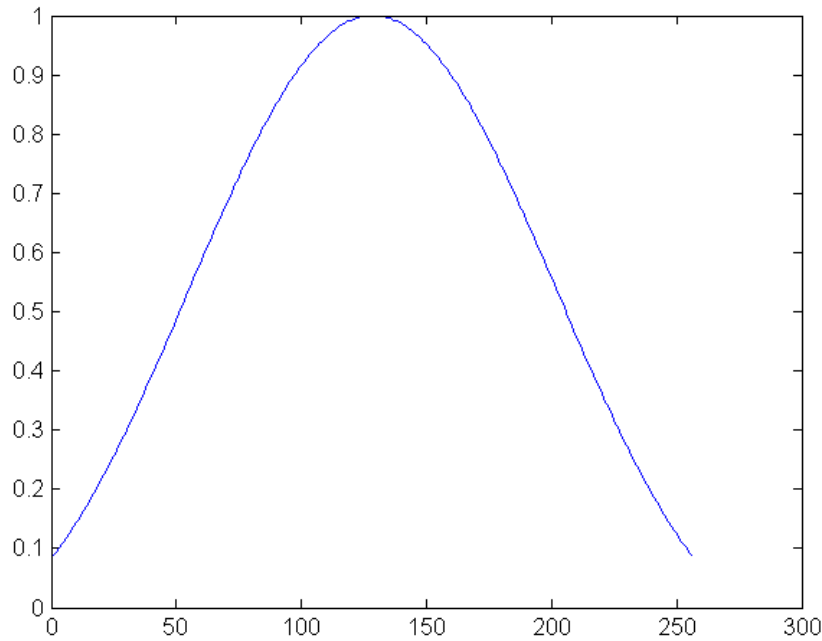
Figura 2 – Diagrama de Blocos da aplicação da STFT.



Source: Elaboração do autor.

O sinal passa inicialmente por uma janela (neste trabalho foi utilizada a janela de Kaiser), como apresentado na figura 3, cujo objetivo é garantir que o sinal discreto tenha duração limitada. Tal limitação permite o uso da STFT e garante que os trechos do sinal sejam aproximadamente estacionários. Também é necessário definir a taxa de sobreposição das janelas, geralmente sobrepostas com uma repetição de metade das amostras, ou seja, uma taxa de sobreposição de 50%.

Figura 3 – Janela de Kaiser com largura de banda de 256 amostras.



Source: Elaboração do autor.

A etapa de Detecção de Voz e Silêncio (DVS) apresentada na figura 2 tem por objetivo avaliar os intervalos de silêncio para que seja realizada a estimativa da potência do ruído. Para a próxima etapa escolhe-se o tipo de filtro que será implementado no processamento. Após o processamento será aplicada a ISTFT. Como não há estimativa da fase, é utilizada a fase do sinal ruidoso. Para a reconstrução do sinal aplica-se algum método de síntese, sendo que um dos mais empregados no processamento baseado na STFT é o *Overlap Add* (OLA) (SMITH, 2003).

3.2 ESTIMAÇÃO DA POTÊNCIA DO RUÍDO

Para que seja realizada a aplicação dos filtros redutores de ruídos, é necessária a estimação da potência do ruído. Essa estimação será realizada a partir dos intervalos de silêncio detectados pela DVS, fazendo-se uma filtragem recursiva de primeira ordem definida pela seguinte equação:

$$\sigma_r^2(w_i, f_k) = \beta \cdot \sigma_r^2(w_{i-1}, f_k) + (1 - \beta) \cdot |Y(w_i, f_k)|^2 \quad (22)$$

onde w_i é o i -ésimo quadro analisado, f_k a frequência processada, $\sigma_r^2(w_{i-1}, f_k)$ representa a estimação da potência do ruído do quadro anterior, $|Y(w_{i-1}, f_k)|^2$ representa a potência do sinal ruidoso e β é uma constante que define a dependência do ruído na janela atual de análise em relação às janelas anteriores.

Esta estimação é de grande importância para que o filtro tenha um bom desempenho na redução de ruído. Por exemplo, uma variação brusca de potência de uma janela para outra pode resultar no surgimento de ruído musical.

O valor de β será determinado com base no tamanho da janela, na frequência de amostragem do sinal e na constante de tempo que garanta o esquecimento das informações passadas, como segue:

$$\beta = e^{\left(\frac{-L}{2f_a T_c}\right)} \quad (23)$$

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 usual para T_c é 140 milissegundos.

3.3 ESTIMAÇÃO DA SNR A POSTERIORI

A SNR a posteriori será calculada a partir do sinal ruidoso e da estimativa do espectro de potência do ruído, como segue:

$$\text{SNR_p\acute{o}st}(w_i, f_k) = \frac{|Y(w_i, f_k)|^2}{\sigma_r^2(w_i, f_k)} \quad (24)$$

Pode-se observar na equação (24) que a relação sinal/ruído a posteriori acompanha diretamente as variações do sinal ruidoso, mostrando sua sensibilidade com as mudanças bruscas de fase do ruído. Já a estimativa da SNR a priori, por representar um sistema não-causal, demanda técnicas específicas para seu cálculo, não podendo ser obtida de forma direta como a SNR a posteriori. Alguns métodos para estimativa são apresentados no próximo capítulo.

4 MÉTODOS PARA ESTIMAÇÃO DA SNR A PRIORI

Como apresentado no capítulo anterior, a obtenção da SNR a priori depende dos valores do sinal de voz limpo e do ruído, que não estão disponíveis durante o processamento. Sendo assim, é necessário o uso de algum método de estimação para que se obtenha a SNR a priori. Dentre os métodos existentes um dos mais conhecidos e utilizados é o da decisão dirigida (*decision directed*) proposto por Ephraim e Malah (EPHRAIM; MALAH, 1984). Baseado nesta abordagem, foram desenvolvidas diversas adaptações para o método original, buscando-se melhorar o resultado da estimativa da SNR a priori e assim melhorar o resultado obtido após o processamento. Dentre as várias possibilidades, neste trabalho serão apresentadas e discutidas quatro: A Técnica de Redução de Ruído em Duas Etapas (*Two-step Noise Reduction Technique*) apresentada em Plapous et al. (2004) e Plapous, Marro e Scalart (2006); A Técnica de Redução de Ruído em Duas Etapas Modificado (*Modified Two-Step Noise Reduction Technique*) apresentada em Ou, Zhao e Gao (2007); o método da decisão dirigida modificado com um valor de α autoajustável apresentado em Hasan, Salahuddin e Khan (2004); e a decisão dirigida utilizando o valor de α adaptativo usando a função sigmóide definida em Park e Chang (2007).

4.1 DECISÃO DIRIGIDA (DD)

Para obtenção da estimativa da SNR a priori será utilizado o próprio sinal de voz estimado, o que o torna um sistema não-causal. Para contornar este problema, é utilizado um estimador que usa a potência do sinal de voz que foi estimada na janela anterior. Este processo pode ser realizado de forma satisfatória, pois, em intervalos curtos de tempo, a variação de potência de uma janela para a outra não é significativa, justificando assim o uso do estimador recursivo proposto por Ephraim e Malah (1984).

Sendo SNR_prio e SNR_post a relação sinal ruído a priori e a relação sinal ruído a posteriori, respectivamente; $|V(w_i, f_k)|^2$ e $\sigma_r^2(w_i, f_k)$ o espectro de potência do sinal de voz puro e o espectro de potência do ruído puro, respectivamente, sendo w_i o i -ésimo quadro analisado e f_k a frequência processada, o método de decisão dirigida deriva da definição da SNR_prio e da sua relação com a SNR_post, de modo que (EPHRAIM; MALAH, 1984):

$$\text{SNR_prio}(w_i, f_k) = \frac{E\{|V(w_i, f_k)|^2\}}{\sigma_r^2(w_i, f_k)} \quad (25)$$

$$\text{SNR_prio}(w_i, f_k) = E\{\text{SNR_post}(w_i, f_k) - 1\} \quad (26)$$

Usando (25) e (26) pode-se escrever a seguinte relação:

$$\text{SNR_prio}(w_i, f_k) = E\left\{\frac{1}{2} \frac{|V(w_i, f_k)|^2}{\sigma_r^2(w_i, f_k)} + \frac{1}{2} [\text{SNR_post}(w_i, f_k) - 1]\right\} \quad (27)$$

O estimador $\text{SNR}_{\text{prio}_{dd}}^{\wedge}(w_i, f_k)$ é deduzido a partir de (27) e é definido por:

$$\text{SNR}_{\text{prio}_{dd}}^{\wedge}(w_i, f_k) = \alpha \frac{|V(w_{i-1}, f_k)|^2}{\sigma_r^2(w_i, f_k)} + (1 - \alpha)P[\text{SNR}_{\text{post}}^{\wedge}(w_i, f_k) - 1] \quad (28)$$

onde α representa o grau de dependência da SNR_{prio} em relação à SNR_{post} e $P[\cdot]$ é o operador da retificação de meia onda para a transformação da SNR_{post} , já que seu valor pode ser menor do que 1, sendo definido por Boll (1979b):

$$P[x] = \begin{cases} x & \text{se } x \geq 0 \\ 0 & \text{caso contrário.} \end{cases} \quad (29)$$

Comparando-se (27) e (28) nota-se que o estimador $\text{SNR}_{\text{prio}}^{\wedge}(w_i, f_k)$ é obtido a partir da equação (27), substituindo-se o espectro de potência do quadro atual ($|V(w_i, f_k)|^2$) pelo espectro obtido no quadro anterior ($|V(w_{i-1}, f_k)|^2$). São incluídos ainda um fator de ponderação α nos dois termos e aplica-se o operador $P[\cdot]$ para garantir que o estimador seja sempre positivo caso, $\text{SNR}_{\text{post}}^{\wedge}(w_i, f_k) - 1$ seja negativo.

Observando a equação (28), nota-se que a contribuição da SNR_{post} no cálculo da SNR_{prio} está vinculada ao valor de α . Nos estudos apresentados por Ephraim e Malah (1984) foi observado que fortes variações na SNR_{post} podem afetar o desempenho da SNR_{prio} . No entanto, a ausência da SNR_{post} na estimação da SNR_{prio} provoca distorções perceptíveis do ponto de vista auditivo. Geralmente, os valores utilizados para α variam entre 0,9 e 1, sendo tipicamente $\alpha = 0,98$ (EPHRAIM; MALAH, 1984).

4.2 TÉCNICA DE REDUÇÃO DE RUÍDO EM DUAS ETAPAS (TSNR)

A SNR a priori calculada pela equação (28) introduz um atraso, pois o valor de α é próximo de 1. Assim, o ganho estimado na janela atual corresponde ao ganho da janela anterior. Para melhorar a estimação da SNR, o trabalho apresentado em Plapous, Marro e Scalart (2006), propõe o cálculo da SNR a priori em duas etapas. Para a primeira etapa, usa-se a decisão dirigida para calcular o ganho espectral $G_{dd}(w_i, f_k)$ definido pela seguinte equação (PLAPOUS et al., 2004; PLAPOUS; MARRO; SCALART, 2006):

$$G_{dd}(w_i, f_k) = \frac{\text{SNR}_{\text{prio}_{dd}}^{\wedge}(w_i, f_k)}{\text{SNR}_{\text{prio}_{dd}}^{\wedge}(w_i, f_k) + 1} \quad (30)$$

O ganho multiplicativo obtido nesta etapa é utilizado para então na etapa seguinte aprimorar a estimativa da relação sinal ruído a priori, por meio da seguinte equação:

$$\text{SNR}_{\text{prio}_{2step}}^{\wedge}(w_i, f_k) = \frac{|G_{dd}(w_i, f_k)Y(w_i, f_k)|^2}{\sigma_r^2(w_i, f_k)} \quad (31)$$

A partir da nova estimativa $\text{SNR}_{\text{prio}_{2step}}^{\wedge}(w_i, f_k)$, calcula-se novamente a função de ganho multiplicativo ($G_{2step}(w_i, f_k)$), que será então utilizada para o aprimoramento do sinal ruidoso por meio da seguinte equação:

$$\hat{V}(w_i, f_k) = G_{2step}(w_i, f_k)Y(w_i, f_k) \quad (32)$$

4.3 TÉCNICA DE REDUÇÃO DE RUÍDO EM DUAS ETAPAS MODIFICADO (MTSNR)

A Técnica de Redução de Ruído em Duas Etapas Modificado (MTSNR) traz uma mudança para a segunda etapa da TSNR convencional, alterando a forma de aprimoramento da estimativa da SNR_prio (OU; ZHAO; GAO, 2007). A proposta de alteração na segunda etapa da TSNR vem como uma forma de contornar o problema com o uso da função de ganho multiplicativo, que nem sempre é estável, o que pode afetar a performance do sistema de redução de ruídos. Para isto será feita a estimativa baseada diretamente no método do Erro Quadrático Médio Mínimo (MMSE), utilizando-se o resultado obtido pela decisão dirigida ($\hat{SNR}_{prio_dd}(w_i, f_k)$).

A segunda etapa então passará a ser realizada pela seguinte equação (OU; ZHAO; GAO, 2007):

$$\hat{SNR}_{prio_MTSNR}(w_i, f_k) = \frac{2\Gamma(1.5)\hat{SNR}_{prio_dd}(w_i, f_k)}{\sqrt{\pi}\hat{SNR}_{prio_dd}(w_i, f_k)+1} + \frac{(\hat{SNR}_{prio_dd}(w_i, f_k))^2|Y(w_i, f_k)|^2}{(\hat{SNR}_{prio_dd}(w_i, f_k)+1)^2\sigma_r^2(w_i, f_k)} \quad (33)$$

onde $\Gamma(\cdot)$ é a função gama representada pela equação:

$$\Gamma(h) = \int_{-\infty}^{\infty} t^{h-1} e^{(-t)} dt \quad (34)$$

4.4 DECISÃO DIRIGIDA COM VALOR DE ALFA AUTOAJUSTÁVEL

Na sua forma tradicional, o método de estimação de decisão dirigida utiliza um valor α para definir a dependência entre o primeiro e segundo membro da equação (28). O valor de α é determinado no intervalo $0 \leq \alpha \leq 1$, sendo que um valor mais próximo de 1 prioriza o uso do primeiro membro que é a parcela referente à estimativa do quadro anterior. Já um valor próximo de 0 prioriza o uso do segundo membro, que é a contribuição da SNR_prio ao quadro atual através da sua definição pela SNR_post.

Foi apresentado em Cappe (1994) que quanto mais próximo de 1 for o valor de α menor é o ruído musical. No entanto, isto também significa mais distorções no sinal processado. Como forma de balancear estes dois efeitos, resultados apresentados na literatura mostram que um valor aceitável para α se encontra no intervalo $0.95 \leq \alpha \leq 0.99$. Porém, um valor de α constante pode trazer certos prejuízos, pois em intervalos em que a SNR_post é mais estável, o uso de um valor menor para α leva a uma melhor contribuição do atual intervalo de análise na estimação SNR_prio.

Como uma forma de determinar um α autoajustável, como apresentado em Hasan, Salahuddin e Khan (2004), propõem-se uma expressão baseada na MMSE dada pela seguinte equação:

$$\alpha_{opt}(w_i, f_k) = \frac{1}{1 + \left(\frac{P[\hat{SNR}_{post}(w_i, f_k)-1] - \hat{SNR}_{prio}(w_{i-1}, f_k)}{P[\hat{SNR}_{post}(w_i, f_k)-1]+1} \right)^2} \quad (35)$$

O valor de α obtido pela equação (35) tende a acompanhar as variações da relação sinal/ruído, de forma a ajustar seu valor para um melhor aproveitamento da SNR_post nos intervalos mais estáveis ou apenas manter a estimativa da janela anterior em intervalos instáveis.

4.5 VALOR DE α ADAPTATIVO USANDO A FUNÇÃO SIGMÓIDE

O atraso introduzido pelo algoritmo de decisão dirigida diminui o desempenho do filtro redutor de ruído durante os intervalos de sinal que apresentam variações abruptas. Para contornar esse problema, em Park e Chang (2007) é proposto o uso da função sigmóide para o cálculo de α , de acordo com a variação da SNR a posteriori. Assim, tem-se:

$$\alpha_{sig}(w_i) = \frac{ke^{(-\beta'(s(w_i)-s_0))}}{1 + e^{(-\beta'(s(w_i)-s_0))}} + \sigma \quad (36)$$

onde

$$s(w_i) = \log \left(\frac{1}{\Delta \text{SNR}_{\text{post}}(w_i)} \right) = \log \left(\frac{1}{|\text{SNR}_{\text{post}}(w_i) - \text{SNR}_{\text{post}}(w_{i-1})|} \right) \quad (37)$$

Através de experimentos, os autores determinaram um $\alpha_{sig}(w_i)$ usando declive $\beta' = 0,4$, offset $s_0 = -9$ e constantes, $k = -1,5$ e $\sigma = 0,99$, respectivamente.

A função sigmóide torna α inversamente proporcional às variações da SNR a posteriori, enquanto a mantém dentro de um intervalo ($\alpha_{\min}$, $\alpha_{\max}$). As mudanças abruptas são caracterizadas por um aumento em $|\gamma(w_i) - \gamma(w_{i-1})|$. Esse aumento faz com que a função logarítmica diminua e com isso o valor de $\alpha_{sig}(w_i)$ se aproxime de $\alpha_{\min}$. Assim, um valor baixo de α diminui o impacto da janela anterior na atual. Por outro lado, se $|\gamma(w_i) - \gamma(w_{i-1})|$ diminui, a função logarítmica aumenta, e um alto valor de α é aplicado mantendo a vantagem do algoritmo da decisão dirigida, como a alta redução do ruído musical.

5 IMPLEMENTAÇÕES E AVALIAÇÕES

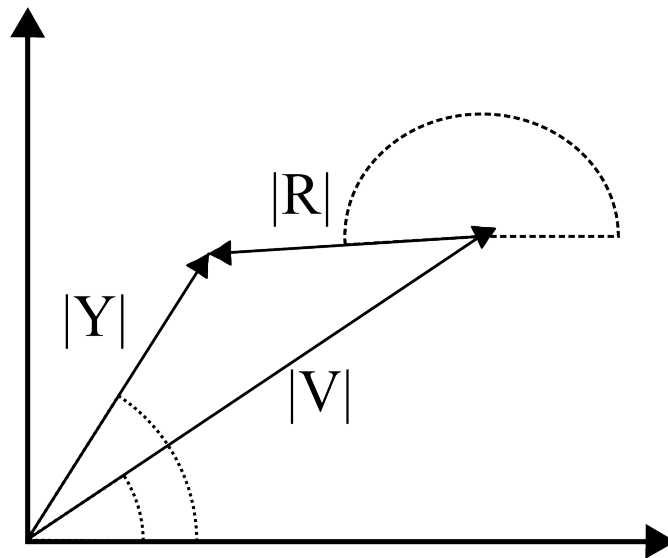
Neste capítulo é feito um breve estudo sobre a influência que a fase tem no momento da geração do sinal ruidoso. A partir das conclusões obtidas são apresentadas duas técnicas que buscam implementar a fase no processamento, sendo que essas técnicas são posteriormente aplicadas aos métodos de processamento estudados no capítulo 2.

5.1 MECANISMO DE FORMAÇÃO DE UM SINAL RUIDOSO

O modelo de geração do sinal ruidoso foi definido na equação (1). A forma mais usual de se realizar o processamento é transformar os sinais para o domínio da frequência através da transformada de Fourier, como definido na equação (3). Uma vez no domínio da frequência é possível separar a amplitude e a fase dos sinais, como visto na equação (8).

Quando o ruído é adicionado ao sinal de voz limpo, ele produzirá alterações tanto na amplitude quanto na fase do sinal, como ilustrado na figura 4:

Figura 4 – Mecanismo de formação do sinal ruidoso.



Source: Elaboração do autor.

Na figura 4, pode-se verificar o efeito que a fase do ruído pode causar na magnitude do sinal de voz puro, que é a alteração aleatória de sua potência.

De forma a ilustrar o efeito que a aleatoriedade da fase causa na magnitude de um sinal, considerando a equação (3), ela é reescrita de forma a separar magnitude e fase, como segue:

$$|Y(\omega)| e^{j\alpha(\omega)} = |V(\omega)| e^{j\theta(\omega)} + |R(\omega)| e^{j\gamma(\omega)} \quad (38)$$

Para simplificação do equacionamento, as respectivas amplitudes e fases serão representadas da seguinte forma: $|Y|$, $|V|$, $|R|$, α , θ e γ . Utilizando a fórmula de Euler $e^{jx} =$

$\cos(x) + j\sin(x)$, a equação (38) é reformulada como segue:

$$|Y|(\cos(\alpha) + j\sin(\alpha)) = |V|(\cos(\theta) + j\sin(\theta)) + |R|(\cos(\gamma) + j\sin(\gamma)) \quad (39)$$

Manipulando matematicamente a equação (39), é possível separar a parte real e a parte imaginária da equação, como segue:

$$|Y|\cos(\alpha) + j|Y|\sin(\alpha) = (|V|\cos(\theta) + |R|\cos(\gamma)) + j(|V|\sin(\theta) + |R|\sin(\gamma)) \quad (40)$$

Apesar da clareza nas modificações, uma exemplificação numérica mostra de forma simples o efeito da fase. Assim, considere três vetores v , r , e y com seis elementos, representando o sinal puro, o ruído e o sinal ruidoso, respectivamente, como seguem:

$$\begin{aligned} v &= [0.2, 0.5, -0.3, -0.1, 0.8, -0.4] \\ r &= [-0.6, 0.3, -0.1, 0.4, 0.5, -0.9] \\ y &= [-0.4, 0.8, -0.4, 0.3, 1.3, -1.3] \end{aligned} \quad (41)$$

O vetor y foi obtido a partir da equação (1). Aplicando-se a FFT, têm-se os respectivos vetores no

domínio da frequência:

$$\begin{aligned} V &= [0.7, (0.1 + 0.1732j), (-0.2 - 1.7321j), 0.7, (-0.2 + 1.7321j), (0.1 - 0.1732j)] \\ R &= [-0.4, (-1.5 - 0.5196j), (-0.1 - 1.5588j), 0, (-0.1 + 1.5588j), (-1.5 + 0.5196j)] \\ Y &= [0.3, (-1.4 - 0.3464j), (-0.3 - 3.2909j), 0.7, (-0.3 + 3.2909j), (-1.4 + 0.3463j)] \end{aligned} \quad (42)$$

A partir desses vetores, pode-se obter suas magnitude e fase, para posterior comparação da magnitude obtida com a equação (40) e a magnitude obtida com a soma direta das magnitudes individuais dos sinais V e R .

A magnitude do vetor Y obtido pela equação (40) é dada por:

$$|Y_1| = [0.3, 1.4422, 3.3045, 0.7, 3.3045, 1.4422] \quad (43)$$

Já a magnitude obtida pela soma direta entre as magnitudes de V e R é dada por:

$$|Y_2| = [1.1, 1.7875, 3.3056, 0.7, 3.3056, 1.7875] \quad (44)$$

Pode-se observar que existem variações nas magnitudes obtidas de formas diferentes. De forma análoga, pode-se esperar que o uso da fase no processo de recuperação do sinal possa ajudar na obtenção de uma estimativa mais precisa do sinal original.

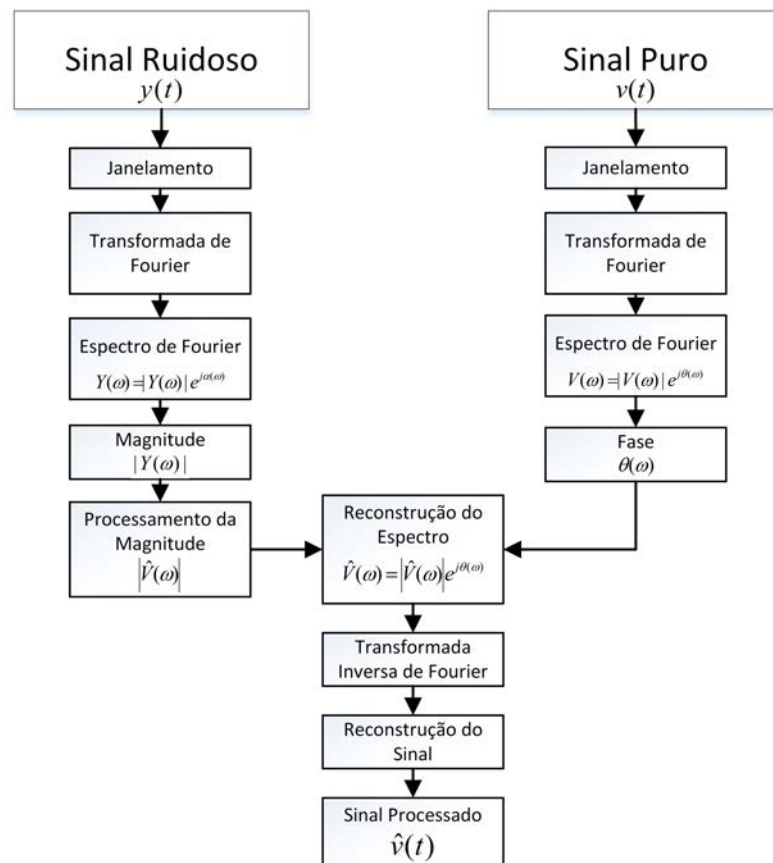
5.2 USO DA SUBSTITUIÇÃO DA FASE NO PROCESSAMENTO

O objetivo nesta seção é apresentar o método que foi empregado para a reconstrução da fase, em conjunto com os métodos de processamento clássicos discutidos nas seções 2.1 e 2.2, como uma forma de melhorar o processamento do sinal de voz.

Os métodos tradicionais de redução de ruído que operam da forma apresentada no diagrama da figura 2, realizam o processamento apenas na magnitude do sinal, mantendo a fase do sinal ruidoso no momento da reconstrução do sinal. Este procedimento é justificado na maioria dos casos porque, as variações aleatórias da fase não são percebidas pelo ouvido humano. Porém, estudos recentes têm apresentado resultados que podem alterar o paradigma da redução de ruído em sinais de voz (PALIWAL; WÓJCICKI; SHANNON, 2011; KRAWCZYK; GERKMANN, 2012; GERKMANN; KRAWCZYK; REHR, 2012).

Neste trabalho, o objetivo foi avaliar a influência da fase em diferentes aspectos, do ponto de vista de implementação de um filtro redutor de ruído. Os testes foram feitos com uma estimativa precisa da fase, de forma que é possível avaliar o nível de melhoria na qualidade do sinal processado, determinando o quanto a fase pode contribuir na qualidade do sinal. Com estas considerações, o diagrama da figura 2 foi reestruturado para permitir uma análise adequada da influência da fase, como apresentado na figura 5.

Figura 5 – Diagrama de blocos do procedimento aplicado nos testes de reconstrução da fase.



Source: Elaboração do autor.

De acordo com a figura 5, a nova estrutura de processamento consiste em dois ramos diferentes. O primeiro ramo é utilizado para análise do sinal ruidoso $y(t)$, realizando-se a estimativa de sua magnitude, enquanto o segundo ramo é utilizado na análise do sinal limpo $v(t)$, de onde é obtida a fase do sinal original (sem ruído).

Para realizar a reconstrução do espectro, combina-se a magnitude do sinal ruidoso processado, por qualquer um dos métodos estudados, com a fase original do sinal limpo, de acordo com a seguinte equação (KRAWCZYK; GERKMANN, 2012) :

$$\hat{V}(\omega) = \left(|Y_p(\omega)| e^{j\theta(\omega)} \right) e^{j\phi(\omega)L} \quad (45)$$

onde, $|Y_p(\omega)|$ é a magnitude do sinal ruidoso já processado, $\theta(\omega)$ é a fase do sinal original sem ruído, $\phi(\omega)$ é a frequência angular normalizada e L é um fator de subamostragem.

Com base no trabalho apresentado em Scalart e VIEIRA FILHO (1996), o algoritmo da SE baseado na SNR a priori e o filtro MMSE, de Ephraim e Malah (1984), foram implementados e avaliados com as seguintes configurações:

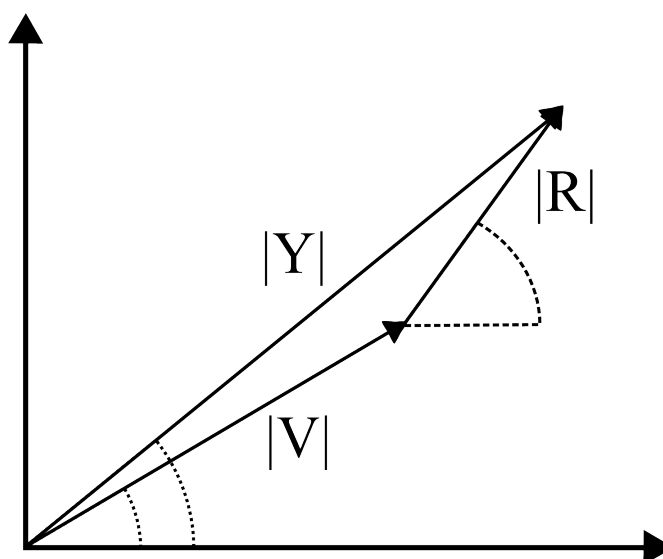
1. Original, como propostos pelos autores;
2. Modificados, com estimação da SNR a priori feita com o uso das potências dos sinais de voz limpo e ruído separadas.

A partir dessas configurações, incorporou-se um teste de alteração da fase no processamento final. Os resultados são importantes e serão apresentados no próximo capítulo.

5.3 COMPARAÇÃO ENTRE FASE DO RUÍDO E SINAL PURO

As técnicas para redução de ruídos apresentadas têm como base a magnitude do sinal, sempre desconsiderando o componente da fase. Isso se deve tanto ao fato de o ouvido humano não perceber possíveis variações que ocorram na fase, do ponto de vista auditivo, quanto a inexistência de um método eficiente para estimativa da fase. No entanto, a fase apresenta influência no momento da criação do sinal ruidoso. No caso da subtração espectral, a aleatoriedade da fase pode causar situações semelhantes à apresentada na figura 6, onde a magnitude resultante da adição é maior que a do sinal puro e casos como o da figura 4, onde a magnitude resultante é inferior à do sinal puro.

Figura 6 – Mecanismo de formação do sinal ruidoso.

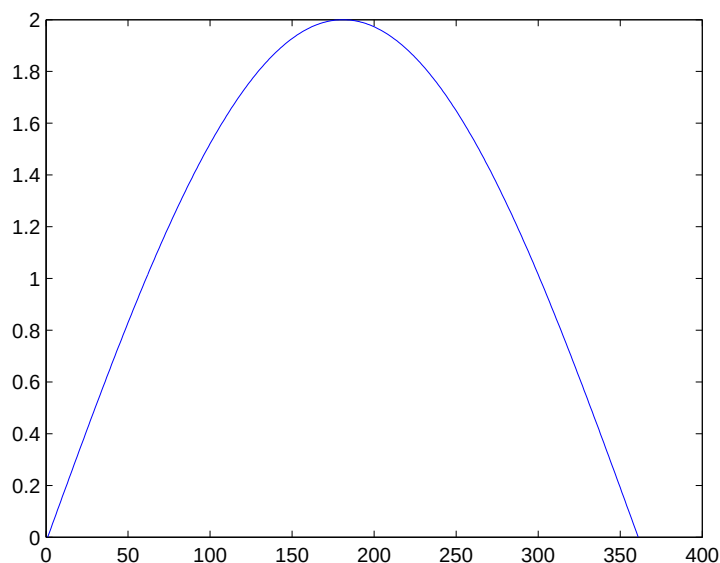


Source: Elaboração do autor.

Como o método clássico desconsidera a fase dos sinais, não é possível saber quando, no decorrer da criação do sinal ruidoso, as magnitudes do sinal de voz e do ruído são adicionadas ou subtraídas. Assim, ao aplicar-se a subtração espectral, os componentes da magnitude do sinal ruidoso, que possivelmente são resultantes da subtração entre o sinal puro e o ruído, passariam por uma nova subtração, reduzindo ainda mais a sua magnitude se comparada à do sinal original. Tendo em vista que a fase do sinal é aleatória, isso pode ocorrer com certa frequência, sendo um possível fator que cause redução na eficiência do método. De forma a avaliar o quanto a aleatoriedade da fase na geração do sinal ruidoso pode influenciar o processamento através da subtração espectral, foram feitas comparações entre a fase do ruído e a fase do sinal puro, observando-se pontos, no domínio da frequência, em que a magnitude do sinal foi gerada através da subtração das magnitudes originais.

Para avaliar os momentos em que os vetores são subtraídos, foi feito um teste onde dois vetores foram adicionados, sendo que, um deles foi mantido fixo enquanto o segundo foi rotacionado em 360° . Os dois vetores possuem magnitude igual a 1, portanto, no decorrer da adição, os valores resultantes variam no intervalo $[0, 2]$, como pode ser verificado na figura 7. Com o resultado obtido é possível determinar um intervalo angular no qual a resultante da soma dos vetores é menor do que 1, indicando uma subtração.

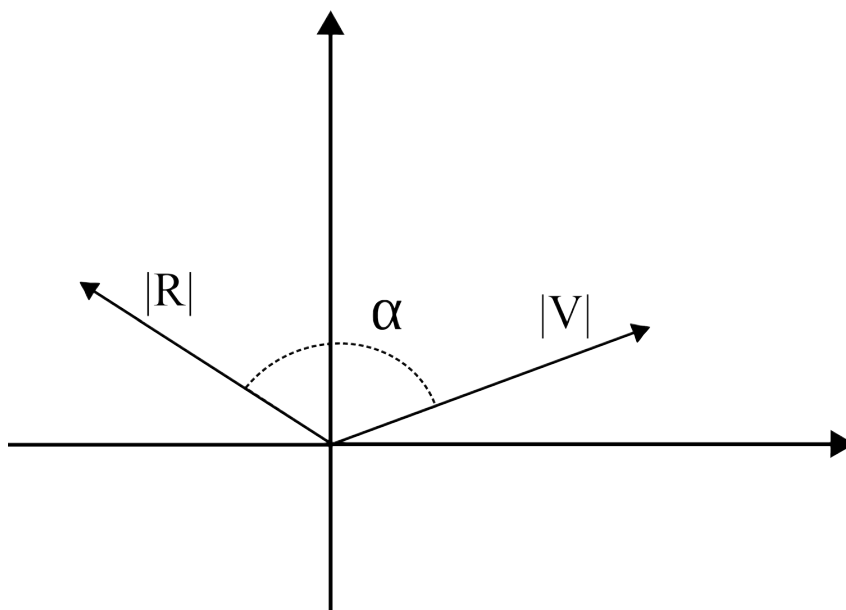
Figura 7 – Experimento com a soma dos vetores.



Source: Figura criada a partir do software Matlab®.

Dessa forma, constatou-se que os vetores são subtraídos quando o ângulo α entre eles estiver no intervalo $120^\circ \leq \alpha \leq 240^\circ$, com α variando no intervalo de $0^\circ \leq \alpha \leq 360^\circ$. Um exemplo é apresentado na figura 8.

Figura 8 – Exemplo de ângulo entre os vetores onde ocorre subtração.

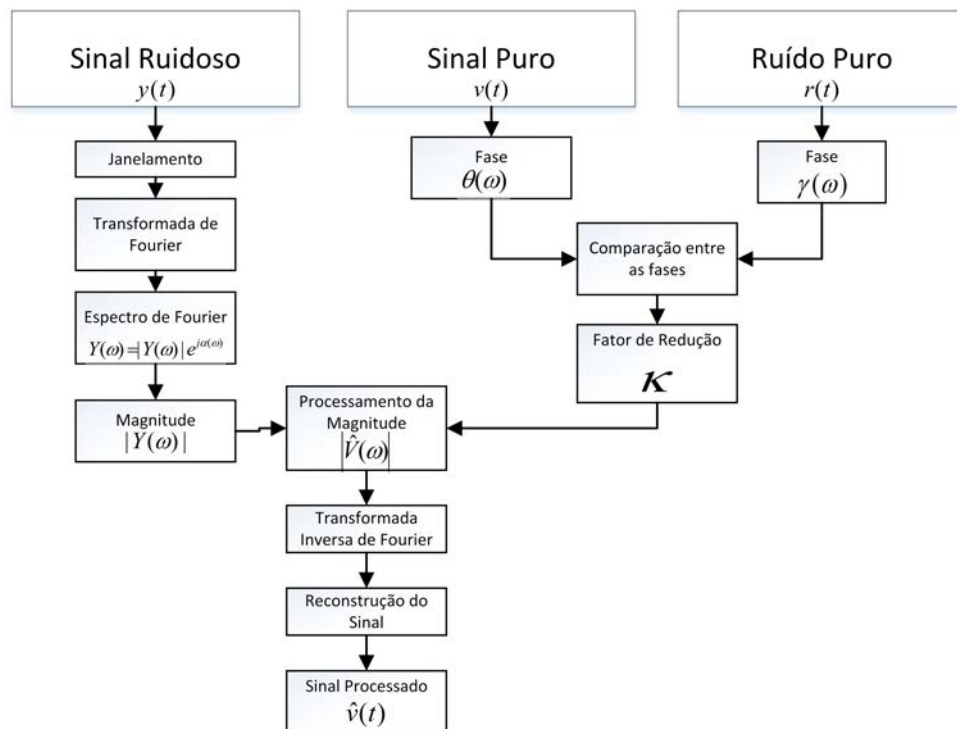


Source: Elaboração do autor.

De forma a avaliar uma possível redução na eficácia da aplicação da subtração espectral devido a aleatoriedade da fase, é proposto nesse trabalho uma adaptação para o método clássico

da subtração espectral. Na adaptação proposta, em cada janela processada é feita uma análise para determinar se a adição entre a magnitude do sinal e do ruído gerou, na realidade, uma subtração entre os vetores. Determinando-se os pontos em que a subtração ocorre pode-se, no momento da aplicação do filtro redutor de ruído gerado pela subtração espectral, aplicar-se um fator de redução κ para reduzir o efeito do filtro. Vários testes foram realizados e um valor de κ que permitiu gerar os melhores resultados foi 0,8. Um diagrama de blocos com as etapas da adaptação proposta é apresentado na figura 9.

Figura 9 – Adaptação para avaliação da fase.



Source: Elaboração do autor.

6 RESULTADOS

Neste trabalho, foram implementadas as técnicas clássicas de redução de ruído SE (BOLL, 1979b) e MMSE (EPHRAIM; MALAH, 1984). As implementações foram realizadas em varias formas sendo que, para comparação da estimativa da SNR a priori, foram utilizados 5 métodos diferentes, utilizando-se sempre o cálculo da SNR real como forma de estimar quais são os melhores resultados que podem ser alcançados.

De forma à avaliar como a fase pode interferir no processamento dos sinais, foram empregadas duas variações para a SE e uma para MMSE, o que permitiu uma boa avaliação da influência que a fase pode ter no processamento. Dessa forma, foram feitas as implementações da substituição da fase como visto na seção 5.2, tanto na SE como na MMSE, e a comparação entre as fases como apresentado na seção 5.3, sendo aplicada apenas na SE. Ambas as aplicações foram comparadas com os métodos tradicionais, sendo possível verificar quais obtiveram os melhores resultados. Neste capítulo serão apresentados alguns resultados obtidos através das simulações.

6.1 SINAIS UTILIZADOS

Os sinais utilizados foram amostrados a uma taxa de 16 kHz e o ruído de fundo é o barulho de um carro de porte pequeno rodando a 90 km/h. O sinal ruidoso foi obtido a partir da adição do ruído de fundo ao sinal de voz puro.

O sinal foi processado com as seguintes configurações para a STFT: janela de Kaiser com intervalos de sobreposição de 50%, 256 amostras e FFT (Fast Fourier Transform) de 256 pontos. A constante de tempo usada no cálculo da potência média do ruído foi de 140 ms. Para a realização da substituição da fase foi utilizado um fator de subamostragem $L = 32$ e para aplicação da correção baseada na fase foi utilizado um fator de redução $\kappa = 0,8$.

Para a simulação, foram utilizados quatro sinais livres da presença de ruídos, que foram corrompidos adicionando-se um ruído de carro. Os sinais utilizados foram:

Ruído: Ruído constante do interior de um carro pequeno com velocidade de 90 km/h.

Sinal 1: Gravado por uma voz masculina o sinal contém a seguinte frase: “A bolsa ficará estável ou sofrerá uma pequena queda”.

Sinal 2: Gravado por uma voz feminina o sinal contém a seguinte frase: “As três safras de grãos por ano são, arroz, trigo e num pomar de laranja, batata, soja e jasmim e amendoim”.

Sinal 3: Gravado por uma voz masculina o sinal contém a seguinte frase: “A speedy man can beat this track mark. He broke a new shoelace that day. The coffee stand is too high for the couch”.

Sinal 4: Gravado por uma voz feminina o sinal contém a seguinte frase: “Heave the line over the port side. A lathe cuts and trims any wood. It’s a dense crowd in two distinct ways.”.

Para realizar uma melhor comparação, foram feitos testes em dois diferentes níveis de relação sinal/ruído: 5 dB e 15 dB.

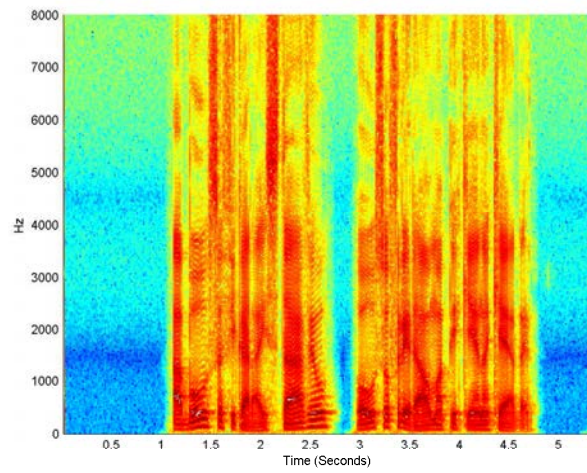
6.2 RESULTADOS OBTIDOS

6.2.1 Espectrogramas

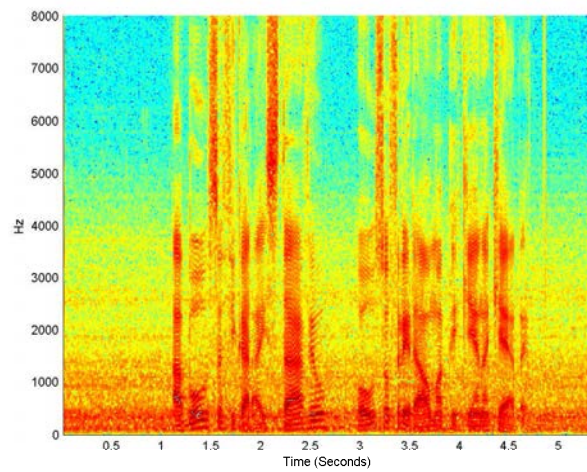
Como forma de avaliar os resultados obtidos nos testes, foram gerados os espectrogramas dos sinais, apresentados nas figuras 10, 11, 12 e 13, sendo que estas representam os sinais 1 a 4 em sua forma pura, ruidosa com SNR de 15 dB e ruidosa com SNR de 5 dB, respectivamente.

Figura 10 – Espectrogramas do sinal 1.

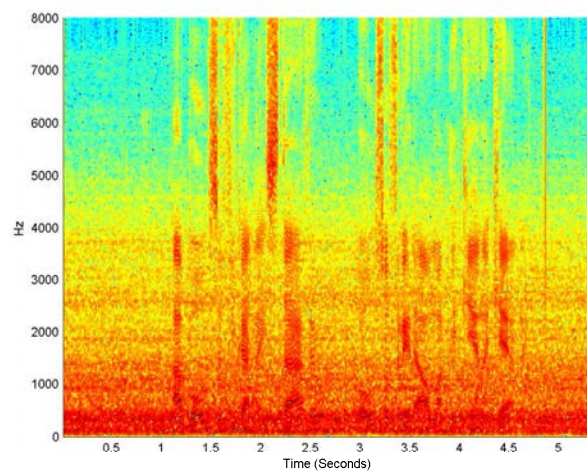
(a) Sinal puro



(b) Sinal ruidoso com SNR de 15 dB



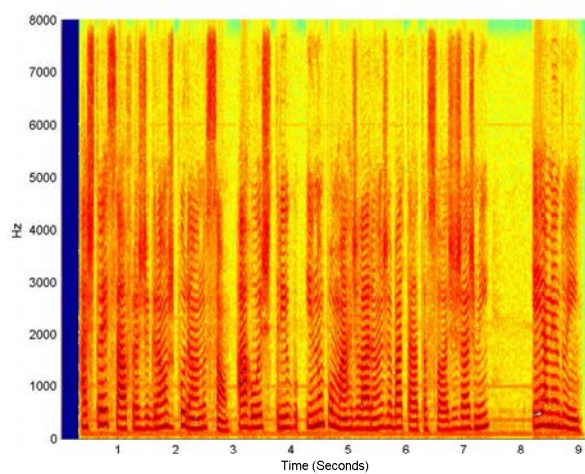
(c) Sinal ruidoso com SNR de 5 dB



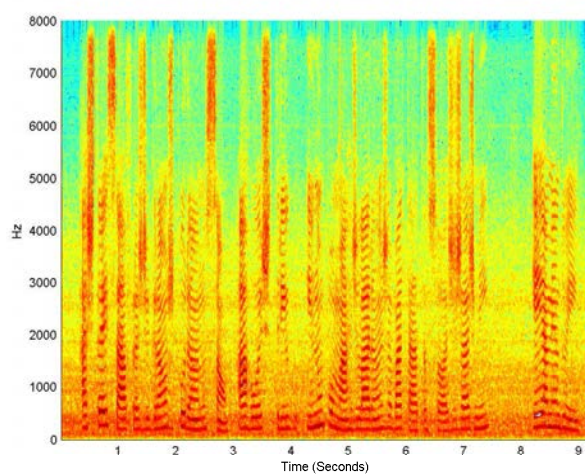
Source: Dados da pesquisa do autor.

Figura 11 – Espectrogramas do sinal 2.

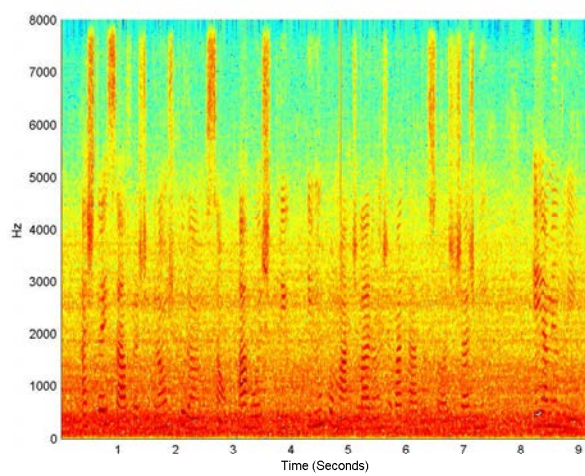
(a) Sinal puro



(b) Sinal ruidoso com SNR de 15 dB



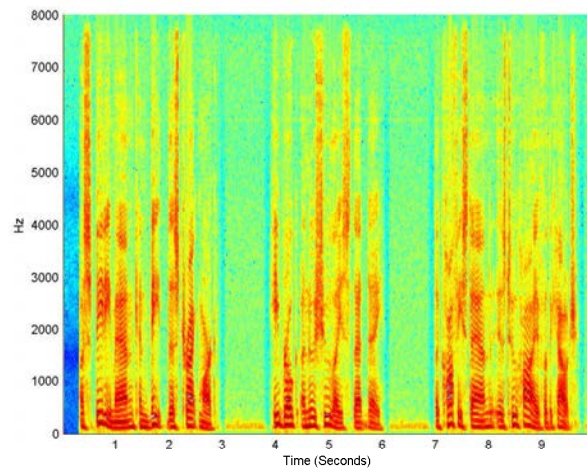
(c) Sinal ruidoso com SNR de 15 dB



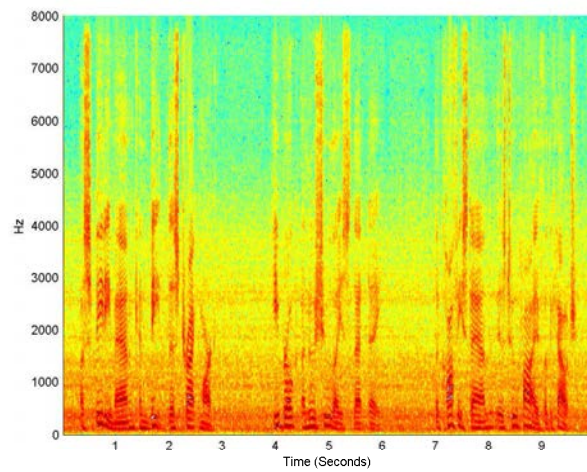
Source: Dados da pesquisa do autor.

Figura 12 – Espectrogramas do sinal 3.

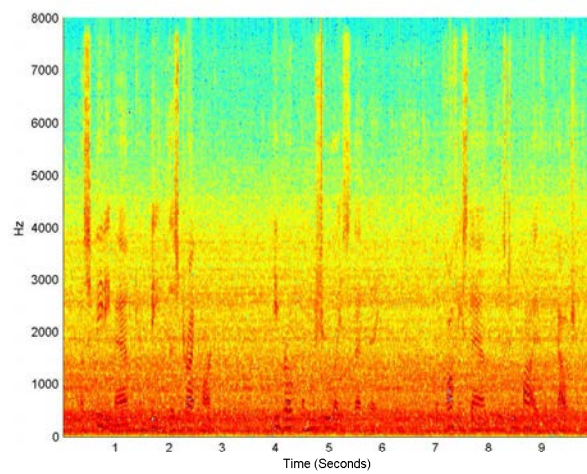
(a) Sinal puro



(b) Sinal ruidoso com SNR de 15 dB



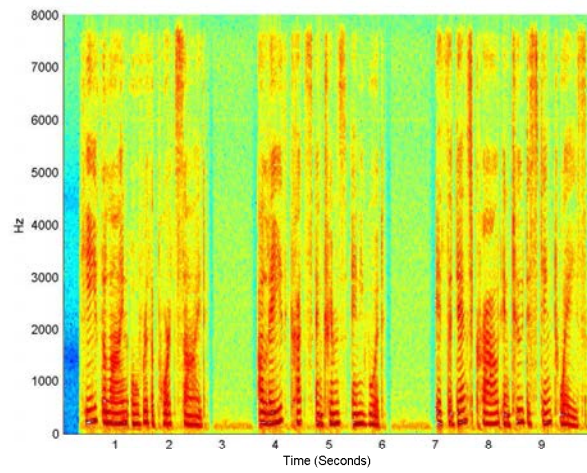
(c) Sinal ruidoso com SNR de 5 dB



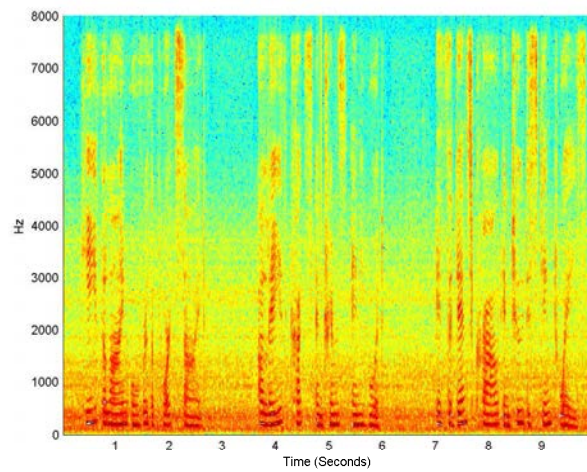
Source: Dados da pesquisa do autor.

Figura 13 – Espectrogramas do sinal 4.

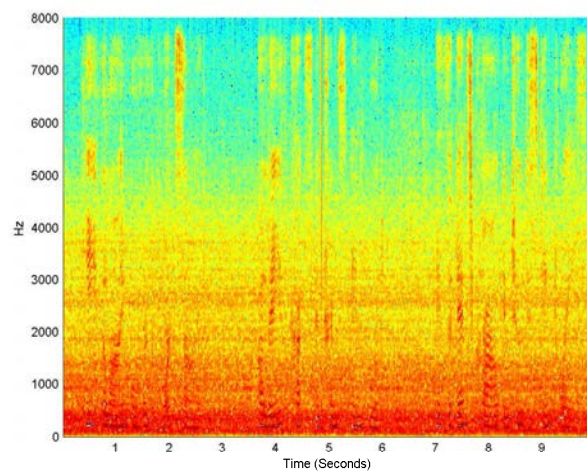
(a) Sinal puro



(b) Sinal ruidoso com SNR de 15 dB



(c) Sinal ruidoso com SNR de 5 dB



Source: Dados da pesquisa do autor.

A distorção causada pelo ruído no sinal de voz é identificável nos espectrogramas. No caso do ruído de carro, observa-se nos espectrogramas, que a maior interferência causada pelo ruído está concentrada nas baixas frequências, na região de 0 a 3000 Hz, que também é a região do espectro onde se concentra parte importante do sinal de voz. Isso não só afeta de forma significativa a qualidade do sinal de voz, mas também cria dificuldade para a redução do ruído.

6.2.1.1 Resultados da Subtração Espectral

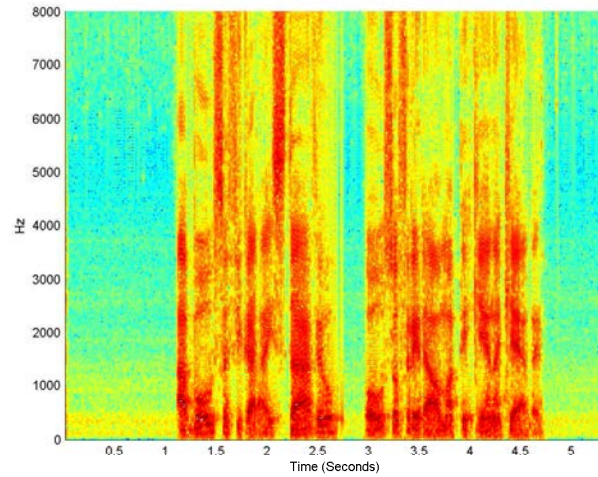
Os espectrogramas apresentados nas figuras 14 a 21, mostram os resultados obtidos a partir da aplicação da SE nos quatro sinais que foram avaliados, sendo que, cada sinal foi processado de três formas diferentes: na primeira, a SE foi aplicada de maneira direta; na segunda, a SE foi aplicada e a fase foi alterada de acordo com a técnica proposta na seção 5.2; por fim, uma terceira aplicação da SE foi realizada e novamente a fase foi alterada, agora de acordo com proposta definida na seção 5.3. Tais procedimentos foram adotados também para as diferentes técnicas de estimação da SNR a priori e aplicados nos sinais com dois diferentes níveis de ruído.

Para observação dos resultados foram selecionados os espectrogramas com os melhores resultados obtidos, para cada sinal testado tem-se um resultado do processamento com SNR de 5 dB e 15 dB, sendo composto por três espectrogramas, apresentando: o resultado do processamento tradicional da SE, uso da fase através da substituição e a comparação entre as fases. Como um resultado interessante, todos os espectrogramas selecionados foram obtidos com a SE tendo a SNR a priori estimada pelo método da TSNR, verificando-se a eficiência da estimação da estimação através dessa técnica.

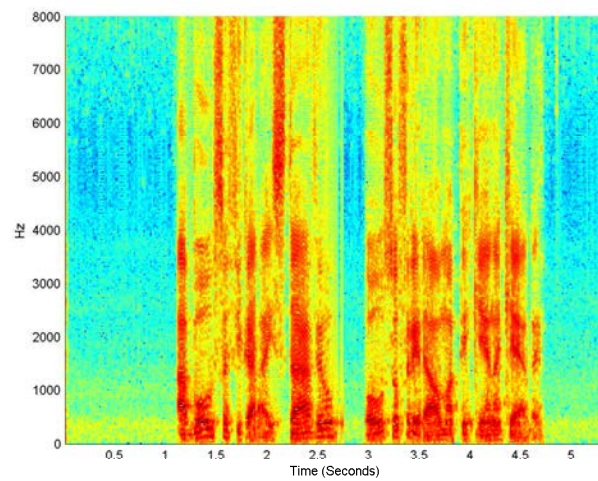
Os espectrogramas obtidos mostram as alterações que ocorrem ao utilizar-se a fase no processamento, verifica-se que a implementação da fase não só causa alterações nos resultados, como também possibilita uma melhoria a SE, podendo notar-se que principalmente uma redução na interferência que o ruído causa nas baixas frequências, uma das características do ruído que foi utilizado. Esses resultados se mostram importantes, pois é possível verificar que a fase pode ser aplicada como uma forma de incremento em uma técnica clássica.

Figura 14 – Espectrograma do sinal 1 com SNR de 5 dB processados pela SE e estimativa da SNR a priori obtida pela TSNR .

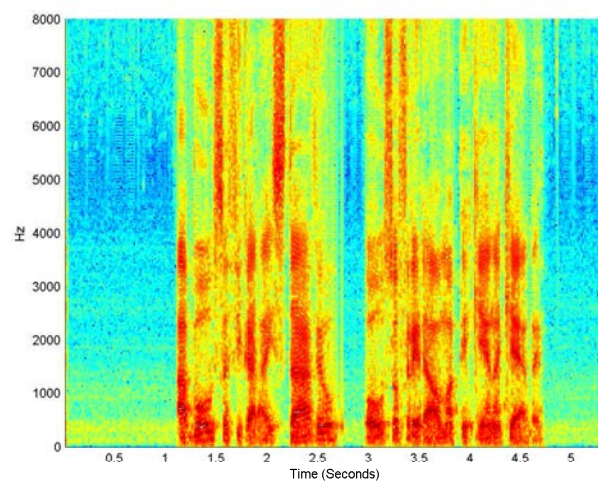
(a) Resultado da aplicação tradicional da SE.



(b) Resultado da aplicação da SE juntamente da substituição da fase



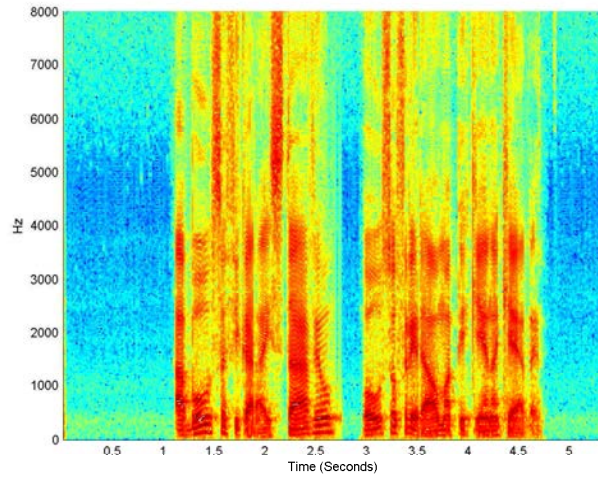
(c) Resultado da aplicação da SE com a correção feita pela fase.



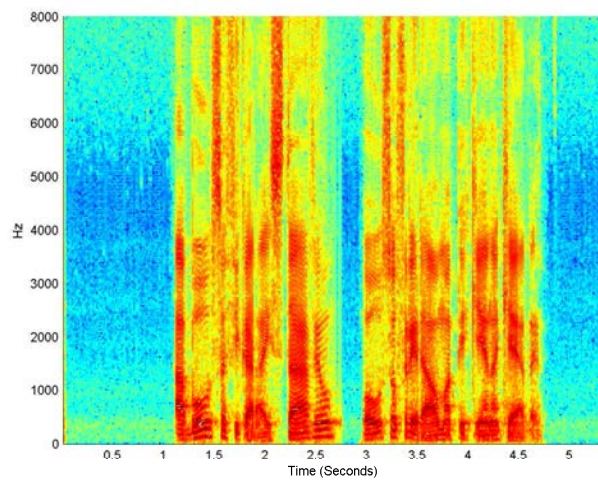
Source: Dados da pesquisa do autor.

Figura 15 – Espectrograma do sinal 1 com SNR de 15 dB processados pela SE e estimativa da SNR a priori obtida pela TSNR .

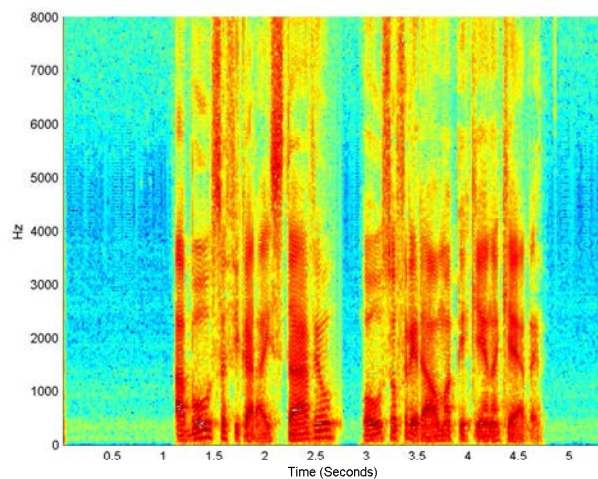
(a) Resultado da aplicação tradicional da SE



(b) Resultado da aplicação da SE juntamente da substituição da fase



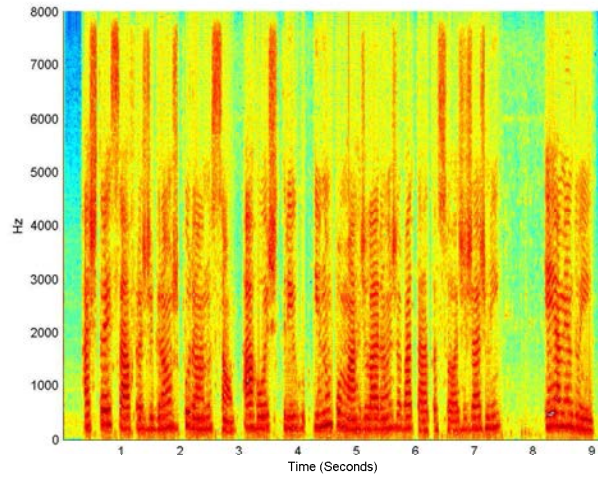
(c) Resultado da aplicação da SE com a correção feita pela fase.



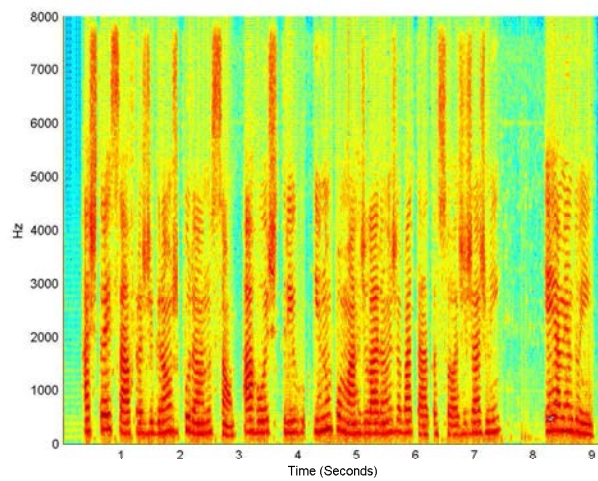
Source: Dados da pesquisa do autor.

Figura 16 – Espectrograma do sinal 2 com SNR de 5 dB processados pela SE e estimativa da SNR a priori obtida pela TSNR .

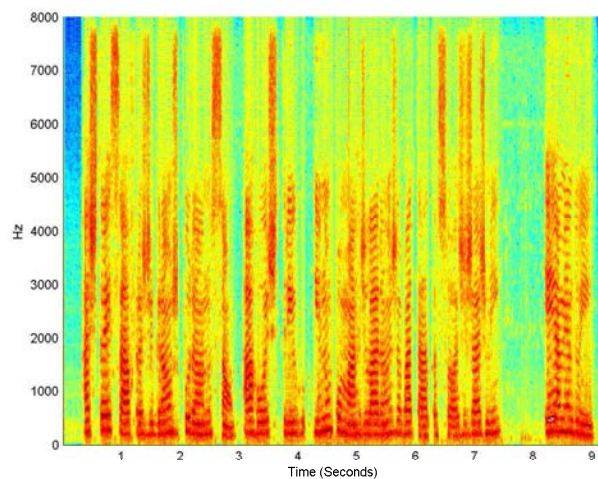
(a) Resultado da aplicação tradicional da SE



(b) Resultado da aplicação da SE juntamente da substituição da fase



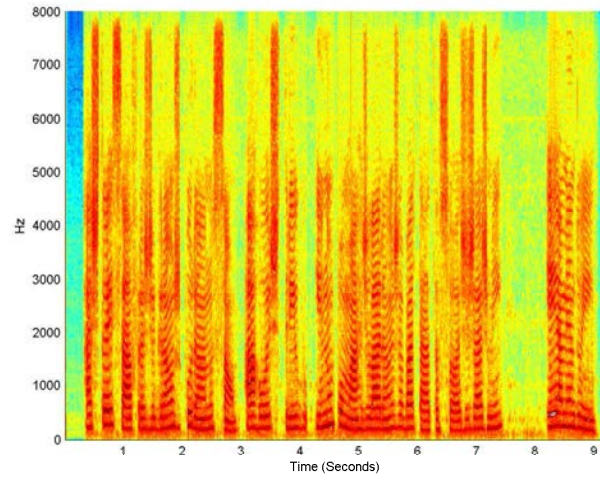
(c) Resultado da aplicação da SE com a correção feita pela fase.



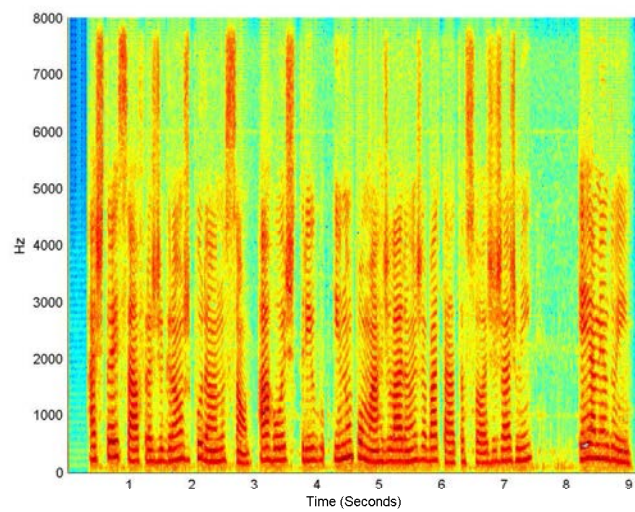
Source: Dados da pesquisa do autor.

Figura 17 – Espectrograma do sinal 2 com SNR de 15 dB processados pela SE e estimativa da SNR a priori obtida pela TSNR .

(a) Resultado da aplicação tradicional da SE



(b) Resultado da aplicação da SE juntamente da substituição da fase



(c) Resultado da aplicação da SE com a correção feita pela fase.

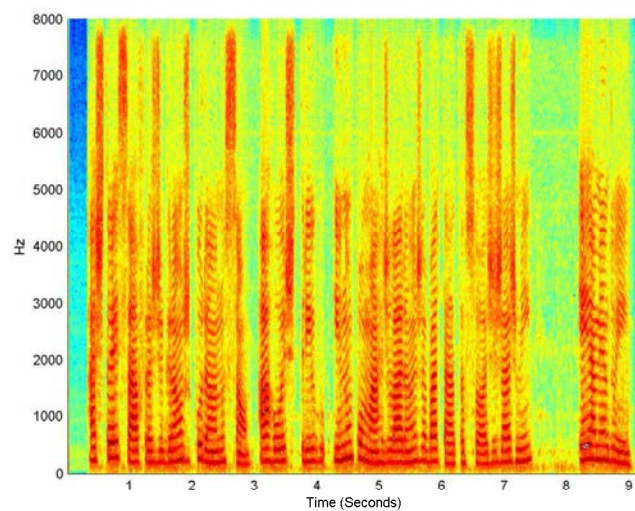
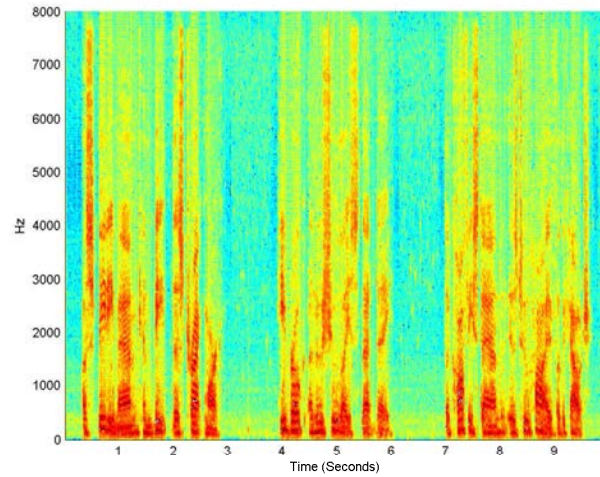
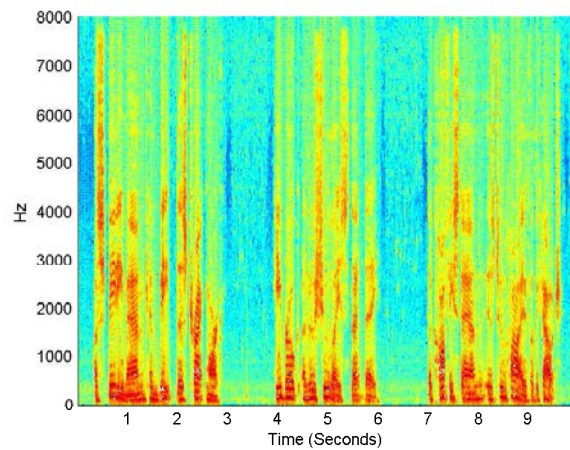


Figura 18 – Espectrograma do sinal 3 com SNR de 5 dB processados pela SE e estimativa da SNR a priori obtida pela TSNR .

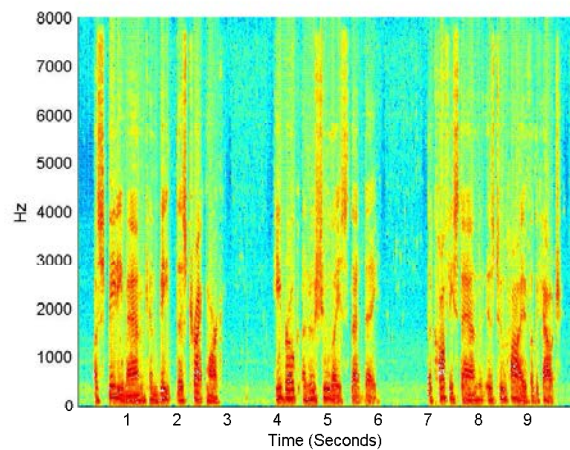
(a) Resultado da aplicação tradicional da SE



(b) Resultado da aplicação da SE juntamente da substituição da fase



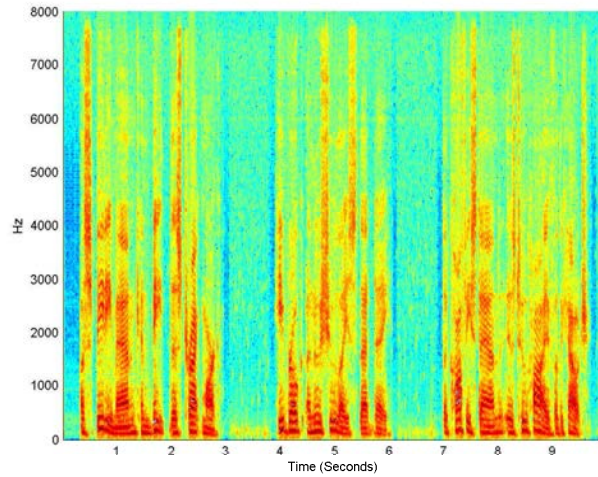
(c) Resultado da aplicação da SE com a correção feita pela fase.



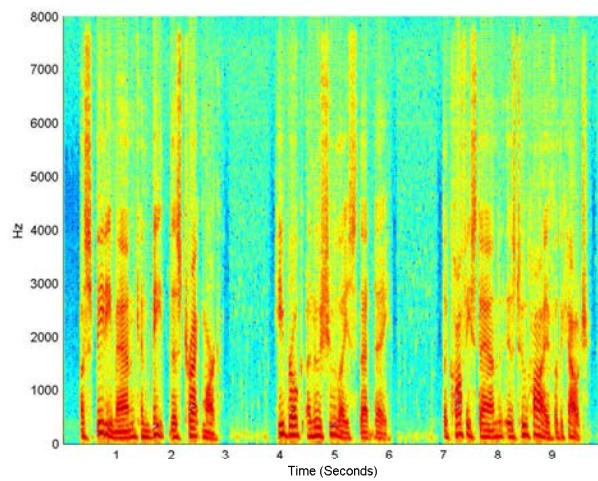
Source: Dados da pesquisa do autor.

Figura 19 – Espectrograma do sinal 3 com SNR de 15 dB processados pela SE e estimativa da SNR a priori obtida pela TSNR .

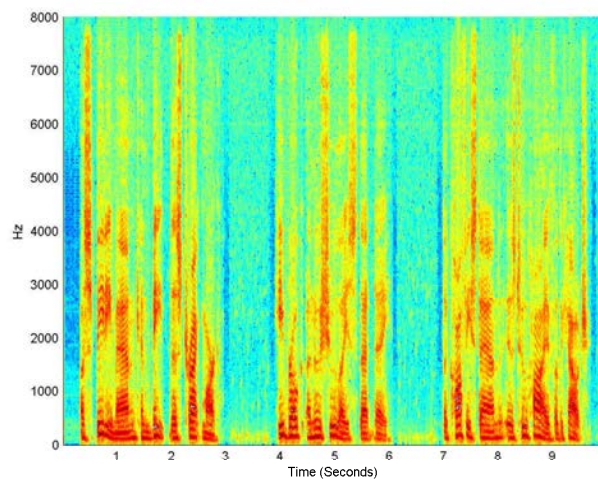
(a) Resultado da aplicação tradicional da SE



(b) Resultado da aplicação da SE juntamente da substituição da fase



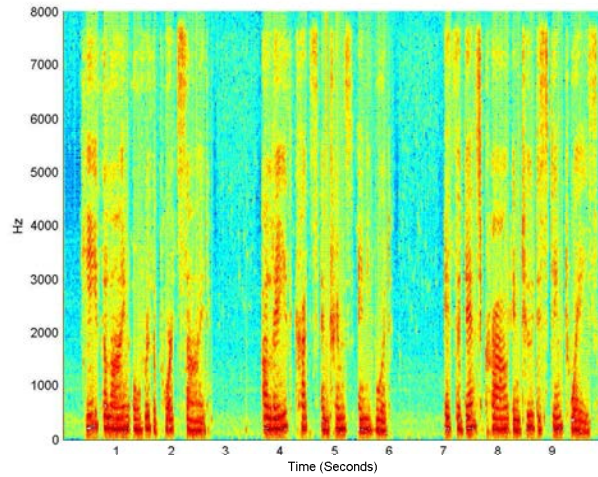
(c) Resultado da aplicação da SE com a correção feita pela fase.



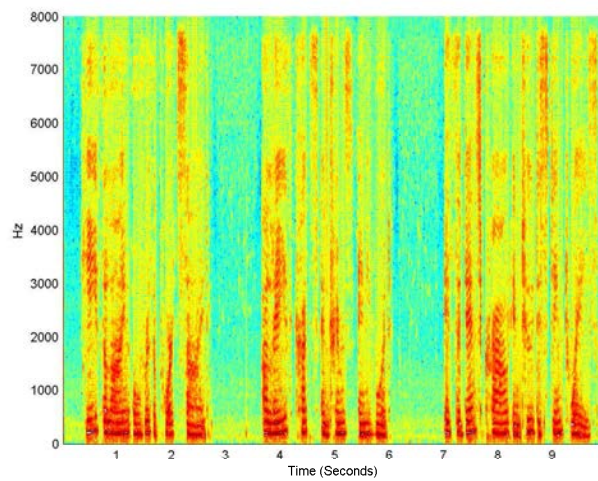
Source: Dados da pesquisa do autor.

Figura 20 – Espectrograma do sinal 4 com SNR de 5 dB processados pela SE e estimativa da SNR a priori obtida pela TSNR .

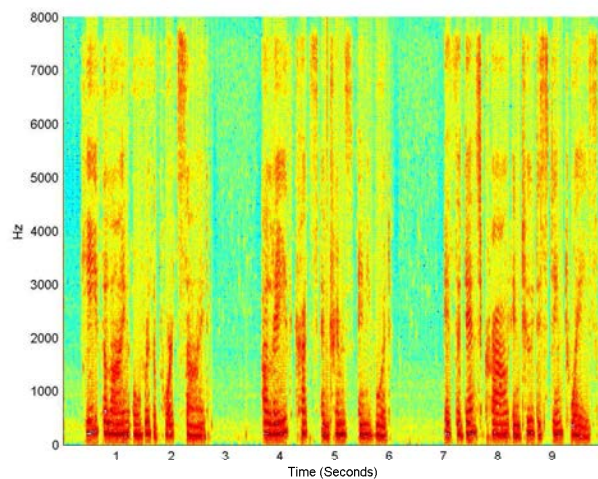
(a) Resultado da aplicação tradicional da SE



(b) Resultado da aplicação da SE juntamente da substituição da fase



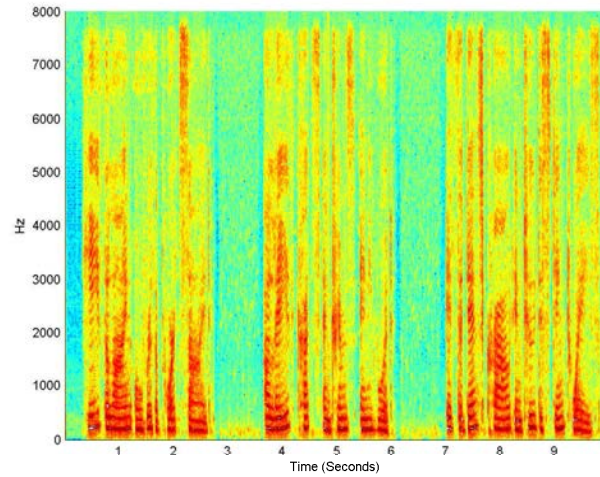
(c) Resultado da aplicação da SE com a correção feita pela fase.



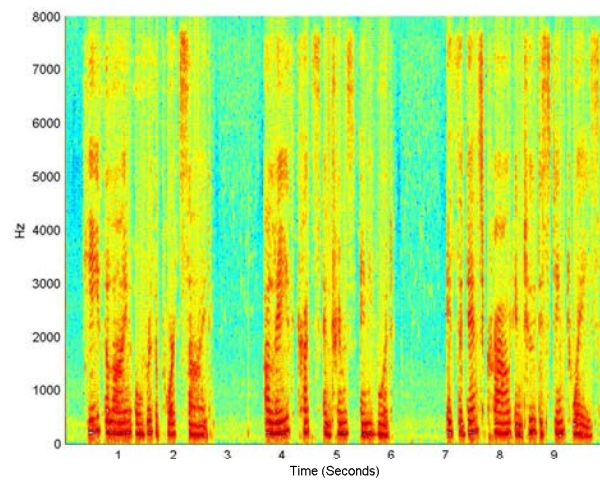
Source: Dados da pesquisa do autor.

Figura 21 – Espectrograma do sinal 4 com SNR de 15 dB processados pela SE e estimativa da SNR a priori obtida pela TSNR .

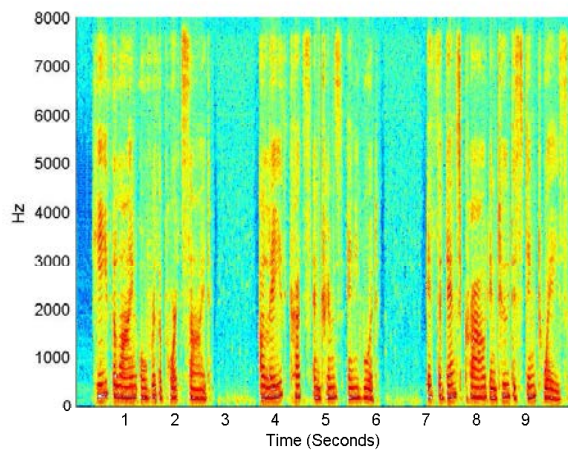
(a) Resultado da aplicação tradicional da SE



(b) Resultado da aplicação da SE juntamente da substituição da fase



(c) Resultado da aplicação da SE com a correção feita pela fase.



Source: Dados da pesquisa do autor.

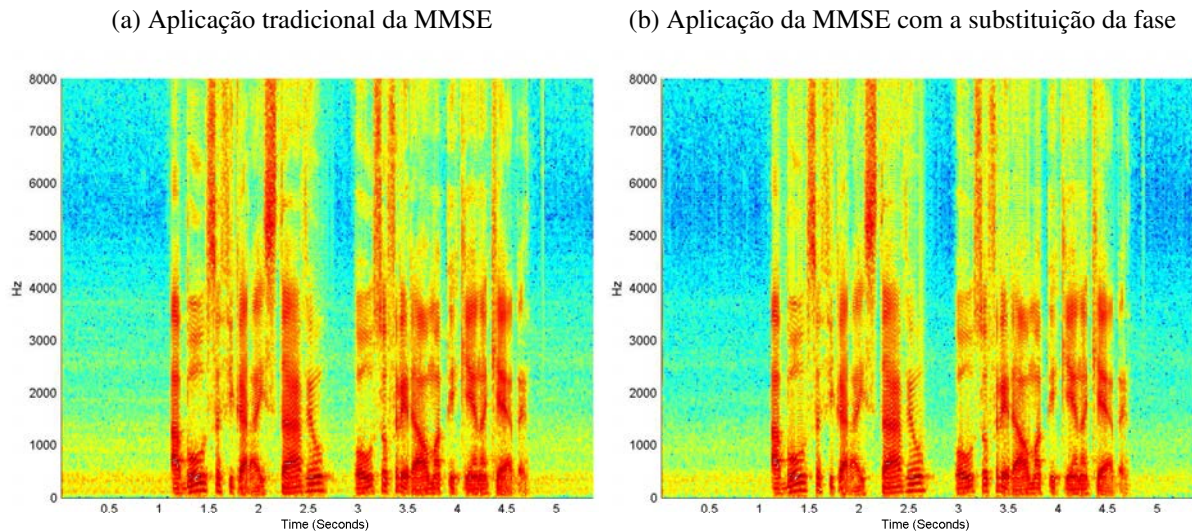
6.2.1.2 Resultados da Minimização do Erro Quadrático Médio

A técnica MMSE foi avaliada de forma similar à SE, utilizando todas as técnicas para estimação da SNR a priori. No entanto, somente a substituição da fase foi aplicada nela, tendo em vista que a correção proposta na seção 5.3 foi baseada no funcionamento da subtração espectral. Os espectrogramas das figuras 22 a 29 mostram o resultado da aplicação no sinal com relação sinal/ruído de 5dB.

Já na aplicação da MMSE observa-se que os resultados em geral foram similares, sendo que em todos os casos manteve-se uma forte interferência do ruído nas baixas frequências. De forma semelhante ao que foi feito com os resultados da SE, foram selecionados os melhores obtidos, novamente o uso da TSNR para estimativa da SNR a priori conseguiu os melhores resultados, indicando mais uma vez que esta técnica é a mais eficiente dentre as que foram avaliadas.

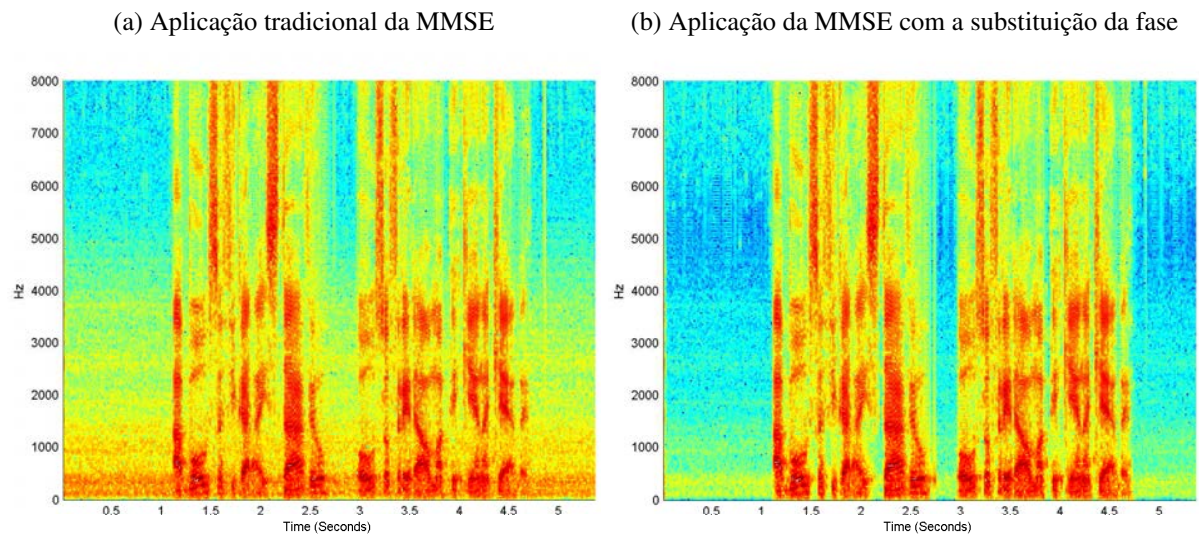
Pode-se observar que os resultados da aplicação da MMSE tradicional mantiveram ainda forte interferência do ruído, aparecendo ainda em grande quantidade nas baixas frequências, a implementação da fase possibilitou resultados com menos interferência, sendo inclusive reduzida nas baixas frequências, mostrando como a fase pode se tornar uma importante ferramenta na obtenção de melhores resultados em técnicas clássicas.

Figura 22 – Espectrograma do sinal 1 com SNR de 15 dB processados pela MMSE e estimativa da SNR a priori obtida pela TSNR .



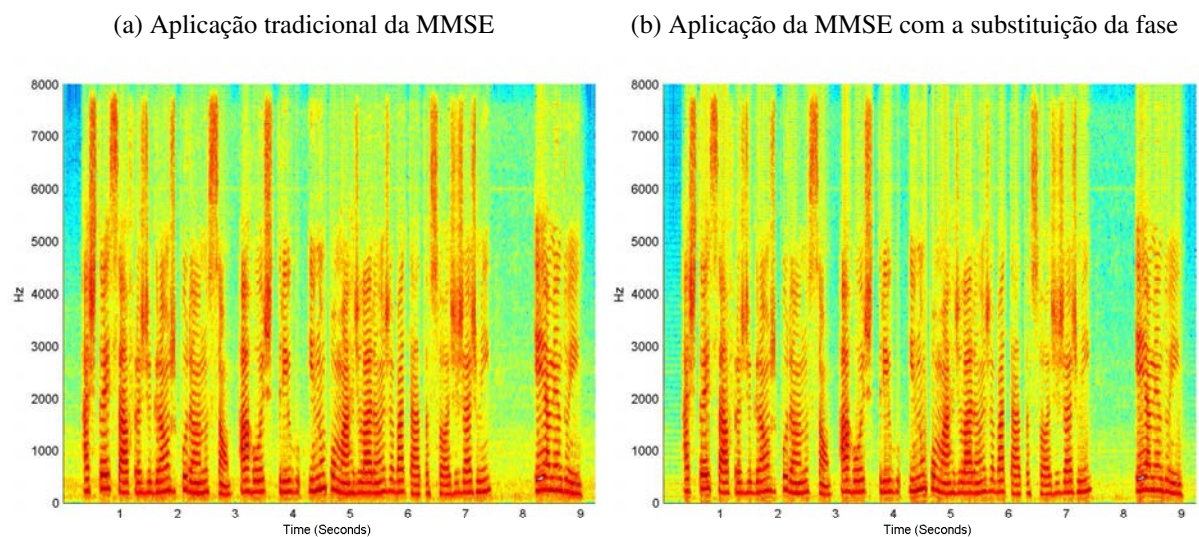
Source: Dados da pesquisa do autor.

Figura 23 – Espectrograma do sinal 1 com SNR de 5 dB processados pela MMSE e estimativa da SNR a priori obtida pela TSNR.



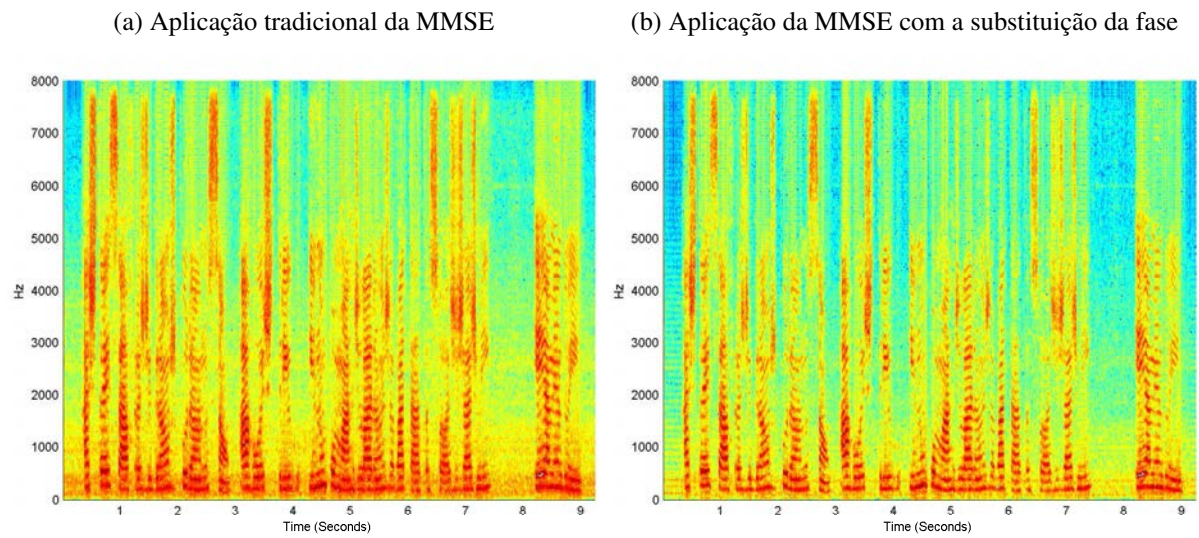
Source: Dados da pesquisa do autor.

Figura 24 – Espectrograma do sinal 2 com SNR de 15 dB processados pela MMSE e estimativa da SNR a priori obtida pela TSNR.



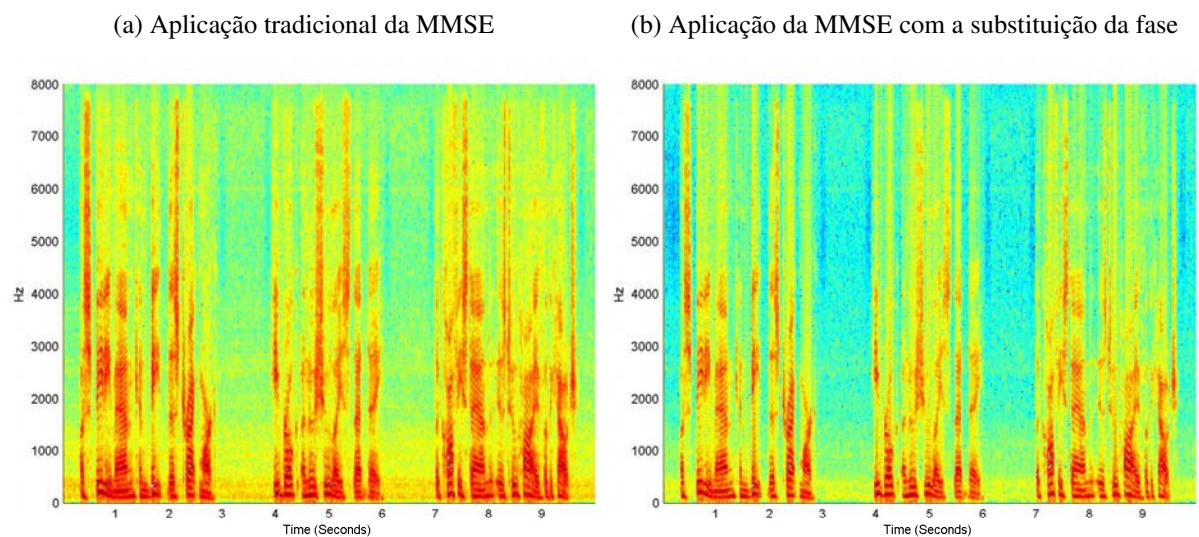
Source: Dados da pesquisa do autor.

Figura 25 – Espectrograma do sinal 2 com SNR de 5 dB processados pela MMSE e estimativa da SNR a priori obtida pela TSNR.



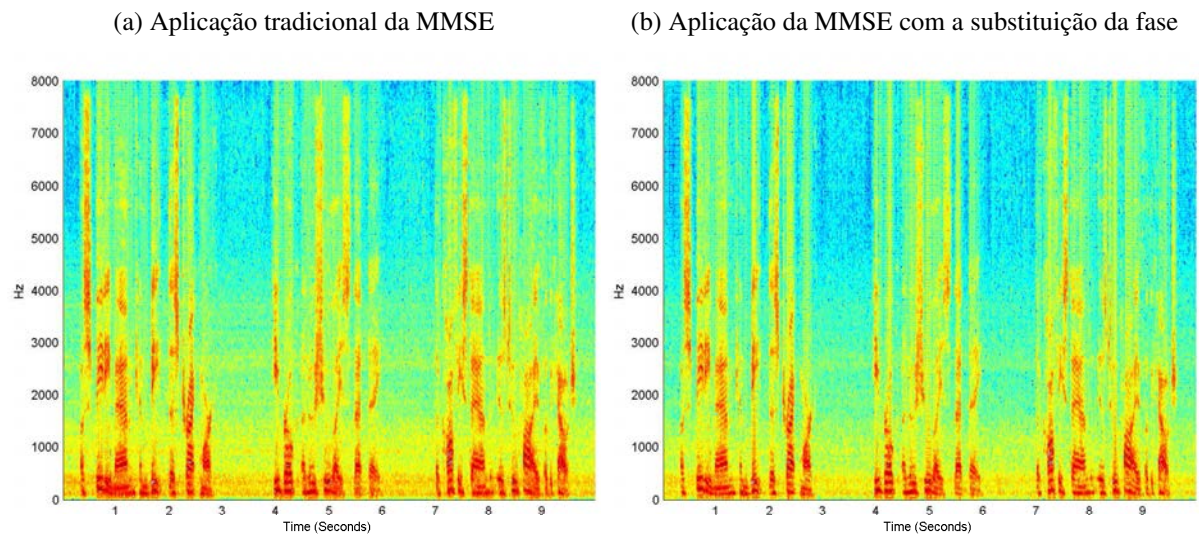
Source: Dados da pesquisa do autor.

Figura 26 – Espectrograma do sinal 3 com SNR de 15 dB processados pela MMSE e estimativa da SNR a priori obtida pela TSNR.



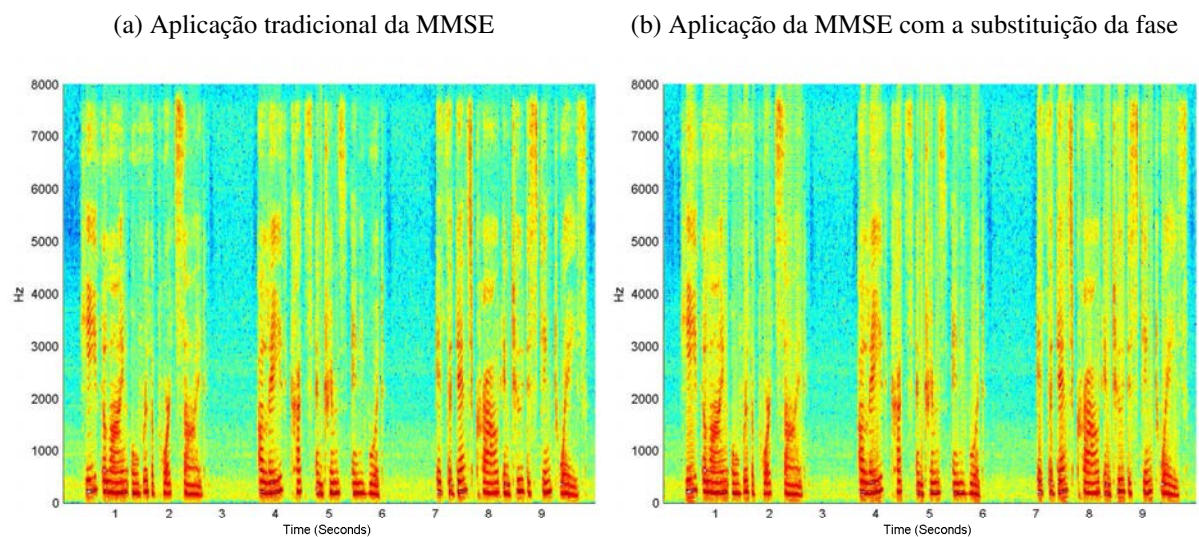
Source: Dados da pesquisa do autor.

Figura 27 – Espectrograma do sinal 3 com SNR de 5 dB processados pela MMSE e estimativa da SNR a priori obtida pela TSNR.



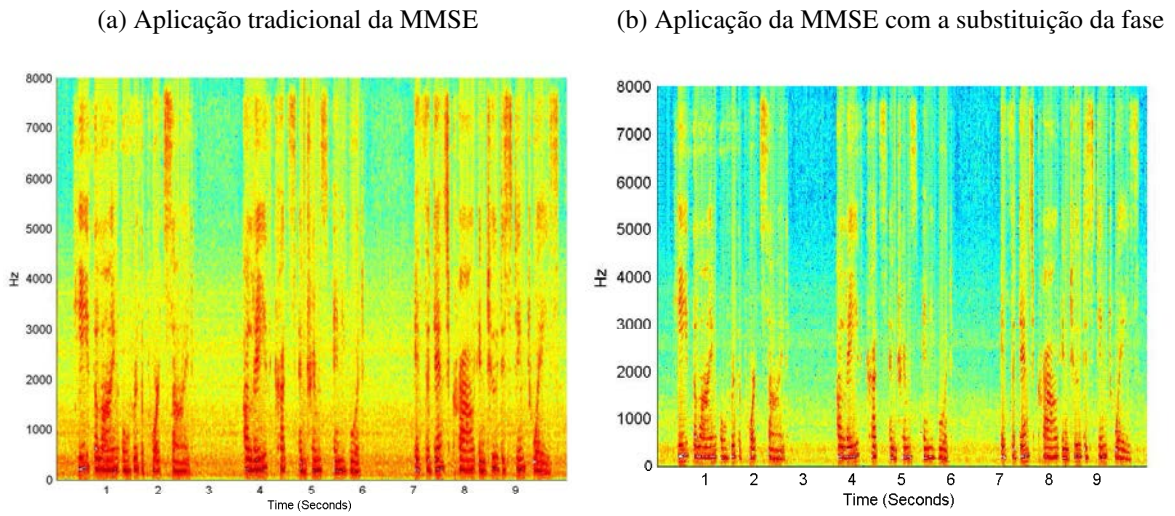
Source: Dados da pesquisa do autor.

Figura 28 – Espectrograma do sinal 4 com SNR de 15 dB processados pela MMSE e estimativa da SNR a priori obtida pela TSNR.



Source: Dados da pesquisa do autor.

Figura 29 – Espectrograma do sinal 4 com SNR de 5 dB processados pela MMSE e estimativa da SNR a priori obtida pela TSNR.



Source: Dados da pesquisa do autor.

6.2.2 Medição de Qualidade Objetiva

Para uma avaliação objetiva dos sinais processados, utilizou-se a medida objetiva PESQ (*Perceptual Evaluation of Speech Quality*), que é um padrão da União Internacional de Telecomunicações para avaliar a qualidade perceptível de sinais de voz. Os resultados são apresentados por meio de uma nota que varia de 0 (baixa qualidade) a 4.5 (livre de distorções) (INTERNATIONAL TELECOMMUNICATION UNION-ITU, 2013).

Nas tabelas que seguem, são apresentados os resultados obtidos pela PESQ na aplicação da SE com todas as formas de estimação da SNR a priori e com as variações de processamento utilizando a fase estudada anteriormente. Nas tabelas 1 a 8 são apresentados os resultados para aplicações da SE em todos os sinais com valores de SNR de 5 dB e 15 dB.

Tabela 1 – Resultados da PESQ para a SE aplicada ao sinal 1 com relação sinal/ruído de 5dB

PESQ	Sinal ruidoso	SE tradicional	SE + sub fase	SE + cor fase
DD	1,673	2,465	2,802	2,515
DD + alfa1	1,673	2,502	2,822	2,553
DD + alfa2	1,673	2,488	2,827	2,537
TSNR	1,673	2,696	2,909	2,771
MTSNR	1,673	2,430	2,725	2,469
Real	1,673	2,947	3,428	3,022

Source: Dados da pesquisa do autor.

Tabela 2 – Resultados da PESQ para a SE aplicada ao sinal 1 com relação sinal/ruído de 15dB

PESQ	Sinal ruidoso	SE tradicional	SE + sub fase	SE + cor fase
DD	2,426	2,878	3,201	2,916
DD + alfa1	2,426	2,904	3,247	2,947
DD + alfa2	2,426	2,889	3,229	2,920
TSNR	2,426	3,059	3,424	3,119
MTSNR	2,426	2,968	3,238	3,010
Real	2,426	3,148	3,556	3,216

Source: Dados da pesquisa do autor.

Tabela 3 – Resultados da PESQ para a SE aplicada ao sinal 2 com relação sinal/ruído de 5dB

PESQ	Sinal ruidoso	SE tradicional	SE + sub fase	SE + cor fase
DD	1,264	2,103	2,484	2,155
DD + alfa1	1,264	2,125	2,509	2,188
DD + alfa2	1,264	2,105	2,498	2,165
TSNR	1,264	2,325	2,593	2,409
MTSNR	1,264	1,998	2,442	2,037
Real	1,264	2,265	2,783	2,320

Source: Dados da pesquisa do autor.

Tabela 4 – Resultados da PESQ para a SE aplicada ao sinal 2 com relação sinal/ruído de 15dB

PESQ	Sinal ruidoso	SE tradicional	SE + sub fase	SE + cor fase
DD	2,306	2,766	3,059	2,837
DD + alfa1	2,306	2,789	3,064	2,866
DD + alfa2	2,306	2,767	3,062	2,839
TSNR	2,306	2,936	3,141	3,021
MTSNR	2,306	2,740	3,080	2,795
Real	2,306	2,949	3,253	3,029

Source: Dados da pesquisa do autor.

Tabela 5 – Resultados da PESQ para a SE aplicada ao sinal 3 com relação sinal/ruído de 5dB

PESQ	Sinal ruidoso	SE tradicional	SE + sub fase	SE + cor fase
DD	1,515	2,261	2,443	2,345
DD + alfa1	1,515	2,281	2,461	2,364
DD + alfa2	1,515	2,265	2,447	2,348
TSNR	1,515	2,512	2,617	2,642
MTSNR	1,515	2,286	2,539	2,335
Real	1,515	2,369	2,569	2,451

Source: Dados da pesquisa do autor.

Tabela 6 – Resultados da PESQ para a SE aplicada ao sinal 3 com relação sinal/ruído de 15dB

PESQ	Sinal ruidoso	SE tradicional	SE + sub fase	SE + cor fase
DD	2,325	2,676	2,778	2,747
DD + alfa1	2,325	2,694	2,802	2,770
DD + alfa2	2,325	2,678	2,781	2,751
TSNR	2,325	2,828	2,866	2,912
MTSNR	2,325	2,841	2,998	2,889
Real	2,325	2,771	2,870	2,850

Source: Dados da pesquisa do autor.

Tabela 7 – Resultados da PESQ para a SE aplicada ao sinal 4 com relação sinal/ruído de 5dB

PESQ	Sinal ruidoso	SE tradicional	SE + sub fase	SE + cor fase
DD	1,408	2,212	2,546	2,259
DD + alfa1	1,408	2,229	2,549	2,286
DD + alfa2	1,408	2,218	2,554	2,267
TSNR	1,408	2,415	2,619	2,485
MTSNR	1,408	2,132	2,532	2,152
Real	1,408	2,531	2,940	2,606

Source: Dados da pesquisa do autor.

Tabela 8 – Resultados da PESQ para a SE aplicada ao sinal 4 com relação sinal/ruído de 15dB

PESQ	Sinal ruidoso	SE tradicional	SE + sub fase	SE + cor fase
DD	2,376	2,789	2,973	2,861
DD + alfa1	2,376	2,821	2,986	2,900
DD + alfa2	2,376	2,789	2,974	2,864
TSNR	2,376	2,918	3,023	3,012
MTSNR	2,376	2,858	3,120	2,887
Real	2,376	2,987	3,123	3,089

Source: Dados da pesquisa do autor.

Observando os resultados, nota-se que em praticamente todos os casos a nota obtida com os diferentes métodos para estimação da SNR a priori foram superiores aos da decisão dirigida tradicional, onde em alguns casos a nota obtida utilizando-se o MTSNR foi inferior ao da decisão dirigida. Conclui-se também, que de forma semelhante ao que ocorreu com os espectrogramas, a aplicação da SE com a estimativa da SNR a priori feita pelo método TSNR obteve os melhores resultados, tanto na aplicação tradicional da SE, quanto nas aplicações utilizando a fase.

Avaliando os resultados obtidos no processamento utilizando a fase, em todos os sinais processados houve um incremento na nota, sendo que, os resultados da aplicação em sinais com relação sinal/ruído de 5 dB obtiveram um maior incremento na nota, obtendo uma melhora média de 13%, já os sinais com relação sinal/ruído de 15 dB obtiveram um incremento médio de 8%. De forma similar o processamento utilizando a correção baseada na fase obteve melhores resultados nas implementações dos sinais mais ruidosos, em média, um resultado 3% superior ao da aplicação tradicional, de forma que a aplicação nos sinais menos ruidosos conseguiram um incremento de 2% em média.

Comparando as notas dos métodos que utilizam estimativa da SNR a priori com a aplicação feita a partir dos valores reais, nota-se no caso da aplicação onde o sinal possuía uma relação sinal/ruído de 5dB, o resultado da aplicação da SNR real conseguiu uma nota substancialmente superior aos métodos estimados, sendo que, o alto nível de ruído provavelmente causou grande instabilidade no momento da estimativa. Já os resultados da aplicação em 15dB, mostram que houve uma menor variação entre os resultados obtidos por meio de estimativas e o resultado dos valores reais.

Nas tabelas a seguir, tem-se os resultados do processamento de todos os sinais através da MMSE com todos os métodos para estimação da SNR a priori estudados e a aplicação da MMSE juntamente da substituição da fase, onde, nas tabelas tabelas 10 a 16 são apresentados os resultados para aplicações da MMSE em todos os sinais com valores de SNR de 5 dB e 15 dB.

Tabela 9 – Resultados da PESQ para a MMSE aplicada ao sinal 1 com relação sinal/ruído de 5dB

PESQ	Sinal ruidoso	MMSE tradicional	MMSE+ sub fase
DD	1,673	2,540	2,817
DD + alfa1	1,673	2,582	2,864
DD + alfa2	1,673	2,561	2,844
TSNR	1,673	2,572	2,782
MTSNR	1,673	2,478	2,734
Real	1,673	2,699	3,026

Source: Dados da pesquisa do autor.

Tabela 10 – Resultados da PESQ para a MMSE aplicada ao sinal 1 com relação sinal/ruído de 15dB

PESQ	Sinal ruidoso	MMSE tradicional	MMSE + sub fase
DD	2,426	2,911	3,123
DD + alfa1	2,426	2,943	3,179
DD + alfa2	2,426	2,926	3,142
TSNR	2,426	2,955	3,148
MTSNR	2,426	3,049	3,161
Real	2,426	3,102	3,266

Source: Dados da pesquisa do autor.

Tabela 11 – Resultados da PESQ para a MMSE aplicada ao sinal 2 com relação sinal/ruído de 5dB

PESQ	Sinal ruidoso	MMSE tradicional	MMSE+ sub fase
DD	1,264	2,176	2,424
DD + alfa1	1,264	2,195	2,472
DD + alfa2	1,264	2,187	2,446
TSNR	1,264	2,286	2,449
MTSNR	1,264	2,000	2,326
Real	1,264	2,406	2,744

Source: Dados da pesquisa do autor.

Tabela 12 – Resultados da PESQ para a MMSE aplicada ao sinal 2 com relação sinal/ruído de 15dB

PESQ	Sinal ruidoso	MMSE tradicional	MMSE+ sub fase
DD	2,306	2,843	2,873
DD + alfa1	2,306	2,867	2,898
DD + alfa2	2,306	2,846	2,889
TSNR	2,306	2,945	3,015
MTSNR	2,306	2,768	2,795
Real	2,306	2,957	3,153

Source: Dados da pesquisa do autor.

Tabela 13 – Resultados da PESQ para a MMSE aplicada ao sinal 3 com relação sinal/ruído de 5dB

PESQ	Sinal ruidoso	MMSE tradicional	MMSE+ sub fase
DD	1,515	2,304	2,602
DD + alfa1	1,515	2,328	2,645
DD + alfa2	1,515	2,309	2,620
TSNR	1,515	2,354	2,606
MTSNR	1,515	2,334	2,541
Real	1,515	2,705	2,823

Source: Dados da pesquisa do autor.

Tabela 14 – Resultados da PESQ para a MMSE aplicada ao sinal 3 com relação sinal/ruído de 15dB

PESQ	Sinal ruidoso	MMSE tradicional	MMSE+ sub fase
DD	2,325	2,702	2,925
DD + alfa1	2,325	2,732	2,947
DD + alfa2	2,325	2,714	2,934
TSNR	2,325	2,746	2,962
MTSNR	2,325	2,938	2,918
Real	2,325	3,013	3,347

Source: Dados da pesquisa do autor.

Tabela 15 – Resultados da PESQ para a MMSE aplicada ao sinal 4 com relação sinal/ruído de 5dB

PESQ	Sinal ruidoso	MMSE tradicional	MMSE+ sub fase
DD	1,408	2,234	2,408
DD + alfa1	1,408	2,269	2,482
DD + alfa2	1,408	2,258	2,446
TSNR	1,408	2,326	2,435
MTSNR	1,408	2,120	2,368
Real	1,408	2,403	2,789

Source: Dados da pesquisa do autor.

Tabela 16 – Resultados da PESQ para a MMSE aplicada ao sinal 4 com relação sinal/ruído de 15dB

PESQ	Sinal ruidoso	MMSE tradicional	MMSE+ sub fase
DD	2,376	2,823	2,878
DD + alfa1	2,376	2,859	2,899
DD + alfa2	2,376	2,837	2,894
TSNR	2,376	2,884	2,977
MTSNR	2,376	2,840	2,911
Real	2,376	2,905	3,084

Source: Dados da pesquisa do autor.

Avaliando-se as notas obtidas, percebe-se que, de forma semelhante ao que ocorreu com a SE, os resultados dos diferentes métodos para estimação da SNR a priori conseguiram melhores notas do que o método decisão dirigida tradicional, sendo que apenas a MTSNR obteve notas inferiores em algumas aplicações. A substituição da fase apresentou, de forma similar ao obtido pela SE, um melhor rendimento em sinais com relação sinal/ruído de 5 dB, obtendo um incremento médio de 11%. Nas aplicações em sinais com relação sinal/ruído de 15 dB o incremento médio apresentado foi de 5%. Comparando as notas das estimativas com a

nota da aplicação com valores reais, nota-se que ocorreu uma menor variação, tanto na aplicação com um nível maior de ruído quanto com o nível menor, o que provavelmente se deve a uma menor sensibilidade da MMSE às variações no ruído.

Observando as notas da SE e da MMSE, pode-se concluir que as notas da MMSE foram superiores, com exceção de alguns resultados da SE com SNR a priori estimada pela TSNR que obteve nota superior tanto na aplicação com 5dB como na aplicação em 15dB e da aplicação da SE no sinal 4, onde a aplicação com a substituição da fase obteve notas superiores.

Os resultados apresentados neste trabalho são importantes e mostram que as técnicas de redução de ruído em sinais de voz ainda podem melhorar. Tentar integrar uma metodologia para estimação da fase e melhorar a estimação da relação sinal/ruído a priori são pontos importantes para trabalhos futuros.

7 CONCLUSÕES

Neste trabalho estudou-se técnicas clássicas para redução de ruídos em sinais de voz, bem como diferentes técnicas para estimação da SNR a priori, um dos principais parâmetros empregados nas técnicas estudadas.

Os estudos envolveram duas técnicas clássicas de redução de ruído: o método MMSE e a SE. Aplicados em sua forma tradicional, os métodos apenas trabalham efetuando a correção da magnitude do sinal em questão, mantendo-se a fase ruidosa. Embora este procedimento seja amplamente aceito, em função da aleatoriedade da fase ser imperceptível ao ouvido humano, estudos recentes mostram que a fase ruidosa do sinal pode ser responsável por resultados inferiores ao fim do processamento.

Dessa forma, foram feitos testes utilizando técnicas onde a fase dos sinais são consideradas. Embora a fase do sinal não esteja disponível em aplicação real, os resultados são válidos pois, através deles foi possível verificar a real influência da fase nas técnicas de redução de ruído. Os testes foram realizados de duas formas diferentes: no primeiro teste, que foi aplicado tanto na SE quanto na MMSE, a fase ruidosa que estava presente no sinal após o processamento era substituída pela fase limpa que foi extraída do sinal antes de ser contaminado pelo ruído aditivo; já no segundo teste, apenas o método da SE foi analisado. Neste caso, foram utilizadas as fase do sinal limpo e do ruído, de forma a se realizar uma comparação entre as duas, avaliando-se os instantes em que o sinal ruidoso é obtido através de uma subtração das magnitudes do sinal limpo e do ruído.

Outro estudo importante considerou diferentes formas de estimação da SNR a priori. Como foi possível avaliar através dos testes que envolveram tanto valores estimados para a SNR a priori, quanto os valores reais, como já esperado o resultado da implementação onde utilizou-se valores reais é superior aos demais, no entanto, essa diferença se mostra maior nos casos onde o nível de ruído no sinal é maior, indicando que a estimativa da SNR pode sofrer maior instabilidade em sinais muito ruidosos. Dentre as técnicas avaliadas, a técnica da TSNR se mostrou eficiente na estimação mais correta da SNR a priori. Uma combinação dessas técnicas com os estudos da fase é um futuro promissor na área de redução de ruído em sinais de voz.

REFERÊNCIAS

- BOLL, S. A spectral subtraction algorithm for suppression of acoustic noise in speech. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING - ICASSP, 3., 1979, Washington-DC. *Proceedings...* Piscataway: IEEE, 1979. v. 4, p. 200–203. Citado 2 vezes nas páginas 17 e 19.
- BOLL, S. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech and Signal Processing*, Piscataway, v. 27, n. 2, p. 113–120, abr. 1979. ISSN 0096-3518. Citado 6 vezes nas páginas 14, 15, 17, 19, 28 e 38.
- CAPPE, O. Elimination of the musical noise phenomenon with the ephraim and malah noise suppressor. *IEEE Transactions on Speech and Audio Processing*, Piscataway, v. 2, n. 2, p. 345–349, abr. 1994. ISSN 1063-6676. Citado 3 vezes nas páginas 20, 22 e 29.
- DELLER, J.; HANSEN, J.; PROAKIS, J. *Discrete-time processing of speech signals*. Upper Saddle River: Prentice Hall, 1993. ISBN 0023283017. Citado na página 14.
- EPHRAIM, Y. A bayesian estimation approach for speech enhancement using hidden markov models. *IEEE Transactions on Signal Processing*, Piscataway, v. 40, n. 4, p. 725–735, 1992. ISSN 1053-587X. Citado na página 14.
- EPHRAIM, Y. Statistical-model-based speech enhancement systems. *Proceedings of the IEEE*, Piscataway, v. 80, n. 10, p. 1526–1555, out. 1992. ISSN 0018-9219. Citado na página 20.
- EPHRAIM, Y.; MALAH, D. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech and Signal Processing*, Piscataway, v. 32, n. 6, p. 1109–1121, dez. 1984. ISSN 0096-3518. Citado 9 vezes nas páginas 14, 15, 17, 20, 21, 27, 28, 34 e 38.
- FARDKHALEGHI, P.; SAVOJI, M. New approaches to speech enhancement using phase correction in wiener filtering. In: INTERNATIONAL SYMPOSIUM ON TELECOMMUNICATIONS, 5., 2010, Tehran. *Proceedings...* Piscataway: IEEE, 2010. p. 895–899. Citado na página 15.
- GERKMANN, T.; KRAWCZYK, M.; REHR, R. Phase estimation in speech enhancement: Unimportant, important, or impossible? In: CONVENTION OF ELECTRICAL & ELECTRONICS ENGINEERS IN ISRAEL (IEEEI), 27., 2012, Eilat. *Proceedings...* Piscataway: IEEE, 2012. p. 5. Citado 3 vezes nas páginas 14, 19 e 33.
- HASAN, M.; SALAHUDDIN, S.; KHAN, M. A modified a priori snr for speech enhancement using spectral subtraction rules. *IEEE Signal Processing Letters*, Piscataway, v. 11, n. 4, p. 450–453, abr. 2004. ISSN 1070-9908. Citado 2 vezes nas páginas 27 e 29.
- INTERNATIONAL TELECOMMUNICATION UNION-ITU. P.862: perceptual evaluation of speech quality (PESQ): an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. Geneva, 2013. Disponível em: <<http://www.itu.int/rec/T-REC-P.862/>>. Acesso em: 13 de maio de 2013. Citado na página 57.
- KRAWCZYK, M.; GERKMANN, T. Stft phase improvement for single channel speech enhancement. In: INTERNATIONAL WORKSHOP ON ACOUSTIC SIGNAL

ENHANCEMENT - IWAENC, 2012, Aachen. *Proceedings...* Aachen: IWAENC, 2012. p. 4. Citado 4 vezes nas páginas 14, 19, 33 e 34.

LANGARANI, M.; VEISI, H.; SAMETI, H. The effect of phase information in speech enhancement and speech recognition. In: INTERNATIONAL CONFERENCE ON INFORMATION SCIENCE, SIGNAL PROCESSING AND THEIR APPLICATIONS - ISSPA, 11., 2012, Montreal. *Proceedings...* Montreal: ISSPA, 2012. p. 1446–1447. Citado na página 15.

MCAULAY, R.; MALPASS, M. Speech enhancement using a soft-decision noise suppression filter. *IEEE Transactions on Acoustics, Speech and Signal Processing*, Piscataway, v. 28, n. 2, p. 137–145, abr. 1980. ISSN 0096-3518. Citado 2 vezes nas páginas 21 e 22.

OPPENHEIM, A.; SCHAFER, R.; BUCK, J. *Discrete-time signal processing*. 2. ed. Upper Saddle River: Prentice Hall, 1999. (Prentice-Hall signal processing series). ISBN 9780137549207. Citado 3 vezes nas páginas 17, 18 e 24.

OU, S.; ZHAO, X.; GAO, Y. Speech enhancement employing modified a priori snr estimation. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING - ICASSP, 2007, Honolulu. *Proceedings...* Piscataway: IEEE, 2007. v. 3, p. 827–831. Citado 2 vezes nas páginas 27 e 29.

PALIWAL, K.; WÓJCICKI, K.; SHANNON, B. The importance of phase in speech enhancement. *Speech Communication*, Amsterdam, v. 53, n. 4, p. 465–494, 2011. ISSN 0167-6393. Citado 3 vezes nas páginas 14, 19 e 33.

PARK, Y.-S.; CHANG, J.-H. A novel approach to a robust a priori snr estimator in speech enhancement. *IEICE Transactions on Communications*, Tokyo, E-90, n. 8, p. 2182–2185, ago. 2007. Citado 2 vezes nas páginas 27 e 30.

PLAPOUS, C.; MARRO, C.; MAUURY, L.; SCALART, P. A two-step noise reduction technique. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, 2004, Montreal. *Proceedings...* Piscataway: IEEE, 2004. v. 1, p. I. 289–I.292. ISSN 1520-6149. Citado 2 vezes nas páginas 27 e 28.

PLAPOUS, C.; MARRO, C.; SCALART, P. Improved signal-to-noise ratio estimation for speech enhancement. *IEEE Transactions on Audio, Speech, and Language Processing*, Piscataway, v. 14, n. 6, p. 2098–2108, nov. 2006. ISSN 1558-7916. Citado 2 vezes nas páginas 27 e 28.

SCALART, P.; VIEIRA FILHO, J. Speech enhancement based on a priori signal to noise estimation. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, 1996, Atlanta. *Proceedings...* Piscataway: IEEE, 1996. v. 2, p. 629–632. ISSN 1520-6149. Citado 2 vezes nas páginas 23 e 34.

SMITH, S. *Digital signal processing: a practical guide for engineers and scientists*. Amsterdam: Newnes, 2003. 650 p. (Demystifying Technology Series). ISBN 9780750674447. Citado na página 25.

VIRAG, N. Single channel speech enhancement based on masking properties of the human auditory system. *IEEE Transactions on Speech and Audio Processing*, Piscataway, v. 7, n. 2, p. 126–137, 1999. ISSN 1063-6676. Citado na página 15.