

Estimação Robusta do Desvio-Padrão de Ruídos Acústicos Aplicada ao Realce de Sinais de Voz

R. Tavares e R. Coelho

Resumo— Este trabalho apresenta um método de realce de sinal de voz, realizado no domínio do tempo, com o uso de um algoritmo de estimação robusta de desvio padrão. A técnica foi avaliada em três diferentes situações, uma utilizando somente o método e as duas outras como uma proposta de pós-processamento. As técnicas utilizadas para o tratamento inicial aplicado aos sinais de voz corrompidos, no segundo experimento, são a SS (spectral subtraction) e o IMCRA (improved minima controlled recursive averaging). Os resultados apontam que o método apresenta um ganho de SegSNR em relação ao IMCRA e SS quando comparados individualmente. Já nos experimentos de pós-processamento o uso do método em sinais de voz tratados por SS apresentou ganhos de SegSNR, o que não ocorreu quando esse foi aplicado ao método IMCRA.

Palavras-Chave— Realce de sinais, estimação robusta, estimador d-dimensional Trimmed

Abstract— This paper presents a method of speech enhancement performed in the time domain using a technique robust standard deviation. The technique was evaluated from three experiments, one using only the method and the other two as a proposal for post processing. The techniques used for the initial treatment applied to speech signals corrupted in the second experiment are SS (spectral subtraction) and IMCRA (improved minima controlled recursive averaging). The results presented by the gain of SegSNR indicate that the method provides a gain compared to IMCRA and SS compared individually. Already in experiment post-processing using the method in speech signals processed by SS showed gains, which did not occur as this method was applied to IMCRA.

Keywords— speech enhancement, robust estimation, D-dimensional trimmed estimator

I. INTRODUÇÃO

O aprimoramento ou realce de sinais de voz [1] [2] é importante para diversos campos de pesquisa. Os ruídos acústicos aditivos podem causar distorções dificultando o desempenho de aplicações como, por exemplo, o reconhecimento de voz ou locutor [3] e a supressão de ruído em aparelhos auditivos [4]. Os ruídos sonoros podem ser originados por máquinas funcionando, som de pessoas conversando, tons de aparelhos celulares ou tráfego intenso de veículos.

Os métodos de realce de sinais de voz são geralmente baseados na atenuação ou ajuste na densidade espectral de potência em períodos curtos. Um dos grandes desafios é a estimação precisa das estatísticas espectrais do ruído acústico adicionado ao sinal de voz. Para esta, adota-se comumente o uso de um algoritmo de detecção de atividade de voz (VAD - *voice activity detector*). Assim, as estimativas das

estatísticas do ruído, considerado estacionário, são obtidas nos períodos de ausência de voz. Todavia, os ruídos acústicos reais apresentam características não-estacionárias [1] [2]. Na literatura, a subtração espectral (SS - *spectral subtraction*) [5] e a estimativa de erro mínimo quadrado de amplitude do espectro de tempo curto (MMSE-STSA - *minimum mean-square error short-time spectral amplitude estimator*) [6] são os principais métodos utilizados para realce de sinais estacionários.

Em [7], Cohen apresentou a técnica IMCRA (*improved minima controlled recursive averaging*) para a atualização da estimativa do espectro de potência do ruído. O autor demonstra que os valores mínimos de potência obtidos considerando-se um certo número de quadros (≈ 120 quadros [2]), permite uma definição melhor das estatísticas do ruído. No entanto, a grande limitação desta proposta é que, devido à dependência de uma quantidade grande de quadros, a estimação torna-se imprecisa para ruídos não-estacionários com bruscas variações espectrais [2].

Este trabalho apresenta uma proposta de realce de sinais de voz no domínio do tempo, para ruídos não-estacionários. A técnica adota o algoritmo de estimação robusta de desvio padrão de ruídos (DATE - *D-Dimensional Trimmed Estimator*) apresentado em [8]. No método aqui proposto, o teste de limitante (*thresholding tests*) é utilizado para determinar as regiões com maior probabilidade de ausência de sinal de voz para a extração de amostras do ruído, e assim, alcançar o realce do sinal de voz. A técnica proposta é também avaliada quando aplicada após um pré-processamento com o uso dos SS e IMCRA. Para os experimentos foram considerados quatro ruídos acústicos reais (Avião, Balbúrdia, Fábrica e Navio), um ruído Branco artificial, e diferentes valores de razão sinal-ruído (SNR - *signal-to-noise-ratio*).

O restante deste trabalho está organizado da seguinte forma. A seção II introduz conceitos básicos de estimação robusta e um resumo do estimador de desvio padrão apresentado por [8]. Na seção III é apresentada a técnica de realce de sinais proposta neste trabalho. A descrição dos experimentos realizados são abordados na seção IV. Os resultados são descritos na seção V. Finalmente na seção VI são apresentadas as principais conclusões deste trabalho.

II. MÉTODOS DE ESTIMAÇÃO ROBUSTA DE DESVIO PADRÃO DE RUÍDOS

Os métodos de estimação robusta [9] são aplicados em situações onde é necessária a utilização de estimadores que não sejam indevidamente afetados por valores discrepantes (*outliers*). Dentre os principais métodos destacam-se as combinações lineares (L) de ordens estatísticas [10]. Os

estimadores-L são muito populares por apresentarem uma baixa complexidade computacional e um alto grau de robustez na presença de *outliers*.

Um importante exemplo de estimador-L robusto é o denominado aparador (estimador-T - *trimmed estimator*). Sendo este obtido a partir da exclusão de um determinado número de valores extremos. Geralmente, isto é feito para se obter uma estatística mais precisa. Estimadores-T também têm, frequentemente, maior eficiência para lidar com distribuições misturadas e distribuições com caudas pesadas.

A. O estimador-T

Sejam Y_1, Y_2, \dots, Y_n valores de uma variável aleatória e $Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(n)}$ as observações classificadas em ordem crescente. O i -ésimo maior valor de $Y_{(i)}$ é chamado de i -ésima-ordem estatística. A variância da amostra aparada (R) é calculada removendo-se os k -maiores e os k -menores valores e efetuando o cálculo a partir dos valores restantes,

$$R = \sqrt{\frac{b}{N-2k} \sum_{j=k}^{N-k} \left(Y_{(j)} - \frac{1}{N-2k} \sum_{i=k}^{N-k} Y_i \right)^2} \quad (1)$$

onde, $k = \lceil \gamma n \rceil$ e γ representa a fração de exclusão de *outliers*, $0 < \gamma < 0,5$, a constante b depende do modelo de distribuição e dos valores de γ .

A exclusão de valores extremos se torna mais precisa quando a magnitude dos *outliers* é muito alta e o número destes é conhecido *a priori*. Porém, o desempenho destes caem significativamente quando a proporção de *outliers* aumenta de forma abrupta. Devido a esta limitação, foi desenvolvido em [8] um estimador robusto de desvio padrão, que não requer montantes fixos de amostras de valores extremos para remover, ou o conhecimento *a priori* da quantidade de *outliers* contidos na amostra e a sua distribuição.

B. D-Dimensional Trimmed Estimator (DATE)

O DATE é uma derivação do estimador-T [8]. As principais diferenças entre DATE e o estimador-T convencional são: aplicado a vetores d -dimensionais, enquanto os estimadores convencionais estão limitados a uma única dimensão. A segunda diferença é que o estimador-T convencional necessita do conhecimento *a priori* do montante de discrepantes (*outliers*) para efetuar a exclusão, enquanto o DATE parte de duas premissas estabelecidas *a priori* que assume que as normas de sinal estão acima de um limitante (τ) e que a probabilidade de ocorrência de sinais abaixo deste limitante é menor que 50% [8]. A partir destas informações são excluídos os *outliers*. Quanto ao modelo de observação, o sinal é de presença e distribuição desconhecida e está adicionado a um ruído branco Gaussiano independente (AWGN - *additive white Gaussian noise*) e pode ser modelado por

$$Y_n = \varepsilon_n \Lambda_n + X_n \quad (2)$$

onde o Y_n representa a sequência de observações, $\varepsilon_n \in [0; 1]$ é uma variável aleatória que indica se o sinal de voz limpa Λ_n , com distribuição desconhecida, está adicionado ou não ao

ruído branco X_n que é Gaussiano com média $\mu = 0$ e desvio padrão σ_0 deverá ser estimado a partir das observações.

$$M_{\{x_1, x_2, \dots, x_N\}}(\tau) = \frac{\sum_{n=1}^N Y_n^r I(Y_n \tau)}{\sum_{n=1}^N Y_n^s I(Y_n \tau)} \quad (3)$$

O estimador DATE, $M_{\{x_1, x_2, \dots, x_N\}}(\tau)$, pode ser entendido como um método de momentos estatísticos oriundo da razão entre a r -ésima e a s -ésima ordem dos momentos das amostras de N sendo $r > s \geq 0$ e utilizando um limitante τ para selecionar as amostras que serão utilizadas no cálculo do desvio padrão. O limitante τ é uma razão de verossimilhança entre uma distribuição Gamma e uma distribuição hipergeométrica generalizada. Todas as amostras do sinal são comparadas a esta razão e as que ficam acima do limiar são excluídas para o cálculo final do desvio padrão.

Neste trabalho, esta detecção robusta considera sinais de voz com presença e distribuições desconhecidas como *outlier*, e é utilizado como uma proposta alternativa ao VAD. O algoritmo de detecção de *outliers*, contido no DATE, foi modificado, para ao invés de detectar a presença da voz, determinar no domínio do tempo, regiões onde somente a presença do ruído é verificada. Além disso, o desvio padrão será calculado utilizando o método, para que seja aplicado a subtração deste ao sinal de voz corrompido.

III. TÉCNICAS DE REALCE DE SINAL DE VOZ

As técnicas de realce de voz existentes na literatura, geralmente utilizam a detecção das regiões com ausência de sinal de voz para estimar o espectro de potência dos ruídos nestes intervalos. Este procedimento é realizado no domínio da frequência, onde o sinal é decomposto em módulo e fase. No caso da subtração espectral, por exemplo, a informação de fase é armazenada na memória e utilizada para compor o sinal de voz realçado juntamente com o módulo estimado das componentes em frequência.

Nesta seção são apresentadas inicialmente as técnicas SS e IMCRA, que são utilizadas para análise comparativa e como pré-processamento para a proposta de realce dos sinais de voz neste trabalho.

A. Subtração Espectral

O objetivo da subtração espectral [5] é a supressão de ruído aditivo a partir de um sinal corrompido, neste caso, um sinal de voz. O sinal de voz corrompido por um ruído aditivo $r(t)$ pode ser representado por

$$d(t) = s(t) + r(t) \quad (4)$$

em que $d(t)$ representa a voz $s(t)$ degradada ou corrompida. A partir da transformada discreta de Fourier (DFT - *discrete Fourier transform*) em janelas tipicamente na ordem de 20-40 ms, obtém-se

$$D(j, l) = S(j, l) + R(j, l) \quad (5)$$

onde j é o índice da sub-banda de frequência e l é o índice da duração do quadro. Uma estimativa da magnitude da voz original limpa $|\hat{S}(j, l)|$ é então obtida pela subtração das

componentes em frequência do ruído estimado $|\hat{R}(j, l)|$ a partir do espectro de potência do sinal corrompido:

$$|\hat{S}(j, l)| = |D(j, l)| - |\hat{R}(j, l)|. \quad (6)$$

A estimativa do ruído utilizado em (6) é convencionalmente obtida no domínio da frequência durante os intervalos onde não há presença de voz. Para tal, a magnitude do espectro de frequência de curto prazo é obtida por

$$|\hat{R}(j, l)| = \frac{1}{|\mathcal{L}|} \sum_{l \in \mathcal{L}} |D(j, l)|, \quad (7)$$

onde \mathcal{L} é um conjunto de quadros consecutivos nos quais a voz está ausente, e $|\mathcal{L}|$ é a sua cardinalidade.

Conforme (6), o pressuposto do método SS é que a redução de ruído seja alcançada suprimindo o seu efeito na magnitude dos espectros. Para a reconstrução do sinal de voz, ou seja, a sua representação no domínio do tempo, uma estimativa complexa (magnitude e fase) de $\hat{S}(j, l)$ é necessária. Esta é alcançada através da combinação da amplitude obtida em (6) com a fase $\theta_D(j, l)$ do espectro do sinal corrompido,

$$\hat{S}(j, l) = |\hat{S}(j, l)|e^{j\theta_D(j, l)}. \quad (8)$$

B. IMCRA

Assim como na subtração espectral, a técnica IMCRA [7] também assume a relação aditiva entre as transformadas de Fourier de tempo curto (STFT - *short-time Fourier transform*) apresentadas em (6). Diferentemente da SS, o ruído é estimado pela média recursiva dos valores de potência espectral do passado da medição do sinal ruidoso, durante os períodos de ausência de voz. Para isto, é utilizado um parâmetro de incerteza da presença de voz. Seja $p(j, l)$ a probabilidade condicional de presença da voz em cada quadro l e em cada componente de frequência j , uma média recursiva pode ser obtida pela expressão

$$\bar{\lambda}_r(j, l+1) = \tilde{\alpha}_d \bar{\lambda}_r(j, l) + [1 - \tilde{\alpha}_d(j, l)] |D(j, l)|^2, \quad (9)$$

onde,

$$\tilde{\alpha}_d(j, l) \triangleq \alpha_d + (1 - \alpha_d) p(j, l) \quad (10)$$

é o parâmetro variável no tempo da suavização da frequência dependente e α_d ($0 < \alpha_d < 1$) é um fator de suavização. Através da introdução de fator de compensação de viés β , a estimativa do ruído é dada por

$$\hat{\lambda}_r(j, l) = \beta * \bar{\lambda}_r(j, l). \quad (11)$$

A fim de calcular a probabilidade da presença de voz, duas iterações de suavização e acompanhamento mínimo são realizadas. A primeira iteração é realizada por uma média recursiva de primeira ordem como

$$S(j, l) = \alpha_s S(j, l-1) + (1 - \alpha_s) S_f(j, l) \quad (12)$$

onde, α_s ($0 < \alpha_s < 1$) é parâmetro de suavização e $S_f(j, l)$ é obtido pela suavização da frequência da potência espectral do ruído

$$S_f(j, l) = \sum_{i=-\omega}^{\omega} b(i) |D(j-i, l)|^2, \quad (13)$$

onde b representa uma função de janela normalizada. O tempo de suavização da segunda iteração é similar ao da primeira, e utiliza o mesmo parâmetro de suavização. O método de mínima localização no IMCRA está em conformidade com o utilizado pelo método de mínimas estatísticas, onde os locais de mínimo de $S(l, k)$ são pesquisados dentro de uma janela de duração finita de tamanho D , para cada posição de frequência.

$$S_{min}(j, l) \triangleq \min \{S(j, l') | l - D + 1 \leq l' \leq l\}. \quad (14)$$

Para reduzir a complexidade computacional, a janela de D amostras é geralmente dividida dentro de sub-janelas U de tamanho V . O processamento da probabilidade de presença de voz também é atribuído como um estimador de valor de SNR obtido a priori. No IMCRA, o SNR a priori é estimado por

$$\hat{\xi}(j, l) = \alpha G_{H_1}^2(j, l-1) \gamma(j, l-1) + (1-\alpha) \max \{\gamma(j, l) - 1, 0\} \quad (15)$$

onde, α é um fator de ponderação, $G_{H_1}(j, l)$ é uma função espectral de ganho, e $\gamma(j, l) = |D(j, l)|^2 / \gamma(j, l)$ representa o valor de SNR obtido a posteriori.

C. Proposta de método de realce de sinal de voz

A técnica proposta neste trabalho, ao contrário das apresentadas, é realizada no domínio do tempo, utilizando o algoritmo de estimação robusta de desvio padrão [8]. O processo de realce com o uso deste método tem por objetivo aprimorar os sinais de voz corrompidos por ruídos não-estacionários. A seleção do método proposto por [8], se deve a existência de um detector de presença de sinais em seu algoritmo. Esta característica é muito útil, pois serve como uma alternativa ao VAD.

A proposta de realce aplicada diretamente ao sinal de voz corrompido, consiste em três passos abaixo:

Passo 1: Parametrização

- Entrada da sequência de amostras do sinal de voz corrompida de tamanho (N): (Y_1, Y_2, \dots, Y_N) ;
- Escolha do tamanho do janelamento (J).

Passo 2: Estimação dos parâmetros

- A sequência é dividida em C sub-amostras e a partir daí se inicia a estimação para cada um $C_{(i)}$ do desvio padrão (σ_i) e do limitante (τ_i);
- Os valores $(\sigma_1, \sigma_2, \dots, \sigma_I)$ e $(\tau_1, \tau_2, \dots, \tau_I)$ são armazenados nos vetores \mathbf{S} e \mathbf{G} , respectivamente.

Passo 3: Processamento

- Os sinais algébricos das amostras são armazenados em:

$$\mathbf{F}(\mathbf{n}) = \begin{cases} 1 & ; Y_{(n)} \geq 0 \\ -1 & ; Y_{(n)} < 0 \end{cases} \quad (16)$$

- As amostras (Y_1, Y_2, \dots, Y_N) são normalizadas ($|Y_1|, |Y_2|, \dots, |Y_N|$);
- Cada $|Y_n|$ da sub-amostra $C_{(i)}$ é subtraída do respectivo valor σ_i (calculado no passo 2) $|W_n| = |Y_n| - \sigma_i$;
- Cada $|W_n|$ gerado pela subtração realizada no item anterior, é comparado com o limitante τ_i :

$$\mathbf{Real}(\mathbf{n}) = \begin{cases} |W|_N & ; |W|_N \geq \tau(i) \\ 0 & ; |W|_N < \tau(i) \end{cases} \quad (17)$$

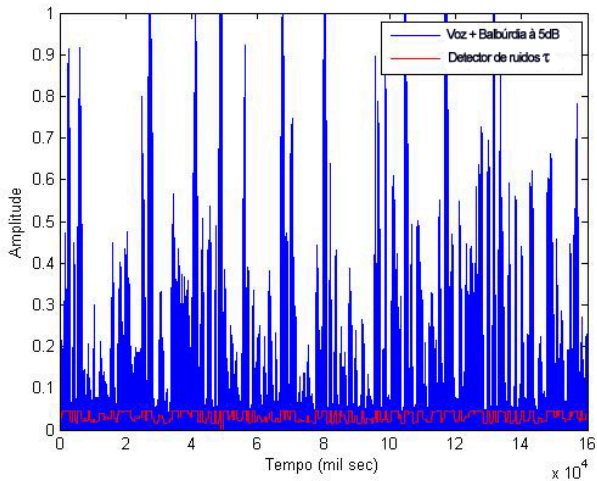


Fig. 1. Limitante τ indicando a região que será eliminada na fase de processamento.

- Finalmente o vetor $\mathbf{F}(\mathbf{n})$ é multiplicado pelo vetor $\mathbf{Real}(\mathbf{n})$ gerando o sinal realçado.

IV. DESCRIÇÃO DOS EXPERIMENTOS

Três experimentos foram realizados para a avaliação do método proposto de realce de sinais (Pro). Este método foi aplicado como realce de sinal de voz e foi avaliado como alternativa de pós-processamento ao método de subtração espectral (SS+Pro) proposto por [5], e por último como pós-processamento do método IMCRA+Pro proposto por [7].

Os experimentos foram realizados utilizando a base de voz TIMIT [11], com duração média de 30 segundos, e taxa de amostragem de 8KHz. Estes sinais de voz foram corrompidos por quatro ruídos acústicos reais, Avião, Balbúrdia, Fábrica e Navio, e um ruído Branco artificial extraídos da base NOISEX-92 [12]. Antes de serem adicionados ao sinal de voz, os ruídos foram subamostrados para a taxa do sinal de voz (8KHz) e foram gerados sinais com SNR de 0 dB, 5 dB, 10 dB, 15 dB e 20 dB. Os resultados foram analisados utilizando o SNR Segmental (segSNR) [13]. A Fig.1 mostra a fronteira gerada (em vermelho) pelo limitante τ , valores abaixo desta região são considerados pela técnica como somente compostos por ruídos, logo estes serão retirados. Neste caso, o sinal analisado está misturado ao ruído Balbúrdia a uma razão sinal ruído de 5 dB.

V. RESULTADOS E DISCUSSÃO

Nesta seção são apresentados os resultados principais dos experimentos realizados utilizando a técnica de realce proposta neste trabalho. O primeiro experimento foi realizado aplicando diretamente a técnica de realce (Pro) ao sinal corrompido. Os resultados do ganho de realce são apresentados na terceira coluna da Tab. I ou da Tab. II, estes são comparados com a aplicação aos mesmos sinais, dos métodos de SS e IMCRA, que se encontram na quarta coluna da Tab. I e II respectivamente. O resultado geral do Pro foi melhor do que dos outros

TABELA I

GANHO DE SEGSNR OBTIDO COM A PROPOSTA, SS E PÓS PROCESSAMENTO AO MÉTODO SS

Ruído	SNR	SegSNR		
		Pro	SS	SS+Pro
Avião	20dB	0,98	0,04	1,07
	15dB	1,21	0,43	1,33
	10dB	1,78	1,02	1,96
	5dB	2,67	2,23	2,99
	0dB	3,82	3,27	4,18
	Média	2,09	1,40	2,31
Balbúrdia	20dB	-0,10	-0,12	0,09
	15dB	0,14	0,09	0,13
	10dB	0,32	0,29	0,29
	5dB	1,88	1,32	1,79
	0dB	2,13	2,15	2,03
	Média	0,87	0,75	0,87
Branco	20dB	1,13	0,14	1,23
	15dB	1,91	0,83	2,13
	10dB	2,85	1,41	3,14
	5dB	3,33	2,89	3,73
	0dB	3,84	3,27	4,21
	Média	2,61	1,71	2,89
Fábrica	20dB	2,52	0,73	2,85
	15dB	3,18	1,64	3,69
	10dB	3,43	2,85	4,02
	5dB	3,67	3,23	4,37
	0dB	3,72	3,53	4,46
	Média	3,30	2,40	3,88
Navio	20dB	0,82	0,44	0,90
	15dB	1,72	0,58	1,91
	10dB	2,01	0,97	2,24
	5dB	2,49	1,04	2,76
	0dB	3,25	2,87	3,54
	Média	2,06	1,18	2,27
Média		2,19	1,49	2,44

métodos, destacando-se em relação ao SS com uma diferença de quase 1 dB. O melhor resultado atingido pela técnica foi no realce do ruído Fábrica, aumentando em 3,72 dB a razão sinal ruído segmental.

O pior desempenho ocorreu no realce de sinal de voz corrompido por ruído Balbúrdia com -0,10 dB. Este baixo desempenho também foi verificado no SS e no IMCRA. O resultado do ganho de segSNR do Pro em todos os ruídos ficaram acima do SS e IMCRA, a diferença percentual média de ganho é da ordem de quase 25%.

O resultado para este experimento mostrou que a utilização do método proposto como pós-processamento para o método de SS, foi melhor do que o do SS aplicado sozinho, tendo o melhor resultado no ruído acústico Navio com mais de 1 dB de diferença. Considerando o resultado geral, composto pela média de todos os ruídos, do SS+Pro em relação ao SS o ganho é de aproximadamente 1 dB. O pior resultado nos testes realizados ocorreu com o ruído Balbúrdia no método IMCRA -1,92 dB. O melhor resultado razão sinal ruído foi no realce do ruído acústico Fábrica com um aumento de segSNR de 3,88 dB de média no SS+Pro e 5,16 em 0 dB para IMCRA.

O resultado da aplicação do método proposto como pós processamento do sinal realçado pelo método IMCRA (IMCRA+Pós) encontra-se na Tab II, assim como o experimento anterior este pós processamento foi comparado ao método proposto (Pro) e ao método IMCRA (IMCRA). O SegSNR

TABELA II
GANHO DE SEGSRN OBTIDO COM A PROPOSTA, IMCRA E PÓS
PROCESSAMENTO AO MÉTODO IMCRA

Ruído	SNR	SegSNR		
		Pro	IMCRA	IMCRA+Pro
Avião	20dB	0,98	-1,15	-0,23
	15dB	1,21	0,45	0,81
	10dB	1,78	1,25	1,42
	5dB	2,67	2,26	2,16
	0dB	3,82	2,82	3,48
	Média	2,09	1,13	1,53
Balbúrdia	20dB	-0,10	-1,92	-1,34
	15dB	0,14	0,32	0,05
	10dB	0,32	0,56	0,29
	5dB	1,88	0,94	0,86
	0dB	2,13	1,53	1,38
	Média	0,87	0,29	0,25
Branco	20dB	1,13	1,15	0,91
	15dB	1,91	1,33	1,93
	10dB	2,85	2,43	2,34
	5dB	3,33	3,46	3,03
	0dB	3,84	3,46	3,53
	Média	2,61	2,36	2,35
Fábrica	20dB	2,52	0,34	0,82
	15dB	3,18	2,14	2,57
	10dB	3,43	3,62	3,40
	5dB	3,67	4,51	4,04
	0dB	3,72	5,16	4,84
	Média	3,30	3,15	3,13
Navio	20dB	0,82	0,63	0,16
	15dB	1,72	0,21	1,11
	10dB	2,01	1,43	1,81
	5dB	2,49	2,21	1,99
	0dB	3,25	3,42	2,60
	Média	2,06	1,58	1,54
Média		2,19	1,70	1,76

geral, composto pela média de todos os ruídos, indica que Pro apresentou um desempenho melhor que o IMCRA. Este resultado se deve ao ganho alcançado por Pro no realce do arquivo de voz corrompido por ruído Avião, a diferença obtida foi de quase 1 dB quando comparado ao IMCRA, e no ruído Balbúrdia, que mesmo tendo o pior resultado geral para todos os métodos (Pro, IMCRA, IMCRA+Pro) colaborou para esse melhor desempenho de Pro. No IMCRA+Pro o resultado geral é pior do que o apresentado na SS+Pro. O pós-processamento neste caso não trouxe resultado significativo.

Comparando o resultado do pós-processamento com Pro, observa-se que ao se aplicar o sinal tratado pelo IMCRA, o método apresentado realiza uma insignificante melhora. Indicando que a melhor opção é utilizar o pós processamento empregado no método de subtração espectral, ou usar somente o Pro, pois apesar de, o melhor resultado ter sido alcançado com SS+Pro, dado o tempo de processamento maior deste método seja mais indicado utilizar somente o método Pro para o aprimoramento ou realce de sinal de voz corrompido por ruído acústico.

VI. CONCLUSÃO

Este trabalho apresentou um novo método de realce de voz corrompida por ruído acústico, com características não-estacionárias, realizado no domínio do tempo com utilização do algoritmo robusto de estimação de desvio padrão proposto

por [8]. Optou-se pela utilização deste método dado a sua robustez na estimação e por apresentar dentro de seu algoritmo uma técnica de detecção de *outliers*.

Para avaliar o seu desempenho foram realizados três experimentos, o primeiro com a aplicação direta do método ao sinal de voz, e os outros dois que, antes da aplicação da técnica, submetem os sinais de voz a um pré-processamento utilizando o método SS e IMCRA.

Os resultados dos experimentos mostraram que, em presença de sinais não-estacionários, a nova proposta de realce, quando comparada com os métodos SS e IMCRA, apresenta um ganho de segSNR.

O primeiro teste utilizando a técnica como pós-processamento aplicado ao sinal aprimorado pelo método SS, apresentou o maior segSNR de todos os experimentos, já nos resultados do pós-processamento do IMCRA verificou-se que, a técnica não acrescentou ganho quando aplicada com pré-processamento ao método proposto.

Quando os resultados da proposta são comparados percebe-se que, o método de pós-processamento aplicado ao SS apresenta um desempenho melhor do que o método aplicado sozinho. Mas se o custo computacional for levado em consideração os ganhos gerados pela primeira proposta apresentam uma melhor relação realce e tempo de processamento.

REFERÊNCIAS

- [1] I. Cohen, B. Berdugo, "Speech enhancement for non-stationary noise environments," *Signal Processing*, Vol. 81, n. 11, pp. 2403-2418, November 2001.
- [2] K. Manohar, P. Rao, "Speech enhancement in non-stationary noise environments using noise properties," *Speech Communication*, Vol. 48, n. 1, pp 96-109, January 2006.
- [3] C.W. Maina, J.M. Walsh, "Joint Speech Enhancement and Speaker Identification Using Approximate Bayesian Inference Audio, Speech, and Language," *IEEE Transactions on*, vol.19, no.6, pp.1517,1529, August. 2011.
- [4] A. Spriet, M. Moonen, J. Wouters, "Robustness analysis of multichannel Wiener filtering and generalized side lobe cancellation for multi-microphone noise reduction in hearing aid applications," *Speech and Audio Processing, IEEE Transactions on*, vol.13, no.4, pp.487,503, July 2005.
- [5] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech and Signal Processing* vol. 27, pp. 113-120, April 1979.
- [6] Y. Ephraim, D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 32, pp. 1109-1121, December 1984.
- [7] I. Cohen, "Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging", *IEEE Transactions on Speech and Audio Processing*, vol. 11, pp. 466-475, September 2003.
- [8] D. Pastor and F. Socheleau, "Robust estimation of noise standard deviation in presence of signals with unknown distributions and occurrences," *IEEE transactions on signal processing*, n. 99, January 2012.
- [9] F. Hampel, E. Ronchetti, P. Rousseeuw, and W. Stahel, "Robust Statistics: the Approach based on Influence Functions," John Wiley and Sons, New York, 1986.
- [10] P. Huber and E. Ronchetti, "Robust Statistics", second edition, John Wiley and Sons, 2009.
- [11] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, N. Dahlgren, and V. Zue, "Timit acoustic-phonetic continuous speech corpus, Linguistic Data Consortium, Philadelphia, 1993.
- [12] A. Varga and H. Steeneken, "Assessment for automatic speech recognition ii: Noisex-92: a database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Communications*, vol. 12, no. 3, pp 247-251, 1993.
- [13] Y. Hu and P. Loizou "Evaluation of objective measures for speech enhancement," *Proceedings of Interspeech 2006* pp. 1-4, September 2006.