Realce de Sinais de Voz em Presença de Ruídos Acústicos Não-Estacionários utilizando o Método EMD

L. Zão e R. Coelho

***Resumo*— Este trabalho apresenta uma técnica de realce de sinais de voz baseada no método EMD. A técnica EMDF é avaliada em locuções corrompidas por seis ruídos acústicos, em seis diferentes níveis de relação sinal-ruído. Os ruídos utilizados possuem diferentes valores para o índice de não-estacionariedade. Como referência, uma outra técnica de realce, baseada nos estimadores IMCRA e OMLSA, é também aplicada sobre os sinais de voz. Duas medidas são utilizadas na avaliação de desempenho das técnicas de realce: a relação sinal-ruído segmental e a taxa de acertos obtida em uma aplicação de identificação automática de locutor. Os resultados mostram que o desempenho das técnicas é fortemente dependente das características de estacionariedade dos ruídos.**

***Palavras-Chave*— Realce de sinais de voz, ruídos não-estacionários, decomposição EMD.**

***Abstract*— This paper presents a speech enhancement technique based on the empirical mode decomposition method. The EMD-based filtering is evaluated in speech utterances corrupted with six acoustic noises and six different signal-to-noise ratios. The noises present different values of INS (index of nonstationarity). For comparison, another speech enhancement technique, based on the use of IMCRA and OMLSA estimators, is used in the experiments. The comparison among the techniques considers two different measures: the segmental signal-to-noise ratio and the accuracies in a speaker identification task. The results show that the enhancement performance is strongly dependent on the INS of the noises.**

***Keywords*— Speech enhancement, nonstationary noises, empirical mode decomposition.**

I. INTRODUÇÃO

A degradação sofrida pelo sinal de voz, quando captado em ambientes acusticamente ruidosos, é um dos principais desafios da área de processamento de voz. Técnicas de realce espectral de sinais foram propostas para reduzir os efeitos dos ruídos acústicos. A maioria delas utiliza uma estimação do espectro de frequências do ruído para, suprimindo as suas componentes, reconstruir o sinal limpo. Um dos principais obstáculos consiste em estimar as variações no espectro quando os ruídos acústicos são não-estacionários.

A técnica clássica de subtração espectral [1] estima o espectro do ruído nos momentos em que não há presença de voz. Para isto, utiliza-se um detector de presença da voz (VAD - *voice activity detector*). O espectro estimado para o ruído é então subtraído na reconstrução do sinal de voz, assumindo que o mesmo não varia ao longo do tempo, ou seja, que o ruído é estacionário.

Leonardo Zão, Programa de Pós-Graduação em Engenharia de Defesa, Instituto Militar de Engenharia (IME), Rio de Janeiro, Brasil, E-mail: zao@ime.eb.br. Rosângela Coelho, Programa de Pós-Graduação em Engenharia Elétrica, Instituto Militar de Engenharia (IME), Rio de Janeiro, Brasil, E-mail: coelho@ime.eb.br.

Em [2], os autores propõem o estimador MCRA (*minima controlled recursive averaging*) para estimar o espectro de ruídos acústicos não-estacionários. O espectro é obtido a partir da média das estimações em momentos passados, ajustada por um parâmetro dependente da probabilidade da presença de voz. A técnica IMCRA (*improved* MCRA) [3] é apresentada como uma alternativa de estimação para situações severas de relação sinal-ruído (RSR) e longos períodos de atividade da voz. Apesar dos estimadores MCRA e IMCRA assumirem a não-estacionariedade dos ruídos, a estimação torna-se imprecisa quando o espectro apresenta bruscas variações [4].

O método EMD (*empirical mode decomposition*) [5] foi proposto para análise não-linear de sinais no domínio do tempo. Em [6], os autores demonstram que o método, quando aplicado sobre ruídos fGn (*fractional Gaussian noise*), resulta em uma decomposição semelhante àquela obtida por um banco de filtros diádicos. Esta constatação levou à proposta de uma técnica de supressão de ruído [7]. O método EMD foi também adotado em um algoritmo de pós-processamento [8] para aplicação em sinais de voz corrompidos por ruídos de baixas frequências. Nesta abordagem, é considerada a hipótese de que o sinal de voz é previamente realçado pelo uso da técnica IMCRA.

Neste trabalho, três técnicas de realce de sinais de voz são avaliadas para situações de ruídos acústicos não-estacionários. Primeiramente, é utilizada a proposta de Cohen [3], que utiliza o estimador IMCRA. Em seguida, a técnica EMDF (*EMD-based filtering*) [8] é aplicada sobre os sinais previamente realçados. Finalmente, ela é também aplicada diretamente sobre os sinais de voz corrompidos, ou seja, sem nenhum processamento prévio. A classificação dos ruídos adotados nos experimentos é realizada utilizando-se o índice de não-estacionariedade (INS - *index of nonstationarity*) proposto em [9]. Neste artigo, as locuções de voz realçadas são ainda utilizadas em um sistema de identificação automática de locutor. O objetivo é verificar se o realce dos sinais de voz ocasiona aumento na acurácia dos testes.

Os resultados dos experimentos demonstram que, para os ruídos não-estacionários, os maiores incrementos de RSR foram obtidos com a técnica de Cohen. Por outro lado, nos testes de identificação de locutor, a melhor acurácia foi obtida com as locuções realçadas pela técnica EMDF.

O restante deste trabalho está organizado da seguinte forma. A Seção II apresenta o algoritmo do método EMD. Na Seção III, são descritas as técnicas para realce da voz. Os experimentos com as técnicas de realce são discutidos na Seção IV. Ainda nesta Seção, são estudados os índices de não-

estacionariedades dos ruídos acústicos adotados neste trabalho. Na Seção V, os sinais de voz realçados são avaliados na tarefa de identificação de locutor. Finalmente, a Seção VI apresenta as principais conclusões deste trabalho.

II. MÉTODO EMD

O método EMD foi proposto em [5] como uma forma não-linear de análise de sinais não-estacionários. Considere um sinal $x(t)$ contendo dois máximos locais consecutivos nos pontos t_- e t_+. Para valores de t no intervalo $t_- \leq t \leq t_+$, pode-se definir uma componente de altas frequências do sinal que passa por estes máximos, e pelo mínimo local que existe entre eles. Desta componente, chamada de detalhes $d(t)$, identifica-se uma componente de tendência local $m(t)$, tal que $x(t) = d(t) + m(t)$ no intervalo $t_- \leq t \leq t_+$. Quando esta decomposição é aplicada a todas as oscilações presentes no sinal $x(t)$, o conjunto das componentes de detalhes define uma função intrínseca de modo (IMF - *Intrinsic Mode Function*). Analogamente, um sinal residual é definido pelo conjunto de componentes de tendência locais. Aplicando repetidamente o procedimento sobre o sinal residual, chega-se a um conjunto de IMFs e a um resíduo de baixas frequências.

O algoritmo para o método EMD aplicado sobre um sinal $x(t)$ pode ser dividido nos seguintes passos [5] [6]:

1) Identificar todos os extremos (máximos e mínimos locais) de $x(t)$;
2) Obter as envoltórias $e_{max}(t)$ e $e_{min}(t)$, utilizando interpolação por *splines* cúbicas nos pontos de máximo e mínimo, respectivamente;
3) Calcular a componente de tendências como a média entre as envoltórias: $m(t) = (e_{min}(t) + e_{max}(t))/2$;
4) Extrair os detalhes: $d(t) = x(t) - m(t)$;
5) Repetir a iteração sobre o sinal residual $m(t)$.

Em geral, para garantir que a componente de detalhes $d(t)$ extraída no passo (4) seja considerada uma IMF, os passos (1)-(4) são repetidos com $d(t)$ no lugar de $x(t)$. Este processo é repetido até garantir que a nova função $d(t)$ tenha média próxima de zero. Ao final de um número finito N de iterações, o sinal $x(t)$ pode ser escrito como

$$x(t) = \sum_{n=1}^{N} \text{IMF}_n(t) + m(t)\,, \tag{1}$$

onde $\text{IMF}_n(t), 1 \leq n \leq N$, são as funções de detalhes obtidas no passo (4) de cada iteração, e $m(t)$ é o sinal residual obtido na última iteração.

III. TÉCNICAS PARA REALCE DA VOZ

Esta Seção apresenta a descrição da técnica EMDF [8] para filtragem de ruídos, baseada no método EMD. Em seguida, é apresentada a técnica proposta por Cohen [2] [3] para realce de sinais de voz.

A. Técnica EMDF

A filtragem baseada em EMD foi proposta em [8] para realçar sinais de voz corrompidos por ruídos de baixas frequências. Nesta técnica, o método EMD é inicialmente aplicado sobre o sinal ruidoso. Em seguida, o sinal de voz é parcialmente reconstruído segundo (1) mas utilizando apenas os índices dos modos (IMFs) compostos predominantemente pelo sinal de voz.

De forma a determinar quais IMFs devem ser excluídas na reconstrução do sinal de voz, utiliza-se o fato de que a maior parte da energia de um sinal de voz limpo se concentra nas quatro primeiras IMFs [8]. Os autores demonstram que as variâncias dos modos decaem significativamente para índices $n \geq 5$. Assim, o aumento na variância destes modos, isto é, $Var[\text{IMF}_n(t)] > Var[\text{IMF}_{n-1}(t)], n > 4$, indica que estes são fortemente afetados pelas componentes de baixas frequências dos ruídos.

Para a filtragem, o sinal de voz é primeiramente dividido em pequenos quadros e, em cada um destes, é efetuada a busca pelas IMFs mais comprometidas por ruídos. Esta busca quadro a quadro é necessária já que, para sinais não-estacionários, as características do ruído podem se alterar ao longo do tempo. O algoritmo da EMDF pode então ser resumido nos seguintes passos:

1) Dividir o sinal de voz $x(t)$ em quadros $x_l(t), l = 1, \ldots, Q$;
2) Para cada quadro $x_l(t)$, efetuar a decomposição em N funções $\text{IMF}_n(t), n = 1, \ldots, N$;
3) Estimar as variâncias $V_l(n) = Var\,[\text{IMF}_n(t)]$;
4) Identificar os índices dos picos p_l tais que $V(p_l) > V(p_l - 1)$ e $V(p_l) > V(p_l + 1)$, para $p_l > 4$;
5) Determinar o índice v_l do vale imediatamente anterior ao pico p_l, isto é, $V(v_l) < V(v_l - 1)$ e $V(v_l) < V(v_l + 1)$, para $v_l < p_l$;
6) Determinar para qual quadro $\hat{l}$ a diferença $p_{\hat{l}} - v_{\hat{l}}$ é máxima;
7) Reconstruir o sinal de voz $x(t) = \sum_{n=1}^{M} \text{IMF}_n(t)$, onde $M = v_{\hat{l}}$.

B. Técnica Cohen

Seja $y(t)$ um sinal de voz corrompido por um ruído aditivo $r(t)$. Assim, pode-se escrever $y(t) = x(t) + r(t)$, onde $x(t)$ é o sinal de voz limpo. Para o realce do sinal de voz, a estimação do espectro do ruído $\hat{\lambda}_r$ é realizada em duas iterações. Primeiramente, é implementado um VAD dependente do tempo e da frequência. Em seguida, esta detecção é utilizada para refinar a estimação do espectro do ruído nos quadros onde há atividade de voz.

Inicialmente, o janelamento e a análise espectral, via transformada de Fourier em tempo curto (STFT - *short-time Fourier transform*), do sinal $y(t)$ leva à relação

$$Y(k,l) = X(k,l) + R(k,l)\,, \tag{2}$$

onde k é o índice correspondente à sub-banda (frequência) e l é o índice do quadro (tempo). Na primeira iteração, a estimação do espectro de frequências do ruído é obtida por uma suavização nos domínios do tempo e frequência:

$$\begin{aligned} S_f(k,l) &= \textstyle\sum_{i=-w}^{w} b(i)|Y(k-i,l)|^2\,, \\ S(k,l) &= \alpha_s S(k,l-1) + (1-\alpha_s) S_f(k,l)\,, \end{aligned} \tag{3}$$

onde $\alpha_s \in [0,1]$ é um coeficiente de suavização e o janelamento $b(i)$ obdece à restrição $\sum_{i=-w}^{w} b(i) = 1$.

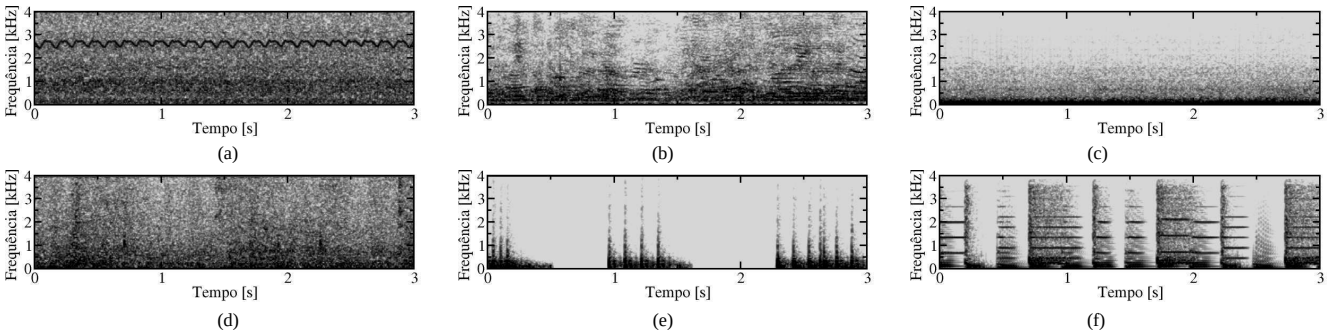

Fig. 1. Espectrogramas dos diferentes ruídos utilizados neste trabalho: (a) Avião, (b) Balbúrdia, (c) Carro, (d) Fábrica, (e) Metralhadora e (f) *Ringtone*.

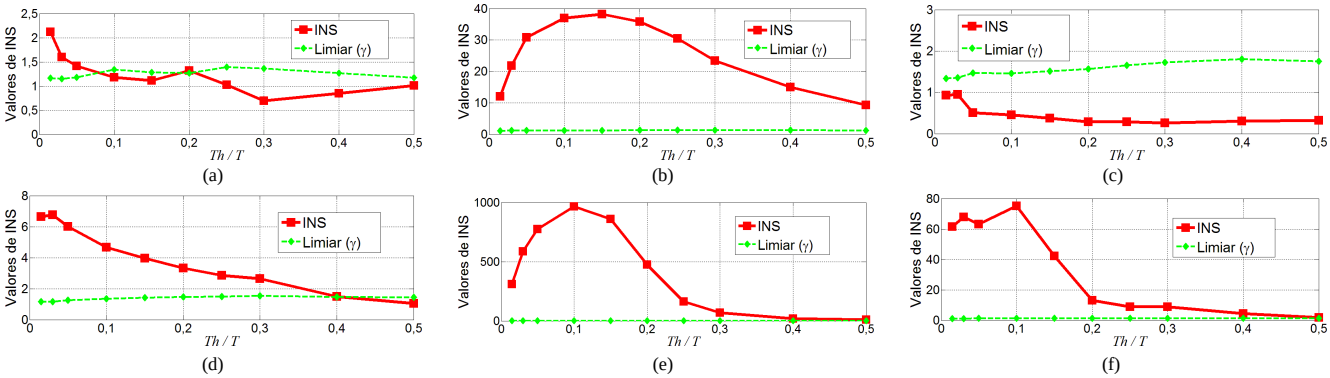

Fig. 2. Curvas dos Índices de não-Estacionariedade (INS) para os ruídos (a) Avião, (b) Balbúrdia, (c) Carro, (d) Fábrica, (e) Metralhadora e (f) *Ringtone*.

Os valores de $S(k,l)$ são comparados com o menor valor $S_{min}(k,l) = \min\{S(k,l'); l-L<l'\leq l\}$, obtido de um conjunto de L quadros passados. A partir desta relação, chega-se a um critério de decisão sobre a presença de voz em cada quadro l e em cada sub-banda k:

$$I(k,l) = \begin{cases} 1, & \text{ausência de voz,} \\ 0, & \text{presença de voz.} \end{cases} \quad (4)$$

Na segunda iteração, é realizada uma nova estimação do espectro suavizado $\tilde{S}(k,l)$, considerando apenas as componentes formadas predominantemente pelo ruído, isto é, $I(k,l)=1$. Da nova relação entre os valores de $\tilde{S}(k,l)$ e seus valores mínimos, define-se a probabilidade $p(k,l)$, contínua no intervalo $[0,1]$, de haver presença de voz no quadro l e na sub-banda k. A estimação do espectro $\bar{\lambda}_r(k,l)$ do ruído é dada pela relação recursiva

$$\bar{\lambda}_r(k,l+1) = \tilde{\alpha}_r(k,l)\bar{\lambda}_r(k,l) + [1-\tilde{\alpha}_r(k,l)]|Y(k,l)|^2\,, \quad (5)$$

onde $\tilde{\alpha}_r(k,l)$ é um coeficiente de suavização variante no tempo e dependente da probabilidade da presença de voz $p(k,l)$. Assim,

$$\tilde{\alpha}_r(k,l) \triangleq \alpha_r + (1-\alpha_r)p(k,l)\,, \quad (6)$$

com α_r é constante. Finalmente, o espectro estimado é obtido pela multiplicação de $\bar{\lambda}_r$ por um fator de compensação β:

$$\hat{\lambda}_r(k,l) = \beta\,\bar{\lambda}_r(k,l)\,. \quad (7)$$

Após a estimação do espectro, a reconstrução do sinal de voz é realizada utilizando o estimador OMLSA (*optimally-modified log-spectral amplitude*), proposto em [2] e alterado em [10]. Neste, o espectro do sinal original é multiplicado por um fator de ganho dependente dos índices de tempo e frequência, resultando numa estimativa do espectro do sinal de voz limpo.

IV. Experimentos e Resultados

Os experimentos para comparação entre as técnicas de realce de voz foram conduzidos com um subconjunto de 168 locutores da base de voz TIMIT [11]. Para análise de desempenho, foram utilizadas 336 locuções, sendo duas por locutor, com duração média de 3 segundos e taxa de amostragem de 16 kHz. Cada uma destas locuções foi corrompida por 6 ruídos acústicos reais: Avião, Balbúrdia, Carro, Fábrica, Metralhadora e *Ringtone*. Com exceção do ruído *Ringtone*, obtido em [12], todos os ruídos foram extraídos da base NOISEX-92 [13]. Antes de serem adicionados, os ruídos foram subamostrados para a mesma taxa dos sinais de voz.

A. Teste de Estacionaridade

A Fig. 1 apresenta os espectrogramas de segmentos de 3 segundos de duração para cada um dos ruídos. Como pode-se verificar, alguns dos ruídos possuem variações ou oscilações nos seus espectros (vide ruído Matralhadora, Fig. 1.e). Outros, como o ruído Carro (Fig. 1.c), possuem espectro com característica estacionária. O índice de não-estacionariedade (INS) [9] foi utilizado para diferenciar os ruídos quanto às suas características.

A Fig. 2 ilustra os valores de INS obtidos para os seis ruídos utilizados neste trabalho. A escala de tempo Th/T indica a relação entre o tamanho da janela de tempo utilizada na análise espectral de tempo curto (Th) e a duração total (T)

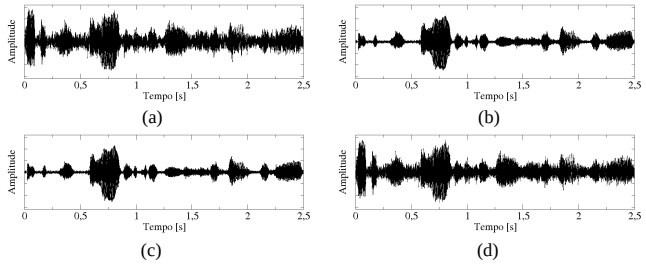

Fig. 3. Amostras de um sinal de voz de um locutor feminino: (a) voz corrompida com ruído Fábrica com RSR de 5 dB, (b) voz realçada com a técnica Cohen, (c) voz filtrada com EMDF após a técnica Cohen, e (d) voz filtrada com EMDF sem a técnica Cohen.

do ruído. Os valores de γ indicam os limiares do teste de estacionariedade, com grau de confiança de 95%. Assim,

$$\text{INS} \begin{cases} \leq \gamma & \text{, ruído é estacionário;} \\ > \gamma & \text{, ruído é não-estacionário.} \end{cases} \quad (8)$$

Os valores de INS obtidos mostram que os ruídos Balbúrdia, Metralhadora e *Ringtone* são não-estacionários para todas as escalas de tempo. Com relação ao ruído Fábrica, seus valores de INS indicam estacionariedade apenas para escala de tempo $Th/T \approx 0,5$. Já o ruído Avião é estacionário, com indicação de não-estacionariedade apenas para as menores escalas de tempo ($Th/T < 0,1$). Finalmente, o ruído Carro é um exemplo de ruído estacionário para todas as escalas de tempo. Ou seja, dos seis ruídos considerados neste trabalho, dois (Avião e Carro) são estacionários e quatro (Babúrdia, Fábrica, Metralhadora e *Ringtone*) são não-estacionários.

B. Resultados das Técnicas de Realce

Experimentos de realce de sinais de voz foram realizados para comparar as técnicas apresentadas na Seção III. Primeiramente, a técnica Cohen é aplicada a todas as locuções corrompidas. Em seguida, os sinais de voz realçados são filtrados utilizando a técnica EMDF. Finalmente, a filtragem é também realizada sobre os sinais de voz sem o emprego da técnica Cohen. A Fig. 3 mostra um exemplo de sinal de voz oriundo de um locutor feminino. Nesta figura, são ilustradas as formas de onda resultantes da adição do ruído Fábrica com RSR de 5 dB, e os sinais resultantes das aplicações das técnicas Cohen e EMDF.

Para análise de desempenho das técnicas adotadas neste trabalho, foi utilizada a medida de relação sinal-ruído segmental (RSRSeg), definida por:

$$\text{RSRSeg} = \frac{10}{|\mathcal{L}|} \sum_{l \in \mathcal{L}} \log \frac{\sum_k |X(k,l)|^2}{\sum_k |R(k,l)|^2} \,. \quad (9)$$

Em (9), X e R são as componentes STFT definidas em (2), $\mathcal{L}$ é o conjunto de quadros onde há presença de voz e $|\mathcal{L}|$ é a sua cardinalidade.

A Fig. 4 mostra os incrementos de RSRSeg médios obtidos com a aplicação da técnica Cohen considerando os diferentes ruídos e valores de RSR. Como pode-se observar, o ruído Carro, que é estacionário, foi o que atingiu os melhores resultados de RSRSeg. Já os ruídos *Ringtone* e Metralhadora, que possuem os maiores índices de não-estacionariedade (vide Fig. 2), foram aqueles que apresentaram menores incrementos

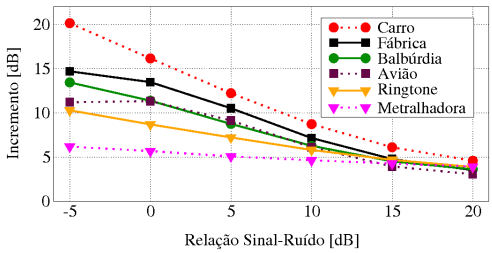

Fig. 4. Curvas com o incremento médio de RSRSeg com a técnica Cohen.

TABELA I
INCREMENTOS MÉDIOS DE RSR SEGMENTAL OBTIDOS PARA CADA UM DOS RUÍDOS (EMDF-1: COM PRÉ-PROCESSAMENTO; EMDF-2: SEM PRÉ-PROCESSAMENTO).

Ruído	Cohen	EMDF-1	EMDF-2
Carro	11,32	10,97	6,83
Fábrica	9,02	8,93	1,58
Balbúrdia	7,99	7,69	0,84
Avião	7,46	7,16	0,44
Ringtone	6,76	5,90	0,12
Metralhadora	4,95	4,18	0,00
Média	7,92	7,47	1,64

de RSRSeg. A queda no desempenho da técnica Cohen para os ruídos altamente não-estacionários é explicado pela dificuldade do estimador IMCRA em captar as bruscas variações ou oscilações do espectro.

A comparação entre os incrementos na RSR segmental, em dB, obtidos nos três experimentos é apresentada na Tab. I. A coluna EMDF-1 corresponde ao experimento onde a filtragem é utilizada como um pós-processamento para a técnica de Cohen. Já a coluna EMDF-2 se refere à filtragem sem qualquer prévio realce. Os valores correspondem às médias para os seis valores de RSR: de -5 a 20 dB.

Note que nenhum dos dois casos onde a filtragem foi utilizada atingiu maiores incrementos em relação à técnica Cohen. Observe ainda que, como nos resultados ilustrados na Fig. 4, o melhor desempenho da EMDF também foi obtido para o ruído estacionário Carro, e os piores foram obtidos para os ruídos não-estacionários Metralhadora e *Ringtone*.

V. IDENTIFICAÇÃO AUTOMÁTICA DE LOCUTOR

A identificação automática de locutor foi utilizada como uma segunda medida de avaliação das técnicas estudadas neste trabalho. Para isto, todas as locuções corrompidas pelos ruídos acústicos, bem como aquelas resultantes das técnicas de realce, foram utilizadas nos testes de identificação. Para a formação das matrizes de atributos, as locuções foram divididas em quadros de 20 ms, com 50% de sobreposição. De cada quadro, foi extraído um vetor de atributos MFCC (*mel-frequency cepstral coefficients*) [14] com 12 componentes. Para a modelagem dos locutores, foram adotados os modelos de misturas Gaussianas (GMM - *Gaussian mixture models*) [15]. O modelo de cada locutor foi obtido a partir da concatenação de oito locuções limpas, diferentes daquelas utilizadas nos experimentos da Seção IV.

A Tab. II apresenta as taxas de acertos obtidas na identificação de locutor, com as locuções corrompidas. Já os resultados mostrados nas Tabs. III, IV e V correspondem à utilização nos testes de identificação dos sinais de voz

TABELA II

TAXAS DE ACERTOS (%) NA IDENTIFICAÇÃO DE LOCUTOR OBTIDOS COM SINAIS DE VOZ CORROMPIDOS POR DIVERSOS RUÍDOS.

Ruído	Relação Sinal-Ruído (dB)						
	-5	0	5	10	15	20	Média
Metralhadora	60,42	76,19	84,23	91,96	96,43	98,21	84,57
Carro	19,05	31,55	49,11	64,29	79,46	90,18	55,61
Balbúrdia	4,76	13,10	35,42	71,13	91,07	96,73	52,03
Ringtone	7,14	17,26	36,61	60,12	80,95	94,64	49,45
Fábrica	0,89	0,60	2,98	15,18	43,75	80,65	24,01
Avião	0,30	0,60	0,60	7,14	26,19	60,42	15,87
Média	15,43	23,21	34,82	51,64	69,64	86,81	46,92

TABELA III

TAXAS DE ACERTOS (%) NA IDENTIFICAÇÃO DE LOCUTOR OBTIDOS COM SINAIS DE VOZ REALÇADOS PELA TÉCNICA COHEN.

Ruído	Relação Sinal-Ruído (dB)						
	-5	0	5	10	15	20	Média
Carro	42,86	46,13	45,83	49,11	51,49	52,98	48,07
Metralhadora	20,83	25,00	29,17	36,01	43,15	44,64	33,13
Balbúrdia	5,06	8,33	15,48	31,55	42,26	50,89	25,60
Ringtone	3,57	6,55	11,31	22,02	36,01	45,24	20,78
Fábrica	2,38	3,27	10,42	21,73	36,90	44,94	19,94
Avião	0,60	1,49	4,76	13,99	30,36	44,64	15,97
Média	12,55	15,13	19,49	29,07	40,03	47,22	27,25

TABELA IV

TAXAS DE ACERTOS (%) NA IDENTIFICAÇÃO DE LOCUTOR OBTIDOS COM SINAIS DE VOZ FILTRADOS PELA TÉCNICA EMDF-1.

Ruído	Relação Sinal-Ruído (dB)						
	-5	0	5	10	15	20	Média
Carro	43,15	43,75	45,83	45,24	44,64	45,54	44,69
Metralhadora	22,32	28,57	33,04	34,52	38,99	42,56	33,33
Balbúrdia	4,76	10,71	18,45	32,44	43,75	47,62	26,29
Ringtone	3,57	6,55	13,99	22,02	36,31	42,26	20,78
Fábrica	1,79	4,17	9,82	23,21	33,63	39,29	18,65
Avião	1,49	2,98	8,33	15,77	31,25	39,88	16,62
Média	12,85	16,12	21,58	28,87	38,10	42,86	26,73

realçados pelas técnicas Cohen, EMDF-1 e EMDF-2, respectivamente.

Como pode-se observar, apesar de apresentar uma melhora nos valores de RSRSeg, a aplicação do realce de voz com a técnica Cohen reduz o desempenho médio da identificação de locutor de 46,92% para 27,25%. Em relação à técnica Cohen, a filtragem aplicada após o realce (Tab. IV) conseguiu aumentar as taxas de acertos para as condições de ruídos mais severas, ou seja, RSR menor que 10 dB.

Já os resultados apresentados na Tab. V mostram que a técnica EMDF aplicada sobre os sinais de voz ruidosos, sem prévio realce, conseguiu aumentar a taxa média de acertos. O melhor resultado foi obtido para o ruído Carro, com um aumento nas taxas de acertos de 55,61% para 83,23%. Note ainda que a EMDF conseguiu obter as maiores taxas para as condições de ruído mais severas, inclusive para os ruídos com maiores índices de não-estacionariedade (Metralhadora e *Ringtone*).

TABELA V

TAXAS DE ACERTOS (%) NA IDENTIFICAÇÃO DE LOCUTOR OBTIDOS COM SINAIS DE VOZ FILTRADOS PELA TÉCNICA EMDF-2.

Ruído	Relação Sinal-Ruído (dB)						
	-5	0	5	10	15	20	Média
Carro	63,99	77,98	84,23	87,50	91,37	94,35	83,23
Metralhadora	65,77	77,68	82,44	86,90	90,77	93,15	82,79
Ringtone	9,82	27,38	45,24	64,58	79,17	86,61	52,13
Balbúrdia	3,57	10,71	30,95	63,39	82,74	92,86	47,37
Fábrica	1,49	0,89	3,27	14,29	43,15	73,51	22,77
Avião	0,89	1,19	1,79	4,76	21,13	51,19	13,49
Média	24,26	32,64	41,32	53,57	68,06	81,94	50,30

VI. CONCLUSÃO

Este trabalho apresentou uma técnica para realce de sinais de voz baseada no método EMD. A técnica foi aplicada sobre locuções da base TIMIT corrompidas por ruídos estacionários e não-estacionários. Como referência, a técnica de realce proposta por Cohen foi também adotada nos experimentos. A técnica Cohen foi a que obteve maiores incrementos nos valores de relação sinal-ruído segmental (RSRSeg). Os melhores resultados foram obtidos para os ruídos estacionários. Por outro lado, os ruídos com maiores índices de não-estacionariedade foram os que resultaram em pior desempenho. Com relação à identificação de locutor, a técnica EMDF, sem o prévio realce, atinge as melhores taxas de acertos, principalmente para os casos mais severos de ruídos. A melhoria nos resultados de identificação com a filtragem foi obtida para ruídos estacionários e não-estacionários.

REFERÊNCIAS

[1] 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.

[2] I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," *Signal Processing*, vol. 81, no. 11, pp. 2403 – 2418, 2001.

[3] 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.

[4] K. Manohar and P. Rao, "Speech enhancement in nonstationary noise environments using noise properties," *Speech Communication*, vol. 48, pp. 96–109, January 2006.

[5] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, pp. 903–995, March 1998.

[6] P. Flandrin, G. Rilling, and P. Goncalves, "Empirical mode decomposition as a filter bank," *IEEE Signal Processing Letters*, vol. 11, pp. 112–114, February 2004.

[7] P. Flandrin, P. Gonçalves, and G. Rilling, "Detrending and denoising with empirical mode decompositions," *Proceedings of the European Signal Processing Conference (EUSIPCO 2004)*, pp. 1581–1584, September 2004.

[8] N. Chatlani and J. Soraghan, "EMD-Based Filtering (EMDF) of Low-Frequency Noise for Speech Enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, pp. 1158–1166, may 2012.

[9] P. Borgnat, P. Flandrin, P. Honeine, C. Richard, and J. Xiao, "Testing stationarity with surrogates: A time-frequency approach," *IEEE Transactions on Signal Processing*, vol. 58, pp. 3459–3470, july 2010.

[10] I. Cohen, "http://webee.technion.ac.il/people/israelcohen/."

[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] FindMIDIs.com, "http://www.findmidis.com."

[13] 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.

[14] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 28, pp. 357–366, August 1980.

[15] D. Reynolds and R. Rose, "Robust text independent speaker identification using gaussian mixture speaker models," *IEEE Transactions on Speech and Audio Processing*, vol. 3, pp. 72–82, 1995.