Variações da Decomposição Empírica de Modos para Realce de Sinais de Voz

L. Zão e R. Coelho

***Resumo*— A decomposição empírica de modos (EMD - *empirical mode decomposition*) tem sido adotada para a supressão de ruídos acústicos de sinais de voz no domínio do tempo. Este artigo investiga o desempenho de duas variações do EMD para realçar sinais de voz coletados em ambientes acusticamente ruidosos. Após a aplicação da decomposição, o expoente de Hurst é utilizado para identificar as componentes mais corrompidas por ruídos. Os resultados obtidos em experimentos de realce demonstram que os métodos alternativos ao EMD podem alcançar melhoria na qualidade e na inteligibilidade dos sinais de voz.**

***Palavras-Chave*— Realce de sinais de voz, decomposição empírica de modos, ruídos acústicos.**

***Abstract*— The empirical mode decomposition (EMD) has been adopted for the suppression of acoustic noise from speech signal in the time domain. This paper investigates two alternative methods to decompose and enhance speech signals collected in noisy environments. After the decomposition, the Hurst exponent is used to select the most corrupted modes. The results obtained in speech enhancement experiments prove that when compared to the EMD the alternative decomposition methods can lead to interesting speech quality and intelligibility improvement.**

***Keywords*— speech enhancement empirical mode decomposition, acoustic noise.**

I. Introdução

A supressão das distorções causadas por ruídos acústicos ambientais é de grande interesse para a área de processamento de voz. Com o objetivo de remover ou reduzir os efeitos causados pelos ruídos aditivos, a maioria das soluções de realce utilizam a a transformada de Fourier de tempo curto para estimar o espectro do ruído. Um dos principais desafios da área consiste em estimar as estatísticas dos ruídos acústicos reais, que podem ser oriundos de diferentes fontes (avião, balbúrdia, carro, trem). Adicionalmente, suas características podem variar ao longo do tempo, ou seja, os ruídos são não-estacionários.

Nos últimos anos, a decomposição empírica de modos (EMD - *empirical mode decomposition*) [1] tem sido utilizada para o realce de sinais de voz no domínio do tempo [2], [3], [4], [5]. O método EMD foi proposto como uma forma não-linear e adaptativa para análise tempo-frequência de sinais não-estacionários. Diferentemente das técnicas espectrais, o realce baseado no EMD não necessita da estimação explícita das estatísticas dos ruídos acústicos, nem de que os sinais analisados sejam estacionários. O método EMD resulta em um conjunto de funções intrínsecas de modo (IMF - *intrinsic mode functions*) que são totalmente dependentes do próprio sinal. Assim, a análise com EMD é adaptativa, o que garante a perfeita reconstrução do sinal pela soma dos modos obtidos na decomposição.

As técnicas de realce de sinais de voz baseadas no EMD são geralmente compostas por três etapas: decomposição do sinal ruidoso, seleção das IMFs mais afetadas pelo ruído, e reconstrução do sinal realçado com os modos restantes. As propostas de realce apresentadas em [2] aplicam filtros e limiares para selecionar e eliminar as IMFs com maior parcela de ruído. Contudo, estas propostas são limitadas a sinais de voz corrompidos por ruído Gaussiano branco. Já as propostas de pós-realce EMD-SRN (*EMD-based suppression of residual noise*) [3] e EMDF (*EMD-based filtering*) [4] foram aplicadas sobre sinais previamente tratados por técnicas espectrais. Ambas identificam as IMFs mais corrompidas baseadas em um estudo dos seus valores de variância. Recentemente, a técnica EMDH [5] adotou o expoente de Hurst (H) para selecionar, quadro a quadro, as IMFs que são mais corrompidas pelo ruído acústico. A proposta EMDH mostrou-se capaz de prover ganho de qualidade e inteligibilidade para sinais de voz corrompidos por ruídos altamente não-estacionários.

Este artigo investiga o realce de sinais de voz com dois métodos recentemente propostos como alternativa ao EMD: CEEMDAN (complete ensemble EMD with adaptive noise) [6] e SeqVMD (sequential variational modal decomposition) [7]. A decomposição CEEMDAN adiciona ruído Gaussiano branco ao sinal para reduzir os efeitos de mistura entre modos (*mode mixing*) do EMD. Já o SeqVMD utiliza ferramentas de otimização para obter as IMFs sem a necessidade da computação de envoltórias, conforme ocorre no EMD original. O principal objetivo do presente trabalho é verificar se a adoção destes métodos é capaz de melhorar a supressão dos ruídos acústicos da proposta de realce EMDH.

As técnicas de realce são avaliadas em experimentos com sinais de voz corrompidos por quatro ruídos acústicos não-estacionários. Os sinais de voz realçados são avaliados em termos de qualidade e inteligibilidade utilizando três medidas objetivas. Os resultados demonstram que o método SeqVMD leva aos melhores ganhos de qualidade para a maioria das situações ruidosas. Quanto à inteligibilidade, os três métodos de decomposição alcançam as melhores taxas de acertos em condições distintas de ruídos.

O restante deste trabalho está organizado da seguinte forma. A Seção II descreve o método EMD e as suas variações CEEMDAN e SeqVMD. Na Seção III, são apresentadas as técnicas de realce baseadas na decomposição EMD, além das medidas de qualidade e inteligibilidade aqui utilizadas. Os resultados dos experimentos de realce de voz são discutidos na Seção IV. Finalmente, a Seção V conclui o presente artigo.

L. Zão e R. Coelho, Laboratório de Processamento de Sinais Acústicos, Departamento de Engenharia Elétrica, Instituto Militar de Engenharia (IME), Rio de Janeiro, Brasil, E-mails: zao@ime.eb.br, coelho@ime.eb.br. Este trabalho foi parcialmente financiado pelo CNPq 304254/2012-6.

II. Decomposição Empírica de Modos

Considere um sinal $y(t)$ 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 detalhe $d(t)$, identifica-se uma componente de tendência local ou resíduo $r(t)$, tal que

$$y(t) = d(t) + r(t)\,,\ t_- \leq t \leq t_+\,. \tag{1}$$

O primeiro modo ($\text{IMF}_1(t)$) é definido pelo conjunto das componentes de detalhes, quando a decomposição é aplicada sobre todo o sinal $y(t)$. O sinal residual é dado pelo conjunto de todas as componentes de tendência local. Aplicando-se repetidamente o procedimento sobre o sinal residual, chega-se a um conjunto de IMFs e a um resíduo de baixas frequências.

A. Algoritmo EMD

O algoritmo para o método EMD aplicado sobre um sinal $y(t)$ pode ser dividido nos seguintes passos [1] [8]:

1) Identificar todos os extremos de $y(t)$, ou seja, os pontos de máximo $y_{max}(t)$ e mínimo $y_{min}(t)$ locais;
2) Obter as envoltórias $e_{max}(t)$ e $e_{min}(t)$, interpolando-se os pontos de máximo e de mínimo, respectivamente. Para isto, adota-se a interpolação polinomial de terceiro grau utilizando o método de *splines*;
3) Calcular o resíduo como a média entre as envoltórias: $r(t) = \left(e_{min}(t) + e_{max}(t)\right)/2$;
4) Extrair as componentes de detalhes: $d(t) = y(t) - r(t)$;
5) Repetir a iteração sobre o sinal residual $r(t)$.

Por definição [1], o número de extremos e de cruzamentos em zero de cada IMF devem ser iguais ou diferir em uma unidade. Adicionalmente, o valor médio definido pelas envoltórias dos seus máximos e mínimos deve ser nulo. Se a componente de detalhes $d(t)$, extraída no passo (4) do algoritmo EMD, não obedecer às propriedades acima, os passos (1-4) são novamente efetuados, com $d(t)$ no lugar de $y(t)$. Este processo, denominado *sifting*, é repetido até garantir que a nova função $d(t)$ seja considerada uma IMF. O algoritmo EMD assegura que qualquer sinal $y(t)$ pode ser decomposto em um número finito (K) de iterações, e pode ser escrito como

$$y(t) = \sum_{k=1}^{K} \text{IMF}_k(t) + r(t)\,, \tag{2}$$

onde $\text{IMF}_k(t), 1 \leq k \leq K$, são as funções de detalhes $d(t)$ obtidas no passo (4) de cada iteração, e $r(t)$ é o sinal residual final decorrente da última iteração.

A redução no número de extremos de um modo para o próximo implica que, localmente, as primeiras IMFs possuem oscilações mais rápidas (altas frequências) que as IMFs de maior índice. Em [8], foi demonstrado que, quando aplicado sobre sinais representados por um processo estocástico fGn (*fractional Gaussian noise*), o método EMD decompõe o sinal em IMFs cujas componentes espectrais são equivalentes às saídas de um banco de filtros diádicos com sobreposição de bandas passantes. Na literatura, propostas alternativas [9], [6] têm utilizado a adição de processos fGn para garantir que uma mesma IMF não possua componentes com oscilações de escalas distintas, ou que IMFs distintas sejam compostas por variações semelhantes, fenômeno este conhecido como mistura entre modos (*mode mixing*).

B. CEEMDAN

O método CEEMDAN [6] é uma evolução da decomposição EEMD (*ensemble EMD*) [9] e utiliza sequências de ruído Gaussiano branco para corromper o sinal $y(t)$. Assim, o primeiro passo consiste em gerar um quantidade I de sinais corrompidos $y^i(t)$, $i = 1, \ldots, I$, pela adição de sequências de ruído Gaussiano branco $w^i(t)$ ao sinal original $y(t)$, i.e.,

$$y^i(t) = y(t) + w^i(t)\,. \tag{3}$$

A mistura entre modos é evitada pela estrutura semelhante a um banco de filtros diádicos das IMFs resultantes. Além disso, como as amostras do ruído adicionado são descorrelatadas, as componentes resultantes do ruído se cancelam quando a média dos modos é considerada como resultado final da decomposição.

A primeira IMF da decomposição CEEMDAN é definida por

$$\widetilde{\text{IMF}}_1(t) = \frac{1}{I}\sum_{i=1}^{I} \text{IMF}_1^i(t)\,, \tag{4}$$

onde $\text{IMF}_1^i(t)$, $i = 1, \ldots, I$, são obtidas pela aplicação do algoritmo EMD sobre cada versão do sinal corrompido $y^i(t)$. O primeiro resíduo é então calculado como $a_1(t) = y(t) - \widetilde{\text{IMF}}_1(t)$.

Em seguida, o primeiro resíduo também é corrompido com um conjunto de I sequências de ruído Gaussiano branco. Cada versão ruidosa de $a_1(t)$ é então decomposta utilizando o algoritmo EMD. A média amostral de todas as IMFs de índice 1 assim obtidas é então definida como a segundo modo $\widetilde{\text{IMF}}_2(t)$. Seja o operador $E_k\{\cdot\}$ cuja saída é o k-ésimo modo decorrente da aplicação do algoritmo EMD, a IMF e o resíduo de índice $k \geq 2$ são dados por

$$\widetilde{\text{IMF}}_k(t) = \frac{1}{I}\sum_{i=1}^{I} E_1\left\{a_{k-1}(t) + E_{k-1}\{w^i(t)\}\right\}\,, \tag{5}$$

$$a_k(t) = a_{k-1}(t) - \widetilde{\text{IMF}}_k(t)\,. \tag{6}$$

Este procedimento é iterativamente repetido até que o último resíduo, de ordem K, tenha menos de dois extremos. Daí, o sinal original pode ser reconstruído com as novas IMFs de maneira análoga à Eq. 2.

C. SeqVMD

O método SeqVMD [7] foi recentemente proposto como outra alternativa à decomposição EMD. Neste, não há necessidade da interpolação dos extremos originalmente presente no algoritmo EMD. A motivação consiste no fato que a interpolação dos pontos de máximos e mínimos locais gera propriedades indesejadas às envoltórias do sinal. Por exemplo, as envoltórias superior e inferior podem se cruzar.

O SeqVMD adota técnicas de otimização para decompor um sinal $y(t)$ em uma sequência de K iterações, resultando em um conjunto de K modos e um resíduo final. Em cada iteração, o último resíduo calculado é decomposto em uma nova sequência de detalhes $d_k(t)$ e tendência $a_k(t)$, tais que

$$a_{k-1}(t) \approx d_k(t) + a_k(t)\,, \quad k = 1, \ldots, K\,. \tag{7}$$

onde $a_0(t) = y(t)$.

O algoritmo SeqVMD obtém a decomposição resolvendo sequencialmente o seguinte problema de otimização

$$(a_k(t), d_k(t)) \in \arg\min_{(a,d)} \|a_{k-1} - d - a\|_2^2\,, \quad k = 1, \ldots, K\,. \tag{8}$$

As seguintes condições são impostas às componentes $a(t)$ and $d(t)$ na Eq. 8 para garantir que elas correspondem às sequências de detalhes e tendência da decomposição EMD:

1) $d(t)$ deve possuir envoltórias com média nula.
2) $a(t)$ deve ser composta por oscilações mais suaves do que $d(t)$.
3) $d(t)$ e $d_j(t)$, para $1 \leq j < k$, devem ser aproximadamente ortogonais.

Seja $(t_k[l])_{1 \leq l \leq L_k}$ a localização dos extremos do k-ésimo resíduo $a_{k-1}(t)$. A primeira condição é aproximada por

$$\left| d(t_k[l]) + \frac{\alpha_l\, d(t_k[l-1]) + \beta_l\, d(t_k[l+1])}{\alpha_l + \beta_l} \right| < \epsilon_{k,l}\,, \tag{9}$$

onde $\alpha_l = t_k[l+1] - t_k[l]$, $\beta_l = t_k[l] - t_k[l-1]$ e $\epsilon_{k,l} > 0$.

Para a segunda condição, adota-se

$$\|Aa(t)\|_p^p \leq \nu_k, \quad k = 1, \ldots, K\,, \tag{10}$$

onde A é o operador diferencial, $p \geq 1$ e $\nu_k > 0$.

Finalmente, a terceira condição imposta pelo SeqVMD é atingida considerando, para cada $k = 1, \ldots, K$,

$$\| \langle d(t), d_j(t) \rangle \|_p^p \leq \zeta_{k,j}, \quad j < k\,, \tag{11}$$

com $\zeta_{k,j} > 0$. Isto assegura que $d(t)$ é aproximadamente ortogonal a todas as sequências de detalhes anteriores. A solução do problema de otimização definido pela Eq. 8 condicionado às Eqs. 9-11 é resolvido com a solução proposta em [10].

III. Realce de Voz

Neste trabalho, os métodos de decomposição CEEMDAN e SeqVMD são avaliados como alternativa ao algoritmo EMD em propostas de realce de sinais de voz. O principal objetivo é verificar se a redução nos efeitos de *mode mixing* e a solução para o cálculo das envoltórias levam a melhores resultados de realce. Nesta Seção, são apresentadas as técnicas de realce e as medidas objetivas de qualidade e inteligibilidade de voz que são utilizadas nos experimentos de avaliação.

A. Técnicas de Realce baseadas no EMD

A técnica de realce EMDH foi proposta em [5] para reduzir o efeito de ruídos acústicos não-estacionários. Para isto, o sinal de voz corrompido é primeiramente dividido em segmentos de curta duração. Em seguida, o expoente de Hurst (H) [11] é estimado para identificar as IMFs mais corrompidas pelas componentes de baixas frequências dos ruídos. Os demais modos são então utilizados na reconstrução de cada um dos quadros do sinal de voz realçado.

A técnica de realce EMDH pode ser resumida nas seguintes etapas:

1) Decomposição do sinal de voz $y(t)$ com o método EMD;
2) Segmentação de cada IMF em Q quadros de curta duração, denotados por w-IMF, i.e.,
$$\text{w-IMF}_{k,q}(t) = \begin{cases} \text{IMF}_k(t + qT_d) & , t \in [0, T_d]\,, \\ 0 & , \text{elsewhere}, \end{cases} \tag{12}$$
onde $q \in \{0, \ldots, Q-1\}$ é o índice dos quadros e T_d é a quantidade de amostras de cada quadro;
3) Para cada quadro q, compor um vetor K-dimensional $\mathbf{H}_q(k)$ com os valores estimados do expoente de Hurst de cada modo w-IMF$_{k,q}(t), k = 1, \ldots, K$;
4) Determinar, para cada quadro q, o índice N_q da última IMF cujo expoente de Hurst é menor que um determinado limiar H_{th}, i.e., $\mathbf{H}_q(N_q) < H_{\text{th}}$;
5) Reconstrução dos quadros $\hat{y}_q(t)$ do sinal de voz realçado
$$\hat{y}_q(t) = \sum_{k=1}^{N_q} \text{w-IMF}_{k,q}(t),\ q = 0, \ldots, Q-1\,. \tag{13}$$
Finalmente, os quadros são concatenados para formar o sinal de voz realçado $\hat{y}(t)$.

No passo (3), o expoente de Hurst é estimado com o método baseado em wavelets [12], utilizando filtros Daubechies [13] com 12 coeficientes. Assim como na proposta do EMDH [5], o limiar $H_{\text{th}} = 0,9$ é definido neste trabalho para selecionar as componentes de baixas frequências dos ruídos acústicos.

Para avaliar os métodos CEEMDAN e SeqVMD para realce dos sinais de voz, estas propostas de decomposição são adotadas na etapa (1) do realce EMDH como alternativa ao algoritmo EMD. As técnicas resultantes destas substituições são aqui denominadas CEEMDAN-H e SeqVMD-H, respectivamente. Todas as demais etapas de realce são mantidas inalteradas.

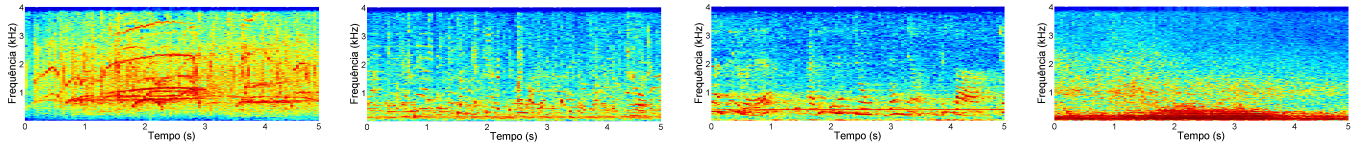
B. Medidas Objetivas de Qualidade e Inteligibilidade

Três medidas objetivas são utilizadas para comparar as técnicas de realce. A razão sinal-ruído segmental (SegSNR - *segmental signal-to-noise ratio*) é calculada no domínio do tempo a partir dos sinais de voz limpos e corrompidos, O aumento de SegSNR é geralmente associado a melhoria na qualidade da voz. Os índices CSII (*coherence speech intelligibility index*) [14] e STOI (*short-time objective intelligibility*) [15] são adotados para predição das taxas de acertos de palavras em testes subjetivos de inteligibilidade.

1) SegSNR: Seja $x(t)$ um sinal de voz limpo, e $y(t)$ uma versão corrompida deste mesmo sinal, o valor de SegSNR de $y(t)$ é estimado por:

$$\text{SegSNR} = \frac{10}{Q} \sum_{\tau=0}^{Q-1} \log \frac{\sum_{t=\tau T_{\text{sh}}}^{\tau T_{\text{sh}} + T_d - 1} x^2(t)}{\sum_{t=\tau T_{\text{sh}}}^{\tau T_{\text{sh}} + T_d - 1} \left[x(t) - y(t)\right]^2}\,, \tag{14}$$

onde T_d representa a quantidade de amostras de cada quadro, T_{sh} é o deslocamento entre quadros consecutivos e Q é o total de quadros.



(a) (b) (c) (d)

Fig. 1. Espectrogramas de segmentos dos quatro ruídos acústicos: (a) Aplausos, (b) Mercado, (c) Plataforma, e (d) Tráfego.

TABELA I

OS QUATRO RUÍDOS ACÚSTICOS UTILIZADOS NOS EXPERIMENTOS.

Ruído	Descrição
Aplausos	Aplausos e assovios ao final de uma apresentação
Mercado	Algumas pessoas conversando em um mercado
Plataforma	Estação ferroviária com anúncio pelo sistema de som
Tráfego	Carros trafegando por uma rua movimentada

2) CSII: A medida CSII foi proposta como uma extensão do índice de inteligiiblidade de voz (SII - *speech intelligibility index*), padrão ANSI S3.5-1997. O SII é calculado pela média ponderada de valores de SNR calculadas em todas as bandas de frequência do sinal. Para aumentar a correlação com os resultados de inteligibilidade, a medida CSII é avaliada separadamente para segmentos de baixo, médio e alto níveis de potência (CSII_B, CSII_M e CSII_A, respectivamente). O valor final de CSII é então definido como $\text{CSII} = 0.155\text{CSII}_\text{B} + 0.845\text{CSII}_\text{M} + 0.0\text{CSII}_\text{A}$ [14].

3) STOI: O STOI foi proposto como um método baseado em correlação para avaliar a degradação da inteligibilidade causado por técnicas de realce de sinais de voz. Os sinais de voz limpo e corrompido são divididos em quadros e sub-bandas de frequência. A correlação entre as magnitudes das sub-bandas destes sinais é avaliada de um conjunto de vários quadros consecutivos. O STOI é então dado pela média dos valores de correlação obtidos de todos os quadros e todas as sub-bandas. Em [15], a medida STOI mostrou alta correlação com os resultados de inteligibilidade de testes reais.

IV. EXPERIMENTOS REALIZADOS

Esta Seção apresenta a descrição e os resultados dos experimentos de realce de sinais de voz. A avaliação das técnicas de realce EMDH, SeqVMD-H e CEEMDAN-H utilizou um conjunto de 24 locutores (16 homens e 8 mulheres) da base de voz TIMIT [16]. As locuções possuem duração média de 3 segundos e taxa de amostragem de 16 kHz. Em todos os experimentos o realce foi realizado com quadros de 32 ms.

Quatro ruídos acústicos não-estacionários foram utilizados para corromper as locuções de voz. Os ruídos são descritos na Tab. I e foram coletados das bases Freesound.org (Aplausos e Trânsito) e Freesfx.co.uk (Mercado e Plataforma). A adição foi realizada considerando valores de SNR entre -10 dB e 10 dB, com intervalos de 5 dB. A Fig. 1 ilustra o espectrograma de segmentos destes ruídos, com frequência no intervalo 0-4 kHz. Note que os quatro ruídos apresentam variações no espectro de frequência, ou seja, são não-estacionários.

A. Resultados de SegSNR

A Fig. 2 ilustra os incrementos de SegSNR obtidos com as três técnicas de realce para os quatro ruídos acústicos. Os

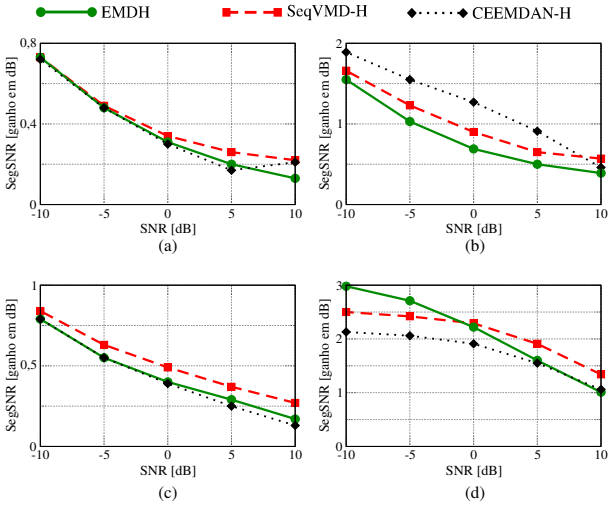


Fig. 2. Ganho de SegSNR obtido com locuções corrompidas pelos quatro ruídos acústicos: (a) Aplausos, (b) Mercado, (c) Plataforma, e (d) Tráfego.

valores correspondem à diferença entre a medida auferida com as locuções realçada e corrompida. Ou seja, incremento de SegSNR corresponde a aumento na qualidade do sinal de voz. Note que os ruídos Mercado e Tráfego levam aos melhores resultados de SegSNR para as três técnicas, atingindo ganho próximo de 2 dB e 3 dB, respectivamente. Este fato pode ser explicado pela maior concentração de energia em baixas frequências destes ruídos (veja a Fig. 1). Por outro lado, os resultados com os ruídos Aplausos e Plataforma são sempre menores que 1 dB.

A técnica SeqVMD-H apresentou o maior ganho de qualidade para três diferentes fontes de ruídos: Aplausos (SNR $\geq$ 0 dB), Plataforma e Tráfego (SNR $\geq$ 0 dB). O realce com CEEMDAN obteve o melhor resultado para o ruído Mercado considerando os valores de SNR $\leq$ 5 dB. Para SNR de 10 dB, o maior incremento foi alcançado pela técnica SeqVMD-H. Já a proposta original EMDH superou as demais técnicas para o ruído Tráfego com SNR $<$ 0 dB. Resultados semelhantes foram apresentados pelas três técnicas de realce para o ruído Aplausos com SNR $<$ 0 dB.

B. Resultados de Predição de Inteligibilidade

Neste trabalho, as medidas CSII e STOI são utilizadas para obter uma predição das taxas de acertos de palavras em testes de inteligibilidade. Para isto, os resultados são transformados por uma função da forma

$$f(d) = 100/(1 + \exp(a\,d + b))\,, \qquad (15)$$

onde d representa a medida correspondente, e a e b são coeficientes constantes. De forma a obter resultados de inteligibilidade similares aos apresentados em [17], os coeficientes

TABELA II

RESULTADOS DE PREDIÇÃO DE TAXA DE ACERTOS DE PALAVRAS (%) OBTIDOS COM A MEDIDA CSII.

Ruído Aplausos			SNR	Ruído Mercado		
EMDH	SeqVMD-H	CEEMDAN-H	(dB)	EMDH	SeqVMD-H	CEEMDAN-H
85,8	**86,0**	85,9	10	84,6	84,6	**84,8**
65,0	**65,4**	63,6	5	53,7	53,8	**54,2**
34,4	**34,9**	33,1	0	21,1	21,1	**21,4**
13,1	**13,3**	12,6	-5	6,4	6,4	**6,5**
5,2	5,2	4,9	-10	2,1	2,1	2,1
40,7	**41,0**	40,0	Média	33,6	33,6	**33,8**
Ruído Plataforma			SNR	Ruído Tráfego		
EMDH	SeqVMD-H	CEEMDAN-H	(dB)	EMDH	SeqVMD-H	CEEMDAN-H
91,4	91,4	91,2	10	99,2	99,2	99,2
70,1	70,1	69,6	5	98,6	98,6	**98,7**
34,5	34,4	34,0	0	96,5	96,5	**96,7**
11,3	11,3	11,2	-5	87,1	87,4	**87,6**
3,7	3,6	3,6	-10	60,1	60,8	60,8
42,2	42,2	41,9	Média	88,3	88,5	**88,6**

TABELA III

RESULTADOS DE PREDIÇÃO DE TAXA DE ACERTOS DE PALAVRAS (%) OBTIDOS COM A MEDIDA STOI.

Ruído Aplausos			SNR	Ruído Mercado		
EMDH	SeqVMD-H	CEEMDAN-H	(dB)	EMDH	SeqVMD-H	CEEMDAN-H
94,5	94,4	93,7	10	86,2	86,3	**86,7**
87,6	87,6	81,7	5	62,7	62,7	**63,6**
70,0	69,9	60,4	0	**27,2**	27,1	26,5
39,4	39,4	30,0	-5	7,3	7,3	6,8
12,8	**12,9**	8,7	-10	2,0	2,0	1,8
60,9	60,8	54,9	Média	37,1	37,1	37,1
Ruído Plataforma			SNR	Ruído Tráfego		
EMDH	SeqVMD-H	CEEMDAN-H	(dB)	EMDH	SeqVMD-H	CEEMDAN-H
91,6	**91,7**	89,6	10	**95,3**	95,2	95,1
79,1	**79,2**	75,6	5	**93,1**	92,7	92,6
50,9	50,9	46,8	0	**89,1**	88,6	88,3
19,3	19,2	17,7	-5	**81,9**	81,0	79,9
5,2	5,1	4,9	-10	**67,1**	66,5	63,4
49,2	49,2	46,9	Média	**85,3**	84,8	83,8

são determinados em testes preliminares como $a = -10,088$ e $b = 4,654$ para o índice CSII, e $a = -13,45$ e $b = 9,36$ para a medida STOI.

A Tab. II apresenta as taxas de predição de inteligibilidade com o índice CSII. Os números em destaque correspondem ao maior valor, quando este ocorre, para uma mesma condição de ruído. Os melhores resultados são novamente obtidos com o ruído Tráfego, enquanto as piores taxas são agora em consequência do ruído Mercado. É possível observar que as taxas de acertos com as locuções realçadas pelas três técnicas são similares. Assim como na Fig. 2, os melhores resultados para os ruídos Aplausos e Mercado são alcançados pelas técnicas SeqVMD-H e CEEMDAN-H, respectivamente. A técnica CEEMDAN-H também supera as demais para o ruído Tráfego. Já para o ruído Plataforma, a técnica EMDH obtém isoladamente as maiores taxas de acertos para SNR de 0 dB e -10 dB.

Os resultados de predição de taxas de acertos com a medida STOI são mostrados na Tab. III. Neste caso, a técnica CEEMDAN-H não obteve resultados próximos às técnicas EMDH e SeqVMD-H para a maioria dos ruídos. A única exceção é o ruído Mercado, onde as três técnicas apresentaram desempenho semelhante. Veja que, na média, a técnica EMDH superou as demais para os ruídos Aplausos e Tráfego. Já para o ruído Plataforma, as técnicas EMDH e SeqVMD-H apresentaram taxas médias de acertos similares.

V. CONCLUSÃO

Este artigo investigou o uso de variações da decomposição EMD para técnicas de realce de voz no domínio do tempo. Estes métodos de decomposição foram originalmente propostos para evitar o efeito de *mode mixing* e o cálculo de envoltórias do EMD original. Neste trabalho, eles foram avaliados como alternativa ao EMD na proposta de realce EMDH. Os experimentos de realce foram realizados com sinais de voz corrompidos por quatro ruídos acústicos não-estacionários. Os resultados demonstraram que o realce com a decomposição SeqVMD leva ao maior ganho de qualidade para a maioria das situações de ruído. Com relação à inteligibilidade, os resultados de CSII e STOI demonstram que as decomposições SeqVMD e EMD alcançam as melhores taxas de acertos para distintas condições de ruídos. Já a decomposição CEEMDAN atinge os maiores valores de CSII para dois ruídos acústicos.

REFERÊNCIAS

[1] N. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, N. Yen, C. Tung, and 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.

[2] K. Khaldi, A. Boudraa, A. Bouchikhi, and M. Alouane, "Speech enhancement via EMD," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, p. 873204, May 2008.

[3] T. Hasan and M. Hasan, "Suppression of residual noise from speech signals using empirical mode decomposition," *IEEE Signal Processing Letters*, vol. 16, pp. 2–5, January 2009.

[4] 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.

[5] L. Zão, R. Coelho, and P. Flandrin, "Speech enhancement with EMD and Hurst-based mode selection," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, pp. 899–911, May 2014.

[6] M. Colominas, G. Schlotthauer, M. Torres, and P. Flandrin, "Noise-assisted EMD methods in action," *Advances in Adaptive Data Analysis*, vol. 04, no. 04, p. 1250025, 2012.

[7] N. Pustelnik, P. Borgnat, and P. Flandrin, "Empirical mode decomposition revisited by multicomponent non-smooth convex optimization," *Signal Processing*, vol. 102, pp. 313–331, 2014.

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

[9] Z. Wu and N. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.

[10] L. Briceño Arias and P. Combettes, "A monotone + skew splitting model for composite monotone inclusions in duality," *SIAM Journal on Optimization*, vol. 21, pp. 1230–1250, October 2011.

[11] E. Hurst, "Long-term storage capacity of reservoirs," *Transactions of the American Society of Civil Engineers*, pp. 770–799, April 1951.

[12] D. Veitch and P. Abry, "A wavelet-based joint estimator of the parameters of long-range dependence," *IEEE Transactions on Information Theory*, vol. 45, pp. 878 –897, April 1999.

[13] I. Daubechies, *Ten lectures on wavelets*. Philadelphia, USA: Society for Industrial and Applied Mathematics, 1992.

[14] J. Kates and K. Arehart, "Coherence and the speech intelligibility index," *The Journal of the Acoustical Society of America*, vol. 117, pp. 2224–2237, April 2005.

[15] C. Taal, R. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, pp. 2125–2136, September 2011.

[16] 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.

[17] P. Loizou and Y. Hu, "A comparative intelligibility study of single-microphone noise reduction algorithms," *The Journal of the Acoustical Society of America*, vol. 22, pp. 1777–1786, September 2007.