

Um Método Baseado em Aprendizado Profundo para Cancelamento de Realimentação Acústica

Fabiano T. Castoldi e Bruno C. Bispo

Resumo—Este trabalho apresenta um método baseado em aprendizado de máquina para o cancelamento de realimentação acústica em sistemas de sonorização. A abordagem proposta emprega representações tempo-frequência dos sinais do microfone e do alto-falante, e redes neurais profundas para estimar dinamicamente o caminho de realimentação. Essa estimativa é integrada ao processo de atualização de um filtro adaptativo responsável por mitigar o efeito de realimentação no sistema. O método proposto alcançou convergência e reconvergência entre 0,1 e 1 s, dependendo do número de janelas temporais utilizadas na estimação, e erros em regime permanente entre -8 e -14 dB. Os resultados demonstram que a identificação do caminho de realimentação acústica por meio de redes profundas constitui uma alternativa promissora frente aos métodos clássicos baseados exclusivamente em filtragem adaptativa linear.

Palavras-Chave—Cancelamento de realimentação acústica, cancelamento profundo de realimentação, redes recorrentes, sistema de sonorização.

Abstract—This paper presents a machine learning-based method for acoustic feedback cancellation in sound reinforcement systems. The proposed approach employs time-frequency representations of both microphone and loudspeaker signals, and deep neural networks to dynamically estimate the feedback path. This estimate is subsequently integrated into the update process of an adaptive filter responsible for mitigating the feedback effect in the system. The proposed method achieved convergence and reconvergence between 0,1 and 1 s, depending on the number of time frames used in the estimation, with steady-state errors ranging between -8 and -14 dB. The results demonstrate that acoustic feedback path identification via deep networks constitutes a promising alternative to classical methods based exclusively on linear adaptive filtering.

Keywords—Acoustic feedback cancellation, deep feedback cancellation, recurrent neural networks, sound reinforcement system.

I. INTRODUÇÃO

Em sistemas de sonorização, o sinal de áudio capturado por um ou mais microfones é amplificado e reproduzido por alto-falantes no mesmo ambiente acústico [1]. Esses sistemas estão presentes em salas de aula e auditórios, carros, aparelhos auditivos, entre outros. O acoplamento acústico entre o alto-falante e o microfone faz o sinal reproduzido pelo alto-falante retornar ao microfone, gerando uma malha de realimentação.

A realimentação acústica afeta o desempenho do sistema de duas maneiras. Em primeiro lugar, o sistema em malha fechada pode apresentar instabilidade, podendo resultar em

oscilações acusticamente percebidas como apito, fenômeno também conhecido como efeito Larsen. Como consequência, a amplificação alcançável é limitada [1], [2]. Em segundo, a qualidade do som é deteriorada por reverberação excessiva [1].

A fim de evitar o efeito Larsen e aumentar a amplificação alcançável, os métodos de cancelamento de realimentação acústica (AFC, do inglês *acoustic feedback cancellation*) utilizam um filtro adaptativo para estimar o sinal realimentado e subtraí-lo do sinal do microfone [3]. Porém, os algoritmos tradicionais de filtragem adaptativa, que aproximam a solução de Wiener através do gradiente ou mínimos quadrados, apresentam desempenho insatisfatório porque os sinais que agem como entrada e interferência para o filtro são correlacionados [3].

Recentemente, técnicas de aprendizado de máquina têm sido exploradas para desenvolver métodos alternativos de supressão da realimentação acústica. Em [4], foi proposto um método que estima o sinal acústico de interesse diretamente a partir do sinal do microfone, suprimindo as distorções causadas pela realimentação. Em [5], foi desenvolvido um método que estima o caminho de realimentação para auxiliar os algoritmos tradicionais de filtragem adaptativa. Em [6], com a nomenclatura DFC (do inglês *deep feedback cancellation*), foi proposto um método para estimar o caminho de realimentação em substituição aos algoritmos de filtragem adaptativa linear.

Os métodos propostos em [5] e [6] se destacam por preservarem a estrutura AFC e seu potencial de anular completamente o efeito da realimentação acústica, buscando melhorar a robustez e velocidade de convergência dos métodos clássicos.

O presente trabalho tem como objetivo implementar e avaliar um método DFC, utilizando como referência a arquitetura de aprendizado profundo desenvolvida em [6]. Uma alteração na função custo do modelo e uma metodologia para aumentar a quantidade de respostas ao impulso do caminho de realimentação, a partir da manipulação de medições reais, são propostas. Este artigo está organizado da seguinte maneira: a Seção II formula o problema e descreve o modelo DFC proposto; a Seção III descreve a configuração das simulações computacionais realizadas e os sinais empregados; na Seção IV são apresentados e discutidos os resultados obtidos; por fim, a Seção V conclui o estudo.

II. METODOLOGIA PROPOSTA

Essa seção apresenta a formulação do problema e a proposta de modelo baseado em aprendizado profundo para AFC.

A. Realimentação Acústica

A Figura 1 apresenta o diagrama de blocos de um sistema de sonorização. O sinal acústico de interesse (ou sinal de entrada), o sinal do microfone, o sinal do alto-falante e o sinal de

Fabiano T. Castoldi, Universidade Federal do Pampa, Alegrete-RS, Brasil, e Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Santa Catarina, Florianópolis-SC, Brasil. Bruno C. Bispo, Departamento de Engenharia Elétrica e Eletrônica, Universidade Federal de Santa Catarina, Florianópolis-SC, Brasil. E-mails: fabianocastoldi@unipampa.edu.br, bruno.bispo@ufsc.br.

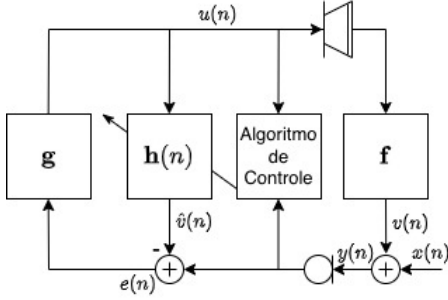


Fig. 1: Cancelamento de realimentação acústica.

realimentação são representados por $x(n)$, $y(n)$, $u(n)$ e $v(n)$ respectivamente, onde n denota o índice de tempo discreto. Por conveniência, assume-se que $x(n)$ inclui as características do microfone e conversor A/D.

O caminho direto modela o sistema de amplificação e qualquer dispositivo de processamento de sinal inserido nesta parte do sistema. Assumindo linearidade e invariância no tempo, esse caminho é representado pela resposta ao impulso

$$\mathbf{g} = [g_0 \ g_1 \ \dots \ g_{G-1}]. \quad (1)$$

O caminho de realimentação modela o acoplamento acústico entre alto-falante e microfone, e inclui as características do conversor D/A, alto-falante, microfone e conversor A/D. Assumindo linearidade e invariância no tempo, o caminho de realimentação é representado pela resposta ao impulso

$$\mathbf{f} = [f_0 \ f_1 \ \dots \ f_{F-1}]. \quad (2)$$

O filtro adaptativo que identifica o caminho de realimentação acústica é representado pela resposta ao impulso

$$\mathbf{h}(n) = [h_0(n) \ h_1(n) \ \dots \ h_{H-1}(n)]. \quad (3)$$

O sinal de realimentação $v(n)$ é estimado como $\mathbf{h}(n) * u(n)$ e subtraído do sinal do microfone $y(n)$, gerando o sinal com realimentação compensada [1]

$$\begin{aligned} e(n) &= y(n) - \mathbf{h}(n) * x(n) \\ &= x(n) + [\mathbf{f} - \mathbf{h}(n)] * u(n), \end{aligned} \quad (4)$$

que é efetivamente o sinal fornecido ao caminho direto.

B. Modelo DFC

O diagrama do modelo DFC proposto é ilustrada na Figura 2, sendo uma variação do modelo descrito em [6]. As entradas do modelo são representações espectrais de tempo curto dos sinais $u(n)$ e $y(n)$, enquanto a saída é uma estimativa de $\mathbf{f}$. A cadeia de processamento é composta por duas etapas: extração de características e aprendizado profundo.

1) *Extração de Características*: As transformadas de Fourier de tempo curto (STFT, do inglês *short-time Fourier transform*) de $u(n)$ e $y(n)$ são representadas por $U(k, m)$ e $Y(k, m)$, respectivamente, onde $k = 0, \dots, K-1$ é o bin de frequência e $m = 0, \dots, M-1$ é o índice da janela temporal.

Os espectros de tempo curto são então normalizados como $\bar{U}(m, k) = U(m, k)/\sqrt{E_U(m)}$ e $\bar{Y}(m, k) = Y(m, k)/\sqrt{E_U(m)}$, onde $E_U(m) = \sum_{k=0}^{K-1} |U(m, k)|^2$, para tornar o modelo robusto à variações de energia ao longo do tempo.

Na sequência, matrizes de representação tempo-frequência de cada sinal são criadas por meio da concatenação da magnitude logarítmica e fase espectrais, resultando em $\mathbf{U}_{inp} = [\bar{\mathbf{U}}_{mag} \ \bar{\mathbf{U}}_{arg}]$, onde $\bar{U}_{mag}(m, k) = \log_{10}\{|\bar{U}(m, k)|\}$ e $\bar{U}_{arg}(m, k) = \arg\{\bar{U}(m, k)\}$, e $\mathbf{Y}_{inp}$ análogo. Desse modo, cada matriz característica é real e tem dimensão $M \times 2K$.

2) *A Rede DFC*: As duas matrizes características formam um tensor, com dimensão $M \times 2K \times 2$, empregado como entrada para uma camada convolucional bidimensional. O objetivo dessa camada é capturar padrões locais no domínio tempo-frequência. A convolução utiliza um *kernel* de tamanho 4×5 sobre tempo e frequência, respectivamente, produzindo a saída $\mathbf{C}$ com dimensão $M \times 2K$. Para garantir essa dimensão na saída, utiliza-se a operação de *zero-padding* no eixo espectral de forma simétrica, e no eixo temporal com *padding* causal, garantindo que o processamento de cada segmento não utilize informações de janelas temporais futuras.

A saída da camada convolucional é concatenada com as duas matrizes características, formando a matriz $[\mathbf{C} \ \mathbf{U}_{inp} \ \mathbf{Y}_{inp}]$ de dimensão $M \times 6K$. Essa matriz é então aplicada a uma camada densa (FC, do inglês *fully connected*) com função de ativação *LeakyReLU*. O objetivo da camada densa é combinar informações espectrais de diferentes janelas temporais para capturar dependências tempo-frequenciais entre todas as características extraídas, enriquecendo as informações obtidas dos sinais $u(n)$ e $y(n)$.

Em seguida, uma camada recorrente do tipo LSTM (do inglês *long short-term memory*) é aplicada para processar a sequência de vetores ao longo das M janelas temporais, produzindo uma representação temporal que incorpora o histórico do sistema e consequentemente do caminho de realimentação. A dimensão do estado oculto da LSTM é fixada em 256 unidades, gerando uma saída de dimensões $M \times 256$.

As representações geradas pela LSTM são processadas por duas novas camadas densas. A primeira, com função de ativação *LeakyReLU*, tem como objetivo reduzir a dimensionalidade dos dados. A segunda, com função de ativação *tanh*, mapeia o espaço latente para uma estimativa da resposta ao impulso do caminho de realimentação em cada janela temporal, representada por $\mathbf{h}'_m \in \mathbb{R}^{M \times 64}$, $m = 0, \dots, M-1$.

Por fim, uma função de média móvel, implementada por uma camada *average pooling 1D*, é aplicada para suavizar os resultados obtidos, gerando o conjunto de estimativas $\mathbf{h}_n \in \mathbb{R}^{(M-(N-1)) \times 64}$, onde $\mathbf{h}_n = N^{-1} \sum_{i=n-N+1}^n \mathbf{h}'_i$ e N é um parâmetro que controla o tamanho da janela de suavização. Destaca-se que as estimativas suavizadas $\mathbf{h}_n$ são obtidas somente a partir da N -ésima janela dos sinais.

3) *Função de Custo*: O treinamento do modelo é realizado de forma supervisionada, utilizando como alvo a resposta ao impulso $\mathbf{f}$ do caminho de realimentação. A função custo é baseada na distância euclidiana normalizada (NESD, do inglês *normalized Euclidian system distance*) entre o alvo e a resposta ao impulso suavizada, a qual é definida na janela n como

$$\text{NESD}(n) = \frac{\|\mathbf{f} - \mathbf{h}_n\|^2}{\|\mathbf{f}\|^2}. \quad (5)$$

Essa métrica, de valor real positivo, também é conhecida na literatura de AFC como desalinhamento normalizado [2]. A

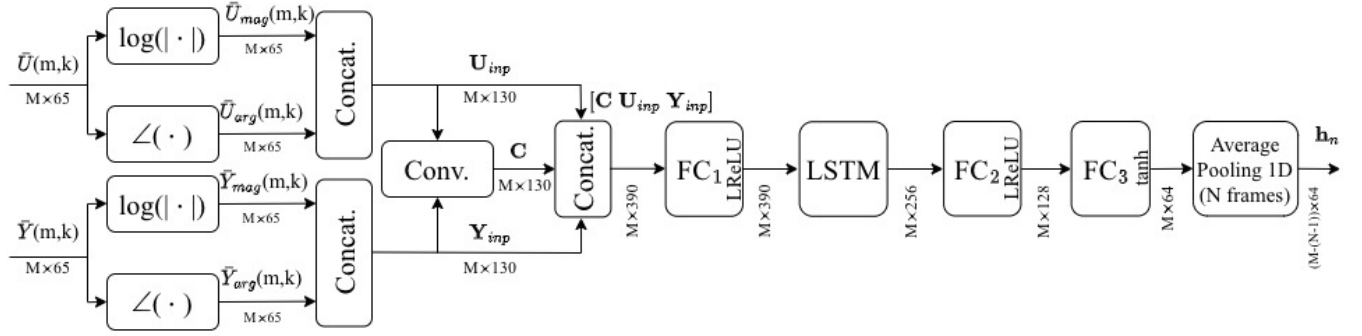


Fig. 2: Diagrama de blocos do modelo DFC proposto.

função custo é então definida como o NESD médio, dado por

$$\text{NESD} = \frac{1}{M - N + 1} \sum_{n=N}^M \text{NESD}(n). \quad (6)$$

Ressalta-se que (6) se difere do NESD utilizado em [6]. No referido estudo, a função custo utiliza $\text{NESD}(n)$ na escala decibel. Porém, como $\text{NESD}(n) \geq 0$ e geralmente $\text{NESD}(n) \leq 1$, os valores de $10 \log_{10}\{\text{NESD}(n)\}$ são tipicamente negativos e com magnitude elevada, o que pode provocar instabilidade numérica e gradientes desbalanceados durante o treinamento de modelos de aprendizado de máquina [7], [8]. Por outro lado, a função custo proposta no presente trabalho tende a satisfazer a condição desejada $0 \leq \text{NESD} \leq 1$.

III. CONFIGURAÇÃO DAS SIMULAÇÕES

Esta seção descreve a configuração das simulações realizadas para treinar e testar o modelo DFC proposto.

A. Caminhos de Avanço e Realimentação

a) *Avanço*: Como em [6], o caminho de avanço foi definido como um atraso de 10 ms e um ganho unitário, i.e.,

$$\mathbf{g} = [0 \ 0 \ \dots \ g_{G-1}], \quad (7)$$

onde $G = 160$ e $g_{G-1} = 1$.

b) *Realimentação*: O caminho de realimentação acústica foi modelado por respostas ao impulso de sala (RIR, do inglês *room impulse response*) reais e sintéticas. O emprego de RIRs sintéticas se deve pela pouca quantidade disponível de medições de RIRs de uma mesma sala e pela necessidade de uma quantidade elevada de RIRs para uma generalização adequada do modelo DFC. Abordagem semelhante foi utilizada em [6].

Foram utilizadas 222 RIRs reais de uma mesma sala, medidas com alto-falante e microfone instalados em posições diferentes, disponíveis em [9]. As RIRs tiveram sua taxa de amostragem reduzida para 16 kHz e em seguida foram truncadas em $F = 64$ amostras. Este comprimento de f é usualmente encontrado em aplicações de aparelhos auditivos [6], [10].

Um total de 778 RIRs artificiais foram geradas ao manipular controladamente as RIRs reais. Para cada seleção aleatória de uma RIR real, foram aplicadas entre uma e três, quantidade também escolhida aleatoriamente, das seguintes operações:

- Deslocamento por uma quantidade escolhida aleatoriamente entre -5 e 5 amostras, seguido de truncamento ou preenchimento com zeros para não alterar o comprimento.

- Modulação por uma exponencial $e^{\alpha n}$, onde α é uma constante escolhida aleatoriamente entre $-0,04$ e $0,04$.
- Adição de ruído branco gaussiano com média nula e desvio padrão igual a $0,002$.

Por fim, cada uma das 1000 RIRs foi aleatoriamente normalizada em amplitude de forma que o máximo ganho estável (MSG, do inglês *maximum stable gain*) [11] varie entre 2 e 10 dB, garantindo assim que o modelo DFC opere em uma condição de estabilidade do sistema de sonorização. Uma RIR real e duas RIRs sintéticas são exibidas na Figura 3.

As RIRs reais foram aleatoriamente separadas em conjuntos de treinamento, validação e teste na proporção 80%, 10% e 10%, respectivamente. Processo idêntico de divisão foi efetuado para as RIRs sintéticas. Desse modo, as RIRs do conjunto de validação não são vistas pelo modelo durante seu treinamento, assim como as RIRs do conjunto de teste não são vistas durante o treinamento e validação. Isso garante que, em relação às RIRs, não há vazamento de dados e a capacidade de generalização do método será avaliada. No total, 800 RIRs foram usadas no treinamento, 100 na validação e 100 no teste.

B. Sinais de Fala e de Entrada

a) *Fala*: A base TIMIT [12] foi utilizada para gerar os sinais de fala. Ela contém 6.300 áudios, contendo uma sentença cada, gravados digitalmente a uma taxa de amostragem de 16 kHz. Cada um dos locutores, de oito regiões dialetais (RD) dos EUA, pronunciou dez sentenças.

De cada áudio das RDs de 1 a 8, um trecho de fala com duração de 2 s foi extraído. Das RDs 1, 2, 3 e 4, 3530 trechos foram aleatoriamente selecionados e utilizados como o sinal $x(n)$ no treinamento do modelo. Das RDs 5 e 6, 1440 trechos foram aleatoriamente selecionados e utilizados como $x(n)$ na etapa de validação. Das RDs 7 e 8, trechos aleatoriamente selecionados foram concatenados para formar 10 sinais $x(n)$ com duração de 10 s para a etapa de teste do modelo.

Essa divisão garante que, também em relação aos sinais de fala, não há vazamento de dados e a capacidade de generalização do método DFC proposto será analisada.

b) *Entradas*: Apesar das entradas do método DFC terem sido definidas como os sinais $u(n)$ e $y(n)$, uma interpretação alternativa é que as entradas são o sinal de fala, caminho de avanço e caminho de realimentação. Dado um trio $\{x(n), \mathbf{g}, \mathbf{f}\}$, todos os sinais que circulam pelo sistema de sonorização podem ser calculados. Ademais, como apenas um caminho de avanço é usado, o trio se torna o par $\{x(n), \mathbf{f}\}$.

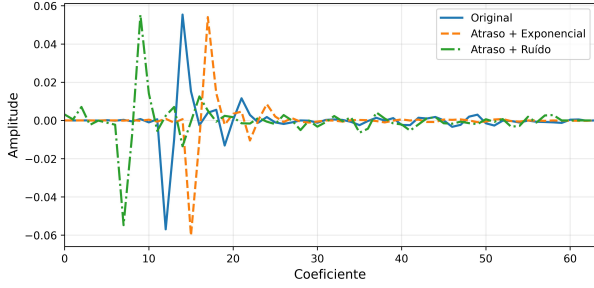


Fig. 3: Exemplos de uma RIR real e duas variantes sintéticas.

Com essa interpretação em mente, para a etapa de treinamento do modelo, 200 sinais $x(n)$ foram individualmente utilizados com cada $\mathbf{f}$, resultando em 160000 pares $\{x(n), \mathbf{f}\}$. Para a etapa de validação, 100 sinais $x(n)$ foram usados com cada $\mathbf{f}$, resultando em 10000 pares $\{x(n), \mathbf{f}\}$.

Por outro lado, para a etapa de teste, os 10 sinais $x(n)$, com duração de 10 s cada, foram empregados em conjunto com pares de $\mathbf{f}$, resultando em 500 trios $\{x(n), \mathbf{f}_1, \mathbf{f}_2\}$. Como será descrito mais à frente, as simulações de teste foram realizadas com o caminho de realimentação inicialmente definido por $\mathbf{f}_1$ e abruptamente alterado para $\mathbf{f}_2$ após o decorrer de 5 s.

C. Treinamento e Validação do Modelo

As etapas de treinamento e validação do modelo utilizaram as mesmas configurações, com exceção dos pares $\{x(n), \mathbf{f}\}$, pois cada etapa possui um conjunto designado de dados.

Ressalta-se que, nestas etapas, a partir do par $\{x(n), \mathbf{f}\}$, os sinais $y(n)$ e $u(n)$ correspondentes foram gerados ao executar o sistema de sonorização sem filtro adaptativo. Desse modo, o modelo DFC proposto é treinado para estimar $\mathbf{f}$ sem considerar os efeitos do cancelamento da realimentação acústica ao longo do tempo. Essa estratégia de treinamento foi inspirada em [6]. Sua motivação é que o modelo não apresenta perdas significativas de desempenho, enquanto o processo de treinamento se torna substancialmente mais rápido [6].

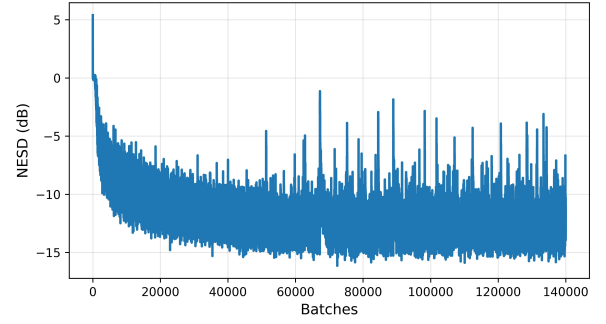
O treinamento foi implementado usando processamento em *mini-batches* de tamanho 16 sinais e o algoritmo de otimização Adam com taxa de aprendizado de 0,001. O modelo foi treinado por até 40 épocas, utilizando o critério de *early stopping* ajustado para interromper o treinamento após 8 épocas consecutivas sem melhora na NESD de validação.

As STFTs $X(m, k)$ e $U(m, k)$ foram calculadas usando uma janela de Hann, com comprimento de 128 amostras e sobreposição de 75%, e a transformada rápida de Fourier (FFT, do inglês *fast Fourier transform*) com 128 pontos, resultando em $K = 65$ e $M = 997$. Esses valores de K e M levam a aproximadamente 656 mil parâmetros treináveis, como em [6].

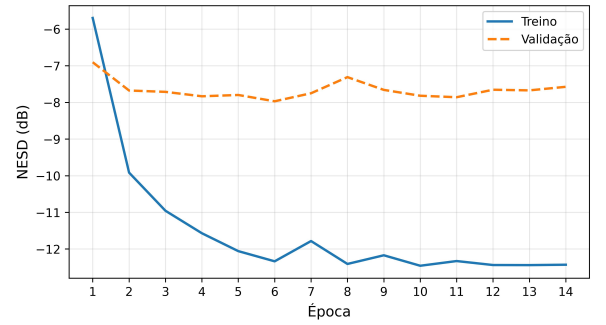
Para o processo de suavização das respostas ao impulso na saída do modelo, optou-se por $N = 50$ para obter resultados comparáveis aos apresentados em [6]. Observa-se, no entanto, que N é um parâmetro importante para controlar a convergência e o erro médio em regime permanente.

D. Procedimento de Teste

Na etapa de teste, para cada trio $\{x(n), \mathbf{f}_1, \mathbf{f}_2\}$, os sinais $y(n)$ e $u(n)$ correspondentes foram gerados a cada iteração, ao executar o sistema de sonorização com o filtro adaptativo.



(a)



(b)

Fig. 4: Convergência da NESD durante treino: (a) por batch; (b) por época, incluindo validação.

O caminho de realimentação foi inicialmente definido por $\mathbf{f}_1$ e alterado abruptamente para $\mathbf{f}_2$ após o decorrer de metade da simulação. Essa configuração, similar à empregada em [1], permite avaliar a reconvergência do método e fornecer uma indicação do seu desempenho em cenários levemente variantes no tempo. As STFTs $Y(m, k)$ e $U(m, k)$ foram calculadas de forma idêntica às etapas de treinamento e validação.

O filtro adaptativo é atualizado recursivamente, a partir da estimativa $\mathbf{h}_n$ proporcionada pelo método DFC, como [6]

$$\mathbf{h}(n) = \alpha \mathbf{h}(n-1) + (1 - \alpha) \mathbf{h}_n, \quad (8)$$

onde $0 \leq \alpha \leq 1$ é um parâmetro que controla a taxa de atualização, introduzido para evitar que oscilações de tempo curto na predição afetem o desempenho geral do sistema. Como consequência, α controla o compromisso entre velocidade de convergência e erro em regime permanente. No presente trabalho, como em [6], adotou-se $\alpha = 0,5$.

IV. RESULTADOS E DISCUSSÃO

Esta seção apresenta e discute os resultados de estimação do caminho de realimentação acústica proporcionados pelo método DFC proposto, tanto no treinamento quanto no teste.

A. Convergência no Treinamento e Validação

A Figura 4 exibe a evolução da função de custo NESD durante o treinamento e a validação do método DFC. Observa-se que o modelo apresentou convergência ao longo das iterações, *batches* ou épocas, com redução progressiva do erro. A evolução do NESD ao longo das épocas apresentou um

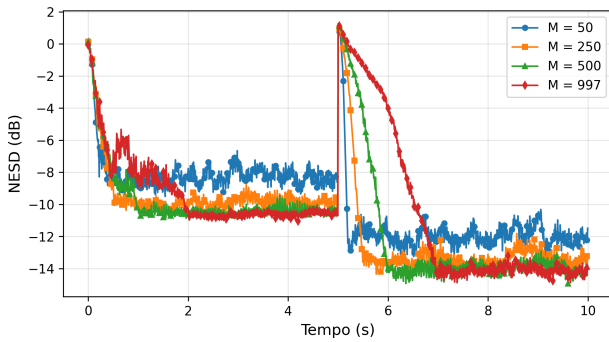


Fig. 5: NESDs para diferentes valores de M .

comportamento menos ruidoso, conforme esperado, uma vez que esses resultados correspondem à média de 16 *batches*.

Após 6 épocas, a função custo atingiu cerca de -12 dB no treinamento e -8 dB na validação. Apesar de mais épocas estarem disponíveis para treinamento, o NESD na validação não apresentou melhoria. Desse modo, houve a interrupção do treinamento pelo critério de *early stop* e o modelo treinado na época 6 foi selecionado para a etapa de teste.

Experimentos adicionais mostraram que a diferença de desempenho entre treinamento e validação foi causada pela divisão das RIRs. Como discutido nas Seções III-A e III-B, os dados foram propositalmente divididos de forma que as RIRs e os sinais $x(n)$ do conjunto de validação não fossem vistos pelo modelo durante seu treinamento. O modelo demonstrou excelente capacidade de generalização em relação aos sinais $x(n)$, mas não tanto quanto às RIRs. A utilização de RIRs não vistas durante o treinamento provocou essa redução de desempenho de aproximadamente 4 dB na validação.

B. Desempenho no Teste

A Figura 5 exibe o resultado médio do método DFC para os 10 trios $\{x(n), f_1, f_2\}$ que compartilham um mesmo par $\{f_1, f_2\}$, e para diferentes valores de M . O desempenho em regime permanente varia com o par $\{f_1, f_2\}$ mas, de forma geral, resultados semelhantes foram obtidos com outros trios.

O método proposto apresentou um bom desempenho para $M \leq 500$, atingindo convergência e reconvergência no intervalo 0,1–1 s e valores de NESD em regime permanente entre -8 e -11 dB antes da troca de caminho de realimentação, e entre -12 e -14 dB após a troca. Os diferentes níveis de estimação se devem à influência das respostas ao impulso do caminho de realimentação sobre o desempenho dos métodos AFC.

Além disso, observa-se a influência do número de janelas temporais no desempenho do método DFC. Com o aumento de M , há uma melhora da estimação em regime permanente, saturando a partir de $M = 500$, mas uma piora na velocidade de convergência e reconvergência. Esse resultado indica que esse parâmetro controla o compromisso entre desempenho em regime permanente e velocidade de rastreamento do método.

Os resultados obtidos com $M = 50$ estão próximos aos reportados em [6] utilizando o método DFC de referência, e levemente superiores aos dos métodos clássicos de AFC apresentados em [6]. Porém, essa comparação deve ser relativizada porque os trabalhos empregaram sinais de fala e caminhos de

realimentação acústica diferentes, e também em quantidades diferentes. Comparações do método DFC proposto com métodos clássicos de AFC serão realizados em estudos futuros.

No entanto, os resultados apresentados no presente trabalho indicam que é possível estimar eficazmente a resposta ao impulso do caminho de realimentação acústica em um sistema de sonorização, utilizando técnicas de aprendizado de máquina. Para tanto, uma quantidade elevada de RIRs deve ser utilizada para garantir um bom desempenho e generalização do modelo.

V. CONCLUSÕES

Este trabalho apresentou um método baseado em aprendizado de máquina para o cancelamento de realimentação acústica em sistemas de sonorização. A abordagem proposta emprega representações tempo-frequência dos sinais do microfone e do alto-falante, e redes neurais profundas para estimar dinamicamente o caminho de realimentação. Essa estimativa é integrada ao processo de atualização de um filtro adaptativo responsável por mitigar o efeito de realimentação no sistema.

O método proposto alcançou convergência e reconvergência entre 0,1 e 1 s, dependendo do valor de M , e erros em regime permanente entre -8 e -14 dB. O desempenho do método proposto foi próximo aos reportado no estudo de referência. Os resultados demonstram que a identificação do caminho de realimentação acústica por meio de redes neurais profundas constitui uma alternativa promissora frente aos métodos clássicos baseados exclusivamente em filtragem adaptativa linear. Comparações do método DFC proposto com métodos clássicos de AFC serão realizados em estudos futuros.

REFERÊNCIAS

- [1] T. van Waterschoot and M. Moonen, "Fifty years of acoustic feedback control: state of the art and future challenges," *Proceedings of the IEEE*, vol. 99, no. 2, pp. 288–327, 2011.
- [2] B. C. Bispo and D. R. S. Freitas, "On the use of cepstral analysis in acoustic feedback cancellation," *Digital Signal Processing*, vol. 44, pp. 88–101, Sep. 2015.
- [3] B. C. Bispo, C. F. Yamamura, W. M. da Silva Nogueira, E. A. R. Theodoro, and P. M. Rodrigues, "Analysis of acoustic feedback cancellation systems based on direct closed-loop identification," *Journal of Communication and Information Systems*, vol. 35, no. 1, pp. 217–229, Aug. 2020.
- [4] H. Zhang, M. Yu, and D. Yu, "Deep ahs: a deep learning approach to acoustic howling suppression," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2023, pp. 1–5.
- [5] X. Zhan, F. Hao, X. Li, and C. Zheng, "Deepem-afc: An improved prediction-error-method-based adaptive feedback cancellation with deep learning for hearing aids," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2025, pp. 1–5.
- [6] E. Lydakis, Z.-H. Tan, J. Jensen, and M. Guo, "Deep feedback cancellation for hearing aids with improved system stability and sound quality," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2025, pp. 1–5.
- [7] Y. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient backprop," in *Neural Networks: Tricks of the Trade*. Springer, 1998, pp. 9–50.
- [8] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [9] M. Jeub, M. Schafer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," in *International Conference on Digital Signal Processing*, 2009, pp. 1–5.
- [10] J. Kates, "Feedback cancellation in hearing aids: results from a computer simulation," *IEEE Transactions on Signal Processing*, vol. 39, no. 3, pp. 553–562, 1991.
- [11] A. Spriet, S. Doclo, M. Moonen, and J. Wouters, *Feedback Control in Hearing Aids*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 979–1000.
- [12] L. F. Lamel, R. H. Kassel, and S. Seneff, "Speech database development: design and analysis of the acoustic-phonetic corpus," 1989.