Performance Evaluation of the Cepstral Method to Estimate the Stable Optimal Solution of Feedforward Occlusion Cancellation in the Presence of Noise

Bruno Bispo and Renata Coelho Borges

October 24, 2022

Performance evaluation of the cepstral method to estimate the stable optimal solution of feedforward occlusion cancellation in the presence of noise

B. C. Bispo[1] and R. C. Borges[2]

[1] Department of Electrical and Electronic Engineering, Federal University of Santa Catarina, Florianópolis-SC, Brazil
[2] Department of Electrical Engineering, Federal University of Technology – Paraná, Cornélio Procópio-PR, Brazil

Abstract— **This work aims to evaluate the performance of the cepstral method to estimate, in the presence of noise, the stable optimal solution of feedforward occlusion cancellation. Simulations showed that the efficiency of the method reduces when decreasing the signal-to-noise ratio (SNR) at the hearing aids external microphone. For an analysis window of 2 s, estimates with mean normalized misalignment (MIS) less than -6, -14 and -22 dB are obtained for SNRs of 30, 40 and 50 dB, respectively. For a window of 4 s, mean MIS less than -16 and -24 dB are achieved for SNRs of 40 and 50 dB, respectively. The results indicate that the calibration process, where the method is used, needs to be carried out in an acoustically controlled environment to maximize the accuracy of the stable solution estimate.**

Keywords— **Occlusion effect, hearing aid, cepstral method, optimal solution, ambient noise.**

I. INTRODUCTION

The occlusion effect in hearing aids occurs when the ventilation opening has insufficient diameter to provide the necessary dissipation of sound energy conducted to the ear canal through the skull and jaw, when the hearing aid user speaks [1, 2].This leads to an increased low-frequency sound pressure level, thereby making the user listen to his own muffled voice [3]. The cause of the occlusion effect is modeled by the impulse response $o(n)$ in Figure 1.

The voice $v(n)$ uttered by the hearing aid user is picked up by the hearing aid's external microphone, after traversing the acoustic path represented by the impulse response $a(n)$, along with the ambient noise $r(n)$, generating the signal $y(n)$. Disregarding the filter $w_1(n)$, the signal of the external microphone, $y(n)$, is amplified by the compensation system represented by the impulse response $g(n)$, resulting in the signal $x(n)$ to be played back by the hearing aid's loudspeaker. The signal $z(n)$, to be picked up by a possible internal microphone of the device, is in fact the signal to be heard by the hearing aid user and is defined as

$$\begin{aligned} z(n) &= x(n) + v(n) * o(n) \\ &= [g(n) * a(n) + o(n)] * v(n) + g(n) * r(n), \end{aligned} \tag{1}$$

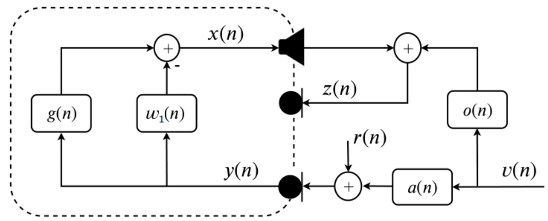


Fig. 1: Feedforward structure for occlusion effect cancellation.

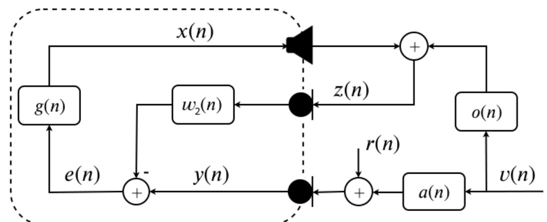


Fig. 2: Feedback structure for occlusion effect cancellation.

where the symbol $*$ denotes the convolution operation. Therefore, the occlusion effect is characterized by the addition of $o(n) * v(n)$ to the desired value of $z(n)$ and, thus, can be interpreted from the reverberation point of view.

Fixed and adaptive controllers have been proposed to, at least, attenuate the occlusion effect. The fixed solutions, proposed in [1, 4–7], ensure the system stability but do not deal with the dynamic changes of the acoustic system, and may suffer performance losses due to variations in the ear channel or displacement of the ear mold. On the other hand, the adaptive solutions proposed to date in [4, 5] present slow convergence of the adaptive filter coefficients and require constant adaptation since the occlusion effect occurs in short time periods, when sound signals are produced by the user.

Both solutions can be implemented in feedforward or feedback structures, which are shown in Figures 1 and 2 respectively. Note that, in the feedback structure, the hearing aid has a internal microphone to pick up and utilize the signal $z(n)$.

Among the mentioned cancellation proposals, the work presented in [7] stands out for being the only one to estimate the optimal solution for feedforward occlusion cancellation. The estimation is performed using a cepstral method and the occlusion cancellation feedback structure in a calibration process, when hearing loss compensation is not necessary and $g(n)$ and $w_2(n)$ can be freely chosen, provided that they do not result in an uncomfortable acoustic environment. However, the estimation method was evaluated considering only the absence of ambient noise in the calibration, a situation that may not be found even in controlled conditions.

This work aims to evaluate the performance of the cepstral method for estimating the stable optimal solution of occlusion cancellation in the presence of noise. This article is organized as follows: Section II discusses the optimal solution $w_o(n)$ for feedforward occlusion cancellation; in Section III, the cepstral method for estimating $w_o(n)$ is described; in Section IV, the configurations of the performed simulations are presented; Section V presents and discusses the results obtained; and, finally, Section VI concludes the work.

II. Feedforward occlusion cancellation system

In the feedforward occlusion cancellation system, illustrated in Figure 1, the optimal frequency response, in the sense of totally removing the signal $o(n) * v(n)$ from $z(n)$, in the absence of ambient noise ($r(n) = 0$) is given by [7]

$$W_o(e^{j\omega}) = \frac{O(e^{j\omega})}{A(e^{j\omega})}, \tag{2}$$

which in the time domain corresponds to

$$w_o(n) = o(n) * a_I(n), \tag{3}$$

where $a_I(n)$ denotes the impulse response of the inverse system to the acoustic path. Due to the properties of the discrete-time Fourier transform, $w_o(n)$ is absolutely summable and therefore this optimal solution is stable.

As the acoustic path models the propagation delay from the user's mouth to the external microphone of the hearing aid, its impulse response can be written as [7]

$$a(n) = \tilde{a}(n) * \delta(n - N_a), \tag{4}$$

where $\delta(n)$ is the unit impulse function, $\tilde{a}(n) = 0$ for $n < 0$, $\tilde{a}(0) \neq 0$ and $N_a > 0$. Consequently, as demonstrated in [7], $a_I(n)$ is in general a two-sided signal composed of left-side increasing exponentials for $n < -N_a$ and right-side decreasing exponentials for $n \geq -N_a$. But, since its energy is concentrated around $n = -N_a$, $a_I(n)$ can be considered of finite length with $a_I(n) \neq 0$ only for $A_1 \leq n \leq A_2$, where $A_1 < -N_a < A_2$. Note that A_1 is always negative.

Combining the above approximation of $a_I(n)$ with the fact that $o(n) \neq 0$ only for $n = 0, 1, \ldots, M-1$, as shown in Section V, $w_o(n) \neq 0$ only for $n = A_1, A_1+1, \ldots, M+A_2-1$ [7]. Hence, the stable optimal solution is non-causal.

III. Cepstral method to estimate $w_o(n)$

In the feedback cancellation system represented in Figure 2, it can be shown that the discrete-time Fourier transform (DTFT) of the error signal $e(n)$ is given by

$$\begin{aligned} E(e^{j\omega}) = {} & \frac{1 - W_o(e^{j\omega})W_2(e^{j\omega})}{1 + G(e^{j\omega})W_2(e^{j\omega})} Y(e^{j\omega}) \\ & + \frac{W_o(e^{j\omega})W_2(e^{j\omega})}{1 + G(e^{j\omega})W_2(e^{j\omega})} R(e^{j\omega}). \end{aligned} \tag{5}$$

In the absence of ambient noise, $R(e^{j\omega}) = 0$ and (5) becomes

$$E(e^{j\omega}) = \frac{1 - W_o(e^{j\omega})W_2(e^{j\omega})}{1 + G(e^{j\omega})W_2(e^{j\omega})} Y(e^{j\omega}). \tag{6}$$

In this ideal case, if $\left|W_o(e^{j\omega})W_2(e^{j\omega})\right| < 1$ and $\left|G(e^{j\omega})W_2(e^{j\omega})\right| < 1$, necessary conditions for Taylor series expansions, then the cepstrum of $e(n)$ can be defined from (6) as [7]

$$\begin{aligned} c_e(n) = {} & c_y(n) - \sum_{k=1}^{\infty} \frac{[w_o(n) * w_2(n)]^{*k}}{k} \\ & - \sum_{k=1}^{\infty} (-1)^{k+1} \frac{[g(n) * w_2(n)]^{*k}}{k}, \end{aligned} \tag{7}$$

where $\{\cdot\}^{*k}$ denotes the kth convolution power.

The cepstral method proposed in [7] explores (7) to estimate $w_o(n)$, in a fashion similar to [8, 9] for acoustic feedback cancellation. The estimation is carried out in a calibration process that occurs before using the hearing aid or when the user finds it convenient. The signal $v(n)$ must be a voiced sound, usually a vowel, emitted by the hearing aid user.

The reasoning behind the method comes from speculating that $w_o(n) * w_2(n)$, the impulse response for $k = 1$ of the first time series in (7), can be extracted from $c_e(n) - c_y(n)$ by properly choosing $g(n)$ and $w_2(n)$. At the calibration process, hearing loss compensation is not required and these impulse responses can be arbitrarily chosen, as long as they do not cause acoustic disturbance to the hearing aid user.

The choice of $g(n)$ and $w_2(n)$ plays a key role as it serves three purposes [7]: ensure that the conditions for Taylor series expansions are met and, therefore, equation (7) is valid; make $w_o(n) * w_2(n)$ causal; and make the non-zero samples of $w_o(n) * w_2(n)$ do not overlap with the non-zero samples of the impulse responses for $k > 1$ of the first time series in (7).

The impulse responses of the compensation and the feedback paths are simply defined as a bandwidth gain and a delay [7], that is,

$$g(n) = k_g \delta(n - N_g) \tag{8}$$

and

$$w_2(n) = k_w \delta(n - N_w), \tag{9}$$

where $N_g > 0$ and $N_w > 0$.

The conditions for Taylor series expansions are met by choosing k_w and k_g such that

$$|k_w| < \max_{\omega} \frac{1}{|W_o(e^{j\omega})|} \tag{10}$$

and

$$|k_g| < \frac{1}{|k_w|}. \tag{11}$$

Causality and non-overlapping of $w_o(n) * w_2(n)$ are achieved by choosing N_w according to [7]

$$N_w \geq M + A_2 - 2A_1. \tag{12}$$

Specified the parameters of $g(n)$ and $w_2(n)$ according to the above discussion, the method starts by computing the real cepstra $c_e(n)$ and $c_y(n)$ from the signals $e(n)$ and $y(n)$, respectively, through the fast Fourier transform (FFT). Then, the method gets $\{w_2(n) * [g(n) - w_o(n)]\}\hat{}$, an estimate of $w_2(n) * [g(n) - w_o(n)]$ which is the resulting impulse response for $k = 1$ in (7), by selecting the first $M + A_2 + N_w - 1$ samples of $c_e(n) - c_y(n)$.

In the sequel, the method computes $\{g(n) - w_o(n)\}\hat{}$, an estimate of $g(n) - w_o(n)$, as follows [7]

$$\{g(n) - w_o(n)\}\hat{} = \{w_2(n) * [g(n) - w_o(n)]\}\hat{} * w_I(n), \tag{13}$$

where $w_I(n) = 1/k_w \delta(n + N_w)$ represents the impulse response of the inverse system to $w_2(n)$ and is known. Note that convolution with $w_I(n)$ consists of a sliding on the time axis and a multiplication.

Finally, the method computes $\hat{w}_o(n)$, an estimate of the stable optimal solution $w_o(n)$ for the feedforward occlusion cancellation, as [7]

$$\hat{w}_o(n) = -\{g(n) - w_o(n)\}\hat{} + g(n). \tag{14}$$

IV. Simulation Configuration

This section describes the configuration of the experiment carried out in a simulated environment to evaluate the performance of the cepstral method in estimating $w_o(n)$ in the presence of noise.

A. Database

The database consists of 20 recordings (12 male and 8 female) of the sustained vowel /a/ sampled at a frequency of 22050 Hz. The vowel /a/ was chosen due to its wide use in the acoustic analysis of voice. It was provided by the Medical Engineering Research Group of the National Council for Scientific and Technological Development (GPEM/CNPq). The recordings were performed at the Hospital das Clínicas, Faculty of Medicine, University of São Paulo (HC-FMUSP), approved by the Human Research Ethics Committee of the Federal University of São Carlos under the protocol number 256/2010.

In this work, the signals were re-sampled to 16 kHz and the active power levels were normalized to -26 dBov through the ITU-T Recommendation P.56 algorithm [10]. Longer speech signals were created by concatenating each signal with itself. Variable length segments of each resulting signal were used as $v(n)$ in order to evaluate the performance of the cepstral method as a function of the vowel length uttered by the hearing aid user.

The speech signals were additively contaminated with zero-mean white noise at the following signal-to-noise ratio levels SNR $= \{\infty, 50, 45, 40, 35, 30, 25, 20\}$ dB.

B. System configuration

B1 Acoustic path

The acoustic path was represented in two ways. First, as in [4, 7, 11], the acoustic path was a time delay defined as

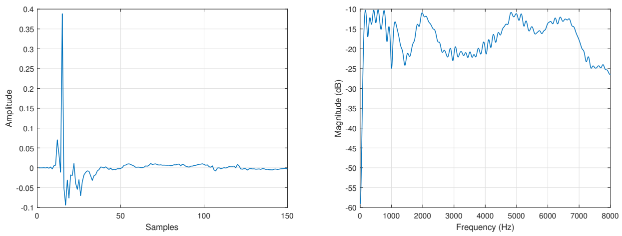
$$a(n) = \delta(n - N_a), \tag{15}$$

where $N_a = 14$. This delay refers to a propagation length of 30.29 cm, assuming an average length of 15 cm between the glottis and the lips [12] and a distance of 15.29 cm between the lips and the external microphone of hearing aids [4]. In this case, the condition (10) becomes $|k_w| < 0.41$.

Closer to a real-world situation, the second acoustic path was modeled by a room impulse response available in [13]. The sampling frequency was reduced to 16 kHz and its first 17 samples were discarded to simulate the typical 14-sample delay from the lips to the hearing aid external microphone. Then it was truncated for computational cost reasons. The impulse response $a(n)$ and the frequency response magnitude of the second acoustic path are shown in Figure 3. In this situation, the condition (10) becomes $|k_w| < 0.14$.

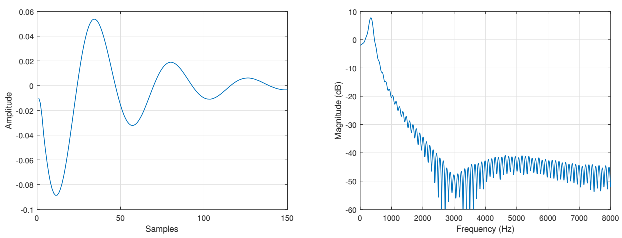
B2 Occlusion path

The occlusion path was modeled by the impulse response available in [4], which was measured in a volunteer with a



(a) (b)

Fig. 3: Second acoustic path: (a) $a(n)$; (b) $|A(e^{j\omega})|$.



(a) (b)

Fig. 4: Occlusion path: (a) $o(n)$; (b) $|O(e^{j\omega})|$.

custom and non-ventilated ear mold and digitally recorded at a sampling rate of 16 kHz. The impulse response $o(n)$ and the frequency response magnitude of the occlusion path are shown in Figure 4, where it is noted that $M = 150$.

B3 Forward path

As in [4, 7, 11], the forward path was modelled as

$$g(n) = \delta(n-1), \tag{16}$$

that is, $N_g = k_g = 1$. It is emphasized that, as discussed in Section III, the forward path does not need to compensate the hearing aids user's loss during the calibration process. And, as discussed in Section IV, the optimal stable solution for feedforward occlusion cancellation is independent of $g(n)$.

C. Misalignment

The estimate of the stable optimal solution obtained by the cepstral method was evaluated through the normalized misalignment (MIS), which is defined as

$$\text{MIS} = \frac{\left\{\sum_n [w_o(n) - \hat{w}_o(n)]^2\right\}^{1/2}}{\left\{\sum_n w_o^2(n)\right\}^{1/2}}. \tag{17}$$

V. Simulation results

The performance of the cepstral method was evaluated for several signal-to-noise ratios (SNR) in the hearing aid external microphone, namely, SNR $= \{\infty, 50, 45, 40, 35, 30, 25, 20\}$ dB, and for different analysis window sizes that is used to calculate the short-time cepstra. Hann's window was used.

A. Scenario 1

The impulse response $a(n)$ of the first acoustic path is defined in (15). Consequently, the impulse responses of its stable inverse system and stable optimal solution are defined as $a_I(n) = \delta(n+14)$ and $w_o(n) = o(n+14)$, respectively, where the non-causality of the stable optimal solution is verified.

The cepstral method was configured as follows: $A_1 = -100$, $A_2 = 100$, $M = 150$, $k_w = 0{,}1$, $N_w = 450$, $k_g = 1$ and $N_g = 1$. The mean MIS obtained for various speech lengths and various SNRs are shown in Table 1. Examples of $\hat{w}_o(n)$, which have MIS close to the mean value, obtained by the method with windows of 2 and 4 s and SNRs of 30 and 40 dB are illustrated in Figure 5.

The results show that the performance of the method improves both with the increase in the analysis window size and with the increase in the SNR. For a window of 2 s, a considerably short time for a person to sustain a vowel sound, SNRs equal to 30, 40 and 50 dB are needed so that, on average, MIS is lower than -12, -20 and -25 dB, respectively.

As initially demonstrated and discussed in [7], the performance improvement with window augmentation is due to the increase in the accuracy of the definition of $c_e(n)$ according to (7), caused by the reduction of the truncation effect of $e(n)$, which theoretically has infinite duration, necessary to compute $c_e(n)$. The performance improvement with the increase in SNR is also due to the increase in the accuracy of (7), but now caused by the reduction of the effect of the second term on the right-hand side of (5), which tends to zero as SNR tends to infinity.

B. Scenario 2

The impulse response $a(n)$ of the second acoustic path is represented in Figure 3. The resulting $w_o(n)$ is shown in Figure 6, where the non-causality of the stable optimal solution is evident. It is important to emphasize that, in this scenario, $w_o(n)$ is an infinite impulse response.

The cepstral method was configured as follows: $A_1 = -300$, $A_2 = 100$, $M = 150$, $k_w = 0{,}1$, $N_w = 850$, $k_g = 1$ and $N_g = 1$. The mean MIS obtained for various speech lengths and various SNRs are shown in Table 2. Examples of $\hat{w}_o(n)$

Table 1: Mean of the normalized misalignment in the first scenario.

SNR	Window length (ms)						
	500	1000	1500	2000	3000	4000	5000
∞	-10.06	-18.21	-20.69	-25.99	-27.89	-29.51	-35.31
50	-10.00	-17.47	-20.65	-25.23	-27.31	-29.82	-33.89
45	-9.69	-16.38	-20.36	-24.15	-26.61	-29.20	-32.21
40	-8.94	-13.57	-18.96	-21.96	-24.28	-27.04	-29.03
35	-7.24	-10.32	-15.90	-17.54	-18.83	-22.25	-24.01
30	-4.97	-6.71	-11.05	-12.21	-13.45	-16.64	-17.78
25	-3.67	-3.93	-6.99	-7.62	-7.76	-10.48	-11.30
20	-2.23	-2.45	-3.76	-4.32	-4.32	-6.22	-7.06

Table 2: Mean of the normalized misalignment in the second scenario.

SNR	Window length (ms)						
	500	1000	1500	2000	3000	4000	5000
∞	-5.90	-16.33	-19.16	-25.42	-26.38	-28.04	-33.96
50	-5.65	-14.07	-17.22	-22.07	-23.59	-24.67	-26.99
45	-5.32	-11.86	-15.67	-18.46	-19.47	-20.36	-20.99
40	-4.89	-9.29	-12.59	-14.08	-15.86	-16.89	-17.02
35	-3.75	-6.70	-9.34	-10.14	-10.80	-11.55	-11.98
30	-2.74	-4.09	-6.04	-6.61	-6.78	-7.83	-8.23
25	-1.84	-2.73	-3.91	-4.20	-4.16	-5.25	-5.70
20	-1.04	-1.58	-2.17	-2.36	-2.39	-3.13	-3.43

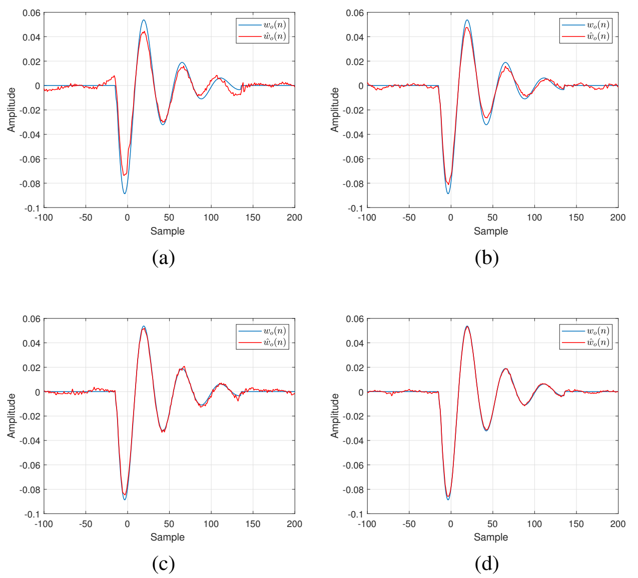


Fig. 5: Estimates of $w_o(n)$ in the first scenario for the following window sizes and SNRs: (a) 2 s, 30 dB; (b) 2 s, 40 dB; (c) 4 s, 30 dB; (d) 4 s, 40 dB.

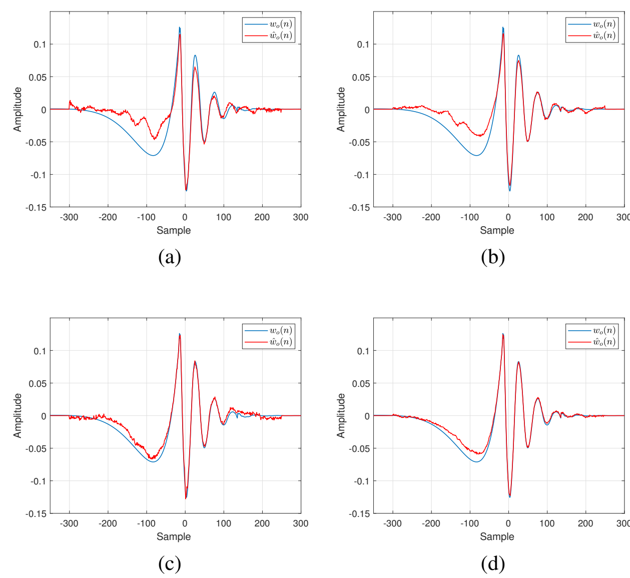


Fig. 6: Estimates of $w_o(n)$ in the second scenario for the following window sizes and SNRs: (a) 2 s, 30 dB; (b) 2 s, 40 dB; (c) 4 s, 30 dB; (d) 4 s, 40 dB.

obtained by the method for speech signals with durations of 2 and 4 s and SNRs of 30 and 40 dB are illustrated in Figure 6.

For the same reasons explained in the first scenario, the method performance improves with the increase in both the analysis window length and the SNR. However, it is observed that the average results are lower than those obtained in the first scenario. This is due to the combination of two factors: the infinite length of $w_o(n)$ inevitably causes overlap of the impulse responses present in $c_e(n) - c_u(n)$, thus impairing their estimation [7]; for the same size of the signal analysis window, the inevitable truncation of $e(n)$ to compute $c_e(n)$ can have an effect on the inaccuracy of (7) greater than in the first scenario since $w_o(n)$ is different.

The results presented in this work show that the performance differences become more significant with the decrease in SNR. For a window of 2 s, SNRs equal to 30, 40 and 50 dB are needed so that, on average, MIS is less than -6, -14 and -22 dB, respectively. In order to obtain mean MIS less than -10 and -20 dB, SNRs equal to or greater than 35 and 45 dB are required, respectively.

VI. Conclusions

This work evaluated the performance of the cepstral method to estimate, in the presence of noise, the stable optimal solution of feedforward occlusion cancellation in hearing aids. The estimation is performed in a calibration process carried out in a controlled environment. The method was originally evaluated considering the absence of ambient noise.

Simulations showed that the efficiency of the method reduces with the decrease in SNR at the hearing aid external microphone. For analysis windows of 2 s, estimates with mean MIS less than -6, -14 and -22 dB are obtained for SNRs of 30, 40 and 50 dB, respectively. SNRs equal to or greater than 35 and 45 dB are required to obtain mean MIS less than -10 and -20 dB, respectively. For windows of 4 s, estimates with mean MIS less than -16 and -24 dB are achieved for SNRs of 40 and 50 dB, respectively.

These results indicate that the calibration process, where the cepstral method is used, needs to be carried out in an acoustically controlled environment to maximize the accuracy of the optimal solution estimate, which can be used in implementing a controller to reduce the occlusion effect.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ACKNOWLEDGMENT

The authors would like to thank Prof. María E. Dajer for providing the voice database.

REFERENCES

1. Mejia J, Dillon H, Fisher M. Active cancellation of occlusion: An electronic vent for hearing aids and hearing protectors *The Journal of the Acoustical Society of America.* 2008;124:235–240.
2. Chung K. Challenges and recent developments in hearing aids: Part II. Feedback and occlusion effect reduction strategies, laser shell manufacturing processes, and other signal processing technologies *Trends in Amplification.* 2004;8:125–164.
3. MacKenzie Douglas J. Open-canal fittings and the hearing aid occlusion effect *The Hearing Journal.* 2006;59:50–52.
4. Borges R C, Costa M H. A feed forward adaptive canceller to reduce the occlusion effect in hearing aids *Computers in Biology and Medicine.* 2016;79:266–275.
5. Borges R C, Costa M H, Cordioli J A, Assuiti L F C. An adaptive occlusion canceller for hearing aids in *Proceedings of 21st European Signal Processing Conference*(Marrakech, Morocco):1–5 2013.
6. Sunohara M, Osawa M, Hashiura T, Tateno M. Occlusion reduction system for hearing aids with an improved transducer and an associated algorithm in *Proceedings of 23rd European Signal Processing Conference*:285–289 2015.
7. Bispo B C, Borges R C. A cepstral method to estimate the stable optimal solution for feedforward occlusion cancellation in hearing aids *Journal of Communication and Information Systems.* 2020;35:113-123.
8. Bispo B C, Rodrigues P M, Freitas D R S. Acoustic feedback cancellation based on cepstral analysis in *Proceedings of 17th IEEE Conference on Signal Processing Algorithms, Architectures, Arrangements and Applications*(Poznan, Poland):205-209 2013.
9. Bispo B C, Freitas D R S. On the use of cepstral analysis in acoustic feedback cancellation *Digital Signal Processing.* 2015;44:88-101.
10. ITU-T P.56 . P.56: Objective Measurement of Active Speech Level 2011.
11. Bispo B C, Borges R C, Costa M H. Método cepstral para estimação da resposta ao impulso do efeito de oclusão em sistemas de auxílio à audição in *Proceedings of the XXXVII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*(Petrópolis, Brazil) 2019.
12. Fant G. A note on vocal tract size factors and non-uniform F-pattern scalings *Speech Transmission Laboratory Quarterly Progress and Status Report.* 1966;1:22–30.
13. ITU-T G.191 . Software tools for speech and audio coding standardization International Telecommunications Union 2016.

Corresponding author: Bruno Catarino Bispo
Institute: Federal University of Santa Catarina
City: Florianópolis
Country: Brazil
Email: bruno.bispo@ufsc.br