A Cepstral Method to Estimate the Stable Optimal Solution for Feedforward Occlusion Cancellation in Hearing Aids

Bruno C. Bispo and Renata C. Borges

Abstract—The occlusion effect is a common complaint from users of hearing aids with narrow or unvented earmolds. This phenomenon makes the user hear his own voice muffled. In the scientific literature, fixed and adaptive controllers have been proposed for occlusion effect reduction. This work proposes a cepstral method to estimate the stable optimal solution for feedforward occlusion cancellation and a fixed controller that utilizes this estimate to reduce the occlusion effect in hearing aids. The cepstral method operates on a feedback structure at a calibration process. Simulations have shown that the performance of the cepstral method improves as the length of the signal uttered by the hearing aid user increases, resulting in estimates with average normalized misalignment less than -19 and -34 dB for signals lasting 1.5 and 5 s. The estimates are significantly more accurate below 500 Hz, which is the frequency range common to the occlusion effect. In addition, results have pointed out that the controller attenuated the occlusion effect, averagely decreasing by 0.17 dB the distortion power and increasing by 0.13 the objective perceptual quality MOS-LQO score.

Index Terms—Occlusion effect, feedforward occlusion cancellation, cepstral analysis, hearing aids.

I. INTRODUCTION

ACCORDING to the World Health Organization, in 2019, around 466 million people, more than 5% of the world population, have disabling hearing loss [2]. It is estimated that by 2050 there will be over 900 million people [2]. And according to the last census conducted in Brazil, in 2010, 9.7 million Brazilian citizens, or 5.1% of the national population, had some type of hearing impairment [3].

Hearing impairment can lead to social isolation, professional difficulties and risk to personal security, resulting in a reduction in the quality of life and productivity of the citizen. There are three types of hearing loss: conductive, sensorineural and mixed [4]. Each type is related to the region of the auditory system where the lesion is located. Compensation depends on the type and degree of the hearing loss and can be performed mainly with hearing aids, cochlear implants or surgery.

Nowadays, the hearing aids are still the most common choice for patients whose inner ear is not highly damaged. They are electronic devices that aims to compensate for hearing loss by amplifying the ambient sound and transmitting it to the external auditory channel through vibro-acoustic systems. Digital hearing aids are basically composed of microphones, analog/digital and digital/analog converters, digital signal processor, loudspeakers, transmission channel and battery.

Although most commercial devices available today use advanced digital signal processing techniques, a large amount of hearing aid users still report significant dissatisfaction [5], which are very often associated to the howling effect [6] and the occlusion effect [7]. The howling effect results from an instability of the closed-loop system generated by the acoustic coupling between the loudspeaker and the microphone of the hearing aid device, which causes the signal played back by the loudspeaker to be fed back into the microphone. To prevent the acoustic feedback, the simplest solution is the placement of a well fitted earmold into the user's ear [8].

A side effect of a well-suited earmold, with a very narrow or a completely closed ventilation opening, is that, when the hearing aid user speaks, the sound pressure produced by the laryngeal vocalizations is not sufficiently dissipated and thus conducted through the bony portions of the skull to the blocked ear. These vibrations result in an increased low-frequency sound pressure level reaching the tympanic membrane that makes the hearing aid user listen to his own voice muffled [9]. Such power magnification typically has the peak centered at around 200–500 Hz [7], [8], [10], [11]. This bone-conducted phenomenon is called the occlusion effect and is especially annoying for people with severe to profound sensorineural losses at high frequencies [9].

Since 2008, fixed and adaptive signal processing strategies have been proposed to cancel or at least attenuate the occlusion effect. Also known as control systems, the fixed solutions ensure the system stability but do not deal with the acoustic dynamics, and may suffer loss of performance due to changes in the ear canal or displacement of the earmold [7], [12], [13], [14]. The adaptive approaches, on the other hand, present a slow convergence of the filter coefficients and require continuous adaptation in search of an optimal solution since the occlusion effect occurs in short time intervals [10]. Both strategies can be implemented in feedforward or feedback structures, where the latter exploits a possible internal microphone of the hearing aid. It is worth mentioning that few models of hearing aid devices possess the internal microphone.

The first proposal of a occlusion cancellation system was presented in [7] using a fixed controller operating on a feedback structure. Afterwards, adaptive solutions were first proposed in [13] and later in [12], [14]. In [12], [13], a

B. C. Bispo is with the Department of Electrical and Electronic Engineering, Federal University of Santa Catarina (UFSC), Florianópolis, Brazil, e-mail: bruno.bispo@ufsc.br. R. C. Borges is with the Department of Electrical Engineering, Federal University of Technology - Paraná (UTFPR), Cornélio Procópio, Brazil, e-mail: renatacoelho@utfpr.edu.br.

feedforward cancellation architecture is used. The adaptive method in [14], on the other hand, employs a feedback scheme. More recently, a fixed controller operating on a feedforward cancellation structure was proposed in [1].

Among the aforementioned cancellation proposals, the work presented in [1] stands out for being the only one to estimate the occlusion path, nomenclature for designating the system that models the bone conduction of the sound pressure produced by the laryngeal vocalizations to the blocked ear. The estimation is performed by assuming knowledge of the acoustic path, system that denotes the acoustic coupling between the user lips and the hearing aid external microphone. However, in practice, this is not true and the acoustic path would also need to be estimated. Moreover, knowledge of the forward path, which represents the user hearing loss compensation system, is also inconveniently required for the controller design.

As a possible answer for these limitations, [1] is expanded in this paper. The goal of this work is twofold: first and foremost, without assuming to know the acoustic path, to propose a cepstral method for estimating the stable optimal solution of the feedforward occlusion cancellation; second, to propose a controller designed based on this estimate and without assuming to know the forward path for attenuating the occlusion effect in the feedforward cancellation structure.

This paper is organized as follows: Section II presents the occlusion effect modeling and the two cancellation structures used in this work as well as demonstrates the stable optimal solution for the feedforward structure; Section III presents a cepstral analysis of the feedback structure, explains in detail the proposed cepstral method for estimating the stable optimal solution of the feedforward occlusion cancellation and describes the proposed controller for attenuating the occlusion effect; Section IV describes the configuration of the simulations carried out; in Section V, the performance results of both proposed methods are presented and discussed; finally, Section VI concludes the paper.

II. OCCLUSION EFFECT MODELING AND CANCELLATION SYSTEMS

The modeling of the occlusion effect is shown in Figure 1 disregarding $w_1(n)$. The acoustic, forward and occlusion paths are represented by the impulse responses $a(n)$, $g(n)$ and $o(n)$, respectively. The signal $v(n)$ denotes the sound signal uttered by the hearing aid user. The external microphone signal $y(n)$ is assumed to be ambient noise-free because of a noise cancellation algorithm. The signal $z(n)$, to be picked up by a possible hearing aid internal microphone, is actually the signal heard by the user and defined as

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

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.

In this work, the feedforward structure adopted for occlusion cancellation is depicted in Figure 1, where the cancellation
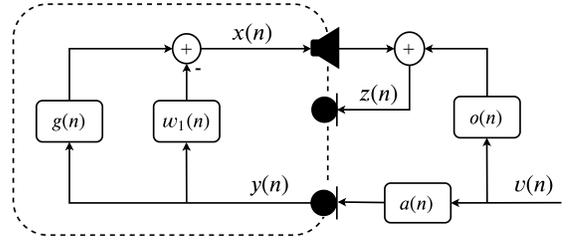
Fig. 1: Feedforward structure for occlusion cancellation.
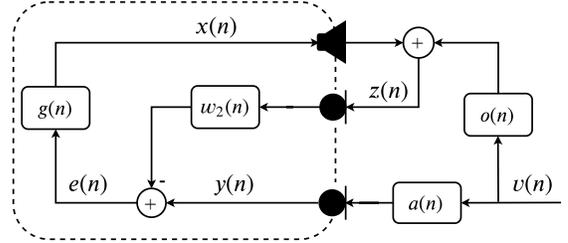
Fig. 2: Feedback structure for occlusion cancellation.

path is represented by the impulse response $w_1(n)$. This is the same structure employed in [12], [13]. And the feedback structure selected for occlusion cancellation is illustrated in Figure 2, where the cancellation path is represented by the impulse response $w_2(n)$. It is evident that the feedback architecture exploits the internal microphone signal $z(n)$.

A. Stable Optimal Solution for the Feedforward Occlusion Cancellation

In the feedforward system for occlusion cancellation, it can be easily demonstrated that the discrete-time Fourier transform (DTFT) of $z(n)$ is given by

$$
\begin{aligned}
Z(e^{j\omega}) &= \left\{ A(e^{j\omega}) \left[G(e^{j\omega}) - W_1(e^{j\omega}) \right] + O(e^{j\omega}) \right\} \\
&\quad \times V(e^{j\omega}),
\end{aligned} \tag{2}
$$

where $A(e^{j\omega})$, $G(e^{j\omega})$, $W_1(e^{j\omega})$, $O(e^{j\omega})$ and $V(e^{j\omega})$ are the DTFT of $a(n)$, $g(n)$, $w_1(n)$, $o(n)$ and $v(n)$, respectively, and ω is the normalized angular frequency.

As ideally $Z(e^{j\omega}) = A(e^{j\omega})G(e^{j\omega})V(e^{j\omega})$, the optimal frequency response for the controller or adaptive filter is

$$
W_o(e^{j\omega}) = \frac{O(e^{j\omega})}{A(e^{j\omega})}, \tag{3}
$$

which in time domain leads to

$$
w_o(n) = o(n) * a_I(n), \tag{4}
$$

where $a_I(n)$ denotes the impulse response of the inverse system to the acoustic path. It is clear that $w_o(n)$ relies directly on $o(n)$ and $a(n)$ but is independent of $g(n)$. The independence of $g(n)$ is an important advantage over the open-loop cancellation structure proposed in [1]. Due to the properties of the DTFT, $w_o(n)$ is absolutely summable and therefore this optimal solution is stable.

As the acoustic path models the inherent propagation delay from the user lips to the hearing aid external microphone, its impulse response can be written as

$$
a(n) = \tilde{a}(n) * \delta(n - N_a), \tag{5}
$$

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 Appendix A, $a_I(n)$ is in general a two-sided signal composed of left-sided increasing exponentials for $n < -N_a$ and right-sided decreasing exponentials for $n \geq -N_a$. But, as its energy is concentrated around $n = -N_a$, $a_I(n)$ can be considered of finite duration 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 IV-A, the impulse response of the stable optimal solution defined in (4) is such that $w_o(n) \neq 0$ only for $n = A_1, A_1 + 1, \ldots, M + A_2 - 1$. Therefore, the stable optimal solution is non-causal, which makes the feedforward occlusion cancellation a very challenging task.

Although it can not be implemented for real-time processing, the knowledge of $w_o(n)$ is of paramount importance either to design a controller with appropriate frequency response or to initialize the coefficients of an adaptive filter for feedforward occlusion cancellation. In the sequel, a cepstral method operating on the feedback structure to estimate $w_o(n)$ and a controller operating on the feedforward structure that exploits this estimate to attenuate the occlusion effect are proposed.

III. METHODOLOGY

A. Cepstral analysis of the feedback structure for occlusion cancellation

The feedback structure for occlusion cancellation is governed by the following frequency domain equations

$$\begin{cases} Y(e^{j\omega}) = A(e^{j\omega})V(e^{j\omega}) \\ E(e^{j\omega}) = Y(e^{j\omega}) - W_2(e^{j\omega})Z(e^{j\omega}) \\ X(e^{j\omega}) = G(e^{j\omega})E(e^{j\omega}) \\ Z(e^{j\omega}) = X(e^{j\omega}) + O(e^{j\omega})V(e^{j\omega}), \end{cases} \quad (6)$$

where $Y(e^{j\omega})$, $E(e^{j\omega})$, $X(e^{j\omega})$, $Z(e^{j\omega})$ and $W_2(e^{j\omega})$ are the DTFT of $y(n)$, $e(n)$, $x(n)$, $z(n)$ and $w_2(n)$, respectively.

Hence, the frequency domain relationship between the error signal $e(n)$ and the external microphone signal $y(n)$ is given by

$$E(e^{j\omega}) = \frac{1 - O(e^{j\omega})W_2(e^{j\omega})/A(e^{jw})}{1 + G(e^{j\omega})W_2(e^{j\omega})} Y(e^{j\omega}), \quad (7)$$

which making use of (3) can be written as

$$E(e^{j\omega}) = \frac{1 - W_o(e^{j\omega})W_2(e^{j\omega})}{1 + G(e^{j\omega})W_2(e^{j\omega})} Y(e^{j\omega}). \quad (8)$$

Applying the natural logarithm in (8) yields

$$\ln\left[E(e^{j\omega})\right] = \ln\left[Y(e^{j\omega})\right] + \ln\left[1 - W_o(e^{j\omega})W_2(e^{j\omega})\right] - \ln\left[1 + G(e^{j\omega})W_2(e^{j\omega})\right]. \quad (9)$$

If $\left|W_o(e^{j\omega})W_2(e^{j\omega})\right| < 1$, the second term on the right-hand side of (9) can be expanded in Taylor's series as

$$\ln\left[1 - W_o(e^{j\omega})W_2(e^{j\omega})\right] = -\sum_{k=1}^{\infty} \frac{\left[W_o(e^{j\omega})W_2(e^{j\omega})\right]^k}{k}. \quad (10)$$

And if $\left|G(e^{j\omega})W_2(e^{j\omega})\right| < 1$, a sufficient condition to ensure the closed-loop system stability, the third term on the right-hand side of (9) can be expanded in Taylor's series as

$$\ln\left[1 + G(e^{j\omega})W_2(e^{j\omega})\right] = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{\left[G(e^{j\omega})W_2(e^{j\omega})\right]^k}{k}. \quad (11)$$

Replacing (10) and (11) in (9) and then applying the inverse DTFT, the cepstral domain relationship between $e(n)$ and $y(n)$ is given by

$$c_e(n) = c_y(n) - \sum_{k=1}^{\infty} \frac{\left[w_o(n) * w_2(n)\right]^{*k}}{k} - \sum_{k=1}^{\infty} (-1)^{k+1} \frac{\left[g(n) * w_2(n)\right]^{*k}}{k}, \quad (12)$$

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

In the feedback system for occlusion cancellation, the cepstrum $c_e(n)$ of the error signal is the cepstrum $c_y(n)$ of the external microphone signal added to time domain series as function of $w_o(n)$, $g(n)$ and $w_2(n)$. The cepstral analysis modified the representation of the system components in relation to the system signals. In (8), they are all in the frequency domain. But in (12), the signals $e(n)$ and $y(n)$ are in the cepstral domain while the system components are actually in the time domain.

However, it should be reminded that $c_e(n)$ is defined by (12) if and only if the conditions for the Taylor series expansions in (10) and (11) are fulfilled. Otherwise, nothing can be inferred about the mathematical definition of $c_e(n)$ as a function of $w_o(n)$, $g(n)$ and $w_2(n)$. And, even though (12) is valid, the practical existence of the impulse responses in $c_e(n)$ depends whether the analyzed time window of $e(n)$ is large enough to include their effects [15].

B. Cepstral method operating on the feedback structure to estimate $w_o(n)$

The proposed method aims at estimating $w_o(n)$ from (12) in a fashion similar to [15], [16], [17] for acoustic feedback cancellation. The estimation would be performed in a calibration process before using the hearing aid or when the user finds it convenient. The source signal $v(n)$ is supposed to be a vowel sound of finite duration uttered by the hearing aid user.

1) Computation of $c_e(n) - c_y(n)$: The availability of both signals $e(n)$ and $y(n)$ enables the computation of their real cepstra through the fast Fourier transform (FFT). The subtraction of $c_y(n)$ from $c_e(n)$ would result in the convolution power series presented in (12). To that effect, however, the signals used in the cepstra computation must be related by (8).

Analysis of (8) as an input-output system reveals that, in general, the signal $e(n)$ is of infinite duration. But, due to the FFT nature, the computation of $c_e(n)$ is inevitably performed using a finite-length segment of $e(n)$. The time windowing of a signal leads to the so-called spectral leakage, where energy from one frequency range leaks into another, resulting in a distorted spectral estimate. The leakage can be reduced by

using a smooth window or increasing the segment length. With this in mind, a Hann window is used in this work for spectral estimation of both $e(n)$ and $y(n)$. In the target application, the spectral leakage makes (12) less accurate and ultimately disrupts the estimation of $w_o(n)$.

2) Estimation of $w_o(n)$ from $c_e(n) - c_y(n)$: The reasoning behind the proposed method comes from speculating that $w_o(n) * w(n)$, the impulse response for $k = 1$ of the first time series in (12), can be extracted from $c_e(n) - c_y(n)$ by appropriately choosing $g(n)$ and $w_2(n)$. At the calibration stage, hearing loss compensation is not required and these impulse responses can be freely chosen as long as they do not result in a very uncomfortable acoustic environment.

The choice of both $g(n)$ and $w_2(n)$ plays a key role as it serves the following three purposes:

1) To ensure that the conditions for Taylor series expansions are met and, consequently, (12) is valid;
2) To make $w_o(n) * w_2(n)$, the impulse response for $k = 1$ of the first time series in (12), causal;
3) To 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 (12).

In this work, the impulse responses of both forward and feedback paths are simply defined as a broadband gain and a delay, i.e.,

$$g(n) = k_g \delta(n - N_g) \qquad (13)$$

and

$$w_2(n) = k_w \delta(n - N_w), \qquad (14)$$

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

The conditions for the Taylor series expansions are met by appropriately setting the broadband gains. After simple algebraic manipulations, it is straightforward that

$$|k_w| < \max_{\omega} \frac{1}{|W_o(e^{j\omega})|} \qquad (15)$$

and

$$|k_g| < \frac{1}{|k_w|}. \qquad (16)$$

Both causality and non-overlapping of $w_o(n) * w_2(n)$ are accomplished by means of an appropriate choice of N_w. Taking into acount that $w_o(n) \neq 0$ only for $n = A_1, \ldots, M + A_2 - 1$, as explained in Section II-A, it follows that $w_o(n) * w_2(n) \neq 0$ only for $n = A_1 + N_w, \ldots, M + A_2 + N_w - 1$. Therefore, in order to $w_o(n) * w_2(n)$ be causal, it is required that

$$N_w \geq -A_1. \qquad (17)$$

And, since $w_o^{*2}(n) \neq 0$ only for $n = 2A_1, \ldots, 2(M + A_2 - 1)$, it turns out that $[w_o(n) * w_2(n)]^{*2} \neq 0$ only for $n = 2(A_1 + N_w), \ldots, 2(M + A_2 + N_w - 1)$. Thus, the non-zero samples of $w_o(n) * w_2(n)$ and $[w_o(n) * w(n)]^{*2}$ do not overlap when

$$N_w \geq M + A_2 - 2A_1. \qquad (18)$$

The non-overlapping condition is sufficient to ensure the causality condition. Therefore, $N_w \geq M + A_2 - 2A_1$ is required.

The parameter N_g is not utilized to extract $w_o(n) * w(n)$ from $c_e(n) - c_y(n)$ because all the impulse responses of the second time series in (12) are known and could be cancelled out. But, since $[g(n) * w_2(n)]^{*k} \neq 0$ for $n = k(N_g + N_w)$ and $N_g > A_1$, it concludes that only $g(n) * w_2(n) = k_w k_g \delta(n - N_w - N_g)$ can be overlapped to $w_o(n) * w_2(n)$. And that will only happen if $N_g \leq M + A_2 - 1$. Assuming this range for N_g, $g(n) * w_2(n)$ needs to be removed as described below.

Specified the parameters of $g(n)$ and $w_2(n)$ in accordance with the above discussion, the proposed method starts by obtaining $\{w_2(n) * [g(n) - w_o(n)]\}\widehat{}$, an estimate of $w_2(n) * [g(n) - w_o(n)]$ which is the resulting impulse response for $k = 1$ presented in (12). This is performed by selecting the first $M + A_2 + N_w - 1$ samples from $c_e(n) - c_y(n)$.

Then, the method computes $\{g(n) - w_o(n)\}\widehat{}$, an estimate of $g(n) - w_o(n)$, as follows

$$\{g(n) - w_o(n)\}\widehat{} = \{w_2(n) * [g(n) - w_o(n)]\}\widehat{} * w_I(n), \quad (19)$$

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 the convolution with $w_I(n)$ consists of a sliding on the sample 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

$$\hat{w}_o(n) = - \{g(n) - w_o(n)\}\widehat{} + g(n). \qquad (20)$$

In addition to make the method feasible, the parameters k_g, k_w, N_g and N_w have influence on its performance. When analyzing (8) considering (13)–(15), it can be demonstrated that the lower $|k_g|$, $|k_w|$, N_g or N_w, the lower the energy ratio of $e(n)$ not included in its finite-length segment used for computing $c_e(n)$. Decreasing this ratio reduces the distortion caused by the spectral leakage and consequently improves $\hat{w}_o(n)$. Therefore, it would be beneficial to set these values as small as possible. In light of this, N_w and N_g are set to their lower bounds, i.e., $N_w = M + A_2 - 2A_1$ and $N_g = 1$.

But some restrictions are imposed on the broadband gains. A very small $|k_g|$ can make the ambient sound inaudible to the hearing aid user during the calibration, which is not acoustically comfortable. With the aim of making the user naturally listen to the ambient sound, $k_g = 1$ is used in this work. This is accomplished by setting k_w such that not only (15) is fulfilled but also $0 < k_w < 1$, resulting in $k_w = 0.1$.

C. Controller operating on the feedforward cancellation structure to attenuate the occlusion effect

The proposed controller for attenuating the occlusion effect operates on the feedforward cancellation structure shown in Figure 1 and uses the estimate $\hat{w}_o(n)$ obtained by the cepstral method described in the previous section.

The controller is defined by the following impulse response

$$w_1(n) = \hat{w}_o(n)u(n), \qquad (21)$$

where $u(n)$ is the unit step function. Therefore, the controller impulse response is actually the estimate of the impulse response of the stable optimal solution, which is non-causal, truncated so as to become causal.

(b)
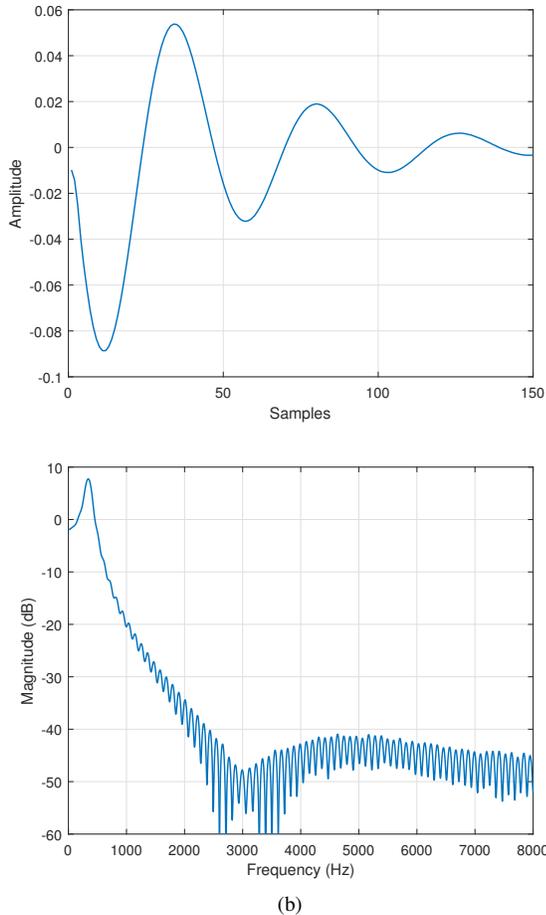
Fig. 3: Occlusion path: (a) impulse response $o(n)$; (b) magnitude of the frequency response $O(e^{j\omega})$.

(b)

Fig. 4: Acoustic path of the second scenario: (a) impulse response $a(n)$; (b) magnitude of the frequency response $A(e^{j\omega})$.

IV. SIMULATION CONFIGURATIONS

This section describes the configuration of the two experiments carried out in a simulated environment. The first one evaluates the performance of the cepstral method for estimating the stable optimal solution $w_o(n)$ of the occlusion effect. The second experiment evaluates the performance of the controller for attenuating the occlusion effect.

A. Simulated Environment

1) Occlusion path: The occlusion path is modeled by the impulse response available in [12], which was measured in a volunteer with a custom and unventilated earmold and digitally recorded at a 16 kHz sampling rate. The impulse response $o(n)$ and frequency response magnitude of the occlusion path are shown in Figure 3, where it is observed that $M = 150$.

2) Acoustic path: The acoustic path is represented in two different ways. First, as done in [1], [12], the acoustic path is a delay line defined as

$$a(n) = \delta(n - N_a), \quad (22)$$

where $N_a = 14$. This delay refers to a propagation length of 30.29 cm, considering an average length of 15 cm between the glottis and lips [18] and a distance of 15.29 cm between the

lips and the external microphone of hearing aids [12]. In this case, (15) becomes $|k_w| < 0.41$ and then $k_w = 0.1$ is used.

Closer to a real-world situation, the second acoustic path is modeled by a measured room impulse response available in [19]. The impulse response was downsampled to 16 kHz and its first 17 samples were discarded to simulate the typical 14-sample delay from 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 4. In this situation, (15) becomes $|k_w| < 0.14$ and $k_w = 0.1$ is used again.

3) Forward path: As in [1], [12], the forward path was modeled as a 1-sample delay, i.e.,

$$g(n) = \delta(n - 1). \quad (23)$$

It is pertinent to emphasize that, as discussed in Section III-B, the forward path is not required to compensate the user hearing loss during the calibration process. And, as discussed in Section II-A, the stable optimal solution for the feedforward occlusion cancellation is independent of $g(n)$.

B. Evaluation Metrics

1) Misalignment: A very common metric in system identification is the misalignment (MIS), which measures the mismatch between the identified and true systems. The performance of the cepstral method for estimating the stable optimal solution of the feedforward occlusion cancellation is evaluated through the normalized MIS defined as [1]

$$\text{MIS} = \frac{\left\{ \sum_n \left[w_o(n) - \hat{w}_o(n) \right]^2 \right\}^{1/2}}{\left\{ \sum_n w_o^2(n) \right\}^{1/2}}. \tag{24}$$

In order to scrutinize the method performance over the frequency range common to the occlusion effect, the normalized MIS is also computed separately for the frequencies below and above 500 Hz, which are respectively defined as

$$\text{MIS}_< = \frac{\left\{ \sum_{\omega=0}^{\omega_1} \left| W_o(e^{j\omega}) - \hat{W}_o(e^{j\omega}) \right|^2 \right\}^{1/2}}{\left\{ \sum_{\omega=0}^{\omega_1} \left| W_o(e^{j\omega}) \right|^2 \right\}^{1/2}} \tag{25}$$

and

$$\text{MIS}_> = \frac{\left\{ \sum_{\omega=\omega_1}^{\pi} \left| W_o(e^{j\omega}) - \hat{W}_o(e^{j\omega}) \right|^2 \right\}^{1/2}}{\left\{ \sum_{\omega=\omega_1}^{\pi} \left| W_o(e^{j\omega}) \right|^2 \right\}^{1/2}}, \tag{26}$$

where ω_1 is the frequency corresponding to 500 Hz.

2) Distortion power: The performance of the fixed feedforward controller for attenuating the occlusion effect is evaluated through the power of the difference between the internal microphone signal $z(n)$ and its ideal value, i.e.,

$$P = \frac{1}{N} \sum_{n=0}^{N-1} \left[z(n) - g(n) * a(n) * v(n) \right]^2, \tag{27}$$

where N is the length of the speech signal. This metric can be interpreted as the power of the distortion caused by the occlusion effect in the sound heard by the user.

3) Wideband Perceptual Evaluation of Speech Quality: Objective measures of speech quality have evolved from those based on purely mathematical criteria towards perceptually salient metrics. The W-PESQ (Wideband Perceptual Evaluation of Speech Quality) is a standard algorithm for objective quality evaluation of wideband (sampled at 16 kHz) speech signals [20], [21], [22], [23]. It compares the psychophysical representations of both a possibly degraded speech signal and its corresponding uncorrupted reference [23].

The raw W-PESQ scores can be mapped to the 1-5 mean opinion score (MOS) scale, resulting in the so-called MOS-Listening Quality Objective (MOS-LQO) score [24]. The correspondence between the five-point scale and the degradation category rating (DCR) is shown in Table I. However, the maximum MOS-LQO score provided by W-PESQ is 4.644 when the reference and degraded signals are identical.

The W-PESQ achieves a correlation of 80% with MOS when assessing speech impairment by reverberation although

TABLE I: Correspondence between five-point score and DCR.

Score	Category
5	Inaudible
4	Audible but not annoying
3	Slightly annoying
2	Annoying
1	Very annoying

TABLE II: Summary of the speech database.

Gender		Age		Duration (s)	
Male	Female	Mean	SD	Mean	SD
12	8	31.31	9.26	2.85	0.15

it was not designed for this purpose [25], [26]. Hence, the W-PESQ was applied to perceptually evaluate the performance of the proposed controller for attenuating the occlusion effect. Thereunto, the ideal and true internal microphone signals worked as the reference and degraded signals, respectively.

In order to fully exploit the W-PESQ, the reference signals are recommended to meet the following specifications [27]: duration between 8 and 12 s, speech activity between 40% and 80%, a minimum active speech of 3.2 s, an active speech level of −30 dBov, a minimum leading and trailing silence of 0.5 s, a maximum leading and trailing silence of about 2 s, utterances separated by a silent period of at least 1 s.

C. Speech Database

The speech database was provided by the Medical Engineering Group of the Brazilian National Council of Scientific and Technological Development (GPEM/CNPq). The database consists of 20 recordings, from Brazilian healthy adult subjects (1 per subject), of the sustained vowel /a/. The vowel /a/ was selected as speech material due to its large use in acoustic analysis of voice [28].

The information about gender and age of the subjects as well as time duration of the speech signals are summarized in Table II, where standard deviation is abbreviated to SD. The active power level of each signal was normalized to −30 dBov through the Recommendation ITU-T P.56 algorithm [29].

The speech signals are used to compose the source signal $v(n)$ in the two experiments performed. In the first experiment, longer speech signals are generated by concatenating each signal with itself. Variable size segments of each resulting signal are used as $v(n)$ in order to evaluate the performance of the proposed cepstral method for estimating $w_o(n)$ as a function of the vowel length uttered by the hearing aid user.

In the second experiment, the source signal $v(n)$ is build from the original speech signals as follows: 1 s of leading silence, speech signal, 1.5 s of silence, replica of the speech signal and 1 s of trailing silence. This temporal structure satisfies the reference signal requirements for the W-PESQ described in Section IV-B3, leading to an appropriate application of the algorithm in the performance evaluation of the proposed controller to attenuate the occlusion effect.

V. SIMULATION RESULTS

This section is devoted to reporting the results of the two experiments performed. First, the performance of the cepstral method for estimating the stable optimal solution $w_o(n)$ of the feedforward occlusion cancellation is addressed. In the sequel, the performance of the controller $w_1(n)$ for attenuating the occlusion effect is presented. In both cases, the simulated environment, the evaluation metrics and the signals described in Section IV are employed.

A. Estimation of the stable optimal solution $w_o(n)$

1) Scenario 1: The impulse response $a(n)$ of the first acoustic path is defined in (22). Consequently, the impulse responses of its stable inverse system and the stable optimal solution are defined as $a_I(n) = \delta(n+14)$ and $w_o(n) = o(n+14)$, respectively, where it is concluded that the stable optimal solution is indeed non-causal.

The method was set up 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 performance results of the proposed cepstral method in estimating $w_o(n)$ for several speech lengths are summarized in Table III. Examples of $\hat{w}_o(n)$ obtained by the method for 1 s and 3 s-long speech signals are illustrated in Figure 5.

As can be seen, the method performance improves as the speech length increases. This behavior is due to the spectral leakage reduction caused by the use of larger windows, which makes (12) more accurate as discussed in Section III-B1. It is also observed that the estimate obtained by the method is significantly more accurate in the range below 500 Hz. The cause of this result is that the distortion in $c_e(n) - c_y(n)$ caused by the spectral leakage is fundamentally of high frequency, as can be noticed in Figure 5a. The highest accuracy of $\hat{w}_o(n)$ in this frequency range is of great value because the occlusion effect is a low-frequency phenomenon.

A careful inspection of Figure 5 reveals that the values selected for A_1 and A_2 causes the proposed method to extract only noise from $c_e(n) - c_y(n)$ for $n < -14$ and $n > 150$. Less noisy $\hat{w}_o(n)$ could be obtained by increasing A_1 or decreasing A_2, ideally $A_1 = A_2 = -14$. But $A_1 = -100$ and $A_2 = 100$ were adopted in order to pretend ignorance of $w_o(n)$ to the point of not overly harming its estimation.

Furthermore, it is noticed that at least 0.5 s of speech is required for the method to obtain a good estimate of $w_o(n)$. For signals longer than 1.5 s, the method achieves average values of MIS and MIS$_<$ less than -21 dB and -31 dB, respectively. Outstanding mean values of MIS and MIS$_<$ approximately equal -35 dB and -47 dB, respectively, are achieved for 5s-long speech signals. The results shown in Table III can serve as a valid indication that, at the first scenario, the proposed method accurately estimates the stable optimal solution $w_o(n)$ of the feedforward occlusion cancellation for speech signals longer than 1 s. And it is reasonable to believe that a person can easily sustain a vowel sound for that time.

2) Scenario 2: The impulse response $a(n)$ of the second acoustic path is the one depicted in Figure 4. The resulting $w_o(n)$ is shown in Figure 6, where the non-causality of the

TABLE III: Performance results of the proposed method in estimating $w_o(n)$ as function of speech length at first scenario.

Speech duration (ms)	Mean of MIS (dB)	Mean of MIS$_<$ (dB)	Mean of MIS$_>$ (dB)
100	14.50	8.19	26.43
250	1.10	-4.49	12.73
500	-10.16	-19.90	2.60
1000	-18.25	-30.58	-5.25
1500	-20.92	-30.96	-8.13
2000	-26.22	-43.67	-13.04
3000	-28.07	-38.03	-15.37
4000	-29.94	-38.54	-17.60
5000	-35.42	-47.38	-22.49

(b)
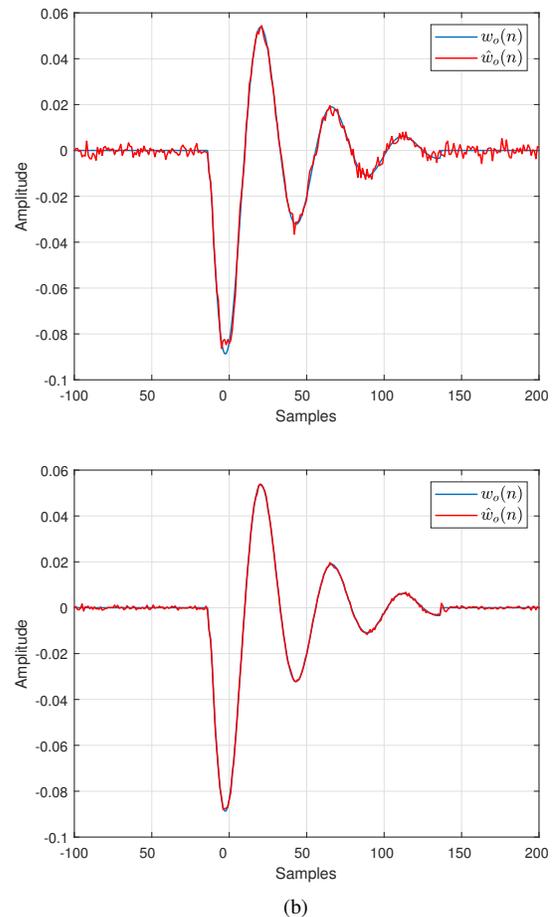
Fig. 5: Estimates of $w_o(n)$ obtained by the proposed method at the first scenario for some speech lengths: (a) 1 s; (b) 3 s.

stable optimal solution is evident. It is relevant to emphasize that, in this scenario, $w_o(n)$ is an infinite impulse response.

The method was set up as follows: $A_1 = -300$, $A_2 = 100$, $M = 150$, $k_w = 10^{-2}$, $N_w = 850$, $k_g = 1$ and $N_g = 1$. The performance results of the proposed cepstral method in estimating $w_o(n)$ for several speech lengths are summarized in Table IV. Examples of $\hat{w}_o(n)$ obtained by the method for 1 s and 3 s-long speech signals are illustrated in Figure 7.

As can be seen, the mean results are slightly lower than those obtained in the first scenario because the infinite length of $w_o(n)$ inevitably causes overlapping of the impulse responses presented in $c_e(n) - c_u(n)$, thereby disrupting its
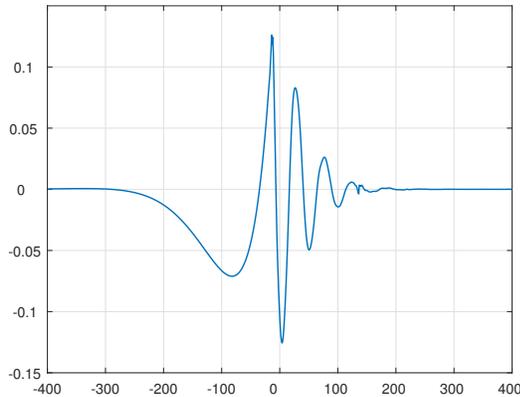
Fig. 6: Impulse response $w_o(n)$ of the stable optimal solution at the second scenario.

TABLE IV: Performance results of the proposed method in estimating $w_o(n)$ as function of speech length at second scenario.

Speech duration (ms)	Mean of MIS (dB)	Mean of $\mathrm{MIS}_<$ (dB)	Mean of $\mathrm{MIS}_>$ (dB)
100	8.43	3.68	21.91
250	5.02	3.20	15.67
500	-6.20	-8.37	5.18
1000	-16.38	-19.00	-3.53
1500	-19.30	-22.34	-6.78
2000	-25.54	-31.49	-11.65
3000	-26.57	-29.62	-14.30
4000	-28.32	-31.23	-16.75
5000	-34.08	-37.41	-21.28

(b)

Fig. 7: Estimates of $w_o(n)$ obtained by the proposed method at the second scenario for some speech lengths: (a) 1 s; (b) 3 s.

estimation. But some observations similar to the first scenario can be taken in the second one. As the use of larger windows reduces the spectral leakage and thus improves the accuracy of (12), the longer the speech signal the better the estimate obtained by the method. As the distortion in $c_e(n) - c_y(n)$ caused by the spectral leakage is fundamentally of high frequency, the method performance is significantly better below 500 Hz which is the frequency range common to the occlusion effect. And at least 0.5 s of speech is required for the method to obtain a minimally acceptable estimate of $w_o(n)$.

Scrutiny of Figure 7 reveals the influence of A_1 and A_2 on the method performance. As can be noted, the distortion caused by the spectral leakage prevents the method from accurately estimating the samples of $w_o(n)$ with low absolute values, especially its tails. Thus an appropriate choice of A_1 and A_2 must take into consideration the decay time of the left $(n < -N_a)$ and right $(n \geq -N_a)$ side of $w_o(n)$, respectively. And, since $N_w \geq M + A_2 - 2A_1$, increasing $|A_1|$ or A_2 increases N_w. Consequently, as discussed in Section III-B, the distortion caused by the spectral leakage intensifies and the method performance worsens. With this in mind, $A_1 = -300$ and $A_2 = 100$ were indiscriminately selected with the intention of estimating the highest absolute values of $w_o(n)$.

Nevertheless, for signals longer than 1.5 s, the method achieves average values of MIS and $\mathrm{MIS}_<$ less than -19 dB and -22 dB, respectively. And mean values of MIS and $\mathrm{MIS}_<$

approximately equal to -34 dB and -37 dB, respectively, are achieved for 5s-long speech signals. The results shown in Table IV reinforce that the proposed method accurately estimates the stable optimal solution $w_o(n)$ of the feedforward occlusion cancellation for speech signals longer than 1 s, which is a considerably short time for a person sustain a vowel sound. This is, as far as we know, a novelty in the literature.

An estimate of $w_o(n)$ can be very useful either in a controller design or at initializing the coefficients of an adaptive filter. In the sequel, the performance of the feedforward controller, proposed in Section III-C, that uses $\hat{w}_o(n)$ to attenuate the occlusion effect is evaluated.

B. Cancellation of the occlusion effect

The performance evaluation of the proposed controller $w_1(n)$, defined in (21), was carried out by using an estimate $\hat{w}_o(n)$ obtained with a 3 s-long speech signal. The frequency response magnitudes of both $w_1(n)$ and $w_o(n)$ for the first and second acoustic scenarios are shown in Figure 8. The frequency response magnitude of the solution of [12], proposed for acoustic paths with 14-sample delay propagation as the two paths used in this work, is also illustrated. A zoomed plot highlights the responses between 150 and 600 Hz, which includes the frequency range common to the occlusion effect.
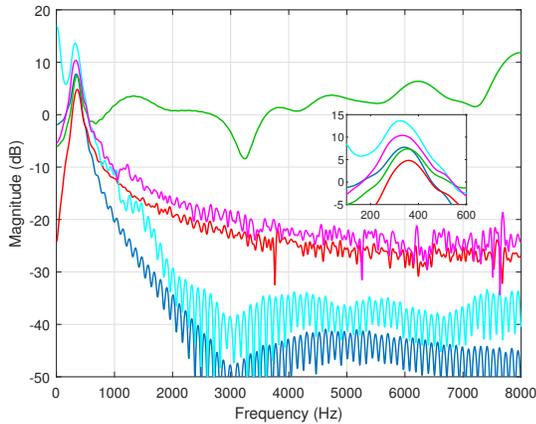
Fig. 8: Frequency response magnitude of the evaluated solutions: (blue) $w_o(n)$ at first scenario; (cyan) $w_o(n)$ at second scenario; (red) proposed $w_1(n)$ at first scenario; (magenta) proposed $w_1(n)$ at second scenario; (green) solution of [11].

1) Scenario 1: At the first acoustic scenario, it is noted from Figure 8 that the solution of [12] is closest to $w_o(n)$ up to approximately 420 Hz, more accurately modeling the peak around 340 Hz. From this frequency onwards, the proposed controller is closest to $w_o(n)$. Unlike $w_o(n)$, the solution of [12] does not intensely attenuate the frequencies above 650 Hz. On the contrary, it amplifies with an increasing average gain from 3600 Hz. This high frequency amplification may not greatly adversely affect its performance because the first three formant of vowel sounds are typically below 2500 Hz [30].

Regarding phase, it was found that the solution of [12] is very close to $w_o(n)$ up to 375 Hz, presenting an error of less than 10 degrees. And the proposed controller is close to the $w_o(n)$, presenting an error of less than 25 degrees, between 250 and 360 Hz, frequency range where the power magnification caused by the occlusion effect is concentrated. From then on, however, both solutions are generally far from $w_o(n)$.

The performance results of the proposed controller at the first scenario are presented in Table V, along with the results obtained with no control, the stable optimal solution $w_o(n)$ and the solution of [12]. As can be seen, $w_o(n)$ is really optimal because its mean P and MOS-LQO are null and maximum, respectively. Compared with none controller, the proposed controller decreased the mean P by 0.29 dB and increased the mean MOS-LQO by 0.14. And the standard deviation of both metrics is significantly reduced. These results reveal that the proposed controller successfully reduces the occlusion effect.

Also comparing with none controller, the solution of [12] raised the mean P and MOS-LQO by 4.49 dB and 0.13, respectively. These results suggest that, on average, this solution perceptually reduces the occlusion effect despite significantly distorting the sound heard by the hearing aid user. The distortion may be a consequence of the high frequency amplification previously discussed. Nevertheless, in light of the high variance associated with MOS-LQO, the audible performance difference may be considered negligible.

The results demonstrate that the proposed controller out-

TABLE V: Performance results of the evaluated controllers in attenuating the occlusion effect at first scenario.

Method	P (dB)		MOS-LQO	
	Mean	SD	Mean	SD
None	−32.77	3.10	4.42	0.15
Stable optimal solution $w_o(n)$	−343.79	2.23	4.64	0
Solution of [12]	−28.28	1.48	4.55	0.11
Proposed controller $w_1(n)$	−33.06	1.59	4.56	0.03

TABLE VI: Performance results of the evaluated controllers in attenuating the occlusion effect at second scenario.

Method	P (dB)		MOS-LQO	
	Mean	SD	Mean	SD
None	−32.77	3.10	4.26	0.23
Stable optimal solution $w_o(n)$	−336.73	2.10	4.64	0
Solution of [12]	−31.95	2.04	4.35	0.34
Proposed controller $w_1(n)$	−32.94	2.23	4.32	0.21

performed the solution of [12]. Although the 0.01 increase in mean MOS-LQO is negligible, the standard deviation is conveniently decreased by 0.08. Nonetheless, the 4.78 dB decrease in mean P is considerable given the associated variances and can serve as valid indication of its superior performance.

2) Scenario 2: At the second acoustic scenario, it is noted from Figure 8 that the proposed controller is closer to $w_o(n)$ than the controller of [12] over the entire frequency range. But the amplification level (higher than 10 dB) at the frequencies below 75 Hz is not modeled. This non-modeled magnitude peak may not adversely affect the sound quality obtained with both controllers because the W-PESQ filters its input signals with a high-pass filter with cutoff frequency at 100 Hz [21], [23]. In relation to phase, it was found that the solution of [12] is not close to $w_o(n)$ at the lower frequencies as in the first scenario. In fact, except at a very few frequencies, both controllers are far from $w_o(n)$ over the entire frequency range.

The performance results of the proposed controller at the second scenario are presented in Table VI, along with the results obtained with no control, the stable optimal solution $w_o(n)$ and the solution of [12]. The P values obtained with none controller are identical to the first scenario because, at this situation, the difference between $z(n)$ and its ideal value is equal to $o(n) * v(n)$. On the other hand, the MOS-LQO values differ because the W-PESQ is not premised on the mere difference between the reference and degraded signals. Furthermore, it is observed again that $w_o(n)$ is really optimal.

Compared with none controller, the proposed controller decreased the mean P by 0.17 dB and increased the mean MOS-LQO by 0.13. And the standard deviation of both metrics is slightly reduced. Such improvements are lower than those obtained in the first scenario but can serve as valid indication that the proposed controller also successfully reduces the occlusion effect in the second acoustic scenario.

The solution of [12] once again raised the mean P and MOS-LQO, now by 0.82 dB and 0.09 respectively, compared with none controller. Although the variations are smaller than in the first scenario, especially in P, these results reinforce the indication that, on average, this solution perceptually reduces the occlusion effect despite distorting the sound heard by user.

Contrasting with the solution of [12], the proposed controller achieved lower mean P and MOS-LQO. The 0.99 dB improvement in P is proportionally much larger than the 0.03 worsening in MOS-LQO. Despite the slight low mean MOS-LQO, the standard deviation is reduced by 0.13. These results indicate that the proposed controller performs better. Nevertheless, bearing in mind the variances associated with both metrics, all the performance differences discussed at the second acoustic scenario may be considered negligible.

VI. CONCLUSIONS

This work proposed a cepstral method to estimate the stable optimal solution for feedforward occlusion cancellation and a fixed controller that utilizes this estimate to reduce the occlusion effect in hearing aids. The estimation is performed at a calibration process, where the hearing aid user utters a sustained vowel, and using a feedback structure with appropriate choices of the forward and feedback paths.

Simulations have shown that the performance of the cepstral method improves as the length of the signal uttered by the user increases, resulting in estimates with average normalized misalignment less than -19 and -34 dB for signals lasting 1.5 and 5 s. The estimates are significantly more accurate below 500Hz, which is the frequency range common to the occlusion effect. In addition, results have pointed out that the controller successfully attenuated the occlusion effect, averagely decreasing by 0.17 dB the power of the distortion produced in the sound heard by the user and increasing by 0.13 the objective perceptual quality MOS-LQO score.

These results have demonstrated that the proposed cepstral method to estimate the stable optimal solution for the feedforward occlusion cancellation has qualities worthy of further exploring in the future, either to utilize its estimate to design an even more efficient fixed controller or as initialization of the coefficients of an adaptive filter.

APPENDIX A
INVERSE SYSTEM TO THE ACOUSTIC PATH

The impulse response of the acoustic path is defined as

$$a(n) = \tilde{a}(n) * \delta(n - N_a), \qquad (28)$$

where $\tilde{a}(n) = 0$ for $n < 0$ and $n > M$, $\tilde{a}(0) \neq 0$ and $N_a > 0$. Considering that the acoustic path is stable, its frequency response is given by

$$A(e^{j\omega}) = \left[\tilde{a}(0) + \tilde{a}(1)e^{-j\omega} + \ldots \tilde{a}(M)e^{-jM\omega} \right] \times e^{-jN_a\omega}. \qquad (29)$$

The frequency respose of the system inverse to the acoustic path is given by

$$A_I(e^{j\omega}) = e^{jN_a\omega} \frac{1}{\tilde{a}(0) + \tilde{a}(1)e^{-j\omega} + \ldots \tilde{a}(M)e^{-jM\omega}}. \qquad (30)$$

Defining p_k as the kth pole of $A_I(e^{j\omega})$ and assuming that there are no multiples poles, (30) can be expanded as

$$A_I(e^{j\omega}) = e^{jN_a\omega} \sum_{k=1}^{M} \frac{C_k}{1 - p_k e^{-j\omega}} \qquad (31)$$

Assuming, without loss of generality, that $|p_k| < 1$, $k = 1, 2, \ldots, M_1$, and $|p_k| > 1$, $k = M_1 + 1, \ldots, M$, then the impulse response $a_I(n)$ of the inverse system is defined as

$$a_I(n) = \delta(n + N_a) *$$
$$\left[\sum_{k=1}^{M_1} C_k p_k^n u(n) - \sum_{k=M_1+1}^{M} C_k p_k^n u(-n-1) \right]$$
$$= \sum_{k=1}^{M_1} C_k p_k^n u(n + N_a)$$
$$- \sum_{k=M_1+1}^{M} C_k p_k^n u(-n - 1 - N_a), \qquad (32)$$

where $u(n)$ is the unit step function.

Therefore, in general, $a_I(n)$ is two-sided signal, i.e., a signal that is of infinite extent for both $n \geq 0$ and $n < 0$, composed of left-sided increasing exponentials for $n < -N_a$ and right-sided decreasing exponentials for $n \geq -N_a$.

ACKNOWLEDGMENT

REFERENCES

[1] B. C. Bispo, R. C. Borges, and M. H. Costa, "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, September 2019. doi: 10.14209/SBRT.2018.160

[2] World Health Organization, "Deafness and hearing loss," https://www.who.int/en/news-room/fact-sheets/detail/deafness-and-hearing-loss, 2019.

[3] Instituto Brasileiro de Geografia e Estatística, "Censo demográfico brasileiro," 2010.

[4] T. Zahnert, "The differential diagnosis of hearing loss," *Deutsches ärzteblatt international*, vol. 108, no. 25, p. 433, 2011.

[5] S. Kochkin, "Marketrak vii: Customer satisfaction with hearing instruments in the digital age," *The Hearing Journal*, vol. 58, no. 9, pp. 30–32, 2005. doi: 10.1097/01.HJ.0000286545.33961.e7

[6] A. Spriet, I. Proudler, M. Moonen, and J. Wouters, "Adaptive feedback cancellation in hearing aids with linear prediction of the desired signal," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3749–3763, 2005. doi: 10.1109/TSP.2005.855108

[7] J. Mejia, H. Dillon, and M. Fisher, "Active cancellation of occlusion: An electronic vent for hearing aids and hearing protectors," *The Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 235–240, 2008. doi: 10.1121/1.2908279

[8] H. Dillon, *Hearing aids*. Hodder Arnold, 2008.

[9] D. J. MacKenzie, "Open-canal fittings and the hearing aid occlusion effect," *The Hearing Journal*, vol. 59, no. 11, pp. 50–52, 2006. doi: 10.1097/01.HJ.0000286218.69092.dd

[10] K. Chung, "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*, vol. 8, no. 4, pp. 125–164, 2004. doi: 10.1177/108471380400800402

[11] M. O. Hansen and M. R. Stinson, "Air conducted and body conducted sound produced by own voice," *Canadian Acoustics*, vol. 26, no. 2, pp. 11–19, 1998.

[12] R. C. Borges and M. H. Costa, "A feed forward adaptive canceller to reduce the occlusion effect in hearing aids," *Computers in biology and medicine*, vol. 79, pp. 266–275, 2016. doi: 10.1016/j.compbiomed.2016.10.016

[13] R. C. Borges, M. H. Costa, J. A. Cordioli, and L. F. Assuiti, "An adaptive occlusion canceller for hearing aids," in *Proceedings of 21st European Signal Processing Conference*. IEEE, 2013, pp. 1–5.

[14] M. Sunohara, M. Osawa, T. Hashiura, and M. Tateno, "Occlusion reduction system for hearing aids with an improved transducer and an associated algorithm," in *Proceedings of 23rd European Signal Processing Conference*. IEEE, 2015. doi: 10.1250/ast.35.318 pp. 285–289.

[15] 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, September 2015. doi: 10.1016/j.dsp.2015.03.003

[16] B. C. Bispo, P. M. L. Rodrigues, and D. R. S. Freitas, "Acoustic feedback cancellation based on cepstral analysis," in *Proceedings of 17th IEEE Conference on Signal Processing Algorithms, Architectures, Arrangements and Applications*, Poznan, Poland, September 2013, pp. 205–209.

[17] B. C. Bispo and D. R. S. Freitas, "A review of the recent cepstrum-based methods for acoustic feedback cancellation," in *Advances in Signal Processing: Reviews*, S. Y. Yurish, Ed. International Frequency Sensor Association Publishing, 2018, vol. 1, pp. 121–154.

[18] G. Fant, "A note on vocal tract size factors and non-uniform f-pattern scalings," *Speech Transmission Laboratory Quarterly Progress and Status Report*, vol. 1, pp. 22–30, 1966.

[19] ITU-T G.191, "Software tools for speech and audio coding standardization," International Telecommunications Union, 2016.

[20] ITU-T P.862, "Perceptual evaluation of speech quality (PESQ): objective method for end-to-end speech quality assessment of narrow band telephone networks and speech codecs," International Telecommunications Union, Geneva, Switzerland 2001.

[21] ITU-T P.862.2, "Wideband extension to recommendation P.862 for the assessment of wideband telephone networks and speech codecs," International Telecommunications Union, Geneva, Switzerland 2005.

[22] A. A. de Lima, F. P. Freeland, R. A. de Jesus, B. C. Bispo, L. W. P. Biscainho, S. L. Netto, A. Said, T. Kalker, R. W. Schafer, B. Lee, and M. Jam, "On the quality assessment of sound signals," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, Seattle, USA, May 2008. doi: 10.1109/ISCAS.2008.4541443 pp. 416–419.

[23] B. C. Bispo, P. A. A. Esquef, L. W. P. Biscainho, A. A. de Lima, F. P. Freeland, R. A. de Jesus, A. Said, B. Lee, R. W. Schafer, and T. Kaller, "EW-PESQ: A quality assessment method for speech signals sampled at 48 kHz," *Journal of the Audio Engineering Society*, vol. 58, no. 4, pp. 251–268, April 2010.

[24] ITU-T P.800.1, "Mean opinion score (MOS) terminology," International Telecommunications Union, Geneva, Switzerland 2006.

[25] A. A. de Lima, F. P. Freeland, P. A. A. Esquef, L. W. P. Biscainho, B. C. Bispo, R. A. de Jesus, S. L. Netto, R. W. Schafer, A. Said, B. Lee, and T. Kalker, "Reverberation assessment in audioband speech signals for telepresence systems," in *Proceedings of 11th International Conference on Signal Processing and Multimedia Applications*, Porto, Portugal, July 2008. doi: 10.5220/0001936602570262 pp. 257–262.

[26] A. A. de Lima, S. L. Netto, L. W. P. Biscainho, F. P. Freeland, B. C. Bispo, R. A. de Jesus, R. W. Schafer, A. Said, B. Lee, and T. Kalker, "Quality evaluation of reverberation in audioband speech signals," in *e-Business and Telecommunications - Communications in Computer and Information Science*, J. Filipe and M. S. Obaidat, Eds. Springer, 2009, vol. 48, pp. 384–396.

[27] ITU-T P.862.3, "Application guide for objective quality measurement based on recommendations P.862, P.862.1 and P.862.2," International Telecommunications Union, Geneva, Switzerland 2007.

[28] A. Castellana, A. Carullo, S. Corbellini, and A. Astolfi, "Discriminating pathological voice from healthy voice using cepstral peak prominence smoothed distribution in sustained vowel," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 3, pp. 646–654, 2018. doi: 10.1109/TIM.2017.2781958

[29] ITU-T P.56, "P. 56: Objective measurement of active speech level," 2011.

[30] J. R. Deller Jr., J. H. L. Hansen, and J. G. Proakis, *Discrete-Time Processing of Speech Signals*. New York: Wiley-IEEE Press, 2000.
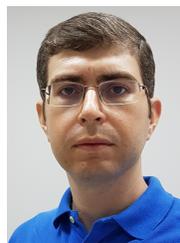
Bruno Catarino Bispo received the B.Sc. degree in electronics and computer engineering from the Federal University of Rio de Janeiro (UFRJ), Brazil, in 2006, the M.Sc. degree in electrical engineering from UFRJ in 2008, and the Ph.D. degree in electrical and computer engineering from the University of Porto, Portugal, in 2015. From 2009 to 2012, he was a Lecturer at the Faculty of Engineering of the University of Porto. From 2015 to 2019, he was an Assistant Professor at the Federal University of Technology - Paraná (UTFPR), Cornélio Procópio, Brazil. Since 2019, he has been an Assistant Professor in the Department of Electrical and Electronic Engineering at the Federal University of Santa Catarina (UFSC), Florianópolis, Brazil. His research interests are in digital signal processing, particularly speech processing and biomedical signal processing.

Renata Coelho Borges received the Technologist degree in telecommunications systems from Federal Institute of Santa Catarina (IFSC), Brazil, in 2009, and the M.Sc. and D.Sc. degrees, both in electrical engineering, from Federal University of Santa Catarina (UFSC), Brazil, in 2012 and 2017, respectively. Since 2018, she has been a Professor in the Department of Electrical Engineering at Federal University of Technology - Paraná (UTFPR), Cornélio Procópio, Brazil. Her research interests are in the area of biomedical signal processing, specifically for hearing aids. It focuses on active noise control and adaptive filters for reducing the occlusion effect and acoustic feedback.