Signal Processing 119 (2016) 153–161

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

# A method of howling detection in presence of speech signal

Soudeh A. Khoubrouy *, Issa Panahi

*Electrical Engineering Department, University of Texas at Dallas, Richardson, USA*

ARTICLE INFO

*Article history:*
Received 31 July 2014
Received in revised form
13 July 2015
Accepted 22 July 2015
Available online 18 August 2015

*Keywords:*
Hearing aid
Acoustic feedback
Adaptive feedback cancellation
Howling detection

ABSTRACT

Hearing aid users suffer from howling sound caused by acoustic coupling between the loudspeaker and the microphone(s) of this device. It is crucial to detect and eliminate the howling before it causes serious irritation to the hearing aid user. This study presents a multiple-feature method which uses voice activity detection (VAD) algorithm to reduce false alarm probability. Experimental results compare the performance of the proposed method with three conventional howling detection techniques in terms of detection probability, false alarm probability, and computational complexity. The proposed method possesses lower false alarm probability and less computational complexity compared to the other methods.

Published by Elsevier B.V.

## 1. Introduction

The small size of hearing aid devices allows a signal leakage, called acoustic feedback, between the loudspeaker and the microphone(s) [1]. Acoustic feedback can cause severe signal distortions and even an annoying howling sound. The closed-loop system produced by the acoustic feedback can become unstable depending on amplification gain, thereby causing the howling phenomenon which is an oscillation or a resonance with unpleasant sound. Several methods have been proposed to cancel the effects of howling [2–6].

Three main characteristics should be considered in a howling detection algorithm, lack of any of them can make the method unreliable (i.e., it is crucial to detect the howling in its initial stages before its high gain makes it intolerable for the user. It is also important to estimate the howling frequency component correctly. Moreover, the detection algorithm should have low computational complexity).

Howling has been shown to occupy certain time or frequency characteristics which are effective for detection [2,3]. Some howling detection methods compute the power of a frequency component and compare it with any of the following references: a threshold value, the average power of the current frame of the input signal, or the powers of different harmonics of that frequency [2,3]. However, due to some similarities, the howling frequency component cannot be easily recognized from the signal tonal components (e.g. formants in speech, or music tones) [3].

Existing howling detection methods could be classified into two major categories, i.e. frame-based and sample-based methods. Frame-based methods often process the Short Time Fourier Transform (STFT) of the input signal and check some frequency domain properties or temporal features of the howling [3]. Peak-to-Threshold Power Ratio (PTPR), Peak-to-Harmonic Power Ratio (PHPR), Peak-to-Neighboring Power Ratio (PNPR), Interframe Peak Magnitude Persistence (IPMP), and Interframe Magnitude slope Deviation (IMSD) are some known methods in this category [3]. Sample-based methods process the input signal sample by sample in time domain. Teager-Kaiser based method [2,5] and Adaptive Notch Filter (ANF) method [4] are classified in this category. Adaptive non-linear rate-level function which uses the model of auditory system can be also used to reduce the effect of howling [6].

This study proposes a frame-based howling detection approach whose false alarm probability (probability of

* Corresponding author. Tel.: +1 9724007218; fax: +1 9728832710.
*E-mail address:* sa.khoubrouy@gmail.com (S.A. Khoubrouy).

http://dx.doi.org/10.1016/j.sigpro.2015.07.027
0165-1684/Published by Elsevier B.V.

detecting a non-howling component as a howling component) is reduced by applying a voice activity detection (VAD) algorithm in the pre-processing step. In addition to applying the VAD algorithm, some properties of the howling components, i.e. large power, increasing power in the early stages, and negligible harmonic power, are also checked to provide an algorithm with lower false alarm probability. Moreover, as the howling detection procedure does not occur unless an alarm is received from the VAD unit, the computational complexity of the system is low.

Performance of the proposed method is compared with three known howling detection methods. The first one is the single-feature PHPR approach which recognizes a howling sound based on one of the howling main properties, (i.e., not exhibiting significant power at the harmonics and sub-harmonics of the howling frequency). Among the single-feature howling detection algorithms explained in [3], the PHPR approach has previously been shown to achieve good performance. Therefore, PHPR is chosen as a reference method for comparison. The second method is another frame-based method which is a multiple-feature approach proposed by Waterschoot and Moonen [3]. ANF method, a frame-based method, is selected as the third reference approach for comparison. Experimental results show lower false alarm probability and less computational complexity for the proposed method.

The paper is organized as follows. Section 2 explains acoustic feedback and the instability it can cause in the hearing aid. Section 3 presents some conventional howling detection methods. The proposed method and the benefits of using VAD are explained in Section 4. Section 5 compares the computational complexity of the proposed algorithm with the other methods. Sections 6 and 7 present the experimental results and conclusion.

## 2. Acoustic feedback

Fig. 1 shows a simple diagram for a typical hearing aid. The input signal, $y(n)$, including the desired signal, $s(n)$, and the feedback signal, $\hat{s}(n)$, is collected by the microphone. This signal is amplified by a gain, $G(\omega)$, which is usually a function of frequency and is compatible to the hearing loss level of the patient. The amplified signal, $u(n)$, is fed into the loudspeaker from which the signal leaks back to the microphone as $\hat{s}(n)$. The acoustic feedback path creates a closed-loop structure in Fig. 1. The closed loop system can be written as

$$U(\omega)/S(\omega) = G(\omega)/[1 - G(\omega)F(\omega)] \quad (1)$$

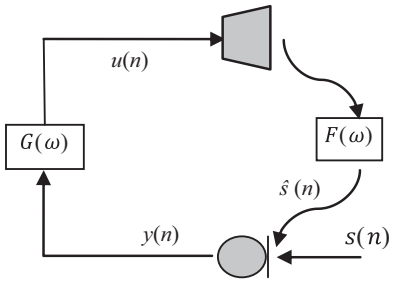


**Fig. 1.** A simple block diagram of a hearing aid device.

$U(\omega)$ and $S(\omega)$ are Fourier transforms of $u(n)$ and $s(n)$. $G(\omega)$ and $F(\omega)$ are the transfer functions for the forward path (gain of hearing aid) and the acoustic feedback path, where in this paper both are assumed fixed over the algorithm implementation time. According to (1) and the Nyquist instability criterion [7], the system is unstable at frequency $f$ if

$$\begin{cases} |G(\omega)F(\omega)| \geq 1 \\ \measuredangle(G(\omega)F(\omega)) = 2\pi K \end{cases} \quad (2)$$

where, $\omega = 2\pi f/f_s$ and $K$ is an integer. If the unstable system is excited by an input signal which contains a non-zero frequency component at the above frequency, then an oscillation or howling occurs [7].

The howling has an oscillatory/sinusoidal nature, which appears as a large peak in the frequency domain [3]. Using sampling frequency of 16000 Hz, Fig. 2 shows the time domain representation of a speech signal containing the howling and oscillation at frequency of 2100 Hz. The initial stage of the howling is better shown in the lower panel and its frequency can be seen in the next figure.

Fig. 3 is the spectrogram of the previous signal. It shows the frequency domain representations of:

- Some phonemes especially vowels (e.g., around instances 0.2 s and 1.2 s) [8].
- Initial stage of howling (around instance 1.3 s).
- Final stage of howling (howling with a dominant peak around instance 1.5 s).

According to this figure, the formants may be mistakenly detected as the howling frequency because of having large magnitudes. Moreover, waiting to recognize the howling component when its magnitude gets larger than the typical magnitude of formants causes some irritation to the hearing aid users. Therefore, it is essential to recognize these two different components by considering other characteristics of the howling.

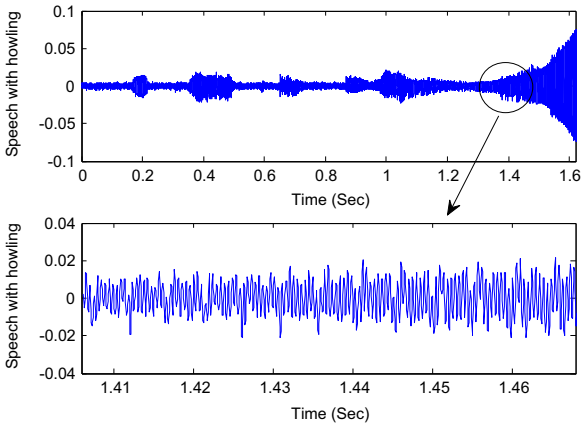


**Fig. 2.** Upper panel: Speech signal with howling which starts around instance 1.35 s. Lower panel: Early stage of howling.

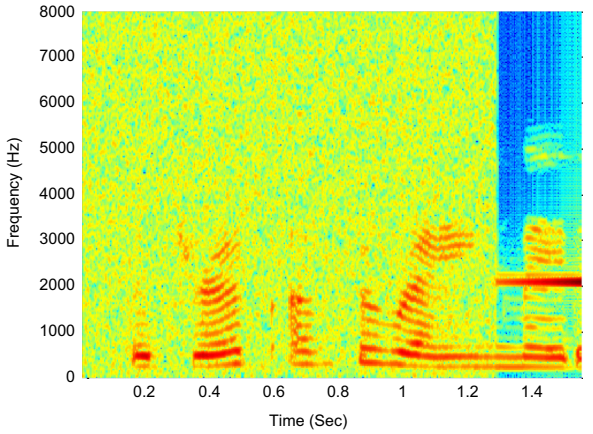


**Fig. 3.** Spectrogram of the speech signal which contains howling around instance 1.35 s.

## 3. Howling detection methods

The howling frequency component can be recognized because of its large frequency magnitude which is also one of the characteristics of formants in speech signals. The howling sound has specific properties in both the frequency and time domains that can be used for detection. Normally, it does not have harmonic or sub-harmonic frequency components [2,3]. In addition, it has increasing magnitude in the time domain until the hearing aid (or sound reinforcement system) saturates [3].

### 3.1. Frame-based methods

Howling detection algorithms usually process the input signal on a frame by frame basis. The STFT is computed for each frame where $N_{Peak}$ peaks of the spectrum are selected as the "howling component candidate" in the set $D_{\breve{\omega}}(l)=\{\breve{\omega}_i\}_{i=1}^{N_{Peak}}$, where $N_{Peak}$ is normally in the range of 1–10 [2,3]. The final howling component is chosen by checking one of the time or frequency domain properties mentioned above. Some of the algorithms are briefly described below.

#### 3.1.1. Peak-to-threshold power ratio (PTPR)

The candidate $\breve{\omega}_i$ is chosen as the howling component if its power is larger than a pre-defined threshold $P_{\mathrm{PTPR}}$ [3]:

$$PTPR(\breve{\omega}_i,l)=10log_{10}\left(\left|Y(\breve{\omega}_i,l)\right|^2\right)>T_{\mathrm{PTPR}}. \tag{3}$$

$Y(\breve{\omega}_i,l)$ is the value of the STFT for frame $l$ and frequency $\breve{\omega}_i$.

#### 3.1.2. Peak-to-harmonic power ratio (PHPR)

$\breve{\omega}_i$ is indicated as the howling component if the powers of its harmonics are negligible. The power of the candidate $\breve{\omega}_i$ is compared with the power of its $m$th harmonic by [3]:

$$PHPR(\breve{\omega}_i,l,m)=10log_{10}\left(\left|Y(\breve{\omega}_i,l)\right|^2/\left|Y(m\breve{\omega}_i,l)\right|^2\right). \tag{4}$$

Then, $\breve{\omega}_i$ is recognized as the howling component if

$$\cap_{m\in\{0.5,1.5,2,3,4\}}\left[PHPR(\breve{\omega}_i,l,m)\geq T_{PHPR}\right]=1 \tag{5}$$

where $\cap$ is the intersection operator and $T_{PHPR}$ is a predefined threshold [3].

#### 3.1.3. Peak-to-neighboring power ratio (PNPR)

Howling components have larger power compared to their neighboring frequency components. Therefore, $\breve{\omega}_i$ is selected as a howling component if it fulfills the following condition [3]:

$$PNPR(\breve{\omega}_i,l,m)=10log_{10}\left(\left|Y(\breve{\omega}_i,l)\right|^2/\left|Y\left(\breve{\omega}_i+\frac{2\pi m}{M},l\right)\right|^2\right)>T_{PNPR} \tag{6}$$

where $T_{PNPR}$ is the pre-defined threshold [3].

#### 3.1.4. Interframe peak magnitude persistence (IPMP)

It is a temporal feature indicating the number of frames out of $N_{Frame}$ frames where the frequency $\breve{\omega}_i$, is considered as a howling frequency candidate [3].

$$IPMP(\breve{\omega}_i,l)=\frac{1}{N_{Frame}}\sum_{j=0}^{N_{Frame}-1}\left[\breve{\omega}_i\in D_{\breve{\omega}}(l-j)\right] \tag{7}$$

#### 3.1.5. Interframe magnitude slope deviation (IMSD)

It is a temporal feature which shows the deviation of the slope over $N_{Frame}$ successive frames. Howling components exhibit nearly linear magnitude increase in time (dB scale). This corresponds to nearly constant slope or small values of the IMSD [3].

$$\begin{aligned}IMSD(\breve{\omega}_i,l)&=\frac{1}{N_{Frame}-1}\sum_{m=1}^{N_{Frame}-1}\\&=\left[\frac{1}{N_{Frame}}\sum_{j=0}^{N_{Frame}-1}\frac{1}{N_{Frame}-j}\Big(10\log_{10}\left|Y(\breve{\omega}_i,l-j)\right|^2\right.\\&\quad-10\log_{10}\left|Y(\breve{\omega}_i,l-N_{Frame})\right|^2\Big)\cdot\\&\quad\left.-\frac{1}{m}\sum_{j=0}^{m-1}\frac{1}{m-j}\Big(10\log_{10}\left|Y(\breve{\omega}_i,l-j)\right|^2-10\log_{10}\left|Y(\breve{\omega}_i,l-m)\right|^2\Big)\right]\end{aligned} \tag{8}$$

#### 3.1.6. Multiple-feature howling detection criteria

Multiple of aforementioned criteria or methods can be combined to obtain a more robust howling detection method. Waterschoot et al. apply PHPR, PNPR, and IMSD algorithms together to obtain lower false alarm probability [3].

### 3.2. Sample-based methods

A few howling detection methods process the input signal in time domain and sample by sample, where the ANF is one of them [4].

#### 3.2.1. Adaptive notch filter (ANF)

Spectral peaks of the speech signal can be tracked and identified by the ANF approach. In this method, the input signal is filtered by a second-order ANF with the following transfer function.

$$H_a(z)=\frac{1-a(n)z^{-1}+z^{-2}}{1-\rho a(n)z^{-1}+\rho^2z^{-2}} \tag{9}$$

where $\rho$ and $a(n)$ are constant and variable parameters, respectively. The ANF adjusts the center frequency of the

notch filter by adjusting parameter $a(n)$ in a way that at time $n$ the power of the output of the filter is reduced [4]. The update equations and parameters are described in [4]. The variability of $a(n)$ is small if the system is tracking a strong frequency component. Hence, Pandey et al. use this property to detect the howling and to identify its frequency component.

## 4. Proposed howling detection method

### 4.1. Proposed VAD application

To reduce the chance of detecting a formant as a howling component, a VAD algorithm has been used by Pandey et al. [4] to detect the silence parts and activate the ANF method in these parts. As described below, the VAD application is different in this paper.

Fig. 2 has shown that the howling starts with a low-energy oscillation whose energy/magnitude increases gradually. Although the magnitude of the howling is small at the beginning, once howling happens there is no silence part in the microphone signal. This can be used as a clue to detect the start of the howling. Using a VAD algorithm specifies if the current frame is speech or silence. If the VAD algorithm does not recognize any silence frame for $N_1$ frames, the system is in one of the following situations:

- A vowel is being pronounced.
- A consonant [8] is being pronounced.
- Howling is occurring.

Consequently, the howling detection algorithm is activated to recognize which one of the above situations is happening. Therefore, the howling detection method is not executed all the time, which means that finding the STFT and controlling one criterion or more (from those mentioned in Section 3), are not performed for every frame of the data. This brings two benefits:

- Lower computational complexity.
- Lower chance of detecting a formant as a howling component.

Therefore, the idea of using VAD can be applied to any of the howling detection techniques described in Section 3 to reduce false alarm probability and computational complexity.

### 4.2. Proposed howling detection method

The howling detection algorithm is activated when the VAD does not recognize any silence frame for $N_1$ frames. If the howling is not in its initial stage, it can be easily detected by its large frequency amplitude. Therefore, if a candidate has had the largest magnitude, also greater than $T_{PTPR}$, over last $N_{PTPR}$ frames (combination of the IPMP and the PTPR criteria) it is recognized as a howling frequency component without checking the other criteria. The value of $N_{PTPR}$ is between 1 and 10. One can see that larger $N_{PTPR}$ improves the accuracy but increases the detection time, too. If any of the PTPR and the IPMP criteria is not satisfied, there is not a common dominant peak in last $N_{PTPR}$ frames; which means that if a howling is occurring it is still in the early stages. Hence, the algorithm starts to distinguish between early stages of the howling and vowel/consonant pronunciation. In this situation, the algorithm checks the power slopes of the candidates in last $N_{Frame}$ frames. In other words, the power of each candidate is compared with its value in the previous frame (if the frequency has been a candidate in both frames). The percentage of power increase/decrease is saved as a positive/negative slope. Apparently, a howling component should have increasing power over time. Therefore, a candidate with maximum positive slope during last $N_{Frame}$ (again, $N_{Frame}$ is an integer between 2 and 10) frames is selected and to ensure that it is not a formant, the PHPR is also checked. The candidate is finally selected if its PHPR has the largest value among the PHPR values of the other candidates. In this method no threshold has been selected for PHPR or power slope criterion and the most reasonable candidate is always selected as a howling frequency component. Whenever the VAD algorithm detects a silence frame, all the counters are reset and the howling detection algorithm is deactivated. Fig. 4 represents the flowchart of the proposed method.

As a result, the proposed algorithm reduces false alarm probability by considering the following policy:

Howling detection method is not executed during the silent parts and first $N_1$ frames of vowel/consonant pronunciation. Since the consonants are typically shorter than the vowels [9], by having proper value for $N_1$ the detection algorithm is not executed during consonant pronunciation. The chance of detecting a formant instead of the howling component is high during the pronuncia-

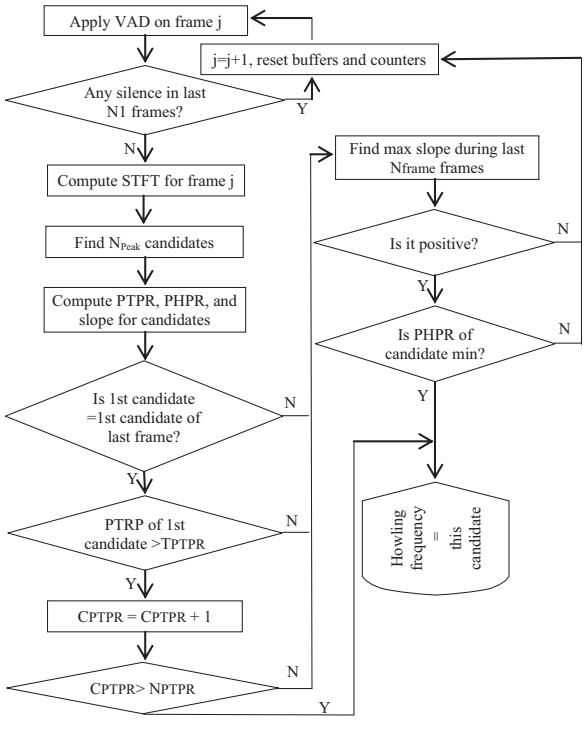


**Fig. 4.** The flowchart for the proposed method.

tion of a vowel. By ignoring first $N_1$ frames of pronunciation of a vowel, these frames will not be considered in the howling detection process.

Selecting a dominant peak in $N_{PTPR}$ frames also reduces the chance of false detection.

- Checking the slope of the power increase and choosing the candidate with the highest positive slope is another test to reduce false detection.
- Finally, checking the power of the second harmonic is another assessment to reduce false detection.

The VAD algorithm originally selected for the proposed method is a simple energy-based VAD algorithm in which the energy is computed for each frame of data by [10].

Assuming that the first few frames of the signal are background noise, the energy of the noise, $E_r$, is estimated by averaging the energies of these frames [10]. For the $l$th frame, the decision is made as follows:

$$\begin{cases} \text{If } E_l > TE_r & \textit{Frame is voice.} \\ \textit{otherwise} & \textit{Frame is silence.} \end{cases} \tag{10}$$

$T$ is a constant threshold and the energy of noise is updated by:

$$E_{rnew} = (1-p)E_{rold} + pE_{silence} \tag{11}$$

where $p$ is a constant in the range of [0,1], $E_{rnew}$ is the updated/new value of the noise energy, $E_{rold}$ is the previous/old value of it, and $E_{silence}$ is the energy of the most recent noise (silence) frame.

For a speech signal (as the desired signal of the hearing aid) without any background noise, this simple VAD algorithm shows enough accuracy; and a more complicated VAD algorithm is not required. However, if the input signal is degraded by noise, the VAD algorithm should check more features from the speech and noise to make more accurate decision, e.g., Sohn et al. [11] and Shin et al. [12] have proposed statistical model-based VAD algorithms. For better comparison, the proposed method is also tested with noisy speech files and the Energy-based VAD and statistical model-based VAD are evaluated.

## 5. Computational complexity

This section compares the computational complexity of the methods. Tables 1–3 present the computational complexities of all the methods in terms of number of multiplications, additions, and comparisons required per sample. Table 4 summarizes these results applying the values used in the simulations, i.e., $M$ (number of samples in each frame) $=320$, $M_{PHPR}$ (number of harmonics and sub-harmonics)$=5$, $M_{PNPR}$ (number of neighboring frequencies)$=4$, $N_{FFT}$ (number of FFT bins)$=512$, $N_{Peak}$ (number of selected peaks)$=4$, $N_{frame}$ (number of frames)$=7$. As it is mentioned before, the proposed method can check different features of the howling but this checking is not necessarily happened in each frame. Therefore, the computational complexity is very low for those frames that VAD algorithm does not activate the howling detection method. As a result, the average computational complexity of the proposed method is lower than the other methods. It should be mention that the VAD algorithm considered in these tables is the Energy-based VAD. Since the statistical model-based VAD algorithm is more complicated to be represented in the tables, we computed its complexity based on its execution time. The average time for the statistical model-based VAD is 6.46 times of the required time for Energy-based VAD algorithm. Therefore, adding a more accurate VAD will increase the complexity of the algorithm.

## 6. Experimental results

The proposed howling detection method is implemented here and compared with three of the methods shown in Section 3. Among single-feature howling detection methods, the PHPR approach is selected to compare the proposed method with. ANF algorithm and multiple-feature howling detection criteria proposed in [3] are also simulated for better comparison. Moreover, the effect of combing the VAD algorithm with the PHPR and ANF methods are tested.

The desired inputs are the files of Noizeus database [13]; i.e. 30 IEEE standard speech files with sampling frequency of 16 kHz. The hearing aid device starts working in the steady situation without any howling occurrence. The gain of

**Table 1**
Number of multiplications per sample for each method.

| | Number of multiplications/Sample | |
|---|---|---|
| PTPR | $\left(\frac{N_{FFT}}{2}\log_2 N_{FFT} + N_{Peak}\right)/M$ | |
| PHPR | $\left(\frac{N_{FFT}}{2}\log_2 N_{FFT} + (4M_{PHPR}-1)N_{Peak}\right)/M$ | |
| PNPR | $\left(\frac{N_{FFT}}{2}\log_2 N_{FFT} + (4M_{PNPR}-1)N_{Peak}\right)/M$ | |
| IPMP | $\left(\frac{N_{FFT}}{2}\log_2 N_{FFT} + N_{Peak}\right)/M$ | |
| IMSD | $\left(\frac{N_{FFT}}{2}\log_2 N_{FFT} + \left(1.5{N_{Frame}}^2 + 0.5N_{Frame}\right)N_{Peak}\right)/M$ | |
| PHPR+ PNPR+ IMSD | $\left(\frac{N_{FFT}}{2}\log_2 N_{FFT} + \left(4M_{PHPR} + 4M_{PNPR} + 1.5{N_{Frame}}^2 + 0.5N_{Frame}\right)N_{Peak}\right)/M$ | |
| ANF | 14 | |
| Proposed method using VAD | VAD | $(N_{FFT}+2)/M$ |
| | VAD+PTPR+IPMP | $\left(\frac{N_{FFT}}{2}\log_2 N_{FFT} + 2N_{Peak} + 1\right)/M$ |
| | VAD+PTPR+IPMP+PHPR+Slope | $\left(\frac{N_{FFT}}{2}\log_2 N_{FFT} + 6N_{Peak} + 1\right)/M$ |

**Table 2**
Number of additions per sample for each method.

| | Number of additions/Sample | |
|---|---|---|
| PTPR | $(N_{FFT}\log_2 N_{FFT})/M$ | |
| PHPR | $(N_{FFT}\log_2 N_{FFT})/M$ | |
| PNPR | $(N_{FFT}\log_2 N_{FFT})/M$ | |
| IPMP | $(N_{FFT}\log_2 N_{FFT}+(N_{Frame}-1)N_{Peak})/M$ | |
| IMSD | $\left(N_{FFT}\log_2 N_{FFT}+\left(3{N_{Frame}}^2-3N_{Frame}\right)N_{Peak}\right)/M$ | |
| PHPR+ PNPR+ IMSD | $\left(N_{FFT}\log_2 N_{FFT}+\left(3{N_{Frame}}^2-3N_{Frame}\right)N_{Peak}\right)/M$ | |
| ANF | 11 | |
| Proposed method using VAD | VAD | $(N_{FFT}+3)/M$ |
| | VAD+PTPR+IPMP | $(N_{FFT}+3+N_{FFT}\log_2 N_{FFT}+(N_{Frame}-1)N_{Peak})/M$ |
| | VAD+PTPR+IPMP+PHPR+Slope | $(N_{FFT}+3+N_{FFT}\log_2 N_{FFT}+N_{Frame}N_{Peak})/M$ |

**Table 3**
Number of comparisons per sample for each method.

| | Number of comparisons/Sample | |
|---|---|---|
| PTPR | $(N_{FFT}\log_2 N_{FFT}+N_{Peak})/M$ | |
| PHPR | $(N_{FFT}\log_2 N_{FFT}+(M_{PHPR}+1)N_{Peak})/M$ | |
| PNPR | $(N_{FFT}\log_2 N_{FFT}+(M_{PNPR}+1)N_{Peak})/M$ | |
| IPMP | $(N_{FFT}\log_2 N_{FFT}+(N_{Peak}N_{Frame}+1)N_{Peak})/M$ | |
| IMSD | $(N_{FFT}\log_2 N_{FFT}+N_{Peak})/M$ | |
| PHPR+PNPR+IMSD | $(N_{FFT}\log_2 N_{FFT}+(M_{PHPR}+M_{PNPR}+3)N_{Peak})/M$ | |
| ANF | 2 | |
| Proposed method using VAD | VAD | $2/M$ |
| | VAD+PTPR+IPMP | $(2+N_{FFT}\log_2 N_{FFT}+(N_{Peak}N_{Frame}+2)N_{Peak})/M$ |
| | VAD+PTPR+IPMP+PHPR+Slope | $(2+N_{FFT}\log_2 N_{FFT}+(N_{Peak}N_{Frame}+4)N_{Peak})/M$ |

**Table 4**
Computational complexity of the methods.

| | Multiplications/Sample | | Additions/Sample | Comparisons/Sample |
|---|---|---|---|---|
| PTPR | $2308/M=7.2$ | | $4608/M=14.4$ | $4612/M=14.4$ |
| PHPR | $2380/M=7.4$ | | $4608/M=14.4$ | $4632/M=14.5$ |
| PNPR | $2364/M=7.4$ | | $4608/M=14.4$ | $4628/M=14.5$ |
| IPMP | $2308/M=7.2$ | | $4632/M=14.5$ | $4724/M=14.8$ |
| IMSD | $2612/M=8.2$ | | $5112/M=16$ | $4612/M=14.4$ |
| PHPR+PNPR+IMSD | $2756/M=8.6$ | | $5112/M=16$ | $4656/M=14.6$ |
| ANF | 14 | | 11 | 2 |
| Proposed method using VAD | VAD | $514/M=\mathbf{1.6}$ | $515/M=\mathbf{1.6}$ | $2/M=\mathbf{0.01}$ |
| | VAD+PTPR+IPMP | $2313/M=\mathbf{7.2}$ | $5147/M=\mathbf{16.1}$ | $4730/M=\mathbf{14.8}$ |
| | VAD+PTPR+IPMP+PHPR+Slope | $2329/M=\mathbf{7.3}$ | $5151/M=\mathbf{16.1}$ | $4738/M=\mathbf{14.8}$ |

hearing aid then increases up to a level that generates howling. The first 200 msec of each file is used for estimating the noise energy. The value of $p$ used in (11) is 0.1. Among the total 31832 frames of speech used in the simulation 2118 frames contain howling. Howling detection methods are typically evaluated using receiver operating characteristic (ROC) curves [4]. An appropriate howling detection method should have high detection probability ($P_D$) and low false alarm probability ($P_{FA}$). In an ROC curve, $P_D$ is plotted against $P_{FA}$; and overall performance of the algorithm is evaluated. Moreover, the appropriate threshold used in the howling detection algorithm can be selected based on this curve. In addition to that, the algorithms are also compared based on detection time which shows the amount of delay (in millisecond) between the howling occurrence and its detection by the algorithms.

Fig. 5 shows the ROC curves for the single-feature algorithm based on PHPR criterion. The curves are obtained by changing the values of $T_{PHPR}$. Therefore, appropriate value of $T_{PHPR}$ is selected based on the ROC curves by compromising between $P_D$ and $P_{FA}$. The upper panel corresponds to the original method explained in Section 3, while the lower panel represents the effect of applying VAD on this algorithm. Three different values, i.e. 10, 20, and 30 are considered for $N_1$. Fig. 6 compares the detection times of these algorithms. Comparing Figs. 5 and 6, one can realize that applying VAD on the PHPR criterion reduces $P_{FA}$ significantly but degrades $P_D$ and detection time at the

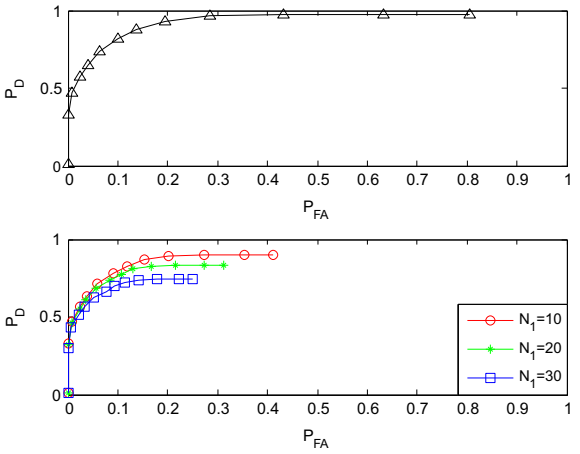


**Fig. 5.** ROC curves corresponding to PHPR algorithm (upper panel) and PHPR algorithm with VAD (lower panel).

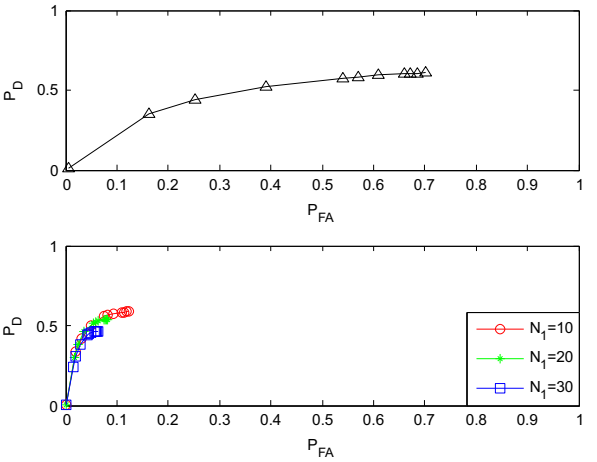


**Fig. 7.** ROC curves corresponding to ANF algorithm (upper panel) and ANF algorithm with VAD (lower panel).

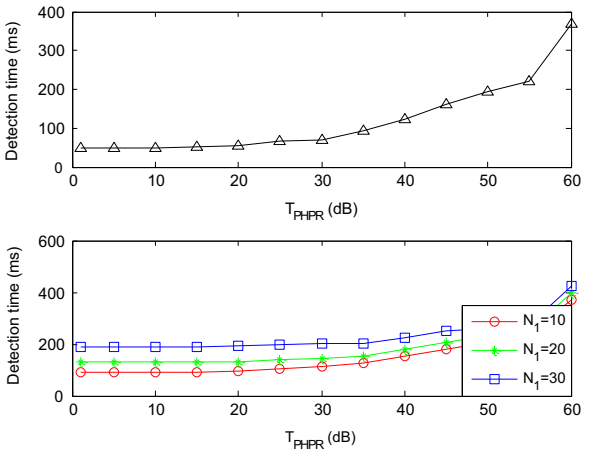


**Fig. 6.** Detection time corresponding to PHPR algorithm (upper panel) and PHPR algorithm with VAD (lower panel).

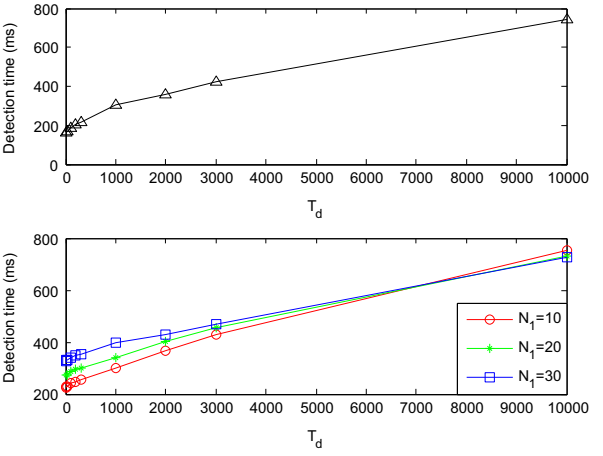


**Fig. 8.** Detection time corresponding to ANF algorithm (upper panel) and ANF algorithm with VAD (lower panel).

same time. Apparently, choosing $N_1 = 10$ gives the best trade-off among different values of $N_1$.

Figs. 7 and 8 correspond to ANF algorithm. This method has several parameters [4] where one of them ($T_a$) whose changes have more effect on $P_D$ and $P_{FA}$ is chosen as the variable. The other parameters are chosen in such a way to maximize the performance ($\rho = 0.95$, $\alpha_a = 0.1$, $\lambda_x = 0.9$, $\delta_q = 0.05$, and $\lambda_m = 0.999$). This algorithm has lower detection probability and longer detection time compared to the PHPR algorithm. Again, applying VAD on this algorithm reduces $P_{FA}$.

Fig. 9 represents the ROC curves for the multiple-feature algorithm introduced in Section 3.1.6. Since the method combines PHPR, PNPR, and IMSD criteria, three different thresholds are involved (i.e., $T_{PHPR}$, $T_{PNPR}$, and $T_{IMSD}$). Values of $T_{PHPR}$ are changed to generate each ROC curve. In this figure, several ROC curves are plotted corresponding to different values of $T_{PNPR}$ and $T_{IMSD}$. The selected values for these thresholds are: $T_{PNPR} = 4$ and 6 dB and $T_{IMSD} = 2$, 5, and 10 dB. The suggested values for $T_{IMSD}$ are typically smaller than 2 dB in other references, i.e. 0.5 dB [3]. However, using 0.5 dB for the speech files under the test reduces $P_D$ dramatically. Therefore, larger threshold is selected to decrease the limiting effect of IMSD criterion. In other words, $T_{IMSD} =$ 10 dB almost eliminates the effect of IMSD criterion. Noticing that the scale on $P_{FA}$ axis is different from Figs. 5 and 7, one can easily observe that this algorithm reduces $P_{FA}$ compared to single-feature methods. Apparently, combining multiple criteria improves$P_{FA}$, but $P_D$ can never be greater than $P_D$ of those single-feature methods [3]. Fig. 10 represents the detection time for this algorithm.

Fig. 11 evaluates the proposed algorithm. Each ROC curve is plotted by changing the values of $T_{PTPR}$. The upper panel in Fig. 11 shows that the proposed algorithm possesses the lowest $P_{FA}$ among the other algorithms (Again notice the scale change on $P_{FA}$ axis). Fig. 12 evaluates the multiple-feature howling detection method and the proposed method using noisy speech signals (i.e., the speech files degraded by 5 dB SNR white Gaussian noise). The upper panel represents the ROC curves for the multiple-feature method. As it can be seen, $P_D$ has not changed significantly but $P_{FA}$ has increased. The middle and lower panels represent the ROC curves corresponding

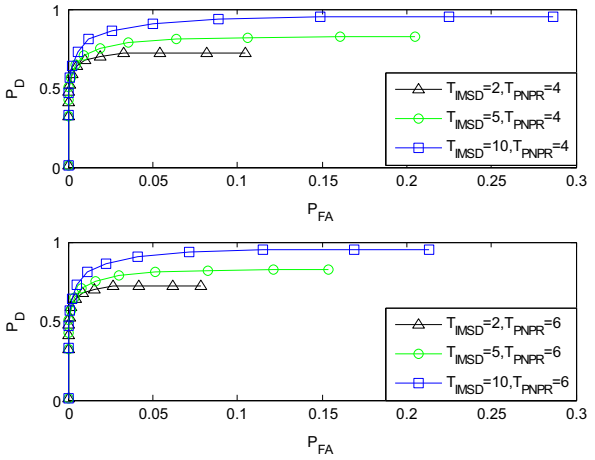


**Fig. 9.** Multiple-feature howling detection method using PHPR, PNPR, and IMSD criteria.

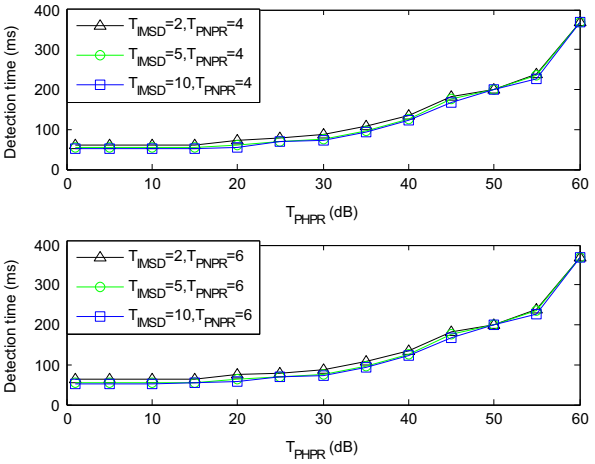


**Fig.10.** Detection time for multiple-feature howling detection method using PHPR, PNPR, and IMSD criteria.

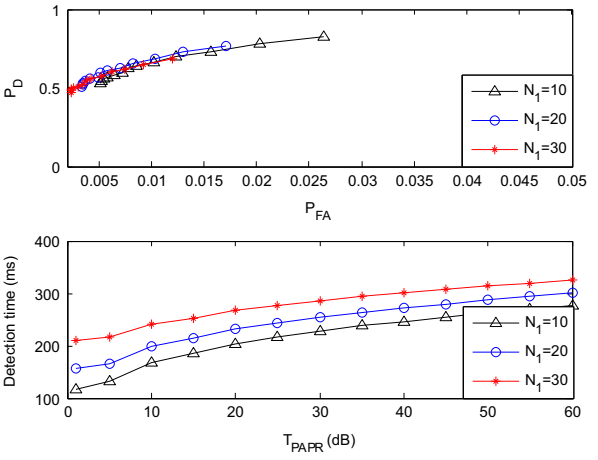


**Fig. 11.** ROC curves (upper panel) and detection time (lower panel) corresponding to the proposed algorithm.

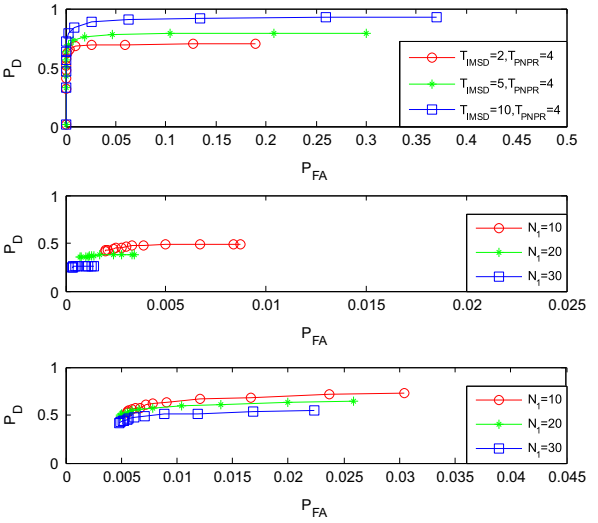


**Fig. 12.** ROC curves for noisy speech signals: upper panel: the multiple-feature howling detection method using PHPR, PNPR, and IMSD criteria, middle panel: Proposed method using Energy-based VAD, and lower panel: the proposed method using statistical model-based VAD.

to the Energy-based VAD and statistical model-based VAD, respectively. Compared to the Energy-based VAD, the statistical model-based VAD improves $P_D$ by about 0.25 (25% ), i.e., from 0.5 to about 0.75 and degrades the $P_{FA}$ by about 0.02 (2%). Again, the proposed method does not outperform the multiple-feature method in terms of $P_D$ but its $P_{FA}$ is about ten times better than the multiple-feature method.

Using VAD and checking different criteria improve $P_{FA}$ considerably while degrade the detection time. Therefore, one should select between the proposed method and multiple-feature method proposed in [3] based on the priority among lower $P_{FA}$, computational complexity, and shorter detection time.

## 7. Conclusion

This study introduced a howling detection algorithm which applied the decision of VAD algorithm to activate/deactivate the howling detection part. Using VAD as an activator significantly reduced the computational complexity. The proposed method also checked different properties of the howling component in a novel order to decrease false alarm probability. Applying the VAD algorithm and checking different criteria improved $P_{FA}$ considerably while degraded the detection time to some extent. The main advantage of the proposed method was its low computational complexity in terms of number of required multiplications, additions, and comparisons.

## Acknowledgments

This research work has been partially supported by Electrical Engineering department of the University of Texas at Dallas.

## References


[1] S.A. Khoubrouy, Issa M.S Panahi, Criteria for estimating an FIR filter for cancelling the feedback path signal in hearing aid system, EURASIP Eur. J. Signal Process. 100 (2014) 101–111.

[2] Soudeh. A. Khoubrouy, Issa M.S. Panahi, John H.L. Hansen, Howling detection in hearing aids based on generalized Teager-Kaiser operator, IEEE Trans. Audio Speech Lang. Process. 23 (2014) 154–161.

[3] T.V. Waterschoot, M. Moonen, Comparative evaluation of howling detection criteria in notch-filter-based howling suppression, J. Audio Eng. Soc. 58 (2010) 923–940.

[4] A. Pandey, V.J. Mathews, Improving adaptive feedback cancellation in digital hearing aids through offending frequency suppression, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Dallas, 2010, pp. 173–176.

[5] S.A. Khoubrouy, I. Panahi, N. Kehtarnavaz, Howling detection in hearing aids using discrete energy separation algorithm-2 and generalized Teager-Kaiser operator, In: Proceedings of the IEEE Conference of Acoustic, Speech, and Signal Processing (ICASSP), May 2014.

[6] Victor Poblete, Néstor Becerra Yoma, Richard Stern, Optimization of the parameters characterizing sigmoidal rate-level functions based on acoustic features, Speech Commun. 56 (2014) 19–34.

[7] T.V. Waterschoot, M. Moonen, Fifty years of acoustic feedback control: state of the art and future challenges, Proc. IEEE (2011) 288–327.

[8] P.C. Loizou, Speech Enhancement Theory and Practice, Taylor and Francis Group, New York, 2007 55–66 and 399–401.

[9] D. Kewley-Port, T.Z. Burkle, J.H. Lee, Contribution of consonant versus vowel information to sentece intelligibility for young normal-hearing and elderly hearing-impaired listeners, J. Acoust. Soc. Am. 122 (4) (2007) 2365–2375.

[10] R.V. Prasad, A. Sangwan,H.S. Jamadagni, M.C. Chiranth, R. Sah, V. Gaurav, Comparison of voice activity detection algorithms for VoIP, In: Proceedings of the 7th International Symposium on Computers and Communications, pp. 530–535, 2002.

[11] J. Sohn, N.S. Kim, W. Sung, A statistical model-based voice activity detection, IEEE Signal Process. Lett. 6 (1) (1999) 1–3.

[12] J.W. Shin, H.J. Kwon, S.H. Jin, N.S. Kim, Voice activity detection based on conditional MAP criterion, IEEE Signal Process. Lett. 15 (2008) 257–260.

[13] Y. Hu, P. Loizou, Subjective evaluation and comparison of speech enhancement algorithms, Speech Commun. 49 (2007) 588–601.