Howling Detection and Gain Control for Speech Reinforcement in a Noisy Car Cabin Environment

Yehav Alkaher and Israel Cohen, *Fellow, IEEE*

***Abstract*—In-car speech communication is particularly challenging due to environmental noise. The speaker's microphone also acquires car and road noises, resulting in a low signal-to-noise ratio and persistent frequency-howls that do not decrease, which degrade the system's output sound quality. In this paper, we address the problem of howling control for in-room speech reinforcement systems under high environmental noise levels, specifically in a car cabin. We control the acoustic feedback via a gain control algorithm based on howling detection. Two detection methods are proposed to detect non-decreasing underdamped frequency-howls. A single-channel optimal Feedback Wiener gain is derived to enhance the desired near-end speech signal, followed by another Wiener filter that is based on the signal magnitude relative to the estimated noise power spectral density. Adjusting the howling energy threshold is proposed to deal with false-detection artifacts arising as the environment's environmental noise level rises. The performance improvements of the howling-detection-based gain control algorithm following the proposed adjustments are evaluated in clean and noisy environments. As detection of non-decreasing underdamped frequency-howls is feasible, it was found that it becomes unnecessary when the environmental noise is successfully reduced via the Wiener filter.**

***Index Terms*—Speech reinforcement, acoustic feedback, howling control, noise reduction, in-car communication.**

I. Introduction

In-car speech communication is particularly challenging due to the environmental noise during a ride [1], [2]. A speech reinforcement (SR) system is commonly used to amplify and play the desired speech back into the cabin. However, due to the relatively small room space of a car cabin, the loudspeaker's delivered speech signal must be of high quality, and the processing time of the SR system needs to be short [1], [3]. In practice, the primary challenges in delivering high-quality speech are acoustic feedback control and environmental noise reduction [4], [5], [6]. As modern car cabins increasingly provide microphone-based applications other than SR, such as hands-free telephony and automatic speech recognition, handling these challenging effects is crucial, as discussed by Müller et al. in [7].

Although the reverberation time in a car cabin is very short (50–60 ms) [2], [3], [8], as the system's amplification gain rises, the small room space makes an employed SR system vulnerable to acoustic feedback [9], [10]. As a result, grating howling noises are emitted at resonance frequencies of the system's closed-loop transfer function (TF) [11]. During silence, howling appears only if the speech harmonics excite the system's howling frequencies. However, high environmental noise may trigger howling regardless of the speech activity. Apart from howling, the acquired environmental noise degrades speech communication and should be filtered out from the reinforced sound.

Many of the recent acoustic feedback control efforts in the literature, mainly for hearing aid devices and public address systems, are attributed to adaptive feedback cancellation (AFC) algorithms [12], [13], [14], [15], [16], [17]. AFC methods use adaptive filters to estimate the loudspeaker-enclosure-microphone paths, utilizing techniques like the normalized least mean squares (LMS) and the Kalman filter, and subtract the estimated acoustic echo from the microphone signal, as in acoustic echo cancellation (AEC). A fundamental challenge of AFC within SR systems is that the loudspeaker signal is correlated with the near-end signal, which serves as a disturbance signal for the adaptive filter. Thus, a standard adaptive filtering algorithm produces a biased estimate and an impaired output speech signal. A common practice is to employ prediction-error-method (PEM)-based decorrelation techniques [10], [12], [15]. As the main challenge of AFC is reducing the computational complexity [18], those works deal with howling by providing low-complexity AFC solutions with fast recovery after changes in the loudspeaker-enclosure-microphone paths. Unfortunately, these works do not concern howling control in noisy environments. Apart from AFC, the use of neural-network-based methods for real-time howling detection and suppression has begun to rise, also aiming to deal with nonlinearities and uncertainties such as in loudspeaker/microphone responses [19], [20].

In our recent work [11], [21], we proposed a dual-microphone SR system with howling control for in-car speech communication, focusing on front-to-rear passenger communication. The dual-microphone SR system includes a speaker's speech acquisition microphone and another intelligently-located microphone that monitors the environment for howling detection. Once a potential frequency-howl is detected [11], the magnitude-slope-deviation (MSD)-based gain-control mechanism from [21] adjusts the amplification gain of the SR system to control the

Manuscript received 8 July 2023; revised 17 January 2024; accepted 1 February 2024. Date of publication 8 February 2024; date of current version 22 February 2024. This work was supported in part by Israel Science Foundation under Grant 1449/23 and in part by Pazy Research Foundation. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Romain Serizel. *(Corresponding author: Yehav Alkaher.)*

The authors are with the Andrew and Erna Viterbi Faculty of Electrical and Computer Engineering, Technion – Israel Institute of Technology, Haifa 3200003, Israel (e-mail: yehava@campus.technion.ac.il; icohen@ee.technion.ac.il).

Digital Object Identifier 10.1109/TASLP.2024.3364091

acoustic feedback and suppress the potential frequency-howls in the reinforced speech signal.

Another well-known approach to dealing with howling in SR systems is employing notch-filter-based howling suppression (NHS) techniques [18], [22], [23], [24]. Hence, the foundation for such howling control mechanisms is the howling detection algorithm, which in practice analyzes the behavior of the signal's frequency components at a given period of time. For example, the MSD-based temporal howling detector we proposed in [11] flags frequency components that increase/decrease exponentially, i.e., change linearly on a dB scale, due to the acoustic feedback [22]. Specifically, these frequency components are termed increasing/underdamped frequency-howls, respectively. Unfortunately, underdamped frequency-howls are not detected when they do not decrease due to environmental noise. Although the amplification gain is adjusted occasionally, the signal model assumes a fixed amplification gain at a given time interval. Therefore, this approach is particularly relevant to "piecewise" time-invariant SR systems. A significant challenge is to suppress frequency-howls before the human ear notices. While howling detection algorithms analyze the behavior of the signal's frequency components, noise reduction filters affect the components' magnitudes over time.

In this paper, we address the problem of howling control in in-room SR systems under high environmental noise levels. Two howling detectors are proposed based on the magnitude's standard deviation and the NINOS^2-T measure. A Feedback Wiener gain is derived to enhance the desired near-end speech signal, followed by another Wiener filter that is based on the signal's magnitude relative to the noise's estimated power spectral density (PSD) and an over-subtraction factor [25]. Then, as low-energy false-alarm-causing artifacts decrease the performance of the temporal howling detector, the howling energy threshold is adjusted based on the measured noise floor. In this manner, the proposed adjustments are added to the SR algorithm described in [21]. For simplicity, in the scope of this paper, the same microphone is used for speech acquisition and monitoring the environment. Thus, the howling-detection-based gain control algorithm's performance improvement following the proposed adjustments is evaluated under clean and noisy environments. Our main contributions are: 1) Analysis of the closed-loop SR system's response to high environmental noise, and formulation of the non-decreasing underdamped frequency-howl. 2) Two detection methods for non-decreasing underdamped frequency-howls. 3) Adjustment of the howling energy threshold to deal with acquired false-alarm-causing artifacts arising from high environmental noise. 4) Analysis of the effective amplification gain in the SR system and the (dis)appearance of the non-decreasing underdamped frequency-howls following the Wiener filter.

This paper is organized as follows: Section II describes the signal model and problem formulation. Section III provides a mathematical analysis of the origins of the howling effect in the presence of environmental noise. Section IV analyzes the magnitude behavior of a frequency-howl induced by high environmental noise, and introduces two detection methods complementary to the temporal howling detector in [11]. Moreover, it addresses the temporal howling detector's sensitivity under high environmental noise. Section V considers single-channel signal enhancement in the frequency domain and addresses associated STFT domain challenges. Section VI demonstrates the advantages of the proposed howling detection methods and the effect of the Wiener filter on the system's effective amplification gain. Then, it demonstrates the SR improvement of the proposed adjustments to the SR algorithm. Finally, Section VII presents the conclusions of the study.

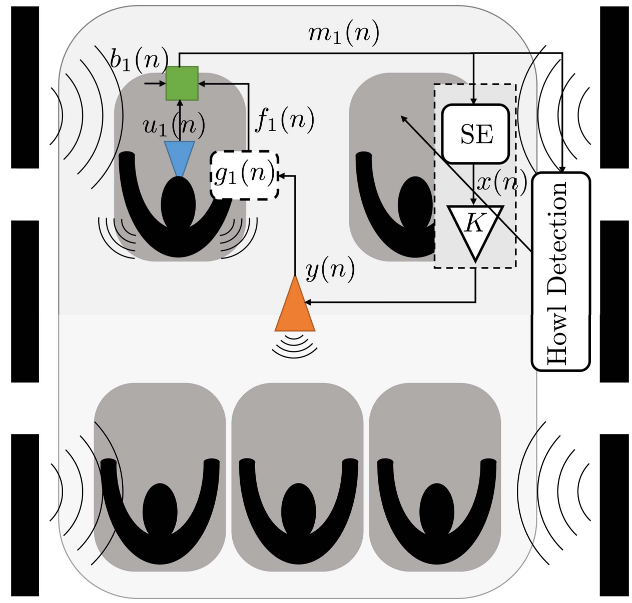


Fig. 1. In-car single-microphone speech reinforcement system with howling-control.

II. Signal Model and Problem Formulation

The in-car SR system model considers a microphone and a loudspeaker in a closed room [21], as illustrated in Fig. 1. The blue triangle is the speaker, the driver in this case, providing the near-end speech to the microphone (green square), denoted as mic1. As mentioned above, mic1 acquires the speech signal to be reinforced and monitors the environment for howling detection. Thus, controlling the amplification gain and the speech enhancement (SE) segment responsible for enhancing the near-end speech. The large orange triangle is the loudspeaker, playing the reinforced speech from mic1 back into the car cabin.

The microphone input signal is composed of the desired near-end speech $u_1(n)$, the environmental noises $b_1(n)$, and the acoustic echo from the loudspeaker $f_1(n)$:

$$m_1(n) = u_1(n) + b_1(n) + f_1(n)\,. \tag{1}$$

The filtered estimated near-end speech $x(n)$ is obtained from $m_1(n)$ using the SE segment to deliver the near-end speech back into the cabin. Then, $x(n)$ is amplified by a factor K, creating the output loudspeaker signal:

$$y(n) = K\,x(n)\,. \tag{2}$$

Note that the microphone and loudspeaker possess internal thermal noises, yet negligible relative to the in-car environmental noises during a ride. The emitted output signal $y(n)$ propagates through the cabin into the speaker's microphone (and the howling-detection microphone) with an RIR $g_1(n)$, generating the echo signal $f_1(n)$, i.e.,

$$f_1(n) = y(n) * g_1(n)\,, \tag{3}$$

where $*$ denotes the convolution operation. This closed-loop behavior is the source of acoustic feedback within the SR system.

Our objective is to provide an SR system that maintains a desired high amplification gain while filtering out the environmental noise and suppressing potential sound artifacts due to acoustic feedback, see Fig. 1. Considering the SE segment that provides the filtered estimated near-end speech $x(n)$ (to be amplified), the near-end speech signal $u_1(n)$ needs to be recovered from the noisy observation $m_1(n)$ for clean speech reinforcement, i.e.,

$$x(n) = m_1(n) * h_{\text{SE}}(n) \approx u_1(n)\,; \tag{4}$$

where $h_{\text{SE}}(n)$ is a causal FIR filter to be implemented in the SE segment. While the reinforced speech is based on the SE segment, the input to the howling detection algorithm is $m_1(n)$. However, it is directly affected by $h_{\text{SE}}(n)$, as explained in the following sections.

III. Response to a Stationary Input Signal

Considering a Linear Time-Invariant (LTI) system with a discrete-time (room) impulse response $h[n]$, it is desired to examine the system response to the input signal $x[n]$. The output of the system, $y[n]$, is thus given by

$$y[n] = x[n] * h[n] = \sum_{k=-\infty}^{\infty} x[n-k]\, h[k]\,. \tag{5}$$

Let $x[n]$ be a wide-sense stationary (WSS) random process, e.g., a white Gaussian noise $x[n] \sim \mathcal{N}(\mu_x, \sigma_x^2)$. Then,

$$\begin{aligned} \mathbb{E}\left[x[n]\right] &= \mu_x\,; \\ R_X[\eta] &= E\left[x^*[n]\, x(n+\eta)\right] \overset{\text{white noise}}{=} \sigma_x^2\, \delta[\eta]\,; \end{aligned} \tag{6}$$

where $E[\cdot]$ denotes mathematical expectation, the superscript * is the complex-conjugate operator, and $R_X[\eta]$ is the autocorrelation function of the time lag η. Therefore,

$$\begin{aligned} \mathbb{E}\left[y[n]\right] &= \sum_{k=-\infty}^{\infty} \mathbb{E}\left[x(n-k)\right]\, h[k] = \mu_x \sum_{k=-\infty}^{\infty} h[k]\,; \\ R_Y[\eta] &= E\left[y^*[n]\, y[n+\eta]\right] \\ &= \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \mathbb{E}\left[x^*[n-k]\, x[n+\eta-l]\right]\, h^*[k]\, h[l] \\ &= \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} R_X[\eta+k-l]\, h^*[k]\, h[l]\,. \end{aligned} \tag{7}$$

Taking the discrete-time Fourier transform (DTFT) of both sides results in the power spectral density (PSD) of $y[n]$, that is,

$$S_Y(\theta) = |H(\theta)|^2\, S_X(\theta) \overset{\text{white noise}}{=} |H(\theta)|^2\, \sigma_x^2\,. \tag{8}$$

Accordingly, the energy of the output signal at frequency θ is proportional to the input signal's energy. Hence, as long as the energy of the input at frequency θ does not decrease, the output signal will correspondingly have a proportional constant energy that does not decrease.

This non-decreasing energy, which may be caused when the environmental noise level is too high, is a particular case of the underdamped frequency-howl and deserves special treatment. Thus, it is termed a non-decreasing underdamped frequency-howl. This type of howling is added to the increasing and underdamped frequency-howls, noted in [11].

IV. Detection of Non-Decreasing Underdamped Frequency-Howls

In an in-room SR system, where feedback is present, the feedback effect on the reinforced speech is determined by the dominant poles of the closed-loop system TF [26]. In [11], the response to an energy burst in a frequency bin, i.e., a sinusoidal windowed input signal, was mathematically analyzed and resulted in distinguishing between increasing and underdamped frequency-howls. With the MSD measure and the temporal approach, the temporal howling detector was designed to immediately detect these frequency-howls, characterized by an energy increase or decrease, respectively, before the human ear notices. For this purpose, the temporal howling detector includes two cascaded stages: Soft Howling Detection and Howling False-Alarm Detection. On the other hand, a research gap arises when an underdamped frequency-howl does not decrease, due to a high level of acquired environmental noise, as shown in Section III. This section proposes two methods to detect non-decreasing underdamped frequency-howls that are not detected earlier by the temporal howling detector proposed in [11]. In each of the following methods, as in [11], the PSD is calculated on subsequent sample frames and inserted into a magnitude history buffer to analyze each frequency component's temporal magnitude behavior. For the sample rate of 16 kHz, the length of the magnitude history buffer is $N = 12$ frames, where the frame length is 512 samples (32 ms — a typical speech analysis frame length), and the frame shift between subsequent sample frames corresponds to 10 ms. In comparison, the length of the temporal howling detector's magnitude history buffer is $N = 6$ frames for the Soft Howling Detection stage (and $N = 120$ frames for the Howling False-Alarm Detection stage). Thus, providing a more accurate assessment of the magnitude's continuity. Additionally, this section also proposes an adjustment to the howling energy threshold with respect to the temporal howling detector, to improve its performance under high environmental noise levels.

A. NINOS^2-T Measure

According to Mounir [27], howling is manifested as a persisting narrowband signal component, and can be observed as a horizontal line in the spectrogram. This corresponds to the phenomena of a non-decreasing underdamped frequency-howl. For this purpose, the NINOS^2-Transposed (NINOS^2-T) was proposed by Mounir [27] to measure spectral non-sparsity in the magnitude history of a specific frequency bin. The NINOS^2-T measure is based on the NINOS^2 feature, proposed by Mounir et al. [28] for the problem of musical note onset detection. The term NINOS^2 stands for Normalized INOS^2, where INOS^2 stands for Identifying Note Onsets based on Spectral Sparsity. Thus, the NINOS^2-T measure at a suspected frequency bin k is defined as

$$\text{NINOS}^2\text{-T}(k,m) = \frac{1}{\sqrt[4]{N}-1}\left(\frac{\|\mathbf{G}(k,m)\|_2}{\|\mathbf{G}(k,m)\|_4} - 1\right), \quad (9)$$

where m denotes the index of the last analyzed frame, the row vector $\mathbf{G}(k,m)$ is the magnitude history buffer at frequency bin k, and N is the number of frames in the magnitude history buffer. As NINOS^2-T measures non-sparsity, at the least sparse case, i.e., when the magnitude in $\mathbf{G}(k,m)$ is constant over time, NINOS^2-T$(k,m) = 1$.

B. Magnitude's Standard Deviation Measure

The environmental noise's energy level (PSD) is assumed to remain constant over time. Therefore, the standard deviation (STD) can be utilized to assess the variance of the magnitude along the history buffer. Accordingly, the magnitude's STD at frequency bin k is calculated as

$$\text{STD}(k,m) = \sqrt{\frac{1}{N-1}\sum_{l=0}^{N-1}|\mathbf{G}(k,m)[l] - \boldsymbol{\mu}(k,m)|^2}, \quad (10)$$

where the mean $\boldsymbol{\mu}(k,m)$ is calculated as

$$\boldsymbol{\mu}(k,m) = \frac{1}{N}\sum_{l=0}^{N-1}\mathbf{G}(k,m)[l], \quad (11)$$

and l is the frame index within the magnitude history buffer.

C. Total Spectrum Noise Floor

When a frequency bin's magnitude settles to a steady energy level (magnitude noise floor) after an energy burst, the non-decreasing energy shall be noticeable only if above the total spectrum noise floor. Otherwise, it is masked by energy in other frequencies and can not be noticed by the human ear. As the two measures above detect frequency bins with steady magnitudes over time, it is possible to determine whether a candidate steady energy level manifests as a frequency-howl by calculating the total spectrum noise floor. Thus, the total spectrum 90% noise floor is calculated as the 90th percentile of all magnitudes in the magnitude history buffer $\mathbf{G}(k,m)$. This measure is essentially a modification of the peak-to-average power ratio (PAPR) feature [23], resulting in the Magnitude-STD-based and NINOS^2-T-based detectors, complementary to the temporal howling detector.

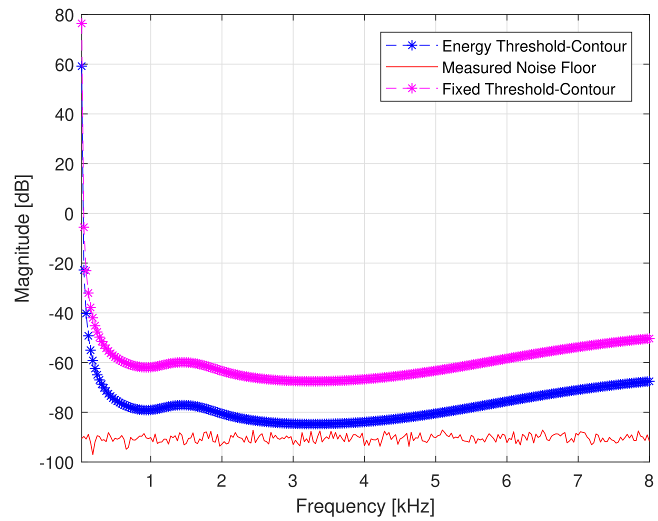


Fig. 2. Howling energy threshold-contour fix, based on the measured input noise floor.

D. Howling Energy Threshold-Contour

Regarding the temporal howling detector in [11], to determine whether a candidate frequency-howl is audible to the human ear and should be flagged, a howling energy threshold-contour is established across the frequency bins based on the equal-loudness contours defined in standard ISO 226:2003 [29]. In the presence of a high noise level, low-energy frequency-howls might be masked by the added noise floor, and therefore can be overlooked by the detector without being noticed by the human ear. Furthermore, as a noise floor is added, low-energy false-alarm-causing artifacts might be "raised" above the energy threshold-contour. Correspondingly, the utilized howling energy threshold-contour is raised uniformly by 20 dB above the calculated PSD of the input noise, see Fig. 2.

V. Single-Channel Frequency Domain Signal Enhancement

The environmental noise reduction is performed using a signal enhancement filter [30]. Calculating the filter gains in the frequency domain allows the formulation of the feedback in the SR system in terms of frequency responses. Accordingly, the terms of mic1 and the loudspeaker can be expressed in the frequency domain as:

$$\begin{aligned} M_1(f) &= U_1(f) + B_1(f) + G_1(f)\,Y(f)\,; \\ Y(f) &= K\,X(f) = K\,H_{\text{SE}}(f)\,M_1(f)\,; \end{aligned} \quad (12)$$

which means that

$$M_1(f) = U_1(f) + B_1(f) + K\,G_1(f)\,H_{\text{SE}}(f)\,M_1(f)\,. \quad (13)$$

Therefore,

$$\begin{aligned} M_1(f) &= \frac{U_1(f) + B_1(f)}{1 - K\,H_{\text{SE}}(f)\,G_1(f)}\,; \\ X(f) &= H_{\text{SE}}(f)\,M_1(f) = \frac{H_{\text{SE}}(f)\,(U_1(f) + B_1(f))}{1 - K\,H_{\text{SE}}(f)\,G_1(f)}\,. \end{aligned} \quad (14)$$

The desired signal $U_1(f)$ is uncorrelated with the environmental noise $B_1(f)$, and needs to be recovered from the noisy observation $M_1(f)$. Hence, the variance of $X(f)$, i.e., its PSD, can then be written as

$$\phi_X(f) = E\left[|X(f)|^2\right] = \frac{|H_{\text{SE}}(f)|^2\,(\phi_{U_1}(f) + \phi_{B_1}(f))}{|1 - K\,H_{\text{SE}}(f)\,G_1(f)|^2}\,; \tag{15}$$

and the cross-correlation between $X(f)$ and $U_1(f)$ is

$$\phi_{X,U_1}(f) = E\left[X(f)\,U_1^*(f)\right] = \frac{H_{\text{SE}}(f)\,\phi_{U_1}(f)}{1 - K\,H_{\text{SE}}(f)\,G_1(f)}\,. \tag{16}$$

A. Single-Channel Optimal Feedback Wiener Gain

The error signal between the estimated and desired signals at frequency f is defined as $\mathcal{E}(f) = X(f) - U_1(f)$. The narrowband MSE criterion is thus

$$\begin{aligned} J\left[H_{\text{SE}}(f)\right] &= E\left[|\mathcal{E}(f)|^2\right] \\ &= \phi_X(f) + \phi_{U_1}(f) - \phi_{X,U_1}(f) - \phi_{U_1,X}(f)\,. \end{aligned} \tag{17}$$

Taking the gradient of $J[H_{\text{SE}}(f)]$ with respect to $H^*_{\text{SE}}(f)$, and equating the result to 0 leads to the optimal Feedback Wiener gain:

$$H_{\text{FW}}(f) = \frac{\phi_{U_1}(f)}{(1 + K\,G_1(f))\;\phi_{U_1}(f) + \phi_{B_1}(f)}\,. \tag{18}$$

Defining the narrowband input signal-to-noise ratio as $\text{iSNR}(f) \triangleq \frac{\phi_{U_1}(f)}{\phi_{B_1}(f)}$, means that

$$H_{\text{FW}}(f) = \frac{\text{iSNR}(f)}{(1 + K\,G_1(f))\;\text{iSNR}(f) + 1}\,, \tag{19}$$

where setting $K = 0$ leads to the known Wiener gain.

Substituting $H_{\text{FW}}(f)$ in $\phi_{M_1}(f)$ results in

$$\begin{aligned} \phi_{M_1}(f) &= \frac{\phi_{U_1}(f) + \phi_{B_1}(f)}{|1 - K\,H_{\text{FW}}(f)\,G_1(f)|^2} \\ &= \frac{\phi_{U_1}(f) + \phi_{B_1}(f)}{\left|\frac{(1+K\,G_1(f))\,\phi_{U_1}(f) + \phi_{B_1}(f) - K\,\phi_{U_1}(f)\,G_1(f)}{(1+K\,G_1(f))\,\phi_{U_1}(f)+\phi_{B_1}(f)}\right|^2} \\ &= \frac{|\,(1 + K\,G_1(f))\;\phi_{U_1}(f) + \phi_{B_1}(f)|^2}{\phi_{U_1}(f) + \phi_{B_1}(f)}\,. \end{aligned} \tag{20}$$

Accordingly, to calculate the Feedback Wiener gain, the variance of the in-car environmental noise $\phi_{B_1}(f)$ can be estimated directly from $\phi_{M_1}(f)$ during a ride if setting K to zero and keeping silence for a while [3], [8], [30]. Then, to obtain the variance of the near-end speech $\phi_{U_1}(f)$, one needs to solve the parabolic equation, derived from (20), and choose one of the roots. Furthermore, besides the known applied amplification gain K, the RIR $G_1(f)$ generally needs to be estimated, e.g., see work by Avargel and Cohen [31], [32], [33], [34].

B. Implementation Considerations of the Noise Reduction Filter

In practice, the optimal filter gain above is calculated in the short-time Fourier transform (STFT) domain, according to [30, Ch. 3], using the multiplicative-transfer-function (MTF) approximation as discussed by Avargel and Cohen [32]. This approximation is only valid when the RIR is short relative to the speech analysis window ($\sim$ 30 ms [35]). Therefore, this model might be invalid above a certain amplification gain since the RIR is "infinite" when the acoustic feedback is high. Hence, it is chosen to set $K = 0$ in the model and utilize the known Wiener gain ($H_{\text{W}}(f) \triangleq H_{\text{FW},K=0}(f)$) within the SR system, due to its simplicity and the fact that the acoustic feedback artifacts should be controlled. Accordingly,

$$\phi_{M_1}(f) \overset{K=0}{\approx} \phi_{U_1}(f) + \phi_{B_1}(f)\,. \tag{21}$$

In the STFT domain, the Wiener gain is thus calculated as

$$H_{\text{W}}(r,k) = \frac{\text{iSNR}(r,k)}{\text{iSNR}(r,k) + 1}\,, \tag{22}$$

where r denotes the frame index and k denotes the frequency bin. Then, for online filtering, an inverse fast Fourier transform (IFFT) is applied to the resulting short-term filter gains.

To obtain a less noisy estimate of $\phi_{U_1}(r,k)$ for designing the Wiener filter gains in (22), it may be desired to use the measured $\phi_{B_1}(k)$ multiplied by an over-subtraction factor $\beta \geq 1$ (as proposed by Berouti et al. [36] and discussed by Pardede et al. [25]), i.e.,

$$\hat{\phi}_{U_1}(r,k) \approx \max\left\{0\,,\ \phi_{M_1}(r,k) - \beta\,\phi_{B_1}(k)\right\}. \tag{23}$$

The over-subtraction factor, which is determined heuristically, is utilized to minimize the gap between the predicted and the actual variances of the environmental noise. When there is no acoustic feedback in the system ($K = 0$) and the variance of the in-car environmental noise $\phi_{B_1}(k)$ is correctly measured, $\beta = 1$ would provide $\hat{\phi}_{U_1}(r,k) = \phi_{U_1}(r,k)$. However, when acoustic feedback is present in the SR system, β must be estimated appropriately.

VI. Results

This section evaluates the SR system with the howling-detection-based gain control algorithm. Section VI-A demonstrates the proposed howling detection methods upon a magnitude analysis over time, under a devised feedback scenario and different iSNRs. Then, Section VI-B demonstrates the effect of the Wiener filter's over-subtraction factor on the effective amplification gain. The RIRs used in simulations are based on the simulated room configuration in [21]. After that, Section VI-C first justifies the adjustment process of the howling energy threshold-contour. Subsequently, it evaluates the overall howling-detection-based gain control algorithm under clean and noisy environments, incorporating the discussed howling detectors.

A. Detection of Non-Decreasing Underdamped Frequency-Howls

To examine the signal magnitude behavior of frequency-howls in a noisy reverberant environment under acoustic feedback, a two-pole TF is applied to a devised input signal under different iSNRs. In [11], the devised input signal comprised a 1.5 s of speech preamble, an energy burst, and silence for analyzing the response. A modified input signal is proposed, which consists of the previously devised signal, using speech samples from the TIMIT speech database [37], with an additive Gaussian noise at chosen iSNR values of 40 dB (Clean) and 10 dB (Noisy). The modified input signal is then followed by 0.5 s of silence, 0.5 s of a sine wave at the poles' frequency, and another replication of the modified input signal. Also, a whistle recording at 1218.25 Hz is added to the signal to review the detectors' performance on a possible false frequency-howl.

To evaluate the presented howling detection methods, the "Recovery Gain-Control Howl" feedback scenario from [11] is applied to the input signal, simulating a gain reduction (as if howling is noticed and thus suppressed) followed by a positive gain change, to a more stable amplification gain. Accordingly, Fig. 3 shows the spectrograms of the input signal and the response signal, with the configured signal-magnitude change rate over time starting from -50 dB/sec, decreasing to -3000 dB/sec (stable poles, low acoustic feedback), and fixing to -150 dB/sec (see (17) in [11]). The howling detections are marked by squares where those around the pole's frequency are marked in red, and those around the whistle are in blue. The rest are marked in gray. Furthermore, the figure shows the signal's magnitude at specific frequency bins over time. On each graph, the hearing threshold of the corresponding frequency bin is depicted by a red dashed line; a magenta dashed line shows the total spectrum 90% noise floor; the detections of the Soft Howling Detection stage of the temporal howling detector are marked by blue circles, and red circles mark the detections that haven't been refuted by the Howling False-Alarm Detection stage. Moreover, the two proposed detection methods of non-decreasing underdamped frequency-howls — the Magnitude-STD-based and the NINOS^2-T-based, are marked by magenta circles and cyan circles, respectively. The fine-tuned thresholds for howling detection are $\mathrm{STD}(k,m) < 3$ and NINOS^2-T$(k,m) > 0.95$.

Following the structure of the devised input signal and the applied feedback scenario, it is interesting to observe the magnitude at the poles' frequency. From 0 to 1 s, for 40 dB iSNR, energy is added mainly due to speech harmonics, and the frequency-howl decays slowly. In contrast, for 10 dB iSNR, constant energy is inserted into the two-pole TF, and the signal maintains a steady magnitude, i.e., a non-decreasing underdamped frequency-howl. In both cases, the frequency-howl is detected by both examined detectors.

From 1.5 to 3 s, both cases exhibit an underdamped frequency-howl, where the frequency bin's magnitude noise floor is higher for 10 dB iSNR, which leads to a shorter decrease period. Then, a non-decreasing underdamped frequency-howl is evident in both cases, although only detected for 10 dB iSNR since the frequency-howl is higher than the total spectrum 90% noise floor. Namely, while the frequency bin's magnitude settles to a steady energy level (noise floor) in both cases, the non-decreasing energy is higher than the total spectrum noise floor for 10 dB iSNR, making it noticeable. In comparison, for 40 dB iSNR, the steady energy level of the frequency bin is unnoticeable due to a masking effect.

From 3 to 3.5 s, the response to the silence part is seen more clearly when it follows the 10 dB iSNR signal. The temporal howling detector detects the underdamped magnitude decrease until it reaches a lower magnitude noise floor. From 3.5 to 4 s, the sine wave is steady in magnitude and detected by all detection methods. From 4 to 7 s, an underdamped response is visual for 40 dB iSNR and detected by the temporal howling detector. For 10 dB iSNR, on the other hand, a non-decreasing underdamped response is visual and detected by the two proposed detection methods. Hence, the temporal howling detector and the two proposed detection methods are complementary in the underdamped scenario of a two-pole TF.

Observing the signal magnitude behavior at a frequency bin around the frequency components of the whistle, while the signal is not affected by the two-pole TF, its magnitude behavior may be perceived (by the two proposed detection methods) as a non-decreasing underdamped frequency-howl. Observing the magnitude of the low speech harmonic, continuing energy bursts are evident, higher than the total spectrum 90% noise floor. Therefore, as a precaution, detected non-decreasing underdamped frequency-howls below 1 kHz are disregarded (as in the temporal howling detector [11]).

B. Wiener Filter's Effect on Signal Behavior

As part of the noise reduction process, it is desired to obtain an adequate estimation of the variance of the near-end speech $\phi_{U_1}(r,k)$, e.g., based on a noise-only period. When setting a desired amplification gain in the SR system, the measured PSD of the noise might be affected by the feedback (and the MTF approximation in the Wiener filter). As mentioned in Section V, the variance of the in-car environmental noise $\phi_{B_1}(k)$ can be initially measured during a ride if setting the amplification gain to zero. Thus, providing the means to apply a Wiener filter under different amplification gains of choice. Fig. 4 shows the effect of using different over-subtraction factors (to compensate for the feedback effect) on the effective amplification gain in the SR system under a configured amplification gain $K = 7$ and iSNR of 10 dB. In practice, the Wiener filter reduces the effective amplification gain per frequency component over time, based on the signal's magnitude relative to the estimated PSD of the noise and the over-subtraction factor β. For $\beta = 0.001$, the Wiener filter stops being effective as the energy increases due to the acoustic feedback. However, for $\beta = 3$, the Wiener filter sufficiently reduces the environmental noise, up to the presence of frequency-howls. Accordingly, in this simulated scenario, the fine-tuned over-subtraction factor for the Wiener filter gains is $\beta = 3$.

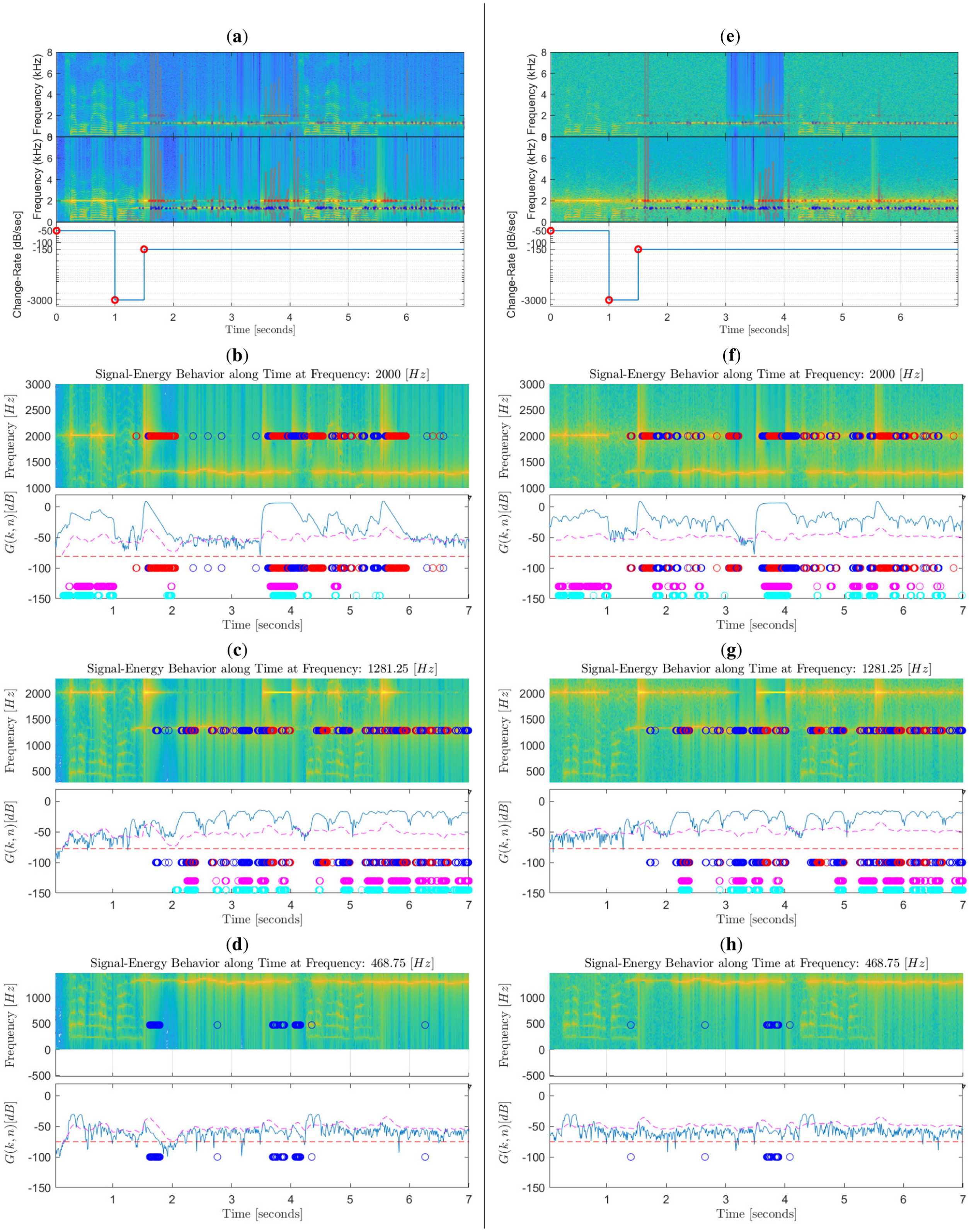


Fig. 3. Comparison of howling detection methods based on the response to a "Recovery Gain-Control Howl" feedback scenario, under iSNR values of 40 dB (Clean) and 10 dB (Noisy). The examined two-pole TF has complex conjugate poles at frequency 2000 Hz, with a magnitude that corresponds to configured signal-magnitude change rates of −50 dB/sec, −3000 dB/sec (neutral two-pole TF), and −150 dB/sec. The left column (**a–d**) refers to a clean input signal (40 dB iSNR), and the right column (**e–h**) refers to a noisy input signal (10 dB iSNR). The top figures (**a,e**) include a spectrogram comparison between the input signal and the response, and the applied pole's magnitude (change rate) over time. The figures underneath show the behavior of the signal's magnitude over time at specific frequency bins: (**b,f**) refer to 2000 Hz (the pole's frequency); (**c,g**) refer to 1281.25 Hz (around the whistle's frequency); (**d,h**) refer to 468.75 Hz (low speech harmonic's frequency).

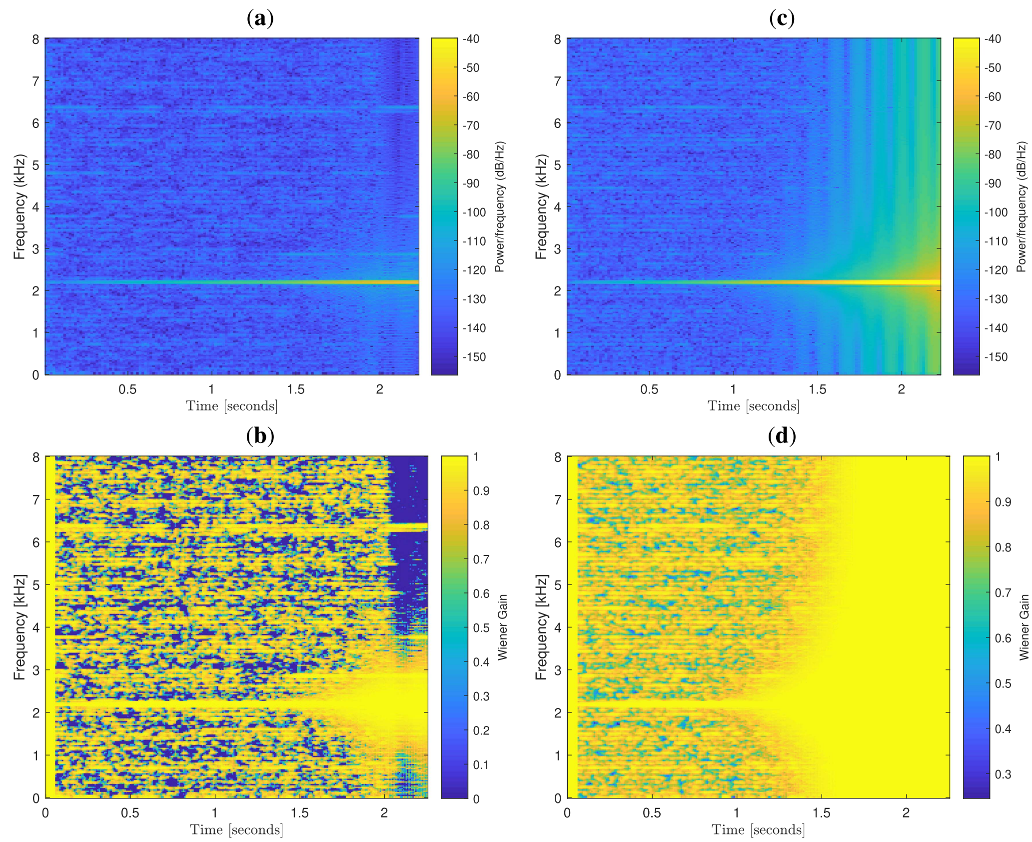


Fig. 4. Spectrogram of the loudspeaker signal (top figures) and the applied Wiener gain over time (bottom figures), under the scenario of $K = 7$ and 10 dB iSNR, using the over-subtraction factors $\beta = 3$ and $\beta = 0.001$. The input signal is a low-power white Gaussian noise, and no gain control algorithm is applied. (**a,b**) relate to $\beta = 3$; and (**c,d**) relate to $\beta = 0.001$.

C. Howling-Detection-Based Gain Control Algorithm Under a Noisy Environment

It is desired to evaluate the performance of the overall howling-detection-based gain control algorithm, together with the speech enhancement process in the SE segment, using the speaker's microphone for both speech acquisition and howling detection.

1) Effect of the Howling Energy Threshold-Contour: Fig. 5 demonstrates the performance of the temporal-howling-detector-based gain control algorithm, under a configured amplification gain $K = 7$ and iSNR of 10 dB, with and without fixing the howling energy threshold-contour. As seen in Fig. 5(a), (b), the howling detection process starts after 1.222 s, that is, once the magnitude history buffer is filled with

$$\frac{L_{\text{frame}} + L_{\text{frame-shift}} \left(N_{\text{history-buffer}} - 1 \right)}{f_{\text{sampling}}}$$

samples [11]. Then, multiple howling detections are marked with no (clearly) seen frequency-howl artifacts, suggesting low energy. Moreover, none of the detected frequency-howls is detected by the temporal howling detector at the dominant howling frequency bin. On the other hand, detected frequency-howls can be seen in Fig. 5(c), (d), i.e., after fixing the threshold contour, as smears of frequency components (horizontal lines) truncated due to gain reduction. In this case, the applied amplification gain does not degenerate to zero but rather fluctuates around a stable (and satisfying) level. Hence, fixing the howling energy threshold-contour is necessary. Nevertheless, the quality degradation in Fig. 5(c) due to the environmental noise can be seen by the vast smearing of frequency components (howling artifacts) not detected by the temporal howling detector.

2) Howling Detection Methods: First, it is desired to evaluate the gain control algorithm via the temporal howling detector, together with the speech enhancement process in the SE segment and the howling energy threshold-contour adjustment process. Fig. 6 shows the spectrogram of the loudspeaker's output and the applied amplification gain over time under a configured amplification gain $K = 7$ and iSNR of 10 dB. The noise reduction in Fig. 6 is clearly visible compared with Fig. 5(c), although there is still some smearing of frequency components (frequency-howls). These residual howling artifacts emphasize the challenge of howling detection in reverberant and noisy environments. In addition, the reduction of the effective amplification gain through the Wiener filter results in no visible non-decreasing underdamped frequency-howls.

Following the approach in [21], the matched effective gain $K_{\text{eff}} \triangleq \arg\min_a \|y(n) - a\, u_1(n)\|^2$ is used to estimate the relative gain reduction $\text{RGR} \triangleq \frac{K - K_{\text{eff}}}{K}$ (aimed to be low). Also,

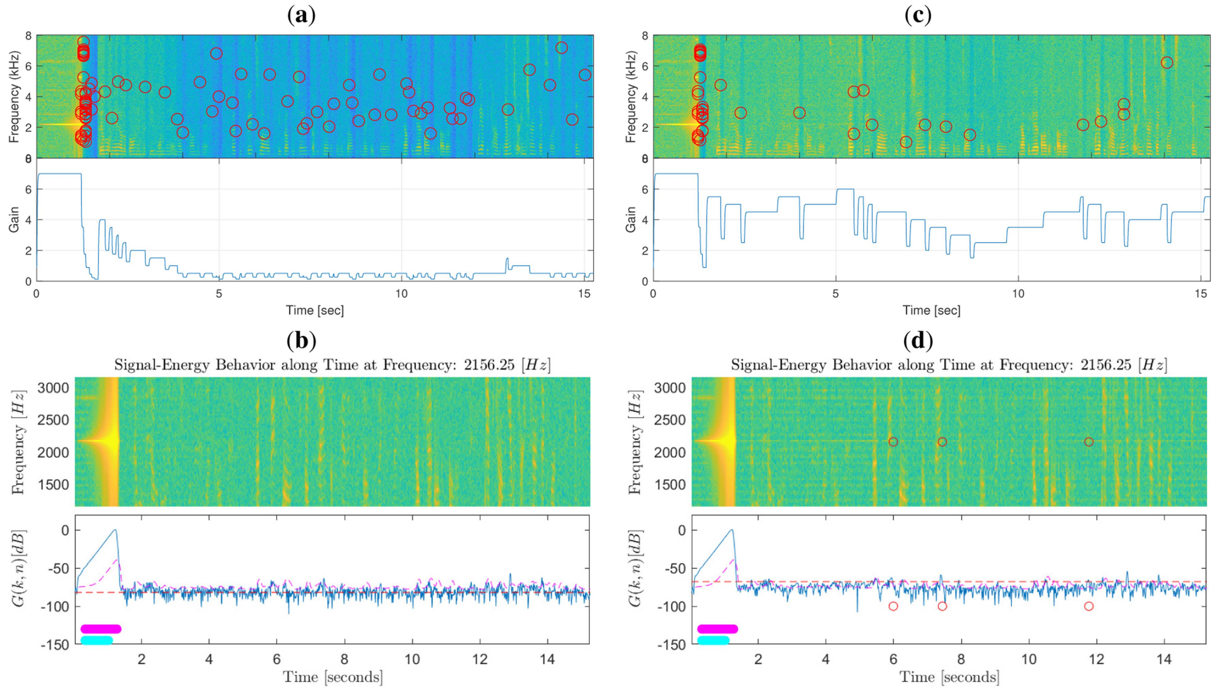


Fig. 5. Performance comparison of the temporal-howling-detector-based gain control algorithm, under a configured amplification gain $K = 7$ and 10 dB iSNR (noisy environment), with and without fixing the howling energy threshold-contour. The top figures include the spectrogram of the loudspeaker signal and the applied amplification gain over time. Red circles mark the detected frequency-howls. The figures underneath show the behavior of the microphone signal's magnitude over time at the frequency bin of the dominant pole of the closed-loop TF. (**a,b**) relate to using the original threshold-contour; and (**c,d**) relate to using the fixed threshold-contour.

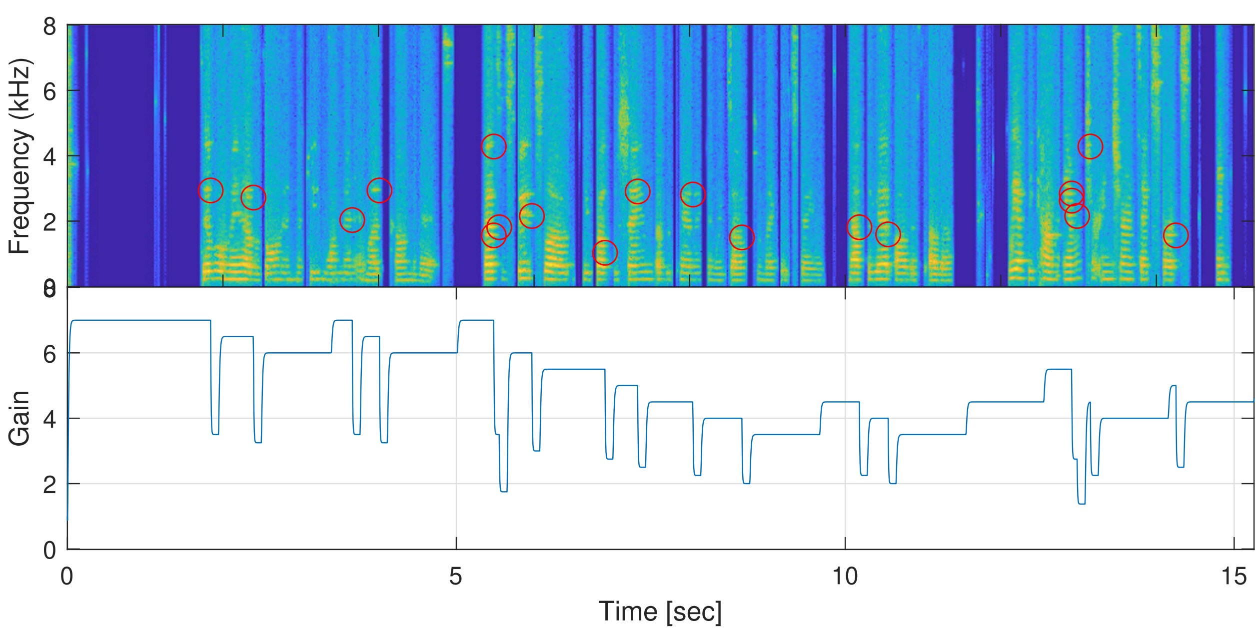


Fig. 6. Performance of the temporal-howling-detector-based gain control algorithm, under a configured amplification gain $K = 7$ and 10 dB iSNR (noisy environment), using a Wiener filter in the SE segment and fixing the howling energy threshold-contour. The figure includes the spectrogram of the loudspeaker signal and the applied amplification gain over time. Red circles mark the detected frequency-howls.

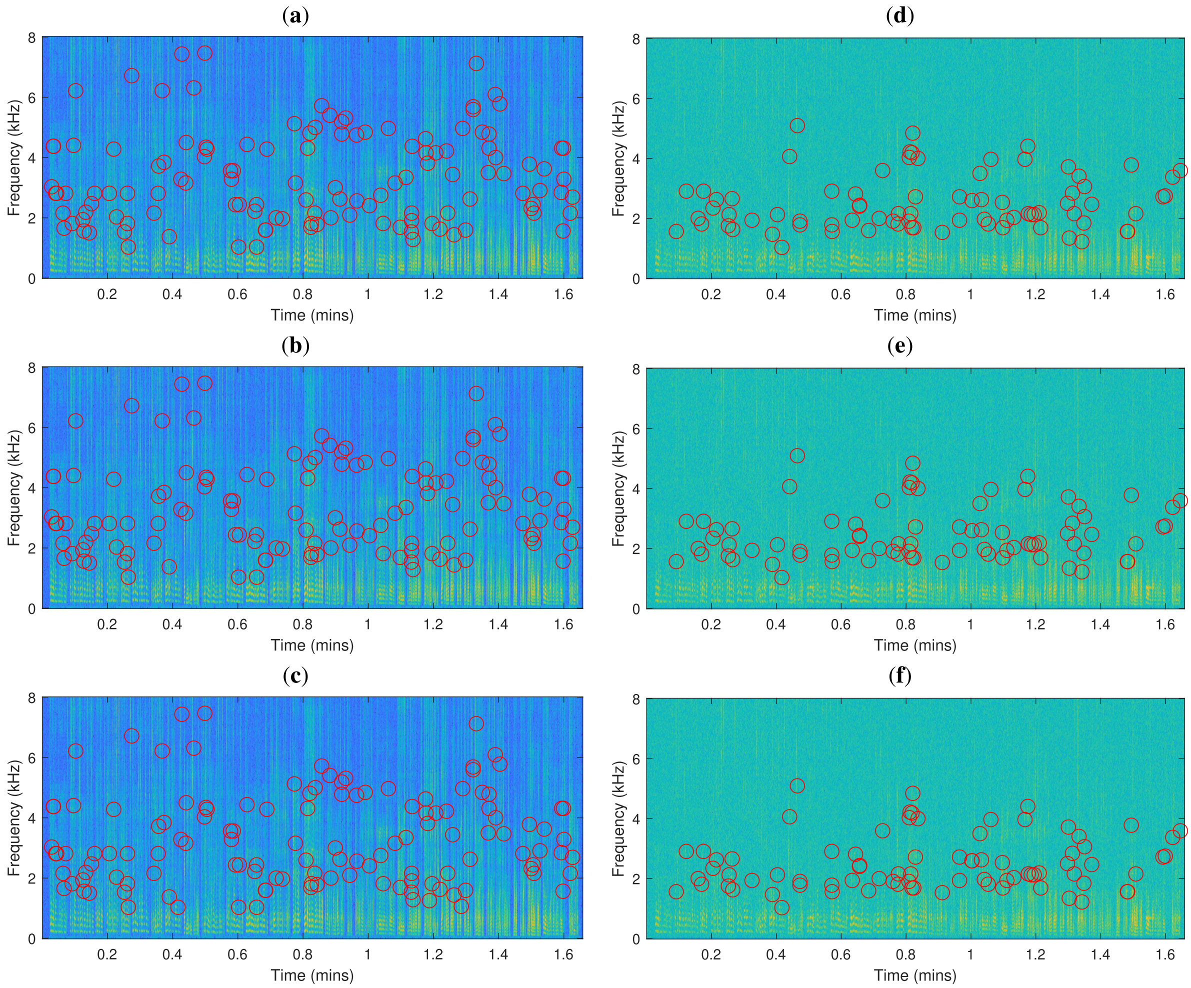


Fig. 7. Method comparison via retrospective howling detections on the spectrograms of the SR system's microphone input under a configured amplification gain $K = 7$ and iSNR values of 40 dB (Clean) and 10 dB (Noisy). The gain control algorithm is controlled by the temporal howling detector, using a Wiener filter in the SE segment and fixing the howling energy threshold-contour. The left column (**a–c**) refers to a clean input signal (40 dB iSNR), and the right column (**d–f**) refers to a noisy input signal (10 dB iSNR). On top of the spectrograms, retrospective howling detections are applied, marked by red circles: (**a,d**) via the temporal howling detector only (as used by the gain control algorithm); (**b,e**) via the temporal and the Magnitude-STD-based detectors; and (**c,f**) via the temporal and the NINOS^2-T-based detectors.

the speech distortion (due to howling) is measured via the median short-term Itakura-Saito distance (Med-IS-Dist) [21], [38]. Based on that, it is desired to compare the performance of the temporal-howling-detector-based gain control algorithm under a configured amplification gain $K = 7$ and 10 dB iSNR, with and without using a Wiener filter, see Figs. 6 and 5(c), respectively. Evaluating the performance over a long concatenated speech input signal of about 98 s, $K_{\text{eff}} = 3.85$ and Med-IS-Dist $= 2.14$ without filtering out the noise from the microphone signal using a Wiener filter, and $K_{\text{eff}} = 2.6$ and Med-IS-Dist $= 1.07$ when using one. Thus, the noise reduction filter has improved the signal distortion in the noisy scenario, even though the calculated matched effective gain is lower.

Fig. 7 shows spectrograms of the SR system's microphone input, where the gain control algorithm is controlled by the temporal howling detector on the microphone signal $m_1(n)$, under a configured amplification gain $K = 7$ and iSNR values of 40 dB (Clean) and 10 dB (Noisy). On top of the spectrograms, retrospective howling detections are applied via the temporal howling detector only, the temporal and the Magnitude-STD-based detectors, and the temporal and the NINOS^2-T-based detectors — all marked by red circles. Over the concatenated speech signal of about 98 s, the temporal howling detector detected 127 frequency-howls for 40 dB iSNR in Fig. 7(a) and 80 frequency-howls for 10 dB iSNR in Fig. 7(d). Both Fig. 7(a), (b) and (d), (e) show identical retrospective howling detections, which means that the Magnitude-STD-based detector detected no frequency-howls. Then, as shown in Fig. 7(c), (f), together with the NINOS^2-T-based detector, 130 frequency-howls were detected for 40 dB iSNR in Fig. 7(c), and 81 frequency-howls for 10 dB iSNR in Fig. 7(f). However, these additionally detected frequency-howls are suspected to be speech harmonics. Hence, it can be inferred that the proposed detectors lose their relevance when the environmental noise is reduced via the Wiener filter, as they no longer provide an advantage in detecting non-decreasing underdamped frequency-howls.

VII. Conclusion

We have investigated the challenges of howling control for in-car speech reinforcement systems in the presence of high environmental noise. The new term of non-decreasing underdamped frequency-howl relates to underdamped frequency-howls that do not decrease due to high environmental noise levels. Two detection methods are proposed to detect non-decreasing underdamped frequency-howls. These are the Magnitude-STD-based and NINOS^2-T-based detectors, which determine whether a steady energy level manifests as a frequency-howl with respect to the calculated total spectrum 90% noise floor. A Wiener noise reduction filter is utilized within the SR system to reduce the effective amplification gain (and acoustic feedback) per frequency component over time, based on the signal's magnitude relative to the estimated PSD of the noise. An optimal value of the over-subtraction factor was fine-tuned such that the Wiener filter sufficiently reduces the environmental noise, up to the presence of frequency-howls. Adjusting the howling energy threshold-contour is proposed to deal with acquired false-alarm-causing artifacts arising as the environmental noise level rises. The performance of the howling-detection-based gain control algorithm is evaluated in clean and noisy environments, incorporating the proposed howling detectors, noise reduction, and howling energy threshold adjustment. Reducing the amplification gain through the Wiener filter resulted in no visible non-decreasing underdamped frequency-howls. The Wiener filter has improved the signal distortion in the noisy scenario, even though the calculated matched effective gain is lower. Overall, the proposed detectors have performed well under both clean and noisy scenarios, yet the detection of non-decreasing underdamped frequency-howls becomes unnecessary when the environmental noise is successfully reduced via the Wiener filter. These findings can be exploited in any howling control mechanism operating under high environmental noise. Future work may address the estimation of an optimal over-subtraction factor for the Wiener filter and other noise reduction methods incorporated with the howling detection-based gain control.

References

[1] A. Ortega, E. Lleida, E. Masgrau, and F. Gallego, "Cabin car communication system to improve communications inside a car," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, 2002, pp. IV-3836–IV-3839.

[2] P. Bulling, K. Linhard, A. Wolf, and G. Schmidt, "Acoustic feedback compensation with reverb-based stepsize control for in-car communication systems," in *Proc. 12th ITG Conf. Speech Commun.*, 2016, pp. 337–341.

[3] C. Lüke, G. Schmidt, A. Theiß, and J. Withopf, *In-Car Communication*. New York, NY, USA: Springer, 2014, pp. 97–118, doi: 10.1007/978-1-4614-9120-0_7.

[4] A. Ortega, E. Lleida, and E. Masgrau, "Speech reinforcement system for car cabin communications," *IEEE Speech Audio Process.*, vol. 13, no. 5, pp. 917–929, Sep. 2005.

[5] G. Reuven, S. Gannot, and I. Cohen, "Joint noise reduction and acoustic echo cancellation using the transfer-function generalized sidelobe canceller," *Speech Commun.*, vol. 49, no. 7/8, pp. 623–635, 2007.

[6] J. Benesty, J. Chen, Y. Huang, and I. Cohen, *Noise Reduction in Speech Processing*, vol. 2. Berlin, Germany: Springer, 2009.

[7] K. Müller, S. Doclo, J. Østergaard, and T. Wolff, "Model-based estimation of in-car-communication feedback applied to speech zone detection," in *Proc. Int. Workshop Acoustic Signal Enhancement*, 2022, pp. 1–5.

[8] J. H. L. Hansen, K. Takeda, G. Schmidt, and H. Abut, *Vehicles, Drivers, and Safety*. Berlin, Boston: De Gruyter, 2020, ch. 11, 12, 13, pp. 165–229.

[9] F. Faccenda, S. Squartini, E. Principi, L. Gabrielli, and F. Piazza, "A real-time dual-channel speech reinforcement system for intra-cabin communication," *J. Audio Eng. Soc.*, vol. 61, no. 11, pp. 889–910, Nov. 2013.

[10] A. Spriet, S. Doclo, M. Moonen, and J. Wouters, *Feedback Control in Hearing Aids*. Berlin, Germany: Springer, 2008, pp. 979–1000, doi: 10.1007/978-3-540-49127-9_48.

[11] Y. Alkaher and I. Cohen, "Temporal howling detector for speech reinforcement systems," *Acoustics*, vol. 4, no. 4, pp. 967–995, 2022.

[12] F. Albu, L. T. Tran, and S. Nordholm, "The hybrid simplified Kalman filter for adaptive feedback cancellation," in *Proc. Int. Conf. Commun.*, 2018, pp. 45–50.

[13] M. Gimm, P. Bulling, and G. Schmidt, "Residual feedback suppression with extended model-based postfilters," *EURASIP J. Audio, Speech, Music Process.*, vol. 2021, no. 1, pp. 1–15, 2021.

[14] R. T. Reas, "Performance evaluation of a low order LMS based adaptive feedback canceller for public address system," in *Proc. IEEE 12th Int. Conf. Humanoid, Nanotechnol., Inf. Technol., Commun. Control, Environ., Manage.*, 2020, pp. 1–7.

[15] F. Albu, H. G. Coanda, and D. Coltuc, "Low complexity acoustic feedback cancellation algorithm with frequency shifting and dichotomous coordinate descent iterations," in *Proc. IEEE Int. Symp. Electron. Telecommun.*, 2022, pp. 1–4.

[16] K. Patel and I. M. Panahi, "Efficient real-time acoustic feedback cancellation using adaptive noise injection algorithm," in *Proc. IEEE 42nd Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2020, pp. 972–975.

[17] K. Patel, *Acoustic Feedback Cancellation and Dynamic Range Compression for Hearing Aids and Its Real-Time Implementation*. Dallas, TX, USA: Univ. Texas Dallas, 2020.

[18] T. van Waterschoot and M. Moonen, "Fifty years of acoustic feedback control: State of the art and future challenges," *Proc. IEEE*, vol. 99, no. 2, pp. 288–327, Feb. 2011.

[19] Z. Chen, Y. Hao, Y. Chen, G. Chen, and L. Ruan, "A neural network-based howling detection method for real-time communication applications," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2022, pp. 206–210.

[20] H. Zhang, M. Yu, and D. Yu, "Deep AHS: A deep learning approach to acoustic howling suppression," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2023, pp. 1–5.

[21] Y. Alkaher and I. Cohen, "Dual-microphone speech reinforcement system with howling-control for in-car speech communication," *Front. Signal Process.*, vol. 2, 2022, Art. no. 819113.

[22] M. C. Green, J. Szymanski, M. Speed, and U. Penryn, "Assessing the suitability of the magnitude slope deviation detection criterion for use in automatic acoustic feedback control," in *Proc. 19th Int. Conf. Digit. Audio Effects*, 2016, pp. 85–92.

[23] T. Van Waterschoot and M. Moonen, "Comparative evaluation of howling detection criteria in notch-filter-based howling suppression," *J. Audio Eng. Soc.*, vol. 58, no. 11, pp. 923–940, 2010.

[24] Y. Li, X. Huang, Y. Zheng, Z. Gao, L. Kou, and J. Wan, "Howling detection and suppression based on segmented notch filtering," *Sensors*, vol. 21, no. 23, 2021, Art. no. 8062.

[25] H. Pardede, K. Ramli, Y. Suryanto, N. Hayati, and A. Presekal, "Speech enhancement for secure communication using coupled spectral subtraction and Wiener filter," *Electronics*, vol. 8, no. 8, 2019, Art. no. 897.

[26] R. C. Dorf and R. H. Bishop, *Modern Control Systems*, 11th ed. Upper Saddle River, NJ, USA: Pearson Prentice-Hall, 2008, ch. 2, 4, 5, 7, pp. 50–57.

[27] M. Mounir, "Acoustic event detection: Feature, evaluation and dataset design," Ph.D. dissertation, Katholieke Universiteit Leuven, Kasteelpark Arenberg, Leuven, Belgium, Sep. 2020.

[28] M. Mounir, P. Karsmakers, and T. Van Waterschoot, "Guitar note onset detection based on a spectral sparsity measure," in *Proc. IEEE 24th Eur. Signal Process. Conf.*, 2016, pp. 978–982.

[29] *Acoustics—Normal Equal-Loudness-Level Contours*, ISO 226: 2003(E), International Standardization Organization, Geneva, Switzerland, Sep. 2003.

[30] J. Benesty, I. Cohen, and J. Chen, *Fundamentals of Signal Enhancement and Array Signal Processing*. Hoboken, NJ, USA: Wiley, 2018.

[31] Y. Avargel and I. Cohen, "System identification in the short-time fourier transform domain with crossband filtering," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 4, pp. 1305–1319, May 2007.

[32] Y. Avargel and I. Cohen, "On multiplicative transfer function approximation in the short-time Fourier transform domain," *IEEE Signal Process. Lett.*, vol. 14, no. 5, pp. 337–340, May 2007.

[33] Y. Avargel and I. Cohen, “Adaptive system identification in the short-time Fourier transform domain using cross-multiplicative transfer function approximation,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 1, pp. 162–173, Jan. 2008.

[34] Y. Avargel and I. Cohen, “Adaptive nonlinear system identification in the short-time Fourier transform domain,” *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3891–3904, Oct. 2009.

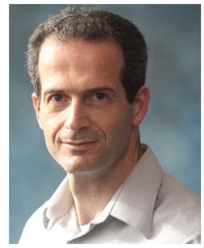
[35] Y. Ephraim and D. Malah, “Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.

[36] M. Berouti, R. Schwartz, and J. Makhoul, “Enhancement of speech corrupted by acoustic noise,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1979, pp. 208–211.

[37] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, and D. S. Pallett, “DARPA TIMIT acoustic-phonetic continous speech corpus CD-ROM. NIST speech disc 1-1.1,” *NASA STI/Recon Tech. Rep.*, vol. 93, 1993, Art. no. 27403.

[38] J. R. Deller Jr., J. G. Proakis, and J. H. Hansen, *Discrete Time Processing of Speech Signals*. New York, NY, USA: Wiley,ch. 5, 2000.

Yehav Alkaher received the B.Sc. degree in electrical engineering in 2016 from the Technion – Israel Institute of Technology, Haifa, Israel, where he is currently working toward the Ph.D. degree (direct track) in electrical engineering. Since 2016, he has been specializing as a signal processing, communications, data-science and algorithms R&D Engineer. His research interests include statistical signal processing, speech and audio processing, array signal processing, analysis and modeling of acoustic signals, speech communications, speech enhancement, probabilistic models, adaptive filtering, trade-off analysis, and control systems theory.



Israel Cohen (Fellow, IEEE) received the B.Sc. (*Summa Cum Laude*), M.Sc., and Ph.D. degrees in electrical engineering from the Technion – Israel Institute of Technology, Haifa, Israel, in 1990, 1993, and 1998, respectively. He is currently the Louis and Samuel Seidan Professor of electrical and computer engineering with the Technion – Israel Institute of Technology. From 1990 to 1998, he was a Research Scientist with RAFAEL Research Laboratories, Haifa, Israel Ministry of Defense. From 1998 to 2001, he was a Postdoctoral Research Associate with the Computer Science Department, Yale University, New Haven, CT, USA. In 2001, he joined the Electrical Engineering Department of the Technion. He is a coeditor of Multichannel Speech Processing Section of *Springer Handbook of Speech Processing* (Springer, 2008), and coauthor of *Fundamentals of Signal Enhancement and Array Signal Processing* (Wiley-IEEE Press, 2018). His research interests include array processing, statistical signal processing, deep learning, analysis and modeling of acoustic signals, speech enhancement, noise estimation, microphone arrays, source localization, blind source separation, system identification, and adaptive filtering. He was the recipient of the Honorary Doctorate from the Karunya Institute of Technology and Sciences, Coimbatore, India (2023), Norman Seiden Prize for Academic Excellence (2017), SPS Signal Processing Letters Best Paper Award (2014), Alexander Goldberg Prize for Excellence in Research (2010), and Muriel and David Jacknow Award for Excellence in Teaching (2009). He was as Associate Editor for IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING and IEEE SIGNAL PROCESSING LETTERS, a Member of IEEE Audio and Acoustic Signal Processing Technical Committee and IEEE Speech and Language Processing Technical Committee, and a Distinguished Lecturer of IEEE Signal Processing Society.