Design of speech leakage-suppressed audio-spot based on auditory masking area control with active masker cancellation using parametric array loudspeakers

Tomoki Hashida*, Yuting Geng*, Masato Nakayama*, and Takanobu Nishiura*
* Ritsumeikan University, Osaka, Japan
E-mail: {is0629es@ed, geng@fc, mnaka@fc, nishiura@is}.ritsumei.ac.jp

***Abstract*—A parametric array loudspeaker (PAL) utilizes ultrasound to reproduce highly directional sound, which can form an area referred to as an "audio-spot." By emitting ultrasound separately from different PALs, the audio-spot is formed in the overlapping area of the ultrasonic beams. However, for signals with harmonic structures such as speech, speech components may leak outside the audio-spot, resulting in unintended auditory perception. To address this issue, we propose a method for designing a speech leakage-suppressed audio-spot based on auditory masking. This method employs two PALs: one PAL emits a speech signal with a masker signal, while the other emits the same speech signal with an anti-masker signal. The anti-masker is generated based on active noise control (ANC) principles to cancel the masker at the control point, where ultrasonic beams from the two PALs overlap. Consequently, the auditory masking area is spatially controlled. Specifically, at the control point, the masker is canceled, and the speech is enhanced. Outside the control point, the masker and anti-masker become dominant, and speech leakage is effectively suppressed by auditory masking. As a result, a speech leakage-suppressed audio-spot is formed at the control point. Experimental results show that the proposed method suppresses speech leakage more effectively than the conventional method.**

I. Introduction

Loudspeakers are commonly used to reproduce sound in various settings [1]. Conventional electro-dynamic loudspeakers emit sound over a wide area. Therefore, there is an issue that the speech content may be leaked to unintended listeners. A PAL can achieve sharp directivity by using ultrasound, thereby confining sound to a limited spatial area [2], [3]. A PAL reproduces audible sound by using nonlinear interactions in the air. When ultrasound containing multiple frequency components is emitted at high intensity as a primary wave, nonlinear interactions in the air generate secondary wave, including sum frequencies, difference frequencies, and harmonics. Based on this principle, the target signal can be self-demodulated by emitting the modulated ultrasonic signal so that the difference frequency matches the target signal. [4]. By emitting the modulated ultrasonic signal from a single PAL, the target signal is reproduced along the propagation axis of the ultrasonic beams. This forms an area known as an audio-spot, where the demodulated sound of the target signal is perceivable [5].

When the modulated ultrasonic signal is separated into the carrier and sideband signals and each is emitted from different PALs, the target signal is reproduced only in the area where the ultrasonic beams overlap [6]. This method allows the audio-spot to be designed for a narrower area. However, when the target signal has a harmonic structure, such as speech, nonlinear interactions can unintentionally generate target signal components along the propagation axis of the sideband signal. [7]. This results in speech leakage to unintended listeners. Furthermore, the audible sound pressure level decreases due to the limited demodulation area [8].

Another approach to suppressing speech perception by unintended listeners is the use of auditory masking [9]. Auditory masking is a phenomenon in which the perception of a sound is hindered by another sound [10]. The interfering sound that prevents perception is referred to as the "masker." For instance, the perception of a speech signal can be hindered by superimposing a masker signal, such as pink noise, which shares similar acoustic characteristics with speech [11]. By exploiting this effect, speech perception by unintended listeners can be effectively obstructed, thereby suppressing the leakage of speech information.

In this paper, we propose a method for designing a speech leakage-suppressed audio-spot based on the control of auditory masking. The proposed approach is inspired by the fundamental principle of ANC, in which unwanted noise is canceled by introducing an anti-noise signal with the same amplitude and opposite phase [12], [13]. In the proposed method, auditory masking is controlled by employing a masker signal and an anti-masker signal that is designed to cancel each other out at the control point. Although the anti-masker signal is intended to cancel the masker signal, it functions in the same way as a regular masker when used alone and can interfere with speech perception. This method employs two PALs: one emits the modulated signal of the target speech signal combined with the masker signal, while the other emits the modulated signal of the same speech signal combined with the anti-masker signal. At the control point, where the ultrasonic beams from both PALs overlap, the anti-masker cancels the masker via ANC-like interaction, allowing clearer perception of the speech

979-8-3315-7206-8/25/$31.00

signal. Outside the control point, where cancellation does not occur, the masker and anti-masker signals each act as maskers, suppressing the intelligibility of the speech through auditory masking. This results in a spatially controlled, speech leakage-suppressed audio-spot.

II. Proposed Speech Leakage-Suppressed Audio-Spot Design

A. Overview of the proposed method

We propose a method for designing a speech leakage-suppressed audio-spot based on auditory masking area control with active masker cancellation. Figure 1 shows an overview of the proposed method, and Fig. 2 shows an overview of the design process for the signals emitted for each PAL. Here, n denotes time index, z denotes the complex variable in the z-domain, $H_{\mathrm{A}}(z)$ and $H_{\mathrm{B}}(z)$ denote the transfer functions from $\mathrm{PAL_A}$ and $\mathrm{PAL_B}$ to the control point, respectively. In the proposed method, the anti-masker signal is generated by applying the anti-masker control filter to the masker signal. The design of this control filter is described in Section II-B. Denoting the filter as $g(n)$, the anti-masker signal is given by:

$$v_{\mathrm{M_{anti}}}(n) = g(n) * v_{\mathrm{M}}(n), \tag{1}$$

where $*$ denotes the convolution operation. The masker and anti-masker signals are mixed with the speech signal $v_{\mathrm{S}}(n)$ to generate two mixed signals $v_{\mathrm{T_A}}(n)$ and $v_{\mathrm{T_B}}(n)$. These mixed signals are expressed as:

$$v_{\mathrm{T_A}}(n) = A_{\mathrm{S}} v_{\mathrm{S}}(n) + v_{\mathrm{M}}(n-\tau), \tag{2}$$
$$v_{\mathrm{T_B}}(n) = A_{\mathrm{S}} v_{\mathrm{S}}(n) + v_{\mathrm{M_{anti}}}(n), \tag{3}$$

where τ denotes the samples of delay time to ensure the causality of the filter, A_{S} denotes a scaling coefficient that adjusts the power ratio between the speech signal and the masker signal. When R_{SM} [dB] is defined as the power ratio between the masker signal and the speech signal, the coefficient A_{S} can be expressed as:

$$A_{\mathrm{S}} = \frac{P_{\mathrm{M}}}{P_{\mathrm{S}}} \times 10^{R_{\mathrm{SM}}/10}. \tag{4}$$

Each of the mixed signals $v_{\mathrm{T_A}}(n)$ and $v_{\mathrm{T_B}}(n)$ is treated as the target signal for $\mathrm{PAL_A}$ and $\mathrm{PAL_B}$, respectively, and is emitted after lower sideband (LSB) amplitude modulation [14]. Let the carrier signal be denoted by $c(n)$. Then, the LSB amplitude-modulated signals are expressed as:

$$v_{\mathrm{LSB_A}}(n) = c(n) + v_{\mathrm{T_A}}(n)c(n) + \mathcal{P}[v_{\mathrm{T_A}}(n)]\mathcal{P}[c(n)], \tag{5}$$
$$v_{\mathrm{LSB_B}}(n) = c(n) + v_{\mathrm{T_B}}(n)c(n) + \mathcal{P}[v_{\mathrm{T_B}}(n)]\mathcal{P}[c(n)], \tag{6}$$

where $\mathcal{P}[\cdot]$ denotes the operator that shifts the phase of a signal by $\pi/2$ using the Hilbert transform [15].

Along the propagation axis of each PAL, the masker and anti-masker interfere with the perception of the speech. In contrast, at the control point where the two ultrasonic beams overlap, the anti-masker cancels the masker, enabling clear perception of the speech. As a result, a speech leakage-suppressed audio-spot is formed, where the speech is heard

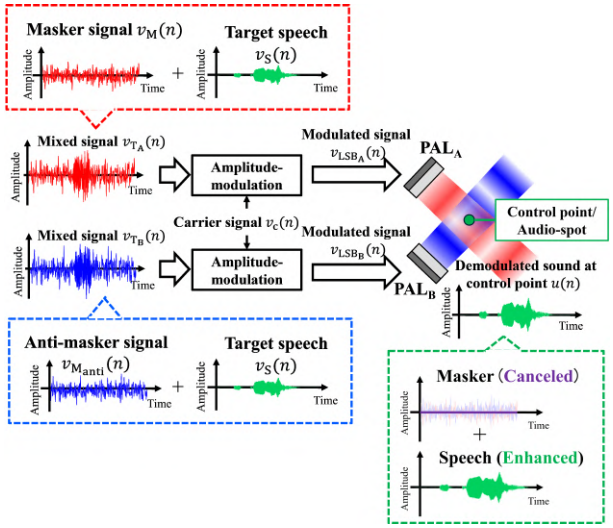

Fig. 1. Overview of the proposed method.

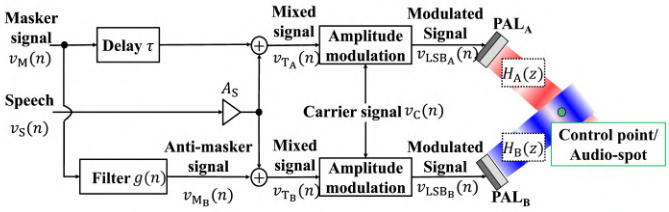

Fig. 2. Overview of the design process for the mixed signals used for each PAL.

only at the control point. The proposed method applies the fundamental concept of ANC to design the anti-masker signal. Specifically, the masker is regarded as the noise in the ANC framework. Based on this, the anti-masker is designed as an anti-noise with the same amplitude and opposite phase at the control point. In the proposed method, the signals undergo LSB amplitude modulation and self-demodulation, which are not present in conventional ANC systems. However, when amplitude modulation is applied, inverting the phase of the target signal also inverts the phase of the demodulated sound [16]. Therefore, the anti-masker signal can still be designed based on the principle of ANC.

B. Design of the anti-masker control filter

The anti-masker control filter $g(n)$ is designed so that the anti-masker cancels the masker at the control point. In this case, by considering the masker as the noise in an ANC framework, the design of the filter of $g(n)$ can be discussed in a manner analogous to control filter design in ANC systems that include a secondary path. Although the demodulation process in PALs involves nonlinear acoustic propagation, distortion components arising from this nonlinearity are relatively minor compared to the linear components. Therefore, the system is approximated as linear in this analysis. The demodulated sound at the control point can be represented in the z-domain as follows:

$$
\begin{aligned}
U(z) &\approx H_{\mathrm{A}}(z)\left[A_{\mathrm{S}}V_{\mathrm{S}}(z)+V_{\mathrm{M}}(z)z^{-\tau}\right] \\
&\quad +H_{\mathrm{B}}(z)\left[A_{\mathrm{S}}V_{\mathrm{S}}(z)+G(z)V_{\mathrm{M}}(z)\right] \\
&= \left[H_{\mathrm{A}}(z)+H_{\mathrm{B}}(z)\right]A_{\mathrm{S}}V_{\mathrm{S}}(z) \\
&\quad +H_{\mathrm{A}}(z)V_{\mathrm{M}}(z)z^{-\tau} \\
&\quad +H_{\mathrm{B}}(z)G(z)V_{\mathrm{M}}(z), \qquad (7)
\end{aligned}
$$

where $V_{\mathrm{S}}(z)$, $V_{\mathrm{M}}(z)$, $G(z)$, and $U(z)$ denote the z-transforms of the speech signal $v_{\mathrm{S}}(n)$, the masker signal $v_{\mathrm{M}}(n)$, the anti-masker control filter $g(n)$, and the demodulated sound $u(n)$ at the control point, respectively.

The condition under which the masker and anti-masker are canceled at the control point is derived from Eq. (7) as:

$$
H_{\mathrm{A}}(z)V_{\mathrm{M}}(z)z^{-\tau}+H_{\mathrm{B}}(z)G(z)V_{\mathrm{M}}(z) = 0. \qquad (8)
$$

From Eq. (8), the anti-masker control filter $G(z)$ is obtained as:

$$
G(z) = -\frac{H_{\mathrm{A}}(z)}{H_{\mathrm{B}}(z)}z^{-\tau}. \qquad (9)
$$

The anti-masker control filter $G(z)$ was adaptively designed using the filtered-x normalized least mean square (FxNLMS) algorithm [17]. In the FxNLMS algorithm, a pre-estimated secondary path model is used to compensate for the effects of the secondary path during filter coefficient updates. This improves both the convergence speed of the adaptive algorithm and the estimation accuracy of the transfer characteristics [18]. Figure 3 shows an overview of the adaptive filter design based on the FxNLMS algorithm. In this study, the transfer function estimation is performed offline using pre-recorded signals. To obtain the desired signal $u_{\mathrm{A}}(n)$ for the adaptive filter, the masker signal with delay $v_{\mathrm{M}}(n-\tau)$ is emitted from $\mathrm{PAL_A}$. The recorded signal is used as the desired signal after downsampling and removing the ultrasonic components.

The filter coefficients are updated using the normalized least mean squares (NLMS) algorithm [19] to minimize the error signal. The input signal $r(n)$ to the adaptive filter is the convolution of the masker $v_{\mathrm{M}}(n)$ and the pre-identified secondary path model $\hat{H_B}(z)$, and is given by:

$$
r(n) = v_{\mathrm{M}}(n) * \hat{h_{\mathrm{B}}}(n), \qquad (10)
$$

where $\hat{h_{\mathrm{B}}}(n)$ is the impulse response corresponding to $\hat{H_B}(z)$. The error signal $e(n)$ of the adaptive filter is defined as:

$$
e(n) = v'_{\mathrm{M}}(n)+u_{\mathrm{A}}(n). \qquad (11)
$$

The update equation for the adaptive filter using the NLMS algorithm is given by:

$$
\boldsymbol{g}(n+1) := \boldsymbol{g}(n)+\frac{\alpha e(n)}{||\boldsymbol{r}(n)||^2+\beta}\boldsymbol{r}(n), \qquad (12)
$$

where $\boldsymbol{g}(n)$ is the filter coefficient vector, $\boldsymbol{r}(n)$ is the input signal vector, α ($0 < \alpha < 2$) is the step size, β is the regularization parameter, and $||\cdot||^2$ denotes the squared norm of a vector. The vectors $\boldsymbol{g}(n)$ and $\boldsymbol{r}(n)$ are defined as:

$$
\boldsymbol{g}(n) := \left(g^{(n)}[0], g^{(n)}[1], \cdots, g^{(n)}[N-1]\right)^T, \qquad (13)
$$

$$
\boldsymbol{r}(n) := \left(r[n], r[n-1], \cdots, r[n-N+1]\right)^T, \qquad (14)
$$

where N denotes the tap length of the adaptive filter.

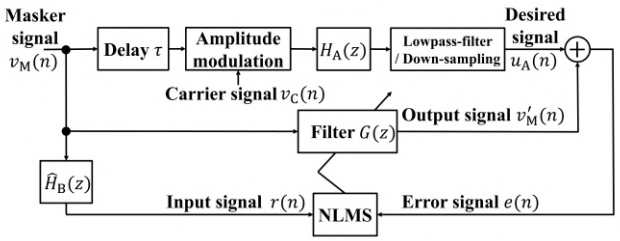

Fig. 3. Overview of the design process for the anti-masker control filter.

At each time index n, the filter coefficients are updated, and n is incremented by one for the next update. Through this iterative process, the filter coefficients converge to values that minimize the error signal. During this process, the vectors $\boldsymbol{g}(n)$ and $\boldsymbol{r}(n)$ are updated.

By convolving the masker signal $v_{\mathrm{M}}(n)$ with the masking control filter designed using the adaptive algorithm, the anti-masker signal $v_{\mathrm{M_{anti}}}(n)$ is obtained that has the same amplitude but opposite phase with respect to $v_{\mathrm{M}}(n)$ at the control point. Each signal is used in the respective mixed signal used for the PALs. As a result, the masker components are canceled at the control point, enabling clear perception of the speech.

III. Evaluation Experiment

In this evaluation experiment, we first investigated the cancellation performance of the anti-masker signal against the masker signal. Next, the effectiveness of the proposed method was verified by comparing the speech quality and the sound pressure level of the speech component in the demodulated sound with those obtained using the conventional method.

A. Common experimental conditions

Table I summarizes the experimental conditions. The equipment used is listed in Table II, and the arrangement of the experimental setup is shown in Fig.4. In the setup, the two PALs were arranged such that their propagation axes were orthogonal to each other. A photograph of the setup is presented in Fig. 5, and Table III provides the conditions for designing the anti-masker control filter.

In this paper, band-limited pink noise was used as the masker signal $v_{\mathrm{M}}(n)$. Pink noise exhibits a frequency characteristic that attenuates at -3 dB/octave from low to high frequencies, which resembles the spectral sensitivity of human auditory perception [20], [21]. Furthermore, the primary frequency components of speech are concentrated within the range of 300 Hz to 3400 Hz. Therefore, by limiting the pink noise to this frequency range, it is possible to achieve effective auditory masking focused on the dominant frequency band of speech.

TABLE I
Experimental conditions

Environment	Soundproof room (T_{60}=0.15 s)
Ambient noise level	L_{A} = 20.1 dB
Temperature / Humidity	17.4 °C / 30.1%
Sampling frequency	192000 Hz
Quantization	24 bits
Modulation method	LSB amplitude modulation
Carrier frequency	40000 Hz

TABLE II
EXPERIMENTAL EQUIPMENT

Parametric array loudspeaker	MEE, PS-60E
Power amplifier	Victor, PS-A2002
Audio interface	RME, Fireface UFX
Microphone	SENNHEISER, MKH 8020
Sound level meter	RION, NL-27

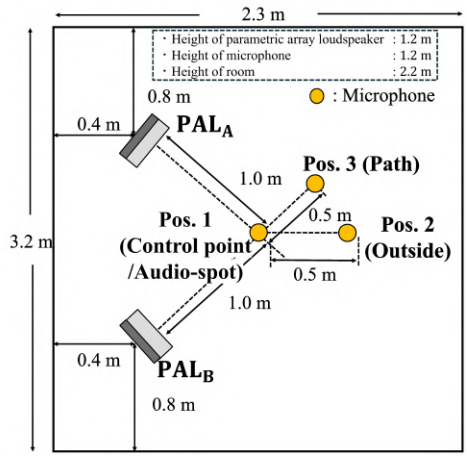

Fig. 4. Arrangement of the experimental equipment.

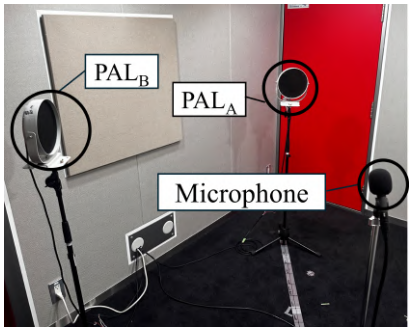

Fig. 5. Photograph of the experimental setup.

TABLE III
CONDITIONS FOR DESIGNING THE ANTI-MASKER CONTROL FILTER

Sampling frequency	8000 Hz
Frequency range	300 - 3400 Hz
Tap length of $G(z)$	4000 samples (500 ms)
Tap length of $\hat{H}_2(z)$	1000 samples (125 ms)
Delay τ	2000 samples (250 ms)
Update algorithm	FxNLMS
Step size parameter	0.05
Regularization parameter	1.0×10^{-6}

B. Evaluation of Masker Cancellation Performance

To evaluate the cancellation performance of the anti-masker at the control point, the sound pressure level (SPL) of the demodulated sound was compared.

The SPL of the demodulated sound L [dB] was calculated as follows:

$$L = L_{\mathrm{pis}} - 10\log_{10}\left(\frac{\sigma_{\mathrm{rec}}^2}{\sigma_{\mathrm{pis}}^2}\right), \tag{15}$$

where L_{pis} [dB] is the SPL of the acoustic calibrator sound measured by the sound level meter, and σ_{rec} and σ_{pis} are the standard deviations of the recorded signals of the demodulated sound and the acoustic calibrator sound, respectively. The comparison conditions for evaluating the masker cancellation performance at the control point are as follows:

- **w/o anti-masker:** Only the modulated signal of the masker signal $v_{\mathrm{M}}(n)$ is emitted from $\mathrm{PAL_A}$, while no signal is emitted from $\mathrm{PAL_B}$.
- **w/ anti-masker:** The modulated signal of the masker signal $v_{\mathrm{M}}(n)$ is emitted from $\mathrm{PAL_A}$, and the modulated signal of the anti-masker signal $v_{\mathrm{M_{anti}}}(n)$ is simultaneously emitted from $\mathrm{PAL_B}$.

In this experiment, ten recordings were made in each condition, and the average sound pressure was compared.

Figure 6 shows the waveform of the demodulated sound at the control point under both conditions. The SPL of the demodulated sound at the control point was 51.6 dB under the w/o anti-masker condition and 38.1 dB under the w/ anti-masker condition. This indicates that the masker signal was reduced by 13.5 dB through cancellation by the anti-masker generated using the anti-masker control filter. The proposed method suppresses auditory masking at the control point by using both the masker and anti-masker signals. This enables the perception of the speech signal and the formation of an audio-spot.

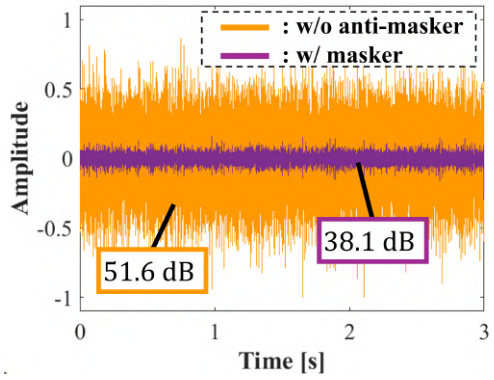

Fig. 6. Waveform of the demodulated sound

C. Evaluation of speech leakage-suppressed audio-spot formation

To verify the effectiveness of the audio-spot formed by the proposed method, the perceptual evaluation of speech quality (PESQ) [22] was computed for the demodulated sound at the control point and at locations outside of it. Additionally, the sound pressure level (SPL) of the speech component was calculated at the control point.

PESQ is based on a model of human auditory perception and is known to correlate highly with subjective speech quality evaluations. PESQ scores range from -0.5 to 4.5, with higher values indicating higher speech quality and lower values indicating poorer quality. Evaluations were conducted at three positions: the overlap point of the ultrasonic beams of the PALs (Pos.1 (Control point/Audio-spot)), a location on the propagation axis (Pos.2 (Path)), and a location outside the propagation axis (Pos.3 (Outside)). In this study, higher PESQ scores are

desirable at the Pos.1 (Control point/Audio-spot), while lower scores are expected at Pos.2 (Path) and Pos.3 (Outside). For the SPL of the demodulated sound in speech component L_{S} [dB], since the demodulated sound contains both speech and masker components, it was approximately calculated using the following equation:

$$L_{\mathrm{S}} \approx 10 \log_{10} \left(10^{L_{\mathrm{mix}}/10} - 10^{L_{\mathrm{masker}}/10} \right), \qquad (16)$$

where L_{masker} [dB] denotes the SPL of the demodulated sound containing only the masker, and L_{mix} [dB] denotes the SPL of the demodulated sound containing both the masker and the speech signal. We evaluated the following comparison conditions:

- **Prop. condition:** The modulated signal $v_{\mathrm{LSB_A}}(n)$ containing the speech signal $v_{\mathrm{S}}(n)$ the masker signal $v_{\mathrm{M}}(n)$ is emitted from $\mathrm{PAL_A}$, while the modulated signal $v_{\mathrm{LSB_B}}(n)$ containing the speech signal $v_{\mathrm{S}}(n)$ and the anti-masker signal $v_{\mathrm{M_{anti}}}(n)$ is emitted from $\mathrm{PAL_B}$.
- **Conv. condition:** The modulated signal of the speech signal is divided into two components. One is emitted from $\mathrm{PAL_A}$, and the other is emitted from $\mathrm{PAL_B}$.

In the proposed method, the power ratio between the speech signal and the masker signal R_{SM} was set to four levels: -20 dB, -15 dB, -10 dB, and -5 dB. As noted in Section II-A, a higher R_{SM} indicates a stronger speech component in the mixed signal. As speech signals, 20 words were selected from the ATR phonetically balanced 216-word set [23], such that the phoneme occurrences were as evenly distributed as possible. All emitted signals were normalized to have the same peak amplitude.

First, Fig. 7 shows the mean and standard deviation of the PESQ scores at Pos.1 (Control point/Audio-spot). As shown in Fig. 7, the PESQ scores under the Prop.($R_{\mathrm{SM}} = -10$ dB), and Prop.($R_{\mathrm{SM}} = -5$ dB) conditions were higher than those under the Conv. condition. These results indicate that, under certain conditions, the proposed method provides better speech intelligibility at the control point than the conventional method.

Next, Figs. 8 and 9 show the mean and standard deviation of the PESQ scores at Pos.2 (Outside) and Pos.3 (Path), respectively. As shown in Figs. 8 and 9, the PESQ scores under the Prop.($R_{\mathrm{SM}} = -20$ dB) and Prop.($R_{\mathrm{SM}} = -15$ dB) conditions were lower than those under the Conv. condition at locations outside the control point. These lower PESQ scores are attributed to the absence of masker cancellation in these areas. As a result, auditory masking effectively occurs, which reduces the perceptual clarity of the speech signal.

Based on the PESQ results at the three evaluation points, under certain conditions, it is confirmed that the proposed method is capable of maintaining high speech intelligibility at the control point while reducing intelligibility at outside that point, compared to the conventional method. These findings demonstrate the potential of the proposed method to form a speech leakage-suppressed audio-spot. However, the results also reveal a trade-off: too low masker level results in insufficient speech leakage suppression at outside of the control point, whereas too high masker level may lead to degradation in intelligibility at the control point.

Finally, Fig. 10 shows the SPL of the speech component at the control point. As shown in Fig. 10, under the Prop.($R_{\mathrm{SM}} = -15$ dB), Prop.($R_{\mathrm{SM}} = -10$ dB), and Prop.($R_{\mathrm{SM}} = -5$ dB) conditions, the SPL of the speech component was higher than that under the Conv. condition. These results indicate that, under certain conditions, the proposed method forms an audio-spot with a higher SPL of the speech component than the conventional method.

Based on both the PESQ and SPL evaluations, it can be concluded that the proposed method is capable of forming a speech leakage-suppressed audio-spot that not only effectively reduces speech leakage, but also improves the SPL of the speech component compared to the conventional method at the control point.

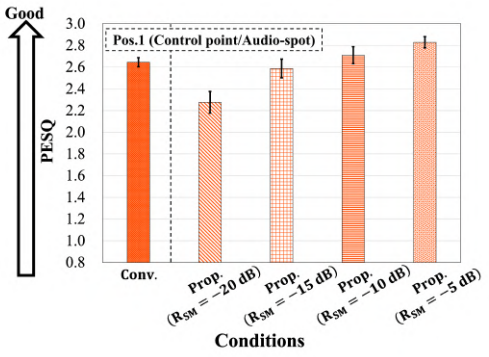

Fig. 7. PESQ scores at Pos.1 (Control point/Audio-spot).

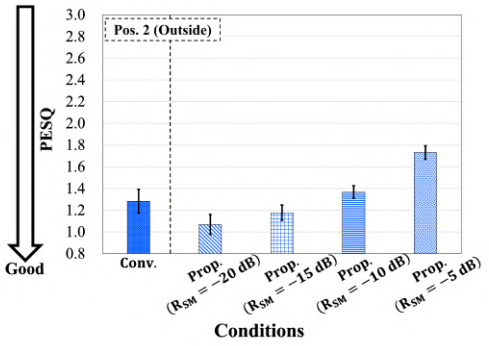

Fig. 8. PESQ scores at Pos.2 (Outside).

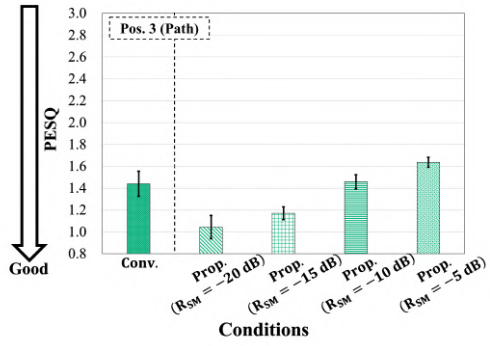

Fig. 9. PESQ scores at Pos.3 (Path).

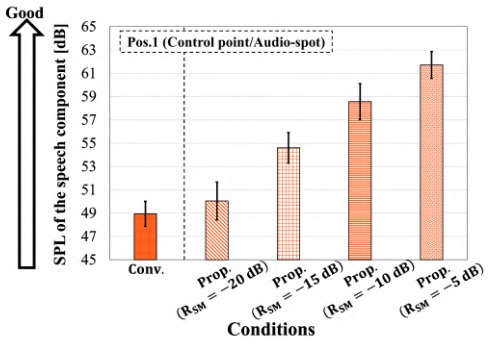

Fig. 10. SPL of the speech component at Pos.1 (Control point/Audio-spot).

IV. Conclusions

This paper proposed a method for designing a speech leakage-suppressed audio-spot based on auditory masking control. In the proposed method, two PALs are used: one PAL emits the modulated signal of the speech signal superimposed with the masker signal, and the other emits the modulated signal of the same speech signal superimposed with an anti-masker signal. The anti-masker is generated based on the principle of ANC to cancel the masker at the control point. As a result, speech is clearly perceived at the control point due to masker cancellation and speech enhancement. Outside this point, the masker and anti-masker dominate the demodulated sounds, and speech leakage is effectively suppressed by auditory masking. Thus, a speech leakage-suppressed audio-spot is successfully formed. Experimental evaluations demonstrated that the masker was successfully canceled by the anti-masker at the control point, resulting in the formation of an audio-spot with suppressed speech leakage. In addition, it was confirmed that the proposed method increases the SPL of the speech component. In future work, we intend to investigate the optimal masker level to achieve both effective leakage suppression and high intelligibility at the control point. Furthermore, we plan to implement real-time filter control to adapt to noisy environments, similar to ANC systems. This approach aims to cancel both the masker and background noise, thereby further enhancing speech intelligibility within the audio-spot.

Acknowledgment

This work was partly supported by Ritsumeikan University R-GIRO, ARC and RARA, and JSPS KAKENHI Grant Numbers JP23K28115, JP24K20803 and JP25H01158.

References

[1] R. H. Small *et al.*, "Closed-box loudspeaker systems part i: Analysis," *Journal of the Audio Engineering Society*, vol. 20, no. 10, pp. 978–808, 1972.

[2] C. Shi and W.-S. Gan, "Development of parametric loudspeaker," *IEEE potentials*, vol. 29, no. 6, pp. 20–24, 2010.

[3] K. Aoki, T. Kamakura, and Y. Kumamoto, "Parametric loudspeaker —characteristics of acoustic field and suitable modulation of carrier ultrasound," *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)*, vol. 74, no. 9, pp. 76–82, 1991.

[4] W.-S. Gan, J. Yang, and T. Kamakura, "A review of parametric acoustic array in air," *Applied Acoustics*, vol. 73, no. 12, pp. 1211–1219, 2012.

[5] T. Matsui, D. Ikefuji, M. Nakayama, and T. Nishiura, "A design of audio spot based on separating emission of the carrier and sideband waves," in *Proceedings of Meetings on Acoustics*, vol. 19, Paper ID: 1pSPc27, AIP Publishing, 2013.

[6] T. Matsui, D. Ikefuji, M. Nakayama, and T. Nishiura, "Multiple audio spots design based on separating emission of carrier and sideband waves," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, vol. 249, pp. 2924–2928, Institute of Noise Control Engineering, 2014.

[7] M. Iwagami, Y. Geng, M. Nakayama, and T. Nishiura, "Speech-leakage reduction for pin-spot audio based on multiple sideband waves using comb filters," in *2024 IEEE 13th Global Conference on Consumer Electronics (GCCE)*, pp. 407–410, IEEE, 2024.

[8] R. Uemura, D. Ikefuji, T. Fukumori, M. Nakayama, and T. Nishiura, "Sound pressure improvement in audio-spot-design based on separating emission with double sideband modulation," in *Proceedings of WESPAC2015*, pp. 473–478, 2015.

[9] S. A. Anand, P. Walker, and N. Saxena, "Compromising speech privacy under continuous masking in personal spaces," in *2019 17th International Conference on Privacy, Security and Trust (PST)*, pp. 1–10, IEEE, 2019.

[10] G. Kidd Jr, C. R. Mason, V. M. Richards, F. J. Gallun, and N. I. Durlach, "Informational masking," *Auditory perception of sound sources*, pp. 143–189, 2008.

[11] H. Masuda, Y. Hioka, C. J. Hui, J. James, and C. I. Watson, "Performance evaluation of speech masking design among listeners with varying language backgrounds," *Applied Acoustics*, vol. 201, p. 109122, 2022.

[12] Y. Kajikawa, W.-S. Gan, and S. M. Kuo, "Recent advances on active noise control: open issues and innovative applications," *APSIPA Transactions on Signal and Information Processing*, vol. 1, p. e3, 2012.

[13] S. M. Kuo and D. R. Morgan, "Active noise control: a tutorial review," *Proceedings of the IEEE*, vol. 87, no. 6, pp. 943–973, 1999.

[14] Y. Wang, X. Li, L. Xu, and L. Xu, "SSB modulation of the ultrasonic carrier for a parametric loudspeaker," in *2009 International Conference on Electronic Computer Technology*, pp. 669–673, 2009.

[15] L. Marple, "Computing the discrete-time "analytic" signal via FFT," *IEEE Transactions on signal processing*, vol. 47, no. 9, pp. 2600–2603, 1999.

[16] Y. Geng, M. Nakayama, and T. Nishiura, "Narrow-edged beamforming based on individual phase inversion in amplitude-modulated wave for parametric array loudspeaker," *Applied Acoustics*, vol. 200, p. 109060, 2022.

[17] S. Gaur and V. Gupta, "A review on filtered-x LMS algorithm," *Int. J. Signal Process. Syst*, vol. 4, no. 2, pp. 172–176, 2016.

[18] E. Bjarnason, "Analysis of the filtered-X LMS algorithm," *IEEE transactions on speech and audio processing*, vol. 3, no. 6, pp. 504–514, 1995.

[19] S. S. Haykin, *Adaptive filter theory*. Pearson Education India, 2002.

[20] S.-Y. Lu, Y.-H. Huang, and K.-Y. Lin, "Spectral content (colour) of noise exposure affects work efficiency," *Noise and Health*, vol. 22, no. 104, pp. 19–27, 2020.

[21] N. Naal-Ruiz, L. Alonso-Valerdi, and D. Ibarra-Zarate, "Neurophysiological and psychoacoustical changes after exposure to modified pink noise by frequency responses of headphones," *Applied Acoustics*, vol. 205, p. 109265, 2023.

[22] A. Rix, J. Beerends, M. Hollier, and A. Hekstra, "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*, vol. 2, pp. 749–752, 2001.

[23] A. Kurematsu, K. Takeda, Y. Sagisaka, S. Katagiri, H. Kuwabara, and K. Shikano, "ATR Japanese speech database as a tool of speech recognition and synthesis," *Speech communication*, vol. 9, no. 4, pp. 357–363, 1990.