Acoust. Sci. & Tech. **41**, 4 (2020)

PAPER

Selective distance recording using acoustic lenses

Maho Kobayashi[1,*] and Yoshinori Takahashi[2,†]

[1] *University of Electro-Communications, 1–5–1 Chofugaoka, Chofu, 182–8585 Japan*
[2] *Tokyo Metropolitan College of Industrial Technology, 8–17–1, Minamisenju, Arakawa-ku, Tokyo, 116–8523 Japan*

(*Received 30 May 2019, Accepted for publication 6 November 2019*)

Abstract: The purpose of this study is to emphasize sound from a specific place by using an acoustic lens in sound recording. Signal processing related to sound source separation using a microphone array has been studied as a robust sound recording technique against ambient noise; however, real-time processing is still difficult to implement because of its calculation cost. On the other hand, a directional microphone can realize high sensitivity in a specific direction; but, it is unable to show such selectivity for the distance. This paper has proposed a method of removing the arrival time difference of sound waves from a specific position by utilizing an acoustic lens composed of multiple horns. The distance selectivity and speech intelligibility were evaluated for a prototype acoustic lens made by a 3D printer. As a result, it was confirmed that sound recorded with the acoustic lens had a higher sound pressure level than that recorded with a conventional microphone for a desired position, and high speech intelligibility was constantly maintained up to the desired sound source distance.

Keywords: Acoustic lens, Directional microphone, Recording, Delay and sum, Acoustic horn, MTF-STI

PACS number: 43.58.Ls, 43.38.Kb, 43.20.Mv [doi:10.1250/ast.41.654]

1. INTRODUCTION

This research is aimed at sound recording (selective distance recording) using an acoustic lens having sensitivity for a specific distance and sharp directivity owing to a delayed sum of sound waves arriving from a specific position.

As a method of extracting a specific sound source in an environment that has many noises souses, one possible strategy would be to combine of a microphone array and spectral subtraction [1,2]. However, this combination would require a huge computational cost to implement the whole system. Although another method may be to use a directional microphone with spectral subtraction [3] or Wiener filtering [4], there would be a limit to the use of simple spectral subtraction in a multi-source environment.

In contrast to the above-mentioned signal processing techniques, a practical slant-plate-type acoustic lens [5], which is used in front of a loudspeaker, has been developed to diffuse high-frequency sound. In addition, it has been reported that the directivity for the types of acoustic lens that are used in front of a conical horn becomes sharper than that when using a parabolic recording device [6].

The authors have proposed a sound recording method by which the arrival time difference of the sound waves coming from a specific position is removed by utilizing an acoustic lens consisting of speaking tubes or horns. As a result of an investigation, it has been confirmed that this method has sharper directivity than that when using conventional super cardioid microphones, and recording with high sensitivity is possible at the position specified when the acoustic lens was designed [7]. This type of acoustic lens is expected to be applicable to speech recognition systems [8] whose positional relationship between the microphone and the sound source is essentially constant [9].

In this paper, an acoustic lens composed of multiple horns and the selective distance effect are introduced. Selective distance recording by means of acoustic lenses using the delayed sum of sound waves is also described. Next, the acoustic lens composed of multiple horns is evaluated in terms of speech intelligibility, and its applicability to a speech recognition system is discussed.

*e-mail: k1830045@edu.cc.uec.ac.jp
†e-mail: yoshinori@ieee.org

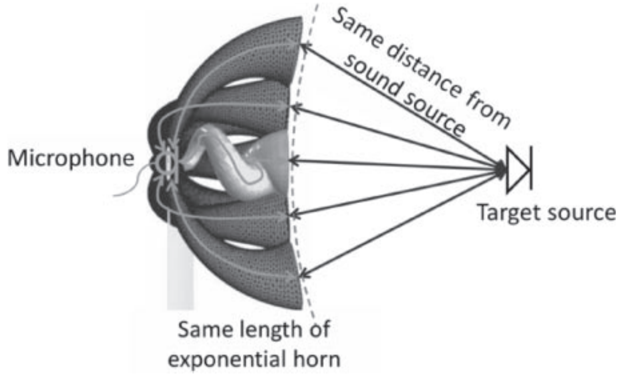

Fig. 1 Structure of an acoustic lens with multiple exponential horns.

2. ACOUSTIC LENS COMPOSED OF MULTIPLE HORNS

By allocating the mouths of horns having the same length at equidistant positions from a sound source and concentrating the throat of horns at a microphone, it is possible to add the sound waves synchronously so that their arrival time differences from the sound source to the microphone can be aligned. This section describes the principle of selective distance recording utilizing an acoustic lens composed of multiple horns and explain the structure of a prototype acoustic lens prepared for this research.

2.1. Selective Distance Recording Utilizing Acoustic Lens Composed of Multiple Exponential Horns

The acoustic lens shown in Fig. 1 has a structure in which the mouths of each horn are positioned on a sphere so that exponential horns with the same length are equidistant from the target sound source. By using this acoustic lens, it is expected that the arrival time difference of wavefronts from the target sound source at the microphone can be removed and the target sound source at the optimum recording position can be emphasized.

2.2. Structure of the Prototype Acoustic Lens

To realize an acoustic lens composed of multiple horns, the structure is described using an actual prototype. In this study, an acoustic lens was prototyped with certain parameters: the horn cutoff frequency f_c of 1 kHz, the horn throat diameter of 5 mm, the angle of the opening wall surface toward the center line of 20°, and the optimal sound recording distance of 0.4 m. The three types of multiple horns of prototype acoustic lens composites are shown as A, B, and C in Fig. 2. The straight horns in each panel are the shape before being bent and are of the same size.

The type "A" horns located at the outermost area of the acoustic lens are bent so that the centerline of the horn becomes

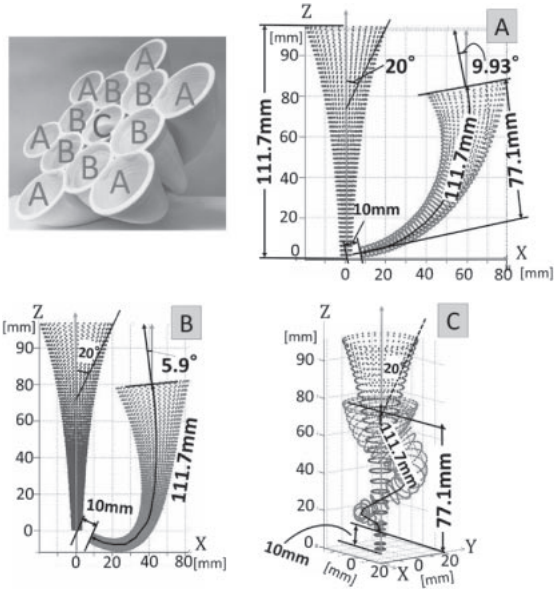

Fig. 2 Shapes of curved exponential horn used in the acoustic lens.

$$X = 72.24\cos\theta + 10, \quad (0 \leq \theta \leq \pi) \tag{1}$$

$$Z = 69.97\sin\theta \tag{2}$$

and the horn rotated by an angle of 9.93° around the Z axis.

Similarly, the type "B" horns are bent so that the centerline of the horn becomes

$$X = 2t(1-t) \times (-49.5) + t^2 \times 49.5 \tag{3}$$

$$Z = 2t(1-t) \times (41.3) + t^2 \times 74.5, \tag{4}$$

where parameter t is given as $0 \leq t \leq 1$. This curve is calculated using a 2nd-order Bezier curve with three control points, $(0, 0)$, $(49.5, -41.3)$, and $(49.5, 74.5)$. The type "B" horns are rotated by an angle of 5.9° around the Z axis.

In the type "C" horn, the centerline is bent into a three-dimensional spiral shape. The coordinates of the centerline, (X, Y, Z), are given as

$$X = f(z)\cos\phi \tag{5}$$

$$Y = f(Z)\sin\phi \tag{6}$$

$$Z = 77.1 \times \phi/(2\pi) + 10, \tag{7}$$

where ϕ is defined as $0 \leq \phi \leq 2\pi$. The function $f(Z)$ follows a curve that rises by 12.9 mm vertically along the Z axis from the throat of the horn, then bulges outward smoothly, and returns to the Z axis at the mouth of the horn. Assuming that the length of the centerline is 111.7 mm, which is the same length as the centerline of the other horns, it is necessary for the centerline not to interfere with the surrounding horns.

To design these horns, the 3D model design method in MATLAB was employed [10]. After outputting an STL file, the horns were allocated using CAD software, as in the

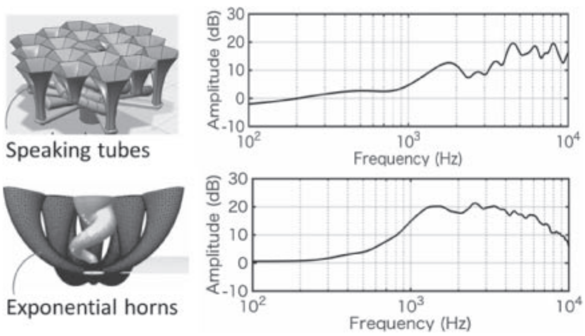

Fig. 3 Amplitude frequency characteristics of acoustic lens with exponential horns and acoustic lens with speaking tubes.

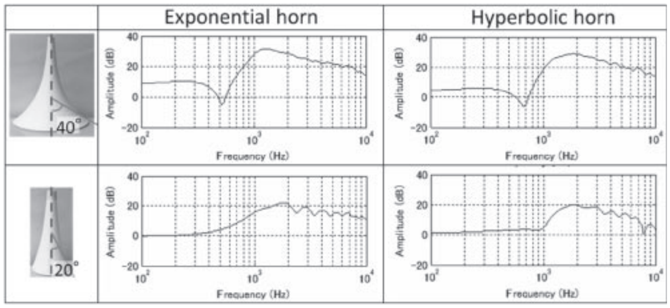

Fig. 4 Amplitude frequency characteristics of exponential horn and hyperbolic horn.

photo shown in Fig. 2, and then manufactured using a 3D printer. The throat of each horn is positioned 10 mm from the origin to enable the installation of a microphone. The thickness of the horn is designed to be 2 mm.

2.3. Horn Shape and Amplitude Frequency Characteristics

The cross-sectional area of the exponential horn expanding along the direction of the x axis is expressed as

$$S = S_0 e^{mx}. \tag{8}$$

This area expands exponentially in accordance with the length x from the initial value S_0 (namely, the cross-sectional area at the throat). In Eq. (8), the parameter m governs the shape of the horn and is given as $m = 4\pi f_c/c$ with the speed of sound c and the cutoff frequency of the horn f_c [11,12]. By designing an acoustic lens with horns, smooth frequency characteristics with less reflection from the open end can be expected. Figure 3 shows the amplitude frequency characteristics of the acoustic lens composed of speaking tubes [7] and the acoustic lens composed of exponential horns proposed in this paper. Because the size and the number of acoustic tubes differ between the two acoustic lenses, the energy of the entire sound waves collected cannot be compared. The total energy of the measured impulse responses is normalized and only the frequency characteristics can be compared. It is confirmed that the amplitude frequency characteristics of the horns are much smoother than those of the tubes because the horns have less resonance. The shape of the acoustic tube is an important determinant of the amplitude frequency characteristics of the acoustic lens. This section describes the difference between the frequency characteristics in accordance with the shape of the horn.

It is known that a hyperbolic horn is better than an exponential horn in terms of the characteristics around the cutoff frequency [11,12]. The cross-sectional area S of a hyperbolic horn starting from the initial area S_0 at the throat and expanding as a hyperbolic function along the x axis direction, can be expressed by

$$S = S_0\left(\cosh mx + \frac{1}{2}\sinh mx\right), \tag{9}$$

where m is a parameter that determines the shape of the horn and is given as $m = 4\pi f_c/c$. Also, the larger the opening angle is in horns of any shape, the larger the horn becomes. However, the reflection and/or the diffraction at the open end is negligible. In this study, the amplitude frequency characteristics of the exponential horn and the hyperbolic horn were evaluated under the condition that the opening angle was 20° and 40° at $f_c = 1$ kHz. Figure 4 shows the amplitude frequency characteristics for each horn. Above the cutoff frequency, the amplitude frequency characteristics become smoother as the opening angle increases. If it is possible to make the horn larger, it would be desirable for the opening angle to be more than 40° [11,12].

According to the result for the exponential horn and the hyperbolic horn, there was no significant difference in the frequency characteristics when the opening angle was 40°. On the other hand, when the opening angle was 20°, the sound pressure level below the cutoff frequency in the exponential horn decreased gently. Therefore, the hyperbolic horn is effective in designing an acoustic lens that intentionally eliminates sound with frequency lower than the cutoff frequency deliberately.

The Bessel horn family (including the exponential horn) is another well-known horn shape. However, the exponential horn has the highest transmission efficiency in the Bessel horn family [11]. Therefore, it was not investigated any other Bessel horns than the exponential horn in this study.

The horns in the acoustic lens proposed in this paper are curved horns. The amplitude frequency characteristics of the curved horns are also discussed in this section. Figure 5 shows the amplitude frequency characteristics for three types of curved horns utilized in the prototype acoustic lens in this study. Above the cutoff frequency, no

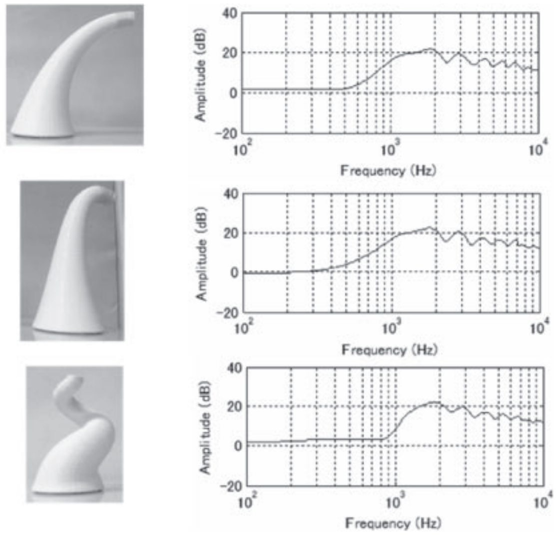

Fig. 5 Amplitude frequency characteristics of bent horn.

significant difference is observed in the frequency characteristics of the straight exponential horn with the opening angle at 20° shown in Fig. 4. For the prototype acoustic lenses in this study, the influence of the horn curve on the sound quality is considered to be minimal. The bending acoustic tubes can be regarded as straight tubes if the radius of curvature is greater than the radius of the tube [13]. When the acoustic lens is miniaturized by bending the horn in a complicated manner, it is necessary to pay attention to the relationship between the radius of curvature and the radius of the tube.

2.4. Wavefront Incident on the Horn

The plane wave is assumed in the theory of sound propagation inside the acoustic horn. Therefore, this section describes whether sound waves from point sources can be considered plane waves if they are collected at a distance r from the source.

The ratio of the sound pressure p to the particle velocity v for a spherical wave at a distance r from a point source is expressed as

$$\frac{p}{v} = \left(\frac{k^2r^2}{k^2r^2+1} + \mathrm{j}\,\frac{kr}{k^2r^2+1}\right)z_0 \tag{10}$$

using the specific acoustic impedance z_0 and wavenumber k [14]. Therefore, since the phase difference between the sound pressure p and the particle velocity v is

$$\phi = \arctan\frac{1}{kr}, \tag{11}$$

the time delay of the particle velocity waveform from the sound pressure waveform is

$$\phi = \arctan\frac{1}{kr}, \tag{12}$$

using the sound speed c. In the frequency band recorded by the horn used for this acoustic lens, the time difference between the particle velocity waveform and the sound pressure waveform is less than one sample under the assumption that the sampling interval is $T = 1/44{,}100\,\mathrm{s}$, source distance is $r = 0.4\,\mathrm{m}$, the sound speed is $c = 340\,\mathrm{m/s}$ and the cutoff frequency of the horn is $f = 1\,\mathrm{kHz}$.

Furthermore, from the ratio of the sound pressure p to the particle velocity v for the spherical wave in Eq. (10), the ratio in effective value of $\bar{p}/\bar{v}$ is given as

$$\frac{\bar{p}}{\bar{v}} = \frac{z_0}{\sqrt{1+\dfrac{1}{k^2r^2}}}. \tag{13}$$

Equation (13) can be expressed as

$$z_0\,\frac{\bar{v}}{\bar{p}} = \sqrt{1+\frac{1}{k^2r^2}}. \tag{14}$$

If the distance from the sound source is sufficiently large in contradiction to the wavelength, that is, $\bar{p}/\bar{v} \approx z_0$, will be considered a plane wave. Here, when representing $\alpha = \{1 + 1/(k^2r^2)\}^{1/2}$, $\varepsilon = (1-\alpha)\times 100\%$ indicates the relative error from $\bar{p}/\bar{v} = z_0$. In the frequency band of the horn used in this work, the relative error is less than 1% and can be regarded as $\bar{p}/\bar{v} \approx z_0$ at the sound source distance $r = 0.4\,\mathrm{m}$ (e.g., $\varepsilon = 0.9\%$ for $f = 1\,\mathrm{kHz}$, $\varepsilon = 0.06\%$ for $f = 4\,\mathrm{kHz}$). If the distance from the sound source is sufficiently far from the wavelength, that is, $\bar{p}/\bar{v} \approx z_0$, the wavefront will be considered a plane wave.

From the above considerations, it can be concluded that the sound wave entering the horn can be regarded as a plane wave in this work.

3. DIRECTIVITY AND DISTANCE SELECTIVITY FOR ACOUSTIC LENS WITH MULTIPLE HORNS

In this section, the measurement results of directivity and distance selectivity for the prototype acoustic lens are described. The characteristics are calculated from impulse responses measured in a sound-absorbing room using TSP signals. The impulse responses are cut out at the time 2.94 ms, which should be sufficiently shorter than the time taken for reflection from the wall to return.

3.1. Directional Characteristics of the Acoustic Lens

An acoustic lens was attached to an omnidirectional microphone and the impulse responses were measured every 30° while changing the sound source direction θ from 0 to 180°, as shown in Fig. 6. The sound source distance was set as 0.4 m since this is the sound recording distance of the acoustic lens. Figure 7 shows the directional characteristics of the prototype acoustic lens. The symbols are the measured values, and the solid lines are the spline

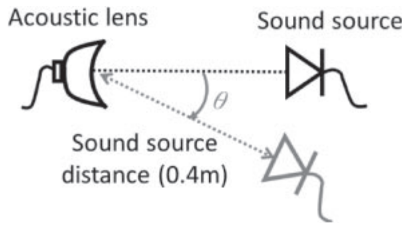

Fig. 6 Conditions for directivity measurement in the acoustic lens with multiple exponential horns.

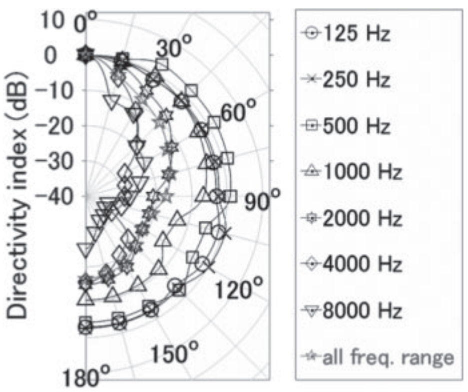

Fig. 7 Directivity in the acoustic lens with multiple exponential horns.

function. A very strong directivity can be confirmed at frequencies higher than the cutoff frequency of 1 kHz.

3.2. Distance Selectivity of the Acoustic Lens

This section describes theoretical and measured distance selectivity values of an acoustic lens.

It is assumed that there is a sound source f in front of the acoustic lens. Designating the distance from the horn at the center of the acoustic lens (horn "C" in Fig. 2) to the sound source as r_c, the distances from the sound source to horn "B" and horn "A" are expressed as

$$r_a = \sqrt{r_c + 4(L^2 - r_c L)\sin^2\frac{\theta_a}{2}} \tag{15}$$

$$r_b = \sqrt{r_c + 4(L^2 - r_c L)\sin^2\frac{\theta_b}{2}}, \tag{16}$$

respectively, where L is the optimal sound recording distance. For the prototype acoustic lens in this study, $L = 0.4$ m. From Fig. 2, $\theta_a = 9.93^\circ$ and $\theta_b = 5.9^\circ$. It is assumed that when a sound source generates a sine wave $s(t) = \sin(2\pi ft)$ of f Hz, the sound waves input to horns "A," "B" and "C" are expressed as

$$s_a(t) = \frac{\mathfrak{R}}{r_a}\sin(2\pi f(t - r_a/c)) \tag{17}$$

$$s_b(t) = \frac{\mathfrak{R}}{r_b}\sin(2\pi f(t - r_b/c)) \tag{18}$$

$$s_c(t) = \frac{1}{r_c}\sin(2\pi f(t - r_c/c)) \tag{19}$$

using the speed of sound c and the directivity factor $\mathfrak{R}$ of the horn. The acoustic lens prototyped in this study has six horns each of "A" and "B" and one horn "C," so the recorded sound wave is represented by

$$s_{sum}(t) = 6(s_a(t) + s_b(t)) + s_c(t). \tag{20}$$

The directivity factor $\mathfrak{R}$ of the horn can be approximated by that of a tablet sound source with the same area as the horn mouth [15]. The directivity factor $\mathfrak{R}$ of the horn is written as

$$\mathfrak{R} = \frac{2J_1(\chi)}{\chi}, \quad \chi = \frac{2\pi a}{\lambda}\sin\phi. \tag{21}$$

In this equation, it is considered that the angle ϕ between the horn mouth normal and sound source direction,

$$\phi = \arccos\frac{L^2 + r^2 - (L - r_c)^2}{2Lr}, \tag{22}$$

where r is r_a or r_b, λ is the wavelength, and $J_1(\ )$ denotes the Bessel function of the first kind.

Next, the measurement distance selectivity of the acoustic lens is described. As shown in Fig. 8, the impulse responses were measured by setting the acoustic lens in front of the loudspeaker and changing the sound source distance R from 0.1 to 0.8 m in steps of 0.05 m in an anechoic room. Figure 9 shows the distance selectivity of the acoustic lens. The symbols indicate the measured values, and the solid lines are the theoretical values based on simulation using Eq. (20). However, the distance selectivity was not clearly observed in the result because the attenuation of sound pressure was dominant owing to the distance.

Accordingly, the results of distance selectivity were normalized by the theoretical value of the sound pressure level based on the distance attenuation. The theoretical value of the sound pressure level E at the sound source distance R is expressed as

$$E = E_0 - 20\log_{10}\frac{R}{R_0}, \tag{23}$$

where E_0 is the sound pressure level at the sound source distance R_0. Figure 10 shows the normalized distance

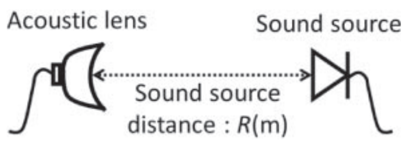

Fig. 8 Conditions for distance selectivity in the acoustic lens with multiple exponential horns.

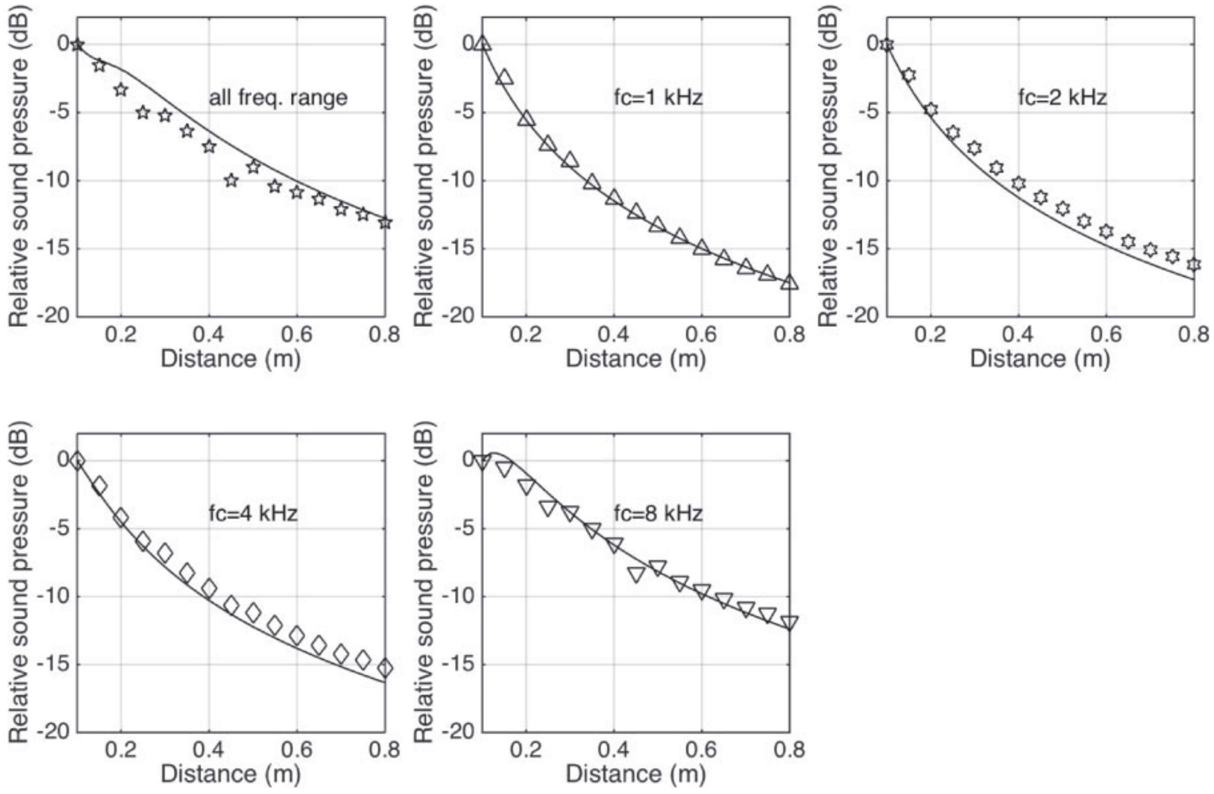

Fig. 9 Distance selectivity in the acoustic lens with multiple exponential horns.

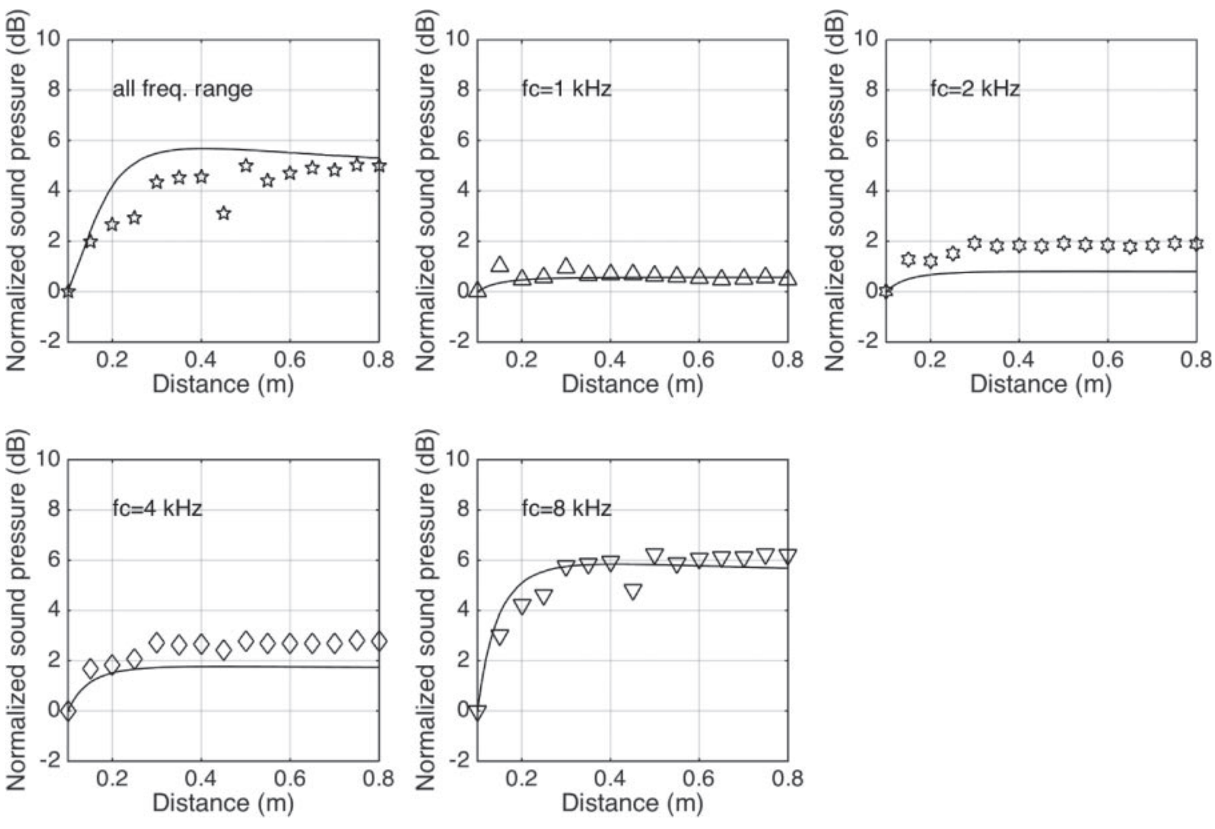

Fig. 10 Distance selectivity of the acoustic lens normalized by the theoretical attenuation in accordance with the sound source distance.

selectivity of the acoustic lens. It can be confirmed that the measurement values are in close around with the simulated trends. This result means that the increase in the sound pressure level shown in Fig. 10 can be expected in the case of recording using the acoustic lens but not the general omnidirectional microphone. It is found that the sound pressure level over the whole band increases by about 5 dB in the vicinity of 0.4 m. It can also be confirmed that the sound pressure level in the frequency bands higher than the cutoff, even in a narrow band, increases uniformly. This should be the result of an effect of the acoustic lens since the optimal sound recording distance was designed to be 0.4 m.

Although the curve of the theoretical value form the simulation has a maximum around $R = 0.4$ m, this is difficult to confirm from Fig. 10. These results suggest that although an acoustic lens was prototyped with an optimum sound source distance of $R = 0.4$ m, no marked distance selectivity effect was obtained. Therefore, some simulations of the distance selectivity were attempted for an acoustic lens with different optimum sound source distance (SSD) conditions. Figure 11 shows the simulation results of the distance selectivity law for the octave bands from 1 to 16 kHz and the entire frequency range. However, the results of distance selectivity were normalized by the theoretical value of the sound pressure level based on the distance attenuation, and the horizontal axis is shown on a logarithmic scale. Figure 11 shows that for an acoustic lens made with an SSD of 0.15 m, the distance selectivity is attenuated by about 3 dB when the acoustic lens is 0.15 m behind the optimal recording position, and for an acoustic lens made with an SSD of 0.1 m, the attenuation is 4 dB in the entire frequency range. The simulation results show that the smaller the SSD, the greater the distance selectivity effect that can be expected. As an example, in the results for all frequency bands, an acoustic lens with SSD = 0.15 m achieves damping of about 3 dB at a distance of 0.15 m from the optimum recording position, and an acoustic lens with SSD = 0.1 m achieves damping of about 4 dB.

4. SPEECH INTELLIGIBILITY OF THE ACOUSTIC LENS

This section describes the estimation of speech intelligibility, defined by the Speech Transmission Index (STI), in the case of using the acoustic lens. Impulse responses were measured in an irregularly shaped reverberation room with a reverberation time of 6.0 s while the distance between the sound source and the acoustic lens was changed from 0.1 to 1.6 m. When estimating the STI from reverberant impulse responses, the modulation transfer function (MTF) for each frequency band is typically

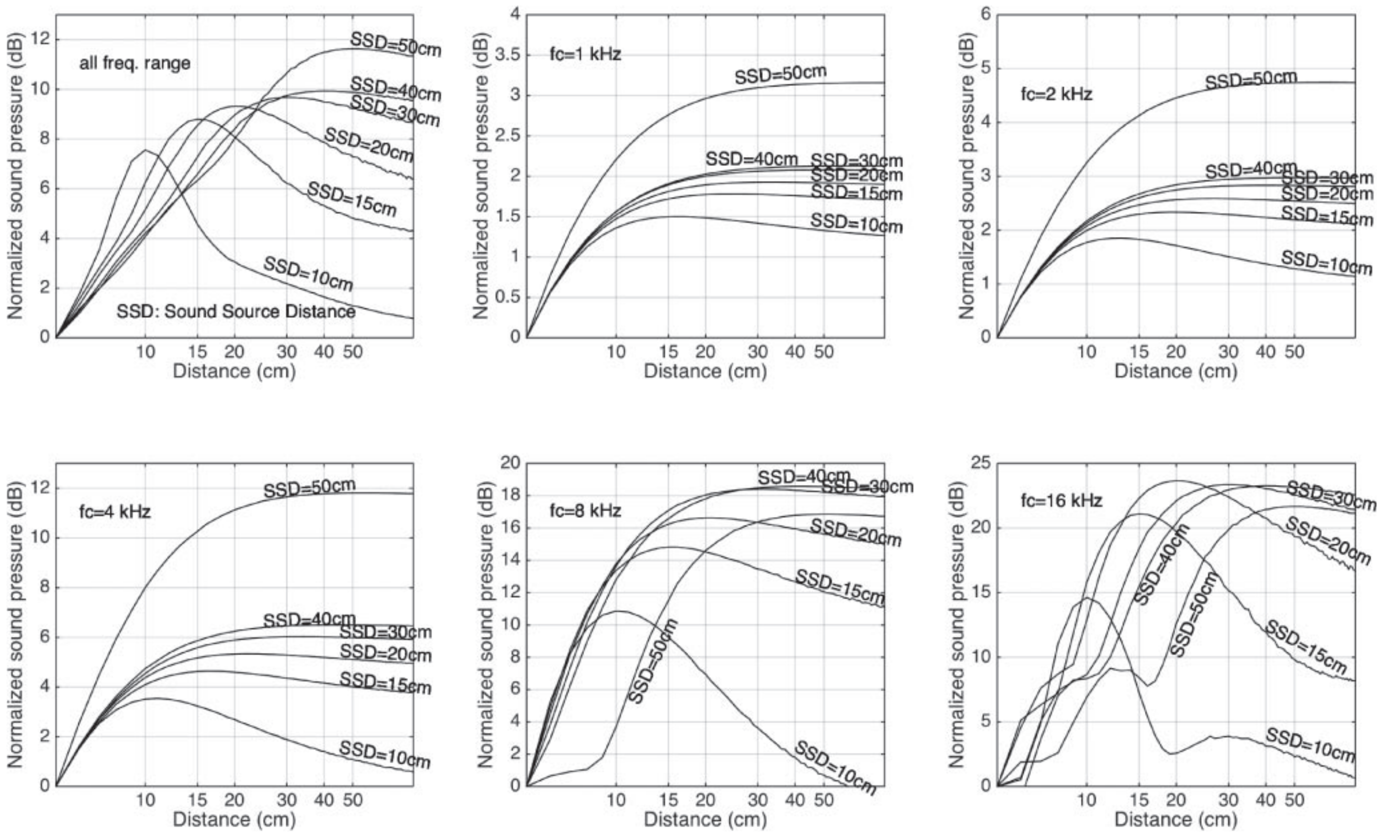

Fig. 11 Simulation results of distance selectivity for the acoustic lens with different sound collection distances.

Table 1 Octave band weighting factors of modulation transfer function used to obtain the speech transmission index.

Oct. band (Hz)	128	250	500	1 k	2 k	4 k	8 k
Weights for general MTF	0.129	0.143	0.114	0.114	0.186	0.171	0.143
Weights for over 1 kHz MTF	—	—	—	0.190	0.300	0.280	0.230

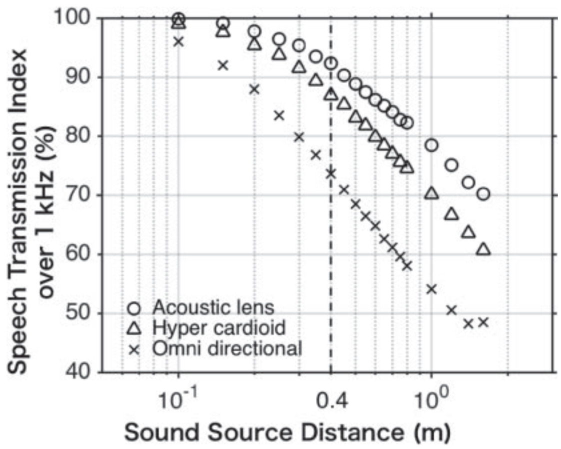

Fig. 12 Calculation result of speech transmission index for each distance.

calculated from the impulse response and then weighted for the contribution rate of each frequency band to speech intelligibility [16]. For this process, the frequency bands from 128 Hz to 8 kHz were used as the octave frequency band. However, since the cutoff frequency of the acoustic lens was 1 kHz, MTFs were calculated for four frequency bands between 1 and 8 kHz and the weighted values are shown in Table 1. The sum of the weights over the four frequency bands is normalized to be 1. The same measurements were made for the omnidirectional microphone and the hypercardioid microphone (RODE, NTG-3) for comparison.

Figure 12 shows the results of STI estimation for each sound source distance. When using the omnidirectional microphone, STI decreases with increasing the sound source distance. On the other hand, when using the acoustic lens, STI values of 90% or more were obtained up to 0.4 m, which is defined as the optimal distance of this acoustic lens. At 0.4 m, STI is 5.5% higher than that of the hypercardioid microphone. In speech recognition under a noisy or reverberating condition, it is well known that consonants, which have lower energy than vowels, are particularly masked [17]. If speech intelligibility at frequencies higher than 1 kHz can be improved, the proposed acoustic lens will be applicable to a speech recognition system. Considering the results shown in Fig. 11, it is expected that an acoustic lens with an optimal sound source distance of 0.1 to 0.15 m is more practical for speech recognition.

5. CONCLUSION

This paper discussed the use of acoustic lens for emphasizing a sound source at a specific position. In the acoustic lens with horns, a gain of about 5 dB was obtained at a desired sound recording distance, from which it can be concluded that the distance selectivity was confirmed. Furthermore, it was found that a high STI was maintained up to the recording distance in the evaluation of speech intelligibility. Therefore, an acoustic lens with horns is promising for application to a speech recognition system in which the sound source distance is relatively constant.

In this study, an acoustic lens with a sound recording distance of 0.4 m was manufactured to consider its application to speech recognition. It is also possible to create acoustic lenses with different recording distances using the design method described in this paper. However, when the sound recording distance is long, it is desirable to increase the diameter of the acoustic lens and to ensure the delay of the sound wave among horns. Therefore, effective distance selectivity can be expected by shortening the optimum sound source distance.

With the acoustic lens prototyped in this study, the desired sound recording distance was determined at the time of acoustic lens design. To make the recording distance variable, it is necessary to make the distance between the sound source position and the mouth of each horn changeable.

In this trial manufacture, a fused-deposition-modeling 3D printer and an ABS-filament was used to hold down the production cost. However, there is room for discussion on the difference in acoustic characteristics owing to the difference in materials used and the output method of manufacturing.

ACKNOWLEDGMENT

This work was supported by a grant from Ono Acoustics Research Fund.

REFERENCES

[1] Y. Kaneda, "Directivity characteristics of adaptive microphone-array for noise reduction (AMNOR)," *J. Acoust. Soc. Jpn. (E)*, **12**, 179–187 (1991).

[2] Y. Haneda, K. Furuya, S. Koyama and K. Niwa, "Close-talking microphone arrays based on spherical harmonic expansion," *IEICE Trans. Fundam.*, **J97-A**, 264–273 (2014) (in Japanese with English figures).

[3] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust. Speech Signal Process.*, **ASSP-27**, 113–120 (1979).

[4] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications* (The MIT Press, Cambridge, Mass., 1949).

[5] W. E. Kock and F. K. Harvey, "Refracting sound waves," *J. Acoust. Soc. Am.*, **21**, 471–481 (1949).

[6] M. A. Clark, "An acoustic lens as a directional microphone," *J. Acoust. Soc. Am.*, **25**, 1152–1153 (1953).

[7] Y. Takahashi, M. Kobayashi and J. Tanaka, "Slective distant recording based on delay-and-su, beamforming using acoustic lens," *J. Acoust. Soc. Jpn.* (*J*), **72**, 761–763 (2016) (in Japanese with English figures).

[8] M. Nishimori, T. Saitoh and R. Konishi, "Voice controlled intelligent wheelchair," *Proc. SICE Annu. Conf.*, pp. 336–340 (2007).

[9] M. Kobayashi and Y. Takahashi, "Acoustic lens with exponential horns for selective distant recording," *5th Jt. Meet. Acoust. Soc. Am. and Acoust. Soc. Jpn.*, Vol. 140, No. 4, p. 3033, Nov. (2016).

[10] Y. Takahashi, "Investigation of subject related to acoustic engineering introducing 3D model programing," *Education in Acoustics Technical Report*, Vol. 9, No. 1, EDU-2017-2, pp. 9–15 (2017) (in Japanese with English figures).

[11] T. Hayasaka and S. Yoshikawa, *Onkyo Shindo Ron* (Maruzen, Tokyo, 1974), pp. 665–680 (in Japanese).

[12] H. Olson, *Acoustical Engineering* (Van Nostrand, 1957), pp. 100–112.

[13] D. Keefe and A. Benade, "Wave propagation in strongly curved ducts," *J. Acoust. Soc. Am.*, **74**, 320–332 (1983).

[14] M. Tohyama, *Sound and Signals* (Springer, Berlin, 2011), p. 148.

[15] T. Hayasaka and S. Yoshikawa, *Onkyo Shindo Ron* (Maruzen, Tokyo, 1974), p. 585, Eq. 18.64 (in Japanese).

[16] T. Houtgast and H. J. M. Steeneken, "Evaluation of speech transmission channels by using artificial signals," *Acustica*, **25**, 355–367 (1971).

[17] V. O. Knudsen, "The hearing of speech in auditoriums," *J. Acoust. Soc. Am.*, **1**, 56–82 (1929).

Maho Kobayashi graduated in Advanced Course of Tokyo Metropolitan College of Industrial Technology in 2018. She received a M.E. degree from University of Electro-Communications in 2020. Currently she is an engineer of Canon Inc. She is a member of Acoustic Society of Japan.

Yoshinori Takahashi graduated in Engineering from Kogakuin University in 2001 and completed his master's course in 2003. He received a Doc.Sc. degree from Waseda University in 2006. He was an associate professor at Tokyo Metropolitan College of Industrial Technology from 2009 to 2019. Currently he is an associate professor in the Kogakuin University Faculty of Informatics Department of Information Design. He is a member of Acoustic Society of Japan, the Institute of Electronics, Information and Communication Engineers, the Institute of Electrical and Electronics Engineering and an associate member of the Audio Engineering Society and Acoustic Society of America.