Hat-type hearing aid system with flexible sound directivity pattern

Takayuki Inoue1,* Yusuke Ikeda2,† and Yasuhiro Oikawa1,‡

1Department of Intermedia Art and Science, Waseda University, 3–4–1 Okubo, Shinjuku-ku, Tokyo, 169–8555 Japan
2Department of Information Systems and Multimedia Design, Tokyo Denki University, 5 Senju-asahi-cho, Adachi-ku, Tokyo, 120–8551 Japan

(Received 2 February 2017, Accepted for publication 28 July 2017)

Abstract: For speech communication, it is useful to emphasise the sound coming from a target direction. A large-scale microphone array enables a flexible design of the directivity pattern. However, it is difficult to implement a large-sale microphone array onto an ear-type hearing aid system. In this paper, we propose a hat-type hearing aid system composed of 48 MEMS microphones that can increase the flexibility of the sound directivity pattern. We found that it is possible to reduce the number of microphones to be used by considering the hat effect, without affecting the system performance. By compensating the time delays and reducing the number of microphones based on impulse response measurements, the directivity range of the sound directivity pattern was improved by 5 dB. In addition, we conducted a qualitative evaluation that resulted in an intelligibility improvement when using our hearing aid system.

Keywords: Delay-and-sum beamforming, Single-bit signal processing, MEMS microphone array, FPGA

PACS number: 43.72.+q [doi:10.1250/ast.39.22]

1. INTRODUCTION

We have the ability to distinguish a target voice in an environment where many people are talking, in a situation commonly called the cocktail party effect. It is useful for speech communication aids to improve the ability to discriminate the target among different sound sources. Methods such as beamforming [1], source separation [2], and reflector-based technique [3] have been proposed for this purpose, resulting in improved speech intelligibility.

In the 1990s, digital hearing aids were developed, and several signal processing techniques have been applied to them, such as single-microphone speech enhancement [4], multiple microphone, and adaptive processing [5]. Recently, more advanced techniques for hearing aids have been studied and developed worldwide [6–8].

In a real environment, however, the user’s surrounding conditions are complex; hence, it is required for a hearing aid system to have high performance, low-latency processing, and other characteristics to guarantee adequate operation.

Many types of hearing aids have been proposed, including Behind-The-Ear (BTE), In-The-Ear (ITE), eyeglass type, and hair-band type hearing aids. Both BTE and ITE hearing aids are mainly used now, but they commonly use only a few microphones because of the limitations in the hardware scale and available mounting space. However, if a large-scale microphone array could be installed in these hearing aids, an increased flexibility regarding sound localization could be achieved.

In this paper, we propose a hat-type hearing aid system using a MEMS microphone array with high-speed single-bit signals. To use several microphones, the electric circuits must be simple and small, and the system must provide enough space to mount the microphone array. To achieve this, we used MEMS microphones with single-bit signal output, a field-programmable gate array (FPGA) system, and a hat [9]. To verify the diffraction effect produced by the hat and user’s head, impulse responses were measured in an anechoic chamber. Based on the measured impulse responses, sound directivity patterns were achieved with compensation of time delays and a reduction in the number of microphones used for beamforming.

*e-mail: wsd-t.ino@asagi.waseda.jp
†e-mail: yusuke.ikeda@mail.dendai.ac.jp
‡e-mail: yoikawa@waseda.jp
2. HAT-TYPE HEARING AID SYSTEM

2.1. System Configuration

Figures 1 and 2 show the diagram and physical structure of our hat-type hearing aid system, respectively. The microphone array is composed of 48 MEMS microphones (SPM0405HD4H, Knowles Acoustics), and each microphone outputs a single-bit digital signal with a 3.125 MHz sampling frequency. To change the direction of the sound directivity pattern in a range of 360 degrees, the microphones were placed on the same plane above the hat brim. The microphones were placed at 1.3 cm intervals along an elliptical in the horizontal plane of the hat, because the MEMS microphones were placed equidistant on a flexible substrate which was then rolled into an ellipse and placed on the hat.

The digital circuit was placed inside the hat in the upper part without interfering with the user’s head. An FPGA (EP4CE15F17C8N, Altera) gets the input signal from each microphone and returns a mixture signal by using the delay-and-sum beamforming technique with high-speed single-bit signal processing. The beamforming technique is described in Sect. 2.2. The system can control the time delays for the beamforming with high accuracy and without interpolation due to the very high sampling frequency of the single-bit signal. The resulting mixture signal is converted from a digital to an analogue signal by a low-pass filter and amplified for hearing.

In our system, the user can arbitrarily change the sound directivity pattern to emphasise the speech from a target source. In order to easily change the sound directivity pattern, the user can modify the settings of the time delays and microphones to be used by a 6-bit control signal, which is sent from a microcontroller board (Arduino Fio) to the FPGA. The microcontroller board is controlled by a smartphone application via a wireless module (XBee Wi-Fi). Thus, we developed an Android application to select the beam direction in real time. Only by tapping the display, the user can easily change the beam direction depending on the direction the user wants to listen to.

Our system is operated using a 9 V battery. This battery is small enough to carry on or wear with the system. Overall, the weight of the proposed system including the hat is 500 g.

2.2. Delay-and-sum Beamforming with Single-bit Signal Processing

In our system, the delay-and-sum beamforming technique is used to change the sound directivity pattern. In addition, to simplify the system configuration, the beamforming is achieved through single-bit signal processing.

The beamforming technique using high-speed single-bit signals has been studied in various works [10–12]. Takeoka et al. show that the delay-and-sum beamforming technique using high-speed single-bit signals helps to achieve a high spatial resolution of sound [12]. In these researches, a single-bit signal is converted into a multi-bit signal during the beamforming processing because the sum of delayed single-bit signals results in a multi-bit signal. Then, the multi-bit signal is converted again into a single-bit signal using ΔΣ modulation [13,14]. If the beamforming processing without multi-bit conversion regarding single-bit signals is performed, it is expected that various directivities can be achieved only by reordering the samples of the single-bit signals.

A conventional delay-and-sum beamforming technique and its signals flow are illustrated in Fig. 3. This technique can emphasise a sound coming from a target direction. The incident plane wave is measured by a microphone array composed of $M$ microphones. Signal $x_m(t)$ measured by the $m$-th microphone was delayed to be synchronized. The sum of the delayed signals, represented by $y(t)$, is the emphasised sound from the target direction [15]. Therefore,

$$y(t) = \frac{1}{M} \sum_{m} x_m(t - \tau_m),$$  

(1)

where $x_m(t)$ is the microphone signal and $\tau_m$ is the time delay in the $m$-th microphone.
In the conventional delay-and-sum beamforming technique, to delay the signal with higher accuracy than the sampling rate, it is required to interpolate the original signal. However, in high-speed single-bit signal processing, the sampling rate is high enough to achieve the delay-and-sum beamforming without interpolation. Figure 4 shows the delay-and-sum beamforming for an array of single-bit signals. In this method, the beamforming can be realized by simply reordering the samples of the single-bit signals.

The number of samples that should be delayed is calculated as follows:

\[ k_m = \left\lceil \frac{r_m \cdot f_s + 0.5}{f_s} \right\rceil \]

where \( f_s \) is the sampling frequency, which is equal to the clock frequency of the MEMS microphones. The corresponding signal delayed by the FPGA is represented by \( x_m[i - k_m] \).

Figure 5 shows the timing diagram of the two clock signals and the mixture data signal for the delay-and-sum beamforming using single-bit signals [16]. The clock signal for the mixture data is generated by multiplying the original sampling frequency \( f_s \) by the number of microphones \( M \). The microphone signals are mixed by sequentially outputting each microphone signal at the rising edge of the clock signal for the mixture data. When the next samples are output from the MEMS microphone array to the FPGA at the rising edge of the microphone clock signal, the signal of the first microphone is outputted from the FPGA. Thus, the delay-and-sum beamforming is achieved only by ordering the single-bit signals with a high-speed clock signal for the mixture.

3. IMPULSE RESPONSE MEASUREMENT

To design various patterns of sound directivity, it is necessary to measure the acoustical characteristics of the microphones as well as the sound diffraction and reflection caused by the user’s head and the hat. Therefore, impulse responses for each microphone of the proposed system were measured in an anechoic chamber. The measurement setup is shown in Fig. 6, and the experimental conditions are listed in Table 1. The output signal of each microphone was recorded, and the microphone array was rotated in 3-degree intervals. The microphones were numbered anticlockwise starting from the front. The loudspeaker was located at 0.5 m from the centre of the microphone array. As shown in Fig. 1, the output from the FPGA was converted to an analogue signal by low-pass filtering, then amplified to make the output sound. In the measurement of impulse responses, the analogue signal was converted to a multi-bit signal (i.e., digitized) using the audio interface listed in Table 1. An optimized Aoshima’s time-stretched...
pulse (OATSP) signal was used as the source signal for measurements [17]. As shown in Fig. 7, the measurements were conducted under three conditions: when the microphone array is not attached to anything or free condition (Fig. 7(a)), attached to the dummy head (Fig. 7(b)), and attached to the hat on the dummy head (Fig. 7(c)).

Table 1 Measurement conditions.

| Location          | Anechoic chamber, Waseda Univ. |
|-------------------|--------------------------------|
| Measurement signal| OATSP                          |
| Sampling rate [kHz]| 44.1                           |
| Loudspeaker       | Yamaha MSP7 Studio             |
| Height of sp. and mic. [m] | 0.8                     |
| Background noise [dB] | 40                       |
| Audio interface   | MOTU UltraLite-mk4             |
| Dummy head        | KEMAR                          |

Figure 8 shows the impulse response for each microphone when the loudspeaker was located in front of the microphone array. In the condition with neither the hat nor the dummy head (Fig. 8(a)), the loudspeaker sound clearly reached all the microphones. However, in the two other conditions (Figs. 8(b) and 8(c)), the loudspeaker sound is attenuated and the rear microphones capture mainly the diffracted sound.

Figure 9 shows the frequency characteristics of the corresponding impulse responses shown in Fig. 8. As these figures (b) and (c) show, the sound was notably reduced at the microphones on the side opposite to the loudspeaker, especially for frequencies of 10 kHz and above. At these rear microphones, the contribution of the signal at approximately 10 kHz and above was small for controlling the sound directivity pattern. Therefore, to calculate the time delays considering the reflection and diffraction
caused by the user’s head and the hat, it is required to set accurate time delays and select only a part of the microphone array for the delay-and-sum beamforming.

4. INFLUENCE OF USER’S HEAD AND HAT ON DIRECTIVITY

To verify the effect of the user’s head and the hat on the delay-and-sum beamforming performance for the proposed system, the sound directivity patterns were calculated using a convolution of impulse responses and white noise. Figure 10 shows the microphone array and a virtual sound source to calculate the time delays. The microphones are elliptically arranged on the horizontal plane of the hat, with the minor and major axes of the ellipse being 0.21 m and 0.23 m, respectively. We assumed that the virtual sound source is located at 0.5 m from the centre of the microphone array. To obtain the time delays for the beamforming, the distance between the virtual sound source and each microphone was calculated. Then, the delay-and-sum beamforming signal was obtained by using the time delays of all the microphones. We also calculated the sound directivity pattern after moving the sound source to a 45-degree incidence angle.

The obtained sound directivity patterns are shown in Figs. 11 and 12, when the directions of maximum sensitivity are 0 and 45 degrees, respectively. The corresponding full width at half maximum (FWHM) and directivity range are shown in Table 2. The directivity range is defined as the difference between the maximum and minimum values in the directivity pattern. Smaller values of FWHM correspond to sharper system directivity. Furthermore, higher values of directivity range result in reduced sound interference from sources other than the target direction. Therefore, a smaller FWHM and larger directivity range are desirable in order to better emphasise the sound signal from the target direction. The FWHM is affected by the user’s head and the hat by only 1 degree because the sharpness of the sound directivity pattern is determined by the length of the microphone array. However, when using the hat, the directivity range reduces by about 3 dB, showing a notable influence on this parameter.

![Virtual sound source and MEMS microphone array](image1)

**Fig. 10** Microphone array and virtual sound source to calculate time delays.

![Sound directivity patterns](image2)

**Fig. 11** Sound directivity patterns when the maximum sensitivity direction is 0 degrees.

![Sound directivity patterns](image3)

**Fig. 12** Sound directivity patterns when the maximum sensitivity direction is 45 degrees.

| Directivity pattern | Condition     | FWHM [deg.] | Directivity range [dB] |
|---------------------|---------------|-------------|------------------------|
| 0 degrees           | Head and hat  | 9           | 11.3                   |
|                     | Head          | 9           | 12.6                   |
|                     | Free          | 10          | 13.9                   |
| 45 degrees          | Head and hat  | 19          | 9.2                    |
|                     | Head          | 19          | 11.5                   |
|                     | Free          | 18          | 12.4                   |
In the case of the 45-degree loudspeaker rotation, the FWHM increases and the directivity range reduces, because the number of effective microphones to control sound directivity is reduced due to the shape of the hat. Therefore, it is important to suitably select the microphones to design the sound directivity pattern in the proposed system.

5. DELAY-AND-SUM BEAMFORMING FOR THE HAT-TYPE HEARING AID SYSTEM

5.1. Time Delays Based on Measured Impulse Responses

From the results in Sect. 3, the microphones opposite to the target sound source are expected to have only a small contribution to emphasising the target signal. When the effects of noise from sources other than the target are considered, the use of all the microphones does not necessarily lead to an improved sound directivity performance. In this section, we present experiments conducted to investigate the effects on the sound directivity when reducing the number of microphones for the delay-and-sum beamforming.

To improve the sound directivity pattern of the microphone array on the hat, the time delays for the beamforming were determined, and the microphones to be used were selected by analysing the measured impulse responses. These responses were used to calculate the time delays for the beamforming. The time delays were estimated by comparing the peaks of cross-correlation functions between the impulse responses measured at the front microphone and the other microphones.

For comparison, the time delays were calculated in two conditions: using and not using the hat. Figure 13(a) shows the comparison between the calculated time delays using the measured impulse responses and by a simulation, when the maximum sensitivity direction is 0 degrees. The simulation was conducted assuming an elliptical shape of the microphone array in the free condition, as shown in Sect. 4. Figure 14(a) shows a similar comparison of the time delays when the maximum sensitivity direction is 45 degrees. The microphone with the highest time delay of sound arrival was selected as the reference to calculate the time delays. Thus, the arrival time of the microphone placed opposite to the sound source was considered as the time delay of 0 ms.

As Fig. 13(a) shows, the calculated time delays using the impulse responses in the free condition are almost equal to those obtained from the simulation. The difference of the
time delays between these two conditions may be caused by the sound reflections of the electrical board and the stand of the microphone array. The highest difference between the time delays calculated from the impulse responses when using and not using the hat is approximately 0.2 ms, because the arrival times are increased by the sound diffraction due to the hat.

5.2. Reducing the Number of Microphones for the Delay-and-sum Beamforming

The FWHM and directivity range were calculated when using from 1 to 48 microphones. When a single microphone was used, the output signal was calculated without beamforming. The microphones were selected in ascending order according to their sound arrival times. The relationship between the number of microphones and both the FWHM and directivity range are shown in Figs. 13(b) and 14(b), for maximum sensitivity directions of 0 and 45 degrees, respectively. The FWHM values converge to approximately 10 degrees when 12 or more microphones are used, regardless the presence of the hat. The directivity range decreased when the number of microphones was 25 or more. Thus, it is not necessary (in fact harmful) to use microphones on the opposite side to the target direction for controlling the sound directivity pattern. In addition, when using the hat, the equivalent directivity range and FWHM can be obtained with fewer microphones than when not using the hat, because it attenuates the sounds received on the opposite side. As shown in Fig. 9(c), the received sound signals above 10 kHz are attenuated at the 24 microphones on the opposite side to the sound source. Thus, the target signal could not be further emphasised even when the delay-and-sum beamforming was applied. Moreover, the directivity range declined and noise accumulated. Consequently, it is possible to save energy and reduce the computational cost by using less microphones.

Figures 13(c) and 14(c) show the comparison between the sound directivity patterns when using 24 and 48 microphones. The time delays were obtained by using the measured impulse responses. The directivity range when using 24 microphones is improved by approximately 5 dB compared to that when using 48 microphones.

5.3. Qualitative Evaluation

The main purpose of our system is to improve speech intelligibility in a noisy environment. Thus, with the use of our system, a preliminary qualitative evaluation of word intelligibility was conducted.

Figure 15 shows the setup for evaluation. A loudspeaker (MS101 III, Yamaha) for playing speech signals was installed at 0.5 m from the participant’s head, located in the centre of the experimental environment. The other 4 loudspeakers, which played pink noises, were installed every 90 degrees at 0.75 m from the centre, and at 45 degrees from the position of the speech loudspeaker. The equivalent noise levels of the speech signal and pink noise at the centre were 64 dB and 70 dB, respectively. This experiment was conducted in a seminar room, Bldg.59, Waseda University. The background noise level was 30 dB. The speech signal data were taken from the familiarity-controlled word lists (FW03) [18]. The range of familiarity for the used signals was from 5.5 to 7.0. The speech stimuli were 50 words uttered by a Japanese female, which varied in accordance with the experimental conditions. Two versions of the qualitative evaluations were conducted to draw a comparison based on whether our hat-type hearing aid system is used or not. Based on the results in Sect. 5.2, the system used only 24 ch microphones located at the front. After each stimulus was played, the participants reported the perceived word by writing it down. The experiments were conducted with four participants. The number of trials for each participant was 100.

Figure 16 shows the correct answers rate for each participant. With our system, an improvement of the correctness rate by approximately 20% was observed in 3
out of the 4 participants. It is possible that some issues when wearing our system might have affected the participant who showed no improvement. Thus, in future developments we will aim to improve the usability and ease of wearing our system.

6. CONCLUSION

We proposed and developed a hat-type hearing aid system with flexible sound directivity patterns by using high-speed single-bit signal processing. Given that the sampling rate of the single-bit signal is very high, the proposed system can finely control the sound directivity using delay-and-sum beamforming. From the measurement of impulse responses for the proposed system, we conclude that it is important for the design of the sound directivity pattern using delay-and-sum beamforming to compensate the time delays and select a reduced number of microphones to be used. Simulation results with impulse responses showed that reflections from the hat influenced the directivity range but not the FWHM. After the compensation and reduction of the number of microphones, the developed system achieved around 10 degrees of FWHM and 17 dB of directivity range.

It is possible to change the direction of sound directivity patterns in the range of 360 degrees because we installed microphones all around the hat in its horizontal plane. When multiple directivity patterns are combined, it would possible to emphasise sounds coming from various directions at the same time.

In the proposed system, the user selects the beam direction to emphasise a specific speech source. In future developments, we will aim to automate the direction of arrival estimation by using other sensors, cameras, etc. together.

REFERENCES

[1] B. D. Van Veen and K. Buckley, “Beamforming: A versatile approach to spatial filtering,” IEEE ASSP Mag., 5(2), pp. 4–24 (1988).

[2] M. I. Mandel, S. Bressler, B. Shinn-Cunningham and D. P. Ellis, “Evaluating source separation algorithms with reverberant speech,” IEEE Trans. Audio Speech Lang. Process., 18, 1872–1883 (2010).

[3] T. Mikoshiba, Y. Oikawa, Y. Ouchi and Y. Yamasaki, “Hearing aid with curved surface reflector,” Proc. Autumn Meet. Acoust. Soc. Jpn., pp. 1379–1380 (2013) (in Japanese).

[4] J. S. Lim and A. V. Oppenheim, “Enhancement and bandwidth compression of noisy speech,” Proc. IEEE, 67, 1586–1604 (1979).

[5] J. E. Greenberg and P. M. Zurek, “Evaluation of an adaptive beamforming method for hearing aids,” J. Acoust. Soc. Am., 91, 1662–1676 (1992).

[6] J. M. Kates, Digital Hearing Aids (Plural Publishing Inc., San Diego, 2008).

[7] V. Hamacher, J. Chalupper, J. Eggers, E. Fischer, U. Kornagel, H. Puder and U. Rass, “Signal processing in high-end hearing aids: State of the art, challenges, and future trends,” EURASIP J. Appl. Signal Process., 18, 2915–2929 (2005).

[8] S. Doolo, W. Kellermann, S. Makino and S. Nordholm, “Multichannel signal enhancement algorithms for assisted listening devices,” IEEE Signal Process. Mag., 32, 18–30 (2015).

[9] T. Inoue, R. Imai, Y. Ikeda and Y. Oikawa, “Hat-type hearing aid system using MEMS microphone array,” in Proc. 12th West. Pac. Acoust. Conf. (WESPAC), pp. 156–159 (2015).

[10] M. Kozak and M. Karaman, “Digital beamforming using non-uniform oversampling delta-sigma conversion,” Proc. IEEE Ultrason. Symp., pp. 1279–1282 (1999).

[11] B. G. Tomov and J. A. Jensen, “Compact FPGA-based beamformer using oversampled 1-bit A/D converters,” IEEE Trans. Ultrason. Ferroelectr. Freq. Control, 52, 870–880 (2005).

[12] S. Takeoka, R. Kogure and Y. Yamasaki, “Super multi channel microphone array system using high speed 1-bit signal processing,” Proc. Autumn Meet. Acoust. Soc. Jpn., pp. 765–766 (in Japanese).

[13] H. Inose, Y. Yasuda and J. Murakami, “A telemetering system by code modulation — Δ-Σ modulation,” IRE Trans. Space Electron. Telem., SET-8, 204–209 (1962).

[14] H. Inose and Y. Yasuda, “A unity bit coding method by negative feedback,” Proc. IEEE, 51, 1524–1535 (1963).

[15] F. Asano, Array Signal Processing for Acoustics (Corona Publishing, Tokyo, 2011) (in Japanese).

[16] R. Imai, Y. Koyano, Y. Ikeda, Y. Oikawa and Y. Yamasaki, “Single-bit signal processing by high speed system clock,” Proc. Spring Meet. Acoust. Soc. Jpn., pp. 645–646 (2016) (in Japanese).

[17] Y. Suzuki, F. Asano, H. Y. Kim and T. Sone, “An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses,” J. Acoust. Soc. Am., 97, 1119–1123 (1995).

[18] S. Amano, K. Kondo, Y. Suzuki and S. Sakamoto, “Speech dataset for word intelligibility test based on word familiarity (FW03),” NII Speech Resources Consortium (2006).