IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 13, 2019, accepted May 7, 2019, date of publication May 15, 2019, date of current version May 29, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2916723

Design of Novel Field Programmable Gate Array-Based Hearing Aid

BOR-SHING LIN[1], (Member, IEEE), PO-YU YANG[2], CHING-FENG LIU[3], YI-CHIA HUANG[2], CHENGYU LIU[4], AND BOR-SHYH LIN[2], (Senior Member, IEEE)

[1]Department of Computer Science and Information Engineering, National Taipei University, New Taipei City 23741, Taiwan
[2]Institute of Imaging and Biomedical Photonics, National Chiao Tung University, Tainan 71150, Taiwan
[3]Department of Medical Research, Chi Mei Medical Center, Chang Jung Christian University, Tainan 71004, Taiwan
[4]School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China

Corresponding author: Bor-Shyh Lin (borshyhlin@gmail.com)

This work was supported in part by the Ministry of Science and Technology in Taiwan under Grant MOST 107-2221-E-009-017 and Grant MOST 107-2221-E-305-014, in part by the University System of Taipei Joint Research Program under Grant USTP-NTPU-TMU-108-01, in part by the Faculty Group Research Funding Sponsorship by National Taipei University under Grant 2018-NTPU-ORDA-04, and in part by the Higher Education Sprout Project of the National Chiao Tung University and Ministry of Education (MOE), Taiwan.

ABSTRACT Hearing loss is one of the most common chronic diseases. For people with hearing loss, communicating with other people, particularly in an environment with considerable background noise, is difficult. Recently, several hearing aids have been developed to improve speech comprehension in a noisy environment. The use of an adaptive beamformer is one of the alternative methods for improving speech intelligibility. However, the adaptive beamformer requires the location of the desired speaker to estimate the time differences of arrival (TDOAs) of speech sources to numerous spatially separated sensors in acoustics. In general, the technique of steered response power source localization was used to estimate the TDOA; however, this technique was easily affected by environmental reverberation. To overcome the aforementioned concern, a novel hearing aid is proposed in this paper. By using an image processing technology, the location of the desired speaker could be manually selected to provide precise information on the TDOA. Moreover, adaptive signal enhancements were implemented in a field-programmable gate array to enhance the speech of interest in real time. The experimental results indicate that the proposed system could improve speech intelligibility in various noisy environments. Therefore, the proposed system may be employed to improve the daily lives of people with hearing the loss in the future.

INDEX TERMS Hearing aid, adaptive beamformer, field programmable gate array, time difference of arrival, image processing technology.

I. INTRODUCTION

Recently, hearing loss has become one of the most prevalent chronic diseases among elderly people [1]. In the United States, approximately 27%–46% of older adults experience hearing loss. For people with hearing loss, communication with other people is inconvenient, particularly in an environment with considerable background noise, and this results in social isolation and withdrawal, mental illnesses, and reduced quality of life [2]. Therefore, people with hearing loss generally require the assistance of a hearing aid to enhance the speech of interest.

Several single-microphone hearing aids have been developed to enhance the noisy speech of interest. In these hearing aids, noisy speech sounds are divided into several frequency bands. Depending on the environment and the degree of hearing loss, gains corresponding to particular frequency bands are adjusted to enhance speech intelligibility [3]. However, when the spectrum of environmental noise overlaps with that of the speech of interest, the enhancement of speech intelligibility is limited [4].

The microphone array technique is an alternative approach for enhancing speech intelligibility. By using the spatial differences between the speech of interest and noise, the intelligibility of the speech of interest can be improved and unnecessary noise can be eliminated [5], [6]. In 1969, Capon proposed a minimum variance distortionless response

The associate editor coordinating the review of this manuscript and approving it for publication was Chua Chin Heng Matthew.

(MVDR) beamformer to minimize the power of the undesired signal components; however, the performance of the MVDR for eliminating diffusion noise and reverberation is limited [7]. A linearly constrained minimum variance (LCMV) beamformer is the generalization of the MVDR. The difference between the MVDR and LCMV beamformers is the addition of multiple linear constraints to reduce interference and prevent the distortion of the desired signal [8]. However, computations of the LCMV beamformer are highly complex [9].

A generalized sidelobe canceller (GSC) proposed by Buckley and Griffiths [10] is a recognized adaptive beamformer. In the GSC, the output of a fixed delay-and-sum beamformer is employed as a speech reference, a blocking matrix is used to obtain noise references by combining delayed multichannel microphone signals, and an adaptive noise canceller is then used to reduce noise components in the speech reference [11], [12]. The performance of the GSC for noise cancellation is superior to that of the aforementioned fixed beamforming technique [13]. However, the GSC requires the location of the desired speaker to estimate the time differences of arrival (TDOA) of the speech of interest to numerous spatially separated sensors in acoustics and radar [14]. In general, the cross-correlation between multichannel microphone signals is used to estimate information on TDOA. When the correlation has the largest value, delay time can be considered TDOA. However, TDOA estimation using the approach of cross-correlation is easily affected in the reverberant environment [15].

To overcome the aforementioned concerns, a novel field programmable gate array (FPGA)-based hearing aid is proposed in this study. In the proposed hearing aid, the location of the desired speaker can be manually selected, and the effect of environmental noise on TDOA estimation can be effectively reduced using the image processing technology. Furthermore, four microphones are used to collect noisy speech from various locations. An FPGA-based sound processing module was designed to run the algorithm of adaptive signal enhancements (ASEs) for the real-time enhancement of the speech of interest. Finally, the performance of the proposed hearing aid was validated under different environmental conditions and was evaluated through a questionnaire exam. The experimental results indicate that the proposed hearing aid can effectively enhance noisy speech and assist people with hearing loss.

II. METHODS AND MATERIALS

A. SYSTEM ARCHITECTURE AND IMPLEMENTATION

Figs. 1 (a) and (b) present the basic scheme and image of the proposed FPGA-based hearing aid. The system primarily comprises a wide-angle lens webcam, an FPGA-based sound processing module, and a back-end host system. The wide-angle lens webcam was placed at the center of four microphones to collect the environmental image, and this image was then transferred to the back-end host system. A real-time

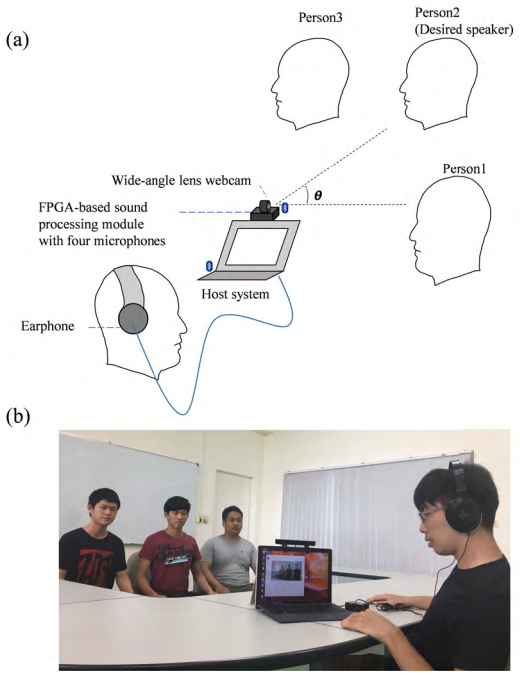


FIGURE 1. (a) Basic scheme and (b) image of proposed FPGA-based hearing aid system.

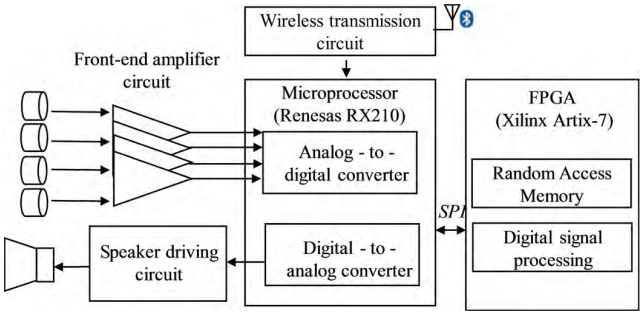


FIGURE 2. Block diagram of FPGA-based sound processing module.

face recognition program developed in the back-end host system detected all people in this environmental image, allowed the user to manually select the desired speaker for TDOA estimation by using the image processing technique, and wirelessly transmitted TDOA information to the FPGA-based sound processing module via Bluetooth. Finally, the algorithm for ASE was employed in the designed FPGA-based sound processing module for the real-time enhancement of the speech of interest obtained from four-channel noisy speech sounds.

Fig. 2 shows the block diagram of the designed FPGA-based sound processing module. The system primarily comprises four microphones, front-end amplifier circuits, a microprocessor, a wireless transmission circuit, a speaker driving circuit, and an FPGA circuit. Speech sound was first collected and converted into an electrical signal using these microphones and was amplified and filtered using the front-end amplifiers (Gain: 30 V/V, 150–4000 Hz). These speech signals were digitized using a 12-bit analog-to-digital

converter developed in the microprocessor with a sampling rate of 20 kHz, and the digitized signals were then sent to the FPGA circuit. The distance between the sound sources and microphone array was short. The speech sounds with a sampling rate of 20 kHz can provide sufficient resolution to distinguish the TDOAs of various microphones. The wireless transmission circuit comprises a printed circuit board antenna and Bluetooth module. In the back-end host system, the time difference of arrival was estimated through an environmental image captured using the wide-angle lens webcam and was then transmitted to the FPGA-based sound processing module via Bluetooth. The FPGA-based sound processing module improved the adaptive signal in real time. After ASE, the enhanced speech was sent to the 12-bit digital-to-analog converter in the microprocessor with a sampling rate of 20 kHz to generate analog speech signals and then sent to the speaker-driving circuit to play the enhanced speech through an earphone.

In this study, a commercial laptop with a Windows 10 operating system was used as the back-end host system platform. A real-time monitoring program in the host system was developed using Microsoft C# to recognize human faces from the environmental image and to allow the user to select the desired speaker. The location of the desired speaker was estimated to obtain TDOA. TDOA information was sent to the FPGA-based sound processing module.

B. ESTIMATION OF TIME DIFFERENCE OF ARRIVAL USING AN IMAGE PROCESSING TECHNIQUE

An Open Source Computer Vision Library (OpenCV) was initially developed in 1999 by Intel and can be used for real-time image processing, computer vision, and pattern recognition. In this study, a Haar cascade classifier in OpenCV was used to detect the faces and eyes of speakers from the image. According to the size and location of the detected eyes, the vertical distance and angle between the webcam and desired speaker could be estimated to further calculate TDOA. Figs. 3 (a) and 3 (b) present the approach used for estimating the vertical and horizontal distances between the webcam and desired speaker, respectively. W_{eyes} is the estimated eye width, and d_{hor} is the distance between the desired speaker and image center. In this example, the eye widths were approximately 198 and 33 pixels when the vertical distances were approximately 80 and 492 cm, respectively. When the vertical distance was approximately 80 cm, the distances between the desired speakers and image center d_{hor} were approximately 270 and 518 pixels when the horizontal distances were approximately 18 and 34 cm, respectively. According to linear geometry, the relationships between the vertical distance D_{vert} and eye width W_{eyes} and between the distances of the desired speaker from the image center d_{hor} and horizontal distance D_{hor} could be estimated as follows. In this study, 30 participants with different heights, weights, and physical statures were included, and information on W_{eyes} and d_{hor} with different distances between the face and camera were obtained. A linear regression method with a

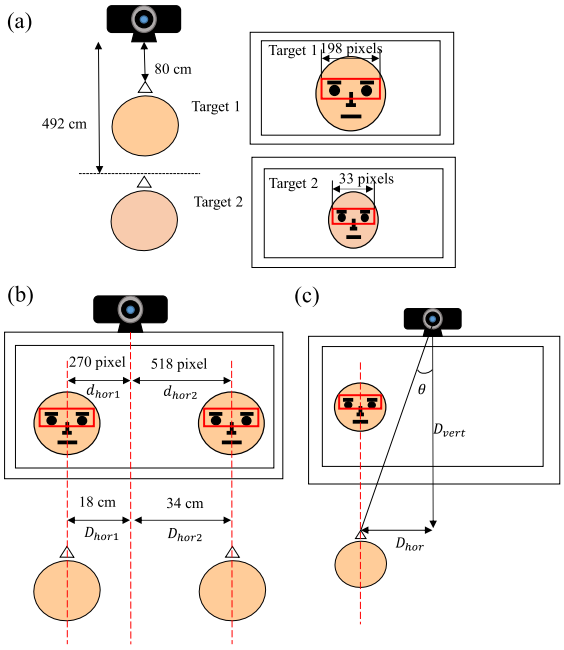


FIGURE 3. Illustration of approach for estimating (a) vertical distance, (b) horizontal distance, and (c) angle from webcam.

minimum mean square error was then used to obtain weights using the following eqs. (1) and (2):

$$D_{vert} = -2.497W_{eyes} + 572.745 \quad (1)$$

$$D_{hor} = 0.066d_{hor} - 0.018 \quad (2)$$

By using the Pythagorean theorem [16], the angle between the vertical and horizontal distances could be calculated as follows:

$$\theta = \tan^{-1}\frac{D_{hor}}{D_{vert}} \quad (3)$$

The actual distances between the desired speaker and various microphones were calculated using the estimated vertical distance, horizontal distance, and angle of the desired speaker to estimate TDOA.

C. ADAPTIVE BEAMFORMER FOR SPEECH ENHANCEMENT

The adaptive beamformer was first developed in 1960s for the military applications of sonar and radar [17] to eliminate uninteresting noises through destructive combination and to improve the desired signal through constructive combination from a particular direction to output. Fig. 4 (a) presents the basic scheme of the adaptive beamformer technique. The speech sounds obtained using various microphones were first enhanced using a delay-and-sum beamformer, and the speech signals enhanced by the beamformer were used as the common reference input of ASEs. The noisy speech was then obtained through various microphones and used as the primary inputs of ASEs. Finally, the averaged output of ASEs was calculated to improve the signal to noise ratio (SNR) of the speech of interest.

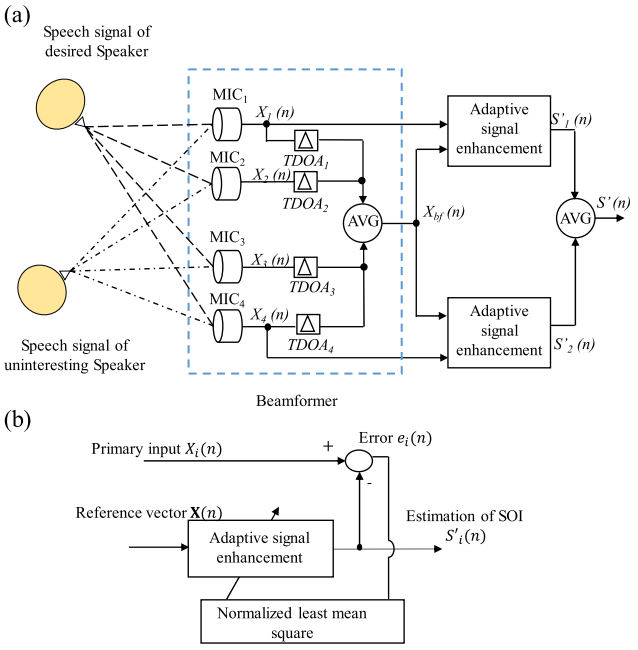


FIGURE 4. Basic scheme of (a) adaptive and (b) ASEs.

Fig. 4 (b) depicts the basic scheme of ASE [18]. $X_i(n)$ is noisy speech at iteration t obtained through the ith microphone MIC_i and used as the primary input of ASE. $\mathbf{X}(n) = [X_{bf}(n), X_{bf}(n-1), \ldots, X_{bf}(n-M)]$ is a $1 \times M$ output vector of the front-end beamformer at iteration n, which was used as the reference input of ASE. Here, M is the window length, and $X_{bf}(n)$ can be calculated as follows in (4), as shown at the bottom of this page, where $TDOA_i$ is the estimated TDOA for the i^{th} microphone. By combining the phase delay of the speech of interest obtained using various microphones, the speech of interest was simply enhanced, and the unwanted speech (i.e., noise) was eliminated. The i^{th} ASE output $S'_i(n)$ at iteration t could be calculated as follows:

$$S_i^{'}(n) = \boldsymbol{w}_i^{\boldsymbol{T}}(n)\boldsymbol{X}(n) \tag{5}$$

where $\boldsymbol{w_i}(n) = [w_{i,0}(n), w_{i,1}(n), \ldots, w_{i,M-1}(n)]^T$ denotes a $M \times 1$ weight vector of the i^{th} ASE filter. Error signal $e_i(n)$ is the difference between the i^{th} ASE output and primary input and can be given as follows:

$$e_i(n) = X_i(n) - S_i^{'}(n) \tag{6}$$

The i^{th} ASE filter weight could be adapted through normalized least mean square algorithm [19]:

$$\boldsymbol{w_i}(n+1) = \boldsymbol{w_i}(n) + \frac{\mu}{1 + ||X(n)||^2} e_i(n)\boldsymbol{X}(n) \tag{7}$$

where μ is the given step size.

D. IMPLEMENTATION OF FIELD PROGRAMMABLE GATE ARRAY-BASED ADAPTIVE BEAMFORMER

Fig. 5 shows the hardware architecture of the FPGA-based adaptive beamformer. A Xilinx FPGA (Artix-7 Xc7a15tcsg324-1 FPGA, Xilinx, United States) with a working frequency of up to 100 MHz was used in this study. A series peripheral interface protocol was used to connect the microprocessor (RX210, Renesas, Japan) and FPGA chip. First, the microprocessor transmitted the data packet of the noisy speech obtained through the four microphones using a master out slave in. A series-to-parallel module converted series data into parallel data to facilitate internal operations in the FPGA chip. The noisy speech data from the received data packet were obtained using Mic1, Mic2, Mic3, and Mic4 and were stored in RAM1. The mean module included an adder and shifter and was designed for beamforming processing. The output was stored in RAM2 and used as the reference input for ASE. The signal normalization module comprised a multiplier and an accumulated adder and was designed to normalize the reference input of ASE. The ASE output module was designed to calculate the outputs of ASE filters. The error estimation module was designed to calculate the difference between the ASE output and primary input. The weight adaption module was designed to adapt the filter weights and store them in RAM3.

The format of the signed fixed-point arithmetic (2's complement) was used in this system to avoid a calculation overflow. In floating-point calculation, the value was multiplied by 1000 to reserve three decimals before calculation. After calculation, the obtained value was divided by 1000 to recover its true value. In this system, the value was left-shifted by 10 bits before calculation, and the obtained value was then right-shifted by 10 bits to recover its true value after calculation.

Figs. 6 (a) and 6 (b) depict the finite-state machine (FSM) and data flow of the FPGA-based adaptive beamformer. The finite-state machine of this system comprised four states and several control signals, including DataIn_En, DataOut_En, and EN_INT. DataIn_En was used to receive new data; that is, when DataIn_En was high, new data was transmitted to the subsequent state. EN_INT was designed to determine the continuous reception of new data. When DataIn_En was high and M new data were received, EN_INT became high to prevent the reception of new data. DataOut_En was used to confirm the completion of calculation; that is, when DataOut_En was high, the calculated result was outputted. In the FPGA chip, the subsequent state was attained at each new clock cycle to perform calculations at each state, and calculations were completed after four state cycles. The system comprised 2834 look up tables (LUTs), 3997 flip flops (FFs), 56 look up table distribution RAMs (LUTRAMs), 6 digital

$$X_{bf}(n) = \frac{X_1(n - TDOA_1) + X_2(n - TDOA_2) + \ldots + X_4(n - TDOA_4)}{4} \tag{4}$$

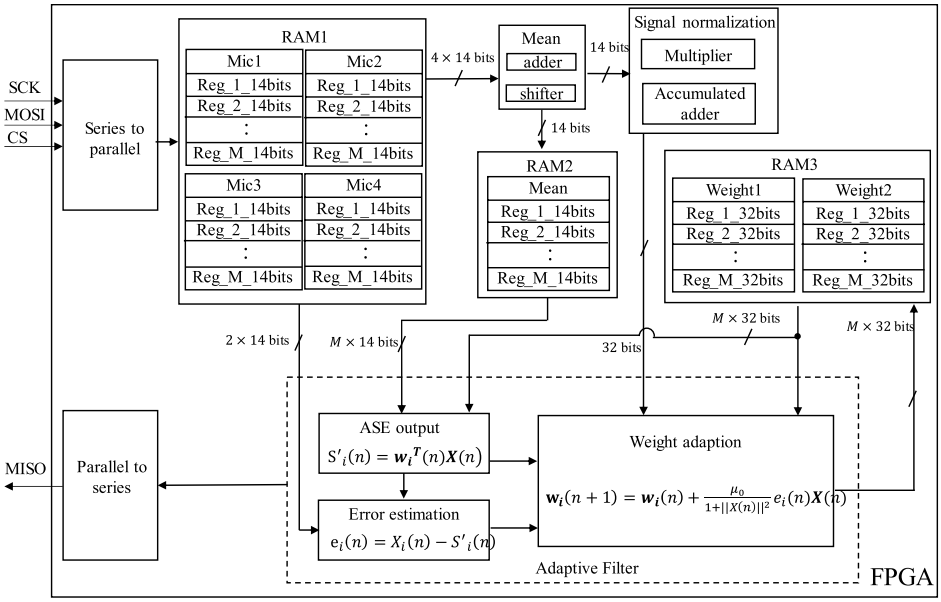


FIGURE 5. Hard architecture of FPGA-based adaptive beamformer.

TABLE 1. A summary of the primary resource usage in FPGA.

Component	FPGA Utilization FPGA Utilization
LUT	2834 (27%)
LUTRAM	56 (1%)
FF	3997 (19%)
DSP	6 (13%)
IO	6 (3%)
MMCM	1 (20%)

signal processors (DSPs), and 1 mixed mode clock manager (MMCM). Six input and output (IO) were used to implement this system. To apply the adaptive beamformer algorithm, 10 adders, 6 multipliers, and 2 dividers were used. Table 1 presents the summary of the primary resource usage in the FPGA. Fig. 7 depicts the internal placement and routing of the FPGA. The power consumption of the FPGA was analyzed, and analysis indicated that the logic and signal power consumption were 0.196 and 0.006 W, respectively, with a total power consumption of 0.202 W.

In state S0, the noisy speech data obtained using the microphones were received and stored to RAM1 through a first in first out register. The average of the four-channel noisy speech data could be calculated by one adder and one shifter, and then stored to RAM2. After receiving M new data, EN_INT became high, and calculation was then performed for the subsequent state. In state S1, the output and error of ASE were calculated, the input vector was normalized, and the used variables were loaded from RAM2 and RAM3. The reference input vector was normalized using a multiplier and an adder. The error of ASE was calculated using a subtractor, and the output of ASE was calculated using an adder and a multiplier. In state S2, the weight was updated. The error of ASE and the normalization of the reference input vector were obtained from the previous calculation. The adaption of the weight vector was calculated using two multipliers, dividers, and adders each. The updated weight vector was then stored in RAM3. Finally, in state S3, the shifter shifted the previous data in RAM1 and RAM2. When the calculations were completed, DataOut_En became high, and thus returned to state S0.

III. RESULTS

A. PERFORMANCE OF PROPOSED HEARING AID FOR SPEECH ENHANCEMENT

In this study, the performance of the proposed hearing aid was investigated. First, the TDOA error was tested. The experimental results indicate that the TDOA error was less than 5×10^{-5} when the angle ranged from $-60°$ to $60°$ and the distance range was 0.5–5 m.

The performance of the proposed hearing aid for speech enhancement was examined. Figs. 8 (a), 8 (b), and 8 (c) depict three environmental conditions in this experiment. The desired speaker and noise source were placed at different locations, and the SNRs of the received noisy speech sounds in conditions 1, 2, and 3 were approximately -4.61, -6.32, and -7.26 dB, respectively. The definition of the SNR is as follows:

$$\text{SNR} = 20log_{10}(\frac{A_{Signal}}{A_{Noise}}) \tag{8}$$

where A_{Signal} and A_{Noise} denote the root mean square of noise-free speech sound and noise, respectively. The filter order and step size of the proposed system were 32 and 0.007, respectively. Fig. 9 shows the noise-free speech of interest, noisy speech sound, and speech sound enhanced using the proposed system in condition 3. The SNRs of the enhanced speech sounds in conditions 1, 2, and 3 were improved to approximately 0.64, 0.16, and -1.42 dB, respectively.

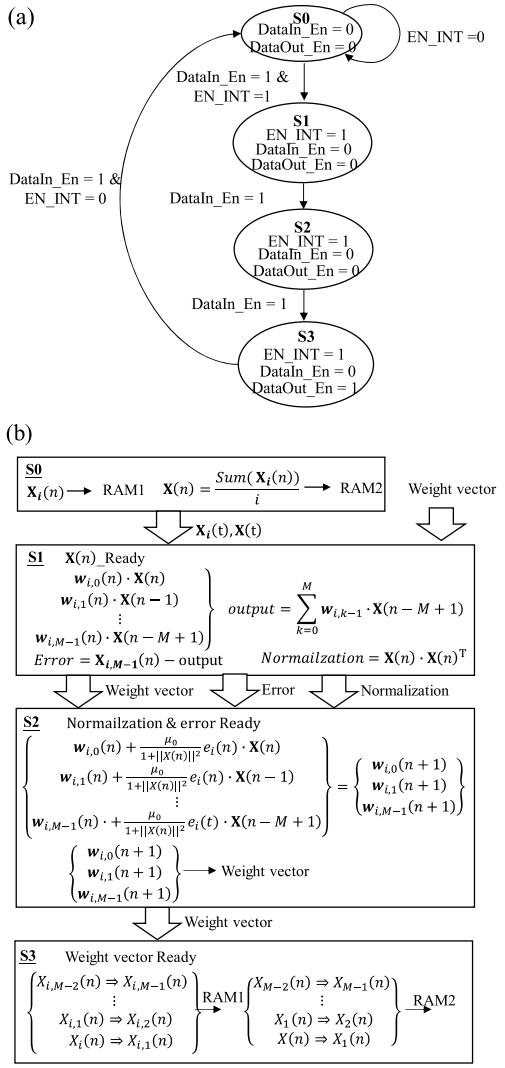


FIGURE 6. (a) Finite-state machine and (b) data flow of adaptive beamformer.

TABLE 2. PESQ subjective analysis.

	Condition 1	Condition 2	Condition 3
Noisy speech sound	1.5384	1.5545	1.5452
Enhanced speech sound	1.9041	1.8461	1.9296

Finally, the perceptual evaluation of speech quality (PESQ) subjective analysis was evaluated, and Table 2 shows the results. The noisy speech sounds and enhanced speech sounds under different conditions were compared with the noise-free speech. The experimental results indicate that the proposed system could effectively increase PESQ.

B. PERFORMANCE OF PROPOSED SYSTEM FOR ENHANCING SPEECH INTELLIGIBILITY

The performance of the proposed system for enhancing speech intelligibility was examined. This experiment was conducted with 20 healthy participants. In this experiment, the participants were instructed to recognize five speech reception thresholds (SRTs) [20] from a speaker.

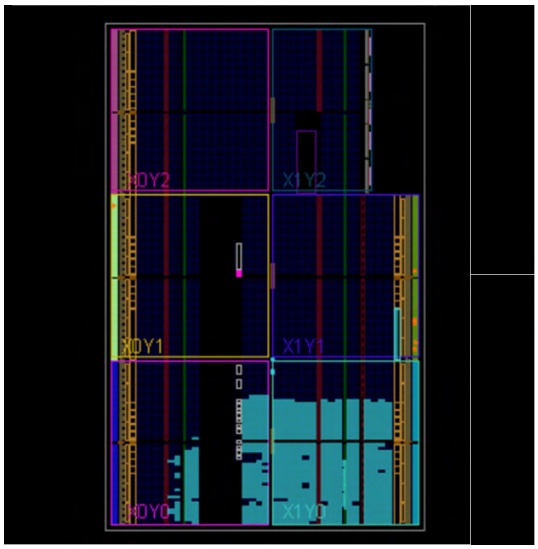


FIGURE 7. FPGA internal placement and routing.

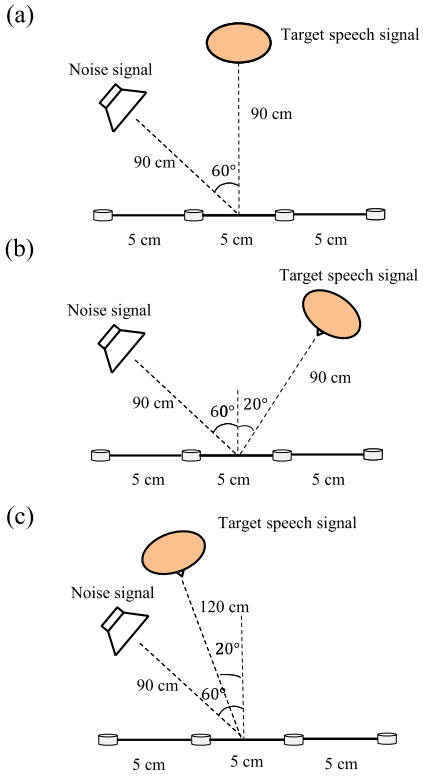


FIGURE 8. Locations of desired speakers and noise sources in conditions (a) 1, (b) 2, and (c) 3.

The participant was then instructed to complete a questionnaire (Table 3) to evaluate the speech intelligibility with or without speech enhancement. Fig. 10 presents the questionnaire result. The significance of the difference was analyzed using a t test and was defined as $p < 0.05$. The distinguishability of the SRT with and without speech enhancements was 96% and 88%, respectively. Moreover, the quality, naturalness, and clarity of noisy speech sounds were improved using the proposed system.

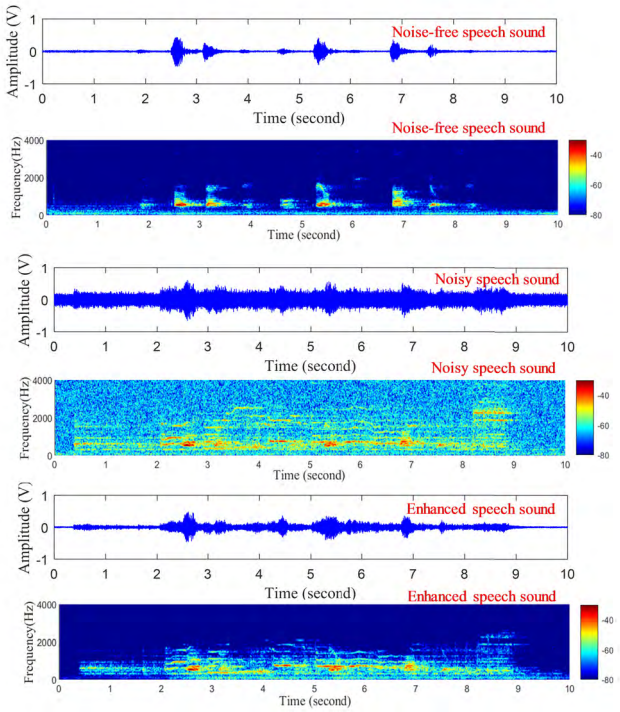


FIGURE 9. Noise-free speech, noisy speech, and enhanced speech and their spectrograms obtained using the proposed system in condition 3.

IV. DISCUSSION

In this study, TDOA estimation was considerably improved by using the image processing technique. The experimental results show that the TDOA error was lower than 5×10^{-5} under various conditions. In the previous study [21], the technique of steered response power phase-transform weighted source localization was used to estimate TDOA and its TDOA error was higher than 0.035 when the SNR of noisy speech was approximately 2 dB. The use of the image processing technique for TDOA estimation provided precise information on TDOA. Moreover, compared with the speech enhancement performance of an adaptive filter with a single channel input (about 3.5 dB), the performance of the proposed system with multichannel inputs for speech enhancement under various conditions was approximately 5 dB. The effect of the locations of the desired speaker and noise source on speech enhancement performance was less noteworthy because of the excellent performance of TDOA estimation.

In this study, the adaptive beamformer was implemented in the FPGA-based sound processing module. In 2007, Halupka *et al.* proposed the FPGA implementation of phase-error-based filtering and sound localization particularly designed for low power consumption [22]. This study used a recognized method of TDOA estimation of the speaker signal between a microphone pair to determine the position of the speaker. The experimental results show that the performance of the proposed system for postprocessing SNR and PESQ in the noisy condition was superior to that of the aforementioned system. In this study, the FPGA architecture provided the advantage of parallel signal processing to perform complex computations in a few cycles. Therefore, implementing a signal or an image processing algorithm in a portable or wearable device is suitable. For the FPGA architecture design, reducing the cost of hardware resources is crucial for limiting the resource utilization of the FPGA chip. The proposed hearing aid comprises 1437 slices (18%) and 2834 LUTs (27%), with a clock frequency of 100 MHz. Other similar FPGA systems comprise 3906 slices and 3513 LUTs with a clock frequency of 25 MHz in a very high-speed hardware description language, 2568 slices and 2536 LUTs with a clock frequency of 12.32 MHz in high-level synthesis (HLS) [23], and 16316 slices and 13396 LUTs with a clock frequency of 50 MHz [24]. Compared with the aforementioned systems, the proposed system requires fewer FSMs and arithmetic logic units; however, it requires a clock with higher frequency.

Several hearing aids have been developed in other recent studies. Table 4 compares the proposed system and other hearing aids. In 2016, Huang *et al.* applied the technique of a reconfigurable sound wave decomposition filter bank

TABLE 3. Questionnaire for evaluating speech quality and speech intelligibility with and without speech enhancement.

Q1: Score the speech quality with speech enhancement (Very Good: 4 points; Good: 3 points; Poor: 2 points Very Poor: 1 point) □Very Good □Good □Poor □Very Poor
Q2: Score the speech quality without speech enhancement (Very Good: 4 points; Good: 3 points; Poor: 2 points Very Poor: 1 point) □Very Good □Good □Poor □Very Poor
Q3: Score the speech naturalness with speech enhancement (Very Good: 4 points; Good: 3 points; Poor: 2 points Very Poor: 1 point) □Very Good □Good □Poor □Very Poor
Q4: Score the speech naturalness without speech enhancement (Very Good: 4 points; Good: 3 points; Poor: 2 points Very Poor: 1 point) □Very Good □Good □Poor □Very Poor
Q5: Score the speech clarity with speech enhancement (Very Good: 4 points; Good: 3 points; Poor: 2 points Very Poor: 1 point) □Very Good □Good □Poor □Very Poor
Q6: Score the speech clarity without speech enhancement (Very Good: 4 points; Good: 3 points; Poor: 2 points Very Poor: 1 point) □Very Good □Good □Poor □Very Poor
Q7: Score the SRT distinguishability with speech enhancement (All correct: 5 points; 4 correct: 4 points; 3 correct: 3 points; 2 correct: 2 points…)
Q8: Score the SRT distinguishability without speech enhancement (All correct: 5 points; 4 correct: 4 points; 3 correct: 3 points; 2 correct: 2 points…)

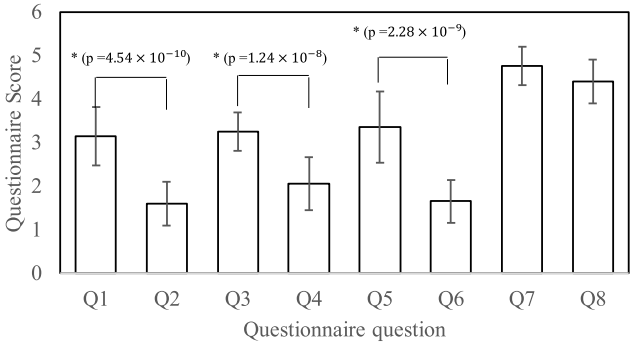


FIGURE 10. Questionnaire results for speech quality and speech intelligibility with and without speech enhancement. Here, ∗ indicates the significance of the difference ($p < 0.05$).

TABLE 4. Systems comparison between proposed system and other hearing aids.

	S. Huang *et al.* [25]	J. Traa *et al.* [26]	Proposed system
Speech enhancement method	Reconfigurable sound wave decomposition filter bank	Robust source localization and enhancement with probabilistic steered response power model	Adaptive signal enhancement filter
Number of microphones	1	8	4
Performance of speech enhancement	MMME: 2.25 dB	SIR: 12 dB	SNR: 5 dB
Computational complexity	Medium	High	Low
Functions	Enhancing hearing loss characteristic of patients	Estimating TDOA by SRP localization to enhance speech	Estimating TDOA by image recognition technology to enhance speech
Manually selecting desired speaker	No	No	Yes
Limitations of use	Influence of spectrum overlap between noise and desired sound	Influence of environmental reverberation on TDOA estimation	Influence of environmental brightness on TDOA estimation

in hearing aids [25]. A cosine modulation technology was used to generate uniform subbands, and the spectrum of noisy speech was extracted using these uniform subbands. These subbands were then converted into nonuniform subbands through nonlinear transformation. By improving the subband of interest and suppressing other subbands, the useful components of the received noisy speech can be enhanced. The minimum maximum matching error of the aforementioned method was approximately 2.25 dB. However, the performance of this method may be affected when the spectrum of the desired sound overlaps with that of noise. Traa *et al.* proposed the MVDR with steered response power (SRP) source localization for multiple sources [26]. SRP localization was used for TDOA estimation by searching the peaks in the output power of the beamformer and for providing the estimated TDOA to the MVDR beamformer for speech enhancement. The SNR of the MVDR with SRP was approximately 12 dB when the SNR of noisy speech was approximately 5 dB. However, the performance of SRP localization on TDOA estimation was easily affected by environmental reverberation, and its performance was evidently worse than that of the image processing technique used in this study. In contrast to the aforementioned hearing aids, the image processing technique used in the proposed system can provide precise TDOA information, and the location of the desired speaker can be manually selected from a graphics user interface of the proposed system. Moreover, the ASE algorithm applied in the FPGA-based sound processing module can provide real-time and effective speech enhancement. The proposed system can currently enhance speech sound from only one speaker. In the future, the software can be modified to rapidly switch different sound sources to overcome the problem caused by multiple speakers. To improve the convenience of use in daily life, four microphones can be embedded in a glass frame, and the image processing technique for TDOA estimation can be implemented as a mobile application for smartphones or smart glasses in the future. When using a smartphone or smart glasses in the future to implement the proposed system, users will not experience a perceived stigma. Therefore, in the future, the proposed system can be modified as a wearable device and used by people with hearing loss to improve their daily lives. This study included participants with normal hearing ability. This study can be extended to people with hearing impairments in the future.

V. CONCLUSION

In this study, the FPGA-based hearing aid was designed to enhance the speech of interest. The image processing technique was used to estimate TDOA and compensate for TDOA in the adaptive beamformer.

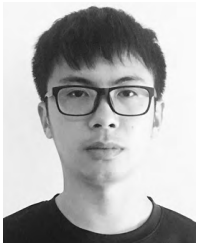
In this study, the most crucial advantage of the proposed system is the ability to directly select the location of the desired speaker in an environment with multiple speakers or with noise by using an image processing technology, thereby enhancing the speech of interest and inhibiting unwanted speech or noise from other speakers or surrounding. By using the image processing technology, the effect of environmental noise (e.g., noisy classroom, traffic noise, and music) on estimating TDOA can be effectively reduced. The adaptive beamformer was implemented in the FPGA-based sound processing module for the real-time enhancement of the speech of interest and elimination of noisy speech. The experimental results indicate that the SNR of the proposed system on speech enhancement was about 5 dB, and the effect of locations of the desired speaker and noise source was less noteworthy. Furthermore, the questionnaire responses demonstrated that the proposed system could effectively improve speech intelligibility. Therefore, the proposed system can be an appropriate prototype of hearing aids and can be used by people with hearing loss to improve their daily lives in the future.

REFERENCES

[1] B. Yueh, N. Shapiro, C. H. MacLean, and P. G. Shekelle, "Screening and management of adult hearing loss in primary care: Scientific review," *JAMA*, vol. 289, no. 15, pp. 1976–1985, Apr. 2003.
[2] M. M. Popelka, K. J. Cruickshanks, T. L. Wiley, T. S. Tweed, B. E. Klein, and R. Klein, "Low prevalence of hearing aid use among older adults with hearing loss: The epidemiology of hearing loss study," *J. Amer. Geriatrics Soc.*, vol. 46, no. 9, pp. 1075–1078, Sep. 1998.
[3] T. Schneider and R. Brennan, "A multichannel compression strategy for a digital hearing aid," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 1, Apr. 1997, pp. 411–414.

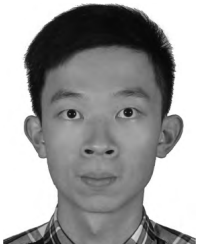
[4] N. R. Council, *Removal of Noise From Noise-Degraded Speech Signals*. Washington, DC, USA: National Academy, 1989.
[5] T. Van den Bogaert, S. Doclo, J. Wouters, and M. Moonen, "Speech enhancement with multichannel Wiener filter techniques in multimicrophone binaural hearing aids," *J. Acoust. Soc. Amer.*, vol. 125, no. 1, pp. 360–371, Jan. 2009.
[6] J. P. Dmochowski and J. Benesty, "Microphone arrays: Fundamental concepts," in *Speech Processing in Modern Communication*, I. Cohen, J. Benesty, and S. Gannot, Eds. Berlin, Germany: Springer, 2010.
[7] K. U. Simmer, J. Bitzer, and C. Marro, "Post-filtering techniques," in *Microphone Arrays*. Berlin, Germany: Springer-Verlag, 2001, ch. 3, pp. 39–60.
[8] E. A. P. Habets, J. Benesty, and P. A. Naylor, "A speech distortion and interference rejection constraint beamformer," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 3, pp. 854–867, Mar. 2012.
[9] Y. Lee and W.-R. Wu, "A robust adaptive generalized sidelobe canceller with decision feedback," *IEEE Trans. Antennas Propag.*, vol. 53, no. 11, pp. 3822–3832, Nov. 2005.
[10] K. M. Buckley and L. J. Griffiths, "An adaptive generalized sidelobe canceller with derivative constraints," *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 311–319, Mar. 1986.
[11] S. Doclo and M. Moonen, "GSVD-based optimal filtering for multi-microphone speech enhancement," in *Microphone Arrays*, M. Brandstein and D. Ward, Eds. Berlin, Germany: Springer-Verlag, 2001, ch. 6, pp. 111–132.
[12] S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, "A consolidated perspective on multimicrophone speech enhancement and source separation," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 4, pp. 692–730, Apr. 2017.
[13] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, no. 2, pp. 4–24, Apr. 1988.
[14] H. Schau and A. Robinson, "Passive source localization employing intersecting spherical surfaces from time-of-arrival differences," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, no. 8, pp. 1223–1225, Aug. 1987.
[15] K. C. Ho and Y. T. Chan, "Solution and performance analysis of geolocation by TDOA," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 4, pp. 1311–1322, Oct. 1993.
[16] D. C. Benson, *The Moment of Proof: Mathematical Epiphanies*. New York, NY, USA: Oxford Univ. Press, 2000.
[17] B. Widrow, P. E. Mantey, L. J. Griffiths, and B. B. Goode, "Adaptive antenna systems," *Proc. IEEE*, vol. 55, no. 12, pp. 2143–2159, Dec. 1967.
[18] E. Ferrara and B. Widrow, "Multichannel adaptive filtering for signal enhancement," *IEEE Trans. Circuits Syst.*, vol. CS-28, no. 6, pp. 606–610, Jun. 1981.
[19] R. H. Kwong and E. W. Johnston, "A variable step size LMS algorithm," *IEEE Trans. Signal Process.*, vol. 40, no. 7, pp. 1633–1642, Jul. 1992.
[20] S. L. Nissen, R. W. Harris, and K. B. Slade, "Development of speech reception threshold materials for speakers of Taiwan Mandarin," *Int. J. Audiol.*, vol. 46, no. 8, pp. 449–458, Aug. 2007.
[21] D. N. Zotkin and R. Duraiswami, "Accelerated speech source localization via a hierarchical search of steered response power," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 499–508, Sep. 2004.
[22] D. Halupka, A. S. Rabi, P. Aarabi, and A. Sheikholeslami, "Low-power dual-microphone speech enhancement using field programmable gate arrays," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3526–3535, Jul. 2007.
[23] A. Rosado-Munoz, M. Bataller-Mompean, E. Soria-Olivas, C. Scarante, and J. F. Guerrero-Martinez, "FPGA implementation of an adaptive filter robust to impulsive noise: Two approaches," *IEEE Trans. Ind. Electron.*, vol. 58, no. 3, pp. 860–870, Mar. 2011.
[24] D. G. Rao, T. K. Kumar, N. S. Murthy, and A. Vengadarajan, "Novel method of realization of scalable VLSI adaptive digital beamforming architecture for phased array radar," *Int. J. Eng. Res. Develop.*, vol. 12, no. 7, pp. 27–35, 2016.
[25] S. Huang, L. Tian, X. Ma, and Y. Wei, "A reconfigurable sound wave decomposition filterbank for hearing aids based on nonlinear transformation," *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 2, pp. 487–496, Apr. 2016.
[26] J. Traa, D. Wingate, N. D. Stein, and P. Smaragdis, "Robust source localization and enhancement with a probabilistic steered response power model," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 3, pp. 493–503, Mar. 2016.

BOR-SHING LIN (M'11) received the B.S. degree in electrical engineering from National Cheng Kung University, Taiwan, in 1997, and the M.S. and Ph.D. degrees in electrical engineering from National Taiwan University, Taiwan, in 1999 and 2006, respectively. He is currently an Associate Professor with the Department of Computer Science and Information Engineering, National Taipei University, Taiwan. His research interests include wearable device, the IoT, biomedical signal processing, and rehabilitation engineering.



PO-YU YANG is currently pursuing the master's degree with the Institute of Imaging and Biomedical Photonics, National Chiao Tung University, Taiwan. His research interests include biomedical circuits and systems, biomedical signal processing, and biomedical image processing.

CHING-FENG LIU is currently with the Department of Medical Research, Chi Mei Medical Center, Tainan, Taiwan, and with the Graduate Institute of Medical Sciences, Chang Jung Christian University, Tainan.



YI-CHIA HUANG is currently pursuing the master's degree with the Institute of Imaging and Biomedical Photonics, National Chiao Tung University, Taiwan. His research interests include biomedical circuits and systems, biomedical signal processing, and biomedical image processing.

CHENGYU LIU received the B.S. and Ph.D. degrees in biomedical engineering from Shandong University, China, in 2005 and 2010, respectively. He has completed a Postdoctoral Training at Shandong University, from 2010 to 2013, Newcastle University, U.K., from 2013 to 2014, and Emory University, USA, from 2015 to 2017. He is currently a Professor with the School of Instrument Science and Engineering, Southeast University, China. He is also the Director of the Southeast-Lenovo Wearable Heart-Sleep-Emotion Intelligent Monitoring Laboratory. His research interests include mHealth and intelligent monitoring, machine learning and big data processing for cardiovascular signals, device development for CADs, and sleep and emotion monitoring.

BOR-SHYH LIN (M'02–SM'15) received the B.S. degree from National Chiao Tung University, Hsinchu, Taiwan, in 1997, and the M.S. and Ph.D. degrees from the Institute of Electrical Engineering, National Taiwan University, Taipei, Taiwan, in 1999 and 2006, respectively. He is currently a Professor with the Institute of Imaging and Biomedical Photonics, National Chiao Tung University. His current research interests include biomedical circuits and systems, biomedical signal processing, and biosensor.