



Detection and direction estimation of a sudden loud sound for the hearing assistive eyeglasses

Ki-Won KIM¹; Jung-Woo CHOI²; Yang-Hann KIM³

¹ Korea Advanced Institute of Science and Technology, South Korea

² Korea Advanced Institute of Science and Technology, South Korea

³ Korea Advanced Institute of Science and Technology, South Korea

ABSTRACT

An assistive device for the hearing-impaired is proposed to detect and visually display the direction of incidence when a sudden loud sound occurs out of sight. The result is intuitively presented in four angular regions, such as the front, back, left, right regions, through the three procedures; decision of the sound occurrence, reduction of reflections, and direction estimation. Basic concepts of the direction estimation is to use the four directional microphones steering toward corresponding angular regions, and each directional microphone is realized by the array signal processing based on the modified LCMV beamformer. The result is obtained by comparing the beamformer output levels, and leads our attention to corresponding region using LEDs only when a sudden loud sound is detected. The detection is performed using the change of the overall sound pressure level with respect to time as a feature, and its performance and latency are affected by the length of temporal integration. To keep the performance with short latency, an amplitude follower is applied.

Keywords: Hearing assistive eyeglasses, Detection and direction estimation
I-INCE Classification of Subjects Number(s): 74.7

1. INTRODUCTION

Hearing has an important role in recognition of the surrounding circumstances such as when a dangerous situation has arisen out of one's sight with an altering sound; sounds such as horn, siren, gunshot, scream, and so on. The hearing impaired persons, who cannot hear but can see, can be confronted with difficulties due to their disability in hearing, and an assistive device is a possible means to substitute their hearing. There are several researches and assistive devices for the hearing-impaired. For example, an electronic system is suggested by Damper and Evans(1) to inform the occurrence of household sounds such as doorbells, telephones, or gas alarms. Ho-Ching et al.(2) attempted to deliver sound information visually by a spectrographic scheme with positional ripples. Their method displays the position, pitch, and volume of the sound. Azar et al.(3) presented different visual displays fused into a single program to enhance the user's awareness of his or her surroundings. Although all these sound visualization methods identify the location of a sound source, they only do so in special spaces where microphone arrays are installed. The systems proposed in these studies are also non-portable; hence, they cannot be used when deaf individuals are outdoors. If an assistive device quickly indicates where an alerting sound occurs, deaf individuals can check their circumstances by turn their head toward indicated direction (Figure 1). In this case, both sound detection and recognition procedures are required to indicate only important information; the results are visually displayed only for important sounds, whereas it does not work for unimportant sounds or background noises. At this time, considering that most alerting sounds cause large changes in sound pressure level (SPL) within a very short time in general, the occurrence of such an alerting sound can be detectable by only sensing the change of SPL without recognition process.

Thus we aim to develop an assistive device which can detect and visually display the incident

¹ kkw5581@kaist.ac.kr

² khepera@kaist.ac.kr

³ yanghannkim@kaist.ac.kr

direction when a sudden loud sound occurs out of one's sight. To be of practical use (4), such a device should be portable and wearable without any unwanted attention to the user. In those respects, an eyeglasses-type device is a good candidate, and it has an additional benefit that the user can see the result without any unnatural attachments and behaviors. In addition, it should sufficiently help the person do the intended task in all environments from anechoic to reverberant fields.

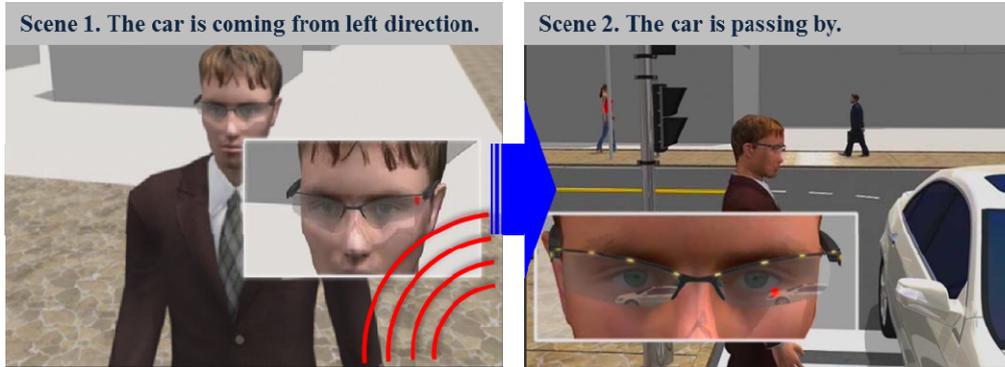


Figure 1 – The direction of the sound source is visually indicated when a car honk a horn out of sight (left), and then the hearing-impaired notice it by turning his head toward the indicated direction (right).

2. PROBLEM DEFINITION

Replacing the disability in hearing, the assistive device helps them recognize their surroundings by rapidly indicating the direction of a sound source when there exists an alerting sound such as a car horn. In many cases, such an alerting sound causes large changes in sound pressure level (SPL) more than 15dB within time duration of 20ms for effective warning (5), and thus we aim to detect and estimate that sound with short latency. The latency is defined as the time delay required to compute the result, and thus reducing such latency is an essential point for an assistive device to rapidly recognize their surroundings and properly react.

To realize the proposed device, two main processes are needed: the detection and the direction estimation of a sudden loud sound. As shown in figure 2, sound pressure signals are measured by the microphones attached on the surface of the eyeglasses, and there are multiple reflections as well as the scattered wave due to the user's head. As the existence of multiple reflections can be a main cause of error in the direction estimation process, an additional process to reduce the reflections should be included. In the detection process, we aim to detect the time frame including the onset for a sudden loud sound and to reduce the reflections in the detected frame. We assume that a sound source is located far enough from the device compared to the Rayleigh distance. This assumption is justified because the glasses are mainly for alerting the wearer of the location of potentially harmful sounds. For direction estimation, we need to define what we mean by “front,” “back,” “left,” and “right” in terms of the azimuth angle φ . When the wearer turns his or her gaze toward the direction where a sudden loud sound occurs, the corresponding region should be included inside the wearer's field of view to understand the situation. A typical human has a field of view of roughly 120° , and this allows us to define the four regions: “front” ($-45^\circ \leq \varphi \leq 45^\circ$), “left” ($45^\circ \leq \varphi \leq 150^\circ$), “back” ($150^\circ \leq \varphi \leq 210^\circ$), and “right” ($210^\circ \leq \varphi \leq 315^\circ$). Once the direction of a sound source is estimated, we can determine the correct region, and the device illuminate the LED corresponding to that region. Therefore, our problem is to accurately estimate the incident region of a sound source with short latency, and we begin by investigating candidate techniques that can detect and estimate the direction of a sound source when a sudden loud sound occurs. Then, the measured pressure signals on the attached microphone array is

$$\bar{p}(\varphi_s, t) = [p(\varphi_s, \bar{r}_1, t) \ \cdots \ p(\varphi_s, \bar{r}_M, t)]^T, \quad (1)$$

where \bar{r}_m is the position of m -th microphone and φ_s is an incident angle of such a sudden loud sound.

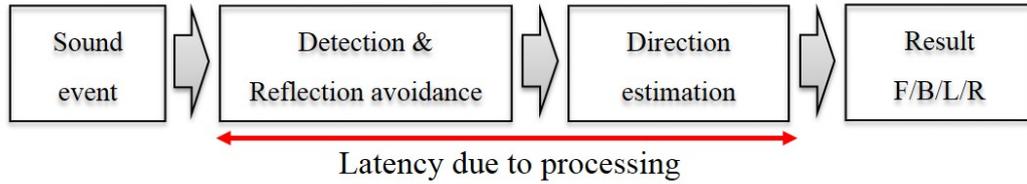


Figure 2 – Procedures for the realization of the proposed assistive device.

3. DETECTION OF A SUDDEN LOUD SOUND

3.1 Objective

Generally, the goal of the detection process is to make a decision when an event of interest occurs. Especially in the case of sound signals, detection algorithms are mainly developed in the field of speech recognition, and pitch or spectral characteristics(6, 7), or linear prediction coefficient(8, 9) are used as features to detect. Unlike methods limited only to voice recognition, the energy based detection algorithms are useful for more general cases. The end-point detection(10) is a widely used pre-processing method to extract the proper section for recognition, but does not reflect changes in the background noise. The median filtering technique(11, 12) is introduced to decide the occurrence of an impulsive sound in slowly varying background noise, but is restricted to only impulsive sounds. In our case, considering that most alerting sounds involved in dangerous situations should be sufficiently loud compared with background noises in order to efficiently notify the situation to people, an event of interest can be defined as a sudden loud sound which causes large changes in SPL more than specific threshold of 10 dB during 20 ms (5). Hence, the detection process aims to make a decision when such a sudden loud sound occurs, and to provide a proper time frame for localization of the sound source including every onset measured at each microphone. In this sense, the change of SPL during ΔT can be a feature to detect the onset when a sudden loud sound occurs. This SPL change with respect to time can be re-expressed as the ratio $r(t)$ between the current and the earlier signal power in linear scale like following equation (2):

$$r(t) = \frac{pow(t)}{pow(t - \Delta T)}. \quad (2)$$

In equation(2), the current signal power $pow(t)$ means the averaged signal power during T_w and is expressed as

$$pow(t) = \frac{1}{T_w} \int_{t-T_w}^t p^2(t) dt, \quad (3)$$

where $p(t)$ is the measured sound pressure signal from one of the attached microphones. As mentioned in equation(1), the background noise, as its level varies slowly with respect to time or environment, is also included in the measured sound pressure. At this time, we can easily expect that little change in $r(t)$ can be observed with respect to time when there are background noises only. On the other hand, if a sudden loud sound occurs, there will be a large change in $r(t)$, and therefore a threshold r_{thr} has to be introduced to discriminate whether the change of $r(t)$ with respect to time is caused by a sudden loud sound or not. Then, decision of the occurrence $d(t)$ is defined as

$$d(t) = \begin{cases} 1, & r(t) \geq r_{thr} \\ 0, & otherwise \end{cases}. \quad (4)$$

3.2 Occurrence decision and selection of the time frame for localization

The maximum latency in detection process corresponds to the time difference ΔT . At this time, the temporal changes in both the signal power $pow(t)$ and the power ratio $r(t)$ are influenced by the

length of the temporal average T_w as defined in equation (3), and thus the detection performance can be affected by the selection of T_w . If the length of temporal average T_w is larger than the time difference, ΔT , the power ratio cannot reach its maximum value because the slow increasing rate owing to the large T_w makes it difficult for the signal power to increase sufficiently within ΔT as shown in figure 3(a). Also, in the case of the selection of small length of T_w (for example, $T_w \rightarrow 0$), the detailed temporal characteristic of the measured pressure signal can be reflected in the signal power $pow(t)$, and cause two types of decision error; one is “false detection” which misjudge even though any sudden loud sound did not occur, and the other one is multiple detection of one sound event as shown in figure 3(b). In this paper, we have selected the length of the temporal average T_w as its upper bound of 20msec, and also used the envelope of $pow(t)$ instead of $pow(t)$ in order to see the tendency only, not the detailed temporal characteristic.

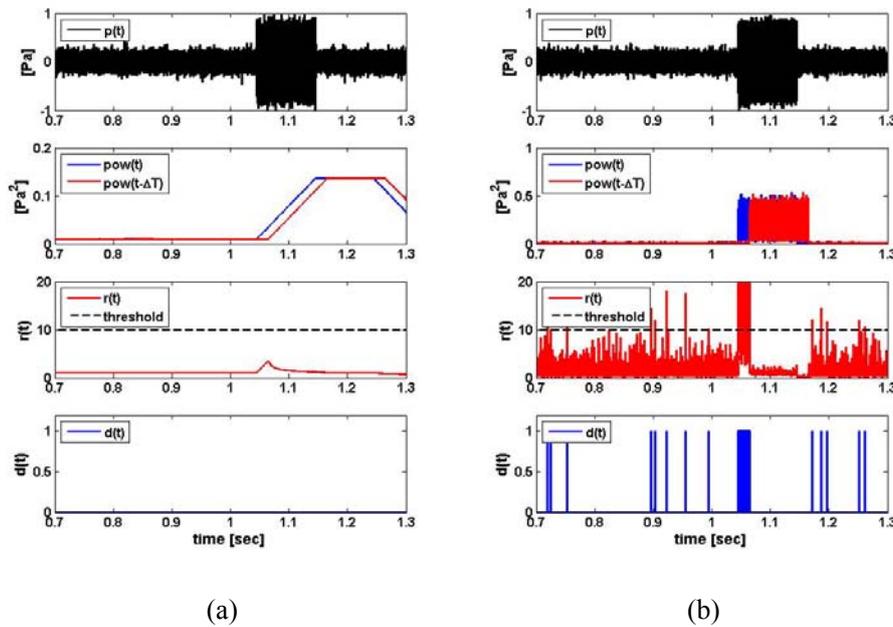


Figure 3. The effect of the length of the temporal average T_w on detection process when a sinusoidal sound(1kHz) is generated from a source during 100ms. (a) The case of large T_w ($T_w=200$ ms), (b) the case of small T_w ($T_w=0.2$ ms).

The current value of envelope $env(t)$ for the current signal power $pow(t)$ can be calculated in real-time by using the amplitude follower(13), and it can be mathematically defined as

$$env(t) = (1 - g) \cdot pow(t) + g \cdot env(t - \Delta t), \quad (5)$$

where Δt is the sampling period in time and g is a weighting. As expressed in equation (5), the amplitude follower requires the current signal power $pow(t)$ and the feedback of envelope signal $env(t - \Delta t)$ to calculate the current value of envelope $env(t)$. Particularly, the weighting g determines the shape of the envelope, and is selectively applied in accordance with the condition below:

$$g = \begin{cases} g_a, & pow(t) \geq env(t - \Delta t) \\ g_r = e^{-\Delta t / \tau_d}, & otherwise \end{cases} \quad (6)$$

Looking at the impulse response of equation (5) as shown in figure 4, if the current signal power $pow(t)$ is larger than the feedback of envelope $env(t - \Delta t)$, the weighting is selected to small value of g_a to rapidly follow the change of $pow(t)$. Otherwise, by selecting the value of g_r , $env(t)$ is exponentially decaying by increasing the weight of the feedback $env(t - \Delta t)$. At this time, we can adjust the rate of decaying by using the time constant τ . In this paper, we have selected $g_a = 0$

and $\tau = 20$ msec as the values of weighting.

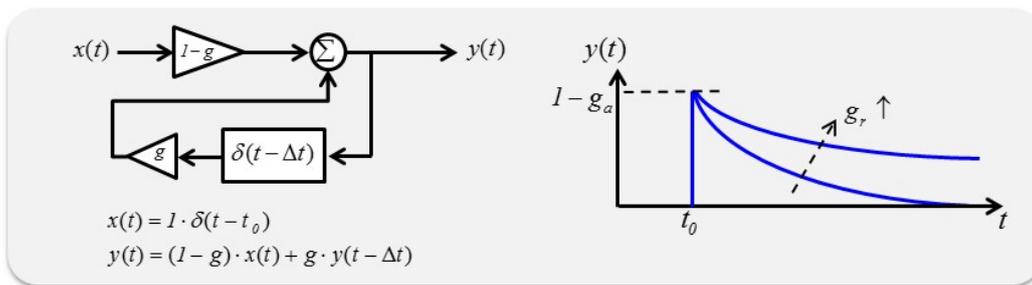


Figure 4 – Block diagram of the amplitude follower and its impulse response.

After we obtain the envelope of the signal power, $env(t)$, the power ratio can be re-defined as equation(7) instead of equation(2),

$$r(t) = \frac{env(t)}{env(t - \Delta T)}, \tag{7}$$

and then the occurrence of a sudden loud sound is determined as explained in equation(4).

As illustrated in figure 5, the direction estimation of the sound source will start when the sudden loud sound is no longer detected (in the signal processing perspective, the moment that $d(t)$ changes from 1 to 0). At this time t , a proper time frame has to be obtained as following equation (8) to estimate where the sound occurs.

$$p_F(\bar{r}_m, t') = p(\bar{r}_m, t'), \quad t - T < t' \leq t \tag{8}$$

An important point is that the obtained time frame should include the onsets of the sound measured at each microphone because several pieces of information required to localize the sound source can be obtained from the onsets. In order to include every onset measured in each microphones, the length of time frame T should be longer than the sum of ΔT and the maximum time difference of arrival between each microphone, and we have determined T as twice of ΔT .

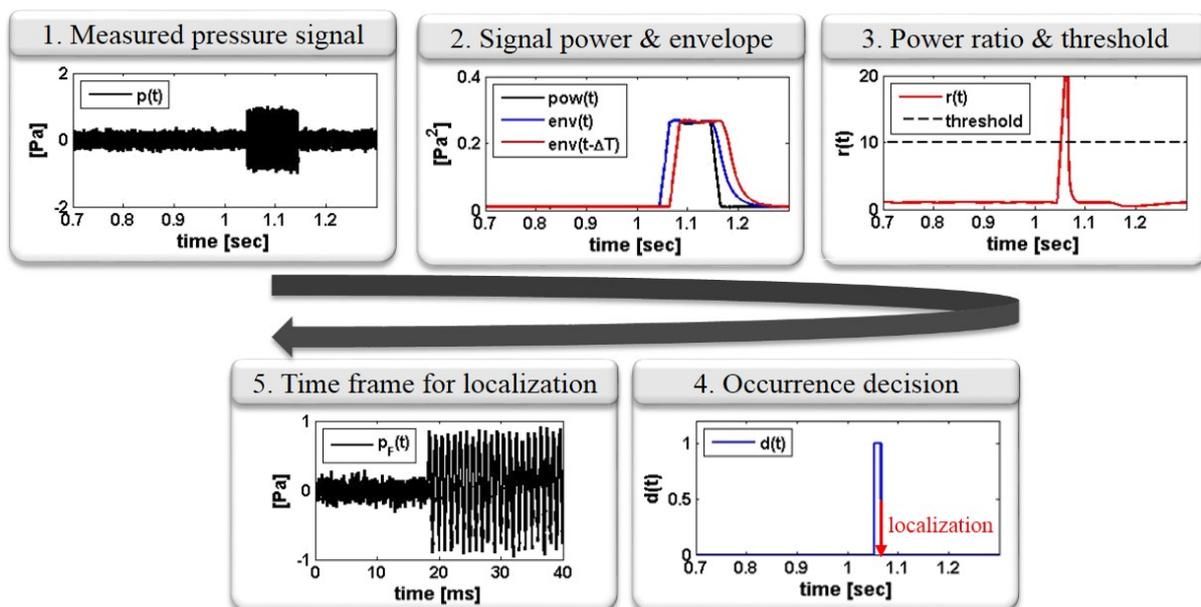


Figure 5 – Procedure for the detection of a sudden loud sound using the amplitude follower and the selection of a proper time frame.

3.3 Window for reflection reduction

As shown in equation(1), several reflected wave components can still be included in the time frame of equation(8), and those reflections can be a main reason for the occurrence of localization errors in real environment. In this paper, we suggest a window function to extract the first-arriving sound component by reducing the reflections in time domain based on the precedence effect(14, 15) of the human hearing system. The window function $w(\bar{r}_m, t)$ for each microphone with respect to time can be designed by using the amplitude follower, and described as

$$w(\bar{r}_m, t') = 1 - \frac{y(\bar{r}_m, t' - \tau)}{\max\{y(\bar{r}_m, t' - \tau)\}}, \quad (9)$$

where τ is the time difference between the first-arriving direct sound and the first-arriving reflected sound, and $y(\bar{r}_m, t' - \tau)$ is the output signal of the amplitude follower, which can be calculated as following:

$$y(\bar{r}_m, t' - \tau) = (1 - g) \cdot |p_F(\bar{r}_m, t' - \tau)| + g \cdot y(\bar{r}_m, t' - \Delta t')$$

$$\text{where } g = \begin{cases} g_a = 0, & |p_F(\bar{r}_m, t' - \tau)| \geq y(\bar{r}_m, t' - \Delta t') \\ g_r = 1, & \text{otherwise} \end{cases} . \quad (10)$$

As shown in figure 6, it is possible to reduce the reflected wave components by multiplying the obtained window with the sound pressure signal as expressed as

$$p_{F,w}(\bar{r}_m, t') = p_F(\bar{r}_m, t') \times w(\bar{r}_m, t'), \quad t - T < t' \leq t, \quad (11)$$

and the subscript w means that the window function is applied.

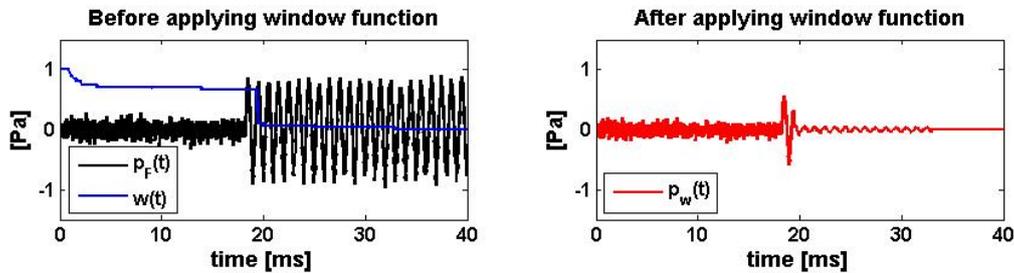


Figure 6 – An abstract illustration of window function design and reflection reduction process.

4. DIRECTION ESTIMATION

When a sudden loud sound is detected and the reflection reduced time frame for localization is given, then the device should indicate where the sound comes from by using LEDs that correspond to front, back, left, and right, as mentioned in chapter 2. The simplest way to achieve this purpose is to use four directional microphones corresponding to the above 4 regions, respectively. Then, the angular region where the sound source is located can be estimated by comparing the signal powers measured at the four directional microphones as shown in figure 7. In this case, the ideal response of a directional microphone is to have flat response in the range of corresponding region, and zero response otherwise. But since it is not easy to physically adjust the response of the directional microphone, a signal processing approach to adjusting directional response can be an alternative way to accomplish our goal. Various kinds of beamformers, the optimal beamformer such as the minimum variance(MV) beamformer(16, 17) or the generalized linear constrained distortionless(LCMV) beamformer, are typical examples. Also, approaches for the regional control

of the directional response are suggested to maximize the ratio of the averaged beamforming power between in specified angular range and otherwise(18). Our objective was to design a fixed beamformer, not an adaptive one, which has a flat unit response in the range of its corresponding angular region and minimum response otherwise. The beamformer also have constant beamwidth with respect to frequency. The output of a beamformer with respect to the incident angle φ_s is generally defined as the sum of the weighted pressure as in the following equation,

$$B(\varphi_s, f) = \sum_{m=1}^M W_m(f) P(\varphi_s, \bar{r}_m, f), \quad (12)$$

and the beampattern, defined as the squared magnitude of the beamformer output, is expressed as

$$\begin{aligned} BP(\varphi_s, f) &= \left| \sum_{m=1}^M W_m(f) P(\varphi_s, \bar{r}_m, f) \right|^2 = \left| \bar{W}(f)^H \bar{P}(\varphi_s, f) \right|^2 \\ &= \bar{W}(f)^H \bar{P}(\varphi_s, f) \bar{P}(\varphi_s, f)^H \bar{W}(f) \\ &= \bar{W}(f)^H \mathbf{R}(\varphi_s, f) \bar{W}(f) \end{aligned}, \quad (13)$$

where \mathbf{R} is $M \times M$ spatial correlation matrix composed of $\mathbf{R}(m, n) = P(\varphi_s, \bar{r}_m, f) \cdot P(\varphi_s, \bar{r}_n, f)^*$. Then, tried to realize the ideal beampattern as shown in figure 7(a) by using the concept of LCMV beamformer. However, the solution based on the original LCMV is dependent on \mathbf{R} as an adaptive method. Thus, we used the spatially averaged correlation matrix $\bar{\mathbf{R}}$ instead of \mathbf{R} in order to avoid the \mathbf{R} dependency of the original LCMV formulation.

$$\begin{aligned} \text{minimize } & \bar{W}(f)^H \bar{\mathbf{R}}(f) \bar{W}(f) \text{ such that } \bar{W}(f)^H \mathbf{C} = \bar{\mathbf{g}}^H \\ \text{where } & \bar{\mathbf{R}}(f) = \frac{1}{2\pi} \oint \mathbf{R}(\varphi_s, f) d\varphi_s \end{aligned}, \quad (14)$$

The objective function of equation(14) includes the following four linearly combined constraints: three distortionless constraints at $\varphi_s = \varphi_C$, $\varphi_s = \varphi_L$ and $\varphi_s = \varphi_H$, and one zero-gradient constraint at $\varphi_s = \varphi_C$ as shown in figure 7(b). These constraints can be mathematically expressed as

$$\begin{aligned} \mathbf{C} &= \begin{bmatrix} \bar{A}(\varphi_L, f) & \bar{A}(\varphi_H, f) & \frac{\partial \bar{A}(\varphi_C, f)}{\partial \varphi_s} \end{bmatrix}, \\ \bar{\mathbf{g}} &= [1 \quad 1 \quad 0]^T \end{aligned}, \quad (15)$$

where \bar{A} means the array function which is determined by the wave propagation in the array. Several angles φ_C , φ_L and φ_H are the center, low bound and upper bound of one angular region, respectively. Then, the weighting function can be expressed as

$$\bar{W}(f)^H = \bar{\mathbf{g}}^H [\mathbf{C}^H \bar{\mathbf{R}} \mathbf{C}]^{-1} \mathbf{C}^H \bar{\mathbf{R}}^{-1}, \quad (16)$$

and the resultant beampatterns for each four angular region are shown in figure 8.

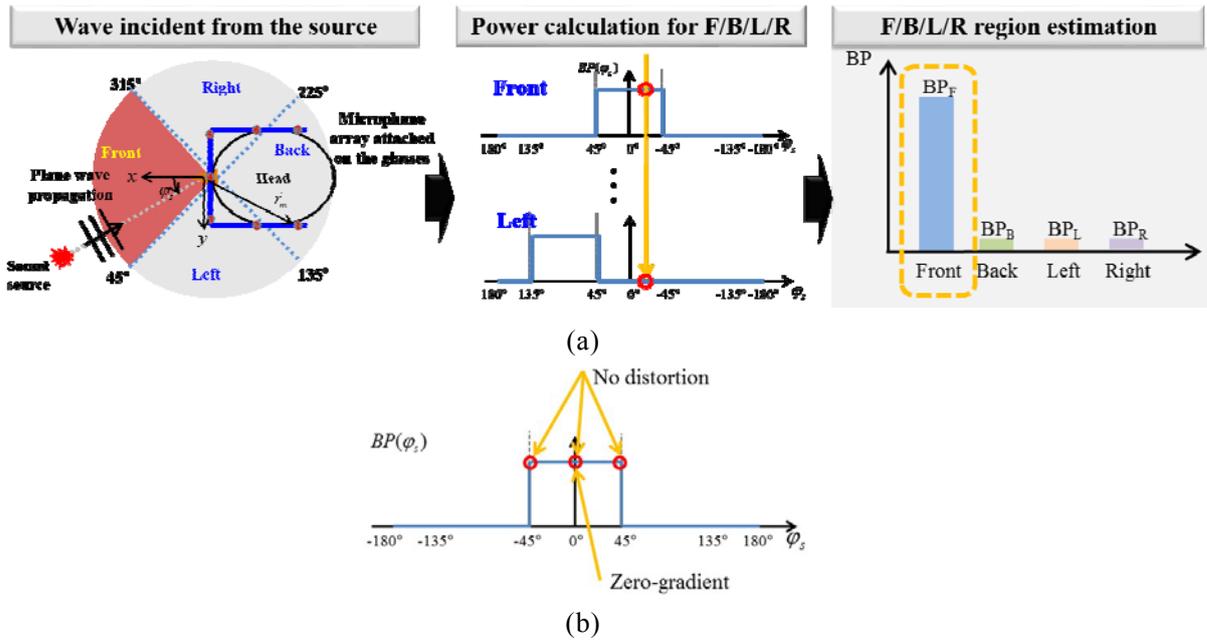


Figure 7 – The concept of F/B/L/R angular region estimation using the fixed beamformer. (a) Estimation procedure, (b) constraints in beamformer design for flat response in the angular range of interest.

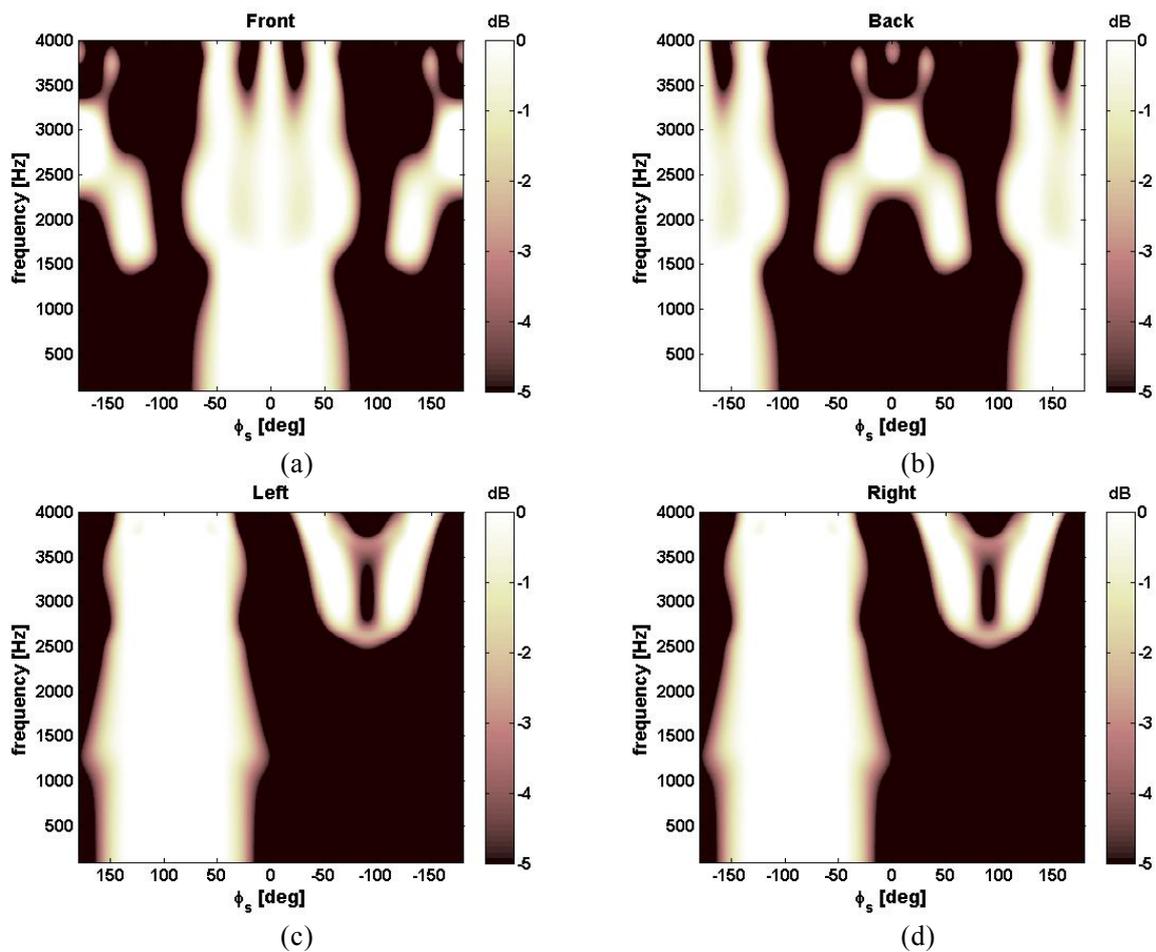


Figure 8 – The result of beampattern for each angular region using 7-microphones attached on the surface of the eyeglasses: (a) front, (b) back, (c) left, and (d) right.

5. CONCLUSIONS

In this research, we aimed to develop an assistive device indicating the direction of incidence when a sudden loud sound occurs for the hearing-impaired, and the device intuitively present the direction in four angular regions; the front/back, left, right regions. To realize this device, we defined three main procedures such as decision of the sound occurrence, reduction of reflections, and direction estimation. In the first procedure, the signal power ratio, the ratio between of the current signal power and previously obtained one with time interval, ΔT , was used to determine the occurrence of the sound, and signal power was replaced into its envelope to improve the penalties due to the short length of the temporal average. Also, we proposed a window function using the amplitude follower in order to reduce the reflections in time frame for localization. Then, four directional microphones steering toward corresponding angular regions were suggested as the simplest way to estimate the direction of incidence, and a modified LCMV method was tried to realize the ideal directivity. But there are several problems in the resultant beamformer that the flatness is broken and aliasing issue is exposed in more than about 3kHz. Therefore, there still remains further studies such as microphone placement for the purpose of anti-aliasing and modification or suggestion of an objective function for an ideal beam pattern.

ACKNOWLEDGEMENTS

This work was supported by the Ministry of Trade, Industry and Energy (MOTIE) grant funded by the Korea government (No. 10037244), and the BK21 (Brain Korea 21) project initiated by the Ministry of Education, and Unmanned Technology Research Center (UTRC) at Korea Advanced Institute of Science and Technology (KAIST), originally funded by DAPA, ADD.

REFERENCES

1. Damper RI, Evans M. A multifunction domestic alert system for the deaf-blind. *Rehabilitation Engineering, IEEE Transactions on*. 1995;3(4):354-9.
2. Ho-Ching F, Mankoff J, Landay JA, editors. Can you see what i hear?: the design and evaluation of a peripheral sound display for the deaf. *Proceedings of the SIGCHI conference on Human factors in computing systems*; 2003: ACM.
3. Azar J, Saleh H, Al-Alaoui M. Sound visualization for the hearing impaired. *International Journal of Emerging Technologies in Learning (iJET)*. 2007;2(1):1-7.
4. MANN WC. SMART TECHNOLOGY FOR AGING, DISABILITY, AND INDEPENDENCE. 2005.
5. Patterson R, Mayfield T. Auditory warning sounds in the work environment [and Discussion]. *Philosophical Transactions of the Royal Society of London B, Biological Sciences*. 1990;327(1241):485-92.
6. Marzinzik M, Kollmeier B. Speech pause detection for noise spectrum estimation by tracking power envelope dynamics. *Speech and Audio Processing, IEEE Transactions on*. 2002;10(2):109-18.
7. Gazor S, Zhang W. A soft voice activity detector based on a Laplacian-Gaussian model. *Speech and Audio Processing, IEEE Transactions on*. 2003;11(5):498-505.
8. Tanyer SG, Ozer H. Voice activity detection in nonstationary noise. *IEEE Transactions on Speech and Audio Processing*. 2000;8(4):478-82.
9. Zhang T, Kuo C-C. Audio content analysis for online audiovisual data segmentation and classification. *Speech and Audio Processing, IEEE Transactions on*. 2001;9(4):441-57.
10. Li Q, Zheng J, Tsai A, Zhou Q. Robust endpoint detection and energy normalization for real-time speech and speaker recognition. *Speech and Audio Processing, IEEE Transactions on*. 2002;10(3):146-57.
11. Dufaux A, Besacier L, Ansorge M, Pellandini F, editors. Automatic sound detection and recognition for noisy environment. *Proc of the X European Signal Processing Conference*; 2000: Citeseer.
12. Dufaux A. Detection and recognition of impulsive sounds signals. *Institute de Microtechnique Neuchatel, Switzerland*. 2001.
13. McNally GW. Dynamic range control of digital audio signals. *Journal of the Audio Engineering Society*. 1984;32(5):316-27.
14. Wallach H, Newman EB, Rosenzweig MR. A Precedence Effect in Sound Localization. *The Journal of the Acoustical Society of America*. 1949;21(4):468-.

15. Litovsky RY, Colburn HS, Yost WA, Guzman SJ. The precedence effect. The Journal of the Acoustical Society of America. 1999;106(4):1633-54.
16. Lacos RT. Data adaptive spectral analysis methods. Geophysics. 1971;36(4):661-75.
17. Capon J. High-resolution frequency-wavenumber spectrum analysis. Proceedings of the IEEE. 1969;57(8):1408-18.
18. Choi J-W, Kim Y-H. Generation of an acoustically bright zone with an illuminated region using multiple sources. Journal of the Acoustical Society of America. 2002;111(4):1695-700.

[APPENDIX] HARDWARE IMPLEMENTATION

The proposed assistive device is composed of two main parts as shown in figure 9: one is the glasses that sensors and LEDs are attached on, and the other one is a processing unit for estimating the direction of sound source. Microelectromechanical systems (MEMS) microphones (Analog devices, ADMP401, 8 mV/Pa) are selected in the present work since their low power consumption (825 μ W) and small size (4.72 mm \times 3.76 mm \times 1.0 mm) characteristics are suitable for such portable device, and total seven microphones are attached to the inner surface of the eyeglasses with equal spacing of 6 cm. Furthermore, the inner surface of the glasses was equipped with four light-emitting diodes (LEDs) to indicate the direction of a sound source, and each LED corresponds to the direction of left, right, front, and back. The sound pressure signals are measured by microphones attached on the eyeglasses, and acquired by an AD converter included in the selected processing module (TMS320F28335, detailed specification is shown in table 1), and then the direction of a sound source is estimated. A signal was then generated using the algorithm and fed into the corresponding LED. Finally, a rechargeable Lithium-Polymer battery is used to operate the device.

Table 1 – Specification of DSP module (TMS320F28335)

Computing power	150 MMACs
Memory	68 kbps (RAM) / 512 kbps (Flash)
ADC	16 channels / 12 bit / 12.5 MHz

Sound-visualization glasses



Processing units



TMS320F28335 processor module
Rechargeable Li-Po battery

Figure 9 – Composition of the assistive device