IEEK Transactions on Smart Processing and Computing

A Beamforming-Based Video-Zoom Driven Audio-Zoom Algorithm for Portable Digital Imaging Devices

Nam In Park[1], Seon Man Kim[1], Hong Kook Kim[1], Myeong Bo Kim[2], and Sang Ryong Kim[3]

[1] School of Information and Communications, Gwangju Institute of Science and Technology (GIST), 1 Oryong-dong, Buk-gu, Gwangju 500-712, Korea {naminpark, kobem30002, hongkook}@gist.ac.kr
[2] Digital Media Business, Samsung Electronics, 416 Meatan 3-dong, Yeongtong-gu, Suwon-si, Gyeonggi-do 443-742, Korea kmbo.kim@samsung.com
[3] Ministry of Science, ICT and Future Planning, 14th Floor, IT Venture Tower, 78 Garak-dong, Songpa-gu, Seoul 138-803, Korea srkim@msip.go.kr

* Corresponding Author: Hong Kook Kim

Received January 8, 2013; Revised February 2, 2013; Accepted February 11, 2013; Published February 28, 2013

Abstract: A video-zoom driven audio-zoom algorithm is proposed to provide audio zooming effects according to the degree of video-zoom. The proposed algorithm is designed based on a super-directive beamformer operating with a 4-channel microphone array in conjunction with a soft masking process that uses the phase differences between microphones. The audio-zoom processed signal is obtained by multiplying the audio gain derived from the video-zoom level by the masked signal. The proposed algorithm is then implemented on a portable digital imaging device with a clock speed of 600 MHz after different levels of optimization, such as algorithmic level, C-code and memory optimization. As a result, the processing time of the proposed audio-zoom algorithm occupies 14.6% or less of the clock speed of the device. The performance evaluation conducted in a semi-anechoic chamber shows that the signals from the front direction can be amplified by approximately 10 dB compared to the other directions.

Keywords: Audio processing, Audio-zoom, Sound focusing, Beamforming, Digital imaging devices

1. Introduction

Most portable digital imaging devices on the market have video zoom capability, which is typically achieved by either optical or digital zoom [1]. In terms of use, video zoom aims to bring clarity when viewing objects as they move in or out of a video viewfinder. On the other hand, video zoom can enhance the user's satisfaction of movement more realistically if the audio contents can also be zoomed in conjunction with the video zoom. In other words, because the video zoom only reflects the images of the objects, original audio signals from the objects are currently recorded, regardless of the video-zoom level. If the same degree of video zoom is maintained with the audio content, the audio signal should be amplified as objects come closer or reduced as objects move further away, which is referred to as *video-zoom driven audio-zoom*.

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the MSIP (Ministry of Science, ICT&Future Planning) (No. 2012-010636).

This paper proposes a video-zoom driven audio-zoom algorithm to be implemented on a portable digital imaging device. In particular, the proposed audio-zoom algorithm was designed based on super-directive beamforming [2], which operates with a 4-channel microphone array. According to [1], increasing the front sound level is sufficient to express the audio zoom effect. Therefore, the front sound level is controlled by an audio gain estimated from the phase difference between the signals reconstructed by super-directive beamforming. To realize the proposed algorithm in real time while recording with a portable digital imaging device, several different levels of optimization were performed, which included the algorithmic level, C-code level and memory optimization. The performance of the proposed video-zoom driven audio-zoom algorithm was then evaluated by comparing the sound pressure level (SPL) of the original audio signal with that of the processed audio signal for each different direction (front, both sides, and back) in a semi-anechoic chamber.

The remainder of this paper is organized as follows.

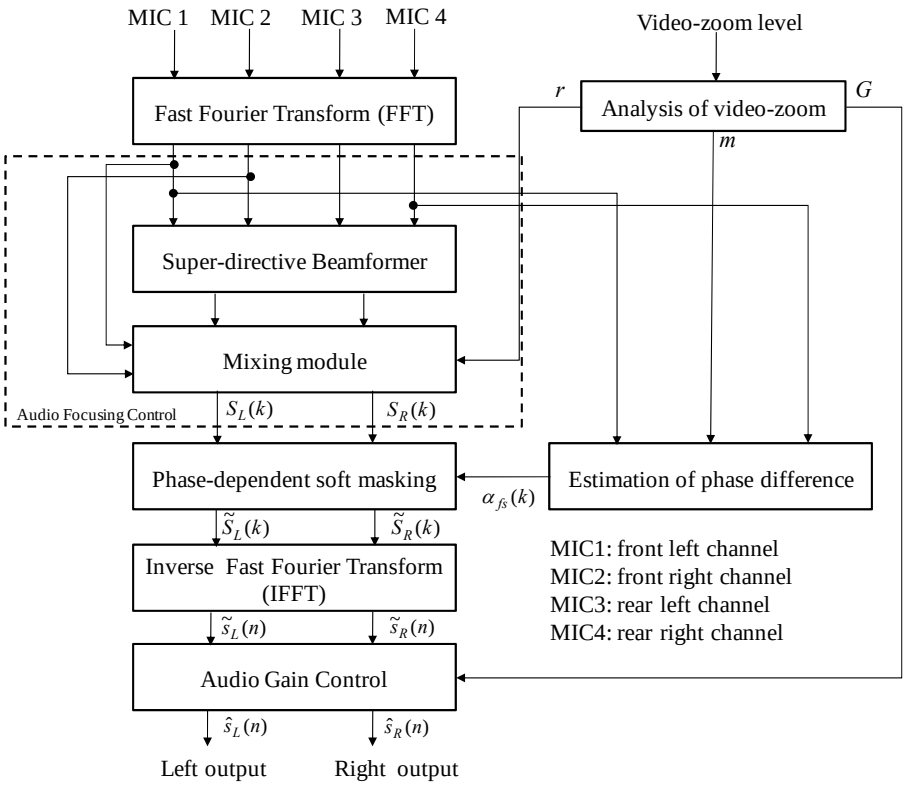


Fig. 1. Procedure of the proposed video-zoom driven audio-zoom algorithm.

Section 2 proposes a video-zoom driven audio-zoom algorithm. Section 3 addresses the implementation issues of the proposed video-zoom driven audio-zoom algorithm on a portable digital imaging device with limited resources. Section 4 reports the performance of the implemented video-zoom driven audio-zoom algorithm, and the paper is concluded in Section 5.

2. Proposed Video-zoom Driven Audio-zoom Algorithm

The proposed video-zoom driven audio-zoom algorithm was designed using a 4-channel microphone array. As shown in Fig. 1, the proposed algorithm consisted of three parts: an audio focus control using a super-directive beamformer, a soft masking process based on the phase difference, and an audio gain control based on the video-zoom level.

First, the input signals from the 4-channel microphone array were transformed into a frequency domain using a fast Fourier transform (FFT), followed by the application of a fixed super-directive beamformer [2, 3]. The signals processed by the beamformer were then mixed with the original signals according to a mixing ratio parameter *r*, which was determined by the video-zoom level. In other words, *r* decreases with increasing video-zoom level. Next, soft masking was used to process the mixed signals, where a masking threshold parameter *m* was determined according to the video-zoom level. Specifically, the mixed signals are multiplied by frequency-dependent mask coefficients that were obtained by the phase-dependent soft masking estimation. The masked signals were then transformed back into the time-domain signals using an inverse FFT (IFFT), and an audio gain *G* corresponding to the video-zoom level was used to amplify the time-domain signals. Finally, a sigmoid function was applied to prevent the audio signal from being clipped when the signal was amplified.

2.1 Audio focusing control using a beamformer

The proposed audio-zoom algorithm adjusts the audio zooming effects according to the degree of video zoom, in which the objects appear to come closer or move further away. A super-directive beamformer was designed to provide more realistic audio-zoom effects [2]

$$W = \frac{\Gamma_{vv}^{-1} d}{d^H \Gamma_{vv}^{-1} d} \tag{1}$$

where d is the delay incurred by the distance between the microphones in the frequency domain, Γ_{vv} is the power spectral density matrix of each microphone input signal [3], and W is the optimal weight of the super-directive beamformer. In (1), the power spectral density matrix is defined as

$$\Gamma_{vv} = \begin{pmatrix} 1 & \Gamma_{v_0v_1} & \Gamma_{v_0v_2} & \cdots & \Gamma_{v_0v_{N-1}} \\ \Gamma_{v_1v_0} & 1 & \Gamma_{v_1v_2} & \cdots & \Gamma_{v_1v_{N-1}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Gamma_{v_{N-1}v_0} & \Gamma_{v_{N-1}v_1} & \Gamma_{v_{N-1}v_2} & \cdots & 1 \end{pmatrix} \quad (2)$$

where $\Gamma_{v_mv_n}$ is the power spectral density of the signals between the m^{th} and n^{th} microphones. For a fixed super-directive beamformer, $\Gamma_{v_mv_n}$ is a function of the angular frequency ω, and it is given by [2]

$$\Gamma_{v_mv_n}(\omega) = \sin c\left(\frac{\omega \cdot l_{mn}}{c}\right) \quad (3)$$

where c is the speed of sound and l_{mn} is the distance between the m^{th} and n^{th} microphones. Note that the weighting vector described in (1) can be obtained by the minimum variance distortionless response (MVDR) algorithm [2].

The original signals were then mixed with the signals obtained after applying the fixed super-directive beamformer. Table 1 lists the mixing ratios according to the audio-zoom level. In the table, the audio-zoom level was divided into five levels according to the video-zoom levels, which range from 1 to 10. As the video-zoom level increases, the audio-zoom level also increases. Therefore, the sound from the objects is expected to become louder.

Table 1. Mixing ratio and audio-zoom level according to the different video-zoom levels.

Item \ Video-zoom level	1-2	3-4	5-6	7-8	9-10
Audio-zoom level	1	2	3	4	5
Mixing ratio	1	0.75	0.5	0.25	0

2.2 Phase-dependent soft masking

A soft mask was applied to the mixed signals obtained in Section 2.1 to further suppress the signals from the other directions compared to the front. The mask value for each frequency bin was estimated by comparing the normalized phase difference between the microphone distances with that of the reference signal [4, 5], such that

$$\phi_{1,4}(i,k) = \theta_1(i,k) - \theta_4(i,k) \quad (4)$$

where $\theta_1(i,k)$, $\theta_4(i,k)$, and $\phi_{1,4}(i,k)$ are the phase of the 1^{st} and 4^{th} microphone and the phase difference between the 1^{st} and 4^{th} microphones for the i^{th} frame and k^{th} frequency bin, respectively. Note that $\phi_{1,4}(i,k)$ in (4) can also be represented by the normalized phase difference as

$$\bar{\phi}_{1,4}(i,k) = \phi_{1,4}(i,k) \cdot \frac{c}{2\pi \cdot d \cdot f} \quad (5)$$

where c is the speed of sound, d is the distance between microphones, and f is the corresponding frequency. Using the normalized phase difference, $\bar{\phi}_{1,4}(i,k)$, the masking value becomes

$$\alpha(i,k) = \begin{cases} 1, & if\ \bar{\phi}_{1,4}(i,k) \le \frac{\pi}{2} \\ 0.2, & otherwise \end{cases} \quad (6)$$

where $\alpha(i,k)$ is the masking value of the k^{th} frequency bin at the i^{th} frame. As indicated in (6), assuming that the normalized phase difference $\bar{\phi}_{m,n}(i,k)$ is less than the difference between the reference signals coming from the front direction, the mask value was set to 1. Otherwise, the mask value was set to 0.2 for the maximum video-zoom level. Subsequently, the mask values were multiplied by the mixed signals in the frequency domain, and then smoothed to prevent the signals from fluctuating over the adjacent frequency bins and frames [6]. Here, a smoothing function in the frequency is defined as

$$\alpha_s(i,k) = \tfrac{1}{4}\left(\alpha(i,k-2) + \alpha(i,k-1) + \alpha(i,k) + \alpha(i,k+1)\right). \quad (7)$$

To prevent the signal from fluctuating over the adjacent frames, a smoothing function for the current frame is also defined as

$$\alpha_{fs}(i,k) = \tfrac{1}{4}\alpha_s(i-1,k) + \tfrac{1}{2}\alpha_s(i,k) + \tfrac{1}{4}\alpha_s(i+1,k) \quad (8)$$

where $\alpha_{fs}(i,k)$ is the finally smoothed masking value for the k^{th} frequency bin in the i^{th} frame.

The masked signal $\tilde{S}_i(k)$ was obtained by multiplying the smoothed masking value, $\alpha_{fs}(i,k)$, into the signal, $S_i(k)$ processed by the super-directive beamformer. The phase components of $\tilde{S}_i(k)$ were maintained by those of $S_i(k)$, i.e. $|\tilde{S}_i(k)| = \alpha_{fs}(i,k)|S_i(k)|$ and $\angle\tilde{S}_i(k) = \angle S_i(k)$. Finally, the masked signal, $\tilde{s}(n)$, was obtained by applying an inverse FFT to $\tilde{S}_i(k)$.

2.3 Audio gain control according to video-zoom level

The gain of the masked signals should be controlled to further emphasize the signals coming from the front direction. In particular, the proposed video-zoom driven audio-zoom algorithm was designed to amplify the input signals to 10 dB. This is because the human ear begins to detect an amplitude change of more than approximately 3 dB, and perceives twice the loudness difference if the amplitude is boosted by 10 dB or more [7]. Fig. 2 shows the audio gain curve according to the video-zoom level. As shown in Table 1, there is a relationship between the video-zoom level and audio-zoom level. The maximum audio-zoom level ($20\log_{10}G$) is bounded to 12 dB. As shown in the figure, if the audio-zoom level approaches an audio-zoom level of 2 or 4, the audio gain is set to 4.86 dB or 10.23 dB, respectively. Therefore, the output signal $s_g(n)$ can be defined as

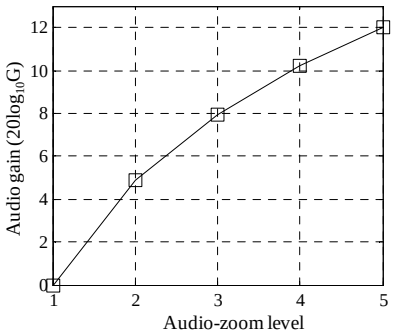


Fig. 2. Audio gain control according to the audio-zoom level.

$$s_g(n) = G \cdot \tilde{s}(n) \tag{9}$$

where G is the audio gain for a given audio-zoom level, as shown in Fig. 2, and $\tilde{s}(n)$ is the masked signal.

On the other hand, the amplified audio signals can be overflown after multiplying the audio gain. Therefore, an audio level control (ALC) based on the sigmoid function is applied as

$$\hat{s}(n) = \begin{cases} s_g(n), & if \ \left|s_g(n)\right| < C \\ \dfrac{(2 \cdot C - 1) \cdot sgn(s_g(n))}{1 + \exp(-\varepsilon \cdot (\left|s_g(n)\right| - C))}, & otherwise \end{cases} \tag{10}$$

where $sgn(x)$ is equal to 1 if x is greater than or equal to 0, otherwise it is equal to -1, and ε and C are set to 0.0001 and 16384, respectively. In other words, $\hat{s}(n)$ is the amplified audio signal obtained after applying a sigmoid function. The magnitude of the amplified signal, $s_g(n)$, is then compared with a predefined value, 16384. If it is indeed greater than 16384, $\hat{s}(n)$ can be obtained by applying the sigmoid function defined in (10).

3. Implementation of the Proposed Audio-Zoom Algorithm on a Portable Digital Imaging Device

This section begins by discussing the implementation issues for achieving real-time operation of the proposed algorithm on a portable digital imaging device equipped with limited resources, where the device has a clock speed and memory size of 600 MHz and 111 MB, respectively. Because the implemented video-zoom driven audio-zoom algorithm should be performed in recording mode, the processing time needs to occupy less than 20% of the system resources, considering the processing time of a video codec. Therefore, the proposed algorithm should be optimized to reduce the computational complexity.

3.1 C-code level optimization

As a first step in optimization, fixed-point arithmetic programming was performed because fixed-point operations are much faster than their corresponding floating-point operations [8]. To this end, most 64-bit or 32-bit floating variables (double, float) are converted to 32-bit or 16-bit integer variables (int, short). Next, a C-code level optimization was carried out, which includes inline functions, intrinsic functions, redundant code removal, reordering operations and loop unrolling. In particular, basic operations were redefined using relatively fast linear assembly instructions [8, 9].

3.2 Algorithmic level optimization

An algorithmic level optimization was carried out to reduce the computation complexity (Fig. 3). In the figure, the audio focusing control module was realized by

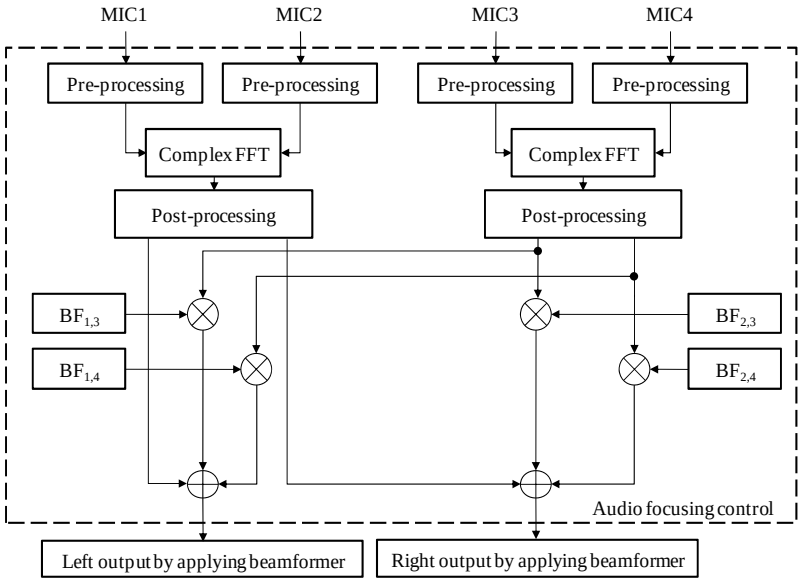


Fig. 3. Illustration of algorithmic level optimization of the proposed algorithm.

applying an FFT to each channel signal followed by performing four different fixed super-directive beamformers in the frequency domain, where the beamformer, $BF_{m,n}$, is designed according to the distance between the m^{th} and n^{th} microphones. By applying an algorithmic level optimization, the computational complexity was reduced by approximately 50% by efficiently dealing with the audio focusing control. That is, the two FFTs for the 2-channel input signals were merged into one complex FFT, which reduces the complexity by approximately 50% [10], as shown in Fig. 3. As a result, because pre-processing and post-processing are composed of adders and shift operations, the structure of the proposed video-zoom driven audio-zoom algorithm has lower complexity than performing four FFTs.

The phase-dependent soft masking requires an arctangent function for each frequency bin to estimate the direction of the audio source. As mentioned above, because a floating-point operation is quite complex, it should first be converted to a fixed-point operation [11].

3.3 Graphical user interface

The graphic user interface (GUI) is an important part of consumer electronics because it provides users with important information and guidelines. Therefore, the GUI for the proposed video-zoom driven audio-zoom algorithm is realized on a portable digital imaging device. Fig. 4(a) shows a snapshot of the main GUI, where the numbers from (i) to (iv) in the top line represent the icons for the imaging mode, operating status, time counter (recording time/remaining recordable time), and storage media, respectively. In addition, the numbers from (v) to (vii) show the image resolution, audio-zoom/volume status, and menu icon, respectively. In this device, the audio-zoom algorithm is activated by touching the 'Zoom MIC' menu, as shown in Fig. 4(b).

Fig. 4. Snapshot of the graphic user interface (GUI) designed for the proposed video-zoom driven audio-zoom algorithm (a) main GUI configuration, (b) audio-zoom mode GUI configuration.

4. Performance Evaluation

The performance of the proposed algorithm implemented on the portable digital imaging device was evaluated in terms of the processing time and relative sound pressure level (SPL) of the audio signals obtained from the front direction.

4.1 Processing time

As described in Section 3, several optimization techniques for reducing the computational complexity were applied to overcome the resource limitations of the device. The average processing time was measured using a series of test data, each of which were 21.3 seconds long and sampled at 48 kHz.

Table 2 lists the average processing time between the floating-point and fixed-point arithmetic implementations of the proposed algorithm. In the table, the fixed-point arithmetic implementation yielded a 65.7% processing time reduction, compared to the floating-point arithmetic implementation.

Table 3 also shows the mean processing time before and after performing the optimization techniques based on the C-code level and algorithmic level for fixed-point arithmetic programming. In C-code level optimization, significant complexity reduction was achieved by reducing the number of FFTs and IFFTs from four to 2. On the other hand, the computations for calculating the beamforming weights, phase differences and sigmoid function were reduced by a table look-up approach in the algorithmic level optimization.

Table 2. Complexity comparison between the floating-point and fixed-point implementation for the proposed audio-zoom algorithm.

Programming	Average processing time (sec)	Percentage of clock speed (%)
Floating-point arithmetic	30.2	141.8
Fixed-point arithmetic	16.2	76.1

Table 3. Complexity comparison of the proposed audio-zoom algorithm before and after optimizing the fixed-point arithmetic programming.

Optimization	Average processing time (sec)	Percentage of clock speed (%)
None	16.2	76.1
C-code level	9.7	45.6
C-code and algorithmic level	3.1	14.6

4.2 Sound pressured level according to audio source directions

The performance of the implemented video-zoom driven audio-zoom algorithm was evaluated in a semi-anechoic chamber to minimize the effects of reverberation, as shown in Fig. 5. A loudspeaker was used as the audio

source, in which white noise was adjusted to 60 dB using a noise level meter. Note that the configuration of a microphone array is not linear and the distances among microphones are not uniform so the user can handle the device easily. Fig. 6 shows the beam patterns at frequencies ranging from 0.5 kHz to 3 kHz when the video zoom is activated at its maximum level in theory. In the figure, the front direction is the most concentrated. The SPL of the processed audio signal was measured depending on the DOA of the audio source, as shown in Table 4. In the table, the SPL in all directions was equivalent when the audio-zoom was switched off. On the other hand, with the audio-zoom turned on, the front audio signal was amplified by approximately 5 dB and 10 dB at the minimum and maximum video-zoom level, respectively.

Tables 5 and 6 show the SPL of the processed audio signal in each direction according to the different SPLs and audio source distances, respectively. Note that the audio-zoom was on with the maximum video-zoom. The SPL of the front audio signal processed by the proposed method increased by approximately 10 dB compared to that of the original front signal for all SPLs and distances. Note that the experimental conditions shown in Fig. 5 were optimized when the distance between the device and sound source was 1.5 m and the sound sources were recorded at 60 dB SPL. This can explain why the performance at 60 dB in Table 5 and at 1.5 m in Table 6 was the best.

Table 4. Measurement of the sound pressure levels (dB) depending on the DOA of the audio source.

Function	DOA of audio source (white noise)			
	0˚	90˚	180˚	-90˚
Audio-zoom off	0.0	0.0	0.0	0.0
Audio-zoom on (minimum video-zoom)	4.73	3.07	2.17	1.54
Audio-zoom on (maximum video-zoom)	10.52	0.38	1.31	0.58

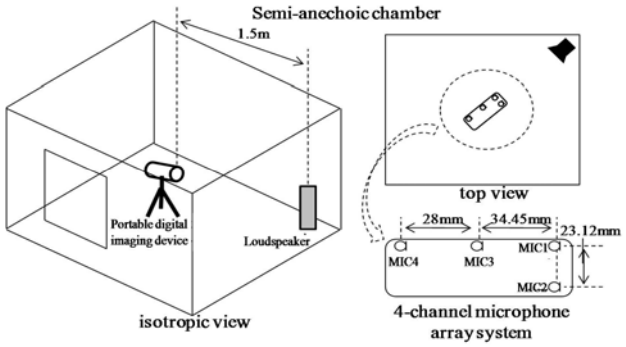


Fig. 5. Experimental environment for measuring the sound pressure levels in a semi-anechoic chamber.

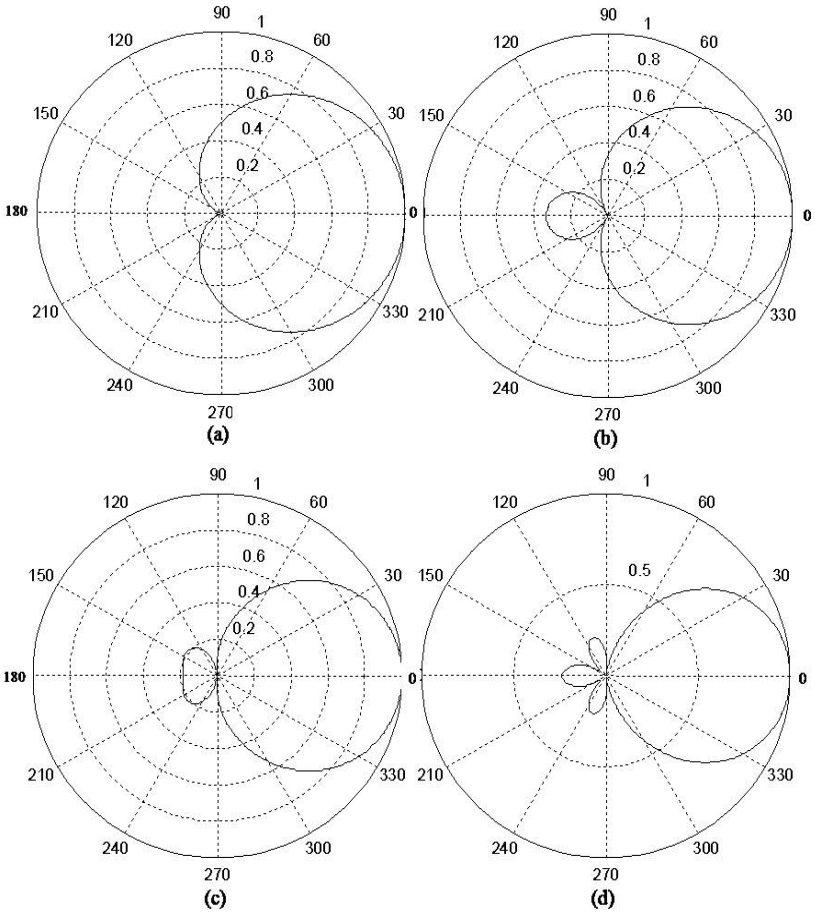


Fig. 6. Beam pattern of a super-directive beamformer at (a) 0.5 kHz, (b) 1 kHz, (c) 2 kHz, (d) 3 kHz when the video-zoom is activated at its maximum level.

Table 5. Measurement of the sound pressure levels (dB) depending on the DOA of the audio source according to the different SPL of the audio source.

SPL (dB)	DOA of audio source (white noise)			
	0˚	90˚	180˚	-90˚
55	9.78	2.16	2.39	1.32
60	10.52	0.38	1.31	0.58
65	9.91	1.67	1.17	1.47

Table 6. Measurement of the sound pressure levels (dB) depending on the DOA of the audio source according to the different distance between the audio source and microphones.

Distance (m)	DOA of audio source (white noise)			
	0˚	90˚	180˚	-90˚
1.0	10.05	1.73	1.23	1.22
1.5	10.52	0.38	1.31	0.58
2.0	9.27	2.62	1.65	2.09

The performance of the implemented video-zoom driven audio-zoom algorithm was evaluated by setting the maximum video-zoom in a different room and noisy environments. Compared to the experimental environment shown in Table 4, a reverberant room was added as a new recording environment. In addition, three noisy environments (white noise, background speech, and background music) were considered, in which the noise signals were obtained from the sound quality assessment material (SQAM) database [12].

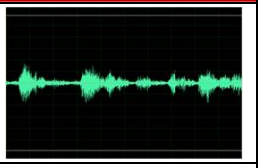
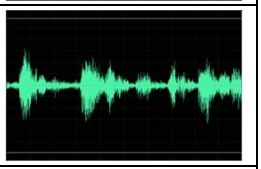
Table 7 compares the sound pressure levels depending on the DOA of the audio source in a different room and noisy environment. When the SPLs of the processed audio signal in the side and back directions were measured in the reverberant room, they were slightly higher than those measured in the semi-anechoic chamber. On the other hand, the front audio signal was still amplified by approximately 10 dB at the maximum video-zoom level in both room environments, regardless of the noisy environment.

The GUI depicted in Fig. 4(b) was operated to show the effect of the proposed algorithm on the recorded video content. Table 8 lists the still images and audio signals captured from the video contents by changing the video-zoom level from 1 to 10, as defined in Table 1. In the table,

Table 7. Measurement of the sound pressure levels (dB) depending on the DOA of the audio source in a different room and noisy environment.

Room	Noise	DOA of audio source			
		0˚	90˚	180˚	-90˚
Semi-anechoic chamber	White noise	10.52	0.38	1.31	0.58
	Speech	10.77	2.10	0.51	2.73
	Music	11.46	2.45	0.01	0.58
Reverberant room (office)	White noise	9.28	3.69	3.32	3.36
	Speech	9.06	5.31	3.25	5.00
	Music	10.25	3.53	4.54	5.00

Table 8. Illustration of the effect of the proposed video-zoom driven audio-zoom algorithm on the video contents.

Video-zoom level	Snapshot of an image	Waveform of the audio signals
1		
5		
10		

the person appeared closer as the video-zoom level was increased. Moreover, the sound pressure level of the person's voice appears to increase by up to 10 dB at a video-zoom level of 10. Therefore, increasing the SPL could provide better user satisfaction because the SPL is strongly correlated with the user's satisfaction [13].

5. Conclusion

A beamforming-based video-zoom driven audio-zoom algorithm was proposed to focus the front audio signal in a portable digital imaging device. The proposed algorithm consisted of three parts: an audio-focus control based on super-directive beamforming, a soft masking process using a DOA estimation, and an audio gain control based on the video-zoom level. To realize the proposed algorithm on a portable digital imaging device with limited resources in real time, several optimization techniques were applied to reduce the computational complexity and required memory size. The performance of the implemented algorithm was then measured in terms of the processing time and the SPL between the front and other directions at the maximum video-zoom level in a semi-anechoic chamber. Subsequent experiments showed that the average processing time was reduced to approximately 14.6% of the entire clock speed. Moreover, the front signal was amplified by approximately 10 dB compared to the other directions.

References

[1] M. Matsumoto, H. Naono, H. Saitoh, K. Fujimura, and Y. Yasuno, "Stereo zoom microphone for consumer video cameras," IEEE Transactions on Consumer Electronics, vol. 35, no. 4, pp. 759-766, Nov. 1989. Article (CrossRef Link)

[2] M. Brandstein and D. Ward, Microphone Arrays, Springer, New York, NY, 2001. Article (CrossRef Link)
[3] J. M. Kates, "Super directive arrays for hearing aids," Journal of the Acoustical Society of America, vol. 94, no. 4, pp. 1930-1933, Oct. 1993. Article (CrossRef Link)
[4] M. Alik, M. Okamoto, S. Aoki, H. Matsui, T. Sakurai, and Y. Kaneda, "Sound source segregation based on estimating incident angle of each frequency component of input signals acquired by multiple microphones," Acoustical Science and Technology, vol. 22, no. 2, pp. 149-157, May 2001. Article (CrossRef Link)
[5] D. L. Wang and G. J. Brown, Computational Auditory Scene Analysis: Algorithms and Applications, IEEE Press. Wiley-Interscience, 2006. Article (CrossRef Link)
[6] S. Y. Jeong, J. H. Jeong, and K. C. Oh, "Dominant speech enhancement based on SNR-adaptive soft mask filtering," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Taipei, Taiwan, pp. 1317-1320, Apr. 2009. Article (CrossRef Link)
[7] E. Zwicker and H. Fastl, Psychoacoustics: Facts and Models, Springer-Verlag, Berlin, Germany, 1990. Article (CrossRef Link)
[8] K. S. Lee, Y. C. Park, and D. H. Youn, "Software optimization of the MPEC-audio decoder using a 32-bit MCU RISC processor," IEEE Transactions on Consumer Electronics, vol. 48, no. 3, pp. 671-676, Aug. 2002. Article (CrossRef Link)
[9] Z. Wei, K. L. Tang, and K. N. Ngan, "Implementation of H.264 on mobile device," IEEE Transactions on Consumer Electronics, vol. 53, no. 3, pp. 1109-1116, Aug. 2007. Article (CrossRef Link)
[10] V. Oppenheim and W. Schafer, Discrete-Time Signal Processing, 2nd Ed., Prentice Hall, 1999. Article (CrossRef Link)
[11] E. Eolder, "The CORDIC trigonometric computing technique," IRE Transactions on Electronic Computers, vol. 8, no. 3, pp. 330-334, Sept. 1959. Article (CrossRef Link)
[12] EBU Tech. Document 3253, Sound Quality Assessment Material (SQAM), Apr. 1998. Article (CrossRef Link)
[13] M. Kompis, P. Feuz, G. Valentini, and M. Pelizzone, "A combined fixed/adaptive beamforming noise-reduction system for hearing aids," in Proc. the 20th Annual International Conference of IEEE Engineering in Medicine and Biology Society, Hong Kong, China, pp. 3135-3139, Nov. 1998. Article (CrossRef Link)

Nam In Park received his B.S. degree in Electronics and Communications Engineering from Kwangwoon University, Korea in 2007, and an M.S. degree in Information and Communications Engineering from Gwangju Institute of Science and Technology (GIST), Korea in 2009. He is now a Ph.D. student at GIST. His current research interests include speech/audio coding and implementation on embedded systems.

Seon Man Kim received his B.S. degree in Mechanical Design Engineering from Chonbuk National University, Korea in 2003, and M.S. degree in Mechatronics Engineering from Gwangju Institute of Science and Tech-nology (GIST), Korea in 2005. He was a visiting scholar with the Hearing Aid Lab at Speech Pathology & Audiology from the University of Iowa, USA in 2004. During 2005-2007, he was a researcher in Samsung Reciprocating Compressor RND Group, Gwangju, Korea. He received his Ph.D. degree in Information and Communications Engineering from GIST in 2013. He is now a post-doctoral researcher at GIST. His current research interests include speech recognition, speech enhancement, source separation, and voice activity detection.

Hong Kook Kim received his B.S. degree in Control and Instrumentation Engineering from Seoul National University, Korea in 1988. He then received both M.S. and Ph.D. degrees in Electrical Engineering from the Korea Advanced Institute of Science and Technology (KAIST), Korea in 1990 and 1994, respectively. He was a senior researcher at the Samsung Advanced Institute of Technology (SAIT), Kiheung, Korea, from 1990 to 1998. During 1998-2003, he was a senior technical staff member with the Voice Enabled Services Research Lab at AT&T Labs-Research, Florham Park, NJ. Since August 2003, he has been with the School of Information and Communications, at Gwangju Institute of Science and Technology (GIST) as a professor. His current research interests include speech recognition and coding, audio coding and 3D audio, and embedded algorithms and solutions for speech and audio processing for handheld devices.

Myeong Bo Kim received his B.S. degree in Electrical Engineering from Yonsei University, Korea in 1991. He then joined the Camcorder R&D Team at Samsung Electronics Co. (SEC) Ltd. in 1991. He received his M.S. degree in Electrical and Electronic Engineering from Yonsei University in 2004. He is now a principal engineer who leads the development of camera products at SEC.

Sang Ryong Kim received his M.S. and Ph.D. degrees in Electrical Engineering from the Korea Advanced Institute of Science and Technology (KAIST), Korea in 1982 and 1989, respectively. He was affiliated with Samsung Electronics, Co. Ltd., as a technical research staff member from 1989 to 1993. During 1993-2006, he was a Vice President of the Human and Computer Interaction (HCI) Lab and the Interaction Lab at the Samsung Advanced Institute of Technology (SAIT), Korea. From 2007 to 2012, he had consecutively filled various posts as senior VP, director of camcorder business development, and R&D and product technology general manager in digital imaging division, at Samsung Electronics, Co. Ltd., Korea. Since November 2012, he has played a role as a Next Generation Broadcasting PM in the Ministry of Science, ICT and Future Planning. His research interests include signal processing for digital imaging devices and advanced network technologies for consumer electronics.