Spatial Sound Reproduction Using Conventional and Parametric Loudspeakers

Ee-Leng Tan, Woon-Seng Gan, and Chiu-Hao Chen
Nanyang Technological University, Singapore
E-mail: etanel@ntu.edu.sg Tel: +65-67906901
E-mail: ewsgan@ntu.edu.sg Tel: +65-67904538
E-mail: chench@ntu.edu.sg Tel: +65-67906901

***Abstract*—The auditory image of a movie or game scene can be decomposed into point-like sources and diffused sources for effective and accurate audio synthesis. By embedding appropriate visual and audio cues into objects in a 2D or 3D visual scene, an immersive and engaging experience can be created. While there are many breakthroughs in the display technology recently, such as the ultra high-definition (UHD) and 3D displays, conventional sound systems (stereo, 5.1, etc) are still being used. Such an audio-visual setup may degrade the overall experience. This degradation is directly linked to the dispersive nature of the conventional loudspeaker, and the rendered auditory image may be perceived to lack sharpness in the spatial imaging due to the reverberant nature of a room. This drawback tends to lead to comparably poor synthesis of point-like sources as compared to diffused sources in the rendered auditory image. On the other hand, the rendered auditory image from a directional loudspeaker, such as the parametric loudspeaker, may seem to lack spaciousness and sound envelopment due to very little influence of the acoustics of a room. Therefore, directional loudspeaker is suitable for rendering point-like sources, but not diffused sources. In this paper, we propose a unique sound system which comprises of conventional loudspeakers and parametric loudspeakers. This setup exploits the high directivity of the parametric loudspeakers to render sharp auditory images while producing the diffused sources of the auditory image using the conventional loudspeaker[1].**

I. INTRODUCTION

The explosion of 3D displays in the consumer space and the fast increasing availability of 3D-TV broadcast, and 3D movies (both in 3D cinema and in the form of 3D Bluray) have garnered a lot of interest in 3D media. Visual experience of digital media can be enhanced using 3D displays, which permit viewers to experience visual objects moving towards or away from them. An immersive and engaging experience is only complete if these visual objects are accompanied with the appropriate audio cues (direction and distance of visual objects from viewer) and the ambience sound of the scene. This observation highlights the importance of accurate spatial sound reproduction in 3D media, but the same observation can also be applied to 2D media.

The current audio-visual experience is delivered using a 2D or 3D display together with a conventional stereo sound system or surround sound system. However, these sound systems may not accurately reproduce the spatial sound content of 2D and 3D media. To illustrate the limitation of the current sound systems, we refer to the example shown in Fig. 1. In this example, the viewer is presented with a scene, where a bee (the sound of bee is regarded as a point-like source) is flying in a grass field (ambience sound of the grass field is regarded as a diffused sound). With a 3D display, the viewer can see a virtual bee flying towards him/her (the bee will look flat on a 2D screen and progressively becoming bigger on the screen). For this example, the viewer expects to hear the ambience sound of the grass field as well as to accurately localize the bee as it flies towards him/her. Such spatial sound is used in movies and gaming, especially in first and third person shooting games. One of the main reasons is such games place the gamer's avatar in the middle of the action. Due to this placement of the gamer's avatar, the gamer experiences large amount of audio cues (such as bullets flying or opponents moving towards or away from the gamer) as well as dynamic ambience sound of the gaming environment (gamer's avatar moves from one location to another).

Surround sound system generates sound around the user to achieve sound envelopment. Sound is usually differentiated between the left and right loudspeakers, as well as between the front and rear loudspeakers. Such sound localization is only a fraction of how sound is perceived by the human listener being in a natural 3D environment, where sound can reach the ears of the human listener from arbitrary directions (defined by azimuth and elevation of the sound source relative to the human head). Ideally, the ambience sound of the grass field provides the sensation of spaciousness and sound envelopment to the listener so as to recreate the experience of "being there". For a stereo sound system, the amount of spaciousness is usually constrained by the amount of space between the left and right loudspeakers. This leads to limited audio depth as the depth dimension of spatial sound extends to the area between the two loudspeakers of the stereo sound system. Hence, spaciousness can be significantly enhanced by a surround sound system as the area between the loudspeakers of the surround sound system tends to be larger than stereo sound system.

To study the dependency of rendered auditory image on the loudspeaker directivity and room acoustics, the rendered

[1] This work is supported by the Singapore Ministry of Education Academic Research Fund Tier-2, under research grant MOE2010-T2-2-040.

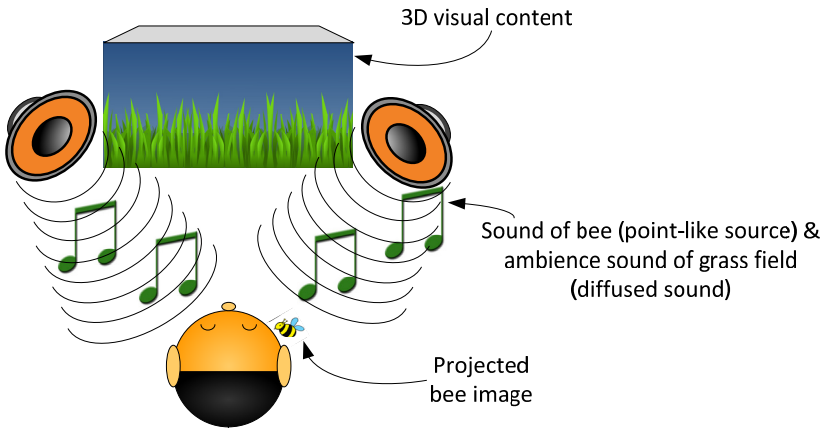


Fig. 1 Typical 3D media with "pop up" visual such as projected bee in this case. Viewers would expect to perceive the closeness of the bee as well as the ambience sound of the grass field.

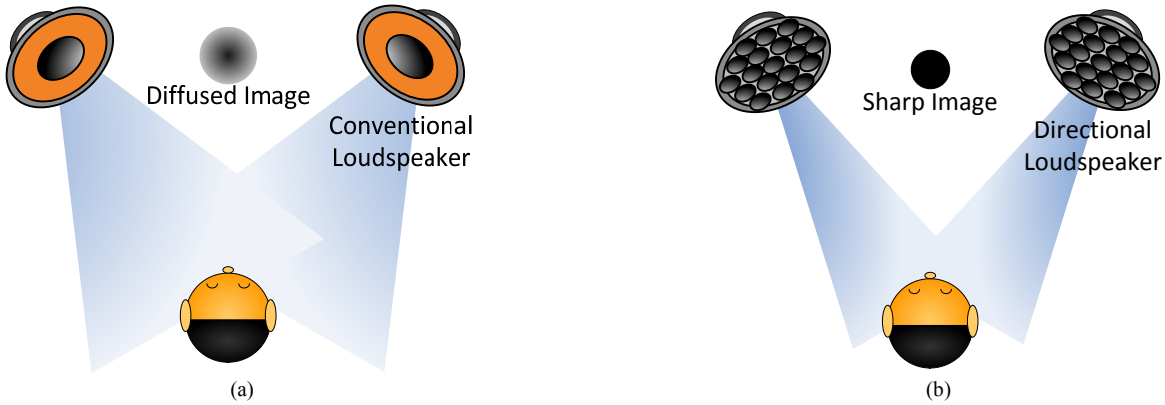


Fig. 2 Auditory images rendered by (a) conventional loudspeaker and (b) directional loudspeaker.

auditory images from directive and dispersive loudspeakers are illustrated in Fig. 2. If the loudspeaker is dispersive, the rendered auditory image from the loudspeaker may be perceived to lack sharpness due to the reverberation of the room. On the other hand, the rendered auditory image from a directional loudspeaker may be perceived to lack spaciousness as the listener is mostly hearing the direct sound from the loudspeaker [1]. Due to significantly reduced crosstalk between the two directional loudspeakers, it has been shown that the audio cues from directional loudspeakers are more accurate than conventional loudspeakers [2]. Lesser reverberant sound is heard from directional loudspeaker as compared to conventional loudspeaker, therefore directional loudspeaker can be used to produce nearer auditory images to its listeners [3].

The directivity of the loudspeakers in surround sound system affects the overall auditory image in a slightly different manner as compared to stereo sound system. Since sound is sent to numerous loudspeakers in the surround sound system, reverberation or spaciousness is rendered in the overall auditory image, even in the case that all the loudspeakers in the surround sound system are fairly directional. For this reason, surround sound system would reproduce better sound envelopment as well as spaciousness as compared to stereo sound system. Therefore, the ambience sound of the grass field in Fig. 1 would be better reproduced by surround sound system as compared to stereo sound system.

The audio cues (from point-like sources) of the bee in Fig. 1 can be well reproduced by a 3D sound system since such a system attempts to render audio cues that are positioned arbitrarily around a listener. This virtual sound localization is achieved by reproducing the acoustic signals that appear at the ears of the listener, and this approach is known as binaural audio or commonly known as 3D audio. The transformation of sound pressure from free field to the listener's ears, including the diffraction of sound by the torso, head, and external ear, is characterized by the head related transfer function (HRTF) [4]. However, producing accurate 3D sound cues using dispersive loudspeakers can be challenging due to the crosstalk between these loudspeakers. Usually, crosstalk cancellation is applied to stereo sound system to improve its 3D sound performance. This approach has limited success as the crosstalk cancellation requires good subtraction of the sound fields produced by the two loudspeakers. In other words, such cancellation is sensitive to the differences of the

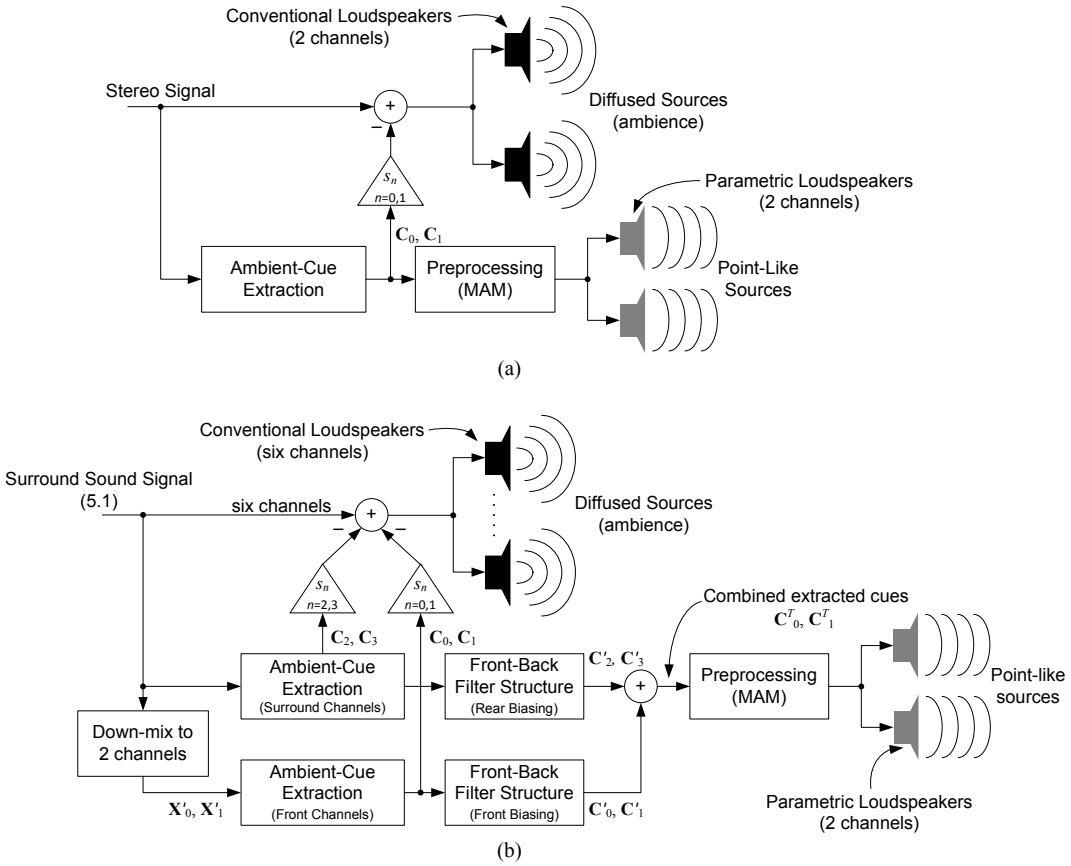


Fig. 3 Two configurations of the i3D sound system for (a) stereo input and (b) multi-channel input.

sound fields produced by the two loudspeakers, and is highly sweet-spot dependent. One common way to eliminate the issue of sweet-spot sensitivity, and avoiding crosstalk cancellation is to use headphones. Since headphones have excellent channel separation, they are commonly used for binaural audio. Unfortunately, binaural audio reproduced by headphones tends to suffer from in-the-head experience and front-back confusion due to the variation of HRTF from one to another [5]. In addition, extended headphones listening can be very uncomfortable. Other shortcomings of headphones listening include limited spaciousness and no sound envelopment due to close proximity of the sound emitters in headphones to the listener's ears.

The rest of the paper is structured as follows. In Section II, the proposed sound system [6] is described. A technique, referred as the modified amplitude modulation (MAM) [7,8], to reduce the distortion in parametric loudspeaker is reviewed in Section III. Experiments are carried out with the proposed system and our observations are summarized in Section IV. Finally, our conclusions are presented in Section V. It may be noted that part of this work has been reported [9] at the Acoustics conference in Hong Kong 2012, and this paper extends on [9] by giving more in-depth simulation results of the proposed sound system.

II. Proposed Sound System

It is difficult to accurately reproduce spatial sound using either the conventional loudspeaker or the directional loudspeaker. To overcome this issue, a sound system that combines the conventional loudspeaker and parametric loudspeaker is developed. The objective of the proposed system is to use the conventional loudspeaker and the parametric loudspeakers to accurately reproduce the diffused sound sources and point-like sound sources, respectively, so as to achieve a better reproduction of spatial sound.

The proposed system is referred as the immersive 3D (i3D) sound system, and two configurations of i3D are shown in Fig. 3(a) and 3(b) for stereo and multi-channel input, respectively. To achieve the sound projection required by i3D, the principal component analysis (PCA) based cue-ambient decomposition (CAD) technique proposed by Goodwin and Jot [10] is adapted in i3D. Their approach is based on the following assumptions: (i) cues have higher energy than ambience, and (ii) cue and ambience are orthogonal in signal space. For a stereo signal, Goodwin and Jot have formulated the extracted cue $\mathbf{C}_n$ of the nth channel as

$$\mathbf{C}_n = \left(\frac{\mathbf{v}^H \mathbf{X}_n}{\mathbf{v}^H \mathbf{v}} \right) \mathbf{v}, \tag{1}$$

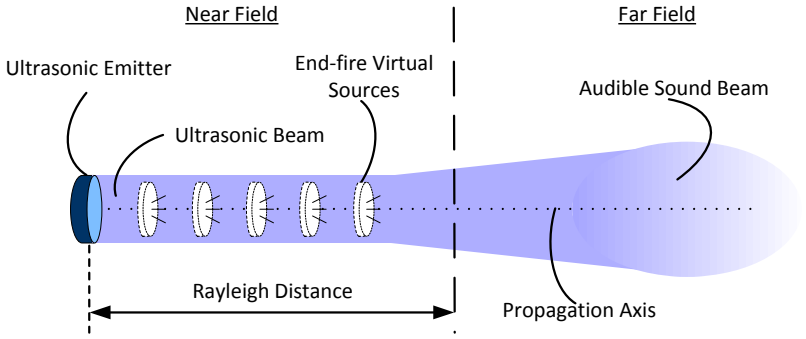


Fig. 4 Formation of the parametric loudspeaker.

where $\mathbf{X}_n$ is the input signal of the nth channel and $\mathbf{v}$ is the eigenvector with the largest eigenvalue. Considering a pair of input signal $\mathbf{X}_0$ and $\mathbf{X}_1$, let $r_{0,0}$, $r_{1,1}$, and $r_{0,1}$ represent the auto-correlation of $\mathbf{X}_0$, the auto-correlation of $\mathbf{X}_1$, and the cross-correlation of $\mathbf{X}_0$ and $\mathbf{X}_1$, respectively. The eigenvector $\mathbf{v}$ is computed as

$$\mathbf{v} = r_{0,1}\mathbf{X}_0 + \left\{0.5\left[r_{0,0} + r_{1,1} + \sqrt{\left(r_{0,0} - r_{1,1}\right)^2 + 4\left|r_{0,1}\right|^2}\right] - r_{0,0}\right\}\mathbf{X}_1. \quad (2)$$

The first configuration of i3D shown in Fig. 3(a) applies CAD to separate the ambience and cue from a stereo signal, and the separated ambience and cue are then channeled to the conventional loudspeakers and parametric loudspeakers, respectively.

Goodwin and Jot's approach is computationally expensive when it is applied to a surround sound input. Instead, a computational efficient scheme to extract cues from surround sound input is adapted from their approach. Our proposed technique extracts cues from the front and surround channels separately. Considering a 5.1 audio signal, the front channels (front left, front right, and center) are down-mixed to two channels before cue extraction, and the surround channels are directed processed by CAD. For this case, we adopt the down-mixing scheme from ITU-R BS.775-2 for the front channels, and the down-mixed front channels are given as

$$\begin{aligned} X_0'(k) &= X_0(k) + 0.7071X_2(k), \\ X_1'(k) &= X_1(k) + 0.7071X_2(k), \end{aligned} \quad (3)$$

where $X_0(k)$, $X_1(k)$, and $X_2(k)$ are the kth sample of the left, right, and center channels, respectively. By applying (1) to the down-mixed front channels and the surround channels (surround left and surround right) $X_3(k)$, and $X_4(k)$, we have two pairs of extracted cues from the down-mixed front channels and the surround channels.

To prevent the reproduced audio output to have over-emphasis of cues (after the acoustic summation of sound from the conventional and parametric loudspeakers), part of the extracted cues are subtracted from the front and surround channels, and the amount of subtraction is defined by S_n, where $0 < S_n < 1$ and $n = 0, 1, \ldots 3$. Let $C_0(k)$, $C_1(k)$, $C_2(k)$, and $C_3(k)$ denote the extracted cues from the down-mixed front channels and the surround channels. In addition, the projected sound is directed to the concha and pinna of the ear to produce individualized cues based on the unique shape and size of the listener's ear [11]. Furthermore, front and back perception of the extracted cues are enhanced by the front-back filter structure proposed by Tan and Gan [11]. Let $C_0'(k)$, $C_1'(k)$, $C_2'(k)$, and $C_3'(k)$ denote the front-back biased cues of $C_0(k)$, $C_1(k)$, $C_2(k)$, and $C_3(k)$, respectively. The combined cues $C_0^T(k)$ and $C_1^T(k)$ obtained with the front-back biased cues are expressed as

$$\begin{aligned} C_0^T(k) &= C_0'(k) + C_2'(k), \\ C_1^T(k) &= C_1'(k) + C_3'(k). \end{aligned} \quad (4)$$

The combined cues are then reproduced by the parametric loudspeakers.

III. Preprocessing For Parametric Loudspeaker

It has been experimentally found by Thuras *et al.* [12] that two collimated sound waves having different frequencies generate two new sound waves which have frequencies of the sum and difference frequencies of the original sound waves. In 1963, Westervelt [13] presented a theoretical model describing the generation of the difference frequency from two high frequency collimated beams of primary sound waves. The nonlinear interaction of the primary sound waves in a medium gives rise to an end-fire array of virtual sources, which is referred as the parametric array. Subsequently, the ultrasonic emitter and the parametric array (in air) are collectively referred as the parametric loudspeaker and are illustrated in Fig. 4. In 1982, Yoneyama *et al.* [14] demonstrated the use of parametric loudspeaker to generate broadband audio. They achieved the reproduction of broadband audio by modulating the input audio onto an ultrasonic carrier and then projecting the modulated carrier into the air using an array of ultrasonic emitters. The reproduced audio (demodulated signal) is the consequence of the generation of the difference frequencies between the modulated ultrasonic carrier and the ultrasonic carrier in air. While their experiments revealed that the demodulated signal

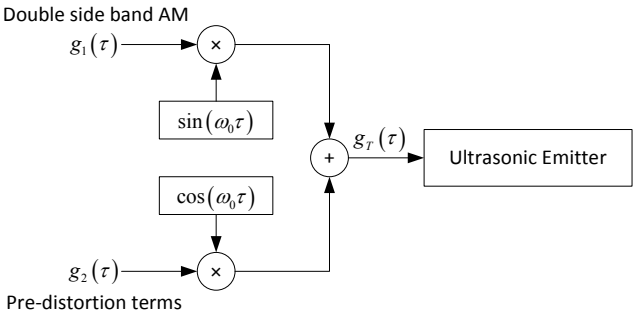


Fig. 5 Quadrature amplitude modulation structure used to reduce distortion in parametric loudspeaker.

generated by the parametric loudspeaker has a very sharp directivity pattern, the demodulated signal is found to exhibit high harmonic distortion and poor frequency response.

In the case of far field, Berktay [15] approximates the demodulated wave pressure p_2 to be proportional to the second time derivative of the squared envelope of a primary wave pressure p_1. Considering the primary wave pressure p_1 is radiated into air from a circular piston source with radius a and the ultrasonic wave at an axial distance z, Berktay's solution is expressed as

$$p_2 \approx \frac{\beta p_0^2 a^2}{16\rho_0 c_0^4 \alpha z} \frac{\partial^2}{\partial \tau^2} E^2(\tau), \tag{5}$$

where $\tau = t - z/c_0$ is the retarded time, c_0 is the small signal sound speed, p_0 is the initial pressure of the primary wave, $E(\tau)$ is the envelope of the primary wave, α is the absorption coefficient, β is the coefficient of non-linearity, and ρ_0 is the ambient density.

It has been long known that the primary cues for sound localization are the time and level differences at the ears of the listener, and these cues are referred as interaural time delay (ITD) and interaural level difference (ILD), respectively. From Rayleigh's duplex theory, the low and high frequencies are localized using time and level cues, respectively. ITD operates at frequencies below 1.5 kHz, but parametric loudspeakers may not produce sufficient SPL at low frequencies (<500 Hz). On the other hand, ILD mainly operates at higher frequencies (>1500 Hz), but also occurs at low frequencies when the sound source is very close to the head. Using a single tone analysis, the experiments conducted in [2] revealed that parametric loudspeakers produce better sound localization as compared to conventional loudspeakers in terms of ITD and ILD.

However, the distortion of the parametric loudspeaker might impair the quality of the reproduced cues. Consider the envelope of an amplitude modulated signal using double sided amplitude modulation (DSBAM) $E_{\text{DSBAM}}(\tau) = 1 + mg(\tau)$, where $g(\tau)$ is the input audio or the modulating signal and m is the modulation index, a second harmonic of $g(\tau)$ will be produced in demodulated signal. The presence of such harmonics may impair the sound localization of the reproduced cues from parametric loudspeaker.

Based on (5), one obvious preprocessing technique to reduce distortion in the demodulated signal from parametric loudspeaker is to square root the envelope of the primary wave. This observation motivated several works in [16,17]. This preprocessing technique is commonly known as the square root amplitude modulation (SRAM), and the envelope of the modulated signal becomes

$$E_{\text{SRAM}}(\tau) = \sqrt{1 + mg(\tau)}. \tag{6}$$

By substituting the envelope of the DSBAM and SRAM signal into (5), the demodulated signal is found to be

$$\begin{aligned} p_{\text{DSBAM},2} &\approx \frac{\beta p_0^2 a^2}{16\rho_0 c_0^4 \alpha z} \frac{\partial^2}{\partial \tau^2} E_{\text{DSBAM}}^2(\tau) \\ &\approx \frac{\beta p_0^2 a^2}{16\rho_0 c_0^4 \alpha z} \left(2m \frac{\partial^2}{\partial \tau^2} g(\tau) + m^2 \frac{\partial^2}{\partial \tau^2} g^2(\tau) \right), \end{aligned} \tag{7}$$

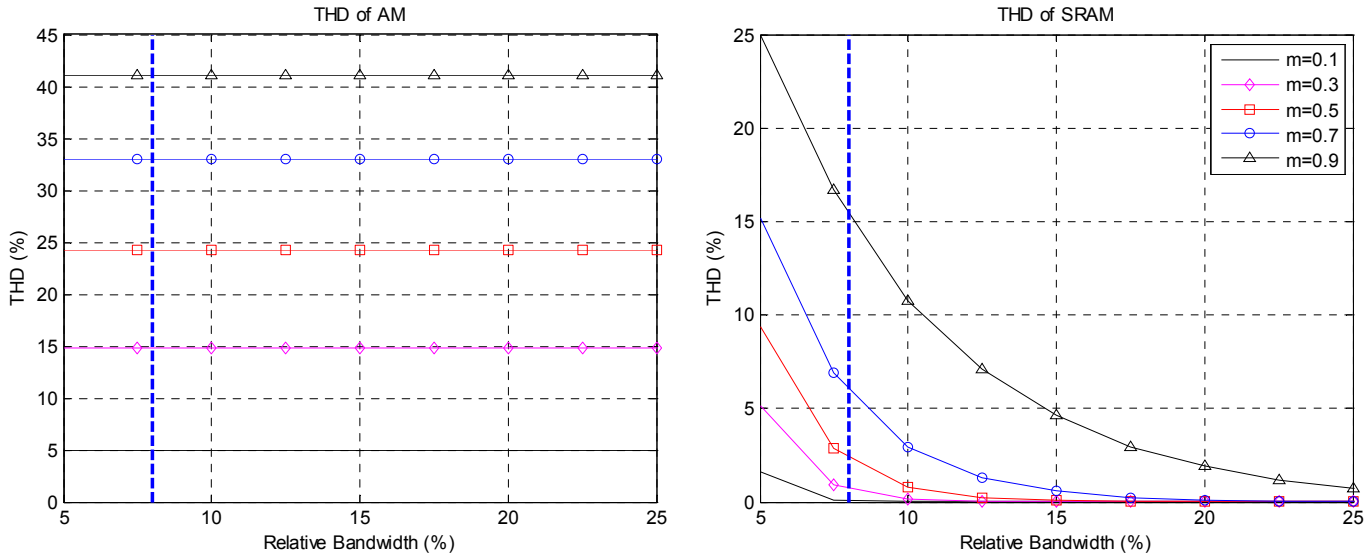
and

$$p_{\text{SRAM},2} \approx \frac{\beta p_0^2 a^2}{16\rho_0 c_0^4 \alpha z} \left(m \frac{\partial^2}{\partial \tau^2} g(\tau) \right), \tag{8}$$

respectively. Equations (7) and (8) reveal that the sound pressure level (SPL) of the demodulated signals from DSBAM and SRAM signal are proportional to m, and the distortion from the DSBAM signal is proportional to m^2. Hence, it is not practical to reduce the distortion by simply reducing the modulation index as it also reduces the SPL of the demodulated signal.

One major shortcoming of SRAM is its requirement of an ultrasonic emitter with wide bandwidth (> 10 kHz), and such an ultrasonic emitter is not realizable with current technology [18]. In the following, we shall briefly describe our technique MAM which provides good reduction of distortion with ultrasonic emitters having limited bandwidth. It is shown in [7] that the quadrature amplitude modulation (QAM) reduces distortion found in parametric loudspeaker. This reduction of distortion is achieved by placing a set of pre-distortion terms in the orthogonal path with DSBAM in the non-orthogonal path of the QAM scheme, as shown in Fig. 5.

We found that the Taylor series of $\sqrt{1 - m^2 g^2(\tau)}$ is particularly effective in reducing distortion in the parametric loudspeaker. From our derivation, we arrived to a q-order MAM scheme given by



(a) (b)

Fig. 6 THD values of (a) AM and (b) SRAM. Dashed blue line indicates the relative bandwidth of 8. Same legend applies to both plots.

$$g_T(q,\tau) = g_1(\tau)\sin(\omega_0\tau) + g_2(q,\tau)\cos(\omega_0\tau)$$
$$= \sqrt{g_1^2(\tau) + g_2^2(q,\tau)}\,\sin\left(\omega_0\tau + \tan^{-1}\left(\frac{g_2(q,\tau)}{g_1(\tau)}\right)\right), \quad (9)$$

where ω_0 is the resonating frequency of the ultrasonic emitter and $g_2(q,\tau)$ is a truncated series of $\sqrt{1-m^2g^2(\tau)}$ which is expressed as

$$g_2(q,\tau) = \sum_{i=0}^{q} \frac{(2i)!}{(1-2i)\,i!^2\,4^i} m^{2i} g^{2i}(\tau), \quad \text{for } \left|m^2g^2(\tau)\right| < 1. \quad (10)$$

Our simulations in [7] revealed that lower and higher order MAM schemes are suitable for ultrasonic emitters having low and high bandwidth, respectively.

To compare the distortion in the demodulated signal from parametric loudspeaker with different preprocessing techniques, the total harmonic distortion (THD) index is used. For a single tone input, THD is defined as

$$\text{THD} = \sqrt{\frac{T_2^2 + T_3^2 + \cdots + T_{n-1}^2 + T_n^2}{T_1^2 + T_2^2 + T_3^2 + \cdots + T_{n-1}^2 + T_n^2}} \times 100\%, \quad (11)$$

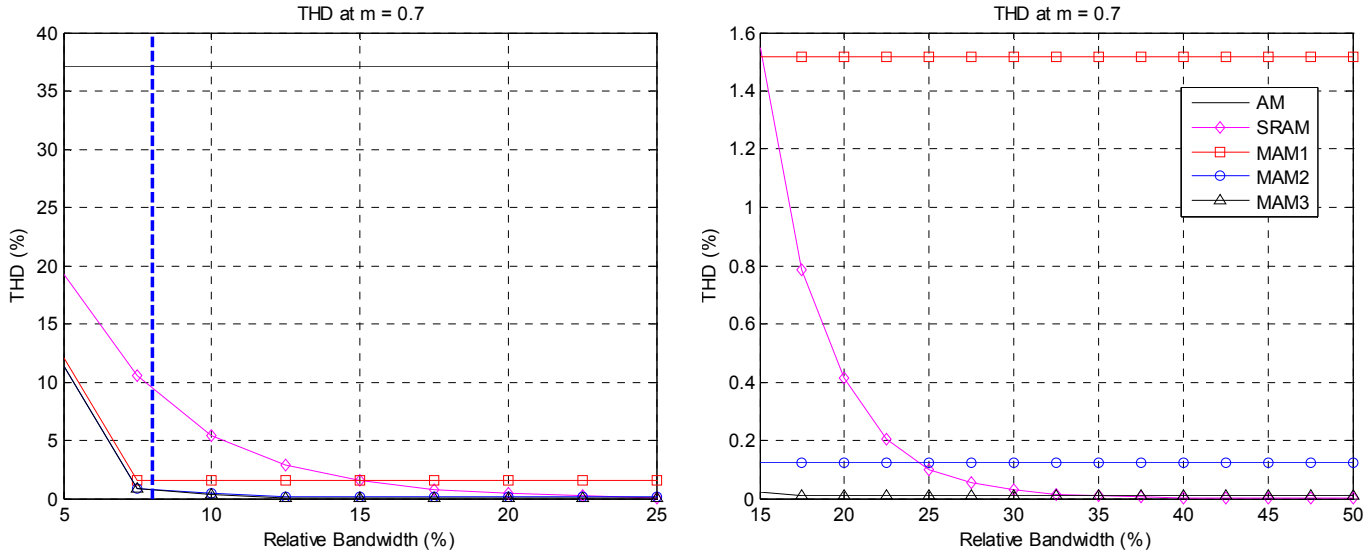
where T_1 and T_j are the amplitude of single tone at ω_1 and its higher harmonics at $j\omega_1$, respectively, for j = 2, 3, ... , n. In addition, we define relative bandwidth as a ratio of $2\omega_c/\omega_0$, where ω_c is the −3 dB bandwidth of the ultrasonic emitter. Considering ω_c = 80π rad/s and a parametric loudspeaker having a −3 dB bandwidth of 10 kHz, we would have a relative bandwidth of 8. The THD values of the AM and SRAM with respect to relative bandwidth are shown in Fig. 6. These THD values are computed with a single tone input at $0.025\omega_0$.

The THD of DSBAM remains unchanged as long as the input signal is not attenuated by the limited bandwidth of the ultrasonic emitter. The THD of demodulated signal using DSBAM is attributed to the second harmonic of the input signal that is generated in the demodulation process. It is also clear that the THD values increase as higher values of m are used. On the other hand, the THD values for SRAM decreases as relative bandwidth increases. As mentioned earlier in this section, SRAM requires ultrasonic emitter with high bandwidth to effectively reduce the distortion in the demodulated signal. This dependency is clearly illustrated in Fig. 6(b). To achieve THD values lower than 5%, the values of m for DSBAM and SRAM would be 0.1 and 0.5, respectively. These values of m are impractical as the SPL of the demodulated signal would be too low for most applications.

Values of m should be sufficiently large for the demodulated sound from parametric loudspeaker to be audible; however, the value of m is directly linked to the amount of distortion in the demodulated signal from parametric loudspeaker. Figures 7 and 8 compare the THD values for various preprocessing techniques used with parametric loudspeaker. From these figures, it is clear that MAM1 outperforms other techniques when the relative bandwidth is lesser than 15% and 17.5% for m = 0.7 and 0.9, respectively. From these figures, it is also clear that MAM3 outperforms the other techniques when the relative bandwidth is lesser than 32.5% and 37.5% for m = 0.7 and 0.9, respectively. We also observed that there is an increase of THD reduction as the order of MAM increases. However, MAM1 requires the lowest computation cost among the MAM scheme.

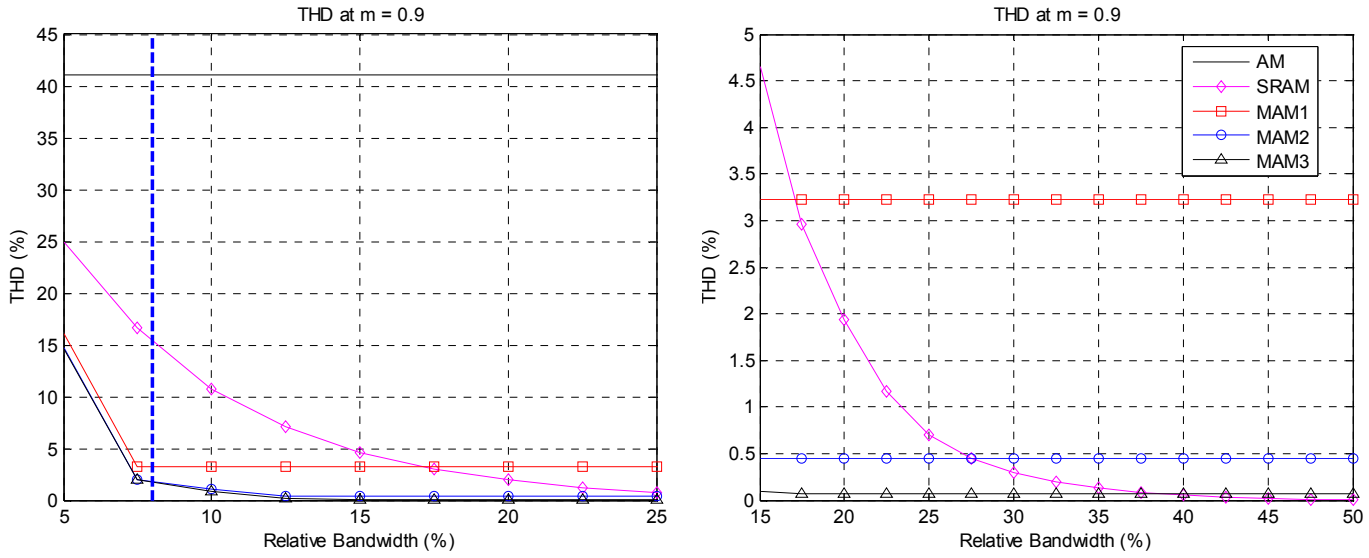
IV. Experimental Results and Observations

The current setup of i3D is shown in Fig. 9, and the i3D loudspeakers consist of a pair of conventional loudspeakers (Creative GigaWorks T3) and a pair of parametric loudspeakers. The proposed system was tested with several



(a) (b)

Fig. 7 THD values of various preprocessing techniques at m = 0.7, and zoomed in view is shown in (b). Dashed blue line indicates the relative bandwidth of 8. Same legend applies to both plots.



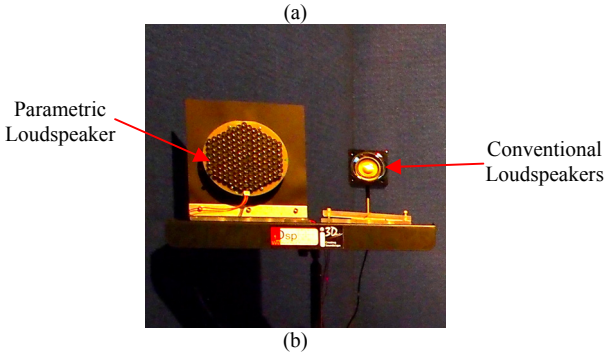
(a) (b)

Fig. 8 THD values of various preprocessing techniques at m = 0.9, and zoomed in view is shown in (b). Dashed blue line indicates the relative bandwidth of 8. Same legend applies to both plots.

game soundtracks (two channels), and CAD was found to perform well in extracting cues in these soundtracks. However, the performance of CAD degrades significantly when the assumptions of CAD do not hold (such as when there is only ambience in the soundtracks). Since ambience contains diffused sounds and have relatively low correlation as compared to cues, these segments of the soundtracks should not be processed by CAD. To mitigate this issue, one simple approach is to avoid processing the soundtrack when it contains ambience only. From our experiments, we found that most cues have at least a cross-correlation of 0.4, hence only the segments of the soundtrack having a cross-correlation of 0.4 or more are processed by CAD.

The cross-correlation plot of a gaming soundtrack obtained from [19] (40 sec video extracted from 5:25 to 6:05) is shown in Fig. 10. This video contains one distinct cue (sound of helicopter panning from center to right of the scene), which should be extracted by CAD. Furthermore, CAD should extract transient cues such as firing of a weapon. In segment A of the soundtrack, the helicopter (considered as a point-like source) moves from the center to the right side of the scene. Within segment B, the cross-correlation of the two channels of the soundtrack exhibits two peaks. This is due to the

(a)



(b)

Fig. 9 Experimental setup of i3D system, (a) view of complete test setup and (b) zoomed in of loudspeakers.

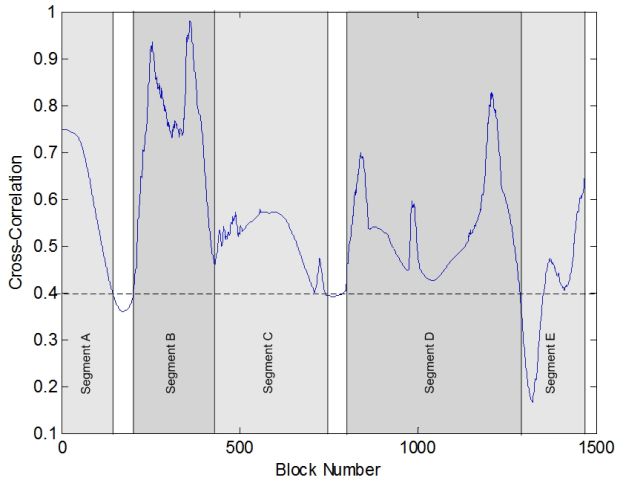


Fig. 10 Cross-correlation of test soundtrack.

continuous panning of the helicopter between the center of the screen and the right side of the scene. In segment C, the helicopter moved to the far right of the scene. The small peak in segment C and the first two peaks in segment D correspond to the firing of weapon (considered as near point-like source) by the gamer's avatar. The third peak in segment D corresponds to a loud explosion in the game soundtrack. In segment E, The helicopter moves from the center to the right of the scene, and then out of the scene. This panning of the helicopter translates to a drop in the cross-correlation in segment E. Subsequently, the helicopter moves back into the scene from the right side to the center of the scene. In segments A to E, the cues (sound of helicopter, firing of weapon, and explosion) are successfully extracted by CAD. Furthermore, the ambience in between segments A and B, segments C and D as well as segment E are avoided, and no cues (as there is no cue) is sent to the parametric loudspeakers.

V. Conclusion

Parametric loudspeaker is capable of rendering accurate audio cues from point-like sources, and the ambience sound (diffused sounds) can be effectively reproduced by conventional loudspeakers. One of the obvious advantages of parametric loudspeakers over conventional loudspeakers is little or no crosstalk exists between the sound fields from parametric loudspeakers. This leads to accurate localization of point-like sources in an auditory scene. Furthermore, the rendered auditory image is relatively robust against head movement as compared to conventional loudspeakers with crosstalk cancellation. Two configurations of i3D were discussed, where the surround sound system is implemented as an extension of the stereo system to avoid prohibitive computational cost. The separation of point-like and diffused sound sources in the input audio source is achieved using CAD which is based on the assumptions that cues have higher energy than ambience as well as cue and ambience are orthogonal in signal space. In our experiments, it is found that the performance of CAD is compromised when the assumptions are not met. Hence, it is highly desirable to detect such cases, and an extension to this work would be to determine suitable signal processing techniques (CAD is bypassed for now) to achieve optimal listening experience for all types of audio content.

References

[1] T. Holman, *Surround sound: Up and running*. Focal Press, Burlington.

[2] S. Aoki, M. Toba, and N. Tsujita, "Sound localization of stereo reproduction with parametric loudspeaker," *Applied Acoust.*, in press.

[3] A. Härmä, S. V. D. Par, and W. D. Bruijin, "On the use of directional loudspeakers to create a sound source close to the listener," *in Audio Engineering Soc. 124th Convention*, Amsterdam, The Netherlands, May 2008.

[4] D. R. Degault, *3D sound for virtual reality and multimedia*. Academic Press, New York, 1994.

[5] G. S. Kendall, ‘‘A 3-D sound primer: Directional hearing and stereo reproduction,’’ *J. Computer Music*, vol. 19, no. 4, pp. 23–46, 1995.

[6] E. L. Tan, and W. S. Gan, A system and method for processing an input signal to produce 3D audio effects, PAT/091/09/10/PCT, 2011.

[7] E. L. Tan, P. Ji, and W. S. Gan, ‘‘On preprocessing techniques for bandlimited parametric loudspeakers,’’ *Applied Acoust.*, vol. 71, no. 5, pp. 486–492, May 2010.

[8] E. L. Tan, and W. S. Gan, Multi-band audio beaming: New method in distortion reduction and its suitability for different ultrasonic emitters, PCT/SG2010/000312, 2010.

[9] E. L Tan, and W. S. Gan, "Reproduction of immersive sound using directional and conventional loudspeakers," *J. Acoust. Soc. Am.*, vol. 131, no. 4, pp. 3215-3215, May 2012.

[10] M. M. Goodwin, and J.-M. Jot, ‘‘Primary-ambient signal decomposition and vector-based localization for spatial audio coding and enhancement,’’ *in IEEE Int'l Conf. on Acoust., Speech, Signal Process.*, vol. 1, April 2007.

[11] C. J. Tan, and W. S. Gan, ‘‘Direct concha excitation for the introduction of individualized hearing cues,’’ *J. Audio Engineering Soc.*, vol. 48, no. 7/8, pp. 642–653, July 2000.

[12] A. L. Thuras, R. T. Jenkins, and H. T. O'Neil, "Extraneous frequencies generated in air carrying intense sound waves," *J. Acoust. Soc. Am.*, vol. 6, no. 3, pp. 173-180, January 1935.

[13] P. J. Westervelt, ‘‘Parametric acoustic array,’’ *J. Acoust. Soc. Am.*, vol. 35, no. 4, pp. 535–537, April 1963.

[14] M. Yoneyama, J.-I. Fujimoto, Y. Kawamo, and S. Sasabe, ‘‘The audio spotlight: An application of nonlinear interaction of sound waves to a new type of loudspeaker design,’’ *J. Acoust. Soc. Am.*, vol. 73, no. 5, 1532–1536, January 1983.

[15] H. O. Berktay, "Possible exploitation of non-linear acoustics in underwater transmitting applications," *J. Sound Vib.*, vol. 2, no, 4, pp. 435-461, April 1965.

[16] T. Kamakura, M. Yoneyama, and K. Ikegaya, "Developments of parametric loudspeaker for practical use," *10th International Symposium on Nonlinear Acoustics*, pp. 147-150, 1984.

[17] F. J. Pompei, "The use of airborne ultrasonics for generating audible sound beams," *J. Audio Engineering Soc.*, vol. 47, no. 9, pp. 726-731, September 1999.

[18] I. O. Wygant, M. Kupnik, J. C. Windsor, W. M. Wright, M. S. Wochner, G. G. Yaralioglu, M. F. Hamilton, and B. T. Khuri-Ya, ‘‘50 kHz capacitive micromachined ultrasonic transducers for generation of highly directional sound with parametric arrays,’’ *IEEE Trans. Ultrasonics, Ferroelectrics, Frequency Control*, vol. 56, no. 1, pp. 193–203, January 2009.

[19] Gameplay Video [online] Available: http://www.youtube.com /watch?v=2r_tuRWxlrE