Dynamic Audio System Based On Listener's Position For Surround Sound Effect

Jagrati Choudhary
Dept of Electronics and Telecom. Engg
AISSMS's Institute of Information Technology
jagratic8@gmail.com

Nachiket Kamod
Dept of Electronics and Telecom. Engg
AISSMS's Institute of Information Technology
nachikamod@gmail.com

Ganeshsai Muppasani
Dept of Electronics and Telecommunications Engg
AISSMS's Institute of Information Technology
ganeshgoldenhawk@gmail.com

N V Roshni
Dept of Electronics and Telecomm. Engg
AISSMS's Institute of Information Technology
nvroshni9@gmail.com

Prof Deepali Yewale
Dept of Electronics and Telecom. Engg.
AISSMS's Institute of Information Technology
deepali.yewale@aissmsioit.org

Abstract— Modern sound systems are increasingly gaining popularity day by day, remarkably since technological advances have lowered their prices, increased their qualities and features. One barrier to the much better experience while using these sound systems is its static nature in the surrounding sound. Surround sound involves three or more speakers surrounding the listener to give a surround sound effect by changing the sound source from various speakers. Although high-end audio systems provide good sound quality, to achieve the surround sound effect, the user must configure the system manually depending upon his current position, which is a tedious task. Whenever we are settling up a complex home theatre bundle, understanding the art and science of placement of the speaker channels and placement is the most crucial step while setting up a sound system. The current sound system needs manual setup according to the ideal sitting position to achieve a good sound effect at the fixed position. This manual setup consists of speaker angles and speaker sound adjusted to create a sound pocket around a fixed position. However, this effect varies when we move away from the surround sound pocket created by speakers.

This project aims to develop a real-time system to determine the listener's position and distance from the speaker system and adjust the orientations and volume levels accordingly.

I. INTRODUCTION

Automation plays a crucial role in the world economy and daily experience. In the last few decades, we have witnessed rapid development in audio systems. The journey of audio systems began with a single channel audio system (monaural audio system) in 1877. Later in the year 1931 a two-channel audio system (Stereo audio system) was introduced and in the year 2005, the most advanced air audio or surround sound system (Multi channel audio system)was introduced.

The current sound system needs manual setup according to the ideal sitting position to achieve a good sound effect at the fixed position.

This manual setup consists of speaker angles and speaker sound adjusted to create a sound pocket around a fixed position.

However, this effect varies when we move away from the surround sound pocket created by speakers. This project aims to develop a real-time system to determine the listener's position and distance from the speaker system and adjust the orientations and volume levels accordingly.

To explain the problem statement briefly, consider a real-life scenario. A person configured the orientation and volume levels of my sound system to get perfect surround sound at some position where he usually sits. However, he wants to change his sitting position or even arrangement to some other part of the room which can be far right or left or may be forward or backward from the last sitting arrangement. In this case,to get perfect surround sound,he will need to reconfigure speakers again (their orientation and volume levels) as per the new seating position, either with the assistance of a technician or on his own, which are mostly manual adjustments.

To overcome this scenario, we experimented with a combination of stereo vision and hardware technology which responds to real-time movements of the listener and dynamically adjusts the sound pocket. This system uses the OpenCV face detection algorithm and simple geometrical formulae to calculate depths and angles for an individual speaker to introduce dynamically adjusted surround sound.

Since the system avoids the heavy usage of hardware, complex algorithms, and machine learning approaches,it can be implemented on low-powered microprocessors and the current processors which are being used by sound systems.

II. LITERATURE SURVEY

Paper 1: **Surround sound systems (United States Patents On, September 16, 2014) [1]**

This paper proposes an idea of the development of a system that comprises a receiver for receiving a multi-channel spatial signal that comprises at least one surround channel. This system comprises a directional ultrasonic transducer for emitting ultrasound towards the surface to reach a listening position via a reflection of the surface and a driver circuit to drive the ultrasonic transducer. The proposed system is capable of producing virtual surround sound without requiring a speaker to be located.

Paper 2: Shadow Sound System Embodied with Directional Ultrasonic Speaker (ICISA.2013 on 2013) [2]

The paper talks about the usage of the ultrasonic speaker and computer vision system installed on a motorized mount that can freely change the speaker's directions and altitude for a specific registered user. The resulting system is proven to be able to track the registered user for providing user-selected sound contents without being interfered with by other people. This method seems promising, but it requires individual hardware for each speaker, and the solution does not cover the implementation on multi-channel audio systems efficiently.

Paper 3: An Efficient Implementation of Acoustic Cross-talk Cancellation for 3D Audio Rendering (IEEE China SIP on July 2014) [3]

The paper Speaks about the development of an efficient real-time system for the reproduction of a spatialized audio field taking the listener's position into account. The system comprises two sections: a sound rendering system based on a crosstalk canceller that is required to have a spatialized audio reproduction and a listener position tracking system to model the cross-talk canceller parameters. The resulting system is proven to be able to track the registered user with an added advantage of Crosstalk Elimination. This method seems promising, but it involves complex recursive ambiophonic crosstalk elimination algorithms. To accurately track the listener position it uses Kinect control whose Hardware is considerably costly**.**

Paper 4: Multi-rate adaptive filtering for immersive audio (IEEE Xplore on February 2001) [4]

This paper describes a method for implementing immersive audio rendering filters for single or multiple listeners and loudspeakers. In particular, the paper focuses on the case of a single or two listeners with different loudspeaker arrays to determine the weighting vectors for the necessary FIR and IIR filters using the LMS (least-mean-squares) adaptive inverse algorithm. 5It describes the transform-domain LMS adaptive inverse algorithm that is designed for cross-talk cancellation necessary in loudspeaker-based immersive audio rendering. The algorithm used in this paper is only for a single listener and for only two loudspeakers.

Review:

High-end audio systems provide very high-quality sound, and it also provides the user with the feasibility to use them for multiple events. But even the best has some drawbacks.

- The complexity of hardware, as per the research study, we can observe the research is based on mono-channel and not multi-channel.
- Research suggests the use of Kinect for object tracking, which comes with its drawbacks.
- The Algorithm's complexity to achieve the effect, researchers suggested very complex algorithms regardless of room geometry. Hence, they are challenging to understand and hard to implement on low-cost hardware.

III . PROPOSED SYSTEM DESIGN

The mathematical model serves the most important role throughout the project, as it is intended to solve the issues that persisted in previous research. This model is further divided into two sub-parts,

1. Listener tracking model-

The listener tracking model is a combination of face detection and stereo vision technique for estimating depth.

Face detection -We must classify and sort out the entities from the rest of the objects from the surroundings to align the speakers properly. In our case, these entities are people who are listening to the system. To classify them from other objects from surroundings, we implement the Haar cascade face detection algorithm to sort and cluster out these entities.

Haar cascades is a cascade classifier that implements a machine learning approach based on the Adaboost meta-algorithm. The rectangular shape of the face is meaningful in initializing the classifier. Further, the algorithm focuses on the property that the eyes region is often darker than the face and nose region. The second feature proposes that the eyes are darker than the bridge of the nose. Similarly, this approach finds the entity's possible relations and features and records the features for further prediction. Once the face is detected, we can obtain the face location from the origin (center of the image).

Stereo Vision - Stereo vision compares the information about a scene from two vantage points and examines relative positions of objects in the two panels. 18 Figure 5.1: Stereo vision showing the deviation from origins of both the cameras An image can be termed as a grid of pixels within some range of indices. Using face detection, we narrowed down the object's position in the grid of pixels (x, y).

Stereo vision gives two images of the same scene from different positions in the same plane. Each image gives the deviation of the image from the origin of that respective image.

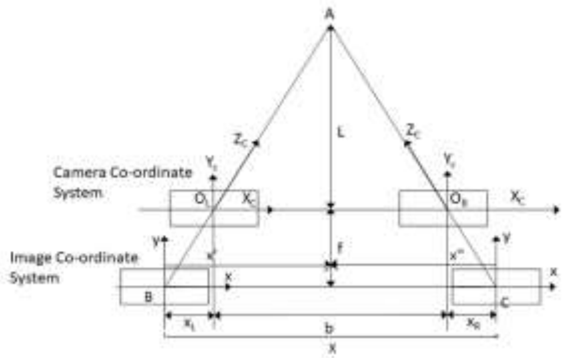

Fig 1. Depth sensing geometrical model

Figure 1 shows us the geometry behind the stereo vision method for depth sensing.

Here, x = Distance between two webcams.

f = Focal length of the webcams.

XL = Deviation of the image from the origin of the left webcam.

XR = Deviation of the image from the origin of the right webcam.

Stereo vision depth-sensing works on the principle of angle-angle symmetry (AA symmetry criterion) of two triangles. Using the AA symmetry criterion, we can state that if the angles of the two triangles are congruent, then the third angle of both triangles must be the same.

Hence, the ratio of each parallel side of triangles is equal. i.e., ab / c

$$\frac{AB}{DE} = \frac{BC}{EF} = \frac{AC}{DF} \qquad (1)$$

Hence from figure 1, we can prove that,

$$\frac{f}{z} = \frac{X_L}{X'} \quad (2)$$

Where, z = f + L,

Similarly,

$$\frac{f}{z} = \frac{X_R}{X''} \quad (3)$$

From equation 2 and 3, we can say that,

$$x' = \frac{z \times X_L}{f} \quad (4)$$

$$x'' = \frac{z \times X_R}{f} \quad (5)$$

Addition of eqn. 4 and 5 gives us x,

$$\therefore x = \frac{z \times X_L}{f} + \frac{z \times X_R}{f}$$

$$\therefore x = \frac{z}{f} \times (X_L + X_R)$$

Finally, we get depth (z),

$$\boxed{z = \frac{x \times f}{X_L + X_R}} \quad (6)$$

2. Orientation angles and Depths for individual speakers

As we discuss throughout the paper, our ultimate goal to achieve optimal surround sound is to implement a mechanism that will change the direction of speakers and sound levels to adjust the sound pocket over the listener's head.

In the previous section, we proved how we achieve the depth-sensing from a reference point, i.e., stereo cam. But we need to translate this depth into tilting and panning angles and depth of listener for an individual speaker. It is achieved using simple Pythagoras and triangle formulas.

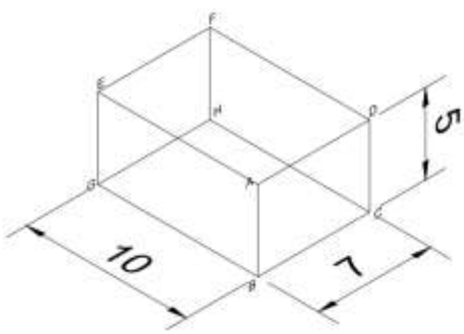

Fig 2. Reference Room

Let's consider a cuboidal room with dimensions as shown in the figure 2, Height = 5m Width = 7m Depth = 10m Denoted as ABCDEFG,

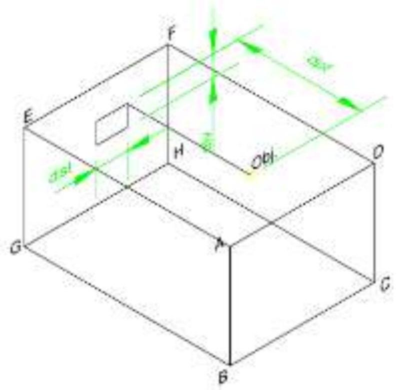

Fig 3. Object

Consider an object Obj. located at some elevation elv, depth dpt and distortion dist from the center from the center of the plane EFGH.

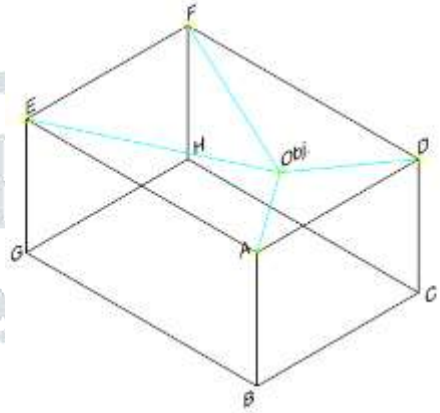

Fig 4. Expected Result

Considering E, F, D, and A are point sources (speakers) that need to be oriented so that the line drawn from their center axis intersects the point Obj.

Hence the audio signals transmitted through speakers focus on the Obj. The distance from each point source to the Obj. is unknown. This above unknown helps speakers adjust their surround sound pocket dynamically around the listener's head according to its position by varying orientations of a speaker and sound levels of each speaker from the unknown depths from each point source.

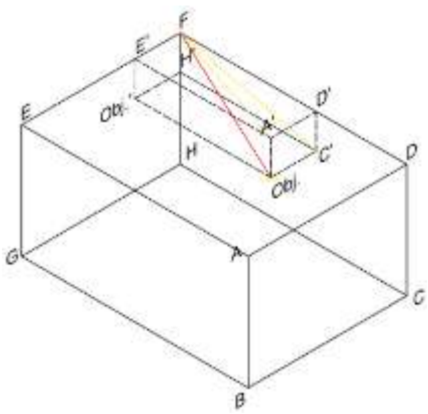

Fig 5. Depth and Angles for an speaker

To find the orientation angle, we need to know the ∠A'FD' and ∠C'FD' so the resultant ∠A'FObj. It will be the orientation angle where the speaker needs to face. Moreover, to adjust the sound level, we need to find the length of the line FObj. These orientation angles and lengths can be calculated using the dimensions of cuboid A'D'C'Obj. E'. Similarly, we can form these types of cuboid for each point source, something like this

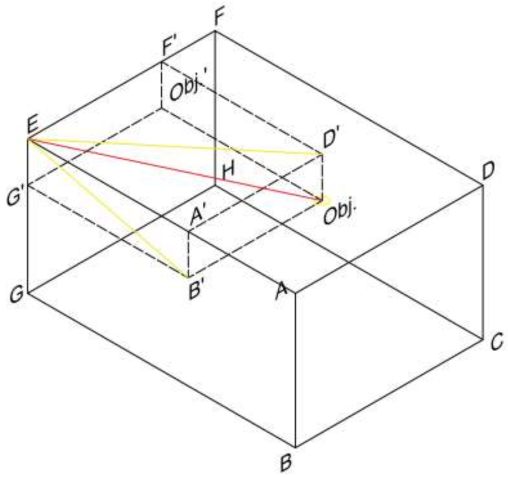

Fig 6. Depth and Angles for an speaker

To get the dimensions of each cuboid, we need to know :distance of each speaker from the center of the plane EFHG. Example, For point-source F (Speaker F), we need speaker elevation distance CamEF0 and Distortion to the right CamFH0 . In point source D (Speaker D), we need three constants: room depth CamCam0 , speaker elevation Cam0AD0 and Distortion to the right Cam0DC0 .

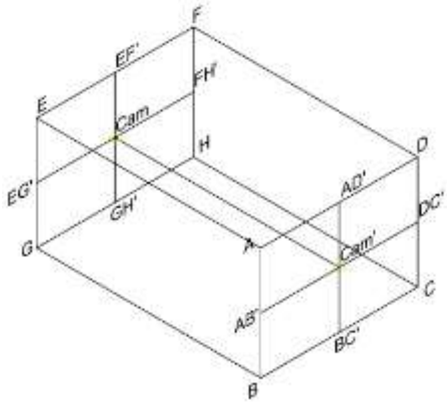

Fig 7. Constants

Moreover, the three variables are defined in Figure 7 (Object). These three variables are nothing but the dimensional distance (XYZ coordinates) from the Center of the plane EF HG. The Center of the plane simulates the stereo camera, which will measure these three coordinates (XYZ) concerning the camera. Now consider, above Figure 7, that we got some constants to calculate the panning and tilting angles. Those are speaker elevations and speaker deviations from the reference point (i.e., point of stereo vision camera). It has four simple geometrical steps

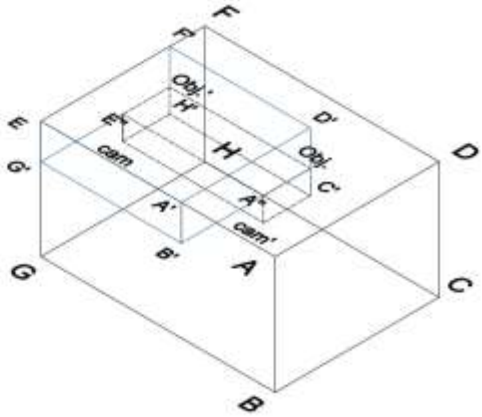

Fig 8. Projection of Object with respect to reference point and speaker

Figure 8 simulates the listener tracking model, from which we get the elevation of the object denoted by projection Obj.0H0 , the deviation denoted by projection Obj.0E00 and the depth Obj.0Obj. from the reference point i.e. stereo cam . We can draw a cuboid,

$$EF''Obj.'G'A'D'Obj.B'$$

This cuboid defines the projection plane and actual plane of the Object (Listener) for the reference point (camera) in the cuboidal room. Using this projection we can draw the projection with respect to the speaker as shown in Figure . Hence we can get the dimensions of projection cuboid of speaker by

$$Obj.B' = A''B' + Obj.A''$$

Where, Obj.D0 can be defined as the distance of object from the ceiling. (Here we are considering that speakers are mounted on the ceiling and the center of axis of speakers are the corners of the room) and Obj.B0 is nothing but distortion of object from the side of the room and last the depth Obj .0Obj. defines the distance of object from the speaker if we draw a perpendicular line from the TV plane. Using above equations we can form a new cuboid,

$$EF''Obj.'G'A'D'Obj.B'$$

Now using the dimensions of cuboid 5.10, we can calculate the tilting ∠A'EB' and panning angles ∠A'ED' and the resultant depth of object from speaker if we draw a direct line from speaker to object. Calculating panning and tilting angles Let's consider a 4A'ED' is a right angle triangle, where ∠A'ED' is panning angle

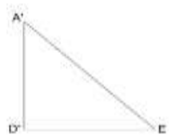

Fig 10. Depth and Angles for an speaker

$$tan\theta = \frac{A'D'}{D'E}$$

Where,

$$A'D' = Opposite of \angle A'ED'$$

$$D'E' = Opposite of \angle A'ED'$$

Similarly we can find the panning angle α using 4A'EB'. After the orientation of individual speakers using above methods, to create a balanced sound pocket we need to adjust the sound levels of individual speakers. To adjust the sound levels we need the distance of the object from the individual speaker which is EObj.. Which we can obtain using 4D'EObj., but first we need to know the length of D'E' ,

$$EObj. = \sqrt{D'E^2 + D'Obj.^2}$$

4. Where D is the maximum Audible Distance of a Speaker. Once the Resistance is Adjusted the output is made available at the input Audio amplifier (LM386) and is amplified by the required Gain factor. The final output is given to Speakers to observe the desired output.

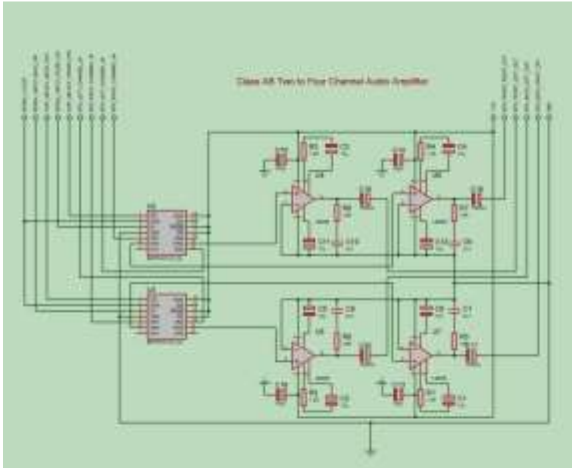

Fig.12. Complete Circuit diagram

A. Circuit Design

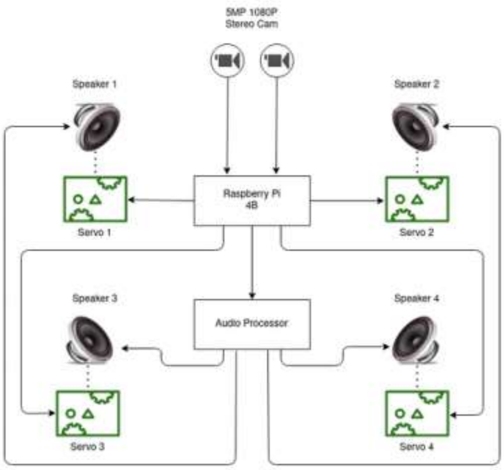

Fig.11 Block Diagram

The controlled output from Raspberry pi is given to the 2-ch digital potentiometer. Here digital potentiometer is used in potentiometer mode also known as the voltage divider mode with input At terminal A and terminal B taken As a reference or vice-versa The calculation of resistances can be analyzed using an example. Suppose the maximum audible range of the speaker is 100 units.

1. MCP42010 Offers 256 contact points with a max resistance of 10k so 1 contact point equivalent to a distance of 0.390625 units and resistance of 39.0625 ohms.

2. For Example, if a person is at a distance of 70 units then we need to keep the wiper at 76.8 contact point can be approximated to 77 and the resistance is equal to 3k.

3. The above statements can be formulated as:- No of contact points (K) = 256 - (D/0.390625) Resistance = K*39.0625 ohms

B.Software Design (Flowchart)

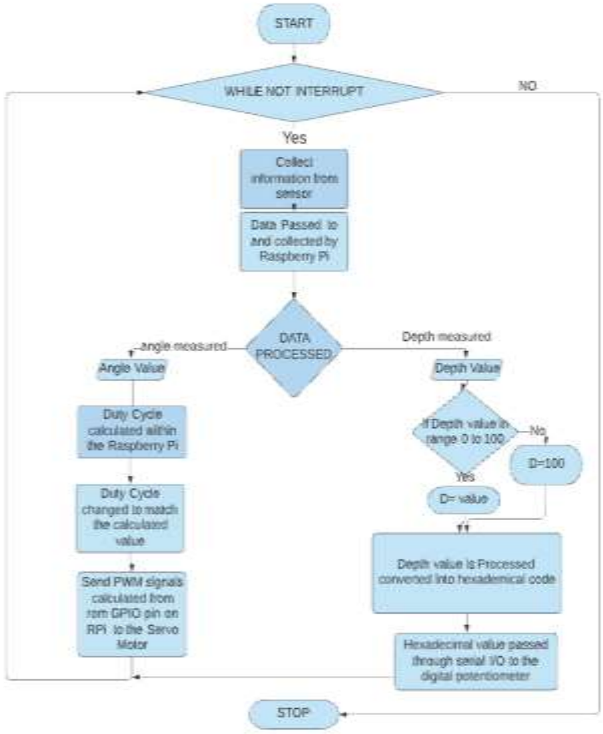

Fig 13. Flowchart for software implementation

IV. TEST RESULT

The constants for the :

1.Room dimensions :11m x 6.5m x 5m
2. Horizontal distance between speakers and the center of the plane: 3.25m
3.Height of the speaker from center :2.5 m

Table 1. Observed sound intensity with respect to different object perpendicular distance from the plane

Actual Depth (m)	Measured Depth (m)	Measured sound intensity(db)
5.5	5.335	52.575
7.8	7.566	65.338
10	10.3	79.849

Table 2 . Observed Panning angle with respect to different object position

Actual Depth (m)	Speaker to centre of the plane distance- Deviation of object from the centre (m)	Angle of panning (Output in degrees)
5..335	3.632	55.75
7.566	1.550	78.42
10.3	0.2	88.89

V. CONCLUSION:

1. The system designed can locate the dynamic position of the user within the room where the system is installed

2. The dynamically mounted speakers direct themselves automatically towards or in the direction of the user based on the user's position.

3.The intensity of sound gets adjusted dynamically depending on the distance of the user from the speakers to provide the user a best surround sound effect.

VI. ACKNOWLEDGEMENT:

It is our pleasure in expressing sincere and deep gratitude towards our guide **Prof. Deepali Yewale** Professor Electronics & Telecommunication Engineering Department for her valuable guidance and constant support throughout this work and help to pursue additional studies in embedded systems. We take this opportunity to thank the Head of Department **Dr. M. P. Sardey** and all staff members of the department of Electronics and Telecommunication Engineering AISSMS IOIT, Pune, for cooperation provided by them in many ways. The motivation factor for this work was the inspiration given by our honorable principal . Lastly, We are thankful to those who have directly or indirectly supported our work.

VII.REFERENCES:

[1] IEEE paper of “CHANG HA LEE" (Location-Aware Speakers for the Virtual reality Environment). Date of Publication: - 22 FEB 2017.

[2] Research paper of " R.M. Aarts, W.P.J. Bruijn , W.J.Lamb , A. Sakari Harma " (Surround sound system) . Date of publication: -16 SEP 2014.

[3] Research paper of "Jahnke, Steven.R" (Dynamic Sound source and listener position (DSSLP)) . Date of publication :-17 AUG 2005.

[4]Surround sound systems (United States Patents On, September 16, 2014)

[5]Shadow Sound System Embodied with Directional Ultrasonic Speaker (ICISA.2013 on 2013)

[6]An Efficient Implementation of Acoustic Cross-talk Cancellation for 3D Audio Rendering (IEEE China SIP on July 2014)