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

Nachiket Kamod (12), N V Roshni (31),
Ganeshsai Muppasani (30), Jagrati Chowdhari (9),

Guided By,
Prof. Deepali Yewale

Academic Year
2020 - 21

Contents

List of Figures

Chapter 1

1.1 INTRODUCTION

Automation plays an important role in the world economy and in daily experience in the last few decades has witnessed a rapid development in audio systems.

Journey of audio systems begins with single channel audio system (monaural audio system) in year 1877, later in the year 1931 two channel Audio system (Stereo audio system) were introduced and in the year 2005 the most advanced air audio or surround sound system (Multichannel audio system) were introduced.

Modern sound systems are increasingly gaining popularity day by day, particularly since technical advances have lowered their prices, increased their qualities and features. One barrier to the greater experience while using these sound systems is its static nature in case of surround sound. Surround sound is a system of stereophony involving three or more speakers surrounding the listener so as to give a more realistic effect.

Even though high end audio systems provide good quality of sound but to achieve the surround sound effect the user has to configure the system manually depending upon his current position which is a very Tedious task. Whenever you are settling up your complex home theatre bundle, understanding the art and science of speaker channel and placement is the most critical step in enjoying your new sound system.

Current sound system needs manual setup according to the ideal sitting position to achieve a good sound effect at fixed position. This manual setup consists of speaker angles and speaker sound adjusted to create and sound pocket around fixed position. But this effect varies when we move away from the surround sound pocket created by speakers.

The aim of this project is to develop a real time system to determine the listener's position and distance from the speaker system. The real time system described here is based on ultrasonic sensors mounted onto servo motors controlled by separate micro controllers (slave). The core of the system is a master micro-controller termed as an audio processing unit which controls the speaker direction as well as sound levels. Distance measured from the rotating ultrasonic system, allows us to map 2 dimensional maps of the room in real time. With less system requirements, the surround sound pocket can be adjusted according to the listener's position.

1.2 ABSTRACT

To explain the problem statement briefly, consider a real life scenario, I configured the orientation and volume levels of my sound system in order to get perfect surround sound at some position where I usually seat. But, now I want to change my seating position or even arrangement to some other part of my room which can be far right or left or may be forward or backward from the last seating arrangement. In this case to get perfect surround sound I need to reconfigure my speakers again, that is, its orientation and volume levels as per the new seating position either with the help of technician or self which is mostly manual adjustments.

To overcome this scenario we experimented a combination of stereo vision and hardware technology which responds to real time movements of listener

and dynamically adjust the sound pocket. This system uses opencv face detection algorithm and simple geometrical formula to calculate depths and angles for individual speaker to introduce dynamically adjusted surround sound. Since the system avoids the heavy usage of hardware, complex algorithms and machine learning approach it can be implemnted on low powered micorprocessors as well same processors used by sound systems.

Chapter 2

LITERATURE SURVEY

A. Surround sound systems

(United States Patents On, September 16, 2014)

This paper propses an idea of developing a system that comprises of reciever for receiving a multichannel spatial signal that comprises at least one surround channel.

This system comprises of a directional ultrasonic transducer for emitting ultrasound towards surface to reach a listening position via a reflection of the surface and a driver ckt to drive ultrasonic transducer.

The proposed system is capable of producing virtual surround sound without requiring a speaker a speaker to be located .

B. Shadow Sound System Embodied with Directional Ultrasonic Speaker

(ICISA.2013 on 2013)

The paper talks about usage of 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 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 system efficiently.

C. An Efficient Implementation of Acoustic Crosstalk Cancellation for 3D Audio Rendering

(IEEE China SIP on July 2014)

In this paper given method the use of 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 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 system efficiently.

D. Multirate adaptive filtering for immersive audio

(IEEE Xplore on February 2001)

This paper describes a method for implementing immersive audio rendering filters for single or multiple listeners and loudspeakers.

In particular, the paper is focused on the case of 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.

It describes transform-domain LMS adaptive inverse algorithm that is designed for crosstalk cancellation necessary in loudspeaker-based immersive audio rendering.

The algorithm used in this paper is only for single listener and only two loudspeakers are used.

2.1 CONCLUSION

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

- Complexity of Hardware, as per the research study we can observe the research is based on mono channel and not multi channel.
- Research suggests use of kinect for object tracking which have its own drawbacks.
- Complexity of Algorithm, to achieve the effect researchers suggested very complex algorithms regardless of room geometry.

Chapter 3

AIM AND OBJECTIVES

3.1 Aim

To develop a real time self adjusting Audio system based on listener's position to achieve high quality air sound effect.

3.2 Objectives

1. To introduce automation into current trend of audio system.

2. To make the audio system compatible of adjusting it's orientation or direction and sound intensity based on user's position.

3.3 Methodology

From abstract we can conclude speaker angles and sound intensities of individual speakers are very much important.

Speaker angles defines how the sound is gonna reach to the listener. Like is it reflecting from any surface or the sound source is directly pointed towards the listener.

Sound is nothing but oscillations of particles (typically air) in vibrational motion, which transports energy through a medium.

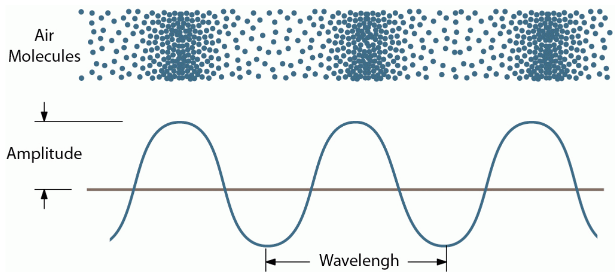

Figure 3.1: Sound Waves

Speakers push and pull surround air molecules in waves to generate a sound wave using a diaphragm. Typically, this diaphragm is in conical shape hence it oscillates the molecules in the oval field.

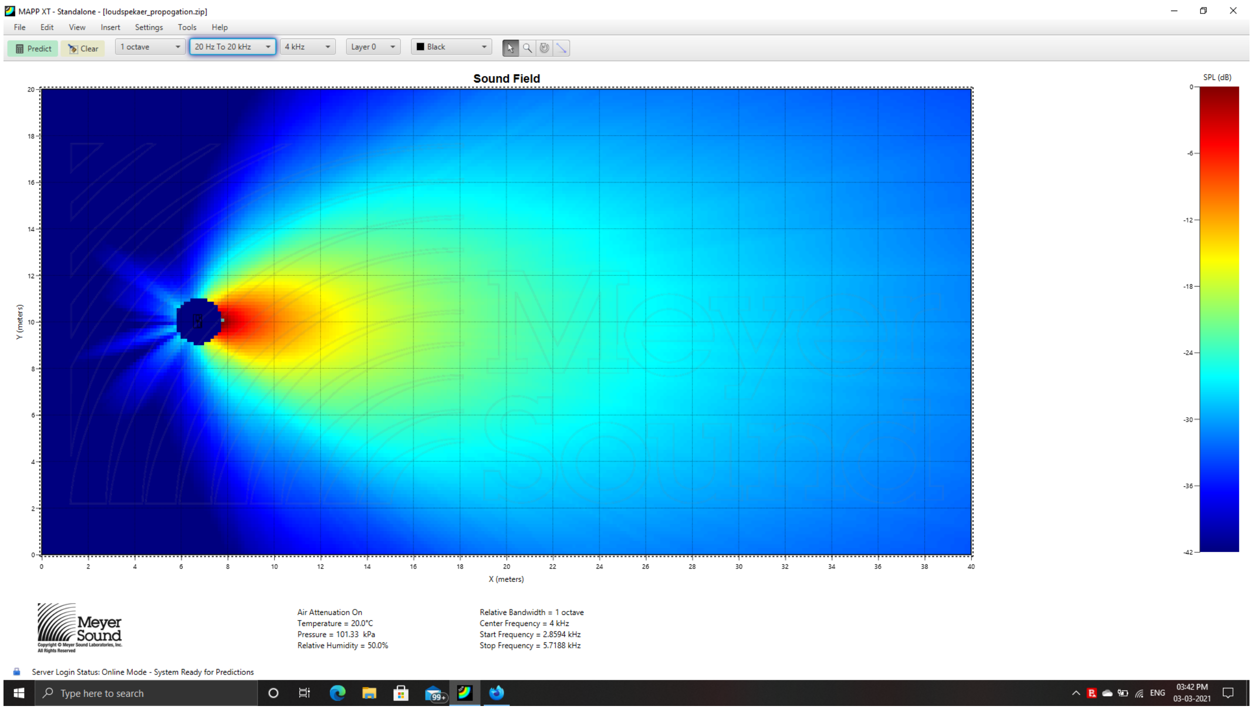

Figure 3.2: Sound propgation of speaker

Figure 3.2 shows the oval propgation of the sound field from the speaker. Where sound intensity of speaker at some depth is denoted with heat map (dB). In front of the speaker the sound intensity is maximum and it fades away as we go far away from the speaker. Where in the other hand it is much less at the back of the speaker. Since, it is oval in nature, the propogation to left and right is also less than that of front.

Figure 3.3 shows the reflection and reverberation of sound due to misalignment and excessive sound intensity of speaker due to collision of sound waves onto walls. These reflections build up with each reflection and decay gradually as they are absorbed by the surfaces of Objects and walls in the room. In this case, listener tends to hear direct sound and the repeated reflected sound waves which might sound muddy and grabbled.

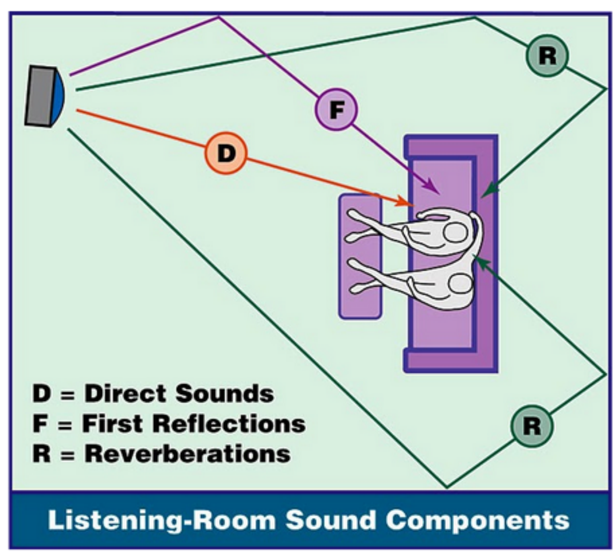

Figure 3.3: Reflection and reverberation of sound

Hence, it becomes necessity to align the speakers as well as adjust the sound levels in proper amounts to get best surround sound.

To overcome this scenario we experimented a five block system. Which consists of,

1. Depth estimation unit (Cameras), to measure the depth of the listener from one reference point and feed these varibles to microporcessor for further calculations.

2. Microprocessor, to measure depth and calculate panning and tilting angles as well as listener's depth from each speaker.

3. Mechanical unit, to pan and tilt the speakers.

4. Audio Processor Unit (Digitally controlled amplifier), to adjust the individual speaker gain using calculated results from microporcessor.

5. Speakers, to sound individual 4 channeled output.

3.4 Specifications of the System

3.4.1 Depth estimation unit

Web Cam

(LAPCARE LAPCAM)

Figure 3.4: Web cam

1. 1280 x 720 pixels @ 720p resoultion

2. Automatic low light correction

3. Plug and play linux compatible, High-Speed USB 2.0

3.4.2 Microprocessor

Microprocessor serves an important role in DAC application, data processing estimation and controlling the response hardware in real time.

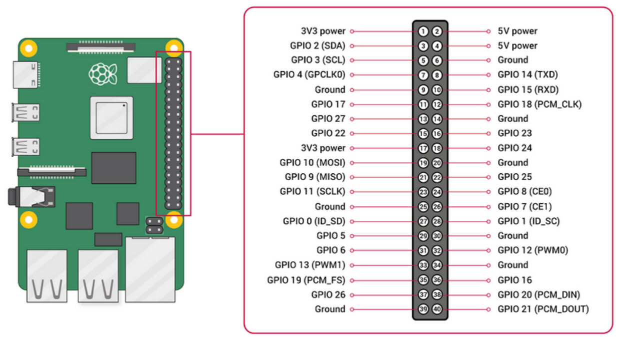

Figure 3.5: Microprocessor (Raspberry Pi 4B 4GB)

Raspberry Pi 4B 4GB RAM model comes packed with,

1. Quad core Cortex-A72 64-bit @ 1.5 GHz clock and uses ARM v8 architecture, with 4GB LPDDR4-3200 SDRAM.

2. 2.4 and 5 GHz IEEE 802.11ac wireless wifi hardware.

3. 2 Micro HDMI ports.

4. H.265 (4kp60 decode), H264 (1080p60 decode, 1080p30 encode).

5. OpenGL ES 3.0 graphics.

6. Micro-SD card slot for loading operating system and data storage.

7. 4 USB ports.

8. Software PWM on all pins and Hardware on GPIO12, GPIO13, GPIO18, GPIO19.

9. SPI

- SPI0 : MOSI (GPIO10), MISO (GPIO09), SCLK (GPIO11), CE0 (GPIO08), CE1 (GPIO07)
- SPI1 : MOSI (GPIO20), MISO (GPIO19), SCLK (GPIO21), CE0 (GPIO18), CE1 (GPIO17), CE2 (GPIO16).

3.4.3 Mechanical unit

Servo Motors

(SG90 Servo)

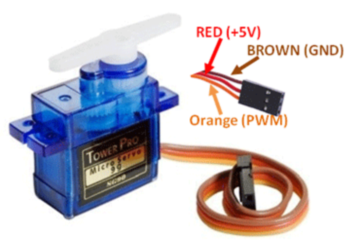

Figure 3.6: SG90 Servo

1. 180° rotation (90 in each direction).
2. Torque 2.5 kg-cm
3. Volatge 4.8-6 V
4. Speed 0.12 sec/60°

3.4.4 Audio processor unit

Audio Amplifier

(LM386N-1)

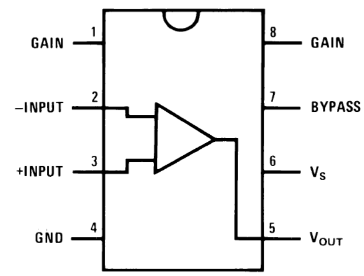

Figure 3.7: LM386N-1 pinout

1. Operating Supply Voltage (Vs) 4 - 12 V
2. Voltage gain 20 - 200
3. Output power 325 mW

Digital Potentiometer

(MCP42010)

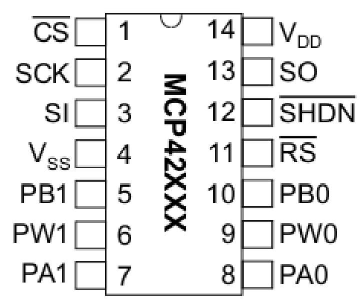

Figure 3.8: MCP42010 pinout

1. Potentiometer values 10 kΩ
2. 256 taps for each potentiometer
3. 2 channel
4. SPI serial interface (mode 0, 0 and 1, 1)
5. Single power opeartion (2.7V - 5.5V)
6. Industrial tempearature range: -40°C to +85°C
7. External tempearature range: -40°C to +125°C

3.4.5 Speakers

Final component is the speakers. Which is mounted on servo and provided with dynamically controlled sound using Audio Processor. They are mounted on four corners of room to form a four channel audio system.

For this application we are using 4 Ω speakers to deliver four channel output.

Chapter 4

BLOCK DIAGRAM OF THE SYSTEM

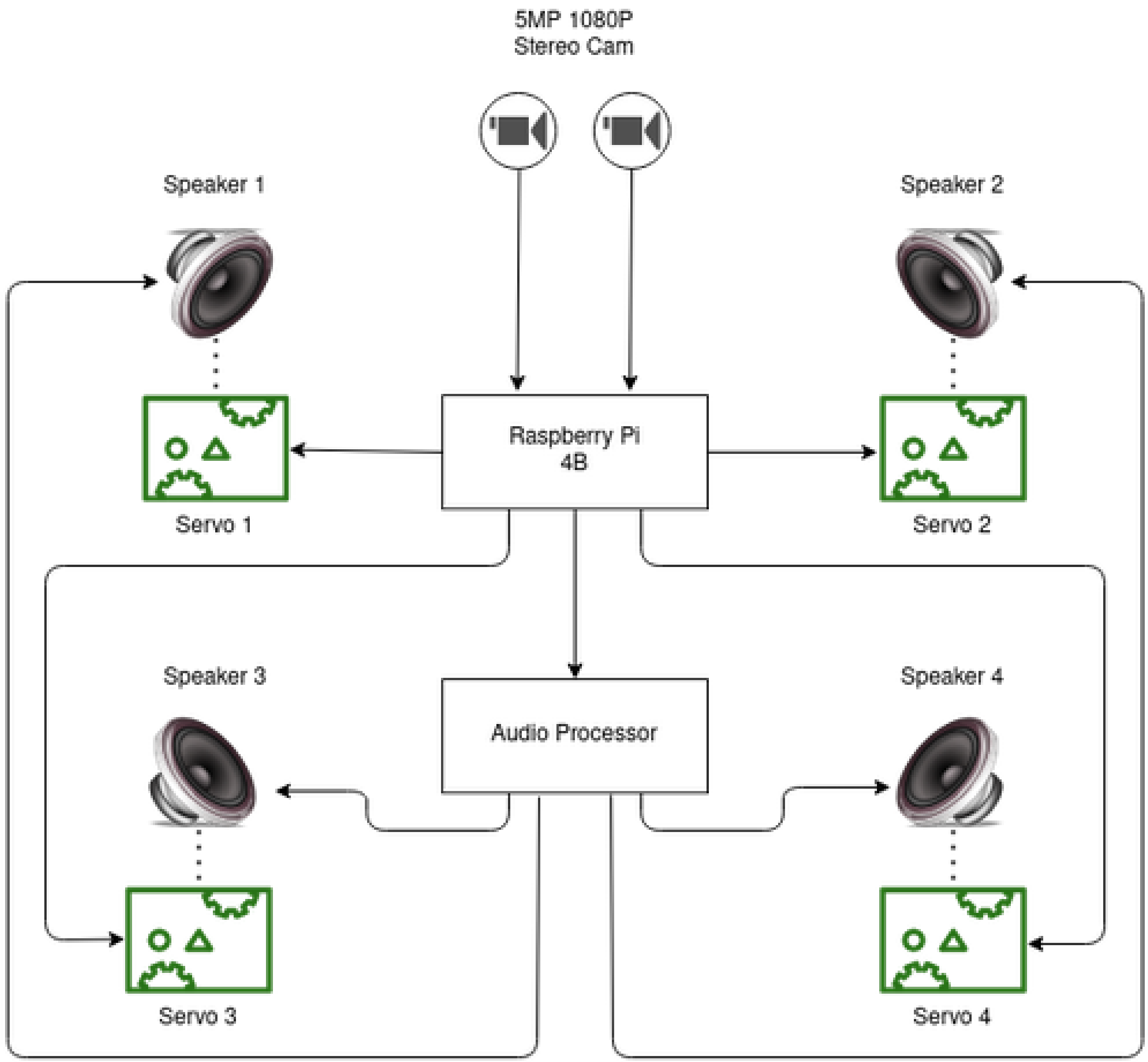

Figure 4.1: Block diagram

Dynamic audio system can be simplified as five measure blocks, each serving it's own application in order to provide dynamic surround pocket over the listener's head,

4.1 Depth estimation unit

Depth estimation unit includes stereo vision assembly of two web cams with similar (known or unknown) focal length (In our case 18 cm).

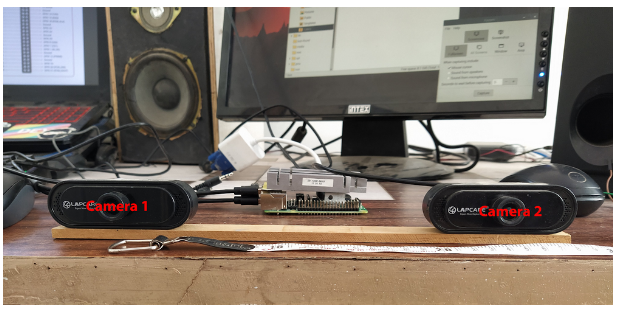

Figure 4.2: Assembly of the depth estimation unit

Placed at some known distance x from each other (24 cm). And implements computer vision (opencv) and geometrical equations to measure the depth, deviation and hieght of a object from a reference point. Where the reference point is the center of the stereo vision system.

4.2 Micro-processor

Micro-processors acts as middleware between DAC and the hardware to control speakers. In this project we are using raspberry pi 4B (Quad core Cortex-A72 64-bit @ 1.5 GHz clock) as our micro-processor. Webcams are connected through USB 2.0 of RPi.

First we find depth, deviation and height of the listener's face from the reference point using opencv's frontal face harrcascade classifier followed by open source AA symmetry algorithm of geometry, where opencv also helps to classify between listener and other objects.

Above process provides us three real-time variables,

1. Depth of the listener's face from the reference point.

2. Deviation of the listener's face from center of the axis.

3. Height of the center point of listener's face from the origin (reference point).

Further using this real-time variables, and some constants (room dimensions and speaker positionings), using customly designed geometrical algorithm we can calculate, panning and tilting angles, depth of the listener from each speaker.

Using panning and tilting angles we can rotate the servos to the required angles to direct the sound field towards the listener. And using depth we can adjust the sound levels of the speakers by controlling input voltage of amplifiers digitally.

4.3 Mechanical Unit

As discussed in methodology, as the speakers propagates sound in oval. Hence shape we need to align the major axis of the sound field towards the listener.

Mechanical units assembles with, two servo motors for each speaker (channel) one for panning and second for tilting.

Servos are connected to hardware PWM pins of the RPi and controlled in real time using feedback of the angle algorithm.

4.4 Audio Processing Unit

Usually an surround sound system contains two or more speakers in order to generate sound effect of moving objects from one place to another.

Even if we direct the speakers towards listener's direction, it is encessary to adjust the sound levels of each speaker according to the depth of the listener from each speaker.

To genrate best surround sound effect, this Audio processor unit assembles with 4 class AB audio amplifiers driven by 2 two-channel digital potentiometers for controlling sound levels of each individual speaker to adjust the sound pocket over listener's head (ears).

4.4.1 Audio amplifier

Basically, audio amplifier is an circuitrary which is designed to increase magnitude of applied signal in order to power a low resistance load (speakers).

Sound signals are applied to non-inverting terminal of an amplifier through an voltage devider circuitrary (potentiometer). This voltage devider adjusts the voltage levels of the input signal resulting the change in volume levels at the output. This change is inversly proportional to the resistance at wiper terminal of the voltage devider.

For this application we are using LM386N-1 as our amplifier.

4.4.2 Digital potentiometer

Digital potentiometers mimics the analog functions of a mechanical potentiometer. Where the resistance is controlled by micro-controllers or micro-processors.

As we discussed, to adjust the sound output of the audio amplifier we adjust the input voltage given to the non-inverting terminal of the amplifier. Hence, we supply the audio signal to amplifier through an digital potentiometer, so we can increase and discrease input voltage and hence the sound levels of speaker digitally using a micro-controller or a microprocessor.

For this application we are using SPI compatible MCP42010 Digital POT,

4.5 Speakers

Speakers serves the 4 channeled dynamically adjusted surround sound to the listener. Usually they are mounted on four corners of the room either at the ear levels of the listener or near the cieling.