

Natural Spatial Image, Envelopment and Depth In 22.2 Multichannel Recordings

ABSTRACT

It is acknowledged that sound arrives at our ears, and that the human brain processes the information in three dimensions. So then why do nearly all previous multichannel systems only playback content in two? The new 22.2 format improves upon these previous multichannel systems by being periphonic or three dimensional, therefore creating greater naturalness and envelopment in playback. The psychoacoustic laws behind spatial imaging within the three dimensional sound field should first be understood, and also the recording techniques devised to capture it. Then the best ways to utilise these techniques relative to this new system are discussed. Finally it is proposed that an omnidirectional microphone array and a multilayered ambience array are used in order to create natural spatial image, envelopment and depth.

0. INTRODUCTION

In the world around us, sound and its sources are based in three dimensions, and contain three mathematical coordinates, x, y and z. This is classed as a periphonic system. On the other hand, conventional multichannel systems such as 5.1 have the sound sources (loudspeakers) arranged in two dimensions and therefore produce sound localisation only in these two dimensions, this is called a pantophonic system [1].

To combat this problem and to create a system that more closely approximates reality, researchers developed a more periphonic system, 22.2 multichannel sound [2]. This system contains loudspeakers arranged in all three dimensions, with nine upper layer channels, ten middle layer channels and finally three lower layer channels (Fig. 1). It can be shown that this can produce a more idealistic representation of a natural sound space [2].

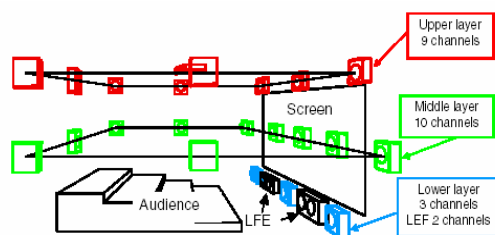


Fig. 1: The 22.2 Sound System [2]

This paper will investigate how to produce recordings for this new system that contain envelopment and natural spatial qualities that more closely relate to this ideal.

The first chapter of this paper will give some background into 22.2 multichannel sound and its advantages over existing multichannel systems. Chapter two will then move on to deal with the psychoacoustics of envelopment and depth, how humans hear these qualities in sound, and will raise the question: What is contained in the audio information of the vertical plane? In chapter three the paper will examine previous recording techniques, and then relate them to use with the 22.2 multichannel system.

1. THE 22.2 SURROUND SOUND SYSTEM

This system was produced by researchers at the NHK Science and Technical Research Laboratories in Tokyo, Japan [2]. It was made to further multichannel audio technologies and also to compliment the “ultrahigh-definition video system” (Appendix i) [15]. The main goals of the system were to create a stable frontal sound image and reproduce a natural three dimensional spatial impression. Also similar to Griesinger [9] it was decided that the system would be for theatre use

and therefore require a large listening area. This is because the listeners would be located at points away from the centre of the loudspeaker array.

The method that the researchers used was to have three layers of speakers. A top layer to give a vertical plane to the sound field created (Fig. 2). The middle layer to produced sound is a similar way to current multichannel system, with the bottom layer further enhancing the vertical localisation [2]. It was shown by Hamasaki [2] that this system will produce a sound field that more approximates the “everyday three dimensional sound experience” stated by Rumsey [5] and expanded in chapter 2 of this paper.

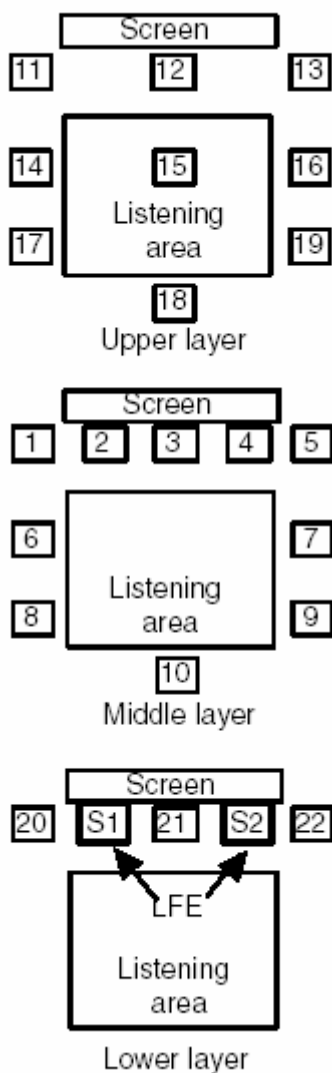


Fig. 2: Diagram showing placement of 22.2 channel loudspeakers [2]

It should be noted though that the three front channels (2, 3 and 4 in the diagram) were produced through phantom imaging. The phantom images were generated by summation of the outputs created by the lower and upper speakers. These speakers were arranged vertically symmetrical, in line with Theile findings on localisation of lateral phantom sources [4]. Phantom sources were used due to need for the three speakers to be located behind the projection screen, and the fact that this screen did not pass sound [2]. Martin [3] showed though that while the focus (“the perceived definition of the source’s boundaries and details” [3]) of this phantom image would be high with non coherent radiating sources, it could lose focus if the sources were coherent.

Also during the development of this system it was found that the production methods already in place for previous multichannel sound were not suitable for this new system. Due to this an integrated surround panning system was produced in parallel with the system [2]. The new panning system consists of a level, distance and pan control. While level (gain) and pan are similar to the existing mixing systems, distance contains formulas to simulate air absorption and attenuation with distance. This models the general effect that source distance from the microphone has on the signal recorded, but fails to cover some of the more subtle effects such as less time between direct sound and reverberate sound, and attenuated reflection from the ground [5].

The findings stated that the advantages that this system holds over existing multichannel systems are:

- More focused “two dimensional frontal sound source localisation” [2]
- “The acoustical impression of elevation” [2]
- “Very natural spatial impression over a wide listening area” [2]
- “Three dimensional movement of multiple sound sources around the viewers” [2]

2. PSYCHOACOUSTICS OF ENVELOPMENT

Natural sound is a very enveloping experience with content being produced by many sources, and containing audio cues in all of the three dimensions [5]. These audio cues supply information that the brain then uses to localise the sound, and also decide what type of space it was produced in.

2.1 LOCALISATION

The sense of envelopment is very much linked to the localisation of sources in a sound space. If every source was to be localised to the same spot in space then the overall perceived sound would be a point source. The brain uses two main audio cues when localising sound; the time or phase difference between the signal at each ear, and the amplitude or spectral difference [6]. These two detection systems work in parallel with priority given to each depending on the source signal and environmental conflicts that may accompany these sources [5].

2.1.1 LOCALISATION THROUGH TIME DIFFERENCE

If a sound source is produced off axis to the listener the arrival time of the two signals reaching each ear will differ, this will be due to the distance that each signal had to travel to the relevant ear. The brain can use this timing information to localise the source that created the original sound (Fig. 3). Taking into consideration movements of the head it can be shown that this same system can lead to some localising of elevation, by tilting the horizontal plane that time difference localisation works within [7], but that most vertical localisation is done through vision.

Interaural time difference (ITD) though cannot clearly localise sources with a frequency higher than 1 kHz. This is because the difference in the distance travelled between the two ears would be higher or equal to half the wavelength of

the signal, therefore the two signals would be 180 degrees plus out of phase with each other. The brain cannot then decide which is the leading signal, so therefore cannot localise the source [12].

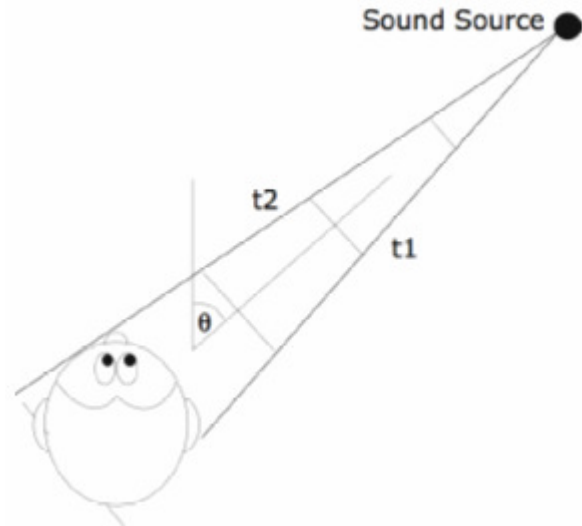


Fig 3: Time Delay Localisation [6]

2.1.2 LOCALISATION THROUGH FREQUENCY AMPLITUDE DIFFERENCE

At high frequencies the size of the head causes it to act as a barrier to the approaching sound (Fig. 4), also the shape of the pinna effects the spectrum of the incoming sound, these spectral changes form what is known as a head related transfer function (HRTF). The brain again uses these differences in signal to localise the source in three dimensional space [5]. In the opposite way to the ITD auditory system the brain cannot clearly localise in the lower frequencies using amplitude difference. This is due to the fact that the head no longer acts as a barrier to these frequencies because it is smaller than a 1/3 of the signals wavelength [12].

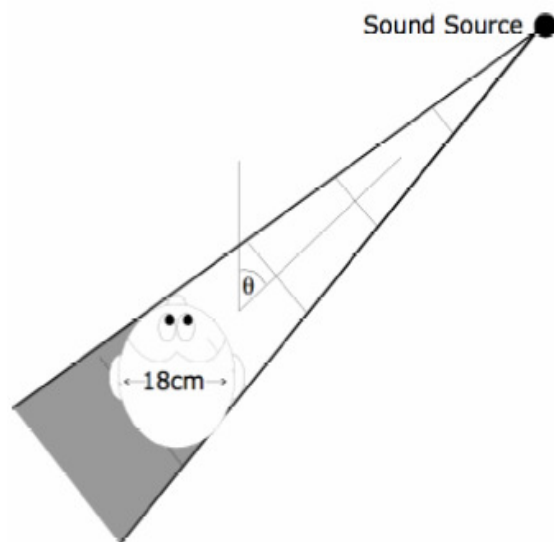


Fig. 4: Effect of head on approaching sound [6]

2.2 SPACIOUSNESS

Environmental spaciousness can be related to two fundamental audio cues, the early reflections and the reverberation tail. These two cues, which arrive in all three dimensions provide the brain with the ingredients to decode the space that the source is located and produce the feeling that the sound is outside or at a distance away from the head [8].

The early reflections are shown to produce the sense of the room type or “space”. These reflections occur between 20ms – 150ms after the initial onset of the source, and have a direct correlation to the original sound [6].

The reverberation tail is produced after the early reflections and is decorrelated with the original signal. It provides the size of the room and helps with the perception of distance from the source.

Griesinger states [9] that the spaciousness (meaning early reflections and reverberation tail) is of vital importance in the psychoacoustics of a listening area. The reflections that arrive during the sound (for example musical note), or within 50ms of it are used to perceive the distance of the sound. In addition to this the reflections that reach your ears in between the sounds gives the sense of envelopment.

It can be shown that the reflections that arrive within the source event (note) are perceived by the brain as being contained in the same sound stream or event as each other, while the brain perceives the spaciousness in-between as a separate background stream, whose spatial properties are detected independently [10].

2.3 DISCUSSION

It can be concluded that the human auditory system is most suited to localisation in the horizontal plane, and that most vertical localisation is done through visual not audio cues. However it was shown that the early reflections and reverb which are vitally important for envelopment and natural spatial quality reach the ears from all three dimensions, and therefore their information is found in the vertical plane. Taking this into account it is possible to see the scope for more enveloping and spacious recordings using a periphonic system that has three dimensions, such as 22.2 multichannel sound.

3. PREVIOUS MULTICHANNEL RECORDING TECHNIQUES

During the time recording of sound has been around, there have been many techniques put forward on how to capture the most natural depth, spaciousness, and envelopment in a recording. In this chapter various multichannel microphone techniques will be discussed, from basic stereophony to more modern multichannel methods. Then their uses for 22.2 multichannel recording will be argued.

3.1 BASIC MULTICHANNEL TECHNIQUES

The main point of any recording is mostly to recreate the natural impression of a sound in a space, whether it is a guitar in a bathroom or an orchestra in a concert hall. Therefore it is important for the recording

techniques used to create a natural depth, source, space and envelopment [10]. The basic techniques for recording a natural multichannel image is a main pair which is then combined with support microphones, and finally extra ambience microphones are added to create envelopment (Fig. 5) [11].

In this sub-section the differing types of microphone techniques for recording each of these elements will be discussed.

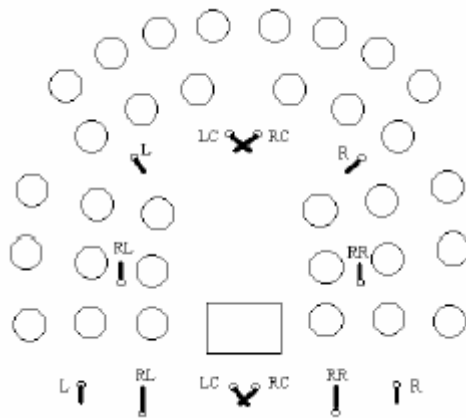


Fig. 5: Basic Multichannel Microphone Set Up [9]

3.1.1 COINCIDENT MAIN PAIRS

Coincident pairs consist of two directional microphones placed close together in space. These two microphones are then angled away from each other, and through changing the settings they can act well in many recording situations [5]. The pair creates a stereo image through level differences between the signal received at each microphone, and phase differences are only noticeable at high frequencies when the small space between the microphones becomes relevant [11].

This technique can be extended further by the use of near coincident pairs which introduce small timing differences by adding a small space in-between the microphones. This has the advantage of increasing the perceived spaciousness of the recording [13].

The advantage of the coincident system where this paper is concerned is that it can produce a good horizontal localisation

even if the listener is not seated in the centre of the loudspeakers. This is because if the source is located on the left, the amplitude of the signal will be greater in the left speaker upon playback. Due to localisation through amplitude difference (mentioned in chapter 2), the brain can still localise this source even if the listener is located nearer to the right loudspeaker. Near coincident techniques such as ORTF have been shown to collapse into the speaker that the listener is seated next to unless they are seated directly in the middle [9].

3.1.2 SPACED MAIN PAIRS

Spaced pairs normally contain two omnidirectional microphones that are separated by a gap. They create the effect of a stereo image by having a time delay between the signals picked up by each microphone as well as an amplitude difference [13]. Spaced pairs are often preferred by engineers because of the extended and flatter frequency response of omnidirectional microphones [5]. Griesinger [9] though concluded that to prevent the collapse of the stereo image mentioned above the pair needed to be widely spaced. This because the signal from the right loudspeaker arrives sooner and is louder than the signal from the right, therefore this counteracts the spatial cues that were recorded by the pair. If the spacing between the microphones is too large an unnatural sounding recording can be produced with the instruments occupying the extremities of the stereo image and a so called “hole” in the centre. The “Deca Tree” method used to counter this by introducing another omnidirectional microphone in-between the pair and slightly forward. This aids to fill in the hole [13] but can again result in an unstable image for listeners not seated in the centre of the multichannel set up [9].

3.1.3 SUPPORT MICROPHONES

If you take the example of a large orchestra in a concert hall, with the main pair located in front of the orchestra [13].

The distance from the front instruments to the pair may be quite short, but the distance between the pair and the rear instruments will be large. Rumsey [11] states that when recording if the microphone is located at a distance greater than the hall radius (the critical distance [6]) from the source that what is picked up by the microphone contain more reverberant energy than direct energy. This causes the source to sound distant and lack clarity.

To counter this problem engineers use spot or support microphones [14] to supply the direct sound, but as Griesinger [9] concludes this means that in the final recording most of the energy is from the support microphones and not the main pair. This seems to contradict that fact that the main pair is placed and used for its spatial image [5], because the main spatial image is created by the support microphones.

3.1.4 AMBIENT ARRAYS

It was proposed by Theile that an array of four (in 5.1 surround) spaced microphones arranged in a cross could be used for the recording of ambient sound. The ambient array could then be placed in a position which contain the most natural sounding reverb, or the so called “reflection hot-spot” [10]. Theile then sends the signal from each microphone to the corresponding left, right, left surround and right surround speakers in order to create a diffuse enveloping reverb. The advantage of this system is that you have control over the amount of ambience in you recording, and you can also easily place the ambience in the surround channels. Hamasaki [2] though proposes that the microphones in the ambience array should be faced away from each other and not in the crossed configuration because it models human hearing more closely.

3.1.5 MIXING ELEMENTS TOGETHER

Mostly the source is to be perceived in front of the listener so will be produced by the front channels (loudspeakers), with the surround channels (if the multichannel system contains then, such as 5.1 and higher) holding the supporting envelopment qualities such as reverb and reflections [14]. Griesinger [9] extended this by concluding that reflections that are located in the opposite front channel (for example front right if the source is located front left) also add to the perceived depth and naturalness of the source’s sound. Theile [10] also added that the natural room response should be preserved by delaying the signals from each microphone element and matching them to the room. He deduced that this technique “supports the naturalness of spatial impression”.

3.2 SUMMARY AND DISCUSSION

22.2 multichannel sound is a periphonic system; therefore the recording techniques for use with the system should mirror it, by also being periphonic. This means that to create a natural, enveloping recording the microphones should be placed in layered arrays.

The main pair needs to create a strong frontal image that will not collapse into the speaker that the listener is located nearest to, and also supply a natural depth and distance. Griesinger [9] put forward that for a recording that require a large listening space upon play back, “amplitude panning work, and time delay panning does not”. This means that the coincident pair would be most suited, the problem is that it does not translate well on to multichannel system, and the near coincident pair is normally used [14]. As mentioned in section 3.1.1 it can be shown that near coincident pairs produce a stable image in a multichannel system only if the listener is located in the centre of the speaker [9], and characteristic which 22.2 does not have [section 1].

The most suitable microphone technique for the main pair would be a spaced array

of omnidirectional microphones. These microphones would be located in a horizontal line, parallel to the source (Fig. 6). This array would work well because the image would spread nicely between the speakers, even for listeners off the axis of the central line.

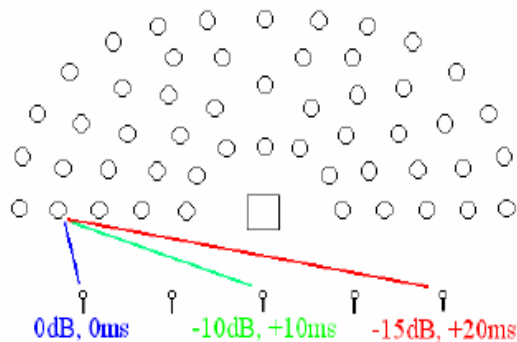


Fig. 6: An array of omnidirectional microphone [9].

The use of support microphones would have to be very carefully dealt with especially in the mixing and panning part, where they are used to build up spatial image [14]. This is because if they were not localised with care they could cause the frontal image to become unstable. They would be needed though because as mentioned in section 3.1.3, they produce the direct sound required to produce a detailed sounding recording. The integrated panning system (mentioned in

section 1) would be very beneficial if used because it could help in the positioning of the sources in the 22.2 surround sound field, which would normal be very difficult [15, 14].

The ambience arrays are the most important element for creating a surround recording with natural envelopment, and due to the fact that reflections arrive at are ears from all direction are extra important for the three dimensional 22.2 system [2]. A multi layered ambience array would suit this system, consisting of nine directional microphones on the upper layer, ten on the middle layer and three on the bottom layer. This would mean that the signal from each microphone could be sent to the corresponding channel. They would have to be widely spaced so that each signal has a large de-correlation to each other to create the most natural recording [9].

When the final mix was made it would also be important to delay the signals that arrive at each microphone so that a natural spatial depth and quality were produced [9, 10]. The creation of a natural image requires special consideration to the “delay situation” of the room (Fig. 7). This is because the lateral reflections lead to the perception of depth within a surround image [10].

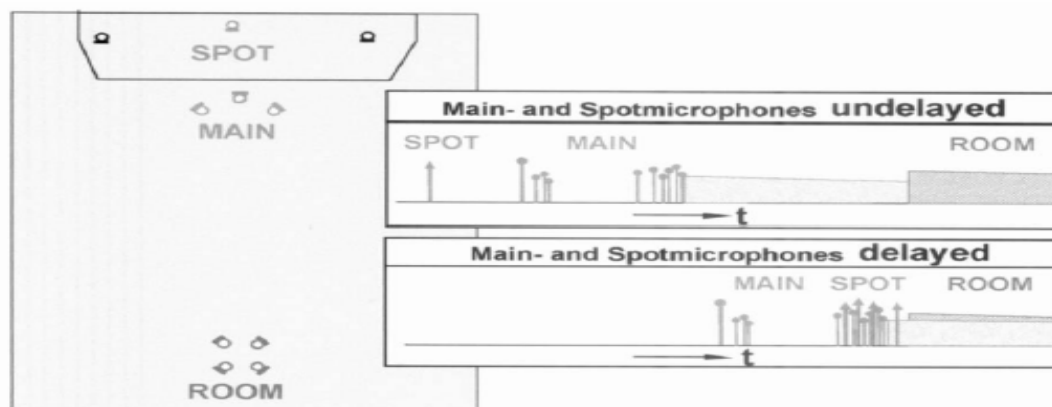


Fig. 7: Delaying each element of the recording to match the room related response. This is to create a pattern of reflections that resembles the original reflection pattern of the hall [10].

4. CONCLUSION AND FURTHER WORK

The 22.2 sound system improves on the previous multichannel systems by providing a technique that more closely models reality (section 1). This system can be used to create recordings that have a more natural sounding spatial image and depth because the reflections can be recreated in all three dimensions that they arrive at the ears in (section 2). It can be shown (section 2) that the frontal two dimensional image will also be improved because even though vertical localisation is mostly perceived through visual cue [5], it is reinforced by audio ones. Finally, due to the fact the sound it plays back is produced from multiple heights, the system creates a more realistic sense of envelopment than previous multichannel systems, such as 5.1.

The previous recording techniques that have been devised for multichannel recording can be further extended so that they are relevant to this new system.

This paper proposes a system in which five omnidirectional microphones act as a main pair, which is recreated on the frontal image of the 22.2 system [9]. Support microphones are then used to more fully define the image [5], which are localised with use of the systems integrated surround sound panning system. A multilayer ambience array provides the additional reflections that the system requires to produce a more enveloping recording [2]. These individual elements are then delayed and mixed together so that the room response can be modelled [10] and a more spacious final mix created.

To further this paper research would be carried out into the sound quality of the recordings created by the proposed recording techniques. Listening tests would then be performed and the recording procedure would then be reviewed and further improved. After this a final proposal of 22.2 multichannel sound system recording techniques would be made, with emphasis on natural localisation, depth and envelopment.

APPENDIX

i. Ultrahigh-definition video system.

This ultrahigh-definition video system was developed to realize a more realistic viewing experience [16]. To achieve this the researchers at the NHK Science & Technical Research Laboratories in Tokyo enlarged the horizontal viewing angle to 100 degrees. One hundred degree was decided after viewing tests into which viewing angle produced the highest sense of reality [17].

The system creates a very high resolution image of 4320 X 7680 which is more than sixteen times the resolution of the current high definition television. The systems specification are shown and compared to high definition television (HDTV) in Table 2 [2].

	HDTV	Ultrahigh-definition TV
Number of pixels	1080×1920	4320×7680
Viewing angle (pixel invisible)	30° Horizontally	More than 100° Horizontally
Comparison with movie	Equivalent to 35mm motion film	More than twice of 70 mm motion film

Table 1: Specifications of Ultrahigh-definition compared to HDTV [2]

REFERENCES

- [1] S. Changer, “*An Investigation into Adding Height Component To Surround Sound*”, Tonmeister Final Year Project (1999)
- [2] K. Hamasaki et al., “*The 22.2 Multichannel Sound System and Its Application*”, AES 118th Convention, Barcelona, Spain. Convention Paper (2005)
- [3] G. Martin et al., “*Controlling Phantom Image Focus in a Multichannel Reproduction System*”, AES 107th Convention, New York, USA (1999)
- [4] G. Theile & G. Plenge, “*Localization of Lateral Phantom Sources*” AES 53rd Convention, Zurich, Switzerland (1976)
- [5] F. Rumsey, “*Spatial Audio*”, Oxford Focal Press (2003)
- [6] T. Brookes, “*Acoustic Lectures*”, Tonmeister Course (2003)
- [7] J. Blauert, “*Spatial Hearing. The Psychophysics of Human Sound Localisation*”, MIT Press (1997)
- [8] D. Begault, “*3D Sound for Virtual Reality and Multimedia*”, London Academic Press (1994)
- [9] D. Griesinger, “*The Psychoacoustics of Listening Area, Depth and Envelopment in Surround Recordings, and Their Relationship to Microphone Technique*” AES 19th International Conference, pp 182-200, Elmau, Germany. (2001 June 21st-24th)
- [10] G. Theile, “*Multichannel Natural Recording Based on Psychoacoustic Principles*”, AES 19th International Conference, pp 201-229, Elmau, Germany. (2001 June 21st-24th)
- [11] F. Rumsey, “*Sound and Recording: An Introduction*”, Oxford Focal Press (2003)
- [12] D. Howard, “*Acoustics and Psychoacoustic: Third Edition*”, Oxford Focal Press (2006)
- [13] M. Gayford, “*Microphone Engineering Handbook*”, Oxford Focal Press (1994)
- [14] T. Holman, “*5.1 Surround Sound: Up and Running*”, Oxford Focal Press (2000)
- [15] K. Hamasaki et al., “*5.1 and 22.2 Multichannel Sound Productions Using an Intergrated Surround Sound Panning System*”, AES 117th Convention, San Francisco, USA (2004)
- [16] F. Okano et al., “*Ultrahigh-definition Television System with 4000 ScanningLines*”, NAB 2004, Las Vegas, USA (2004)
- [17] T. Hatada et al., “*Psychophysical Analysis of The Sensation Of Reality Induced by a Visual Wide-Field Display*”, SMPTE Journal, Vol. 89, pp. 560-569 (1980)

Word Count: 4020 words.