



Audio Engineering Society Convention e-Brief 111

Presented at the 135th Convention
2013 October 17–20 New York, USA

This Engineering Brief was selected on the basis of a submitted synopsis. The author is solely responsible for its presentation, and the AES takes no responsibility for the contents. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Audio Engineering Society.

Subjective Comparison of Surround Microphone Recording Techniques Presented With and Without Video

Kruszielski, Luiz Fernando¹,

Globo TV Network, 22780-083 Rio de Janeiro, BR
Tokyo University of the Arts, 120-0044 Adachi Tokyo, JP
luiz.fk@tvglobo.com.br

ABSTRACT

The comparison of different setups of microphones for surround recordings of music is a topic that has a large interest in the audio community. Although it is known that image has a strong effect on the sound perception particularly in spatial aspects, very little research had been done aiming the surround sound recordings with accompanied video. To compare possible influences in this perception, a test was created using five different surround recording techniques that was done simultaneously at a Brazilian carnival parade. The subjects were presented to the sound with and without video and asked to rate four different aspects: localization, deepness, immersion and preference. The results show that there is a difference in perception depending on the presence or absence of video.

1. INTRODUCTION

Research concerning the differences between recording techniques has been a field that has received a lot of attention during recent decades. There are several studies that compare different aspects of surround sound through subjective evaluations of different arrays of microphones recorded simultaneously [1]. Most of the research in this field concerns the recording of music, more specifically classical music. The main objective of those studies, is to try to understand differences between different techniques, trying to re-create the original sound as close as it can be to the experience of being in a concert hall. However, our ears alone do not give the

experience of being in a concert hall. It is made by a merge of senses [2]. Some research shows significant differences in the perception of audio and overall impressions when said audio occurs with the presence of an image. It is also known that the visual sense is a dominant factor in the localization of sound in space, and both are complementary [3].

Through the years, the use of surround sound in cinema has become a well-established standard. The amount of immersion given by well-done surround sound in a movie is something that plays a great role in the audiences' experience. However, the amount of research concerning the techniques of recording surround for films is quite small, compared to recording of music itself.

The aim of this study is not only to compare recording setups between surround recording techniques for image, but is more to understand the possible effect that the visual aspect has under the preference of a certain sound.

2. METHOD

A test comparing five different microphone arrangements was conducted. The recording was done in the Rio de Janeiro Brazilian Carnival Parade. The reason that this event was selected as the sound sources was because it has a very rich variety of sound instruments being played together. The number of percussionist in each parade was about 250 people, and the instruments varied from surdo, agogo, shakers, repiques and tamborines. Since it is a parade, the sound sources are moving, and that was also considered to be interesting to the test. Also, the sound pressure generated by these instruments is quite high ($>100\text{dBA}$), overcoming possible differences in the recorders preamp noise headroom. Besides the percussion instruments, a PA system was used to reinforce the sound of vocal and harmonic instruments (guitar and cavaquinho). The PA sound was also playing during the recordings for the test.

The arrays were positioned in two different places: one at about 10 meters from the instruments passing by (position 1) and the other one about 2 meters from the instruments passing by (position 2). A GoPro 2 camera fixed and pointed to the same position as the microphones was used to create the visual material for the test (Fig.1).

2.1. Microphones

The microphone selection was based on two criteria: availability and also portability. Since the main idea of this test is to compare microphones for video, the microphones should be able to be easily moveable. Also, due to the space restrictions of the location, bigger arrays would be impossible to be mounted. The five tested microphones setups are:

- Zoom H2
- Zoom H2n
- Holophone H2 pro

- IRT Cross + MHK 60
- NHK Umbrella

Both the Zoom H2n and H2 are self-contained microphones and recorders. The H2 have a XY configuration microphone on the front (90°) and in the back (120°). The zoom H2n recorder has a MS configuration for the front microphones and a XY configuration in the back. In both recorders the capsule distance is less than 5cm.

The Holophone H2 Pro microphone is a device consisting of 6 capsules of microphones displaced in an elliptical shape to emulate the characteristics of the human head. The LFE channel from this microphone was not used in the test.

The IRT cross is an array created by the Schoeps Company. It consists of four Schoeps 64 cardioid microphones equally spaced (in this test at 25 cm) in a 90.0° angle between them. As an experimental design, the shotgun microphone Sennheiser MKH 60 was added for use in the center channel.

The NHK umbrella is an invention from the NHK laboratories, where using an umbrella wireframe five microphones (Sennheiser Me-2) were positioned in a similar position as the surround speaker position (one channel per microphone). The umbrella had an approximate distance between microphones of 60cm.

Each one of those three last microphone surround setups where connected to Fostex pd-606 recorders. All the microphones were positioned as close as possible to each other. The recordings where made in 48kHz and 24bit depth.

2.2. Test methodology

The test consisted of evaluating each microphone in four different aspects, rated from 1 to 10. The aspects and their corresponding questions are:

- Deepness: how distant are the sound sources perceived to be from each other. If the sound sources are perceived to be flat and in the same plane, the score should be lower. If they are perceived to be in different planes of distance, the score should be higher.

- Immersion: to what degree is there an impression of being surrounded and inside the scene
- Sound source localization: how clearly can the sound source be spatially defined, and how closely can it be pinpointed
- Preference: The overall preference rating for each microphone setup



Figure 1 Video frames from the position 1 (top) and position 2 (bottom)

The subject was presented to a 10 second segment loop of the Samba parade. Through a Digi 002 Command 8 controller, sound stimuli corresponding to a microphone set up could be switched during the loop playback. The subjects did not have previous knowledge or information about the microphones used. Each one of the four aspects was evaluated per time in a random order. First, the subjects evaluated the sound alone (sound mode), and then subjects did the same task accompanied by video (video mode). The audio mode was always presented first to avoid a memory effect caused by video in evaluating audio only stimuli. All the tasks resulted in 16 evaluations (4 aspects x 2 video/audio mode x 2 positions). The average time was 55 minutes per subject.

Nine subjects participated in the test. They are all experienced sound designers who work for the Globo Television Network. A room using ITU-R BS.775-2

alignment recommendations was used, and Genelec 8030 speakers were used at a distance of 1.8 meters from the subject. All the sound pressure levels of the stimuli were individually corrected to 79 dB_{L_{Aeq}}. The video display was a Sony TV Bravia, with 60 inches diagonal size.

3. RESULTS AND DISCUSSION

By doing an analysis of variance it was possible to see that all microphone setups had significant differences for every asked aspect ($p < .05$), except in regards to deepness. The difference between the position and the audio/video modes were also valid, except for deepness. This probably occurred because of the very dense sound environment, which may have caused the subjects to not have a clear understand of the task between microphones.

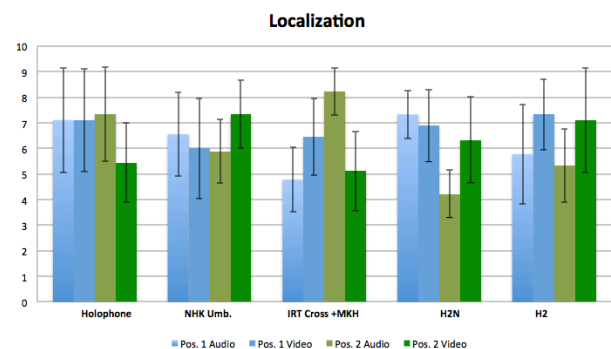


Figure 2 Localization: a higher value represents a result with more defined sound source and localization. The five microphone setups are spaced in groups of four columns corresponding to the recorded position and audio and video mode. The whiskers represent a 95% confidence interval

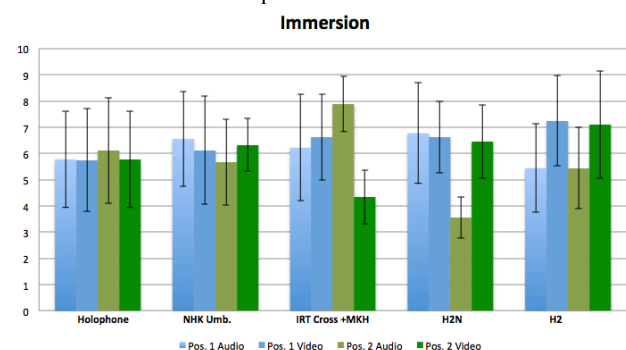


Figure 3 Immersion: a higher value represents a result more immersive.

The localization (Fig.2) and immersion (Fig.3) aspects had very similar response trends from the subjects. There was an interesting variation between the scene with and without video, where microphones that did not have such high scores for the audio, received better scores when presented together with video. There was a clear interaction between the microphones and the recording position. In position one the instruments are much more distant compared to position two. Those cues in audio are not that influential as they are in closer distances. The H2 and H2n microphones had big improvements in both localization and immersion for the video mode compared with the audio only for the second position, where the instruments were located closer. However only H2 had improved for the first image. It is argued that the visual reference cues probably overcame the lack of information from those microphones, making it much more suitable for that kind of situation. The NHK umbrella had an increase in the localization for the video mode in the second position, however this was not observed in the immersion. Also, for the second position, the IRT Cross + MKH 60 had the highest score for the localization and immersion in the audio only stimuli, however, the result was very different for the stimuli presented with video.

The overall preference varied significantly according to both the recording position and the absence or presence of video. This presence or absence of video had bigger influence in position 2.

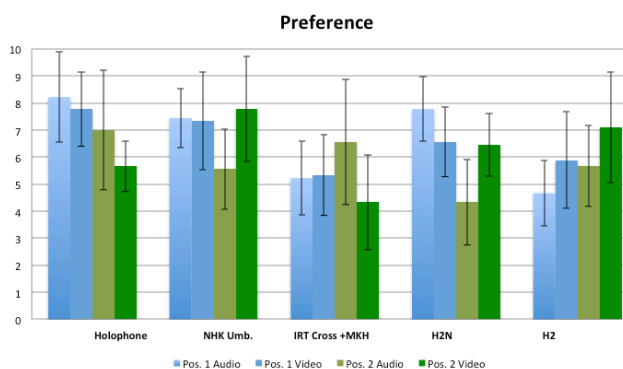


Figure 4 Preference: a higher value represents a result of greater preference.

The localization and immersion did not had direct relation with preference. A microphone could be very immersive and had a clear sound image, however still not be the preferable one, such as in the IRT cross +

MKH60. Although overall preference was also very subjective to the listener personal preference, a difference between video and audio could be observed for the same microphone. In a short interview after the test, some subjects stated that the presence of video had a strong influence in their perception, and audio that was “less defined” became more suitable for video. Also, some of the subject stated that, if they wanted only to listen to music, the Holophone would have the best results, however, when presented with image, the sound was considered “too clean”.

4. CONCLUSION

This study addresses a comparison of different microphones in two different positions, judged with or without the presence of video. These aspects had an influence in the listener responses almost as much as the microphone setup. It should be keep in mind that this test was conducted in a very “polluted” sound scene, with several elements that could interfere with the main result, such as the PA sound that was coming from various speakers on venue. This is very different from most of the surround comparisons were a very clean sound is recorded in a controlled environment, and subtle details are much more evident. Nevertheless, it was evident that there are perceivable differences in the perception of a recorded sound with or without the presence of video. Therefore, this effect should not be neglected, whether for sound experimentation or in creating sound for a movie. Further investigation in this area is needed to understand aspects that are caused by the visual influence.

5. REFERENCES

- [1] Kassier, R.; Lee, H.; Brookes, T.; Rumsey, F.: “An Informal Comparison Between Surround Microphones” In Proceeding of AES 118th International Convention, Barcelona, Spain 1999.
- [2] Jeon, J.Y.; Kim, Y.H.; Cabrera, D; Bassett, J.: “The effect of visual and auditory cues on seat preference in an opera theatre” Journal of the Acoustical Society of America, Vol.123, No.6, pp.4272–4282. 2008.
- [3] Begault, D.: “Auditory and non-auditory factors that potentially influence virtual acoustic imagery”. In Proceeding of AES 16th International Conference on Spatial Sound Reproduction. 1999.