

Examining observer response on environmental audio with ambisonic and stereo sound samples

Folkert de Vries
dr. Tjeerd C. Andringa

August 21, 2014

Abstract

The way humans experience soundscapes is important to understand sound annoyance. This research focuses on ambisonic sounds as opposed to stereo sounds. The foreground and background audio is also considered. The goal of this research is to discover the effect that the type of sound has on the degree of immersion people experience. The hypothesis states that ambisonic sounds are more related to real world sounds and therefore allow for easier immersion. We conducted an experiment in which people listened to ambisonic and stereo audio samples and commented on their feelings. Results show no significant result between ambisonic and stereo samples on the immersion.

1 Introduction

Empirical research hinted current acoustic researchers that ambisonic sound plays a role in observer response as opposed to stereo sound. [1] In current research the role of sound in quality of life has gained more attention. The current focus is on discovering ways to combat sound annoyance in new ways as lowering decibel doesn't always work as expected. It is apparent that some sounds below the current limit of loudness in decibel levels can also cause major annoyance on people. [2] This research aims at discovering differences in observer response to different types of sound. If a sample is presented in an ambisonic condition as opposed to a stereo condition, does it alter the observer response? Also, is the foreground or background information important? The presentation of these samples is conducted inside an ambisonic chamber. It is observed by Guastavino that certain classes of stimuli such as ambisonic sound trigger the brain into thinking it is part of the environment as opposed to being a passive observer. [1] [3] Is ambisound significant in this effect and also present in foreground or background sounds. Further research can use the results of this research in their focus on eliminating sound annoyance.

1.1 Previous research

Regarding the ambisonic sound condition this research is fairly new in its field but there has been a lot of research in soundproofing. Current solutions include

sound absorption or reflection using material science to design sound wave absorbing or reflecting surfaces and more advanced solutions like noise canceling using destructive interference. However the degree of noise reduction these solutions provide aren't always effective, especially at low frequencies. [4] Since people are still experiencing noise annoyance new methods to combat unwanted noises should be discovered. A good example are the unwanted sounds generated by wind turbines. Some people gain serious health issues as a result of constant exposure to the sounds generated by wind turbines even when these sounds are low in decibel levels. [5] If background noises intrude into the foreground part of the soundscape, does this influence observer opinion and cause annoyance? We introduce the ambisound topic into this problem. Guastavino found that observers respond differently on the same stimuli when it is provided in stereo or ambisonic. [1] This was not conducted under lab conditions but on the streets of Paris. Krijnders showed that subjects do not need vision to correctly annotate sounds. [6]

1.1.1 Core Affect

The Core Affect is a method to annotate the mood a person experiences. [7] It is defined as the combination of the current perceived state of emotional feeling and the allocation of resources to maintain or improve this feeling. [8] As seen in figure 1 it leads to four quadrants, each labeled as how humans regard the combination of the horizontal and vertical axis.

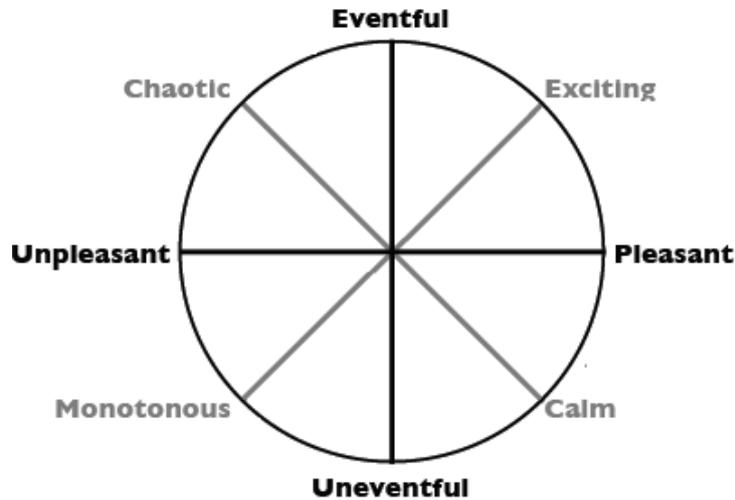


Figure 1: Core Affect figure. It consists of two diagonal axis on which activation and pleasure scale. This research uses this diagram to annotate test subjects emotional connection to the sound environment

This research will use the core affect to gauge test subjects' mood. Test subjects are to mark their feelings on a Core Affect diagram directly. It is more interesting for this research to see the degree of effect they mark as opposed to in which quadrant they place their response, as this may indicate a strong reaction on the sound effects. A stronger effect indicates a stronger emotional

connection and therefore a higher level of immersion.

1.2 Hypothesis

“Observer response is more immersed on an ambisonic as opposed to stereo audio stimuli of a soundscape. Additionally foreground sounds inflict a stronger response as opposed to background sounds in both sound type conditions.” To test this hypothesis subjects are exposed to sound samples in an ambisonic room.

1.3 Type of research

Since ambisound is a new topic in acoustic research this research will be exploratory. The hypothesis predicts a difference in response on ambisonic sounds as opposed to stereo. With the experiments we try to collect data that will support this claim. It is however not enough to use the results to reject or confirm the hypothesis. Furthermore the hypothesis is not strong enough to be used in a hypothesis testing research.

1.4 Uses

The results are relevant in the subject of sound annoyance and general acoustics. If subjects feel significantly more immersed in ambisonic conditions, it allows for more research. The difference in ambisonic and stereo audio environments can then be used in designing methods to counter annoying sounds. For example, people like to make detours through pleasant sound environments like parks. Urban planning can make major walking routes more accessible through parks and away from annoyance like busy traffic roads. Since sounds are always around us at any time, we can improve the quality of our lives by learning about them.

2 Method

The experiment determines if there is a difference between the groups of subjects regarding the type of sound they hear. It is based on their interpreted level of immersion. Furthermore the experiment collects data on the research method itself, as in how the performance of the experiment can improve for further research on the same or similar subjects. 120 male and female subjects with an average age between 18 and 25 are placed in groups of three in an ambisonic room on low, comfortable chairs called “Fatboys”. The room is dimmed and muted with acoustic dampening materials to be as noiseless as possible. There are no light sources. In the room are 8 “Yamaha HS5” monitor speakers in cube format (1 speaker in each corner of the room) and a “Yamaha HS8S” bass, all hidden from view. Audio stimuli are presented using a “Focusrite scarlett 18i20” USB 2.0 external sound card.

2.1 Experimental setup

The subjects are divided in seven groups, of which there are two control conditions. The groups each get a different pair of audio stimuli. These audio samples

are different front and background environments mixed over each other. The stimuli pairs differ in one factor in audio format. The formats are delivered in pairs so it is easier for the subjects to compare the stimuli with each other (see table 2). These pairs are the same sound environments but their audio type is mixed differently (see table 1). The subjects are given 4 trials between 90 seconds and two minutes. The control groups get pairs of the same condition, their trials don't differ in audio type. They hear the exact same sounds twice and thus should not report any differences. The subjects are primed at the beginning of the experiment by the researcher to focus on their "feelings" during the experiment. This makes it easier for them to report in the questionnaire.

2.2 Questionnaire

The questionnaire contains questions in which subjects need to comment on the audio and their feelings. Ideal results should give an indication that ambisonic sounds trick the subject into actually being at the location of the sound fragment, where stereo sounds make the subject an observer. The questionnaire contains 15 closed questions and a open question. The closed questions determine how much the subjects picked up from the audio and in what degree they feel connected with the events. They are divided into two categories: immersion questions and difference questions. The immersion questions focus on the feeling of the subject and allow for some grading of the level of immersion. The difference questions focus on the difference in the trial. They are aimed at the foreground and background audio differences. The open question ask to describe as much as they can about the audio samples where the subjects can comment freely on the audio connected with their feelings. The open question is aimed to improve this research method for future benefits. Lastly the subjects are asked to mark the location of the last sound environment on a core affect diagram (figure 1). This contains an emotional diagram where the subject marks his or her mood during the indicated sound fragment.

2.3 Audio resources

The samples test subjects hear are edited using Audacity¹ and MCtools². From a database of ambisonic sounds, a selection of environmental sounds and single source sounds were selected. These sounds are all in ambisonic B-format. This is a speaker independent format. The audio information is represented in a field of sound source directions, instead of speaker positions. The sound is then cut to an eventful period of circa 1 minute. Using `findecode` from the MCtools kit, the ambisonic format is decoded and down-sampled. It can then be formatted in a stereo configuration using Audacity. The original ambisonic sample is preserved. Next from the selection of sounds, suitable pairs are selected to supersample in order to create realistic sound environments with relatively distinct foreground and background sources. For example, a sea as background "noise" and a colony of seagulls that screech as foreground stimuli. These audio samples are superimposed and mixed in four different configurations (see table 1). The result can be played in the ambisonic room.

¹<http://audacity.sourceforge.net/>

²<http://people.bath.ac.uk/masrwd/mctools.html>

Name\ID	Foreground	Background
FABA	Ambisonic	Ambisonic
FABS	Ambisonic	Stereo
FSBA	Stereo	Ambisonic
FSBS	Stereo	Stereo

Table 1: Sound configuration. The foreground and background parts of the sound environment are mixed separately. In the rest of this paper we will refer to the configuration with the name as represented in column 1.

2.4 Statistic model

The experimental setup is quite complex. The research directive has an exploratory motive. The formal statistics are therefore limited. There is a lot of data available but not all of it can be used to test the hypothesis. However it is possible to conduct variances analysis on the groups.

	Stimulus 1		Stimulus 2	
Group	Foreground	Background	Foreground	Background
1	Ambisonic	Ambisonic	Ambisonic	Ambisonic
2	Stereo	Stereo	Stereo	Stereo
3	Ambisonic	Ambisonic	Ambisonic	Stereo
4	Ambisonic	Ambisonic	Stereo	Ambisonic
5	Stereo	Ambisonics	Stereo	Stereo
6	Ambisonic	Stereo	Stereo	Stereo
7	Ambisonic	Ambisonic	Stereo	Stereo

Table 2: Experimental setup. The subjects are divided in 7 groups. Each group gets a different configuration of the audio mixing in the different stimuli.

For this research the most interesting difference should occur in the FABA-FSBS condition (group 7 in table 2). This would not regard the fore and background as essentially important since the whole sound environment is different and test subjects should report the biggest difference in the questionnaire regarding the audio and their feelings. Other differences occur in the groups 3 to 6, where only one type changes. Groups 1 and 2 are control conditions.

3 Results

Following the questionnaire from the experiment we have gained multiple types of data. These were collected from male and female test subjects with an average age between 18 and 25.

3.1 Immersion Questions

The primary questions regarding the feeling of the subject based on the sound environment resulted in Likert scale data. Using a Manova type analysis of variance on the six questions by group the difference between the four configurations is not significant for any combination ($P > 0.05$), see table 3. However

Dependent Variable	Type III SS	df	Mean Square	F	Sig.
I felt connected to the sound environment.	9.810	3	3.270	1.597	.191
I felt I was at the location of the sound environment.	17.148	3	5.716	2.421	.067
The sound felt right, because it was predictable.	9.464	3	3.155	1.658	.177
I could determine the location of the sound sources.	7.124	3	2.375	1.450	.229
I could place myself between the sound sources.	7.224	3	2.408	1.138	.334
I was a participant and not an observer.	15.859	3	5.286	1.922	.127

Table 3: Results Manova. None of the variances are significant, however two results do indicate subjects might feel immersed differently.

the results did indicate that certain questions particularly interesting to our hypothesis do come close. For example the question “I felt I was at the location of the sound environment” has a significance level of 0.067 and “I was a participant, not an observer” has a significance level of 0.127, which do indicate that the subjects might feel a different level of immersion at different stimuli.

3.2 Core Affect

The test subjects we’re asked to mark their current mood on a core affect diagram, see figure 1 and appendix B. The question was for either the first fragments or the second fragments. The results are diverse. Most subjects experience eventfulness and some mark the environments as calm. These responses correspond with the sound fragments they heard. The differences seem to be slight between groups. The most prevailing is between the FABA-FABS (group 3) and FABA-FSBA (group 4) groups, see figure 2 and figure 3. These subject have heard the same initial condition but a different second condition. In the figures we see that the FABA-FABS group leans towards the pleasant end. Both the FABA and FABS subjects situate themselves to the right of the verticle axis. At the FABA-FSBA group the subjects place themselves towards the eventful end. They experience more chaotic than calm environments in comparison with the FABA-FABS group.

3.3 Difference Questions

The questionnaire contained three questions regarding the difference between the two samples in a trial. These questions were aimed towards the foreground and background differences. The results are slim, since there was not much focus on this topic. The results of a one-way Anova are not significant ($P > 0.05$). A close look indicates that the responses lean towards a more immersive response in stereo conditions as opposed to ambisonic. This is opposite of the hypothesis and against intuition. There is also no clear indication that foreground and

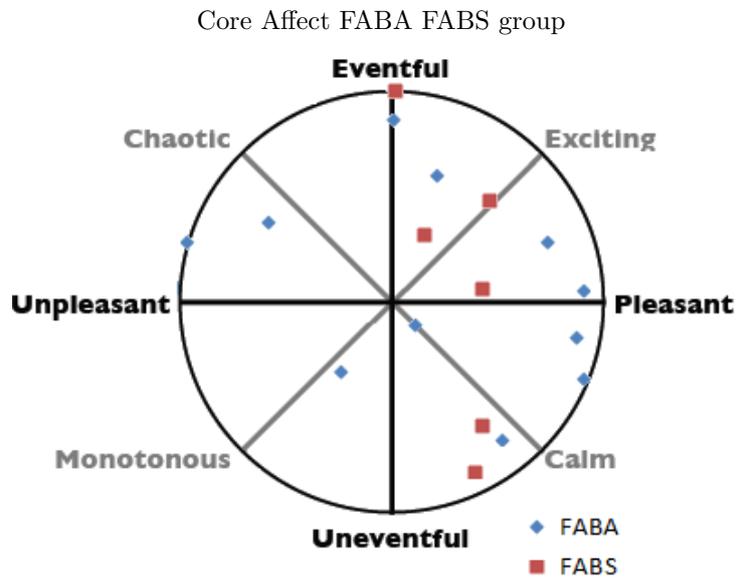


Figure 2: The core affect annotation from the FAB A FABS group. The test subjects listened to the same difference in sound type, but were asked to mark only one type. The FAB A condition are marked from the first sound fragments (blue), the FABS condition are marked from the second sound fragments (red).

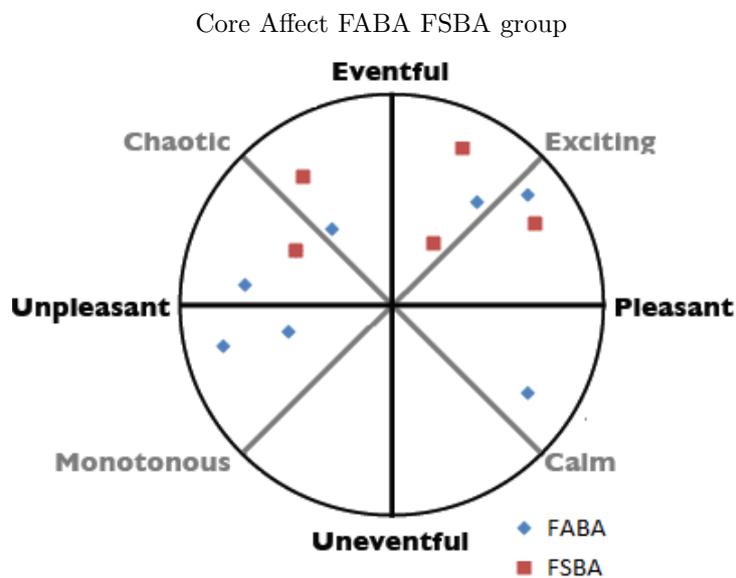


Figure 3: The core affect annotation from the FAB A FSBA group. The FAB A condition are marked from the first sound fragments (blue), the FSBA condition are marked from the second sound fragments (red).

background sounds are causing a difference.

3.4 Open Question

The results of the open question give a lot of insight for the setup of any further experiments. Most interesting for this research are responses where subject mention they feel somewhat physically located at the sound environment. In total 24 out of the 127 subjects mentioned this, however this is completely spread around all 7 testing groups.

Another recurring response are the differences between the two samples. Mostly the subjects indicate a difference in volume or even report the source as added (or removed) to the previous sample. This indicates poor normalizations of the two supersampled audio fragments.

Furthermore there are a lot of indications about recalling the sound environments. Most subjects summarize the whole experiment, but are unable to note all environments correctly. The answers are incomplete or incorrect. This indicates the experiment had a high load on the subjects' short term memory.

4 Discussion

The results indicate both in the closed and open questions that there is no significant difference between the ambisonic and the stereo condition. Moreover they tend to indicate that the stereo condition results in more immersion, which is the opposite of the hypothesis. Following intuition this shouldn't be the case. Here are some possible explanations:

- This experiment used a setup in an ambisonic room. Three subjects were placed together in a line towards the front of the room. The subjects on the side are located closer to one of the speakers used primarily in the stereo case. The only person to reliably interpret the sound environments the same is the one located in the middle of the room. When conducting an experiment similar to this one in an ambisonic room, one should be advised about positioning and consider using one subject at a time. The interpretation of the stereo stimuli from the single speaker can overpower the other speakers used, which is important since the ambisonic stimuli are spread over all speakers.
- Possibly the seats removed the ambisonic effect from the rear speakers as the headrest blocks audio. The subjects we're seated in low chairs with head high closed backrests. One subject (who also noted that he had some knowledge of audio engineering) mentioned the blocking of audio from the rear. Since the stereo case only uses two frontal speakers and the ambisonic case uses eight speakers from which four are behind the listener, this could influence the results, also amplifying the stereo cases.
- Considering that the subjects often mention difference between volume in sound samples, this may have influence the results. It indicates poor normalization of the volume levels during the superimposing of the sounds.

When converting from stereo to ambisonic format (and vice versa) the perceived volume changes significantly. The stereo condition has higher amplitude as it only uses two channels whereas the ambisonic uses eight. The poor correction may indicate these responses.

Besides the results indicating the hypothesis is not significant, there are some other interesting things the subjects note. Some participants report problems recalling the sound fragments. This indicates the trials were quite heavy on subject's short-term memory. It is true that the questionnaire asks for a lot of information. However this is not always a bad thing as it allows for the most pressing information to be reported. For this research it is not important to remember exactly what the sound fragments were, it is more interesting to see responses about feelings. It is however always difficult to report on emotions, feelings and immersion. The closed questions try to extract the most useful topics but one is often not exactly sure about his or her own arousal. Further research can consider using smaller trials for easier recall, or larger trials for a more overall measurement.

The results of the Core Affect question are too diverse to conclude on. Some subjects remarked that the question was vague and I can see why. The question was meant to be a summation of all the trials, but it can be interpreted as only the first (or second) trial. The results are therefore untrustworthy. The experiment was not focused on the Core Affect, but it can be a useful tool. In new research one could use the core affect the same way, but it is better to use it on an isolated trial to get reliable results.

Also some form of circular statistics can be used to analyze the results. This can result in discovery of a correlation between the sound type and the core affect quadrants. For the degree of effect on immersion, this was not relevant.

The audio used in the experiment should be classified by a different (small) set of subjects to classify the front and background elements in the sample. Since foreground sounds are defined by attention [9] they can differ between subjects. Instead of subjects we can also use a computational foreground-background separation model. Both will result in a less subjective foreground-background separation. In this experiment all foreground and background classification was done by the researcher while creating the audio environment.

Further research can also use new ambisonic recordings. The recordings used in this experiment are samples from earlier recordings and then superimposed. The experiment can be done again with real world recordings for (hopefully) more accurate results.

Another thing to take into account is the closed environment. This research assumes that audio is perceived the same without visual input, as Krijnders has shown in his research [6]. The subjects should be able to easily get immersed as if they are located in the real world. It is, however, possible the audio in the closed room is not enough true to nature to result in reliable results. For more accurate research, the validity of a closed room with regard to ambisonic sounds could use additional attention and confirmation.

In the experiment we encountered some practical issues that can be improved for further research. For instance, make sure the subjects can hear all speakers individually and with an equal volume. Furthermore, using an ambisonic room there is only one position in the center of the room where the sound environment is optimal. The position in the room and barriers between the ear and speakers should be considered carefully. It is likely the subjects in this research were exposed to bad ambisonic sounds.

In conclusion it seems that ambisonic sounds do not cause a more immersive experience than stereo sounds. The experiment had some flaws that can be corrected for more accurate results. Regarding sound annoyance ambisonic and stereo are still important. I think that pleasant sounds have a stronger effect on a persons moods if they are presented in an ambisonic environment while at the same time unpleasant sounds cause a stronger effect when they are stereo. This could be an interesting topic for further research.

References

- [1] Guastavino, C., Katz, B.F., Polack, J.D., Levitin, D.J., Dubois, D.: Ecological validity of soundscape reproduction. *Acta Acustica united with Acustica* **91**(2) (2005) 333–341
- [2] Andringa, T., Lanser, J.: Towards causality in sound annoyance. In: *Proceedings of the Internoise Conference*. (2011) 1–8
- [3] Andringa, Tjeerd C, v.d.B.K.A., Vlaskamp, C.: Learning autonomy in two or three steps: linking open-ended development, authority, and agency to motivation. (2013)
- [4] Leventhall, H., et al.: Low frequency noise and annoyance. *Noise and Health* **6**(23) (2004) 59
- [5] Andringa, T.C., van den Bosch, K.A.: Core affect and soundscape assessment: fore-and background soundscape design for quality of life
- [6] Krijnders, J.D., Andringa, T.C.: Differences between annotating a soundscape live and annotating behind a screen. In: *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*. Volume 2010., Institute of Noise Control Engineering (2010) 6125–6130
- [7] Andringa, T.C.: Soundscape and core affect regulation. In: *proceedings of Interspeech*. (2010)
- [8] Russell, J.A.: Core affect and the psychological construction of emotion. *Psychological review* **110**(1) (2003) 145
- [9] Schafer, R.M.: *The soundscape: Our sonic environment and the tuning of the world*. Inner Traditions/Bear & Co (1993)

Appendices

A Questionnaire

A.1 Closed immersion questions for the first fragments

- I felt connected to the sound environment.
- I felt I was at the location of the sound environment.
- The sound felt right, because it was predictable.
- I could determine the location of the sound sources.
- I could place myself between the sound sources.
- I was a participant and not an observer.

A.2 Closed immersion questions for the second fragments

- I felt connected to the sound environment.
- I felt I was at the location of the sound environment.
- The sound felt right, because it was predictable.
- I could determine the location of the sound sources.
- I could place myself between the sound sources.
- I was a participant and not an observer.

A.3 Closed difference questions

- The difference between fragments was clear.
- I felt differently at the second sample compared to the first sample.
- The second sample evokes different feelings than the first sample.

A.4 Core Affect

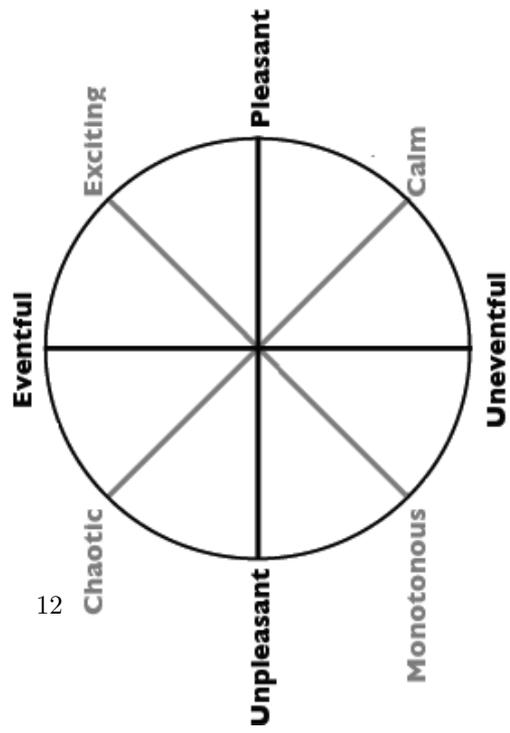
See appendix B. Test subjects received one of the two possible questions.

A.5 Open Question

- Describe the sound environments with a minimum of 50 words. Think about the sound sources, the background, your reaction in the studio and how your reaction would be in the real world.

B Core Affect

Mark the location where you think the first time you heard the fragment is located.



12

Mark the location where you think the second time you heard the fragment is located.

