Perceptual Evaluation of the Diffuseness of Synthetic Late Reverberation Created by Wave Field Synthesis at Different Listening Positions

Jens Ahrens

Quality and Usability Lab, Technische Universität Berlin, Email: `jens.ahrens@tu-berlin.de`

Introduction

We revisit and extend a study presented in [1] that evaluated a method for the creation of artificial diffuse reverberation using Wave Field Synthesis (WFS). The experiment investigated the minimum number of uncorrelated sound field components that are required so that a rotation of the late reverberation tail is not audible.

In the present context, we employ the following simple definition of diffusion/diffusity[1]: *In a diffuse sound field, there is equal probability of energy flow in all directions* [2]. The perceptual equivalent of *diffuseness* is not clearly defined. A number of studies are available that investigate the creation of perceptually diffuse sound fields by means of loudspeakers. Typically, the number of uncorrelated sound sources required to produce a diffuse sound field is determined, which is in the order of 10. Examples are [3, 4, 5]. Diffuseness is detected via auditory localization attributes, e.g. low *locatedness* or *image focus* or, equivalently, significant *blur*.

The situation that we are interested in differs from the one considered in [3, 4, 5] in that the diffuse signal is preceded by a non-diffuse one (the direct sound and the early room reflections). Perceptually, this constitutes a fundamental difference as it is the first arriving signal components that dominate localization. The later signal components contribute primarily to the spatial impression. This circumstance is summarized under the term *precedence effect* [6]. Most results (apart from, e.g., [5]) are only available for a static listener and when all uncorrelated signals impinge synchronously.

The proposition in [1] is to synthesize diffuse late reverberation via an appropriate number of equiangularly spaced virtual plane waves that carry uncorrelated signals as illustrated in Fig. 1(a). It was tested how many plane wave components are necessary so that a rotation of the diffuse reverberation tail by half the angle between adjacent plane waves is not audible. Direct sound and early reflections were kept constant. It is reported that significant systematic coloration was observed for some of the stimuli. Comb filtering was applied to all stimuli to remove timbre as a cue and subjects were instructed to only respond to spatial attributes. For the 8 tested subjects a number between 3 and 10 plane wave components was required in order that the subjects did not perceive a change in the spatial attributes when the diffuse tail was rotated. A likely explanation of the coloration is the fact that a rectangular loudspeaker array was used, which exhibits frequency-dependent corner effects. The loudspeaker directivity might also have played a role.

The most significant restriction for application of the results to practical scenarios is the circumstance that [1] tested exclusively the central listening position in the loudspeaker array where all plane wave components arrive aligned in time. In the present contribution, we replicate the experiment from [1] and additionally test two off-center positions.

Test Paradigm

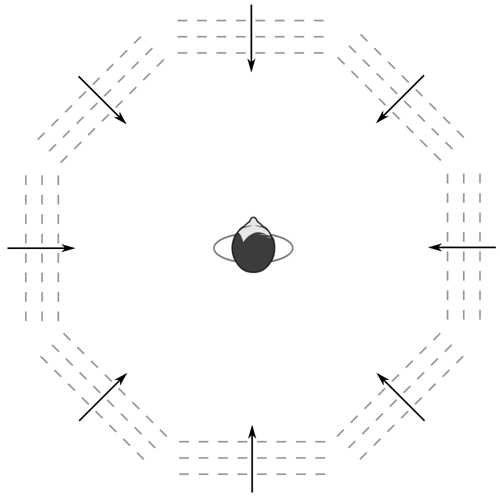
We apply the assumption from [1] that a spatial rotation of a diffuse sound field is not audible. Each of the three listening positions was tested in a separate session. A three-alternative forced choice paradigm with 2-up-1-down adaptation was used. Two of the stimuli in a given triad were always identical whereas the late reverberation of the third stimulus was rotated spatially. Stimulus assignment was randomized. The subjects' task was to identify the stimulus that differs from the other two and report accordingly. If no stimulus appeared to differ then the subjects were asked to report a random response. Each session started with a training phase of seven triads for the subjects to accustom to the interface and the stimulus differences. The actual experiment started with one single virtual plane wave presenting the late reverberation. The number of plane waves was increased after two successive correct responses from the subject for a given number of plane waves. Any incorrect response resulted in a reduction of the number of plane waves. We assumed that a threshold exists above which spatial rotation of the late reverberation is not detectable and that the adaptation converges to that threshold. The experiment ended after eight reversals of the response tendency (up vs. down).

Pilot studies suggested that increasing the number of plane waves in steps of one might lead to an unacceptably long experiment duration for the two off-center listening positions as the expected threshold seemed to be significantly higher than for the center position. The number of plane wave was therefore increased in steps of 1 for the center position, in steps of 2 for lateral position 1, and in steps of 3 for lateral position 3 (cf. Fig. 1(b)).

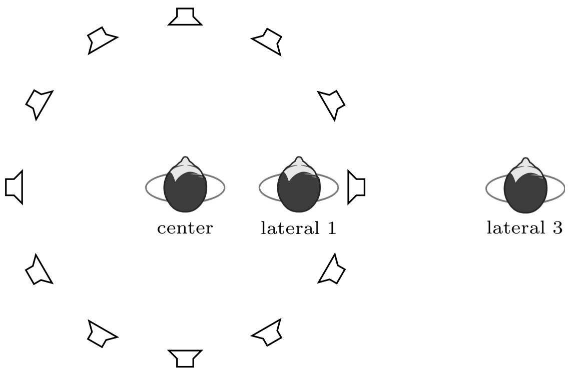
Stimulus Creation

The experiment was conducted using dynamic (head-tracked) binaural simulation of a real loudspeaker array of which binaural room impulse responses (BRIRs) were measured in order to achieve controllable and repeatable

[1]In this paper, we differentiate between the physical property of *diffusion/diffusity* and the perceptual attribute *diffuseness*.



(a) An example set of 8 equiangularly spaced plane waves for the creation of the late reverb tail



(b) The tested listening positions; *lateral position 1* is 1 m off-center, *lateral position 3* is 3 m off-center

Figure 1: Schematics of the geometrical setup

conditions. The SoundScape Renderer was used for the realtime convolution of a dry castanet signal with a duration of 2.6 s with the given impulse responses. The castanet signal was looped continuously and the subjects we able to switch between the stimuli of a given triad seamlessly. A Polhemus Fastrack unit was used for low-latency head tracking. Audio examples can be obtained from [7].

All stimuli were composed of direct sound, early reflections, and late reverberation. The direct sound was implemented using un-manipulated HRTFs. The early reflections were also realized using HRTFs whereby lowpass filtering was applied to mimic absorption at the room boundaries. The mixing time was set to 50 ms; RT60 was 1.6 s. The late reverberation was created as a simulation of a 56-channel circular loudspeaker array with a radius of 1.5 m of which binaural impulse responses were available for two different listening positions (see below) and different head orientations. Care was taken to assure that the time alignment of the individual components is correct. Contrary to [1], we did not detect systematic coloration in the stimuli and did therefore not apply comb filtering. We used the Sound Field Synthesis Toolbox [8] for the stimulus creation.

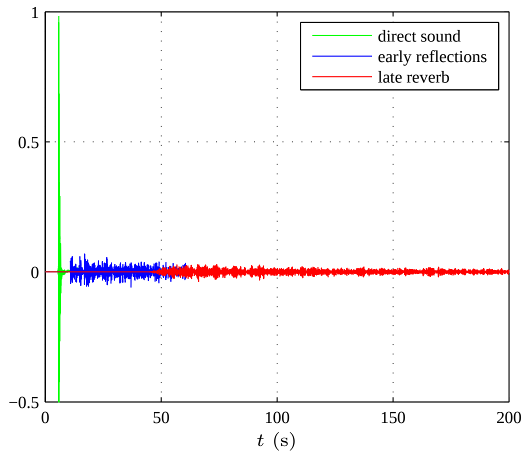


Figure 2: Sample binaural room-related impulse response

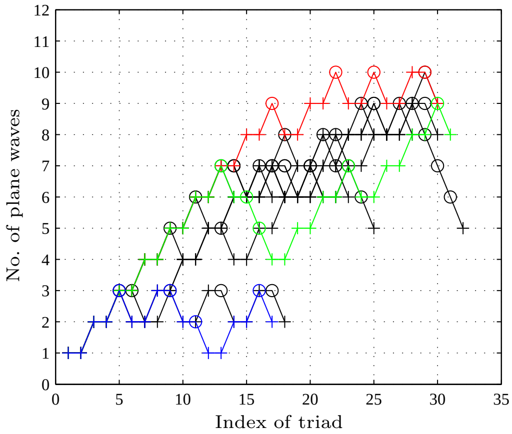
Direct sound and early reflections were identical for all stimulus conditions. The late reverberation was created from uncorrelated white noises with appropriate fade-in and frequency dependent decay. All components were created for different head orientations with a resolution of 1°. The *root mean square* of the late reverberation was constant for all stimulus conditions in the center position. The lateral positions employed the same weighting like the center position. The late reverberation was presented using virtual plane waves impinging from equiangularly distributed directions as illustrated in Fig. 1(a). The arrival directions are distributed such that there is always one plane wave impinging directly from straight ahead. Each stimulus was created additionally with all plane waves rotated by half the angle between adjacent plane waves.

Fig. 1(b) illustrates the listening positions that were tested. Lateral position 3 was created by using the BRIRs of lateral position 1 and modifying the timing of the virtual planes waves. No modifications of the amplitudes were applied based on the assumption that the differences between lateral position 1 and lateral position 3 are negligible in a very large system.

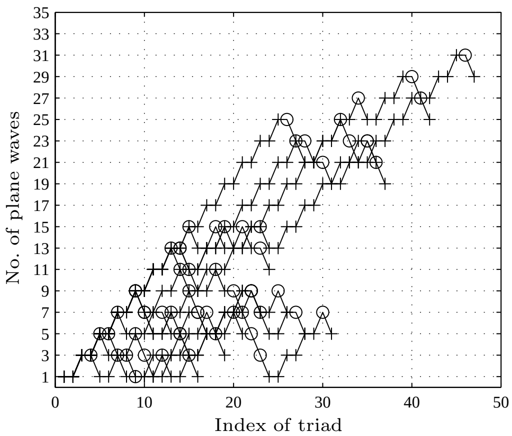
Results

Each listening position was evaluated in a separate session by 12 subjects each. 4 subjects participated in more than 1 session. The average time spent on a triad was 23 s with significant variance. No session lasted longer than 20 min.

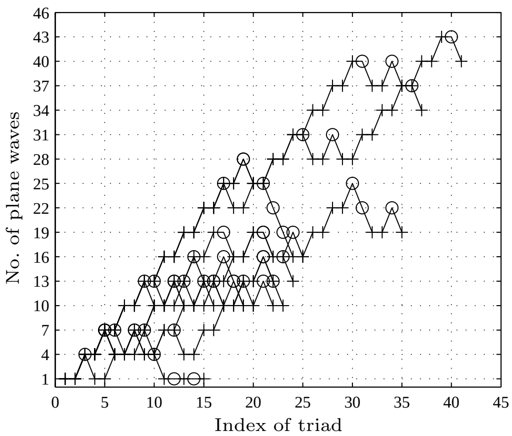
The results are depicted in Fig. 3. We found significant individual differences in the ability to identify rotation of the diffuse reverberation tail. Three individual results are highlighted in color in Fig. 3(a) in order to illustrate the range of performance. The situation is very similar for the other positions (Fig. 3(b) and (c)). It was rare that an actual convergence of the adaptation around a threshold occurred. The red curve in Fig. 3(a) is one of the exceptions. We comment on this aspect in the Conclusions.



(a) Center position; three individual results are highlighted in color to illustrate the range of performance



(b) Lateral position 1



(c) Lateral position 3

Figure 3: Results for the three different listening positions; crosses indicate a correct response; circles indicate an incorrect response

The results presented in [1] were replicated closely (cf. Fig. 3(a)): The required number of uncorrelated plane wave components for the center position varies dependent on the subject between very few and around 10. These numbers are significantly higher for the lateral listening positions whereby the farther off-center position led to higher numbers of plane waves. Note that the highest threshold apparent in Fig. 3(c) is almost as high as the number of loudspeakers of the system (56).

The following attributes where reported by the subjects as having been used for discriminating the rotated stimuli from the non-rotated ones: 1) spatial impression, 2) timbre differences, and 3) smearing of transients.

Conclusions

The most important conclusions that we draw from the presented experiment are actually not directly related to the initial question of what number of uncorrelated plane wave components is required for achieving perceptual rotation invariance. Although we can measure a threshold for each listening position, informal listening suggests that the resulting reverberation sounds different for the different listening positions. The property of perceptual rotation invariance is obviously not sufficient for proving diffuseness. Note that the statistics of a diffuse sound field are also translation invariant by definition.

Measuring diffuseness (recall footnote 1) seems an unsolved task. A major challenge is the fact that not all spatial attributes of an auditory event are fully independent of other attributes such as timbre. Comments by the subjects suggested that the most critical subjects used timbre in the most critical triads as means of differentiation of the rotated stimuli. Similar observations are reported in [9] and in other locations. It is unclear at this point what property of the ear signals causes the observed changes in perceived timbre in the present case nor is it clear whether all subjects actually perceived similar timbre changes in the according situations.

Interestingly, the subjects reported that the difficulty of the task did not increase monotonously with the number of plane waves. It is evident especially in Fig. 3(b) that more incorrect responses were given for certain numbers of plane waves than for others (e.g., 15 plane waves vs. 17). It is therefore not surprising that convergence of the adaptation did not occur. We do not have an explanation for this observation.

The experiment design allowed for the identification of even the slightest differences between a rotated and a non-rotated stimulus. It seems that this design is overcritical for the present research question.

References

[1] J.-J. Sonke. *Variable acoustics by wave field synthesis.* PhD thesis, Delft University of Technology, 2000.

[2] F. Jacobsen. The diffuse sound field – Statistical considerations concerning the reverberant field in the steady state. Report No. 27, 1979.

[3] K. Hiyama, S. Komiyama, and K. Hamasaki. The minimum number of loudspeakers and its arrangement for reproducing the spatial impression of diffuse sound field. In *113th Convention of the AES*, Los Angeles, CA, USA, October 2002.

[4] O. Santala and V. Pulkki. Directional perception of distributed sound sources. *J. Acoust. Soc. Am.*, 129(3):1522–1530, 2011.

[5] F. Melchior, C. Sladeczek, D. de Vries, and B. Fröhlich. User-dependent optimization of wave field synthesis reproduction for directive sound fields. In *124th Convention of the AES*, Amsterdam, The Netherlands, May 2008.

[6] J. Blauert. *Spatial Hearing*. The MIT Press, 1997.

[7] J. Ahrens. Audio examples. `http://www.soundfieldsynthesis.org/audio_examples/DAGA2015`, 2015. last accessed: Oct. 31, 2015.

[8] H. Wierstorf. The sound field synthesis toolbox. `https://github.com/sfstoolbox/sfs`, 2012. last accessed: Oct. 16, 2014, commit 46962f7.

[9] S. Klockgether and S. van de Par. Limits for the perception of directional dependence of the reverberant tail in binaural room impulse responses (BRIR). In *DAGA*, Darmstadt, Germany, 2012.