Multichannel source directivity recording in an anechoic chamber and in a studio

Roland Jacques, Bernhard Albrecht, Hans-Peter Schade
Dept. of Audiovisual Technology, Faculty of Electrical Engineering and Information, Technische Universitaet Ilmenau, 98693 Ilmenau, Max-Planck-Ring 14, Germany, roland.jacques@gmx.de, bernhard.albrecht@tu-ilmenau.de, schade@tu-ilmenau.de

Diemer de Vries
Lab. of Acoustical Imaging and Sound Control, Faculty of Applied Sciences, Delft University of Technology, P.O.B.5046, 2600 GA Delft, The Netherlands, diemer@akst.tn.tudelft.nl

Frank Melchior
Fraunhofer Institute for Digital Media Technology (IDMT), 98693 Ilmenau, Langewiesener Straße 22, Germany, mor@idmt.fraunhofer.de

Within the framework of the cooperation between TU Ilmenau, the Fraunhofer Institute for Digital Media Technology (IDMT) and TU Delft, new sound recording concepts are developed dedicated to sound reproduction by Wave Field Synthesis (WFS). It is intended to also take the directivity characteristics of musical instruments into account in this process, instead of representing these instruments as omnidirectional point sources or plane wave generators, as usual in WFS practice until now. As an introductory experiment, the frequency-dependent directivity patterns of five brass instruments, together forming a quintet, have been measured in an anechoic chamber and in a studio by means of a circularly arranged microphone setup. Characteristics have been recorded of the individual instruments as well as of the ensemble. The data give ample information about the influence of the acoustic environment on the recorded radiation properties of the different instruments, as a function of frequency and radiation angle. Besides, the data give insight in how far the radiation pattern of a group of instruments can be resynthesized from the characteristics of the individual instruments in the two researched environments. In the paper it will also be considered how the results can be implemented in WFS music reproduction.

1 Background

Wave Field Synthesis (WFS) aims at reproducing sound sources and their acoustical environment in a fully natural way, in the temporal as well as in the spatial domain. This means that not only the localization of musical instruments should be correct for all listeners, but that also their directivity patterns must be correctly reproduced: a trumpet heard from a frontal direction should be perceived louder and brighter than when heard under a lateral angle. When music is reproduced in its acoustical environment the directivity patterns of the instruments have to be taken into account: for a trumpet oriented forward into a hall, its signal should not, or only weakly, be convolved with the lateral reflections of the room impulse response. Until now, instruments and voices are represented as omnidirectional point sources in WFS reproduction, such that the above effects are discarded. This gives rise to unnatural spatial balance in the reproduced sound, especially in 'dry' acoustical environments. Under reverberant (diffuse) conditions the effects are less critical, since even the sound of strongly directional sources reaches the listener at a certain distance from many directions after some time. Therefore, it is interesting to investigate how the directional distribution of sound from musical instruments and its multichannel synthesis is influenced by the acoustical environment.

2 Conception of the recordings

The measurements of this paper were taken from recordings of a professional brass quintet. The recordings were made in the context of a diploma thesis ([1]) on which this paper is based. The first recording took place in an anechoic chamber at PTB Braunschweig. The second one was made in a large radio studio at NDR Hannover (see Fig. 1), built in 1952 (3000 m³, T₆₀ = 1.2 s, C₄₀ = 3.5 dB), which has an interesting interior design creating bright, diffuse reverberation.
On both locations the whole quintet was recorded with a setup of 15 microphones (see Fig. 2).

Subsequently each instrument, playing the same music, was recorded separately, using a circular microphone arrangement with 22.5° of angular resolution (omitting the position at 112.5° due to microphone availability). The microphones used were aligned Sennheiser MKH800s, set to figure-of-eight characteristics in order to attenuate the other instruments and floor/ceiling reflections (see Fig. 3 and Fig. 4).

3 Playback methods

In the WFS context, there are various playback methods which can take advantage of the recordings and measurements presented in this paper. One general approach is to calculate the long- or short-term average spectra of every direction, as it was done for the plots in Fig. 8 to Fig. 11. These spectra, which can also be found in literature such as [2], can be used to generate appropriate directivity filters for the current position and angle of the virtual sound source, only one source signal of the actual performance is required then. (see Fig. 5)

The measured spectra can also be transferred into room simulation software or into an MPEG-4 scene description to set the source directivity parameters. Another application is monopole synthesis which approximates the original sound field of a source, including its frequency-dependent directivity, by superposition of a group of typically 3 to 5 monopoles ([3]). Position and amplitude of these monopoles is optimised so that the measured real directivity is approximated. This method can easily be used ([4]) in WFS, where creation of monopole sound fields is state-of-the-art technology. Ideally the result is the same as with directivity filters.
However, it can be disadvantageous to use a static, averaged directivity pattern. Most musical instruments exhibit significantly varying directivity characteristics, depending on the current fundamental, playing techniques etc. Severe artefacts can be introduced if expressive movements of the player occur, because the real and the synthesized directivity pattern will incorrectly be added. Another approach, introduced in [5], attempts to overcome these issues by means of multichannel recordings, see Fig. 6.

**Fig. 6 Multichannel directivity recording**

All angles of interest are recorded during the performance, as described in this paper. Each channel will be discretely reproduced during playback, which requires coincident directional sources. These sources must not be confused with the musical instrument itself: they do not represent this instrument directly, they rather create a multichannel image of the instrument and its directivity. Like loudspeakers they should be frequency-independent, so that this image is not impaired. Such sources do not exist in reality, they have to be simulated or synthesized as virtual sources. They can easily be defined in room simulation programs or MPEG-4 scenes, for WFS playback they are currently being developed ([1]). This way every part of the large WFS listening area could be fed with the correct, time-variant signal characteristics, resulting in a highly realistic listening experience for every member of the audience.

The spatial resolution of this second approach depends on the number of available microphones and transmission channels, and sometimes this approach is not applicable for practical reasons. Listening tests have shown that a few strategically placed microphones can yield good results. However, this depends on whether subjective or physical realism is required, since human listening experience regarding details of the source directivity is in most cases rather limited. Sometimes, the mere existence of a change in sound colour while moving around the source can be convincing, while an absolutely lifelike reproduction will not necessarily be perceived as such. This implies that the temporal resolution of the reproduced directivity is just as important as the spatial resolution. A reasonable balance of both can be achieved by multichannel reproduction. For every recording situation these issues have to be considered anew to determine the most efficient proceeding.

Another issue to be considered is the shape of the frequency-independent virtual directional sources (which act as “loudspeakers”, creating a multichannel image of the musical instrument). They should be unidirectional and they should provide some kind of crossfading between the microphone channels so that discontinuities in the directivity pattern are avoided. A suitable solution is the usage of ideal higher order cardioid sources, defined by \((1 + \cos(\phi))^n\). The order \(n\) needs to be chosen so that the crossfading is continuous, but not too wide. Otherwise the spatial resolution is compromised, and also comb filtering artefacts can occur. This type of sources has been used for the investigations presented in this paper.

### 4 Measurements

The following measurements are based on a forte chord of the quintet (noted in Fig. 7) which was FFT-analysed in octave bands of 125, 250, 500, 1000, 2000, 4000, 8000 Hz respectively.

**Fig. 7: Forte chord, ensemble**

Fig. 8 and Fig. 9 show normalised polar plots in dB of the C-trumpet and the French horn from the anechoic recordings. Quite clearly the radiation is concentrated on the bell axes, with the strongest concentration at high frequencies. The acoustic shadowing from the player’s body is also visible.

**Fig. 8: Trumpet, anechoic recording**
Several conclusions for recording practice can be drawn from the above results. When spectral differences between different directions are required in order to derive directivity filters, the use of an anechoic chamber for such measurements is highly recommended. But when recording the actual performance with multiple channels, which will be discretely reproduced later, the room influence is much less problematic, regular recording studios can be used then. This adds to the earlier discussed potential benefits of balanced temporal/spatial resolution of the multichannel approach. However, if several musical instruments, which were recorded separately, will be played simultaneously, then the presence of reverberation will create an unnatural impression of several isolated rooms. In this case anechoic recordings are highly recommended.

Another possibility to analyse the two different recording setups (see chapter 2) is the visualization of the frequency dependent spatial SPL distribution from the ensemble. The visualizations are made in two different ways. The first (“75ch”) is resynthesized from the single instrument recordings with 15 microphones each, which are reproduced from the instrument’s original position in the ensemble and with musically indicated level. The application of the single recordings results in the combination of 75 microphone channels to resynthesize the ensemble completely. The 75ch-version is used as a high-resolution and high-instrument-separation reference here (see Fig. 12).
The second version ("15ch") is resynthesized from the ensemble recording with a total number of 15 irregularly placed microphones (see Fig. 2), which are also reproduced from the corresponding instrument’s position (see Fig. 13).

Now we are able to analyze the directivity synthesis errors, which are caused by the lower resolution (15 instead of 75 channels) and by the instrument crosstalk (ensemble vs. separate recording).

When comparing Fig. 14 and Fig. 15 (250 Hz band, anechoic) it can be seen that low frequencies were radiated rather omnidirectionally in 75ch. The 15ch image is much less homogenous, but it must be kept in mind that the used strongly directional reproduction of each microphone is not quite viable at these frequencies, so this is not expected to be a problem in real implementations. In the 15ch visualizations the black sections between the instruments are caused by the missing lateral microphones.

In Fig. 16 (2 kHz band, anechoic) the directional radiation of the trumpets and the French horn are clearly visible. The corresponding 15ch image (Fig. 17) shows the same characteristics, but less sharp and clear. Errors occur at the sides, where due to the instrument crosstalk the SPL is too high.

In the 8 kHz band only the left trumpet and the French horn contribute significant sound energy, see Fig. 18 (75ch, anechoic). In the 15ch image (Fig. 19) it is interesting to see a beam coming from the tuba’s position. Since the tuba has no energy at these frequencies, this is due to crosstalk from the French horn. Another beam comes from the side of the right trumpet, this time it is caused by crosstalk from the left trumpet, because both trumpets lie on the axis of the right side microphone (compare Fig. 2).

Apparently, the 15ch version is capable of reproducing the basic directivity characteristics of the quintet. Now the question arises how the lower spatial resolution and the crosstalk are perceived. It could be possible that the crosstalk doesn’t compromise, but even enhance the sonic image of the instruments, because it could introduce stereophonic imaging at some sections. These questions are subject of the next chapter’s listening test.
5 Listening test results

The listening test had 28 participants, including expert and non-expert listeners. A 32-channel WFS setup was used in the listening lab of TU Ilmenau, where the same source layout was synthesized as in the visualizations above. Only the 75ch version was reduced to 40 channels for complexity reasons. The test design was double-blind with hidden reference, the participants controlled the playback and entered the ratings with a listening test software of Fraunhofer IDMT.

The first part of the test featured a single instrument, represented by 15, 8 and 4 channels. For each variant a short sequence was repeated 4 times, while the instrument was turned in 30° steps. The ratings show the tendency that the perceived naturalness of the sound colour is proportional to the resolution. But the participants were almost unable to judge how far the instrument had turned. A clear change of the tonal properties was heard, but most listeners were unsure how to interpret it. Some subjects found the lower resolution more convincing, because the tonal changes were smoother than at higher resolutions. This result was most probably a consequence of the limited listening experience for such details and to the various expectations about how the trumpet should sound at the side.

The next part presented the synthesized quintet. Due to the complexity of the issue and the experimental stage of the WFS implementation, not all of the following results are significant, although tendencies can be found in most cases. The main parameter to be judged was “plasticity”, meaning how real and three-dimensional the quintet sounds. The participants were allowed to move freely inside the listening room, so they could judge the sound from different perspectives. On average, the 15ch version was found to sound slightly more plastic and real than the 75ch variant.

The most important conclusion from this is not that one or the other is better, but that the 15ch version is clearly not worse, despite the considerably lower production effort.

A quite different judgement was given on the question of naturalness and authenticity of the sound colour. In some cases the 15ch version was rated significantly worse here. This may be due to the lower resolution, which requires a larger overlapping of the virtual sources. But more important is the reason that the musical balance of the instruments was not optimal. This was the case because for the sake of reproducibility the level of the microphone channels was left unchanged instead of being optimised by ear, which would certainly be done for a musical production. Also the microphone arrangement could be optimised from a musical point of view, but for these recordings the height and distance of the mics were determined in advance for scientific reasons.

The general conclusion from this listening test is that the 15ch version can deliver a realistically sounding image of the quintet. The drawbacks regarding sound colour can most probably be avoided by optimisation of the musical balance. However, all presented results are based on this first experimental case, and should be considered just as hints for the potential of multichannel directivity reproduction.

Many issues which were addressed in this paper deserve further examination. Most importantly, the playback implementation of directional frequency-independent sources needs further development, whether in WFS or any other playback context. Recordings of expressly moving instruments or extended sources like a grand piano are necessary to determine what spatial resolution is needed to realistically recreate these properties. Finally, new applications like interactive, object-oriented auditive scenes can take advantage of isolated source recordings with lifelike directivity characteristics.

References