3D Audio Sound Installation "Linie U54"

ROLAND ERNST, JENS KOTHE, ANDREAS KRAFT, ARNE MORGNER, LUKAS MÜNTER, TIM PHILIPP *and* **CHRISTIAN TOBIAS**
tp057@hdm-stuttgart.de

Master's Project Department of Audio, Study Programme Audiovisual Media, Faculty Electronic Media, Stuttgart Media University (Hochschule der Medien), Nobelstr. 10, 70569 Stuttgart, Germany

This paper contains the concept and development of the 3D sound installation "Linie U54". It examines the recording process, the sound design, and the installation. Furthermore, the mixing process and the interaction between the 3D setup and the headphones are reviewed, as well as the experiences and perceptions of the visitors.

0 CONCEPT

If you look around in the train, most people are listening to music with headphones while starring at their phone screens. They wish to encapsulate themselves from reality and calm down through music. But how can you keep calm if the screen shows new messages and calls every minute? One possibility is to focus on your music through your headphones. The atmosphere and the music gives freedom to the thoughts and the option to settle.

The goal of the installation "Linie U54" is to establish a train atmosphere and extend the imagination of the listener. The motivation is to push the boundaries of acoustic sensations by walking the line between both sound and music, as well as internal and external perceived sensitivities. The music is introduced by the headphones and leads through the experience. Lo-Fi music is chosen because it relaxes most people as there is no way to reflect everyone's preferences. The headphones losing battery life is the only definite interruption of the otherwise freely controllable listening experience during the headphone loudspeaker interaction. Users can create their own experiences by putting the headphones on and off, which results in different impressions of the 3D setup. Sound elements of the train atmosphere are used to develop the music and to externalize it. Therefore, the communication between headphones and 3D setup opens new possibilities of envelopment. Modifying room-size and changing the play-out between headphones and 3D setup are increasing the immersion of the sound. With the end of the headphone experience comes the breakdown part before the highlight hits. This opens the possibility to hear the dynamic range and the immersion of the 3D setup without headphones before the reality returns.

All in all, the listener goes through an extraordinary train ride in "Linie U54" and witnesses the unique sounds of Stuttgart's trams and their evolution through fantasy. What starts with headphones and lo-fi-music to suppress the atmosphere, ends without headphones and a whole orchestra filling the underground.

1 METHODOLOGY

1.1 SETUP

The installation was assembled in the recording room of Stuttgart Media University's own sound studio. Using an acoustically well treated room as a basis for the installation was very beneficial, as frequently seen problems like room modes or echoes could for the most part be avoided from the beginning. This also meant that the loudspeaker's individual frequency responses were very homogenous despite their various placements in the room. This was one of the most important requirements to the system as sound events should sound the same, rega,ess of their positioning within the system. Varying frequency responses among the speakers would also produce noticeable coloration when sound objects are panned continuously between speakers. To further ensure this requirement being met, the system was measured and differences in frequency response were accounted for with equalization filters.

db Audio's E8 passive coaxial loudspeakers were chosen as full range loudspeakers as well as four E12 subwoofers and a B6 subwoofer. Even though being PA loudspeakers and normally used for far field applications, their reproduction was sufficiently balanced and transparent also in the near field, not least due to their coaxial design. Their biggest advantage over the usually more suitable near field studio monitors was their availability, as they could be provided by db in the high quantities necessary for the 3D setup within a limited project time frame. The loudspeakers were positioned around a 2x2 meter listening area in a custom layout which will be detailed in the following.

The loudspeaker layout consisted of four layers as shown in fig. 1. The bottom layer (brown and black) on ground level consisted of four E8s as well as the four E12s, in pairs of one of each and evenly spaced on a circle around the listening area similar to a quadrophonic loudspeaker setup. The middle layer on ear level (blue) consisted of 7 E8s, which were arranged much like a classic 7.1 setup but with slightly different positioning of the rear speakers. The top layer (green) was set up similar to the bottom layer, consisting of another four E8s, in the same positions as their bottom layer counterparts. For the top most layer (light green), two E8s provided a stereo pair pointing directly downward to the listening area, similar to a conventional Voice of God, except consisting of two loudspeakers and thereby offering another stereo basis instead of a mono effect channel. The B6 subwoofer was placed right outside the spherical loudspeaker setup, extending the setup's frequency response toward the low bottom end. The loudspeakers were fed by db Audio's D6, D20 and D80 amplifier units.

There were multiple reasons for extending the more conventional 7.4.1 loudspeaker setup with a lower layer including the additional four subwoofers, as well as the Voice of God stereo pair. On the one hand, one of the most important aims of the loudspeaker system was to create a soundscape which was as immersive and enveloping as possible. The two additional layers provided additional spacial resolution which helped closing the top and bottom gaps in the spherical soundscape and was thereby very valuable in achieving this aim. The four subwoofers additionally helped achieving maximum envelopment, as the lower frequency range could also be reproduced from different directions, unlike a more conventional point source subwoofer. The system was, on the other hand, tailored to specific content, which included the distinctive rattling and screeching noise of the wheels and tracks of the tram. The lower layer of full range speakers helped reproduce these sounds a lot more convincingly as a conventional 3D setup, which doesn't provide any loudspeakers below ear height.

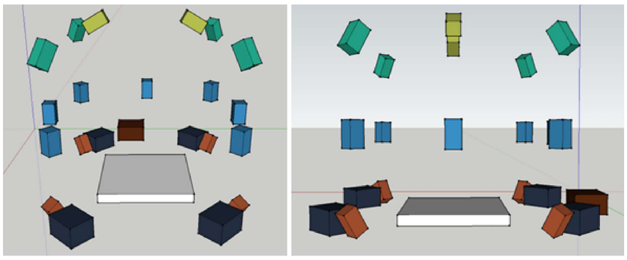

Fig. 1. Loudspeaker placement around the listening area

In addition to the loudspeaker setup, four stereo headphones (AKG K701) were provided, one for each listener, and fed by a Dante enabled stagebox (Yamaha Tio). An open-back headphone design was chosen to allow the loudspeaker's sound to leak through to the listener and mix with the headphone's sound. This helped on the one hand, ensuring that the listener wouldn't miss anything reproduced by the loudspeaker setup and on the other hand enable the two reproduction systems to interfere and mix with each other in the interesting ways we intended them to. Also, two body shaker units were attached below the listening area to reproduce the distinctive vibration of a railcar on tracks and also to complement the sub bass with additional structure-borne sound to further emotionalize the experience. The body shakers got their signal from a stereo amplifier unit, which itself was fed by the Yamaha stagebox mentioned above.

A laptop running Max 8 played back the final 26-channel WAV-files and fed the audio streams into a Dante network via a RME Dante interface. Both the amplifiers and stagebox were connected to the Dante network, and received their corresponding audio channels via the Dante audio network protocol. The amplifier units were situated in the same room as the setup. An acoustic partition helped dampening their fan noise and thereby acoustically separating them from the listening area.

1.2 RECORDING APPROACH

The recording process was approached with two goals in mind. The first was to capture the sounds in manner that depicts them in a realistic way that doesn't distort the acoustic perspective too much. The second was to record signals which have a correlation of around zero, which can generate a sense of envelopment in the listener.[1] This prompted the decision to use two different microphone setups simultaneously.

On the one hand, a Schoeps ORTF-3D was employed, an eight-channel array proposed by Wittek and Theile.[2] It is based on the popular stereophonic microphone technique ORTF and extends it in all directions in the horizontal and vertical plane. The small form factor allows for this array to be used in a very flexible way. In this project the ORTF-3D was mainly utilized to capture a precise image of concise sounds in three-dimensional space, e.g. a door closing. Because of the super-cardioid polar patterns and the close spacing of the microphone-capsules (18cm), the ORTF-3D produced signals that lacked in low frequency content and had a relatively high degree of correlation. These properties, which enable the precise imaging, compromise the capability to produce a sense of envelopment. Therefore, the signals were not serviceable for atmospheric sounds, because they resulted in a rather "narrow image" when listening back.

To compensate this, the second microphone setup was utilized to record the atmospheric sounds. This setup consisted of eight omni-microphones in wide pairs (around 3m) spaced multiple meters apart throughout the entire tram carriage. This technique delivered a set of four simi-

lar sounding stereo pairs, which had a correlation of around zero down to the lowest frequencies. The signals turned out to be very suitable for rendering the spaciousness of the tram and to create the desired sense of envelopment in the listener.
The final recording session took place in cooperation with the Stuttgart local tram services SSB, who provided an exclusive train ride. This enabled the possibility to place microphones inside the tram without restrictions and capture the tram ride free of passengers and other disturbances. Additionally, it opened the opportunity to record a variety of sounds as isolated as possible. Among other things, the tram bell, doors opening and closing, the train resting in silence, a handful of loudspeaker announcements and an emergency braking were recorded.
The combination of these two contrasting microphone setups provided a wide range of possibilities, especially regarding the further processing and mixing of the various tram sounds.

1.3 SOUND DESIGN METHODS

The main goal was to create sounds that would still give off the feeling of a tram but be considered more musical/tonal in nature. To make things easier and more concise, we split the sound designing process into four parts: beginning/ending, first act with music and interplay between headphones and 3D setup, interlude, and the finale. Each sound designer worked in their own DAW and a familiar environment. Initial sound experiments were judged by us and re-arranged in a continuous flow alongside the development of the project itself and its core idea, that underwent multiple changes.
We used the train and tram recordings in stereo and a multitude of library sounds and other samples of our own. Each sound designer was working on their sounds with a different approach. The transition from working in stereo to creating 3D sounds was found to be a large technical and creative challenge. Though, there were some techniques that proved to be very useful to 3D designing, and thus were used broadly: Envelope shaping/following, frequency extracting, EQs and the use of multiple delays. Other used techniques include wavetable synthesis, granular synthesis, and oscillation.
To create interesting new sounds that fit the theme, one needs to find a starting point. For this we scrolled and listened through the recordings we made and separated the parts that caught our interest. From there on one must apply sound designing skills and use the above stated tools/practises to morph the sounds into the desired direction.
As stated, it was desirable that the sounds would still be recognizable as tram sounds to an extent, so a major part of the end results root in our recordings. This excludes real instruments, such as the bass guitar, a drum kit, the piano, and strings, that were additionally recorded and some orchestral midi instruments which were used to create the song/music in the beginning and the finale of the piece. Pads and effects were solely made from extracted frequencies and mono recordings like the tram bell, announcements, or the warning beeping of the door. Additional recordings, that were not related to the tram, e.g., the crushing wave effect, were also recorded by us.
With the construction of the setup, we translated all prior used techniques to the 3D audio system, which proved to be extensively more complex than working in stereo. To get rid of the high correlation that comes with the usage of multiple same files, a lot of tracks must be duplicated, slightly changed, and moved. Multiple instances of reverb and delay with different timings, filters, pitches, sizes, and states of wetness must be created and routed accordingly for each pair of speakers. This creates a feeling of envelopment, discretion, or movement.

1.4 MIXING

Having completed the raw sound design, the next step was the mix of the soundscape. The main goals while mixing were to increase the overall clarity of the soundscape, to accentuate several distinct sounds and to extend the auditive spaciousness, that is to extend the external sound sphere beyond the borders given by the loudspeakers.
In order to work as flexible as possible, we exported the stems of all the sounds from the sound design sessions and loaded them into one new, combined session. We mixed the project in Steinberg's Nuendo, as this DAW offers a lot of possibilities in 3D mixing – or panning, respectively. We grouped the loudspeaker signals into several master busses: one each for the different layers alongside some busses combining loudspeakers from several layers. This made it easier to change the overall volume of the different layers, which proved to be a valuable aid. The separation of the soundscape into several sections, which we established in the sound design approach, was beneficial during mixing as well, as we could work on the mix with several people.
The increase of the clarity and the accentuation of sounds was mainly realised by changes in volume and adjustments with filters. The increase or decrease of the volume or even the muting of single sounds was partially sufficient to make a section of the soundscape clearer, as the layering of many sounds at once was reduced. At times, an adjustment of the overall volume of related sound groups or a whole layer (as stated above) was additionally necessary. This also helped to shift the focus to certain sounds or auditory movements that were not as distinct before the mix. The filtering of sounds or sound groups with equalizers reduced disturbing frequencies and aided the reduction of layered sounds. These filters cleared the soundscape up even more and improved the overall listening experience.
The extension of the auditory spaciousness was realised by changing the channel-based setup into a mixed setup. This mixed setup combined the channel-based and the object-based approach by encoding several sounds in 3^{rd} order Ambisonics: we had a channel-based bed that ensured envelopment, several channel-based sounds that were placed in discrete positions and thus exhausted the potential of the loudspeaker layout, but also a multitude of object-based

sounds in Ambisonics. On the one hand, the latter made it possible to pan these sounds freely in the setup without being bound to the loudspeaker channels. On the other hand, we could place them outside of the auditory space provided by the loudspeakers. This increased the gap between the internal and the external sound sphere and made the difference between headphones and loudspeakers more noticeable. Furthermore, for the object-based sounds at least, the loudspeakers seemed to disappear, that is, it was not possible to pinpoint the loudspeaker or group of loudspeakers responsible for the playback of these sounds, which increased immersion.

1.5 MEASUREMENTS

The following describes and illustrates the measuring setup for the technical examination of the 3D audio installation “Linie U54”. For this, the measuring software SmaartV8, several Dante interfaces and the loudspeakers with amplifiers in the rig came into operation. The measuring environment was the studio D of Stuttgart Media University’s sound studio, in which the audience also experienced the installation. In doing so, the reception environment was not left but was geared as close as possible to the listening impression of the visitor.
Several measurements were planned and performed with different positions, microphones and measuring methods that describe the centre of the installation to make differences in the various loudspeaker positions comparable to a zero point. Similarly, the measuring points were located on the listening positions of the audience, to make differences between the various listening positions visible and to make the influence on the perception comprehensible. Moreover, the utilized headphones of the installation in combination with a dummy head were important measuring instruments, to detect the influence and perception with and without headphones. With impulse measurement and transfer measurement, the utilized software SmaartV8 (version 8.5.1 – 2021) provides two important measuring methods.
In the first measuring setup, it was measured to the zero point of the installation. In doing so, the loudspeakers were delayed to zero to a measuring point in the in centre of the installation and the levels were balanced. This zero point is located in the same distance on the horizontal plane to the four surrounding visitor positions. The measuring height is in 120cm, the average ear height of a sitting visitor on an upholstered chair. It was first measured with a Beyerdynamic MM1, directed vertically to the ceiling. The required reference signal for the transfer measurement was sent in a loop through the analogue-digital and digital-analogue converters to eliminate possible measuring errors during the conversion. In the second passage a Neumann KU100 came into operation. A measurement of the left and right ear was conducted using a dummy head. Like this, the impact of the head on the perception of the directional information can be targeted in the evaluation, without the interference of the headphones. In each of the third and fourth passage the dummy head was equipped with a pair of headphones and the measurement was repeated for every loudspeaker. Hence the influence of different headphones on the frequency response and the thereby connected change of the directional perception can be examined. In the last measuring passage four measuring microphones (Beyerdynamic MM1) were placed in the listening positions of the four visitors. Here, the microphones pointed vertically to the ceiling. With this measurement the level difference across the frequency response between the four listening positions can be evaluated, that is, to what extend the position of the listener affects a varying level- and frequency-related directional perception.

2 OBSERVATIONS

2.1 OBSERVATIONS BY THE AUTHORS

2.1.1 DIRECTIONAL PERCEPTION

When experiencing the 3D audio sound installation, it can be observed that subjective directional perception becomes less accurate with headphones on. As shown in figure 2, this can probably be attributed to the dampening effect of the open headphones.

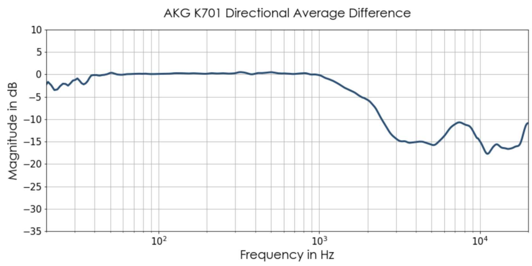

Fig. 2. Differential frequency response of AKG K701 headphone averaged over all speaker positions

The graph shows the differential frequency response of the AKG K701 headphones averaged over all speaker positions. This results from the average frequency response of the transfer measurements of all loudspeaker positions (without subwoofer) to the dummy head without headphones minus the average frequency response of the transfer measurements of all loudspeaker positions (also without subwoofer) to the dummy head with headphones:

$$f_{diff} = f_{avg.\ without\ headphone} - f_{avg.\ with\ headphone}$$

The transfer measurements of the individual loudspeaker positions result in each case from the arithmetic mean of the left and right channels of the dummy head.

2.1.2 BENEFITS OF SEVERAL DIFFERENTLY POSITIONED SUBWOOFERS

The multitude of subwoofers was beneficial especially for the creative arrangement of the soundscape. It was possible to playback sounds with several subwoofers at once as well as alternately. Like that, we were, among other things, able to create a slight spatial movement of low frequencies in the auditory space. Four of the subwoofers were lo-

calised next to the speakers in the lower layer, which made it possible to support several sounds coming from below. These four subwoofers were often used simultaneously or in pairs, the latter meaning especially the two in the front or the two in the back.
The fifth subwoofer localised outside of the loudspeaker setup was mainly used for extremely low-end frequencies below 60 Hz. It partially played as the only subwoofer, foremost serving as an effect-channel, but also was a valuable aid in supporting the other subwoofers by extending the playback to even lower frequencies.
An additional subwoofer-element were the two bodyshakers that were attached to the bottom of the stage. They transferred the low frequencies to the stage using it as a kind of resonant body, which resulted in a shaking or vibration of the stage. This offered a further possibility to play low frequencies that was decoupled from all the other subwoofers. The sounds coming from the bodyshakers were less audible, but more sensible. We found that short, distinct sound impulses were especially effectful for our purposes. For this reason, the bodyshakers often simulated the rumble of the tram on the tracks that extended the sensation of the soundscape to the physical layer and increased immersion.

2.1.3 THE BENEFIT OF THE LOWER LAYER

The use of an additional lower layer, consisting of two pairs of speakers each in the front and in the back, increases the envelopment of the entire sound image. In direct comparison to the setup without lower layer, it is noticeable that the soundscape appears to float in the room at the level of the head. The difference is especially noticeable with sound effects that move vertically from above and with typical train/track noises, as these should be heard from below the head for a more natural sound. The use of two additional body transducers below the stage also helps to make the experience more immersive. They are switched on when particularly strong vibrations are desired, such as during hazard braking or when the train knocks over the tracks.

2.1.4 VOICE OF GOD

The Voice of God speaker is used to display events that happen directly above the listener, such as the announcement at the end of the experience. But a special feature in the installation is the Voice of God in stereo. Testing a mono source worked quite well to picture stationary events. The Difficulties started with moving sound-objects. They jumped from the upper layer to the Voice of God and disrupted the immersion. That is why an inclusion of the Voice of God into the upper Layer was needed.
Splitting the one Voice of God into two sources shortened the distances to the upper layer and resulted in fewer jumps between the speakers. In addition to that, the original Voice of God was not lost, but replaced by a phantom source. Even if a little directness is gone, the improved immersion is more important. Furthermore, the sweet spot is enlarged, which is advantageous for the intended number of listeners.
Now that the speakers are part of the setup, they are used to create the atmosphere and not only as effect-channels. Consequently, the freedom to move sound objects above the listener from left to right, from front to back and vice versa is given without the listener noticing exact sources. Besides, a movement from top to bottom and backwards on the auditory axe is also possible, although the lower layer has no speakers here.
In conclusion, the stereo Voice of God is a benefit for the installation and opens many possibilities that would otherwise not be there.

2.2 RECEPTION BY LISTENERS

The sound installation ”Linie U54” was publicly exhibited on Thursday, June 30 2022 at the ”Media Night” event series. It revealed unbiased and constructive feedback from the audience.
Already in the run-up to the event evening we designed tickets for the visitors in order to be able to control the expected increased number of participants in the best possible way.

Fig. 3. Final print of access authorization for the sound installation at ’Media Night’

We scheduled the time slots in 15 minute segments of four possible participants each. Due to the expected increase in the number of visitors, we additionally offered off-plan shows between the individual time slots and extended the accessibility of the installation by one hour until about 11 p.m. In order to give as many interested people as possible a chance to experience and evaluate the installation, we also extended the number of available seats per run to six.

Table 1. Visitor traffic record for the sound installation at Media Night

Time slot (p.m.)	No. of visitors
6.00	4
6.15	4
6.30	4
6.45	4
6.52	**4**
7.00	4
7.15	4
7.30	4
7.45	4
8.00	4
8.15	4
8.22	**4**
8.30	4
8.45	4
9.00	4
9.07	**4**
9.15	4
9.30	4
9.42	**4**
9.45	4
10.00	**6**
10.08	**6**
10.15	**6**
10.22	**6**
10.30	**6**
10.43	**6**
10.45	**6**
10.52	**6**

Note: The time series **highlighted in bold** were carried out on an unscheduled basis.

Thus, in the course of the evening, there were an additional 64 users for 64 pre-arranged visitor seats, resulting in a total number of 128 test subjects.

The following evaluation was created based on participants' individual subjective perceptions statements following their experience with the sound installation. It is based on spontaneous survey interviews and has therefore no empirical significance.

Dynamics:

The attack of drum beats, percussion and various train sounds was basically described by the participants as accurate and clear. The bass precision also felt accurate and clear. In this context, the resonance of the floor plates was often highlighted as an increased immersion factor. Only during the introduction to the climax some users did report a certain head booming and attributed this to the increased use of the bass. Nevertheless, most participants described the dynamics of the composition balanced and powerful.

The loudness was described by most respondents as rather loud, although it was noted here and there that the softer parts were also found pleasant and fit well into the overall composition. Especially the loud climax was often described with the feeling of goose bumps.

Timbre:

The strength of low, mid and high frequencies was described sharp. The pure and clear brilliance of the train sounds was hereby especially emphasized. Most testers felt clarity in the upper frequencies without being sharp or shrill and without distortion. It was also noted that the bass temporary extended down in the low end of the spectrum and thus offered a good impression of depth. Overall, the balance between bass and treble was described neutral and not disturbing. In General, the presence of simultaneous high and low frequencies was described with a full sound impression.

Spatiality:

The spatial extent was especially praised. The interplay between flatness and depth as well as tightness and width many users considered particularly successful. In addition, many testers felt very enveloped in the scene, especially in the second half of the composition. The strong contrast to the flat, two-dimensional headphone mix was felt very innovative and interesting. When asked about where the sound was actually coming from, participants partly reported the feeling of irritation, but also curiosity. In the group of six, the far left and right sitting users reported feeling skewed to one side. Nevertheless, the spatial balance was overall described neutral. The distance interplay of near and far sound sources was found very exciting several times. Most participants were not able to describe exactly whether the individual sounds played were to be located distinctly internally or externally.

Envelope vs. headphones experience:

The feeling of envelopment was reported both at points of increased intensity and during rapid direction-changing auditory events. In the opinion of the users, the envelopment was credibly staged. As explained in chapter 2.1.1, the subjective directional perception decreased when the headphones were put on, but the enveloping capacity of the sound installation was hardly noticeably affected.

In addition to the ability to extend depth through the surrounding multidirectional speaker system, the headphones, in contrast, provided an internal, almost intimate sense of closeness according to attendees. And this is where our sound experience manages to push the boundaries of listening experiences. The expansion of the listening depth spectrum on the one hand, but also the interplay between internal and external stimuli plus very individual listening experiences in particular characterize our approach here and make it unique and iconic.

According to the listeners, the confrontation and interaction of two worlds awards our idea significantly: The uncontrollable listening experience, which flows seamlessly back and forth between a private, explicit inner world and an abstract, inspiring outer world, constructs a very individual listening reality for each visitor. It also varies with each run-through and this is how the audience, as the central recipient of the listening experience, makes this sound installation really effective and complete.

2.3 ACKNOWLEDGEMENTS

First and foremost we are extremely grateful to our supervisors, Prof. Curdt and Prof. Dr. Melchior for their invaluable advice, continuous support, and patience during planning, implementation and execution of the Master's project. Their immense knowledge and plentiful experience have encouraged us all the time. In particular, the constant close coordination through weeklys but also spontaneous status meetings in between gave us a great added value. They always showed us the scope of possibilities and continuously challenged us.
We would also like to thank Felipe Sanchez and Lasse Nipkow for their technical support on event installations and 3D audio development. We would also like to thank Matthias Reusch for demonstrating his measurement expertise and Sebastian Bartmann for asking the right questions in the early development phase of the composition.
A warm thank you also goes to Lena Meinhardt and Eva Dörr, who provided with their art installation "ralentir" a fundamental impulse to our idea approach.
Special thanks also go to the sponsor partners: Stuttgarter Straßenbahnen AG, db audiotechnik GmbH Co. KG, SCHOEPS Mikrofone and LAUTMACHER Veranstaltungstechnik GmbH Co. KG.
It is their kind help and support that gave the project that additional, beneficial extra.

3 REFERENCES

[1] L. Nipkow, G. Zielinsky and T. Ammermann, *Die Bedeutung von 3D bei Immersive Audio*, presented at the 30^{th} Tonmeistertagung – VDT International Convention (2018 Nov.), proceedings.

[2] H. Wittek and G. Theile, *Development and Application of a Stereophonic Multichannel Recording Technique for 3D Audio and VR*, presented at the 143^{rd} Convention of the Audio Engineering Society (2017 Oct.), convention paper 9869.

THE AUTHORS

Arne Morgner, Roland Ernst, Christian Tobias, Jens Kothe, Andreas Kraft,
Lukas Münter, Tim Philipp

Master’s project team in the Audio department of the Audiovisual Media programme at Stuttgart Media University