H. Wierstorf, C. Hold, and A. Raake,
"Listener Preference for Wave Field Synthesis, Stereophony, and Different Mixes in Popular Music,"
J. Audio Eng. Soc., vol. 66, no. 5, pp. 385–396, (2018 May.).
DOI: https://doi.org/10.17743/jaes.2018.0019

Listener Preference for Wave Field Synthesis, Stereophony, and Different Mixes in Popular Music

HAGEN WIERSTORF,[1] *AES Associate Member*, **CHRISTOPH HOLD,**[2, 3] *AES Student Member*, **AND**
(hagenw@posteo.de) (chris.hold@mailbox.tu-berlin.de)

ALEXANDER RAAKE,[3] *AES Associate Member*
(alexander.raake@tu-ilmenau.de)

[1]*Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, UK*
[2]*Assessment of IP-based Applications, Technische Universität Berlin, Berlin, Germany*
[3]*Audiovisual Technology Group, Technische Universität Ilmenau, Ilmenau, Germany*

Stereophony and wave field synthesis are capable of providing the listener with a rich spatial audio experience. They both come with different advantages and challenges. A direct comparison in terms of listener preference has rarely been carried out in previous research, since these methods differ in their production stage. This study generated mixes of four different popular music recordings and introduced systematic changes to the wave field synthesis mix for one song. Listeners rated their preference comparing the different mixes and reproduction systems. Part of the experiment was repeated using a binaural simulation of the involved loudspeaker setups. The results show that a presentation using a high number of loudspeakers is preferred and the differences between reproduction methods can have a larger influence than strong variations of single mixing parameter.

0 INTRODUCTION

The most common loudspeaker reproduction methods in the consumer market are 2- or 2.1-channel stereophony (stereo) and 5.1-channel stereophony (surround / 5.1) [1]. Several multichannel methods that provide a higher number of surrounding loudspeakers are available as well [2]. More loudspeakers can lead to a higher emotional reaction and encourage deep immersion of listeners during reproduction [3]. The question arises if integrating more loudspeakers will also enhance the overall listening experience for listeners. In a study comparing stereo, surround, Ambisonics, 9-channel, and 22-channel stereophony Francombe et al. [4] showed that listeners preferred systems with more than five channels for different types of broadcast content.

The present study assesses if the same preference for systems with a higher number of loudspeakers is present for wave field synthesis (WFS) [5, 6] when compared to stereo and surround in the context of popular music. There is evidence against this as WFS is known to introduce stronger coloration on single sound sources than stereophony [7].

Mixes of popular music for WFS are produced in a slightly different workflow than for stereo and surround. When listeners are asked which reproduction system they prefer for a given song the actual mix might influence their choice as well. The goal of the study is to quantify the influence of the mixing process on listener preference.

The involved reproduction systems can be simulated via headphones, employing binaural re-synthesis to mimic the involved loudspeaker setups [8]. This is especially of interest for investigating loudspeaker-based methods that rely on a large number of loudspeakers, which might be hard to realize in a listening test [9]. The study employs binaural simulations of the loudspeaker reproduction systems and compares listener preferences obtained with those simulations and the actual loudspeaker setups.

We assess those three questions in paired comparison listening tests asking listeners to rate their preference. Experiment I investigates the influence of the reproduction system and four different mixing parameter on listeners' preference ratings for one song. This is achieved by systematically varying the mix for WFS and compare it to stereo and surround reference mixes. Experiment II analyzes if the listening preference for the reproduction systems is constant across songs by employing four different songs. For each song only the reference mix for WFS, stereo, and surround is employed. Experiment III repeats Exp. I, but this time all loudspeakers are simulated by dynamic binaural re-synthesis to investigate the influence of the binaural simulation on listener preference.

1 BACKGROUND

Several studies evaluated multichannel reproduction systems. The interest focused mainly on what attributes describe important underlying perceptual dimensions and how are those related to listener preference.

Nakayama and colleagues [10] were one of the first to study the relationship between listener preference and the number of loudspeakers. They reproduced two pieces of popular music recorded with an eight-channel microphone. Listeners showed a higher preference for systems involving more loudspeakers. They also collected similarity ratings and a multi-dimensional scaling identified three factors explaining 77% of the variance of their preference ratings. They labeled those three factors fullness, clearness, and depth of the image sources.

Later, Letowski [11] hypothesized that timbre and spaciousness are the underlying dimensions of sound quality that was confirmed by several newer studies, e.g., by Rumsey and colleagues [12] and Choisel and Wickelmaier [13].

On the other hand, the finding of higher preference for more loudspeakers was not always confirmed by newer studies. In the context of classical and popular music both of the latter mentioned studies disagree in their results, showing a slight preference for surround over stereo [12] versus a similar preference for both [13]. Similarly, Zacharov and Koivuniemi [14] found no difference between their best two-channel and multichannel systems they tested for a wide range of audio content, including music, speech, and environmental sounds. In the context of reproducing Ambisonic recordings of soundscapes, Guastavino and Katz [15] found a dependency on the content. Surround was more preferred for outdoor recordings and stereo for frontal music scenes.

No direct comparisons of stereo or surround systems with WFS have been done regarding preference. Several studies have shown that WFS provides high spatial fidelity in terms of localization of a single source [9, 16, 17]. It indicates that WFS is able to provide better spatial experience than stereo, but it is still under debate if high localization accuracy alone is sufficient for characterizing the spatial capabilities of a system [18]. On the other hand, WFS introduces stronger coloration to a reproduced single sound source than stereo [19, 7]. Comparing those studies, it is hard to conclude if WFS will be able to provide a preferred listening experience in the context of popular music.

The stimuli employed in previous WFS studies represented rather simple scenes, consisting of single noise or speech sources [17, 19, 20]. Simple scenes are easily transferable to other reproduction systems, but they fail to show the full artistic potential available with the different systems. The goal of this study was to overcome those limitations by creating complex popular music scenes. Popular music has high practical relevance to most listeners. At the same time, with popular music, no reference or live performance exists that the different reproduction systems try to match. It, rather, tries to reach an artistic intent at the end of the production process.

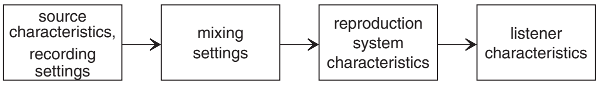
source characteristics, recording settings
mixing settings
reproduction system characteristics
listener characteristics

Fig. 1. Overall steps involved in producing and listening to popular music. All these steps have certain characteristics that have an impact on the resulting listening experience, including the listeners themselves.

The audio signals presented to a listener when listening to popular music or other type of produced audio typically have undergone different processing steps. These include recording the source, mixing the recorded signals, reproduction of the mix, and finally the listening process [21]. As indicated in Fig. 1, these may all have an impact on the listening experience. The sound engineer decides for a tonal appearance when recording and mixing the source signals to a music track. A reproduction system might then not be capable of reproducing the full artistic intent. For example, it might lack in bass or narrow the sound stage. This again can shift the listeners' attention towards unintended details within the music.

For the comparison of stereo and surround the influence of the mixing engineer can be limited by producing a music mix for the surround system and use a down-mixing algorithm to generate the stereo mix. The same approach cannot be meaningfully applied if stereo or surround should be compared with WFS, as the underlying reproduction principle of WFS differs fundamentally. The high number of loudspeakers playing coherent signals in WFS can lead to stronger problems with coloration than in stereo or surround, due to comb-filter-like spectra [7]. Mixing engineers can adjust their mixes to such problems—current up-mixing algorithms cannot. In addition, stereo and surround are channel-based reproduction systems and in most cases their content is produced employing a channel-based panning approach. WFS is independent of the number of applied channels and is in most cases mixed with an object-based approach. This requires the mixing engineers to adapt their techniques to this rather unfamiliar environment, where ideally every sound source comprises only a single sound signal and does not interact.

As a result, the mixing engineer and the reproduction system both control the perception of a song by a listener. For unfamiliar music, the mix can have a significant influence on listener preference [22] and cannot be neglected. In the study described in this paper, we tried to disentangle this influence by varying the reproduction systems on the one hand, and mixing parameters on the other hand, asking listeners for preference ratings in a paired comparison test. As mixing parameters we selected compression, EQ, reverb, and spatial positioning. Mixing engineers would typically focus on these to enhance the balance of instruments within a song.

Compression reduces the dynamic range of audio signals. When used on a single track during mixing, it can smooth out and increase the loudness of that track. It is also known to the audio community from the so-called loudness war [23], which refers to its usage on the final master of the mixing process. The sound of modern pop and rock productions relies on compressed acoustic instrument and vocal recordings. Compression is a non-linear operation and can thus involve a wide selection of parameters, which are usually dependent on the underlying type of compressor. The effects are not always obvious and the parameters are often correlated, e.g., dependent on the amount of compression applied. In addition, it can introduce artifacts like pumping, breathing, and low-frequency distortion, which appear if too much compression is used [24]. Consequently, there exists a best point of compression regarding listener preference [25, 26].

The *Equalizer (EQ)* allows the mixing engineer to enhance and correct the spectral balance of elements within the mix. In addition, filtering (applying EQ) can spectrally unmask content (e.g., separates vocals from other instruments). Shaping the sound with an EQ can impact listener preference, since bass balance [26] and brightness [27] can influence the perceived sound quality.

Reverb is mainly artificial reverb in popular music [28]. It enhances envelopment by providing space and depth information. There seems to be a perfect level between direct sound and reverb regarding listener preference [29], which might depend on the reverberation of the listening environment as well [30].

Position is the most influential mixing parameter for creating spaciousness, a common goal in the mixing process of popular music. It also ensures that there are no gaps between instruments and a spatial balance between left and right [29]. In addition, it allows to separate content by spatial unmasking. In popular music, there exists a well-established pattern for positioning single instruments. For example, lead vocals are placed in the center [31]. A deviation from this pattern might negatively influence listeners' preferences. The same holds for too narrow or too wide arrangements [13].

Spatial audio can also be presented to a listener by headphones, employing binaural synthesis to simulate the involved loudspeaker setups [8]. This is especially of interest for investigating loudspeaker based methods that rely on a large number of loudspeakers, which might be hard to realize in a listening test [9]. In addition, the availability of binaural signals would allow for a direct auditory modeling of the perceived audio quality [32]. Wierstorf and colleagues have shown that dynamic binaural synthesis results in a similar perception in terms of localization of sound scenes consisting only of one object [33]. We wanted to investigate if listeners' preference for a particular loudspeaker reproduction system and mix is affected by using dynamic binaural synthesis to simulate the involved loudspeaker setups.

We split the listening test into three experiments to limit the degrees of freedom in each test. Experiment I systematically varied the mix for WFS and compared it to stereo and surround reference mixes for one song. Experiment II employed only reference mixes for WFS, stereo, and surround for four different songs. Experiment III repeated Exp. I, but this time all loudspeakers were simulated by dynamic binaural re-synthesis.

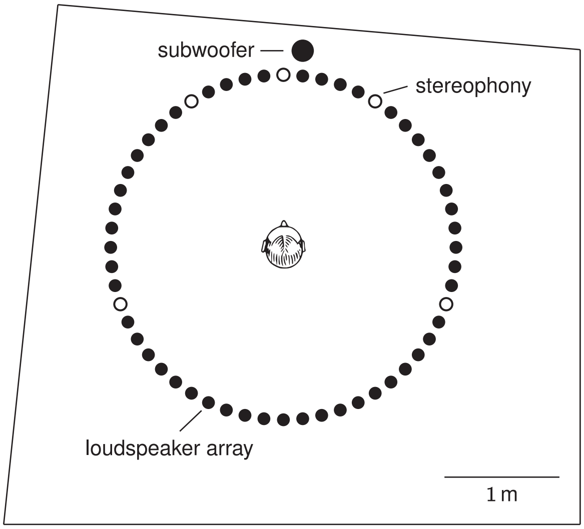

Fig. 2. Setup of the circular loudspeaker array and the subwoofer in room Pinta used for Exp. I and Exp. II. The loudspeaker marked with open circles were used for WFS and the stereophonic methods. Note that head and loudspeaker sizes are not true to scale.

The next section introduces the methods used in all three experiments. The results of the experiments are then presented separately in the results section.

2 METHODS

2.1 Apparatus

Experiments I and II took place in a 54 m^3 acoustically treated room ("Pinta" in the Telefunken building of TU Berlin). The room is equipped with a circular loudspeaker array with a diameter of 3 m consisting of 56 loudspeakers (Elac 301) and one subwoofer (Genelec 7060A) as shown in Fig. 2. The listeners sat on a chair in the center of the loudspeaker array.

Experiment III took place in a 83 m^3 acoustically damped listening room ("Calypso" in the Telefunken building of TU Berlin). The listeners sat on a chair wearing open headphones (AKG K601) with an attached head tracker (Polhemus Fastrak).

In all experiments the listeners sat in front of a flat screen placed on a table and chose between a mouse or keyboard for entering their responses.

In a separate room a computer equipped with a multichannel sound card including D/A converters (RME Hammerfall DSP MADI) was used to play back all sounds. In Exp. I and Exp. II, the active subwoofer received an analog signal directly from the sound card. All remaining 56 channels were sent via MADI to the listening room and D/A converted as well as amplified by custom-made units. In Exp. III the signals traveled through a headphone amplifier (Behringer Powerplay Pro-XL HA 4700) and analogue cable to the headphones in the listening room, a distance of approximately 5 m.

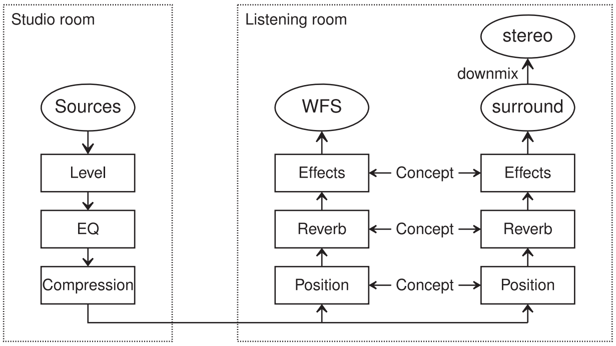

Fig. 3. Block diagram illustrating the basic multi system sound mixing procedure, from the source material to the finished output.

2.2 Stimuli

2.2.1 Audio Material

For Exp. I and Exp. III, the audio material consisted of a multi-track recording session, with double tracking mainly for guitars and vocals. It was a moderate tempo pop music piece including deep male vocals, acoustic and electric guitars, bass, drums, shaker, reverb, and delay effects ((unpublished)—Lighthouse) [34]. The mixing engineer also recorded the track, applying only light and broad tonal shaping with an EQ and compressor. This ensures that the recordings don't restrict further processing.

For Exp. II, three additional stimuli were generated from freely available multi-track recordings. The three recordings consist of a pop-rock song with live feeling and male vocals (The Brew—What I Want)[1], a slightly heavier rock song with female vocals (Hop Along—Sister Cities)[2], and a shorter hip-hop track (Lushlife—Toynbee Suite)[2] with male rap vocals. In addition, Exp. II included two versions of the songs "Lighthouse" and "What I Want," once played completely and once starting at its spatial and quite different sounding guitar bridge part (B).

2.3 General Mixing Concept

The mixing process of popular music involves stages that are, to some degree, independent of the involved reproduction system. On the other hand, some stages like positioning and reverberation are usually adjusted on each individual system to ensure their best possible usage. Applying advanced mixing techniques demands adaptation to WFS, since most modern mixing techniques are based on channel-based stereophony and not on the object-based approach taken by WFS. Particularly, this includes parallel and bus-compression used extensively in modern pop productions.

Fig. 3 shows a block diagram of the basic layout we applied to create comparable mixes for the different systems. The left box describes the system-independent steps, the right box all steps performed on the actual systems used in the listening test. The system-independent mixing took place on a separate stereo reference system in a studio environment, well known by the mixing engineer. There, he applied what he considered to be the optimum amount and type of EQ and compression.

Table 1. Quantification of the mixing parameters.

Condition	EQ Gain	Compression Gain Reduction	Reverb Level	Position Foreground
– / off	off	off	off	very narrow
–	0.5	0.5	−6 dB	∼ stereo
○	1	1	0 dB	reference
+	2	2	+6 dB	wide
++				very wide

The final system-dependent processing included position, reverb, and other effects. It was guided by an underlying joint concept, which means that every processing and actual mixing decision is consistent across systems. This avoids fundamentally different mixing results and ensures comparability. The joint concept was based on currently common mixing techniques in popular music. All main components of a song were positioned in the frontal scene. Lead vocal, snare drum, and bass were always positioned in the center. Elements that likely create envelopment, e.g., ambient sounds or delay effects, were allocated to all directions. The main reverb was a quadrophonic reverb whose outputs were placed in each corner for WFS accordingly. All special-effects processing, such as modulation-based effects, were made as similar as possible between the systems.

Nevertheless, each system needs idiosyncratic adaptations to perform adequately. Those adaptations tackle especially the positioning of single objects in the music scene. The latter followed a conservative handling, which means avoiding extreme and unconventional settings, as well as omitting moving sources. In very few cases this change in position led to small level corrections.

At the end of the mixing process, a validation and correction of all settings—including EQ and compression—was performed by directly comparing the systems. If an adjustment was needed, it was carried out by the same amount for every system.

2.3.1 Variation of Mixing Parameters

For Exp. I and Exp. III and the song "Lighthouse," the mixes for WFS were varied for the following mixing parameters in a systematic way as summarized in Table 1.

Compression There were alterations in both directions, more and less compression. The gain reduction represents an indicator for the amount of compression applied per instance. Starting from the "reference" (○) mix, the variant containing "more compression" (+) applies around twice as much average gain reduction per compressor. Therefore, the ratio of each compressor was doubled. In case of fixed-ratio compressors, or if the amount of compression was not sufficient yet, the corresponding threshold was lowered for the desired amount. The procedure was reversed for the variant "less compression" (–), where half the gain reduction was applied. Bypassing all compressor instances produced "no compression" (off).

[1] Single tracks downloaded from http://bit.ly/telefunken-ea

[2] Single tracks downloaded from http://bit.ly/medleydb [35]

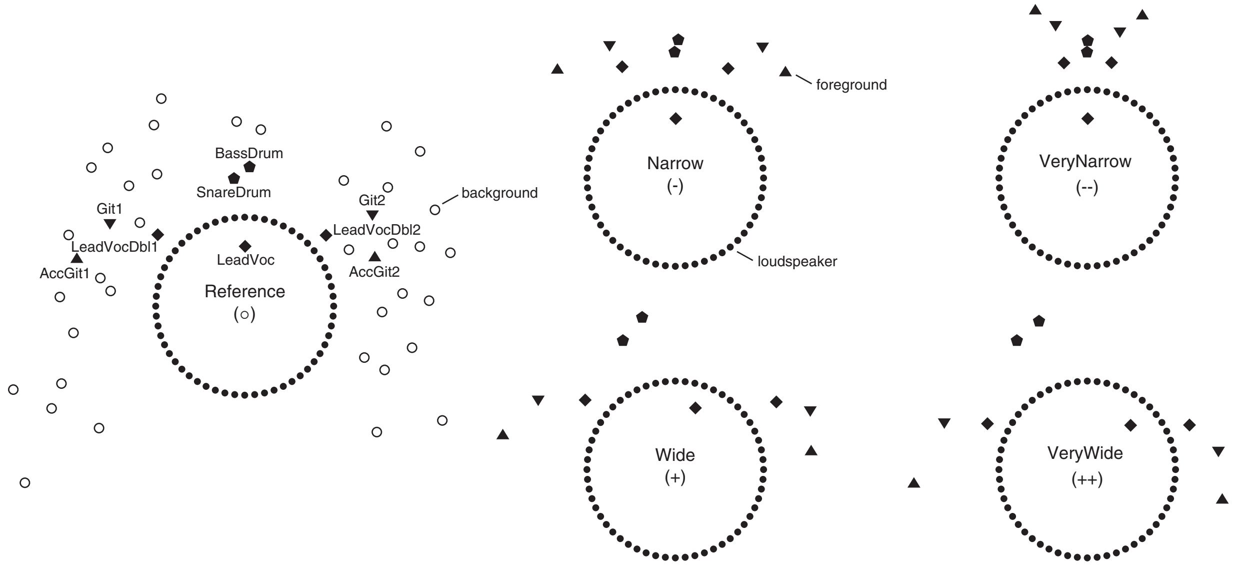

Fig. 4. Arrangement of the different object-based mixes for WFS varying the mixing parameter position. The positions of the different foreground elements (vocals, drums, guitars) are shown by the filled symbols relative to the circular loudspeaker array for the five different cases. For the reference mix the sound objects belonging to the background are indicated by the open circles and omitted for the other conditions as they remained the same.

EQ The EQ settings from the reference mix (○) were altered in both directions, applying more and less filtering. Therefore, the amount of boost or cut per filter band was scaled. For the condition "more EQ" (+), every intervention was exaggerated by doubling the amount of applied filter gain. Respectively, every applied filter gain was halved for the condition "less EQ" (–). Bypassing every equalizer in each channel finally led to the condition "no EQ" (off). Note that this setting also bypassed all involved high-pass filters.

Reverb The condition "more reverb" (+) was introduced by increasing all the associated effect return levels by 6 dB. For "less reverb" (–), those return channels were lowered by 6 dB accordingly. "No reverb" (off) corresponds to those effect returns being muted.

Position Starting from the reference mix, two different mixes with more narrowly spaced foreground elements and two mixes with more widely spaced foreground elements were produced. As audio objects belonging to the musical foreground vocal, drums, and guitar were chosen matching the three most mentioned instruments regarding "like" and "quality" from a previous study by Wilson and Fazenda [22]. Here, the foreground instrument "drum" was represented by the sound objects "bass drum" and "snare drum" and its remaining parts were considered to belong to the background. Besides the lead tracks, common pop music practice includes vocal and guitar/harmony double tracks, which were moved accordingly. The reference WFS mix (○) represents a common and modern variant with lead tracks in the center, guitar tracks positioned to the side, and double tracks spread symmetrically, compare Fig. 4. This arrangement is similar to the stereo mix, however moderately wider. For the "narrow" (–) version all foreground tracks were moved towards the center. The "very narrow" (–) mix consists of a center-foreground base set a little narrower than for stereo. In the "wide" (+) mix, the foreground objects are gently shifted away from the center of the scene, retaining an appropriate and symmetrical impression. Hence, the lead vocal and drum tracks are shifted inversely, with the drums on the left and lead vocals on the right side. In the "very wide" (++) mix, some guitar parts finally appear from behind the listeners and the lead vocal from the right of the listener. The background part of the piece, including reverbs and delays, remain at their reference position in every mix.

The metadata [36] for the different positions are available in the audio scene description format [37]. In addition, the signal feeds [34] and the finished mixes [38] for all songs can be downloaded as well.

2.3.2 Rendering and Binaural Synthesis

An open source WFS renderer (SoundScape Renderer [39]) was used to compute the loudspeaker driving signals. We extended the renderer by an amplitude decay compensation to allow easier positioning in musical mixes [40]. Version 0.4.3 of the SoundScape Renderer will include this extension in a slightly modified version.

Experiment III was not conducted with the real loudspeakers but with an anechoic simulation of the same loudspeakers realized by using dynamic binaural re-synthesis [8]. For the simulation one binaural room scanning (BRS) file was created for each loudspeaker [41] with a resolution of 1° utilizing high resolution head-related impulse responses recorded with a KEMAR dummy head [42]. During playback the binaural synthesis software (SoundScape Renderer [39]) convolved every BRS file with the corresponding loudspeaker driving signals, which were summed and returned as headphone signals. The binaural renderer updated the ear signals depending on the head orientation of the listeners, which was captured by the head tracker with an update rate of 120 Hz. The HRTF switching for the dynamic binaural synthesis was performed on an audio block length of 1024 samples, resulting in an estimated latency of the whole dynamic binaural synthesis of around 70 ms.

2.3.3 Loudness Adjustment

Loudness can easily dominate listeners' preferences [23] and has to be equalized between all conditions. For Exp. I and Exp. II the loudness was adjusted between the systems by means of dummy-head recordings at the listener position. One system represents a reference and the other two systems are adjusted to the level of the reference system. For all recordings, loudness model estimations (non-stationary Zwicker model [43, Sec. 8.7] as implemented in the GENESIS Loudness Toolbox for Matlab 1.2)[3] allowed for an accurate adjustment of the different systems.

For Exp. II, the same loudness of the binaural signals for the different conditions was ensured by correcting the signals for a head orientation of 0° applying the same loudness model.

2.4 Participants

One-hundred-twenty-three participants (age range: 18–72; mean age: 31.2) were recruited for the listening test and were equally distributed to Exp. I, Exp. II, and Exp. III, resulting in 41 participants for each experiment. They self-reported no hearing loss or hearing disturbances. Informed written consent was obtained from each participant and they received a financial compensation. The study received ethical approval from the Technische Universität Berlin Ethics Committee (RA_01_20140422).

2.5 Procedure

Participants were presented with a pair of two temporally aligned clips of music for pairwise comparison. The subjects could switch back and forth between the two stimuli. They were asked which of the two they preferred to listen to. Experiment I consisted of 90 trials of paired comparisons, 21 included only changes to the mixing parameter position, 15 changes in EQ, 15 changes in reverb, 15 changes in compression, and 24 pairs changes across the mixing parameters. In Experiment II, 18 pairs were presented to the listeners, 3 for each song. Experiment III excluded the surround condition and involved 107 trials: 15 for the mixing parameter position, 10 for EQ, 10 for reverb, 10 for compression, and 62 for changes across mixing parameters.

For Exp. I and Exp. III, playback stopped after the end of a 30 s long extract and listeners had to answer to advance to the next trial, following an inter-trial interval of 1 s. In Exp. II, the whole songs were looped and participants were allowed to listen as long as they wanted. Participants could submit their answer before the end of the trial, provided they had heard a minimum of five seconds and had heard each of the two stimuli at least twice. With the limitation of the playback time to 30 s in Exp. I and Exp. III, more pairs could be presented, also restricting the position within the song the participants used for their judgments. Before the start of all experiments, participants practiced the paradigm twice with the experimenter. Here, another extract from a song was played and one of the tracks in the pair was presented at −6 dB.

At the end of Exp. I and Exp. III participants completed a verbal survey asking for average daily hours spent listening to music and favorite music genres. Furthermore, participants were asked:

1) When comparing a pair of stimuli, what did you pay attention to or which attributes of the mix triggered your decision?
2) Try to explain reverberation, compression and equalization with respect to music production. Do you have expertise in sound mixing?

These survey responses were recorded by the experimenter.

2.6 Statistical Analysis

Suppose a number of musical pieces A, B, and C is given that should be assessed by listeners regarding their preferences. The advantage of the paired comparisons method lies in its very few assumptions about the underlying process leading to the choices of the listeners. It is able to measure specific seemingly inconsistent choices by the listeners such as circular triads where A is preferred over B, B over C, and C over A. This can appear to be a reasonable choice for a given listener for stimuli that vary in different aspects. If instead a ranking of the stimuli or a preference rating on a scale is applied, it is already assumed that the rankings lie within one dimensional perceptual scale [44]. The pairwise comparison circumvents this restriction and allows for assessing of a multi-dimensional perceptual space.

An indication of a higher dimensional perceptual space is the systematic appearance of a high number of circular triads. These can also stem from inconsistent individual choice behavior, indicated by a randomized appearance of triads across listeners. Counting the triads only provides a descriptive measure of the underlying choice process. In order to classify whether the appearance of triads is systematic and listeners agree on them, a statistical test is required. This can be achieved by fitting a Bradley-Terry-Luce (BTL) model [45] to the data. The BTL model is a probabilistic choice model that predicts the probability of choosing one option out of a pair. It assumes that this decision is independent of all other paired comparisons. If the model is able to fit the data of a paired comparison test it indicates that only a few systematic triads can be present in the data. It reveals that no systematic deviations from a one-dimensional perceptual preference space occur and estimates the choices of the listener on a ratio scale [13]. The goodness of fit of the BTL model is indicated by the corresponding p-value (H_0: difference to ideal model is zero) of a χ^2 test comparing the estimated BTL model

[3] Downloaded from http://bit.ly/genesis-loudness

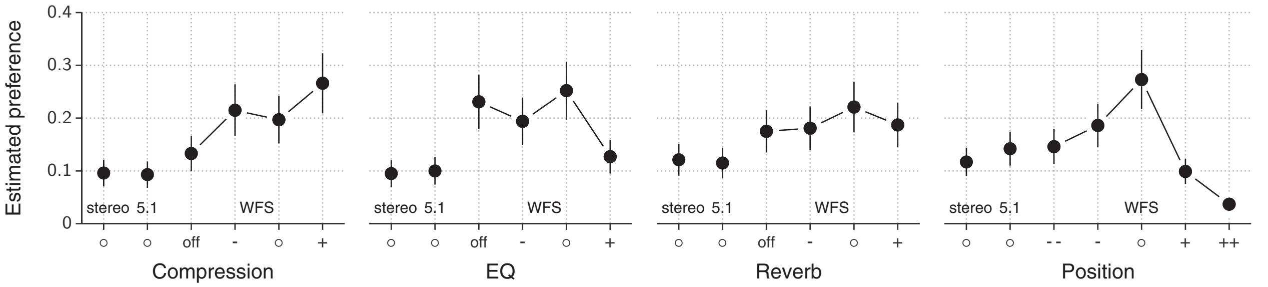

Fig. 5. Listener preferences for stereo, surround, and different WFS mixes with changes in the mix parameters compression, EQ, reverb, and position. The estimated preference is shown as the probability for each condition to be preferred together with its 95% confidence interval. Shown are the results from Exp. I. The different mix settings (off, ○, +, . . .) are quantified in Table 1.

against an ideal saturated model for the paired comparisons. Wickelmaier and Schmid [46] propose to consider models as sufficient with $p > 0.1$.

The current study used the BTL implementation in R (eba-package) after [46]. The BTL values were normalized to sum up to unity and present the probability that a given condition was preferred. All results from the listening tests, together with code for the statistical analysis and creation of all the figures is available for download [47].

3 RESULTS

For the evaluation, the ratings were aggregated over all participants in each experiment. The preference ratings were then transferred to a ratio scale by fitting BTL models to them. The goodness of fit is indicated by the corresponding p-value presented for every model, whereby a p-value of 1 would indicate a perfect fit.

3.1 Experiment I

In Exp. I the song "Lighthouse" was played to the listeners through loudspeakers using stereo, surround, and WFS. For WFS, the mixing parameters compression, EQ, reverb, and position were altered.

Fig. 5 displays the results together with the 95% confidence intervals of the BTL models, fitting the data for compression ($p = 0.92$), EQ ($p = 0.62$), reverb ($p = 0.94$), and position ($p = 0.78$). As the ratio scale resulting from the BTL model can be scaled by a factor without changing the underlying scale properties, the preference ratings are normalized to sum to 1 and can then be interpreted as the probability of a given condition being preferred. Most of the WFS conditions were preferred over stereo or rated similarly as for example the very narrow (–) position mix, which has a narrower foreground spread than the stereo mix. The only condition with a lower estimated preference was the very wide (++) position mix. For compression and reverb, the mixing conditions with no processing for WFS (off) were only slightly preferred over stereo and surround. The condition no EQ (off) was rated the same as the reference. With the exception of the compression parameter, the WFS reference mix was the most preferred condition. Stereo and surround led to the same estimated preference for all presentations.

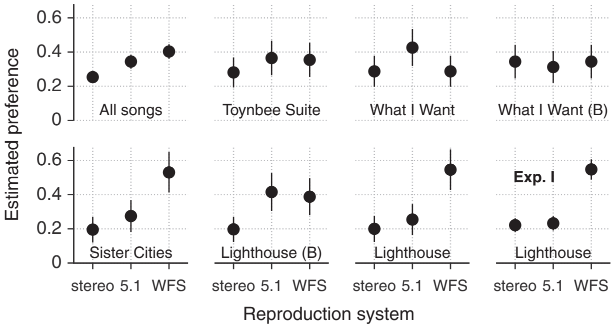

Fig. 6. Listener preferences for stereo, surround, and WFS for six excerpts from four different songs as rated in Exp. II. The estimated preference is shown as the probability for each condition to be preferred together with its 95% confidence interval. The top left graph shows the estimated preference by first pooling the results from all songs. The bottom right graph shows the results for the ratings of the song "Lighthouse" as obtained in Exp. I, which employed only one song.

The mixes with the settings no (off) and more (+) for compression, EQ, reverb and the settings very narrow (–) and very wide (++) for position were also directly compared to each other in the listening test. Fig. 8 shows the estimated preferences from those comparisons together with the results from Exp. III on the same task. Sec. 3.3 will report on the comparison of both experiments, the following will focus on the results from Exp. I only. For Exp. I the results from the within mixing parameters comparisons from Fig. 5 are confirmed. Only the mixing parameters no EQ (E_{off}) and more compression (C_+) were rated better than the reference (○). Whereas no compression (C_{off}), no reverb (R_{off}), very narrow position (P_{--}), and more EQ (E_+) were slightly preferred over stereo. Very wide position (P_{++}) was the least preferred condition and the only one where the changes of the mix had let to a worse rating than for stereo reproducing the reference mix.

3.2 Experiment II

In Exp. II six excerpts from four different songs were played to the listeners through loudspeakers using reference mixes for stereo, surround, and WFS.

Fig. 6 summarizes the results by showing the estimated preference ratings together with the 95% confidence

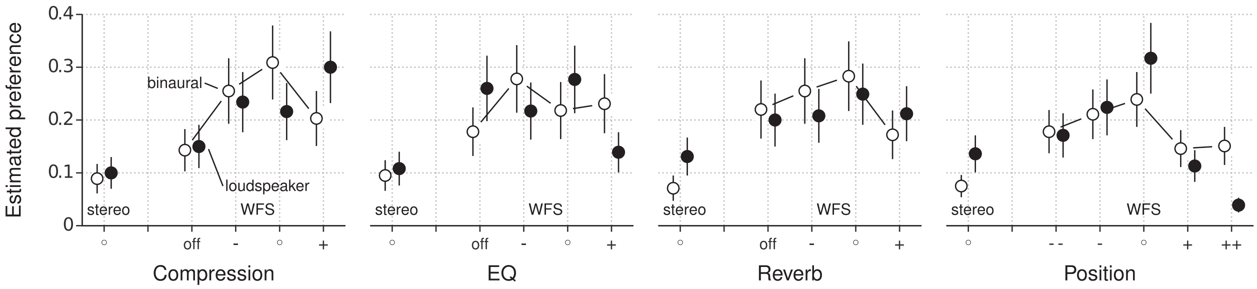

Fig. 7. Listener preferences for stereo, surround, and different WFS mixes with changes in the mix parameters compression, EQ, reverb, and position. The estimated preference is shown as the probability for each condition to be preferred together with its 95% confidence interval. The filled circles are results from Exp. I employing loudspeakers, the open circles are results from Exp. III employing binaural simulation of the loudspeakers.

intervals for the BTL models, fitting the data for "All songs" ($p = 0.69$), "Toynbee Suite" ($p = 0.24$), "What I Want" ($p = 0.78$), "What I Want (B)" ($p = 0.79$), "Sister Cities" ($p = 0.57$), "Lighthouse (B)" ($p = 0.50$), "Lighthouse" ($p = 0.57$), and "Lighthouse Exp. I" ($p = 0.37$). The preference ratings are normalized to 1 and can be interpreted as the probability of a given condition to be preferred. The results were calculated for each of the six song excerpts and for all of them together, labeled "All songs" (top left). In this case, the paired comparison ratings of all six excerpts were pooled before the statistical model was applied to estimate the preferences. In addition, the results for the song "Lighthouse" from Exp. I are presented (bottom right). They were calculated by considering all stereo, 5.1, and WFS ratings for the reference mixes from Exp. I.

The results for the pooled ratings of all songs showed an estimated preference of 0.40 ± 0.04 for WFS compared to 0.34 ± 0.04 for surround, and 0.25 ± 0.03 for stereo. Listeners are more likely to prefer WFS than surround or stereo, and to more likely to prefer surround than stereo. By comparing the ratings for the different songs, an interaction of song and reproduction system on the estimated preferences becomes visible. The songs "Sister Cities" and "Lighthouse" both are preferred for WFS, whereas the song "What I Want" is more preferred for surround.

For the songs "Lighthouse" and "What I Want" alternative excerpts starting at their guitar bridge parts were presented, labeled (B). A difference can be found comparing the results of these excerpts to the ratings for the excerpts starting at the beginning of the song. For "Lighthouse," the strong preference by the listeners for a reproduction with WFS is not existent for the excerpt starting at the bridge part. In this case, the presentations of the song with surround and WFS are equally preferred over stereo.

The results for "Lighthouse" from Exp. I and Exp. II are identical, only the confidence intervals for Exp. I are lower, since a higher number of paired comparisons were rated. The identity was verified by a Wilcoxon signed rank test looking for differences between both experiments. It compares the observations pairwise, and the resulting p-value (H_0: difference between pairs of observations is zero) provides evidence for different ratings. The test did not provide evidence for different ratings obtained in the two listening tests ($p = 0.25$).

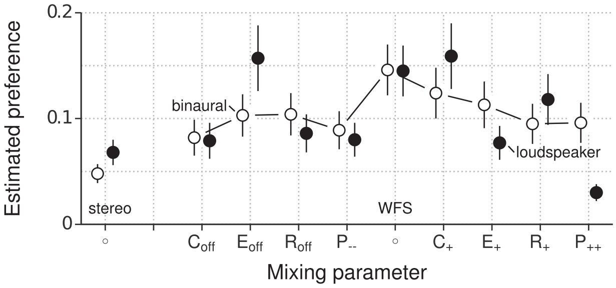

Fig. 8. Listener preferences for stereo and different WFS mixes with changes in the mix parameters compression (C), EQ (E), reverb (R), and position (P). The estimated preference is shown as the probability for each condition to be preferred together with its 95% confidence interval. The filled circles are results from Exp. I employing loudspeakers, the open circles are results from Exp. III employing binaural simulation of the loudspeakers.

3.3 Experiment III

In Exp. III the song "Lighthouse" was played to the listeners through a binaural simulation of stereo and WFS. For WFS, the mixing parameters compression, EQ, reverb, and position were altered. The Experiment repeated Exp. I, now employing a binaural simulation instead of real loudspeakers and omitting the surround condition. This experiment aimed at identifying the influence of the binaural simulation on the preference ratings.

The results are summarized and compared to the ones from Exp. I in Fig. 7 and Fig. 8. For the ratings of Exp. I, the surround condition was excluded and new BTL models were calculated to allow the direct comparison with Exp. III, where the surround system was not used. The goodness of fit of the different BTL models is indicated by their corresponding p-values (H_0: difference to ideal model is zero), which are all above 0.1: $p = 0.75$ and $p = 0.51$ for Exp. I and Exp. III for compression, $p = 0.26$ and $p = 0.94$ for EQ, $p = 0.75$ and $p = 0.27$ for reverb, $p = 0.63$ and $p = 0.41$ for position.

The results for binaural simulation and loudspeaker based reproduction are in agreement for most of the conditions as indicated by Fig. 7 and Fig. 8. The strongest difference occurs for the very wide position (++), which is rated above stereo for the binaural simulation whereas it was rated by far the least preferred condition for loudspeaker based reproduction. Smaller deviations between both experiments are observable for more compression (+), more EQ (+), no EQ (off), and stereo.

To quantify the influence of dynamic binaural simulation instead of actual loudspeaker reproduction, we evaluated the difference between the ratings of both experiments with a Wilcoxon signed rank test. This test compares the observations pairwise, and its p-value (H_0: difference between pairs of observations is zero) provides evidence for different ratings. We found strong evidence for different ratings for position ($p = 0.002$). The test provides no evidence for differences for the ratings of EQ ($p = 0.28$), compression ($p = 0.28$), reverb ($p = 1$), and the comparison across mixing parameters as shown in Fig. 8 ($p = 0.05$). Comparing the pooled results the statistical test indicates a difference between real loudspeaker reproduction and its binaural simulation ($p = 0.03$).

The difference between Exp. I and Exp. III for the mix parameter position can be further analyzed by incorporating the results from the questionnaire. For Question 1 in the binaural Exp. III, 21 participants reported they were dissatisfied when lead tracks—especially the lead vocals—were shifted outside the center. The remaining 20 participants did not report any attributes directly related to position. Two groups named “dissatisfied” and “neutral” were built accordingly. Additionally, a second grouping was analyzed. If participants of Exp. III were able to answer at least three of the four points in Question 2, they were categorized as “experts.” This resulted in 7 “expert” and 34 “naive” listeners.

The likelihood ratio of individually calculated BTL models reveal whether two groups of subjects rated differently. The corresponding p-value (H_0: difference between likelihoods is zero) indicates whether the combination of both group model likelihoods is higher than the likelihood of the model calculated from the entire population, as this distance is approximately χ^2-distributed. For “expert” versus “naive” listeners, the statistical test provided no evidence for a difference between both groups for the mix parameter position ($p = 0.5$). For the two groups of participants that reported to be either “neutral” or “dissatisfied” with laterally shifted lead vocals, the statistical test provided strong evidence for a difference for the mix parameter position ($p < 0.001$).

For the two groups “dissatisfied” and “neutral” the BTL model was fitted independently for each group for Exp. III. The goodness of fit of the different BTL models is indicated by their corresponding p-values (H_0: difference to ideal model is zero) of $p = 0.55$ for the group “dissatisfied” and $p = 0.26$ for the group “neutral.” The results are compared in Fig. 9 to the average results without grouping from Exp. I. The group that reported to be dissatisfied with the placement of the lead vocals to the side rated similar to the average results of participants of Exp. I and showed low preferences for the two wide WFS mixes. The group that did not mention anything related to position in the questionnaire preferred the very wide mix over all other mixes, in strong contrast to the participants of Exp. I.

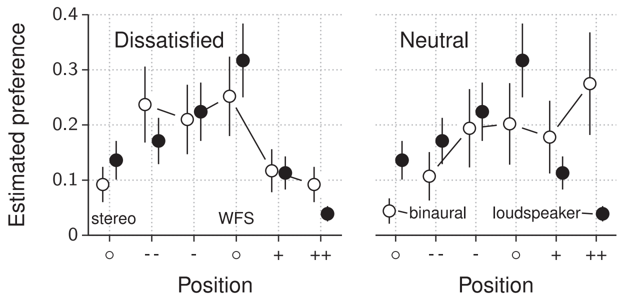

Fig. 9. Listener preferences for stereo, different WFS mixes with changes in the mix parameter position. The estimated preference is shown as the probability for each condition to be preferred, together with its 95% confidence interval. The filled circles show the results from Exp. I employing loudspeakers, the open circles the results from Exp. III employing binaural simulation of the loudspeakers, and presented for the two user groups “dissatisfied” and “neutral.”

4 DISCUSSION

The results show that listeners prefer audio reproduction systems with more loudspeakers. This is supported by all three experiments. For the song used in Exp. I and Exp. III the differences between the reproduction systems are more important than changes to the actual mixes presented by the preferred WFS reproduction system. The average listener preference across the six song excerpts of Exp. II favored WFS over surround and surround over stereo.

In Exp. III none and in Exp. I only two of the 14 available WFS mixes were preferred less than the stereo reference mix, even though the changes to the WFS mixes along a specific parameter were relatively drastic. Still, the listeners were not indifferent to changes in the WFS mix. The highest probability to be preferred was on average 0.28 for the WFS reference mix out of five conditions including stereo. The next highest average probability was found for the mixes “–” with 0.23, followed by “off” with 0.20, “+” with 0.18 and stereo with 0.12, showing a clear influence of the changes to the WFS mix on the BTL ratio scale.

Even if the reproduction systems with a higher number of loudspeakers are preferred, Exp. II also showed a strong dependency on the content. For the six song excerpts presented in Exp. II, only two led to a preferred reproduction by WFS. For two others the listeners were indifferent regarding the reproduction system, and for the remaining two surround was preferred as much or even more than WFS. For these four song excerpts it is very likely that changes to the WFS mix would have resulted in a different result than the one found in Exp. I and Exp. III for the song “Lighthouse.” One explanation for those differences might be that for some songs the WFS inherent comb-filter-like artifacts due to the spatial aliasing [7] lead to a pronounced coloration, impossible to hide in the mixing process of the song. This would then have counterbalanced the higher spaciousness of the generated mix.

Another explanation may be that some of the choices regarding spaciousness or other parameters of the mix were perceived as less suitable for the WFS mix for particular songs. This was confirmed by informal listening to the audio files with the different systems. In general, the sound sources perceived in WFS seem easier to localize and sharper. The overall impression tends to be crisper and more direct in WFS. For different contents, the combination of mix and playout has led to quite different characters of the resulting sound impression. An example for such effects is the interplay between the crispness of percussion and drum elements that may interfere with a higher degree of spaciousness. This may be perceived as less suitable for certain types of music as observed for the rap song “Toynbee Suite,” where a softer and fuller percussion sound might be desired for this song. Other cases that prove the interplay between the mixing choice and the reproduction method in terms of the perception of certain scene feature are cases with a high preference for WFS. An example is the interplay between the basic character of a singer’s voice and the effect that compression may have. In the WFS mix of the song “Lighthouse” the singer was rendered as a focused source and was perceived to sing with more passion and urgency—although the underlying recording remained the same. This aspect will have to be investigated in more detail and more formally in the future, possibly using multidimensional analysis approaches similar to Choisel and Wickelmaier [13].

For instrumental modeling of the data from Exp. I and Exp. II it is important to get the binaural signals for the different conditions as most auditory models have the two ear signals as inputs. The results of Exp. III show that listeners rated the different reproduction systems in a very similar way using binaural simulations of the real loudspeakers. Only the mixing parameter position showed a strong disagreement between the ratings of both experiments. Here, the binaural simulation also led to two different groups of listeners regarding the preferred mix of the arrangement. The disagreement of both groups for position might be explained by the individual differences and pronounced problems binaural simulations show with externalization [48]. Most probably the binaural simulation changed the spatial appearance of the whole scene for some listeners.

5 CONCLUSION

This study investigated listener preference by comparing two-channel with five-channel stereophony and wave field synthesis, using a circular array of 56 loudspeakers. The results show a preference by listeners for wave field synthesis over stereophonic reproduction for popular music. This preference is dependent on the actual content and might vary between different songs or even song excerpts.

The wave field synthesis system requires its own mix, which introduces an influence of the mixing process on the comparison of the different reproduction systems. Although introducing relatively strong changes to the wave field synthesis mix, listeners still preferred that system most of the times. The only condition disliked by listeners was a very wide arrangement of the foreground elements of popular music like vocals, snare and bass drum, and guitars. This highlights that the main advantage of higher spaciousness comes to play for background and ambient parts of a song.

For auditory modeling of listener preference binaural simulations of the loudspeaker setups are desirable. This study found binaural simulations to provide very similar results in a listening test, strong differences only appeared when presenting changes to the spatial arrangement of the scene.

6 ACKNOWLEDGMENT

The authors would like to thank Russel Mason and two anonymous reviewers for their very helpful comments. This research has been supported by EU FET grant TWO!EARS, ICT-618075.

7 REFERENCES

[1] ITU-R BS775-3, “Multichannel stereophonic sound system with and without accompanying picture” (2012).

[2] S. Spors, H. Wierstorf, A. Raake, F. Melchior, M. Frank, and F. Zotter, “Spatial Sound with Loudspeakers and Its Perception: A Review of the Current State,” *Proc. IEEE*, vol. 101, no. 9, pp. 1920–1938 (2013), https://doi.org/10.1109/JPROC.20e3.2264784

[3] D. Västfjäll, “The Subjective Sense of Presence, Emotion Recognition, and Experienced Emotions in Auditory Virtual Environments,” *Cyberpsychol. Behav.*, vol. 6, no. 2, pp. 181–188 (2003), https://doi.org/10.1089/109493103321640374.

[4] J. Francombe, T. Brookes, R. Mason, and J. Woodcock, “Evaluation of Spatial Audio Reproduction Methods (Part 2): Analysis of Listener Preference,” *J. Audio Eng. Soc.*, vol. 65, pp. 212–225 (2017 Mar.), https://doi.org/10.17743/jaes.2016.0071.

[5] J. C. Steinberg and W. B. Snow, “Symposium on Wire Transmission of Symphonic Music and its Reproduction in Auditory Perspective: Physical Factors,” *Bell Syst. Tech. J.*, vol. 13, no. 2, pp. 245–258 (1934).

[6] A. J. Berkhout, “A Holographic Approach to Acoustic Control,” *J. Audio Eng. Soc.*, vol. 36, pp. 977–995 (1988 Dec.).

[7] H. Wierstorf, C. Hohnerlein, S. Spors, and A. Raake, “Coloration in Wave Field Synthesis,” presented at the *AES 55th International Conference: Spatial Audio* (2014 Aug.), conference paper 5–3.

[8] U. Horbach, A. Karamustafaoglu, R. Pellegrini, P. Mackensen, and G. Theile, “Design and Applications of a Data-Based Auralization System for Surround Sound,” presented at the *106th Convention of the Audio Engineering Society* (1999 May), convention paper 4976.

[9] H. Wierstorf, A. Raake, and S. Spors, “Assessing Localization Accuracy in Sound Field Synthesis,” *J. Acoust. Soc. Am.*, vol. 141, no. 2, pp. 1111–1119 (2017), https://doi.org/10.1121/1.4795780.

[10] T. Nakayama, O. Kosaka, M. Okamoto, and T. Shiga, “Subjective Assessment of Multichannel Reproduction,” *J. Audio Eng. Soc.*, vol. 19, pp. 744–751 (1971 Oct.).

[11] T. Letowski, "Sound Quality Assessment: Concepts and Criteria," presented at the *87th Convention of the Audio Engineering Society* (1989 Oct.), convention paper 2825.

[12] F. Rumsey, S. Zielinski, R. Kassier, and S. Bech, "On the Relative Importance of Spatial and Timbral Fidelities in Judgments of Degraded Multichannel Audio Quality," *J. Acoust. Soc. Am.*, vol. 118, no. 2, pp. 968–976 (2005), https://doi.org/10.1121/1.1945368.

[13] S. Choisel and F. Wickelmaier, "Evaluation of Multichannel Reproduced Sound: Scaling Auditory Attributes Underlying Listener Preference," *J. Acoust. Soc. Am.*, vol. 121, no. 1, pp. 388–400 (2007), https://doi.org/10.1121/1.2385043.

[14] N. Zacharov and K. Koivuniemi, "Audio Descriptive Analysis & Mapping of Spatial Sound Displays," presented at the *Int. Conf. Auditory Display*, pp. 95–104 (2001).

[15] C. Guastavino and B. F. G. Katz, "Perceptual Evaluation of Multi-Dimensional Spatial Audio reproduction," *J. Acoust. Soc. Am.*, vol. 116, no. 2, pp. 1105–1115 (2004).

[16] E. W. Start, *Direct Sound enhancement by Wave Field Synthesis*, Ph.D. thesis, Technische Universiteit Delft (1997).

[17] E. Verheijen, *Sound Reproduction by Wave Field Synthesis*, Ph.D. thesis, Technische Universiteit Delft (1997).

[18] R. Mason, "How Important Is Accurate Localization in Reproduced Sound?" presented at the *142nd Convention of the Audio Engineering Society* (2017 May), convention paper 9759.

[19] H. Wittek, F. Rumsey, and G. Theile, "On the Sound Color Properties of Wavefield Synthesis and Stereo," presented at the *123rd Convention of the Audio Engineering Society* (2007 Oct.), convention paper 7167.

[20] M. Boone and W. de Bruijn, "Improving Speech Intelligibility in Teleconferencing by Using Wave Field Synthesis," presented at the *114th Convention of the Audio Engineering Society* (2003 Mar.), convention paper 5800.

[21] J. Berg and F. Rumsey, "Systematic Evaluation of Perceived Spatial Quality," presented at the *AES 24th International Conference: Multichannel Audio: The New Reality* (2003 Jun.), conference paper 43.

[22] A. Wilson and B. M. Fazenda, "Perception of Audio Quality in Productions of Popular Music," *J. Audio Eng. Soc.*, vol. 64, pp. 23–34 (2016 Jan./Feb.), https://doi.org/10.17743/jaes.2015.0090.

[23] E. Vickers, "The Loudness War," *J. Audio Eng. Soc.*, vol. 59, pp. 346–351 (2011 May).

[24] D. Giannoulis, M. Massberg, and J. D. Reiss, "Parameter Automation in a Dynamic Range Compressor," *J. Audio Eng. Soc.*, vol. 61, pp. 716–726 (2013 Oct.).

[25] N. Croghan, K. Arehart, and J. Kates, "Quality and Loudness Judgments for Music Subjected to Compression Limiting," *J. Acoust. Soc. Am.*, vol. 132, no. 2, pp. 1177–1188 (2012), https://doi.org/10.1121/1.4730881.

[26] A. Wilson and B. M. Fazenda, "101 Mixes: A Statistical Analysis of Mix-Variation in a Dataset of Multitrack Music Mixes," presented at the *139th Convention of the Audio Engineering Society* (2015 Oct.), convention paper 9398.

[27] C. V. Fung, "Musicians' and Nonmusicians' Preferences for World Musics: Relation to Musical Characteristics and Familiarity," *J. Res. Music Educ.*, vol. 44, no. 1, pp. 60–83 (1996), https://doi.org/10.2307/3345414.

[28] B. De Man, K. McNally, and J. D. Reiss, "Perceptual Evaluation and Analysis of Reverberation in Multitrack Music Production," *J. Audio Eng. Soc.*, vol. 65, pp. 108–116 (2017 Jan./Feb.), https://doi.org/10.17743/jaes.2016.0062.

[29] P. Pestana and J. D. Reiss, "Intelligent Audio Production Strategies Informed by Best Practices," presented at the *AES 53rd International Conference: Semantic Audio* (2014 Jan.), conference paper S2-2.

[30] B. Leonard, R. King, and G. Sikora, "The Effect of Acoustic Environment on Reverberation Level Preference," presented at the *133rd Convention of the Audio Engineering Society* (2012 Oct.), convention paper 8742.

[31] S. Mansbridge, S. Finn, and J. D. Reiss, "An Autonomous System for Multitrack Stereo Pan Positioning," presented at the *133rd Convention of the Audio Engineering Society* (2012 Oct.), convention paper 8763.

[32] J. Skowronek, L. Nagel, C. Hold, H. Wierstorf, and A. Raake, "Towards the Development of Preference Models accounting for the Impact of Music Production Techniques," presented at the *43rd Ger. Ann. Conf. Acoust. (DAGA)*, pp. 856–860 (2017).

[33] H. Wierstorf, S. Spors, and A. Raake, "Perception and Evaluation of Sound Fields," presented at the *59th Open Sem. Acoust.*, pp. 263–268 (2012).

[34] C. Hold and H. Wierstorf, "Signal Feeds for Creating the Music Mixes for Comparison of Wave Field Synthesis, Surround, and Stereo" (2016 Jun.), https://doi.org/10.5281/zenodo.55718.

[35] R. Bittner, J. Salamon, M. Tierney, M. Mauch, C. Cannam, and J. Bello, "MedleyDB: A Multitrack Dataset for Annotation-Intensive MIR Research," presented at the *15th Int. Soc. Mus. Inform. Retrieval Conf.* (2014).

[36] C. Hold and H. Wierstorf, "Object-Based Audio Scene Files for Variations of the Spatial Arrangement in Pop Mixes for Wave Field Synthesis" (2016 Aug.), https://doi.org/10.5281/zenodo.61110.

[37] M. Geier, J. Ahrens, and S. Spors, "Object-Based Audio Reproduction and the Audio Scene Description Format," *Organised Sound*, vol. 15, no. 3, pp. 219–227 (2010), https://doi.org/10.1017/S1355771810000324.

[38] C. Hold, L. Nagel, A. Raake, and H. Wierstorf, "Variations of Pop Mixes for Wave Field Synthesis" (2016 Aug.), https://doi.org/10.5281/zenodo.61000.

[39] M. Geier, J. Ahrens, and S. Spors, "The SoundScape Renderer: A Unified Spatial Audio Reproduction Framework for Arbitrary Rendering Methods," presented at the *124th Convention of the Audio Engineering Society* (2008 May), convention paper 7330.

[40] C. Hold, H. Wierstorf, and A. Raake, "The Difference between Stereophony and Wave Field Synthesis in the Context of Popular Music," presented at the *140th Convention of the Audio Engineering Society* (2016 May), convention paper 9533.

[41] H. Wierstorf, "Binaural Room Scanning Files for a 56-Channel Circular Loudspeaker Array" (2016 Jun.), https://doi.org/10.5281/zenodo.55572.

[42] H. Wierstorf, M. Geier, and S. Spors, "A Free Database of Head Related Impulse Response Measurements in the Horizontal Plane with Multiple Distances," presented at the *130th Convention of the Audio Engineering Society* (2011 May), eBrief 6.

[43] E. Zwicker and H. Fastl, *Psycho-Acoustics—Facts and Models* , 2nd ed. (Springer, Berlin, 1999).

[44] M. G. Kendall and B. B. Smith, "On the Method of Paired Comparisons," *Biometrika*, vol. 34, no. Pt 3-4, pp. 324–345 (1947), https://doi.org/10.2307/2332613.

[45] R. A. Bradley, M. E. Terry, "Rank Analysis of Incomplete Block Designs: The Method of Paired Comparisons," *Biometrika*, vol. 39, no. 3/4, pp. 324–345 (1952), https://doi.org/10.2307/2334029.

[46] F. Wickelmaier and C. Schmid, "A Matlab Function to Estimate Choice Model Parameters from Paired-Comparison Data," *Behav. Res. Meth. Ins. C.*, vol. 36, no. 1, pp. 29–40 (2004), https://doi.org/10.3758/BF03195547.

[47] H. Wierstorf, C. Hold, and A. Raake, "Code to Reproduce the Figures in the Paper 'Listener Preference for Different Reproduction Systems and Mixes in Popular Music'" (2017 Nov.), https://doi.org/10.5281/zenodo.1054525.

[48] K. Brandenburg, S. Werner, F. Klein, and C. Sladeczek, "Auditory Illusion through Headphones: History, Challenges and New solutions," presented at the *22nd Int. C. Acoust.*, pp. 3063–3072 (2016).

THE AUTHORS

Hagen Wierstorf

Christoph Hold

Alexander Raake

Hagen Wierstorf is a Research Fellow at the Centre for Vision, Speech and Signal Processing (CVSSP) at the University of Surrey since 2017. He studied physics at the Carl von Ossietzky Universität Oldenburg, where he received a master equivalent degree with a thesis in the field of psychoacoustics and neuroscience in 2008. He received a doctoral degree from Technische Universität Berlin on the topic of perceptual assessment of sound field synthesis systems in 2014 and continued as a Postdoctoral Researcher. In 2016 he moved as a Postdoctoral Researcher to Technische Universität Ilmenau and was a Visiting Professor at Filmuniversität Babelsberg KONRAD WOLF.

•

Christoph Hold studies M.Sc. in audio communication technology and holds a B.Sc. in electrical engineering from the Technische Universität Berlin, where he specializes in signal processing and virtual acoustics. He is interested in high quality audio and its perception. As a mix engineer and musician, he is concerned about the underlying technology—from recording and processing music to the final listening experience. From 2015 to 2017 he was a research assistant at TU Berlin, followed by a research internship at Microsoft Research in Redmond, WA, USA. For the Audio Engineering Society, he is the current Chair of the Berlin Student Section and was part of the 142nd AES Convention committee.

•

Alexander Raake has joined TU Ilmenau in 2015 as a full Professor, where he heads the Audiovisual Technology Group. Between 2005 and 2015 he held Senior Researcher, Assistant and later Associate Professor positions at TU Berlin's An-Institut T-Labs, a joint venture between Deutsche Telekom AG and TU Berlin, heading the Assessment of IP-based Applications group. From 2004 to 2005, he was a Postdoctoral Researcher at LIMSI-CNRS in Orsay, France. From the Electrical Engineering and Information Technology Faculty of the Ruhr-Universität Bochum he obtained his doctoral degree (Dr.-Ing.) in early 2005, with the book *Speech Quality of VoIP*. His research interests are in audiovisual and multimedia technology, speech, audio and video signals, human audiovisual perception, and Quality of Experience. Since 1999 he has been involved in ITU-T Study Group 12's standardization work on QoS and QoE assessment methods. He is a member of the IEEE, the Acoustical Society of America, the AES, VDE/ITG, and DEGA.