Tim Ziemer

Perceptual sound field synthesis concept for music presentation

Veröffentlicht in:

Proceedings of Meetings on Acoustics, Volume 30, Issue 1
Verlag: AIP Publishing | Jahr: 2017

DOI: https://doi.org/10.1121/2.0000661

Perceptual sound field synthesis concept for music presentation

Tim Ziemer

Citation: Proc. Mtgs. Acoust. **30**, 015016 (2017); doi: 10.1121/2.0000661
View online: https://doi.org/10.1121/2.0000661
View Table of Contents: https://asa.scitation.org/toc/pma/30/1
Published by the Acoustical Society of America

Proceedings of Meetings on Acoustics

Volume 30 http://acousticalsociety.org/

Acoustics `17 Boston

173rd Meeting of Acoustical Society of America and 8th Forum Acusticum
Boston, Massachusetts
25-29 June 2017

Architectural Acoustics: Paper 5pAAb5

Perceptual sound field synthesis concept for music presentation

Tim Ziemer
Institute of Systematic Musicology, University of Hamburg, Hamburg, 20354, GERMANY; tim.ziemer@uni-hamburg.de

A perceptual sound field synthesis approach for music is presented. Its signal processing implements critical bands, the precedence effect and integration times of the auditory system by technical means, as well as the radiation characteristics of musical instruments. Furthermore, interaural coherence, masking and auditory scene analysis principles are considered. As a result, the conceptualized sound field synthesis system creates a natural, spatial sound impression for listeners in extended listening area, even with a low number of loudspeakers. A novel technique, the "precedence fade", as well as the interaural cues provided by the sound field synthesis approach, allow for a precise and robust localization. Simulations and a listening test provide a proof of concept. The method is particularly robust for signals with impulsive attacks and long quasi-stationary phases, as in the case of many instrumental sounds. It is compatible with many loudspeaker setups, such as 5.1 to 22.2, ambisonics systems and loudspeaker arrays for wave front synthesis. The perceptual sound field synthesis approach is an alternative to physically centered wave field synthesis concepts and conventional, perceptually motivated stereophonic sound and benefits from both paradigms.

Published by the Acoustical Society of America

 https://doi.org/10.1121/2.0000661

1. INTRODUCTION

The physical features of musical instruments have been studied for centuries. Many instrumental sounds share certain features, which dissociate them from other sounds such as spoken language or industrial noise. Not all physical aspects of instrumental sounds are perceived by the listener. Perceived features are affected by absolute thresholds, differential thresholds, and integration times. While conventional audio systems can create a sound that is perceived as equivalent to original musical instruments in many ways, sound field synthesis systems tend to aim at a perfect physical copy of an original sound field. This would make perceptual considerations superfluous for the technical realization of music presentation[1].

To date a perfect physical copy of a sound field is hardly realizable by technical means. Therefore, a compromise is suggested in this paper, taking commonalities of musical instrumental sounds as well as auditory perception into account. The created sound field only partly resembles an original sound field but still the sound is perceived as a wide, natural musical instrument which can be localized well. The paper describes the perceptual and musical aspects. For technical details, simulations and experimental evaluations, refer to the cited literature.

2. BACKGROUND

To understand the proposed perceptual sound field synthesis approach, a background in three fields is necessary. The basics are outlines in the following. First, an overview about the physical features of musical instrumental sounds is given. Then, aspects of auditory perception in terms of psychoacoustics and auditory scene analysis are discussed, followed by a basic treatise of sound field synthesis.

A. MUSICAL INSTRUMENTS

According to [1] the main features of musical instruments include musical scale, dynamics, timbre, time envelope and sound radiation characteristics. The structure of instrumental sounds is presented in [2–5]. Notes often have an initial transient with a high spectral density and a large bandwidth. The spectral energy may be induced by impulsive plucking or hammering, or by continuous bowing or blowing. The high density and bandwidth can be explained by the impulsive nature of musical instrument sound generation, like the string knocking on a bridge which then deflects a top plate or a sound board, or the lips, labia, or reeds that open only shortly every period. In fact, all sound generation of musical instruments can be traced back to impulses, referred to as *impulse pattern formulation* [3]. So the initial transient, or attack, is more or less impulsive. Many frequencies rapidly decay, due to radiation or internal damping, whereas other frequencies sustain for several periods. This sustained period is referred to as *quasi-steady* or *quasi-stationary* part of the instrumental sound. This phase can last for a couple of periods only, as in many percussive instruments. Or it can last for several seconds, like in many blown and bowed instruments. Many instruments exhibit a homogeneous timbre over a wide compass.

Usually, the sound created with the musical instrument is not radiated homogeneously in all directions. An overview about the rough radiation characteristics of musical instruments is given in [2, 6–8]. Musical instruments emit different spectra into each direction and are more or less directive. In many instruments, the radiation characteristic of a frequency depends on the used material, fundamental frequency, excited string, or the playing technique [3,4,8,9]. Recent measurements with large-scale circular microphone arrays reveal that even small musical instruments can create rather incoherent signals at listeners' ears at all angles and distances up to at least 3 m [9–13].

[1] Of course, perceptual considerations may still be useful for the art of composition and performance practice.

B. MUSIC PERCEPTION

Music perception is rather complex. Perceived aspects of instrumental sounds include, e.g., pitch, loudness, spatial location and extent, and timbre, and their variation over time. They are subject of psychoacoustic research [14] and subjective room acoustics [15]. These aspects and additional auditory qualities, like chords, rhythm and melody, are a matter of the psychologic organization principles, referred to as *auditory scene analysis* [16]. This section is mainly based on these extensive books and the literature that they cite. All aspects of auditory perception are affected by several acoustic features and occur due to thresholds and limitations, integration times and just noticeable differences of the auditory system.

Loudness is affected mostly by the sound pressure level and the spectral distribution, density and bandwidth. Furthermore, it is a matter of duration. The loudness of noise bursts with equal spectrum and sound pressure level increases with increasing duration until about 100 ms; a natural integration time of the auditory system. In room acoustics, the *early strength* describes the perceived loudness of a room. The measure integrates the first 80 ms of the squared room impulse response and divides it by a squared anechoic reference impulse. For speech, 50 ms tend to be applied.

Pitch and **pitch strength** are mostly a matter of frequency and periodicity. The lowest audible frequency lies around 20 Hz, i.e., a period of 50 ms. This is also the minimum duration it takes to detect the pitch of pure tones with fair strength. It is assumed that this natural integration time of the auditory system is the reason for the fact that the initial transients of most musical instruments are shorter than 50 ms [3]. Some researchers think that the auditory system uses a time window of about 16 to 20 ms to detect pitch [17]. However, it may take up to 300 ms to detect the pitch of sounds, especially if they are inharmonic or noisy. Psychoacoustic models of noisiness tend to be based on loudness models and assume the same integration times as for loudness [18, 19], i.e., about 100 ms.

Perceived **sharpness** is assumed to be the major contributor to timbre perception. Although being related to the attack of instrumental sounds, it is not a matter of temporal and spectral fine structure but rather the centroid of excitation on basilar membrane, i.e., the mean value of partial loudnesses. Consequently, it can be assumed that the integration time is the same as for loudness perception [18, 19]. Closely related to sharpness is the perception of brightness.

Roughness is another contributor to the perception of timbre. Two slightly different frequencies create regular envelope fluctuations which are perceived as *loudness fluctuations* or *beating*. At a rate between 15 and 300 Hz, i.e., periods between around 3 and 67 ms, these fluctuations sound rough. Psychoacoustic models of roughness assume the ear to integrate sound over 100 to 200 ms to sense roughness [18–20].

Timbre perception is a rather complex. It is often assumed to consist of three psychological dimensions. Sound signals need to have a duration of at least 2 to 5 ms so that the auditory system can derive a specific timbre. One dimension of timbre, besides sharpness or brightness, seems to be closely related with the synchrony of frequencies during the initial transient. These have an effect on the physical steepness and signal envelope and affect the perceived *punch* or *bite* of the attack. This dimension helps, e.g., to differentiate between metallic and wooden sounds, even when they contain a similar spectrum in the quasi-stationary part. [3]

The auditory system does not process sound as a whole but rather divides the spectrum into about 25 **critical bands** [1, 21]. Frequencies that simultaneously fall into the same critical band cannot be discriminated. Either the loudest frequency masks the other frequencies within that critical band or all frequencies create a common contribution to the perception of pitch, loudness, sharpness, beating and roughness.

Simultaneous masking refers to frequencies being inaudible due to the presence of other frequencies with higher amplitude, the so-called *maskers*. The loudest partial within a critical band tends to mask the others, lower bands tend to mask frequencies in higher bands. Depending on the distribution of maskers, frequencies are either detected vaguely by the auditory system or not at all. During the masker onset, the masking threshold is up to 26 dB higher and then decreases towards a steady level after 50 ms [21]. This

effect is referred to as **overshoot phenomenon**.

Pre-masking is the phenomenon that frequencies can even mask other frequencies that had arrived earlier. This is due to the fact that the auditory system takes more time to process signals with low amplitudes compared to signals with high amplitudes. Low-level signals can be masked even when they arrive between 10 and 50 ms before the masker [2, 21].

Source localization is a matter of monaural and binaural cues that the auditory system derives from the head-related transfer function (HRTF), i.e., the change of sound due to the spatio-temporal transfer from the source to the ears. Main binaural cues are interaural time- and level difference (ITD and ILD). Here, ITD between 27 and 640 μs are evaluated by the auditory system to derive lateral source deflection. Note that this is a much lower order of magnitude compared to all other aspects of sound perception. [22]

When multiple wavefronts arrive, the first wavefront tends to dominate source localization. This phenomenon is referred to as **precedence effect** and is even active when signal delays arriving within 5 to 30 ms are 10 dB louder [21].

Apparent source width (ASW) is widely believed to depend on the coherence of ear signals. Wavefronts that arrive within 50 to 80 ms are integrated for the perception of width [13, 15]. Although interaural coherence exhibits strong fluctuations in a room, ASW tends to remain constant, i.e. the auditory system seems to compensate for these fluctuations to keep up the width for an auditory stream.

Auditory qualities, like pitch, timbre, rhythm and melody are no physical properties but result from organization principles referred to as **auditory scene analysis** [16]. Simultaneous sounds are integrated into auditory streams which are segregated from one another, referred to as *simultaneous grouping*. These streams may stay grouped over time or be reorganized, referred to as *sequential grouping*. Elements are grouped, when they share properties. For simultaneous grouping, the principles of harmonicity, synchrony and spatial location are dominant. Synchronous frequencies that share a common fundamental and seem to arrive from the same spatial location tend to be integrated into one auditory stream. Here, not all three conditions need to be fulfilled. When the three principles suggest contradictory grouping, the auditory system makes a guess which is based on the three principles, on sequential grouping principles and comparison with other senses. For example, if frequencies seem to arrive from different spatial locations but share a common fundamental frequency and arrive in synchrony, they are likely to be perceived as one auditory stream, receiving a common group location. This is especially true if the sounds follow the sequential grouping principles of *continuity*, *proximity*, *common fate*, *timbre* and *closure*. Melodies, for example, are usually considered as succession of pitches which is grouped into the same sequential auditory stream and lasts for the duration of short-term memory, i.e., 2 to 5 seconds [3]. Established auditory streams can last for seconds and more [16]. Sounds that are integrated into an auditory stream may lose some of their original salience [16], p. 140.

Thorough consideration and technical implementation of the above-mentioned psychoacoustic and auditory scene analysis principles has led to successful audio compression algorithms [23], music recommendation systems [18], auditory displays [24, 25], synthesizers [26], sound source localization [27] and source width detection [13] applications and is the basis of the perceptual sound field synthesis concept for music presentation.

C. SOUND FIELD SYNTHESIS

The term *sound field synthesis* comprises a number of methods that are implemented in audio systems. They have in common that their objective is to synthesize the desired sound pressure distribution within an extended listening area. An overview about the mathematics, physics and some perceptual aspects of different approaches can be found in [28, 29]. Early formulations of wave field synthesis in terms of *wave front synthesis* are given in [30, 31] and are revisited in [32]. Typically, the Kirchhoff-Helmholtz integral is considered as the core element of wave field synthesis. It describes the relationship between sound pressure

and sound pressure gradient on a volume surface S and the sound field within that source-free volume $P\left(\omega, \vec{X}\right)$, if there was a source outside the source-free volume by including the sound propagation function $G\left(\omega, \Delta\vec{r}\right)$, a Greens' function. However, for a technical implementation, the integral is simplified by many assumption and results in the discrete Rayleigh-Integral

$$P\left(\omega, \vec{X}\right) = \frac{1}{2\pi} \sum_{\vec{r}_Y=-\infty}^{\infty} \left(G\left(\omega, \Delta\vec{r}\right) \frac{\partial P\left(\omega, \vec{Y}\right)}{\partial \vec{n}} \right) \Delta\vec{r}_Y \ . \tag{1}$$

The Rayleigh-Integral describes the relationship between the sound pressure on an infinite separation plane between source- and source-free volume and the sound field within the source-free half space by including the propagation function of a monopole source. Now, loudspeakers can act as monopole sources. These create the calculated sound field both within and outside the source-free volume. A number of compensation methods allows to

- Allow for a finite separation plane
- Install only loudspeakers at the height of the listeners' heads
- Surround a listening area from all sides
- Allow for virtual sources within the source-free volume

resulting, however, in synthesis errors and a smaller listening area. The compensation methods are not perceptually-motivated but engineering solutions that work in practice. Simply put, in wave field synthesis we propagate a source sound from a chosen virtual source location to the separation plane via Greens' function and feed the corresponding loudspeaker with the calculated, i.e., delayed and attenuated, signal. Then we apply the necessary compensation methods. This way, eq. 1 is a forward-problem. Modeling virtual sources as monopole or plane wave and synthesizing the spatio-temporal propagation of the wave fronts is certainly the most widespread wave field synthesis approach. An alternative is to synthesize the sound field that an idealized source would have created along a small circle or sphere around a central listening point, referred to as *ambisonics* [13, 33, 34]. Here, eq. 1 is an inverse problem; the desired sound field is known and the loudspeaker signals are to be found.

In addition to monopoles or plane waves, point sources of higher order have been implemented in wave front synthesis [35, 36] and in ambisonics [37] systems. However, it could be shown that complex point sources of order 64 are necessary to capture the incoherent signals that even small instruments may create at a distance of up to three meters [9,11]. Yet another approach is to synthesize a desired sound field at discrete listening points that sample an extended listening area [38–40]. This approach is the basis of the perceptual sound field synthesis approach for music presentation presented in this paper.

Research on the proposed sound field synthesis method has been carried out over eight years and can be followed in [9–11, 21, 29, 40–45, 51][2].

3. METHOD

The perceptually-motivated sound field synthesis approach for music presentation is based on the aforementioned fundamentals of instrumental sound, music perception and sound field synthesis.

[2]These publications are comprised in a ResearchGate project.

A. CAPTURING OF THE RADIATION CHARACTERISTICS

As described in Sec. 2.1 and 2.2, musical instruments exhibit complicated radiation patterns, which may create incoherent ear signals that affect perceived source width, even in absence of room acoustics. To account for this important aspect of music perception, the radiation characteristics of musical instruments are measured every 5 cm by means of a circular microphone array with a radius of 1 m. Here, a drastic simplification is made: The actual geometry of the instrument, and the presence of the instrumentalist are neglected. Instead, the source is considered as a complex point source in the center of the array, which has the ability to radiate its source sound $P(\omega, \mathbf{r}_Q)$ in a complex manner, i.e., amplitude and phase of each frequency are radiated individually towards each direction φ, according to a complex radiation factor $\Gamma(\omega, \varphi)$. According to this simplification, the spectra at the discrete microphone locations can be reconsidered as the product of one common source sound, modified by a complex transfer function for the distance Δr between source and receiver point $G(\omega, \Delta r)$, and a direction-dependent, complex radiation factor $\Gamma(\omega, \varphi)$, i.e.,

$$P(\omega, \mathbf{r}_{\text{mic}}) = P(\omega, \mathbf{r}_Q)\, G(\omega, \Delta r)\, \Gamma(\omega, \varphi) \ . \tag{2}$$

Since $\Gamma(\omega, \varphi)$ is the only direction-dependent variable, it fully describes the radiation pattern of the complex point source, and the measured spectra are directly proportional to it, i.e. $P(\omega, \mathbf{r}_{\text{mic}}) \propto \Gamma(\omega, \varphi)$. We thus call the measured spectra $\Gamma'(\omega, \varphi)$, which are sometimes referred to as *far field signature function* or *far field directivity pattern* [28, 46]. This principle is referred to as *complex point source* method and is described in detail in [11]. A similar simplification is implicitly made in other studies that analyze the radiation characteristics of musical instruments [2,6–8,47]. It could be shown that the complex point source model yields plausible results concerning interaural cues in the radiated sound field [9,11,13]. Furthermore, circular and spherical arrays are considered a natural choice when it comes to sound field recordings around a source [48], and complex point sources were found to yield more natural sound impressions in auralizations [47].

As a sparse, but perceptually meaningful representation of the radiation characteristic of a musical instrument, we store the radiation pattern of one salient frequency within each critical band, yielding 25 radiation patterns. This is achieved by transforming one second of quasi-stationary instrumental sound into frequency domain via Fourier transform.

B. SOUND FIELD EXTRAPOLATION

The propagation function $K(\omega, \Delta\mathbf{r})$, describing the radiation of a source spectrum $P(\omega, \mathbf{r}_Q)$ from a complex point source at $\mathbf{r}_Q$ to a listening point at $\mathbf{r}_X$ can be described by a function

$$K(\omega, \Delta\mathbf{r}) = G(\omega, \Delta r)\, \Gamma(\omega, \varphi_X) \tag{3}$$

comprising the free field Green's function for the distance Δr between source point and listening point

$$G(\omega, \Delta r) = \frac{e^{-\imath k \Delta r}}{\Delta r} \tag{4}$$

modified by the complex amplitude at each angle $\Gamma'(\omega, \varphi)$. Here, Δr is the distance between source and listening point, φ_X is the angle between the facing direction of the source and the listening angle, ω is the angular frequency, $\imath$ the imaginary unit, $k = \omega/c$ is the wave number and c the sound velocity. Along all measured angles the propagated sound field at a listening points $P(\omega, \mathbf{r}_X)$ can be calculated as

$$P(\omega, \mathbf{r}_X) = P(\omega, \mathbf{r}_Q)\, K(\omega, \Delta\mathbf{r}) \ . \tag{5}$$

Angles between the discrete microphone angles are approximated by linear interpolation. Here, the necessity of the high spatial resolution of the microphone array becomes clear: A listener facing the source can be

placed up to three meters away from the source before both ears lie between two measured angles. i.e. before both ear signals are a result of interpolation between the same pair of measured radiation factors. Further details on the measurement setup and the complex point source model to characterize sound radiation characteristics and to extrapolate the radiated sound can be found e.g. in [9, 11, 13, 21, 44].

Only the loudest frequency of each critical band is extrapolated by means of eq. 5, assuming that they tend to be most prominent and dominate source localization, timbre and width perception.

C. SYNTHESIS OF THE DESIRED SOUND FIELD

The desired sound field $P(\omega, \mathbf{r}_X)$ is supposed to be created by loudspeaker signals $P(\omega, \mathbf{r}_Y)$. Assuming the loudspeakers to be monopole sources, the loudspeaker signals to create the desired sound field can be calculated by solving the linear equation system

$$\mathbf{G}(\omega, \Delta r)\, P(\omega, \mathbf{r}_Y) = P(\omega, \mathbf{r}_X) \ . \tag{6}$$

Note that this formulation resembles the Rayleigh integral, Eq. 1. But instead of a half space, the solution is valid for a distribution of discrete listening points. These listening points sample a listening area. According to the Nyquist-Shannon theorem, the sound field is synthesized correctly within in the whole listening area, for all frequencies whose wave lengths is at least twice the distance between neighboring listening points. When as many listening points as loudspeakers available are chosen, the linear equation system has a unique solution. In wave field synthesis, Eq. 1, is a forward-problem. The sound pressure distribution along the separation plane is calculated by propagating a source sound from any virtual source position towards this plane, assuming a monopole source or a plane wave. The desired sound pressure distribution within the source-free half-space is a result of this forward-propagation. Even though it is formulated in frequency domain, the approach can be calculated in time domain by simple time shifting and gain adjustments, or with spatio-temporal considerations. So the approach even works for transient signals. In the present approach, Eq. 6 is an inverse-problem. The desired sound field is calculated and the loudspeaker driving signals to create this sound field are sought. So the linear equation system is ill-posed. More critical is the fact that the linear equation system can even be ill-conditioned. This is the case when listening points are close to one another. In this case, the contribution of one loudspeaker signal to these two listening points is almost the same. Likewise, two loudspeakers may have a similar contribution to the same listening point. In these situations, the linear equation system is ill-conditioned and the calculated loudspeaker signals may have extremely high amplitudes. Furthermore, the amplitudes may exhibit strong fluctuations when the source position or source sound alters only slightly.

Hence, a regularization method is needed for a relaxation of the linear equation system. Widely applied regularization methods for square matrices and over-determined systems, which describe the desired sound field at more locations than loudspeakers present in the setup, are least-squares methods, Tikhonov regularization, and truncated singular value decomposition. These are revisited in [49], together with an alternative approaches named *golden section search*. Another newly developed regularization method, the *radiation method*, is not mathematically-motivated but originates in physical assumptions. Here, like the musical instruments, the loudspeakers are considered as complex point sources, and their radiation characteristics are measured by means of the microphone array as described in sec. 3.1, yielding a complex transfer function $\mathbf{K}(\omega, \Delta\mathbf{r})$, where $\Delta\mathbf{r}$ describes the path between each loudspeaker and each listening point. This modifies Eq. 6 to

$$\mathbf{K}(\omega, \Delta\mathbf{r})\, P(\omega, \mathbf{r}_Y) = P(\omega, \mathbf{r}_X) \ . \tag{7}$$

So the radiation characteristics of loudspeakers, which tends to deviate slightly from a monopole at low frequencies and strongly at high frequencies, are implemented. This relaxes the linear equation system and approximates the actual physics better than assuming perfect monopoles. An extensive treatise of this topic,

including a scenario and the respecting condition numbers and loudspeaker amplitudes, can be found in [40]. The solution is a set of loudspeaker signals in frequency domain.

D. RENDERING

In the proposed sound field synthesis approach, time is eliminated. Strictly speaking, a Fourier transform is only valid for stationary signals. It is an integration over time, assuming that the signal has started at $t = -\infty$ and will last forever. In all other cases, errors occur, like leakage effects. So already the capturing of the radiation characteristics, Sec. 3.1, does not account for transients. And also in the sound field extrapolation, Sec. 3.2, we neglect time and do as if the propagated spectrum had started at $t = -\infty$ and will repeat forever. For this stationary case, the solution of Eq. 7 is valid. As discussed in Sec. 2.1, musical instruments tend to have quasi-stationary phases which can last for seconds. During most of this time, the sound field is synthesized correctly for the 25 considered frequencies. During transients, the sound field may contain errors. The actual loudspeaker driving signals in time domain are created by the following steps: 1. Dividing the source signal into overlapping time windows, 2. transforming these into frequency domain, 3. solving the linear equation system for the loudest frequency of each critical band, 4. taking the original source spectrum and replacing the complex amplitude of these 25 frequencies with the solutions from the linear equation system, individually for each loudspeaker, 5. performing an inverse Fourier transform, and 6. cross-fading the signals of successive time windows.

i. Precedence Fade

With the above-presented sound field synthesis approach, all loudspeaker signals start st the same time. As a consequence, several wave fronts reach each listening point from a different angle at a different point in time. For a strictly stationary signal this is not a problem. It has started an infinite time ago and will last forever. So it has always been there and no wavefront can be identified. This situation resembles a standing wave, like a vibrating membrane: each point can have its own signal amplitude and phase, but there seems to be no sound propagation direction. In music, of course, every signal is at best quasi-stationary, because it does have a starting point $t_0 \neq -\infty$. To create a wave front that resembles the one of the virtual source, the *precedence fade* is applied: only one loudspeaker, the so-called *precedence speaker*, plays all parts of the source sound, including note onsets. All other loudspeakers start muted at every note onset. Their calculated signal gradually fades in. The fading-duration is chosen long enough, that the first wave front emanating from the precedence speaker passes all listening points before the other loudspeaker signals are faded in completely. But the fading duration also has to be short enough so that the missing contribution of these loudspeakers to the overall loudness of the signal is inaudible and that these signals are not heard as echoes. This implies the maximum of 50 ms mentioned in Sec. 2.2 in the course of the overshoot phenomenon and loudness integration.

4. EXPERIMENTAL EVALUATION

Many experiments and simulations on the presented psychoacoustic sound field synthesis have been carried out [10, 21, 40, 42–45, 51]. Summarized in Fig. 1 are some results concerning the psychoacoustic sound field synthesis approach with and without the precedence fade as well as an application with the pure precedence fade. In the background the loudspeaker setup and lines every 10° can be seen. Plotted on this is the cumulative number of participants that marked the specific angle as source angle. The Q marks the virtual source location. Without the precedence fade the psychoacoustic sound field synthesis approach creates no localizable source. The pure precedence fade only creates a very vague localization cue. Only the combination can be localized well.

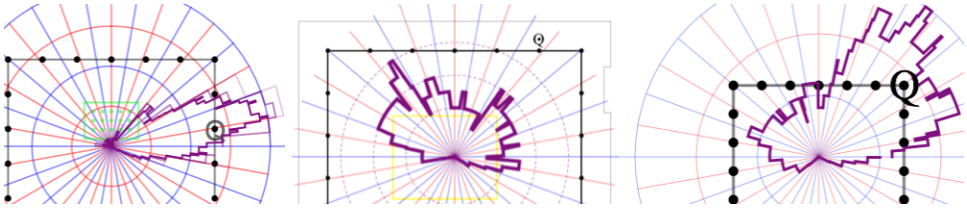

Figure 1: Source localization in Psychoacoustic Sound Field Synthesis including (left) and excluding (center) the precedence fade vs. a pure precedence fade (right).

A. PERCEPTUAL CONSIDERATIONS

As discussed in Sec. 2.2, loudness is barely affected, even if transients are not synthesized correctly, due to the integration time of about 100 ms. Neglecting transients in the sound field synthesis does not affect pitch perception. Pitch perception needs at least around 20 ms to build up and it can take 100 ms or more, until pitch strength reaches its maximum. So the contribution of attack transients to pitch perception is minor. The same is true for perceived roughness. During steady state, pitch and pitch strength are also unaffected by the signal processing due to the fact that the loudspeaker driving signals strongly resemble the original source signal. Only the amplitude and phase of 25 frequencies are altered. Consequently, the fundamental frequency as well as the distribution of overtones is the same as in the original source signal.

The attack transient is the main contributor to the perception of punch, an important timbre quality. An impulsive attack results from a high number of frequencies which start in phase. Many frequencies die out quickly. Filters can destroy the punch, especially when the phase of frequencies is manipulated. In the presented sound field synthesis approach, however, punch is barely affected. Only the phase of 25 frequencies is manipulated. The amplitude and phase of all other frequencies remains unaltered. During the quasi-stationary phase of many musical instruments, the sound may have between 10 and 80 frequencies, depending on the instrument, the playing technique and the fundamental frequency. Here, the manipulation of 25 frequency amplitudes has a stronger impact on the sound, compared to an attack, which may contain hundreds to thousands of frequencies.

The auditory system derives perceived sharpness from partial loudnesses along the Basilar membrane, so again, the signals transients play a minor role here. But signal processing during the stationary phase has a large effect on sharpness. By implementing the radiation characteristics of musical instruments, the spectral distribution is modified individually for each direction. As a consequence, the synthesized sound field may have a different sharpness at different points withing the listening region. The spectrum may be different at the two wars of a listener, and it may change while walking through the listening area. This change of spectral distribution over space is desired. The synthesized spectrum resembles the original spectral distribution radiated by musical instruments. It does affect the perceived naturalness, vividness and spaciousness in terms of width and depth. Even the best electric pianos do not sound as wide as a real grand piano, because they do not recreate the varies sound radiation characteristics that the complicated vibrations of the sound board and the other parts of the mechanical structure create.

Source localization has a temporal and a spectral component. Spectral cues are the interaural level and phase difference, as well as spectral peaks or notches that result from the head-related transfer function. The presented sound field synthesis approach synthesized the natural interaural cues of a musical instrument, including the effect of the sound radiation characteristics. Temporal cues for source localization are interaural time differences and, in the case of multiple wavefronts, the precedence effect. The presented approach creates interaural time differences that depend only on the locations of the loudspeakers and the positions of the listeners, not on the virtual source location. A multitude of wavefronts arrives the listeners within a short amount of time. The situation is comparable to the direct sound and early reflections in a reverberant

room. Consequently, the precedence fade dominates the temporal aspect of source localization.

Both the spectral synthesis as well as the precedence fade are necessary to recreate the desired source location. Each single one is not sufficient. This has been shown in [21,40,42]. The combination is necessary: The precedence fade roughly guides localization towards the location of the precedence speaker. Then, the spectral cues, created by the sound field synthesis approach, deliver the cues necessary for a precise localization, namely interaural amplitude and phase differences.

5. CONCLUSION

A perceptually-motivated sound field synthesis approach for music presentation has been presented. Considering common features of instrumental sounds and aspects of auditory perception allows to calculate and synthesize only a small portion of a desired sound field and still creating the desired sound impression: a natural, spatial sound which can be localized well by listeners within an extended listening area. This is achieved by applying the radiation method and the precedence fade. Like wave field synthesis and ambisonics, the approach can be traced-back to the Rayleigh-Integral. Such a sparse sound field synthesis raises the question how much physical detail is necessary to create a desired auditory scene for listeners.

6. PROSPECTS

In the presented approach not the wave fronts and their spatio-temporal propagation are synthesized. Instead, the sound field is synthesized in a sampled listening area during most of the quasi-stationary periods of instrumental sound. In conventional wave front synthesis approaches the wave field is sampled along one dimension, namely the distribution of active loudspeakers. In the perceptually and musically sound field synthesis approach the sound field is sampled within a two-dimensional listening area. Consequently, with the same number of active loudspeakers, the listening area tends to be much smaller. This is partly compensated by the fact that all loudspeakers can be active at all times, even when their location is opposed to the virtual source location. The same is true for many ambisonics approaches. Still, there may be a need to enlarge the listening area and allow for motion within the whole area, which is bounded by the loudspeakers. One possible solution is to combine the approach with motion capture technology to track an individual listener, as implemented in the wave field synthesis system at the HAW Hamburg [50]. This way, the listening area can move with the listener, who can even control the source location via gesture. Another approach is suggested in [51]. The listening points could be chosen individually for each frequency band. For lower frequencies, the listening points can have large distances without degrading the synthesis. This way, the listening area can have a large extent. It may be beneficial not to consider the loudest frequency of each critical frequency band but to concentrate on the very frequency region, which is dominant for source-localization. Interestingly, this excludes a region which is important for the perception of pitch. The presented sound field synthesis approach could be combined with other wave field synthesis approaches and act as an alternative to optimized phantom source imaging. Above the aliasing frequency of a wave front synthesis setup, the presented approach could create the desired sound field in a subspace whose size depends on the number of available loudspeakers. The psychoacoustic sound field synthesis could also be tested further. It would be interesting to see how well listeners are able to differentiate between different sound radiation characteristics. It should also be experimentally confirmed whether the approach also works for multiple listening areas.

ACKNOWLEDGMENTS

I would like thank Michael Vorländer, Stefan Weinzierl and Ning Xiang who invited me to present my work at the Acoustics ’17 conference as well as the German Academic Exchange Service (DAAD) for their

congress travel funding which enabled me to present my work. I also thank Luiz Naveda, Rolf Bader, Robert Mores and Georg Hajdu, who invited me to demonstrate the psychoacoustic sound field synthesis system and a predecessor at the International Conference of Students of Systematic Musicology (SysMus) 2009, the International Summer School in Systematic Musicology (ISSSM) 2012, the Fachausschuss Musikalische Akustik meeting of the German Acoustical Society (DEGA) 2015 and the Sound and Music Computing (SMC) Summer School in 2016. I thank Ralph Kessler who let me test the approach in his mastering studio for 5.1 and Auro-3D. I also thank Georg Hajdu, Wolfgang Fohl, Steffan Diedrichsen and the Apple Logic Team, Rolf Bader and Thomas Sporer for valuable discussions and ideas bout future work.

REFERENCES

[1] B. Kostek, *Perception-Based Data Processing in Acoustics*. Berlin, Heidelberg: Springer, 2005.

[2] J. Meyer, *Acoustics and the Performance of Music*, 5th ed. Bergkirchen: Springer, 2009.

[3] R. Bader, *Nonlinearities and Synchronization in Musical Acoustics and Music Psychology*. Berlin Heidelberg: Springer, 2013.

[4] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Intstruments*, New York: Springer, 2008.

[5] J. Backus, *The Acoustical Foundations of Music*. New York: W. W. Norton & Co., 1969.

[6] J. Pätynen and T. Lokki, "Directivities of symphony orchestra instruments", *Acta Acoust United Ac*, vol. 96, pp. 138–167, 2010.

[7] F. Zotter, "Analysis and synthesis of sound-radiation with spherical arrays", Ph.D. dissertation, University of Music and Performing Arts, Graz, 2009.

[8] F. Hohl and F. Zotter, "Similarity of musical instrument radiation-patterns in pitch and partial", in *Fortschritte der Akustik, DAGA, Berlin*, 2010.

[9] T. Ziemer, "Sound radiation characteristics of a shakuhachi with different playing techniques", in *Proceedings of the International Symposium on Musical Acoustics*, Le Mans, 2014, pp. 549–555.

[10] T. Ziemer, "Wave field synthesis. theory and application," (Magister Thesis), University of Hamburg, Hamburg 2011.

[11] T. Ziemer and R. Bader, "Complex point source model to calculate the sound field radiated from musical instruments", in *Proc Mtgs Acoust*, vol. 25, no. 1, Oct 2015.

[12] T. Ziemer, "Exploring physical parameters explaining the apparent source width of direct sound of musical instruments", in *DGM*, Oldenburg, Sep 2015, pp. 40–41.

[13] T. Ziemer, "Source width in music production. methods in stereo, ambisonics, and wave field synthesis", in *Studies in Musical Acoustics and Psychoacoustics*, A. Schneider, Ed. Cham: Springer, 2017, pp. 299–340.

[14] H. Fastl and E. Zwicker, *Psychoacoustics. Facts and Models*, third updated ed. Berlin, Heidelberg: Springer, 2007.

[15] L. L. Beranek, *Concert Halls and Opera Houses: Music, Acoustics, and Architecture*, 2nd ed. New York: Springer, 2004.

[16] A. S. Bregman, *Auditory Scene Analysis*. Massachusetts: MIT Press, 1990.

[17] E. Terhardt, "Calculating virtual pitch", *Hearing Research*, vol. 1, no. 2, pp. 155–182, 1979.

[18] T. Ziemer, Y. Yu, and S. Tang, "Using psychoacoustic models for sound analysis in music", in *Proc 8th Annual Meeting of the Forum on Information Retrieval Evaluation*, Kolkata, Dec 2016, pp. 1–7.

[19] W. Aures, "A model for calculating the sensory euphony of various sounds," *Acustica*, vol. 59, no. 2, pp. 130–141, 1985.

[20] P. Daniel and R. Weber, "Psychoacoustical roughness: Implementation of an optimized model," *Acta Acust united Ac*, vol. 83, no. 1, pp. 113–123, 1997.

[21] T. Ziemer, "Implementation of the radiation characteristics of musical instruments in wave field synthesis application", Ph.D. dissertation, University of Hamburg, Hamburg, July 2016.

[22] J. Blauert, *Spatial Hearing. The Psychophysics of Human Sound Source Localization*, revised ed. Cambridge , MA: MIT Press, 1997.

[23] K. Brandenburg and D. Seitzer, "Ocf: Coding high quality audio with data rates of 64 kbit/sec", in *Audio Eng Soc Convention 85*, Nov 1988.

[24] T. Ziemer, D. Black and H. Schultheis, "Psychoacoustic sonification design for navigation in surgical interventions", in *Proc Mtgs Acoust*, 2017.

[25] T. Ziemer and D. Black, "Psychoacoustically motivated sonification for surgeons," in *Proc. of the 31st International Congress and Exhibition on Computer Assisted Radiology and Surgery (CARS)*, no. (Suppl 1):1, Barcelona, Jun 2017, pp. 265–266.

[26] D. Wessel, "Timbre space as a musical control structure", *Computer Music Journal*, vol. 3, no. 2, pp. 45–52, 1979.

[27] S. Mattes, P. Nelson, F. Fazi, and M. Capp, "Exploration of a biologically inspired model for sound source localization in 3d space", in *Proc EAA Joint Symposium Auralization Ambisonics*. Berlin, Mar 2014, pp. 93–99.

[28] J. Ahrens, *Analytic Methods of Sound Field Synthesis*. Berlin, Heidelberg: Springer, 2012.

[29] T. Ziemer, "Wave field synthesis," in *Springer Handbook of Systematic Musicology*, R. Bader, Ed., Berlin Heidelberg, 2017, ch. 18, pp. 175–193.

[30] A. J. Berkhout, "A holographic approach to acoustic control", *J Audio Eng Soc*, vol. 36, no. 12, pp. 977–995, 1988.

[31] A. J. Berkhout, Diemer de Vries, and P. Vogel, "Wave front synthesis: A new direction in electroacoustics", in *Audio Eng Soc Convention 93*, 10 1992.

[32] S. Spors, R. Rabenstein, and J. Ahrens, "The theory of wave field synthesis revisited", in *Audio Eng Soc Convention 124*, 5 2008.

[33] J. Daniel,"˜hrefhttp://www.aes.org/e-lib/browse.cfm?elib=12321Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format", in *23rd International Audio Eng Soc Conference: Signal Processing in Audio Recording and Reproduction*, Copenhagen, May 2003.

[34] J. Daniel, R. Nicol, and S. Moreau, "Further investigations of high order ambisonics and wavefield synthesis for holophonic sound imaging", in *Audio Eng Soc Convention 114*, 3 2003.

[35] E. Corteel, "Synthesis of directional sources using wave field synthesis, possibilities, and limitations", *EURASIP Journal on Advances in Signal Processing*, vol. 2007, p. Article ID 90509, 2007.

[36] L. Böhlke and T. Ziemer, "Perceptual evaluation of violin radiation characteristics in a wave field synthesis system", *Proc Mtgs Acoust*, vol. 30, no. 1, p. Paper number: 035001, 2017.

[37] D. Menzies and M. Al-Akaidi, "Ambisonic synthesis of complex sources", *J. Audio Eng. Soc.*, vol. 55, no. 10, pp. 864–876, 2007.

[38] M. Kolundzija, C. Faller, and M. Vetterli, "Sound field reconstruction: An improved approach for wave field synthesis", in *Audio Eng Soc Convention 126*, 5 2009.

[39] W.-H. Cho, J.-G. Ih, and M. M. Boone, "Holographic design of a source array achieving a desired sound field", *J Audio Eng Soc*, vol. 58, no. 4, pp. 282–298, 2010.

[40] T. Ziemer and R. Bader, "Psychoacoustic sound field synthesis for musical instrument radiation characteristics", *J Audio Eng Soc*, vol. 65, no. 6, pp. 482–496, 2017.

[41] T. Ziemer, "Wave field synthesis by an octupole speaker system", in *SysMus09 Proceedings*, L. Naveda, Ed., 11 2009, pp. 89–93.

[42] T. Ziemer, "A psychoacoustic approach to wave field synthesis", in *Audio Eng Soc Conference: 42nd International Conference: Semantic Audio*, Ilmenau, Jul 2011, pp. 191–197.

[43] T. Ziemer, "Psychoacoustic effects in wave field synthesis applications", in *Systematic Musicology. Empirical and Theoretical Studies*, A. Schneider and A. von Ruschkowski, Eds. Frankfurt am Main: Peter Lang, 2011, pp. 153–162.

[44] T. Ziemer and R. Bader, "Implementing the radiation characteristics of musical instruments in a psychoacoustic sound field synthesis system", in *Audio Eng Soc Convention 139*, New York, 2015.

[45] T. Ziemer, "Spatial sound impression and precise localization by psychoacoustic sound field synthesis", in *Seminar des Fachausschusses Musikalische Akustik: "Musikalische Akustik zwischen Empirie und Theorie"*, R. Mores, Ed., Hamburg, 2015, pp. 17–22.

[46] E. G. Williams, *Fourier Acoustics. Sound Radiation and Nearfield Acoustical Holography*. Cambridge: Academic Press, 1999.

[47] F. Otondo and J. H. Rindel, "The influence of the directivity of musical instrument in a room," *Acta Acust United Ac*, vol. 90, pp. 1178–1184, 2004.

[48] R. Bader, "Microphone array", in *Springer Handbook of Acoustics*, T. D. Rossing, Ed. Berlin Heidelberg: Springer, 2014, pp. 1179–1207.

[49] M. R. Bai, C. Chung, P.-C. Wu, Y.-H. Chiang, and C.-M. Yang, "Solution strategies for linear inverse problems in spatial audio signal processing", *Applied Sciences*, vol. 7, no. 6, Article 582, 2017.

[50] W. Fohl and M. Nogalski, "A Gesture Control Interface for a Wave Field Synthesis System", in *Proc Int Conf New Interfaces Musical Expression*, Daejeon + Seoul, 2013, pp. 341–346.

[51] T. Ziemer, *Spaciousness in Music. In Auditory Perception, Acoustics, Audio and Sound Field Synthesis*, (under preparation): ser. Current Research in Systematic Musicology. Cham: Springer, 2018, vol. 5.