Abstract

In this paper a psychoacoustic sound field synthesis (pSFS) system for musical applications is presented. It recreates the radiation patterns of musical instruments for an extended listening area. For this purpose several psychoacoustic effects are utilized: Considering the critical bandwidth, the amount of data to be processed can be reduced massively. Applying a “precedence fade” allows for wave fronts to arrive from angles opposing to the virtual source position when regarding the integration time of the auditory system and phenomenons known from the field of auditory scene analysis. A listening test demonstrates that the approach results in a precise source localization together with a spatial sound impression which is known to be the most crucial aspect of the subjective acoustical quality of concert halls. This is achieved even in a reverberant room and with a relatively large distance between loudspeakers compared to other wave field synthesis systems. Implementations in audio systems such as 5.1 or wave field synthesis systems with a large number of loudspeakers are possible.

1 Introduction

Commonly, sound field synthesis formulations and implementations aim at creating wave fronts of virtual monopole sources or plane waves [1, 2, 3]. The focus lies on direct sound and the horizontal plane. Such a wave front synthesis has the advantage that it delivers precise localization cues in an extended listening area. For musical performances, a drawback is that monopole sources sound very narrow. Natural musical instrument, in contrast, have rather complicated direction- and frequency-dependent sound radiation characteristics creating the impression of a certain source extent. Several authors proposed methods to overcome this drawback [4, 5, 6] by either implementing higher order radiators or a distribution of monopole sources in wave front synthesis applications. They report an improved perceived naturalness and an influence on perception of spatial and timbral character as well as a recognition of the orientation of the source. However, they discovered artifacts which could only partly be compensated with high computational demands. In this paper a psychoacoustic sound field synthesis (pSFS) approach is presented, implementing the radiation characteristics of musical instruments. Its aim is to create a natural and pleasurable sound impression for musical applications rather than a physically precise replication of the original sound field. Many researchers see the potential of psychoacoustics in auralization and sound field synthesis applications [7, 8]. The method presented in this paper utilizes a number of psychoacoustic effects to achieve a precise localization and a spatial sound impression. A more detailed description of the signal processing and a more comprehensive listening test can be found in [9].

This paper is structured as follows: First, the fundamentals of spacious sound impressions in music are outlined. Against this background, a method to measure the radiation characteristics of musical instruments is presented. Then it is described how such radiation characteristics are implemented in a sound field synthesis system. Finally, it is discussed how the psychoacoustics are utilized to reduce the computational demands and to compensate physical synthesis errors. Strengths and weaknesses of this psychoacoustic sound field synthesis system are illuminated based on a comparison with mono playback in a listening test. In the prospects potential improvements of the system are discussed as well as potential implementations in other existing audio setups such as 5.1 surround systems and wave front synthesis systems with a much higher number of loudspeakers.

2 Spaciousness

In the field of subjective room acoustics the degree of spaciousness is one of the most distinct indicators concerning quality judgment of concert halls [10]. A rather large apparent source width (ASW) and a high level of listener envelopment (LEV) are ideal conditions for a pleasurable listening experience. Furthermore, a precise source localization is desirable for a clear and natural auditory image.

Spaciousness is especially a matter of interaural differences of direct sound and early reflections, often quantified by the interaural cross correlation (IACC) [10, 11]. The term, however, is comprehensive, comprising aspects such as “liveness/Halligkeit”, “spatial impression/Räumlichkeit”, “Ambience”, “Spatial Responsiveness” [11] as well as “intimacy”, ASW and LEV [10]. Most of these terms are not used consistently throughout the literature. Still, subjects agree to a large degree on how “spatial” a sound scene is, as will be seen in the results of the listening test in sec. 6. Therefore, the term is not defined more distinctly at this point but is believed to be understood intersubjectively in a listening situation, especially in a direct comparison. Localization is also a matter of direct sound and very early reflections. It is affected, among others, by interaural level and time differences within critical frequency bands [12]. In the case of several wave fronts the precedence effect plays a
crucial role in localization[12, 24]. Both source localization and the ASW are subject to auditory scene analysis principles. In music production practice a large number of recording and mixing strategies exists to increase the perceived spaciousness by recording instruments several times and/or from different angles, applying complementary comb filters or individual compressors or reverb effects on the channels etc., to create decorrelated stereo channels to increase IACC [13, 14]. This is, however, not a systematic approach. If we ask the question how musical instruments create the psychological impression of spaciousness, we can focus on those aspects of the sound field in a pSFS system.

3 Radiation characteristics of musical instruments

The radiation characteristics of musical instruments have been investigated with circular or spherical microphone arrays by [15] and more recently by [16, 17]. Even under anechoic conditions the radiation patterns of musical instruments create interaural signal differences which lead to the impression of a certain source extent as demonstrated e.g. in [18]. This is especially true in the near field. In the far field a wavefront more and more resembles a narrow sounding plane wave. For this pSFS system a method is chosen to record and recreate interaural signal differences for listeners up to distances of 3 meters. The radiated sound of musical instrument is recorded simultaneously by 128 microphones which are distributed equidistantly along a circle with a radius of 1 m, yielding $M = 128$ time series

$$p(t, r_m), m = 0, \ldots, M - 1$$

where $r_m$ is the position vector of the $m$th microphone in polar coordinates consisting of the horizontal angle $\varphi = \frac{2\pi m}{28}$ and the distance $r = 1$ m. Fig. 1 is a photo of the measurement setup. A discrete Fourier transform (DFT) over one second of quasi-stationary sound of each microphone recording yields 128 complex spectra

$$P(\omega, r_m) = \text{DFT}[p(t, r_m)].$$

When considering each musical instrument as complex point source $Q$ originated at the center of the microphone array, i.e. ignoring its actual extent and geometry, the measured spectra $P(\omega, r_m)$ equal the (unknown) source spectrum at the origin $P(\omega, r_0)$ modified by a complex transfer function which consists of a function of distance $G(\omega, r)$ and a function of angle $\Gamma(\omega, \varphi)$:

$$P(\omega, r_m) = P(\omega, r_0) G(\omega, r) \Gamma(\omega, \varphi)$$

Here, $G(\omega, r)$ is the well-known free field Green’s function for point sources describing a spherical wave front propagation

$$G(\omega, r) = \frac{e^{-ikr}}{r},$$

where $\text{i}$ is the imaginary unit. From eq. 3 we can see that the recorded spectra $P(\omega, \varphi_m, 1 \text{ m})$ are directly proportional to $\Gamma(\omega, \varphi)$, which is the only direction-dependent function and thus fully characterizes the sound radiation characteristics of the complex point source. This way we simplify musical instruments to point sources while maintaining their complex sound radiation characteristics to a certain extent. Any arbitrary source spectrum can be propagated to all measured angles at all distances by eq. 3. The interaural signal correlation is important for the ASW, as mentioned in section 2, so both amplitude and phase need to be considered. 128 microphones at a distance of 1 m yields one microphone every 0.05 m. Propagating the wave front to a radius of 1.5 m the wave front is sampled every 0.075 m and at 3 m every 0.15 m. The complex sound pressure in between the measured angles need to be gained in terms of interpolation. Note that assuming a distance of 0.15 m between a listener’s ears, we have gained interaural information for distances up to three meters away from the source. The setup and equipment, the exact mathematics as well as some exemplary results of measured radiation patterns and propagated sound fields will be available soon in detail in [19].

4 Sound Field Synthesis

In general, the presented sound field synthesis works in two steps: First, it is calculated how the sound field would be within a listening area if a complex point source with a chosen position and orientation radiated an arbitrary source spectrum. This is achieved by propagating the source spectrum by means of eq. 3 to discrete listening points which sample the listening area. The result is the complex amplitudes at the listening points, the desired sound field. According to the Nyquist-Shannon theorem for a band-limited signal the sound field is reconstructed correctly within the whole listening area up to the critical wavelength which is twice the distance of the listening points. The approach is similar to the “sound field reconstruction” principle presented in [20]. As a second step, loudspeaker spectra are calculated which create the desired sound field by superposition of their signals. This step can be written as simple matrix equation:

$$K_{Y_l - X_n} P_{X_n} = P_{X_n}, n = 1, \ldots, N, l = 1, \ldots, L$$

Here, $P_{X_n}$ are the calculated complex spectra at the chosen listening points $X_n$. $Y_l$ are the loudspeaker posi-
tions, \( \mathbf{P}_Y \), the (yet unknown) loudspeaker spectra needed to create the desired sound field. \( \mathbf{K}_Y \cdot \mathbf{x}_c \) is the propagation matrix containing the complex transfer function which describes the sound propagation from each loudspeaker to each listening point. The linear equation system, eq. 5, has a unique solution if as many listening points as loudspeakers are chosen, i.e. if \( L = N \). Eq. 5, however, is ill-posed. Furthermore, assuming the loudspeakers to be monopole sources, the propagation matrix may be ill-conditioned: If listening points lie close to another and if loudspeakers have similar distances to the listening points, compared to the considered wavelengths, the contribution of each individual speaker to the sound field at one listening point is very similar and thus the rows of the linear equation system are not independent of one another. A badly conditioned matrix could cause extremely high amplitudes and strong variations of the solution even when listening points, source position, orientation or radiation pattern are only altered slightly. To prevent from these undesired cases the loudspeakers are not considered as monopoles. Instead, they are considered as complex point sources just as the musical instruments. This means the measurements described in sec. 3 are also performed for the used loudspeakers. This does not guarantee for a robust solution but in practice is has proven to deliver good results. Naturally, an inverse Fourier transform of the calculated spectra is necessary for driving the loudspeakers.

5 Psychoacoustics

The sound field synthesis approach described above is formulated in frequency domain and could thus be called *spatial sound spectrum synthesis*. It is valid for stationary signals, i.e. stationary sound of motionless sources. The approach neglects transients in the audio signal and the propagation times of wave fronts. For music audio files the linear equation systems is solved for overlapping time blocks, assuming the sound to be quasi-stationary during one block. Furthermore, the sound field synthesis approach needs to be adapted based on the following psychoacoustical considerations.

Critical Bands

Human listeners are mostly not able to perceive all frequencies of instrumental sound or even a complex musical piece. Rather, frequencies within one critical band are processed together. Critical bands are assumed to have a width of about 20 % of a center frequency and what more towards lower frequencies [21]. The loudest frequency within a critical band either masks the others or they are perceived as one tone with a beating or as narrow band noise. Also, frequencies within one frequency band are localized together. Therefore, the audible frequency region is divided into 25 critical bands for the presented sound field synthesis approach. Only one radiation pattern per critical band is recorded and stored. In the sound field synthesis process, the measured radiation pattern is reconstructed only for the loudest frequency of each critical band. This is a good compromise between synthesis precision and computational costs. Several notes need to be played by each instruments so one radiation pattern per critical band can be stored. As the radiation pattern of a frequency may depend on the played fundamental, fingering etc. [16, 22], the stored patterns are only exemplary and do not deliver a full description of the radiation characteristics.

Auditory Scene Analysis

Auditory scene analyses (ASA) [23] is the perceptual organization of sound. It is the attempt of the auditory system to group those parts of sound which originated in the same physical happening into one auditory stream, the pendant to the visual object. Parts of sound simultaneously reaching the listener are segregated into individual auditory streams. This is called simultaneous grouping. These streams are sustained over time in terms of sequential grouping. From all sounds reaching the listener within a short time window, those parts are likely integrated into a stream which are similar in terms of harmony (i.e. tones are harmonics of a common fundamental) and synchrony (e.g. of note onset or offset). Sequential sounds are likely to be integrated into a common auditory stream if they are related in terms of timbre, proximity (of frequency, pitch, and/or location), continuity (like pitch glides or smooth spatial motions) and common fate (e.g. modulations of amplitude, frequency or beat rate).

Simultaneous grouping is necessary to understand a number of frequencies as being one complex tone, sequential grouping to understand a sequence of complex tones as being a melody. Streams belong together and are likely to get one perceived group location, source extent etc. Another aspect which increases the probability of one common group location is that all wave fronts arrive within the integration time of the auditory system.

Integration Time

The integration time \( T = 50 \text{ ms} \) is a threshold which is well-known in the field of psychoacoustics [24]. Successive acoustical events happening quicker than that are not discriminated but perceived as one sound or noise. \( T \) is also the echo threshold: Reflections arriving within the first 50 ms after a direct sound have an amplifying effect whereas later arriving reflections are heard as echoes [11, 12, 24]. Even if a reflection, arriving with a delay of 5 to 30 ms, is 10 dB louder than the first arriving wave front, it won’t affect the localization.

As propagation time is neglected in the sound field synthesis approach, all loudspeaker signals start at the same time. Only amplitudes and initial phases of the 25 lowest frequencies are altered according to the solution of eq. 5. To arrive at all potential listening positions within the integration time of the auditory system, the listening area must not have dimensions larger than \( 7c \approx 17 \text{ m} \), where \( c \) is the sound velocity. Furthermore, the distance between the loudspeakers closest and farthest to any potential listening position must not exceed 17 m.
Precedence Effect

Those parts of sound belonging to the same auditory stream and arriving at the listener within the integration time are likely to have one common group location. This location is largely determined by the precedence effect [24], i.e. the location of the first wavefront which arrives at the listener. This is even true if later arriving wave fronts have a largely deviant origin and a higher amplitude.

Considering the above mentioned perception principles—ASA, the integration time and the precedence effect—it is obvious that with the sound field synthesis principle every listener will localize the virtual source at the position of the nearest loudspeaker. However, these principles can also be utilized to create the perception of a distinct source location which is the same for all listeners: If only one loudspeaker actively plays the onsets of all notes, the perceived source position will coincide with this loudspeaker position. The signal amplitudes of the other loudspeakers are faded in after this precedence loudspeaker. The fading duration has to be chosen in a way that the signals are completely faded in after the first wave front emanating from the precedence speaker has reached every listening point. This precedence fade at the note onset directs the perceived source position towards the precedence loudspeaker for all listeners. Again, the fading duration must not exceed the integration time of the auditory system.

6 Comparison with Mono

A listening test reveals if the implementation of the psychoacoustic principles results in a proper localization and a natural sound that is not corrupted by the fading. The subjective impression of the presented psychoacoustic sound field synthesis approach is investigated in a listening test with 24 subjects.

Test setup

A loudspeaker setup consisting of 15 loudspeakers with a distance of $\Delta Y = 0.65$ m between them is installed in a reverberant room of dimensions 2.5 m/4.8 m/4.3 m (h/w/d). The discrete listening points which sample the listening area have a distance of 0.17 m, so the aliasing frequency lies around 1 kHz. Compared to that, a wave front synthesis would have a worst-case aliasing frequency of $f_{\text{max}} = \frac{1}{2\Delta Y} \approx 264$ Hz. The speaker setup surrounds a listening area from three sides. In groups of 2 to 3, subjects stand in the listening area. They are free to turn around to improve localization performance. 40 seconds of a fugue are played first in mono and then by means of the pSFS including the measured radiation characteristic of a violin. During playback subjects mark the perceived source location on a “map” which shows the loudspeaker setup and a straight line every 10°. These lines help the subjects to orientate on the map. They are advised to mark the source position with a cross, if the source is perceived as very narrow, or with a line or curve, if the source sounds broad or cannot be localized more precisely. Subjects were explicitly allowed to mark several source positions when localizing sources in different directions. The loudspeaker setup, the lines and the listening area are illustrated as background of the test results in fig. 2 and 3. Subjects furthermore judge, on a scale from 1 to 7, how well the source can be localized (localizability), how natural it sounds (naturalness) and how spatial the sound impression is (spaciousness). These terms were not further described to not draw the attention of the listeners to specific sound features. Finally, the subjects judge whether they heard one or several source point(s) or region(s).

Results

The subjective impressions of the subjects are listed in tab. 1. The mono playback is localized more distinctly and reported as sounding similarly natural but less spatial than the pSFS. Most subjects had the impression that the source location was one point in the mono playback and one region in the pSFS. Eight listeners heard more than one source point or region in the pSFS. Apparently, some notes or some spectral components seemed to be localized separately by these subjects. The reported evaluations are in good agreement with the marked perceived source angles which are plotted in fig. 2 and 3. The polar plots show the number of subjects who marked an angle as source position or region in steps of 0.5°. The two additional plots with a lower opacity show the cumulated number of subjects within steps of 5° and 10°. Concentric circles indicate 5, 10 and 15 subjects. The Q marks the source position which is the only active loudspeaker at 170.5° (mono) and the virtual source position at 9.5° (pSFS). The right hand side is considered as 0°. It can be seen that the presented pSFS approach is localized almost as good as the mono source; the lobe is only slightly broader. It is conspicuous that in this test even the mono source is localized much worse than reported in the literature, e.g. [12]. Possible causes are the parallel plane walls creating strong room modes. Furthermore, the distance between loudspeaker membranes and walls is about 0.45 m, so the first reflection can corrupt the localization. It can be observed throughout most listening test that perceived sources are shifted towards the room corners which is probably due to the fact that most early reflections arrive from there [25].

Discussion

The virtual complex point source created by the pSFS system is rated as sounding about as natural and more spatial than mono playback but localizable with a lower precision. The localization task confirms this subjective impression. Most listeners report to hear an extended source region rather than a narrow point source. Introducing the precedence fade, a fairly precise localization is possible although loudspeaker sounds from 15 directions arrive at the listeners. With the present listening test conditions it remains uncertain if the increased spaciousness results from including the radiation characteristics of the instrument, from the distributed wave fronts arriving
Table 1: Mean values (ar. mean ¤) of perceived source localizability, naturalness and spaciousness of mono and the presented psychoacoustic sound field synthesis (pSFS) system and absolute number of subjects who had the impression of 1 or several (>1) source point(s) or region(s).

<table>
<thead>
<tr>
<th></th>
<th>Mono</th>
<th>pSFS</th>
</tr>
</thead>
<tbody>
<tr>
<td>© Localizability</td>
<td>6.33</td>
<td>5</td>
</tr>
<tr>
<td>© Naturalness</td>
<td>4.5</td>
<td>4.62</td>
</tr>
<tr>
<td>© Spaciousness</td>
<td>3.125</td>
<td>4.75</td>
</tr>
<tr>
<td>1 Point</td>
<td>17</td>
<td>2</td>
</tr>
<tr>
<td>&gt;1 Point</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>1 Region</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>&gt;1 Region</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

Figure 2: Number of subjects who marked the specific angle as source location or area in the mono presentation.

Figure 3: Number of subjects who marked the specific angle as source location or area in the presented sound field synthesis approach.

at the listeners, or from the more decorrelated reverberation due to the higher number of unequal loudspeaker signals. Although reflections from the walls and the wave shadow created by listeners are not considered in the calculation, the principle still works for several listeners in a reverberant room.

7 Conclusion

A sound field synthesis approach has been introduced. Its core is a linear equation system whose solutions are loudspeaker spectra which together create a desired sound spectrum in an extended listening area by means of superposition of radiated sounds. The calculation for the desired sound field includes the measured radiation characteristics of musical instruments which are simplified as complex point sources. This way interaural signal differences resemble those of natural instruments rather than of monopole sources. For the synthesis to be correct, audio signals need to be stationary. At transients the superposition of the loudspeaker signals does not create the desired sound field as arrival times of the wave fronts may deviate. The different arrival times also prohibit a correct localization of the virtual source. These shortcomings are compensated by regarding psychoacoustics. Calculating new complex amplitudes only for the 25 lowest frequencies, it is likely that the loudspeaker signals fall into the same auditory stream as long as their radiated wave fronts arrive within the integration time of the auditory system. Finally, the precedence fade ensured that all listeners localize the virtual complex point source at the same position. The result is a natural sound with a precise source location and a spatial sound impression.
8 Prospects

In principle, the described pSFS can be applied directly on conventional and widespread loudspeaker setups such as 5.1 system. The only requirements are that the radius of the circular loudspeaker arrangement is known as well as the radiation characteristics of the speakers. In an informal listening test on a 5.1 system in a studio with heavy damping, the results were very much the same as the results in the listening test: The source could be localized well even beyond the listening area, which was of course rather small due to the limited number of loudspeakers. The spatial sound impression remained despite the damping of the room. Different orientations of the source especially lead to different spectral colorations. The pSFS principle can also be applied on wave front synthesis systems above the aliasing frequency to increase spaciousness and decrease aliasing artifacts. Setting up and solving an over-determined linear equation system with more listening points than available loudspeakers may even increase the extent of the listening area or increase the aliasing frequency. So far, the pSFS approach has been tested for static virtual sources at the loudspeaker positions only. Further research may concentrate on sources beyond the loudspeaker positions and moving sources. Listening tests under anechoic conditions may reveal to what degree the radiation pattern influences the perceived spaciousness.

References