ARTIFACTS IN THE SOUND FIELD OF A MOVING SOUND SOURCE RECONSTRUCTED FROM A MICROPHONE ARRAY RECORDING

Jens Ahrens and Sascha Spors

Deutsche Telekom Laboratories, Technische Universität Berlin
Ernst-Reuter-Platz 7, 10587 Berlin, Germany
`{jens.ahrens,sascha.spors}@telekom.de`

ABSTRACT

We present an analysis of the sound field of a moving sound source reconstructed from recordings of a virtual dual-radius open-sphere microphone array. As a consequence of the discrete property of such microphone distributions artifacts arise, most notably spatial aliasing and spatial bandwidth limitation artifacts. We show that these artifacts are much more pronounced for moving sound sources than for static ones. We analyze the artifacts with a focus on a possible perceptual impairment when such recordings are used for audition purposes.

Index Terms— Spatial audio, spherical microphone array, spherical harmonics, spatial aliasing, wave field extrapolation

1. INTRODUCTION

In the context of sound field reproduction, it was discovered that the artifacts inherent to a loudspeaker setup under consideration exhibit a very prominent quality for moving sources [1]. For stationary scenarios it is such that artifacts like spatial sampling artifacts and loudspeaker array truncation artifacts always exhibit the same temporal frequency like the desired signal. This circumstance makes them perceptually tolerable in many situations.

However, for dynamic scenarios, especially moving sources, above mentioned artifacts can exhibit a different temporal frequency than the desired signal. This constitutes a perceptually very disturbing impairment.

In the capturing of sound field by microphone arrays, spatial sampling artifacts similar - but not equal - to those occurring in multichannel sound field reproduction arise [2]. In this contribution we investigate the properties of the sample scenario of a single moving sound source with respect to artifacts introduced by the capturing process. To our awareness an analytical treatment of moving sound sources is not available in the spatial frequency domain. We therefore employ numerical simulations based on an analytical formulation in the time domain in order to get more insight into the subject.

In [3], a setup for realtime sound field capture and playback employing 60 microphones on a rigid sphere with radius 0.1 m is presented. We are not aware of a realtime implementation providing higher resolution. For convenience, we therefore consider a setup comparable to the one described above. However, due to the fact that we have only knowledge of the sound field emitted by a moving source under free-field conditions, we do not employ a rigid sphere array but a so-called open design [4]. As described below, we actually assume microphones distributed over two spherical surfaces centered at the same position but with two different radii. Due to practical problems like the limitedly available space for wiring and microphone placement we do not expect that such a setup is implementable for realtime operation. However, the fundamental properties of such a dual-radius open array are comparable to rigid sphere arrays and the findings derived in this paper hold qualitatively [5].

Nomenclature

In the remainder of the paper, we use the following notational conventions: For scalar variables, lower case denotes the time domain, upper case the temporal frequency domain. Vectors are denoted by lower case boldface. The three-dimensional position vector in Cartesian coordinates is given as $\mathbf{x} = [x \ y \ z]^T$. The Cartesian coordinates are linked to the spherical coordinates via $x = r \cos\alpha \sin\beta$, $y = r \sin\alpha \sin\beta$, and $z = r \cos\beta$. α denotes the azimuth, β the zenith angle. Refer also to Fig. 1.

The acoustic wavenumber is denoted by k. It is related to the temporal frequency by $k^2 = \left(\frac{\omega}{c}\right)^2$ with ω being the radial frequency and c the speed of sound. Outgoing spherical waves are denoted by $\frac{1}{r}e^{-i\frac{\omega}{c}r}$. i is the imaginary unit ($i = \sqrt{-1}$).
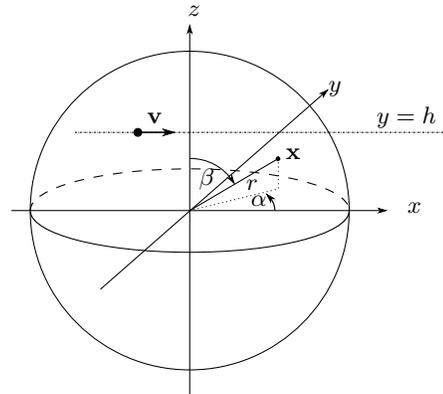
Figure 1: The coordinate system used in this paper. The sphere indicates one of the microphone arrays. The trajectory of the moving sound source is indicated by the dash-dotted line.

2. OPEN-SPHERE MICROPHONE ARRAYS

A sound field $S(\mathbf{x}, \omega)$ which is source-free in the region $r < r_{\max}$ can be described in that region by its spherical harmonics expan-

sion as [6]

$$S(\mathbf{x}, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \breve{S}_n^m(\omega) \, j_n\left(\frac{\omega}{c}r\right) Y_n^m(\alpha, \beta) \,, \qquad (1)$$

whereby $j_n\left(\frac{\omega}{c}r\right)$ denotes the n-th order spherical Bessel function of first kind [6].

The spherical harmonics $Y_n^m(\alpha, \beta)$ are defined as [6]

$$Y_n^m(\alpha, \beta) = \sqrt{\frac{(2n+1)}{4\pi} \frac{(n-m)!}{(n+m)!}} \cdot P_n^m(\cos\beta) \cdot e^{im\alpha} \,, \quad (2)$$

with $P_n^m(\cdot)$ denoting the m-th order associated Legendre polynomial of n-th degree.

The coefficients $\breve{S}_n^m(\omega)$ as defined in (1) can be obtained from the sound field $S(\mathbf{x}, \omega)$ as [6]

$$\breve{S}_n^m(\omega) = \frac{1}{j_n\left(\frac{\omega}{c}r_{\text{ref}}\right)} \int_{\Omega \in S^2} S(\Omega, \omega) \, Y_n^m(\Omega)^* \, d\Omega \,, \qquad (3)$$

whereby the asterisk $*$ denotes complex conjugation. Ω denotes the surface of a sphere of radius r_{ref} which is centered around the origin of the coordinate system. By arranging pressure microphones on such a spherical surface, the coefficients $\breve{S}_n^m(\omega)$ can be approximately obtained from the recording by approximating the integral in (3) by a summation as [2]

$$\breve{S}'^m_n(\omega) = \frac{1}{j_n\left(\frac{\omega}{c}r_{\text{ref}}\right)} \sum_{l=1}^{L} w_l \, S(\Omega_l, \omega) \, Y_n^m(\Omega_l)^* \,, \qquad (4)$$

whereby $\breve{S}'^m_n(\omega) \approx \breve{S}_n^m(\omega)$. w_l represent the weights applied to the individual measurement points which are dependent on the quadrature employed. L is the number of sampling points and Ω_l denotes the position of the measurement point with index l.

Two properties of (4) have to be emphasized:

1) The Bessel function in (4) exhibits zeros and therefore (4) is undefined for certain combinations of frequency and radius.

2) The approximation of an integration by a summation over discrete points introduces various artifacts [2].

Ad 1): In order to avoid undefined expressions and numerical instabilities, the microphones in open-sphere designs are typically arranged on two concentric spherical surfaces. The radii of the latter are chosen such that problematic frequencies on one surface can be safely captured by the other one [4].

It is furthermore such that the Bessel functions in (4) exhibit zeros for zero argument for all $n > 0$ [6]. For small arguments, Bessel functions of order $n > 0$ exhibit very small values. This results in heavy amplifications at low frequencies and therefore in numerical instabilities which can not be avoided by switching to a different radius. In practical implementations, this heavy amplification is also applied on noise present in the microphone signals which limits the operating frequency range. However, this circumstance is not of relevance in the present investigation due to ideal assumptions.

Since we are not particularly interested in very low frequencies in the present study, we exclude them from the analysis avoiding above described instabilities.

Ad 2): It is generally impossible to distribute discrete microphones on a spherical surface such that the summation in (4) is exact when

$S(\cdot)$ is not bandlimited. Depending on the situation, we employ a grid of 64 or 324 nodes arranged according to [7] on each of the two spherical surfaces. These nodes represent a nearly uniform distribution and have been found to provide a suitable choice for the problem under consideration [3].

Employing the 64-node grid we can not reliably extract higher orders than $N = 7$ from the recordings, the 324-node grid allows for orders up to $N = 17$ [3]. Since the recorded sound fields are typically of infinite spatial bandwidth, only a spatially bandlimited approximation $S'(\mathbf{x}, \omega)$ of the recorded sound field $S(\mathbf{x}, \omega)$ can be obtained which is given by

$$S'(\mathbf{x}, \omega) = \sum_{n=0}^{N-1} \sum_{m=-n}^{n} \breve{S}'^m_n(\omega) \, j_n\left(\frac{\omega}{c}r\right) Y_n^m(\alpha, \beta) \,. \quad (5)$$

$S'(\mathbf{x}, \omega)$ is said to be of N-th order respectively to exhibit a bandwidth of N. A pair (n, m) is referred to as *spatial mode*.

The fact that the recorded sound field $S(\mathbf{x}, \omega)$ is generally not spatially bandlimited in combination with the spatially discrete distribution of the microphones leads to a leakage of higher orders into lower orders which is referred to as *spatial aliasing* [2]. At temporal frequencies below a certain limit typically referred to as *spatial aliasing frequency*, the energy of the spatial aliasing artifacts is very low and the recording is considered as not corrupted. The spatial aliasing frequency is dependent on the order n and on the frequency f which are aimed to be extracted, on the radius r of the microphone array under consideration, and on the number of microphones employed and their distribution. The higher the frequency is for which we try to extract $\breve{S}'^m_n(\omega)$ the more aliasing we have to expect [2].

3. SIMULATION FRAMEWORK

In this section, we outline the framework which we have set up in order to carry out the present investigation.

We assume a single monopole sound source under free-field conditions. The source emits a monochromatic signal $s_0(\tilde{t})$ of frequency $\omega_s = 2\pi f_s$ and amplitude a_0. In complex notation $s_0(\tilde{t})$ reads

$$s_0(\tilde{t}) = a_0 \cdot e^{j\omega_s \tilde{t}} \,. \qquad (6)$$

The sound field $s(\mathbf{x}, t)$ of a monopole source emitting $s_0(\tilde{t})$ and moving uniformly at a speed $v < c$ on a trajectory in the x-y-plane parallel to the x-axis in positive x-direction and at a distance h (refer to Fig. 1) reads then [1]

$$s(\mathbf{x}, t) = \frac{1}{4\pi} \cdot \frac{s_0(\tilde{t}(\mathbf{x}, t))}{\Psi(\mathbf{x}, t)} \,, \qquad (7)$$

whereby $\tilde{t}(\mathbf{x}, t)$ denotes the time instant, where the signal arriving at the point $\mathbf{x}$ at time t has been emitted by the sound source. $\tilde{t}(\mathbf{x}, t)$ is given by

$$\tilde{t}(\mathbf{x}, t) = t - \frac{M\Phi(x, t) + \Psi(\mathbf{x}, t)}{c(1 - M^2)} \,,$$

$$\Psi(\mathbf{x}, t) = \sqrt{\Phi^2(x, t) + ((y - h)^2 + z^2)(1 - M^2)} \,,$$

$$\Phi(x, t) = x - vt - x_s(0) \,.$$

We are not aware of an analytical transformation of $s(\mathbf{x}, t)$ neither into the temporal frequency domain nor into the spherical harmonics domain. We therefore perform a numerical Fourier transform

on $s(\mathbf{x}, t)$ in order to apply (4). The result is then numerically transformed back to the time domain.

We assume the sound source to move uniformly along the above specified trajectory at distance $h = 2$ m and at speed $v = 30$ m/s.

We simulate a double-radius open-sphere microphone array composed of 64 respectively 324 microphones on each radius nearly uniformly distributed according to [7]. The microphone array is centered around the origin of the coordinate system (refer to Fig. 1).

The radii r_1 and r_2 of the two spherical surface where chosen for each set of f_s and N such that numerical instabilities are avoided.

4. RESULTS

In this section, we summarize the conclusions which we have drawn from numerical analysis of the scenario under consideration. We employ different ways of illustrating the findings, such as spectrogram analysis and inspection of the time domain signal, dependent on the specific artifacts to investigate. Especially spectrogram analysis has shown to be a suitable tool for the investigation of spatial aliasing artifacts of time-variant systems [1].

4.1. Spatial aliasing

Fig. 2(a) and Fig. 2(b) depict spectrograms of the sound field under consideration reconstructed at the center of the microphone array, i.e. at $\mathbf{x} = [0\ 0\ 0]^T$. In both figures, a microphone array with 64 microphones on each spherical surface is employed. Note that only the zeroth order ($N = 1$) is extracted. The higher modes do not contain information since $j_n(0) = 0\ \forall\ n > 0$ [6] (refer also to (5)).

Fig. 2(a) indicates the result for $f_s = 1000$ Hz. The reconstructed
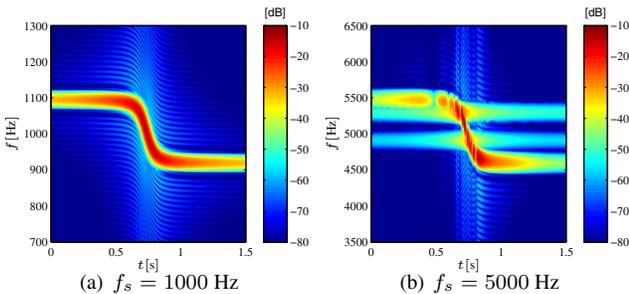
Figure 2: Spectrograms of the reconstructed sound fields at $\mathbf{x} = [0\ 0\ 0]^T$ for different frequencies. In both cases $r_1 = 0.2$ m, $r_2 = 0.22$ m, $N = 1$, and 64 microphones are employed on each spherical surface.

sound field does not exhibit considerable artifacts. It is nearly indistinguishable from perfect reconstruction. The low energy artifacts apparent in Fig. 2(a) are artifacts inherent to the spectrogram analysis and can not be attributed to the signal analyzed.

However, the signal with $f_s = 5000$ Hz in Fig. 2(b) is heavily corrupted. The energy of the artifacts as well as their prominence increase with increasing frequency f_s. It can be shown that an increase of the number of microphones and thereby an increase of the spatial aliasing frequency makes these artifacts vanish. We therefore conclude that the latter are a consequence of spatial aliasing.

With moving sound sources it is such that higher orders describe components of the sound field which generally exhibit a different temporal frequency than the components which are described by the lower orders. Therefore a leakage of higher orders into lower ones can lead to energy at frequencies different from that of the desired signal. Note that this is not the case with static sound fields like plane and spherical waves.

The results deduced from figure Fig. 2(b) indicate that for applications with a setup comparable to that of the present investigation, e.g. [3], considerable corruption has to be expected already at moderate frequencies. Informal listening suggests that this type of artifacts can be perceptually very disturbing.

4.2. Spatial bandwidth limitation

So far, we have only investigated the reconstruction of the moving source's sound field in the center of the microphone array. However, it is widely known that the limited spatial bandwidth of the reconstructed sound field (refer to (5)) concentrates the energy of the latter around the center of the spherical harmonics expansion respectively along a channel in direction of instantaneous propagation [8]. The concentration of energy is more pronounced the higher the temporal frequency is. Note that in the context of sound field reproduction spatial bandwidth limitation is also referred to as order truncation [8]. In the remainder of this section, we employ 324 microphones on each spherical surface.

In order to get insight into the consequences of the spatial band-
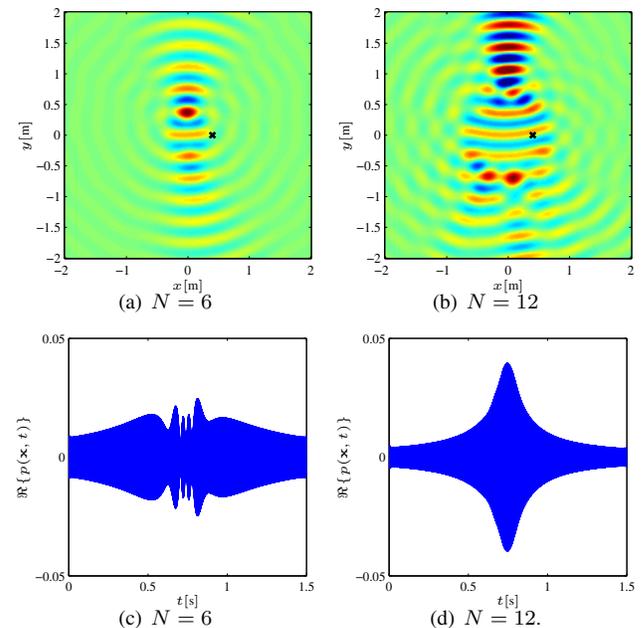
Figure 3: Snapshot of the reconstructed sound field for $f_s = 1000$ Hz for different spatial bandwidths in (a) and (b) at $t = 0.74$ s. The marks in (a) and (b) indicate the location $\mathbf{x} = [0.4\ 0\ 0]^T$ for which the according time domain signal is depicted in (c) and (d). In both cases $r_1 = 0.2$ m, $r_2 = 0.36$ m, and 324 microphones are employed on each spherical surface.

width limitation, we reconstruct the sound field under consideration in a location at a given distance from the center of the micro-

phone array. Fig. 3(a) depicts a snapshot of the reconstruction of the moving source's sound field for $N = 6$.

When inspecting the location $\mathbf{x} = [0.4\ 0\ 0]^T$, we find that the reconstructed sound field is corrupted by undesired amplitude fluctuations. The time domain signal at $\mathbf{x}$ is depicted in Fig. 3(c). The amplitude fluctuations are apparent around $t = 0.7$ s. Note that these fluctuations are also observed in spectrograms. We chose a time domain representation with linear amplitude scaling in order to make the fluctuations clearly visible.

A snapshot of the reconstruction from the same recording is depicted in Fig. 3(b) for $N = 12$. It can be seen that the signal at $\mathbf{x}$ is nearly perfectly reconstructed. For 6-th order reconstruction it is such that the observed location $\mathbf{x}$ does not lie inside the region of high energy at all considered instants of time. Note that the instantaneous direction of propagation of the reconstructed signal, and therefore also the extent of the channel of high energy described above, change over time. This variation of energy of the reconstructed sound field over time leads to the observed amplitude fluctuations. For the 12th-order reconstruction, $\mathbf{x}$ always lies within the region of high energy and therefore no amplitude fluctuations arise. Note however that the general amplitudes of the signals in Fig. 3(c) and Fig. 3(d) do not match but are both similar to the general amplitude of the perfect reconstruction. For far locations, more deviation has to be expected.

Since we assume that no considerable spatial aliasing is apparent in the above described scenario, we conclude that the amplitude fluctuations apparent in Fig. 3(c) are a consequence of the spatial bandwidth limitation occurring in the reconstruction.

The fact that we have not found comparable amplitude fluctuations in the reconstructions of static sound fields like plane and spherical waves suggests that the fluctuations are a consequence of the time-variant property of the sound field under consideration.

4.3. Extrapolation

Applications which employ loudspeaker arrays to reconstruct the captured sound field over an extended receiver area, e.g. [8], implicitly extrapolate the captured sound field to the positions of the loudspeakers. The latter can be positioned at a distance of several meters from the center of the loudspeaker arrangement (which virtually coincides with the center of the microphone array). To our awareness a thorough analysis of the limitations of the extrapolation of such spatially discrete data is not available.

However, an analysis of the consequences of spatial aliasing on such a wave field extrapolation can not be performed within the scope of this paper. This mostly due to the fact that artifacts of the spatial bandwidth limitation described in Sec. 4.2 superpose to the extrapolation artifacts. Recall from Sec. 4.2 that severe artifacts due to spatial bandwidth limitation occur already at moderate distances for a reconstruction bandwidth of $N = 12$. Note that such a 12-th order reconstruction is already far beyond what typical real-world implementations of microphone arrays are capable of delivering.

As a consequence, a simulation framework is required which can extract enough spatial modes to assure that no considerable spatial bandwidth limitation artifacts arise at the location under consideration. Such a simulation framework has to provide significantly higher numerical precision than the current framework does. This high numerical precision is required due to the fact that already at moderate frequencies and low orders, the reciprocal of the Bessel function in (4) becomes very large. We emphasize again

that, to our awareness, currently no real-world realtime system exists which is capable of extracting more than a few orders from a recording.

Note that the reproduction of a sound field by a loudspeaker array also introduces severe artifacts, most notably spatial aliasing, which are superposed to the artifacts present in the recording which is auditioned [8].

5. CONCLUSIONS

We have presented an analysis of the properties of the sound field of a moving sound source when it is reconstructed from a microphone array recording. It was shown that the occurring artifacts can be expected to be perceptually much more prominent for moving sources than for static scenarios.

When the sound field is reconstructed at the center of the microphone array, the reconstruction is free of considerable artifacts for relatively low temporal frequencies. Above a certain frequency, spatial aliasing arises. These spatial aliasing artifacts can exhibit a temporal frequency which is different from the frequency of the desired signal. The human ear is very sensitive towards this type of artifacts.

The reconstruction of the moving source's sound field is always limited with respect to the spatial bandwidth. This spatial bandwidth limitation can lead to artifacts when the sound field is reconstructed at locations at a given distance from the center of the microphone array. The artifacts become apparent as fluctuations of the amplitude of the time domain signal of the reconstructed sound field at a given location. These amplitude fluctuations can become pronounced for far locations.

The investigation of possible consequences of the arising artifacts on human spatial perception was beyond the scope of the present study and is subject to future work.

6. REFERENCES

[1] J. Ahrens and S. Spors, "Reproduction of moving virtual sound sources with special attention to the doppler effect," in *124th Conv. of the AES*, Amsterdam, The Netherlands, May 17–20 2008.

[2] B. Rafaeli, B. Weiss, and E. Bachmat, "Spatial aliasing in spherical microphone arrays," *IEEE Trans. on Signal Proc.*, vol. 55, no. 3, pp. 1003–1010, March 2007.

[3] R. Duraiswami, D. N. Zotkin, Z. Li, E. Grassi, N. A. Gumerov, and L. S. Davis, "High order spatial audio capture and its binaural head-tracked playback over headphones with HRTF cues," in *119th Conv. of the AES*, New York, NY, Oct. 7–10 2005.

[4] I. Balmages and B. Rafaeli, "Open-sphere designs for spherical microphone arrays," *IEEE Trans. on Audio, Speech, and Language Proc.*, vol. 15, no. 2, pp. 727–732, Feb. 2007.

[5] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Trans. On Speech and Audio Process.*, vol. 13, no. 1, pp. 135–143, Jan. 2005.

[6] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustic Holography*. London: Academic Press, 1999.

[7] J. Fliege and U. Maier, "A two-stage approach for computing cubature formulae for the sphere," in *Ergebnisberichte Angewandte Mathematik 139T, Universität Dortmund*, 1996. [Online]. Available: http://www.personal.soton.ac.uk/jf1w07/nodes/nodes.html

[8] J. Ahrens and S. Spors, "An analytical approach to sound field reproduction using circular and spherical loudspeaker distributions," *Acta Acustica utd. with Acustica*, vol. 94, no. 6, pp. 988–999, Nov./Dec. 2008.