EMBEDDING DISTANCE INFORMATION IN BINAURAL RENDERINGS OF FAR FIELD RECORDINGS

César Salvador, Shuichi Sakamoto, Jorge Treviño and Yôiti Suzuki

Advanced Acoustic Information Systems Laboratory,
Res. Inst. Electrical Communication and Grad. Sch. Information Sciences, Tohoku University
Sendai, Japan
{salvador@ais., saka@ais., jorge@ais., yoh@}riec.tohoku.ac.jp

ABSTRACT

Traditional representations of sound fields based on spherical harmonics expansions do not include the sound source distance information. As multipole expansions can accurately encode the distance of a sound source, they can be used for accurate sound field reproduction. The binaural reproduction of multipole encodings, though, requires head-related transfer functions (HRTFs) with distance information. However, the inclusion of distance information on available data sets of HRTFs, using acoustic propagators, requires demanding regularization techniques. We alternatively propose a method to embed distance information in the spherical harmonics encodings of compact microphone array recordings. We call this method the Distance Editing Binaural Ambisonics (DEBA). DEBA is applied to the synthesis of binaural signals of arbitrary distances using only far-field HRTFs. We evaluated DEBA by synthesizing HRTFs for nearby sources from various samplings of far-field ones. Comparisons with numerically calculated HRTFs yielded mean spectral distortion values below 6 dB, and mean normalized spherical correlation values above 0.97.

1. INTRODUCTION

The primary cues for distance perception are the intensity and the direct-to-reverberant energy ratio [1]. Recent studies suggest that listeners are also able to use binaural cues to determine the range of lateral sound sources for distances within 1 m [2, 3, 4, 5, 6]. Binaural cues can hence be used to determine directions and distances of nearby sound sources. However, it is difficult to include distance information on available far field HRTFs. The simplest approximation uses a head-sized sphere to model distance variations [7]. Better approximations require to solve an acoustic propagation problem using demanding regularization techniques [8, 9, 10].

We alternatively propose a method to edit distance information in the spherical harmonics encodings of distant sources. Our method is intended to make sounds appear closer or farther than their original distance during its binaural rendering (see Figure 1). At the recording stage, we assume sound fields captured by a compact spherical microphone array. At the reproduction stage, we rely on the use of a surrounding distribution of virtual secondary monopole sources rendered with far field HRTFs. A discrete distribution of this kind of virtual sources rendered with HRTFs can be understood as an array of virtual loudspeakers [11]. Hence, we refer to this reproduction scheme as the virtual loudspeaker approach. To match the sound field at the central area in the virtual loudspeaker positions to the field in the microphone positions, we perform spherical re-samplings based on spherical harmonics and distance manipulations based on Hankel functions. Binaural signals are finally rendered using the distance-edited encodings.

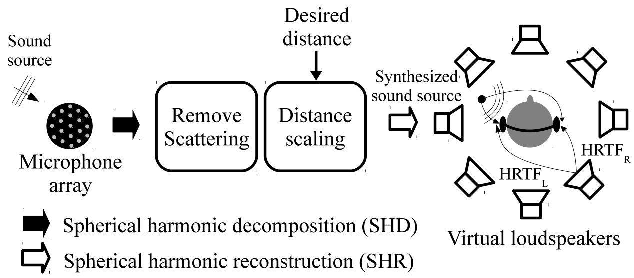

Figure 1: *Overview of the binaural synthesis method.*

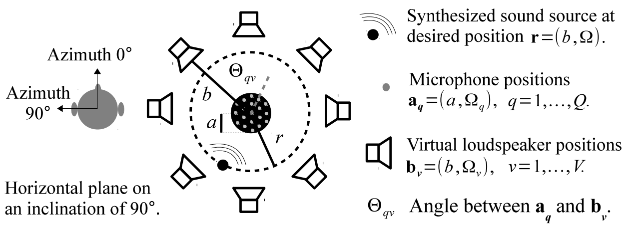

Figure 2: *Geometry used for the binaural synthesis method.*

A top view of the assumed geometry is shown in Figure 2. A point in space $\mathbf{r} = (r, \theta, \phi) = (r, \Omega)$ is specified by its distance r, its inclination $\theta \in [0°, 180°]$, and its azimuth $\phi \in [0°, 360°]$. The listener's ears lie on an inclination $\theta = 90°$. The front direction lie on an azimuth $\phi = 0°$.

Section 2 overviews sound field analysis and binaural synthesis techniques. Section 3 overviews the synthesis of HRTFs for arbitrary positions from continuously available far field HRTFs. Section 4 describes the continuous formulation of our proposal. Section 5 evaluates our proposal in a practical scenario, where microphones and virtual loudspeakers are placed on spherical samplings. Conclusions are presented in Section 6.

2. BINAURAL AMBISONICS

2.1. Binaural rendering from spherical harmonics encodings

The Schmidt semi-normalized spherical harmonics, of order n and degree m, are denoted by $Y_{nm}(\theta, \phi) = Y_{nm}(\Omega)$. They form an orthonormal basis for the set of square-integrable functions on the unit sphere $\mathbb{S}^2$. The sound pressure $f(\Omega)$ on the unit sphere is a function in this set. It can be expanded as [12]:

$$f(\Omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} f_{nm} Y_{nm}(\Omega), \tag{1}$$

where the coefficients f_{nm} define its spherical spectrum [12]:

$$f_{nm} = \int_{\Omega' \in \mathbb{S}^2} f(\Omega') Y_{nm}^*(\Omega') d\Omega'. \tag{2}$$

Eqs. (1) and (2) are respectively called the spherical harmonic reconstruction (SHR) and decomposition (SHD). A captured sound pressure field can thus be encoded with the SHD and decoded with the SHR. This defines the traditional High Order Ambisonics (HOA) format, a scalable way to render sound fields by decoupling the directions of the recording (Ω') and reproduction (Ω) setups.

Binaural reproduction of sound fields encoded by Eq. (2) is also possible. Encodings are decoded for a surrounding array of V virtual secondary sources using Eq. (1). The secondary source driving signals D_v derived in this way are then rendered with HRTFs H_v for the corresponding directions. Binaural signals B consist on superposing the resulting signals from all directions Ω_v:

$$B = \sum_{v=1}^{V} D_v H_v \alpha_v, \tag{3}$$

where D_v is decoded from existing encodings f_{nm} as follows:

$$D_v = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} f_{nm} Y_{nm}(\Omega_v), \tag{4}$$

and the normalization factor α_v is applied to the virtual loudspeakers, so that they cover almost equal areas.

2.2. Distance manipulation of multipole encodings

The multipole expansion extends Eq. (1) to include distance information. The pressure $g(\mathbf{r}) = g(r, \theta, \phi)$ on a sphere of radius r can be expanded by [12]:

$$g(\mathbf{r}) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} g_{nm} h_n(kr) Y_{nm}(\Omega), \tag{5}$$

where h_n is the spherical Hankel function, and the coefficients g_{nm} can also be derived from the pressure $g(\mathbf{r}')$ on a sphere of different radius r', as follows [12]:

$$g_{nm} = \frac{1}{h_n(kr')} \int_{\Omega' \in \mathbb{S}^2} g(\mathbf{r}') Y_{nm}^*(\Omega') d\Omega'. \tag{6}$$

Eqs. (5) and (6) are the basis for the treatment of distances in sound field analysis. They have been applied to recording and reproduction technologies like Near Field Compensated High Order Ambisonics (NFC-HOA) [13]. These equations relates the pressure on a recording sphere of radius r' and the pressure on a reproduction sphere of radius r. Binaural rendering with virtual loudspeakers can also be done in a similar way to Section 2.1. However, spherical harmonics encodings cannot be easily converted into the NFC-HOA format, since this requires determining the reference distance r' established during recording. Furthermore, NFC-HOA systems seek accurate reproduction, while some recordings may be enhanced by making sounds appear closer or farther than their original distance.

3. USING HRTFS FOR CONTINUOUS DISTRIBUTIONS OF FAR SOURCES

In this section, the ideal case where HRTFs are continuously available for distant sound sources at a fixed radius is introduced. In this continuous case, the binaural synthesis of nearby sound sources is formulated as an acoustic propagation problem. We do not consider reverberant fields. Hence, we use the term *far field* to refer to spherical sound fields for which HRTFs hardly depend on distance, that is, to sound sources beyond 1 m distance from the listener's head [2, 3, 4, 5, 6].

The Helmholtz' principle of reciprocity allows to formulate the measurement of HRTFs as an acoustic radiation problem [14]. Two original sound sources are assumed to be located at the listener's ears, and a measurement sphere of radius b is centered on the listener. Here, all the sources of scattering from the head and torso of the listener, together with the original sound source, all of them constitute the source field. When torso is not considered, the head's radius r_h is defined as the smallest sphere's radius containing the head, hence containing the source field too.

Given an initial set of HRTFs denoted by $H(\mathbf{b}, k)$, measured on the sphere $\mathbf{b} = (b, \Omega)$ enclosing the head for a source field of wave number k, the HRTFs denoted by $\hat{H}(\mathbf{r}, k)$ on any other sphere $\mathbf{r} = (r > r_h, \Omega)$ containing the source field are completely defined by the simple source formulation [12]:

$$\hat{H}(\mathbf{r}, k) = \int_{\Omega \in \mathbb{S}^2} G(\mathbf{r}, \mathbf{b}, k) H(\mathbf{b}, k) d\Omega, \tag{7}$$

where $G(\mathbf{r}, \mathbf{b}, k)$ are the Green functions of wave number k characterizing the sound transmission in free space, from all monopole sources located at $\mathbf{b}$ to each desired position $\mathbf{r}$.

Multipole expansions of the Green function in Eq. (7) has been used to synthesize HRTFs for arbitrary positions from the initial set of HRTFs at a single radius [8, 9, 10]. Accurate synthesis is obtained following this approach. However, the source positions of the initial set need to be distributed almost uniformly on the sphere, for the radiation problem is formulated on the spherical harmonics domain. Otherwise, the multipole expansions requires regularization techniques according to particular geometries, for which an appropriate selection of the regularization parameter can become a demanding task.

4. DISTANCE EMBEDDING FOR HIGH ORDER AMBISONICS WITH BINAURAL RENDERING

In this section, our alternative proposal to embed distance information on recordings of distant sound sources is described.

4.1. Using far field recordings by a rigid continuous sphere

An alternative approach to the multipole expansion of the Green functions assumes a surrounding and continuous distribution of monopole secondary sources. The surrounding secondary sources are placed at the same radius $b > 1$ m as the initial set of far field HRTFs and binaurally rendered with them. This reproduction scheme is called the virtual secondary source approach [11]. The requirement of expanding the Green functions $G(\mathbf{r}, \mathbf{b}, k)$ is thus relaxed, and the signals $D(\mathbf{r}, \mathbf{b}, k)$ to drive the virtual secondary sources are computed instead. The driving signals are typically derived so as to match the sound pressure field due to sound sources in the far field, on a radius $r > 1$ m, using sound field analysis techniques [15, 16]. However, sound field techniques typically decompose the sound field into plane waves, thus neglecting the distance-related effects, which may be important for a binaural rendering with high levels of realism. Our proposal follow the virtual secondary source approach considering the distance effects. In fact, we will derive the driving signals from far field recordings, but assuming point-like sources instead of plane waves. Therefore, in our proposal, the distance information can be further edited.

We next derive the signals to drive the continuous distribution of virtual secondary sources from a captured sound pressure field. By $D(\mathbf{r}, \mathbf{b}, k)$ we denote the driving signal of a virtual secondary source placed at $\mathbf{b}$, associated to a sound field generated by a sound source of wave number k placed at $\mathbf{r}$. In particular, we assume that the sound sources are on the same radius where the virtual secondary sources are continuously distributed ($r = b > 1$ m).

4.1.1. Spherical spectra of recordings and driving signals

On the recording side, we consider an ideal scenario where the pressure field is captured by a continuous, rigid and spherical sensing surface of radius a. In other words, the far field recordings, which we denote by $M(a, k)$, are available at an infinite number of points $\mathbf{a} = (a < b, \Omega')$. We characterize the recorded signals using the model of the acoustic scattering from the rigid sphere due to a point-like sound source. The total pressure on the surface of the rigid sphere reads [17]

$$S(\mathbf{a}, \mathbf{b}, k) = -\frac{1}{ka^2} \sum_{\nu=0}^{\infty} \frac{h_\nu(kb)}{h'_\nu(ka)} (2\nu + 1) P_\nu(\cos \Theta_{\mathbf{ab}}), \quad (8)$$

where $\Theta_{\mathbf{ab}}$ is the angle between the measurement point $\mathbf{a}$ and the source position $\mathbf{b}$, P_ν is the Legendre function, h_ν is the spherical Hankel function and h'_ν its derivative. In addition, we consider the recording spherical spectrum coefficients $S_{nm}(a, k)$, of order n and degree m, which reads [12]:

$$S_{nm}(a, k) = \int_{\Omega'} S(\mathbf{a}, \mathbf{b}, k) Y^*_{nm}(\Omega') d\Omega'. \quad (9)$$

On the virtual reproduction side, though, we assume driving signals whose spherical spectrum coefficients $D_{nm}(b, k)$ vanish for orders greater than N. Expansions of the driving signals in terms of spherical harmonics, evaluated in the secondary source directions Ω, are therefore defined by [12]

$$D(\mathbf{b}, k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} D_{nm}(b, k) Y_{nm}(\Omega). \quad (10)$$

The spherical harmonics encodings are independent of the decomposing directions. Hence, the spherical spectra of the recording and virtual reproduction signals can be related by means of propagating filters from the radius a to the radius b.

4.1.2. Filters on the spherical spectrum

By F_n we denote the distance propagation filters. To derive F_n, we replace $D_{nm}(b, k)$ in Eq. (10) by the product of F_n with the spherical spectrum $S_{nm}(a, k)$ of Eq. (9). We then proceed to use the orthonormality property of spherical harmonics, and their addition theorem to decompose the Legendre polynomials into a sum of spherical harmonics products [12]. Assuming infinite recording points, it can be shown that the driving signals become

$$D(\mathbf{r}, \mathbf{b}, k) = \sum_{n=0}^{N} \frac{-F_n h_n(kb)}{ka^2 h'_n(ka)} (2n + 1) P_n(\cos \Theta_{\mathbf{rb}}), \quad (11)$$

where $\Theta_{\mathbf{rb}}$ is the angle between the source position $\mathbf{r}$ and the virtual secondary source position $\mathbf{b}$.

The filters F_n in Eq. (11) are chosen in such a way that they compensate for the distance effects. These filters therefore read

$$F_n(a, b, k) = -ka^2 \frac{h'_n(ka)}{h_n(kb)}, \quad (12)$$

whose factors compensate for the scattering effects introduced by the rigid sphere of radius a, and propagate the recordings on the radius a to the radius b where the secondary sources are. These filters are typically used to capture sound fields with rigid spherical microphone arrays [12, 18].

4.1.3. Distance-embedding filters

In addition, the theory of acoustic holography [12] allows to compute the near field compensation filters $\frac{h_n(kb)}{h_n(kr)}$ to estimate the pressure field at a new distance r. The driving signals for an arbitrary distance r can therefore be synthesized by applying the filters

$$F_n(a, r, k) = \frac{h_n(kb)}{h_n(kr)} F_n(a, b, k) = -ka^2 \frac{h'_n(ka)}{h_n(kr)} \quad (13)$$

to the spherical spectrum of the far field recordings.

The filters proposed in Eq. (13) do not depend anymore on the distance b of the original sound source, as long as the original source is placed beyond 1 m distance from the center of the listener's head. According to the acoustic radiation problem in Section 3.1, the minimum desired distance r that can be synthesized is the radius r_h of the smallest sphere containing the listener's head.

Application of Eq. (13) to sound fields recorded by compact microphone arrays and encoded with spherical harmonics enables the binaural rendering of sound sources at any distance $r > r_h$. We call this method the Distance-Editing Binaural Ambisonics (DEBA) hereafter.

5. APPLICATION OF DEBA

We proceed now to formulate and evaluate DEBA in a practical scenario, where microphones and virtual secondary sound sources are placed in almost regular samplings of the sphere.

5.1. Using spherical microphone arrays

In practice, a finite number Q of microphones is used on the recording side. The microphones are assumed to be placed at discrete points $\mathbf{a}_q = (a, \Omega_q)$ on the spherical surface. We denote each microphone signal by $M(\mathbf{a}_q, k)$, which arises from the discretization of $M(\mathbf{a}, k)$. We replace D_{nm} in Eq. (10) by the product of F_n with a quadrature over q of the recording spherical spectrum $\int_{\Omega'} M(\mathbf{a}, k) Y^*_{nm}(\Omega') d\Omega'$. We proceed to use the addition theorem of spherical harmonics [12] to deal with relative directions. The signals to drive the continuous distribution of secondary sources, necessary to binaurally render nearby sound sources from the compact microphone array recordings, now read

$$D(\mathbf{r},\mathbf{b},k) = \sum_{n=0}^{N}(2n+1)F_n(a,r,k)\sum_{q=1}^{Q}P_n(\cos\Theta_q)M(\mathbf{a}_q,k)\beta_q, \quad (14)$$

where r is the desired distance, and Θ_q is the angle between the microphone at $\mathbf{a}_q$ and the virtual secondary source at $\mathbf{b}$. In particular, we considered almost constant integration quadratures β_q.

5.2. Using actual data sets of HRTFs

Measured sets of HRTFs are generally available for only some surrounding source positions at a fixed radius on the far field. Their spatial resolution is generally lower that the minimum audible angle of human auditory perception [19, 20]. To implement DEBA with such HRTF data set, an integral over the surface of the unit sphere similar to Eq. (7) need to be approximated by a weighted sum of a finite number of initial far field HRTFs. We refer to this kind of discrete distributions of secondary sources as virtual loudspeaker arrays. We therefore assume a finite number V of virtual loudspeakers placed at discrete points $\mathbf{b}_v = (b, \Omega_v)$ on the far field. We denote by $B(\mathbf{r}, k)$ the binaural signals for a desired position $\mathbf{r}$. Hence, the binaural signals are synthesized as follows:

$$B(\mathbf{r},k) = \sum_{v=1}^{V} D(\mathbf{r},\mathbf{b}_v,k)H(\mathbf{b}_v,k)\alpha_v, \quad (15)$$

where α_v is the normalized quadrature weight that approximates the differential $d\Omega$ at each sampled point $\mathbf{b}_v$. In particular, we will use quadrature weights that are proportional to the area of each sampled point's neighborhood. We define the neighborhood of a sample as all points on the sphere that are closer to it than to other samples.

The driving signal $D(\mathbf{r}, \mathbf{b}_v, k)$ in Eq. (15) arises from the discretization of Eq. (14). The driving signal for a virtual loudspeaker at $\mathbf{b}_v$ intended to render binaurally nearby sources from the microphone array recordings finally reads

$$D(\mathbf{r},\mathbf{b}_v,k) = \sum_{n=0}^{N}(2n+1)F_n(a,r,k)\sum_{q=1}^{Q}P_n(\cos\Theta_{qv})M(\mathbf{a}_q,k)\beta_q, \quad (16)$$

where $F_n(a, r, k)$ is the distance-embedding filter of Eq. (13), and Θ_{qv} now represents the angle between the microphone position $\mathbf{a}_q$ and the virtual loudspeaker position $\mathbf{b}_v$.

The filters F_n in Eq. (13) show high gains at low frequencies and high orders n, specially when using a rigid sphere of small radius a. In order to avoid low frequency distortion, spatial modes and frequencies are typically related. Hence, the reconstruction order N was chosen according to the wave number k and the scatterer size a as proposed in [20]:

$$N = \min(\lceil \frac{eka}{2} \rceil, \lfloor \sqrt{Q} - 1 \rfloor), \quad (17)$$

where e is the base of the natural logarithm and the number of microphones Q imposes the upper limit to the order.

Virtual loudspeakers should be placed on regular samplings of the sphere to avoid spatial aliasing. Regular spherical samplings,

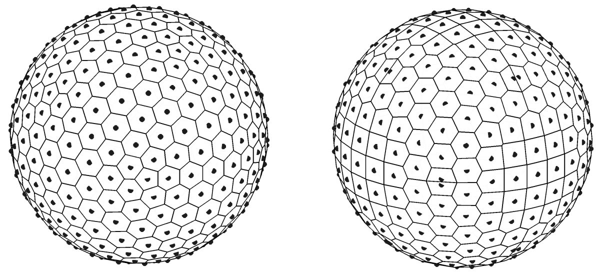

Figure 3: *Spherical grids to distribute the virtual loudspeakers.*

though, are only possible for the platonic solids. Among existing almost-regular samplings of the sphere, we have chosen the constructions based on the octahedron and the icosahedron. Icosahedral grids are constructed by subdividing the icosahedron's edges. They provide almost constant quadrature weights. In contrast, octahedral grids are constructed so to have octahedral rotation and inversion symmetry. They provide exact quadratures for numerical integration on the sphere [21] and, therefore, are suitable for computations with spherical harmonics. Figure 3 shows examples of icosahedral and octahedral grids, where dots indicate the positions of virtual loudspeakers and the lines enclose their neighborhoods.

5.3. Conditions for the evaluation of the numerical accuracy

We need to know the effect of the number of virtual loudspeakers on the synthesis accuracy. For this purpose, microphone signals denoted by $M(\mathbf{a}_q, k)$ were characterized with Eq. (8) and the algorithm provided in [22]. The microphone signals correspond to 360 far field sound sources equiangularly distributed on the horizontal plane at a radius $b = 1.5$ m. Initial sets of far field HRTFs denoted by $H(\mathbf{b}_v, k)$ were computed numerically for a dummy head using the Boundary Element Method (BEM) [23]. The sound sources used to compute the far field HRTFs were arranged on icosahedral and octahedral grids, at a radius $b = 1.5$ m. Transfer functions for the whole binaural synthesis process, denoted by $B(\mathbf{r}, k)$, were therefore characterized by using Eqs. (15) and (16), for several frequencies and desired positions in the horizontal plane. A reference set of near-field HRTFs, denoted by $H_{ref}(\mathbf{r}, k)$, was also numerically computed using BEM. The resulting transfer functions for the whole binaural synthesis process were finally compared with the reference near-field HRTFs.

For each desired distance r, accuracy along azimuth θ was calculated by means of the spectral distortion (SD), defined by the logarithmic spectral distance between $H(\theta, f)$ and $B(\theta, f)$ [24]:

$$SD(\theta) = \left(\frac{1}{I}\sum_{i=1}^{I}\left(20\log_{10}\left|\frac{H_{ref}(\theta,f_i)}{B(\theta,f_i)}\right|\right)^2\right)^{\frac{1}{2}}. \quad (18)$$

Also for each desired distance r, accuracy along frequency f was calculated by the normalized spherical correlation (SC) between $H(\theta, f)$ and $B(\theta, f)$ [10]:

$$SC(f) = \frac{\sum_{j=1}^{J} H_{ref}(\theta_j,f)B(\theta_j,f)}{\left(\sum_{j=1}^{J} H^2_{ref}(\theta_j,f)\right)^{\frac{1}{2}}\left(\sum_{j=1}^{J} B^2(\theta_j,f)\right)^{\frac{1}{2}}}. \quad (19)$$

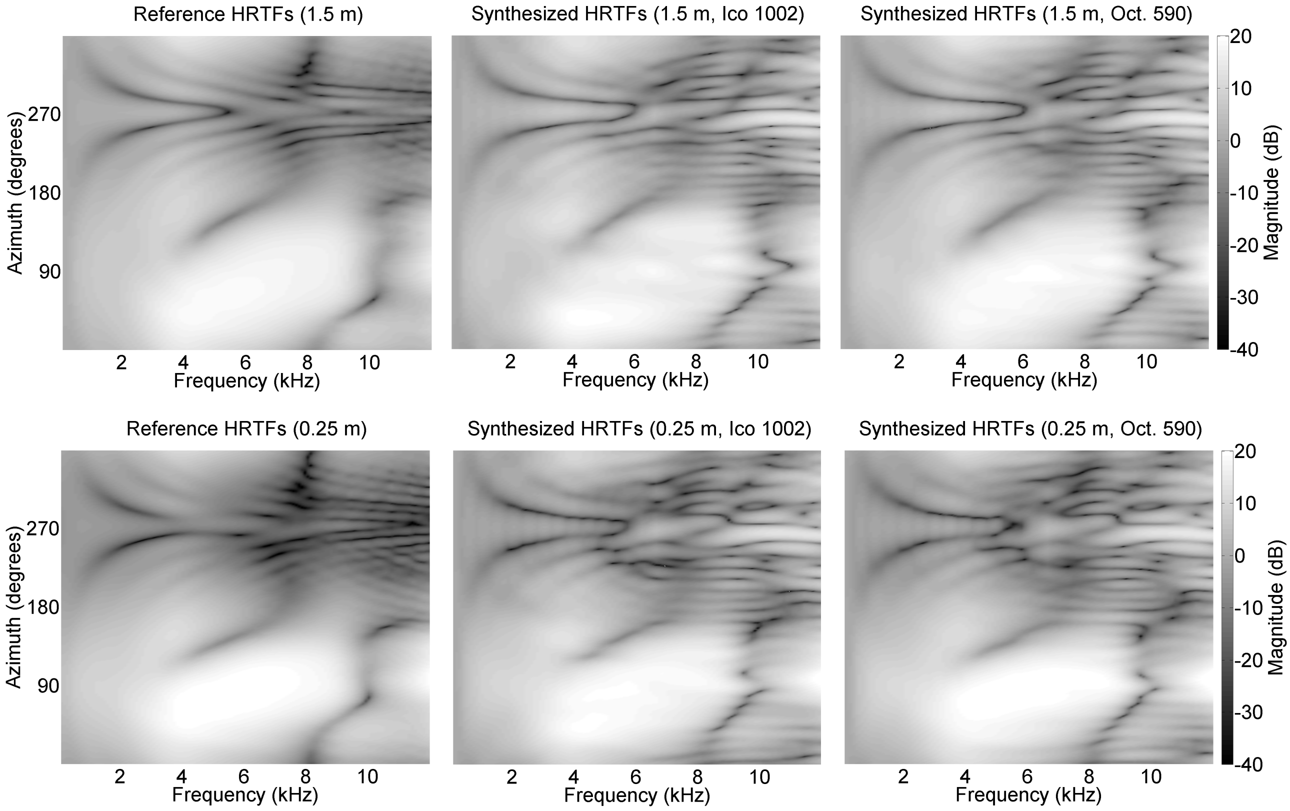

Figure 4: *Reference and synthesized HRTFs for distant (top) and nearby (bottom) sources on the horizontal plane (inclination of 90°). Reference HRTFs were numerically computed for a dummy head (left). We assumed 252 microphones and, hence, a spherical harmonics decomposition of order N = 14. Synthesis was performed with 1002 virtual loudspeakers on an icosahedral grid (middle) and 590 virtual loudspeakers on a octahedral grid (right), in both cases at a 1.5 m distance. These numbers of virtual loudspeakers correspond to the best accuracies (see Figure 5).*

We assumed microphones placed on a spherical scatterer of $a = 8.5$ cm radius, which we consider is the size of an average human head. According to [20], binaural synthesis in the entire audible frequency range, from 20 Hz to 20 kHz, would require an order $N = 43$, and therefore, a recording array of at least $Q = (43+1)^2 = 1936$ microphones. However, the practical number of microphones in existing compact arrays imposes a limited spatial bandwidth. At this stage, our evaluations were particularly focused on the recording setup available at the Research Institute of Electrical Communication in Tohoku University [25]. We therefore assumed $Q = 252$ microphones distributed in an icosahedral grid over the scatterer of $a = 8.5$ cm radius. This allowed for spherical harmonic expansions up to an order $N = 14$, and hence, accurate synthesis was only expected up to a spatial aliasing frequency of around 6.7 kHz.

5.4. Accuracy evaluation by computer simulations

Figure 4 shows some examples of HRTFs synthesized for the left ear and sound sources on the far (top panels) and near (bottom panels) regions. A visual comparison with the reference HRTFs on the right panels shows that the synthesis for sound sources placed on the same side of the ear (azimuth from 0° to 180°) can be performed with good accuracy up to around 8 kHz. Nevertheless, clearly decreasing accuracies appear for sound sources placed on the opposite side of the ear (azimuth from 180° to 360°). We noticed that the low-order spherical harmonics expansion does not yield a good approximation for the HRTFs. This was specially noticed for sound sources on the contralateral side of the ear, where signals of rapid variations along frequency and azimuths are caused by the head shadowing. In addition, discontinuity lines at some frequencies were due to the order limitation set by Eq. (17). Discontinuities are more prominent on the contralateral side and for desired distances near the head. On the other hand, slight decreasing accuracies appeared for distant and nearby sound sources of frequencies below 1.5 kHz. These particular observations suggested to focus the spectral distortion evaluations along azimuth on the contralateral side, and the spherical correlation evaluations along frequencies below 1.5 kHz.

Figure 5 shows the results of the numerical accuracy evaluation of the binaural synthesis performed with virtual loudspeakers on icosahedral (left panels) and octahedral (right panels) grids. The top panels show the mean values of the spectral distortion for sound sources on the opposite side of the left ear, along azimuths from 180° to 360° and frequencies below 8 kHz. The bottom panels show the mean values of the spherical correlation along all azimuths and frequencies below 1.5 kHz. Spectral distortions for contralateral sound sources yielded monotonically decreasing ac-

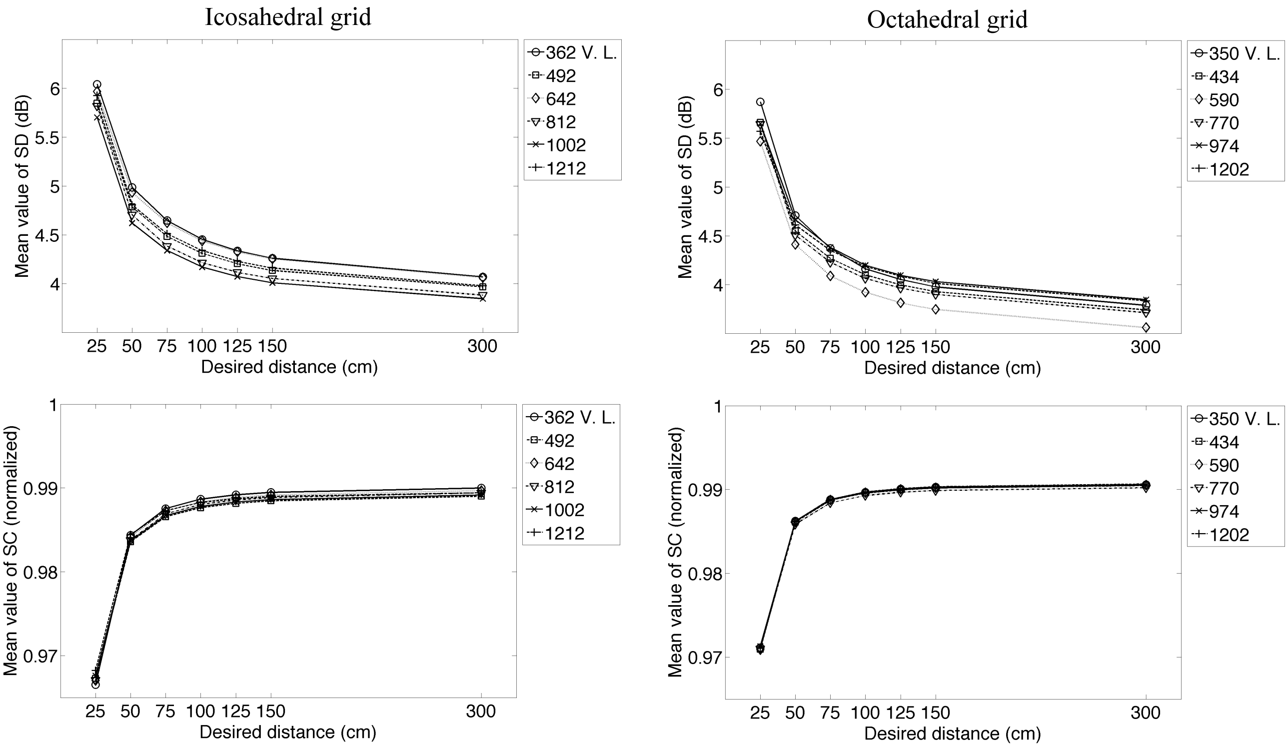

Figure 5: *Mean values of the spectral distortions on the contralateral side (top) and spherical correlations below 1.5 kHz (bottom) between the reference and synthesized HRTFs. Virtual loudspeakers (V. L.) were arranged on icosahedral (left) and octahedral (right) grids.*

curacies with decreasing desired distance. Regarding the number of virtual loudspeakers, the best accuracies were obtained using 1002 points in icosahedral grids and 590 points in octahedral grids, for which common minimums clearly appeared at all distances. On the other hand, spherical correlations for low frequency sound sources showed that accuracy is not affected by the number of virtual loudspeaker, but decreases monotonically with the desired distance. In general, evaluation using the spectral distortion yielded mean values below 6 dB, and using the spherical correlation, mean values above 0.97.

Our simulations were based on the addition theorem of spherical harmonics and, therefore, we did not consider the effects of matrix inversion based on regularization techniques, which are commonly applied in existing implementations of sound field encoding and decoding techniques [13]. In addition, typical sets of HRTFs are measured for non-uniform distributions of sound sources, making it necessary to use regularization techniques to match the virtual loudspeaker signals to sound field recordings. Although at this stage our evaluations were focused on the number of uniformly distributed loudspeakers, an extended study would require to add regularization techniques.

6. CONCLUSIONS

We proposed DEBA (Distance Editing Binaural Ambisonics), a method to synthesize the binaural signals at arbitrary sound source positions. We synthesized the binaural signals from the recordings made with microphones placed on the surface of a rigid sphere. For this purpose, we considered a surrounding array of virtual loudspeakers driven with head-related transfer functions. DEBA can accurately synthesize binaural signals due to sound sources placed on the horizontal plane. Accurate synthesis is possible up to the spatial aliasing limit imposed by the use of a finite number of microphones.

For evaluation, we relied on spherical harmonics encodings derived from the computer simulation of a compact, spherical microphone array. Transducers for both, recording and reproduction arrays were positioned in almost regular samplings of the sphere. Transfer functions for the whole process were characterized and compared with a set of near-field HRTFs computed numerically for a dummy head. Comparisons using the spectral distortion yielded mean values below 6 dB, and using the spherical correlation, mean values above 0.97. The accuracy cannot be improved by increasing the number of loudspeakers beyond the spatial aliasing limit imposed by the number of microphones. For lateral sources below 1 kHz, the accuracy decreased monotonically as the synthesized sound sources approaches the listener's head.

7. ACKNOWLEDGMENTS

This study was supported by Grant-in-Aid of JSPS for Scientific Research (no. 24240016), the Foresight Program for "Ultra-realistic acoustic interactive communication on next-generation Internet", and the Cooperative Research Project Program of RIEC Tohoku University (H24/A14). The authors wish to thank Makoto Otani for his efforts in developing the BEM solver used to generate the reference HRTF data.

8. REFERENCES

[1] Georg von Békésy, *Experiments in hearing*, McGraw-Hill, New York, NY, USA, 1960.

[2] M. Morimoto, Y Ando, and Z Maekawa, "On head-related transfer function in distance perception," in *Proceedings of the Congress of the Acoustical Society of Japan*, Japan, 1975, pp. 137–138, (in Japanese).

[3] Douglas S. Brungart and William M. Rabinowitz, "Auditory localization of nearby sources. hear-related transfer functions," *Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1465–1479, Sept. 1999.

[4] David R. Moore and Andrew J. King, "Auditory perception: The near and far of sound localization," *Current Biology*, vol. 9, no. 10, pp. R361–R363, May 1999.

[5] Hae-Young Kim, Yôiti Suzuki, Shouichi Takane, and Toshio Sone, "Control of auditory distance perception based on the auditory parallax model," *Applied Acoustics*, vol. 62, no. 3, pp. 245–270, Mar. 2001.

[6] Pavel Zahorik, Douglas S. Brungart, and Adelbert W. Bronkhorst, "Auditory distance perception in humans: A summary of past and present research," *Acta Acustica united with Acustica*, vol. 91, pp. 409–420, 2005.

[7] Alan Kan, Craig Jin, and Andre van Schaik, "A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function," *Journal of the Acoustical Society of America*, vol. 125, no. 4, pp. 2233–2242, Apr. 2009.

[8] Ramani Duraiswami, Dmitry N. Zotkin, and Nail A. Gumerov, "Interpolation and range extrapolation of HRTFs," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 2004, vol. 4, pp. 45–48.

[9] Wen Zhang, Thushara Abhayapala, and Rodney A. Kennedy, "Insights into head-related transfer function: spatial dimensionality and continuous representation," *Journal of the Acoustical Society of America*, vol. 127, no. 4, pp. 2347–2357, Apr. 2010.

[10] Martin Pollow, Khoa-Van Nguyen, Olivier Warusfel, Thibaut Carpentier, Markus Müller-Trapet, Michael Vorländer, and Markus Noisternig, "Calculation of head-related transfer functions for arbitrary field points using spherical harmonics," *Acta Acustica united with Acustica*, vol. 98, no. 1, pp. 72–82, Jan. 2012.

[11] Jean-Mar Jot, Scott Wardle, and Veronique Larcher, "Approaches to binaural synthesis," in *Audio Engineering Society 105th Convention*, Paris, France, Sept. 1998, Audio Engineering Society.

[12] Earl G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*, Academic Press, London, UK, 1999.

[13] Jerome Daniel, "Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format," in *Audio Engineering Society Conference: 23rd International Conference: Signal Processing in Audio Recording and Reproduction*, Denmark, May 2003.

[14] Dmitry N. Zotkin, Ramani Duraiswami, Elena Grassi, and Nail A. Gumerov, "Fast head-related transfer function measurement via reciprocity," *Journal of the Acoustical Society of America*, vol. 120, no. 4, pp. 2202–2215, Oct. 2006.

[15] Markus Noisternig, Markus Sontacchi, Alois Musil, and Robert Holdrich, "A 3D ambisonic based binaural sound reproduction system," in *Audio Engineering Society 24th International Conference: Multichannel Audio, The New Reality*, Graz, Austria, June 2003, Audio Engineering Society.

[16] Sascha Spors and Jens Ahrens, "Generation of far-field head-related transfer functions using virtual sound field synthesis," in *German Annual Conference on Acoustics (DAGA)*, Mar. 2011.

[17] J. J. Bowman, T. B. A. Senior, and P.L.E. Uslenghi, *Electromagnetic and acoustic scattering by simple shapes*, Hemisphere, New York, NY, USA, 1987.

[18] Jens Meyer and Gary Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *Proceedings of the 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Orlando, FL, USA, May 2002, vol. II, pp. 1781–1784.

[19] A. W. Mills, "On the minimum audible angle," *Journal of the Acoustical Society of America*, vol. 30, no. 4, pp. 237–246, Apr. 1958.

[20] Wen Zhang, Mengqiu Zhang, Rodney A. Kennedy, and Thushara Abhayapala, "On high-resolution head-related transfer functions measurements: an efficient sampling scheme," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, no. 2, pp. 575–584, Feb. 2012.

[21] V. I. Lebedev, "Quadratures on a sphere," *{USSR} Computational Mathematics and Mathematical Physics*, vol. 16, no. 2, pp. 10 – 24, 1976.

[22] Richard O. Duda and William L. Martens, "Range dependence of the response of a spherical head model," *Journal of the Acoustical Society of America*, vol. 104, no. 5, pp. 3048–3058, Nov. 1998.

[23] Makoto Otani and Shiro Ise, "Fast calculation system specialized for head-related transfer function based on boundary element method," *Journal of the Acoustical Society of America*, vol. 119, no. 5, pp. 2589–2598, May 2006.

[24] Takanori Nishino, Naoya Inoue, Kazuya Takeda, and Fumitada Itakura, "Estimation of HRTFs on the horizontal plane using physical features," *Applied Acoustics*, vol. 68, pp. 897–908, Feb. 2007.

[25] Shuichi Sakamoto, Satoshi Hongo, R. Kadoi, and Yôiti Suzuki, "SENZI and ASURA: new high-precision sound-space sensing systems based on symmetrically arranged numerous microphones," in *Proceedings of the Second International Symposium on Universal Communication*, 2008, pp. 429–434.