

EVALUATION OF A NEW AMBISONIC DECODER FOR IRREGULAR LOUDSPEAKER ARRAYS USING INTERAURAL CUES

Jorge TREVINO^{1,2}, Takuma OKAMOTO^{1,3}, Yukio IWAYA^{1,2} and Yôiti SUZUKI^{1,2}

({jorge,okamoto}@ais.riec.tohoku.ac.jp,{iwaya,yoh}@riec.tohoku.ac.jp)

¹ Research Institute of Electrical Communication, Tohoku University

² Graduate School of Information Sciences, Tohoku University

³ Graduate School of Engineering, Tohoku University

Abstract: Conventional ambisonic decoders do not work well with irregular loudspeaker configurations. Distributions of loudspeakers inadequately sampling the spherical harmonic functions lead to ill-conditioned re-encoding matrices and, therefore, decoders that exhibit numerical instability. Furthermore, when the number of loudspeakers in the array exceeds the number of ambisonic channels, simple decoders relying on the Moore-Penrose pseudo-inverse can lead to solutions that are not optimal for a human listener. To tackle these problems, we have proposed a new ambisonic decoder. Our decoding scheme, when compared with the pseudo-inverse approach, exhibits a slight drop in the accuracy with which the sound field is re-created at the sweet spot; however, it maintains low error levels throughout a wider listening region. We evaluate the performance of our proposed decoder using two physical variables that are important for sound localization: interaural level difference (ILD) and interaural phase difference (IPD). Simulation results using HRTF measurements show that our decoder can more accurately convey interaural cues to a human listener; particularly when an irregular loudspeaker array is used.

Key words: Ambisonics, decoding of ambisonics, irregular loudspeaker array, regularization, sound field reproduction, virtual loudspeaker array, binaural rendering, spatial sound, spherical acoustics

1 INTRODUCTION

In ambisonics[1], sound fields are encoded using the spherical harmonic decomposition of the sound pressure field observed on a spherical boundary. The decomposition is truncated to an arbitrary degree called the *ambisonic order*, which we will denote with the symbol N . The expansion coefficients captured by ambisonics can be used to reproduce sound fields using almost any surrounding loudspeaker array. Accurate reconstruction, however, depends on the loudspeaker positions defining a regular sample of the spherical boundary surrounding the listener. A decoding stage generates the loudspeaker signals from the expansion coefficients.

Mainstream ambisonic decoders rely on the Moore-Penrose pseudo-inverse of a re-encoding matrix:

$$\mathbf{B}(k) = \mathbf{C}\mathbf{p}(k), \quad (1)$$

where $\mathbf{B}(k)$ is the vector of ambisonic signals, shown here in the frequency domain as functions of the wavenumber k . The components of vector $\mathbf{p}(k)$ are the loudspeaker signals. The re-encoding matrix \mathbf{C} , therefore, has $(N + 1)^2$ rows and one column for every loudspeaker in the array. The elements of \mathbf{C} are given by the spherical harmonic functions Y_{mn} [2] evaluated in the directions of the loudspeakers, (θ_s, φ_s) , as follows:

$$c_{m^2+m+n,s} = Y_{mn}(\theta_s, \varphi_s). \quad (2)$$

The loudspeaker signals needed to reconstruct a particular ambisonic-encoded sound field can be computed by inverting the linear system of Eq. (1). For the re-encoding matrix to be invertible, however, the number of loudspeakers in the array must match the count of ambisonic channels. In practice, it is desirable to use larger arrays to improve the reproduction accuracy; this leads to an underdetermined linear system. It is common to rely on the Moore-Penrose pseudo-inverse to invert the re-encoding matrix. The decoding equation can be written in terms of the pseudo-inverse of \mathbf{C} , denoted by \mathbf{C}^+ , as [3]

$$\mathbf{p}(k) = \mathbf{C}^+\mathbf{B}(k). \quad (3)$$

Our previous research highlights two drawbacks of this kind of conventional ambisonic decoders: numerical instability and suboptimal solutions[4]. Numerical instability is especially problematic when a decoder based on the pseudo-inverse is used to re-create sound fields using an irregular loudspeaker array. Round-off errors can cause some of the spherical harmonic functions to become almost indistinguishable if they are inadequately sampled. In addition, the pseudo-inverse does not guarantee the optimal presentation

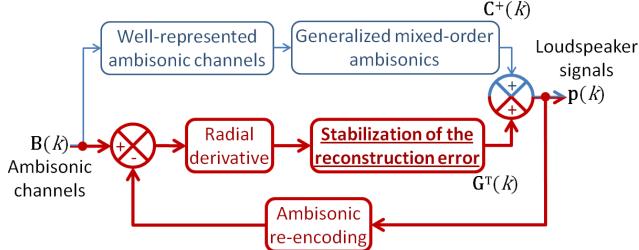


Figure 1: Block diagram of the new ambisonic decoder for irregular loudspeaker arrays. Our main innovation is summarized by the lower loop.

of sound to human listeners if the number of loudspeakers in the array exceeds the number of ambisonic channels. Decoders based on the pseudo-inverse yield the solution that achieves perfect reconstruction at a listening point and minimizes the output power of the loudspeakers. Human listeners, however, require an extended listening region and would benefit more from a decoder that attempts to mitigate reconstruction artifacts in the neighborhood of the listening position. A comfortable listening experience should allow the listener to move within a reasonably large region without an audible drop in sound quality.

2 A NEW APPROACH TO AMBISONIC DECODING

Our previous research introduced a new method to decode ambisonic data for reproduction over irregular loudspeaker arrays[4]. The new method, summarized in Fig. 1, attempts to mitigate any reconstruction artifacts throughout a wide listening region. Our proposal can be decomposed in two parts: a generalization of the mixed-order ambisonics technique and a regularization term that attempts to smooth out the error field within the listening area.

The mixed-order ambisonics technique consists of ignoring the expansion coefficients corresponding to spherical harmonic functions that are not adequately sampled by the positions of the loudspeakers. Decoding for arrays with different horizontal and vertical spatial resolutions can be achieved by first considering the full multipole expansion up to a low degree, and then complementing it with the expansion coefficients corresponding to the horizontally oriented spherical harmonics of higher degrees. A trivial extension is to discard only the ambisonic channels that are not well-represented in the array, irrespective of their relation to horizontal spherical harmonics. A useful heuristic is to rotate the loudspeaker array using the point group transformations (i.e. rotations by a fixed angle $\theta_r = 360^\circ/n$). The spherical harmonic functions of a given degree are invariant under some of these transformations, thus, an array that samples them uniformly should preserve this symmetry to a certain extent. Alternatively, one can exhaustively calculate the *condition number* (the ratio between the largest and the smallest singular values) of the re-encoding matrix for all combinations of ambisonic channels and choose the

most complete, well-conditioned one.

The generalized mixed-order ambisonics approach is numerically stable and yields an initial set of decoding gains that can already approximate an ambisonic encoded sound field. The accuracy of this initial solution is highly dependent on the regularity of the array, since only the channels that are uniformly sampled by the loudspeaker positions were decoded. Rather than attempting to improve the reconstruction accuracy at the center of the array, our focus is on expanding the listening region.

Assuming ideal monopole radiators, the reconstruction error at the position \vec{r} can be written as

$$\epsilon(k, \vec{r}) = \tilde{\psi}(k, \vec{r}) - \phi(k, \vec{r}) - \sum_s \sum_{m=0}^N \sum_{n=-m}^m G_{mn}^s(k) \frac{e^{-ik|\vec{r}-\vec{r}_s|}}{|\vec{r}-\vec{r}_s|} Y_{mn}(\theta_s, \varphi_s), \quad (4)$$

where $\tilde{\psi}(k, \vec{r})$ represents the sound field encoded in the ambisonic stream, $\phi(k, \vec{r})$ stands for the sound field reconstructed through the generalized mixed-order ambisonics approach. The first sum runs over the loudspeakers in the array; the position of the s -th loudspeaker is given, in spherical coordinates, as $\vec{r}_s = (r_s, \theta_s, \varphi_s)$. The gains $G_{mn}^s(k)$ are initially set to zero, yielding only the mixed-order ambisonics approximation.

We now seek gains $G_{mn}^s(k)$ that result in an expansion of the listening region. The behavior of the reconstruction error as the listening point moves away from the center of the array can be described by the radial derivative of Eq. (4)

$$\begin{aligned} \frac{\partial}{\partial r} \epsilon(k, \vec{r}) &= \nabla \epsilon(k, \vec{r}) \cdot \hat{r} \\ &= \nabla [\tilde{\psi}(k, \vec{r}) - \phi(k, \vec{r})] \cdot \hat{r} - \sum_s \frac{\partial}{\partial r} \left[\frac{e^{-ik|\vec{r}-\vec{r}_s|}}{|\vec{r}-\vec{r}_s|} \right] \\ &\quad \sum_{m=0}^N \sum_{n=-m}^m G_{mn}^s(k) Y_{mn}(\theta_s, \varphi_s). \end{aligned} \quad (5)$$

The first term is the radial derivative of the reconstruction error when using only the mixed-order ambisonics decoding method. This term can be regarded as a constant $\mathbf{d} \equiv \nabla[\tilde{\psi}(k, \vec{r}) - \phi(k, \vec{r})] \cdot \hat{r}$ since it is independent of the choice of gains $G_{mn}^s(k)$. The radial derivative of the monopole field can be expressed as:

$$\frac{\partial}{\partial r} \left[\frac{e^{-ik|\vec{r}-\vec{r}_s|}}{|\vec{r}-\vec{r}_s|} \right] = D_s(k) \left[\frac{e^{-ik|\vec{r}-\vec{r}_s|}}{|\vec{r}-\vec{r}_s|} \right], \quad (6)$$

with the operator

$$D_s(k) \equiv -\frac{|\vec{r}| - |\vec{r}_s| \cos(\vec{r}, \vec{r}_s)}{|\vec{r} - \vec{r}_s|} \left(\frac{1}{|\vec{r} - \vec{r}_s|} + ik \right). \quad (7)$$

Using these definitions, Eq. (5) can be rewritten as

$$\begin{aligned} \frac{\partial}{\partial r} \epsilon(k, \vec{r}) &= \mathbf{d} - \sum_s \sum_{m=0}^N \sum_{n=-m}^m \left[D_s(k) \frac{e^{-ik|\vec{r}-\vec{r}_s|}}{|\vec{r}-\vec{r}_s|} \right. \\ &\quad \left. Y_{mn}(\theta_s, \varphi_s) \right] G_{mn}^s(k). \end{aligned} \quad (8)$$

By taking the norm of Eq. (8), we can now impose the following constraint on the radial derivative of the reconstruction error:

$$\left| \frac{\partial}{\partial r} \epsilon(k, \vec{r}) \right| = \|\mathbf{L} \cdot \mathbf{G} - \mathbf{d}\| \leq \rho. \quad (9)$$

Here, ρ is some threshold limiting the permissible variation of the reconstruction error. The entries of \mathbf{G} are the loudspeaker gains $G_{mn}^s(k)$, while the entries of the operator \mathbf{L} are defined as

$$L_{mn}^s(k) = D_s(k) \frac{e^{-ik|\vec{r} - \vec{r}_s|}}{|\vec{r} - \vec{r}_s|} Y_{mn}(\theta_s, \varphi_s). \quad (10)$$

Equation (9) can be used to calculate a set of decoding gains by defining a target radius for the listening region and a maximum allowed variation for the reconstruction error. These gains can be used to generate loudspeaker signals through the following decoding equation:

$$\mathbf{p}(k) = [\mathbf{G}^T(k) + \mathbf{C}_{Mixed-order}^+] \mathbf{B}(k). \quad (11)$$

3 EVALUATION

The performance of our proposal has been previously compared with that of a conventional decoder using a 157-channel, irregular loudspeaker array[4]. Computer simulations of the array yield promising results; however, our previously presented analysis has focused exclusively on the physical reconstruction of sound fields. We now consider two perceptually meaningful parameters for sound localization and compare our proposed decoder with a conventional one.

Two fundamental parameters in sound localization are the interaural level difference (ILD) and interaural phase difference (IPD)[5]. Both parameters are accurately captured by a linear model like the head-related transfer function (HRTF). The present study uses the HRTF of a dummy head (SAMRAI, Koken Co., Ltd.) as measured at the Research Institute of Electrical Communication in Tohoku University. Measurements were made using a spherical loudspeaker array housed in an anechoic chamber. Our HRTF measurements span all azimuth angles from -175° to 180° in increments of 5 degrees, and are available for elevation angles between -80° and 90° in increments of 10 degrees.

HRTF measurements can be used to derive binaural signals from ambisonic data by defining a virtual loudspeaker array[6]. The resulting signal can be used to evaluate the accuracy of the sound field reconstruction as would be perceived by an actual, human listener.

We designed three virtual loudspeaker arrays and used them to reproduce fifth-order ambisonic data. The loudspeaker distributions include an almost regular, 42-channel configuration, a highly irregular, 42-channel one, and an approximation to the actual, 157-channel loudspeaker array used in our previous research[7].

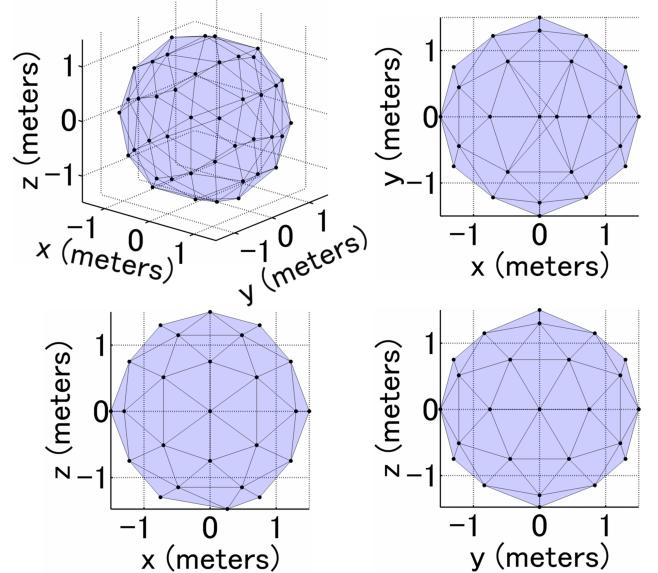


Figure 2: Distribution of loudspeakers for a regular, 42-channel loudspeaker array. The array can be used to reproduce 5th order ambisonic data and is very close to an ideal sampling of the sphere.

Several fifth-order ambisonic encodings of monopole sources were synthesized at different positions. All of the monopole sources were located at a distance of 1.5 meters. We considered azimuth angles between 0° and 90° and elevation angles between -20° and 90° . The monopoles radiated sound at fixed frequencies of up to 5 kHz. Two ambisonic decoders were used in our tests: a conventional ambisonic decoder based on the pseudo-inverse, and our new decoder for irregular loudspeaker arrays. The monopole sound fields were decoded and reproduced using the three virtual loudspeaker arrays. We then calculated the ILD and IPD from the resulting binaural recordings. The interaural cues extracted from our simulation were compared with those measured for the SAMRAI dummy head.

3.1. Regular 42-channel louspaker array

The first array to be considered in the present study consists of a regular distribution of 42 loudspeakers. It was designed from the solution to a simplified Thomson problem [8], constraining the charges to occupy only the positions for which SAMRAI's HRTF was available. The loudspeaker distribution for this array is shown on Fig. 2. The largest source of asymmetry comes from the lack of a measurement of the HRTF from below (elevation angle of -90°); a loudspeaker at an azimuth of 0° and elevation of -80° was used instead. The distance between the loudspeakers and the listener was fixed at 1.5 meters, the same distance at which SAMRAI's HRTF was measured.

A computer simulation of this virtual array was used to calculate the interaural level and phase differences that a listener inside the array would experience. The results were compared to the actual measurements of SAMRAI's HRTF. When the measurement at a particular direction was not

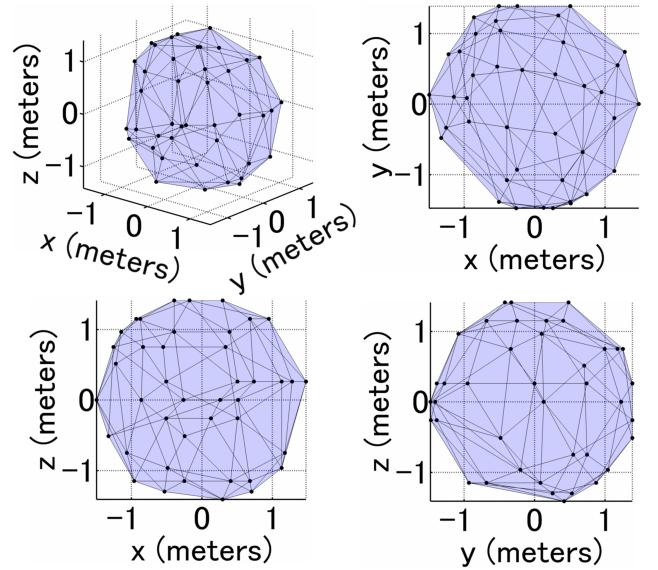
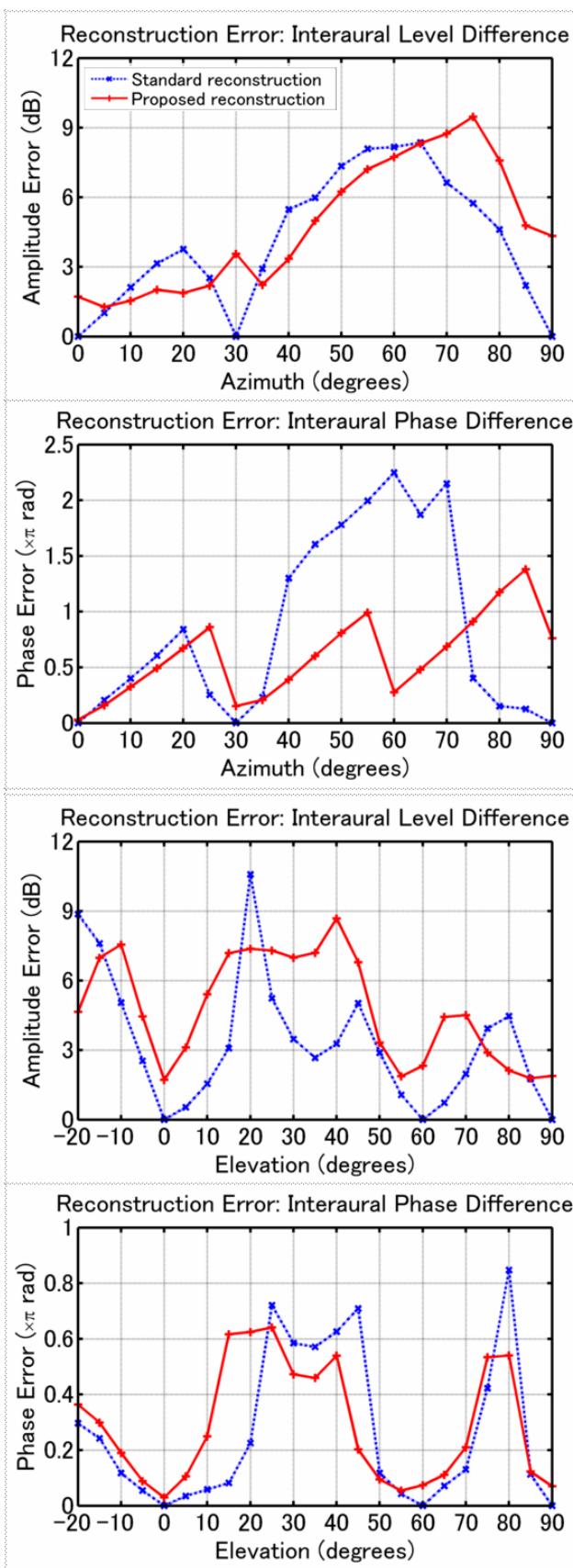


Figure 4: Distribution of loudspeakers for an irregular, 42-channel loudspeaker array. The array can be used to reproduce 5th order ambisonic data and its layout covers the full sphere; however, the angular spacing between contiguous loudspeakers is very irregular.

available, its nearest neighbor was used. Our results for this regular loudspeaker distribution, presented in Fig. 3, show that the conventional decoding method, predictably, re-creates the desired ILD more accurately than the new decoder for this kind of layouts. However, the difference between our proposal and the conventional decoder is not too significant, typically within 3 dB. Meanwhile, the results for the IPD show that our proposal can slightly outperform the conventional decoder even when a regular array is used.

3.2. Irregular 42-channel loudspeaker array

The second loudspeaker array used in the present evaluation exhibits a highly irregular distribution of 42 loudspeakers. The array, shown in Fig. 4, was designed to reproduce 5th order ambisonic data, retaining numerical stability even when a conventional decoder was used. It was also designed to sample all directions; however, care was taken to ensure that the array is not invariant under any point group transformations (except for the trivial identity operation) for all axes. The resulting array should be one of the worse designs that retain numerical stability for ambisonic reproduction.

The performance of the conventional decoder drops significantly when using an irregular loudspeaker array. The results, presented in Figs. 5, show that the ILD reconstruction error can exceed 20 dB when using the conventional decoder with the irregular 42-channel array. Conversely, the accuracy of our proposed decoder, designed specifically for irregular arrangements of loudspeakers, drops by less than 3 dB when changing to an irregular layout. Additionally, our decoder can reproduce sounds incoming from different directions to similar accuracy; in contrast, the performance of the standard ambisonic decoder varies widely for different angles of incidence.

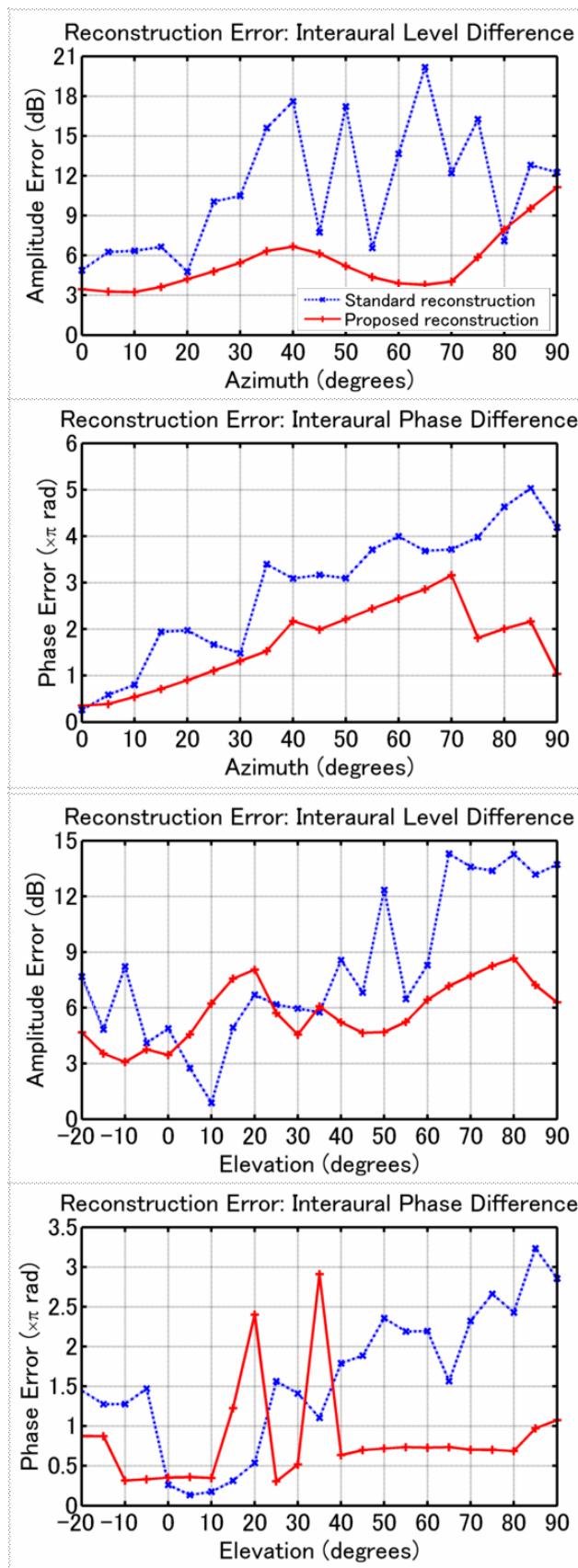


Figure 5: Error in the reconstruction of interaural cues for the irregular 42-loudspeaker array.



The proposed reconstruction method outperforms the standard approach for almost all directions, most of the time by a large margin. It is particularly notable that the ILD error achieved by the new decoding method does not vary too much between the regular and irregular configurations. As long as there are enough loudspeakers distributed throughout a surrounding sphere, the regularity of their distribution has little impact on the performance of the proposed decoder. In contrast, the standard decoding method suffers a large performance degradation when the layout of the array becomes irregular.

3.3. Irregular 157-channel loudspeaker array

Finally, an approximation to the irregular, 157-channel, surrounding loudspeaker array discussed in our previous research[7] was designed from the samples of SAMRAI's HRTF. The loudspeaker array being approximated is physically available in the Research Institute of Electrical Communication of Tohoku University and shown in Fig. 6. It is housed in a semi-anechoic room, covering its walls and ceiling at regular spatial intervals[9]. Each of the loudspeakers in the original array was mapped to the closest available HRTF sampling point after distance compensation. Since the positions of the loudspeakers are meant to sample the spherical harmonic functions, the best results are achieved by using the orthodromic distance rather than the Euclidean distance. The orthodromic distance, or central angle, between two points on the sphere (θ_1, φ_1) and (θ_2, φ_2) can be calculated using the Vincenty formula:

$$\Delta\hat{\sigma} = \arctan \left(\frac{\sqrt{A+B}}{C+D} \right), \quad (12)$$

where the symbols A , B , C and D are defined as follows:

$$A = (\cos \varphi_1 \sin(\theta_1 - \theta_2))^2,$$

$$B = (\cos \varphi_2 \sin \varphi_1 - \sin \varphi_2 \cos \varphi_1 \cos(\theta_1 - \theta_2))^2, \quad (13)$$

$$C = \sin \varphi_2 \sin \varphi_1,$$

$$D = \cos \varphi_2 \cos \varphi_1 \cos(\theta_1 - \theta_2).$$

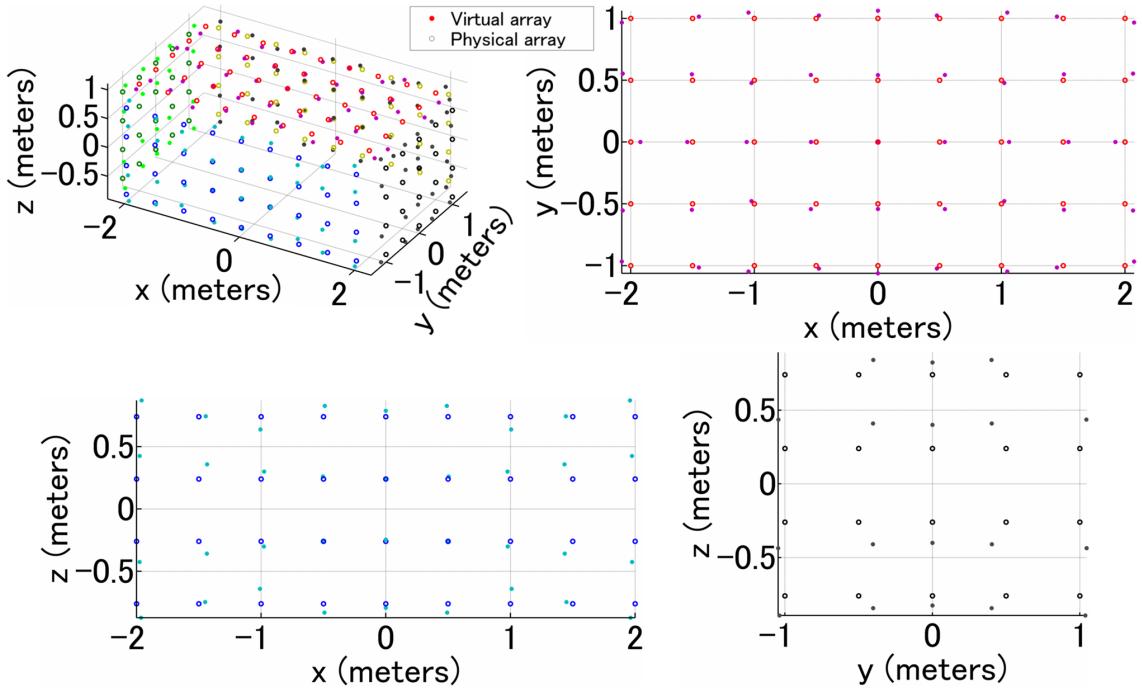


Figure 7: Distribution of loudspeakers for an irregular, 157-channel loudspeaker array. The virtual array, denoted by filled marks, provides the best fit between the actual 157-channel, surrounding loudspeaker array from our previous research[7] and the directions for which Samurai's HRTF was sampled. It can be used to reproduce 5th order ambisonic data, but its layout covers approximately one hemisphere. The distribution of loudspeakers has some regularity, but it is far from being a good sampling of all directions.

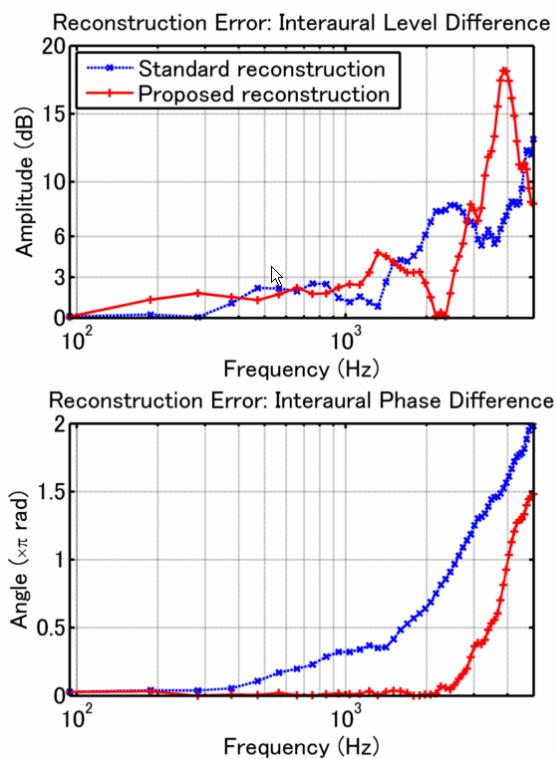


Figure 8: Error in the reconstruction of interaural cues for a monopole source radiating from the front.

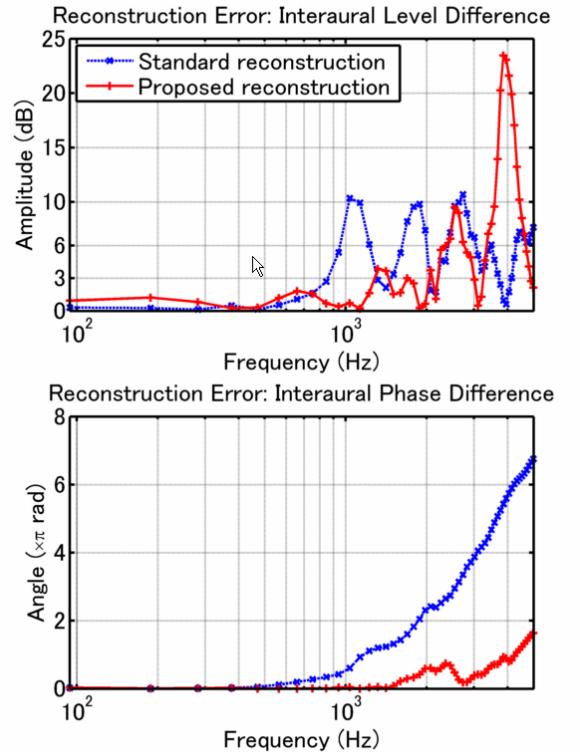


Figure 9: Error in the reconstruction of interaural cues for a monopole source radiating from an azimuth angle of 60° and elevation angle of 0°.

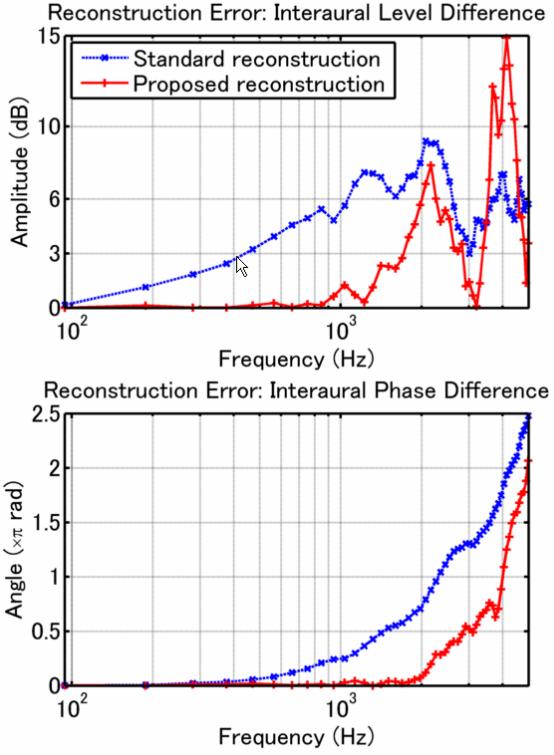


Figure 10: Error in the reconstruction of interaural cues for a monopole source radiating from an azimuth angle of 0° and elevation angle of 45° .

The best possible fit to the actual 157-channel loudspeaker array is shown in Fig. 7. The greatest error incurred in the approximation was of approximately 5.05° which, taking distance into account, results in a loudspeaker located around 20.7 cm away from its original position. The average error, however, was of approximately 2.96° and, after accounting for the distance to the loudspeakers, of about 10.88 cm.

The 157-channel loudspeaker array can be used to reproduce 5th order ambisonic recordings. This remains true even after the approximations made during the design of the virtual array using SAMRAI's HRTF measurements. However, unlike the previous two arrays, this array does not sample all directions. Instead, the 157-channel, irregular loudspeaker array is closer to a hemispherical distribution.

Unlike the previous results for the regular and highly irregular arrays, the 157-channel loudspeaker array exhibits a more varied behavior. Some directions privilege the conventional decoder, while others are better handled by the new decoding method. However, the general behavior for most directions is similar at frequencies below 2 kHz. Figures 8 and 9 show two representative cases of a sound source located at the front and at an azimuth of 60° and elevation of 0° , respectively. Both decoders can reconstruct the ILD with very high precision at low frequencies, although the standard decoder is slightly better. However, this is not a significant difference and the ILD is not an impor-

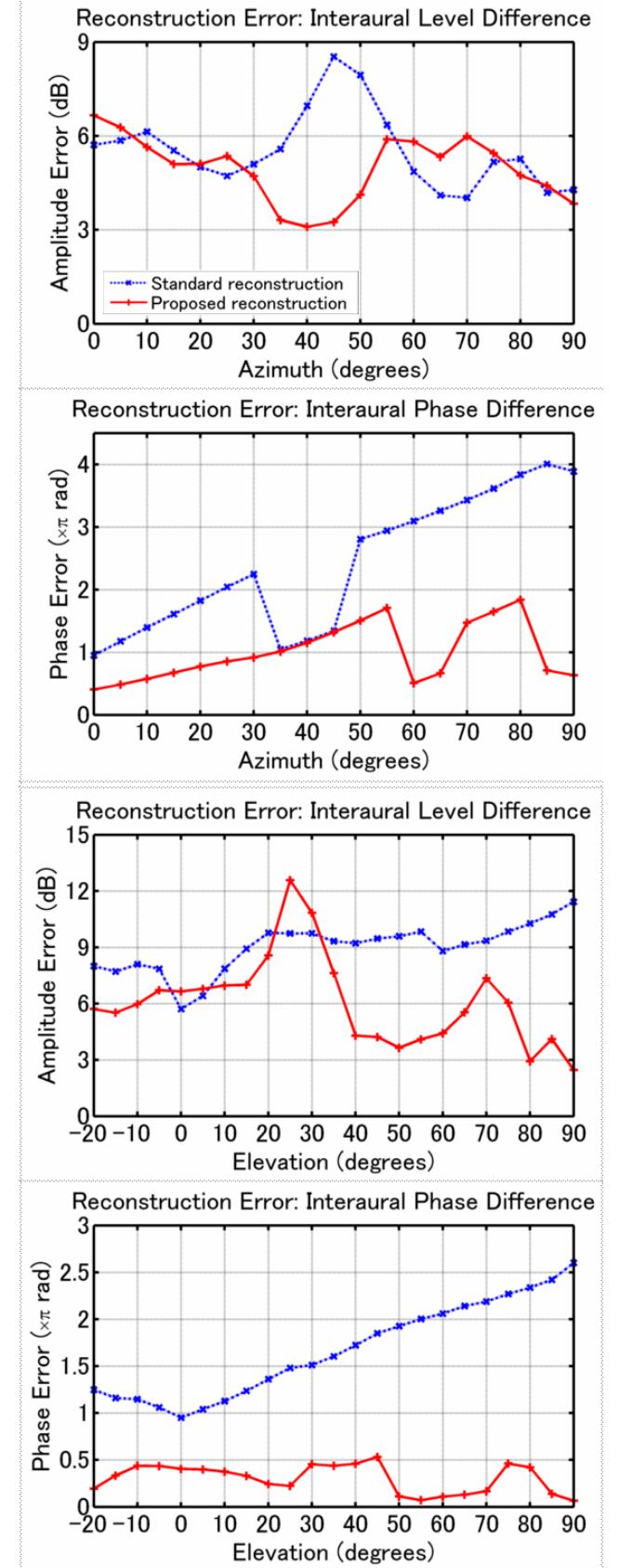


Figure 11: Error in the reconstruction of interaural cues for the irregular 157-loudspeaker array.

tant perceptual cue at low frequencies. As the frequency increases, the conventional decoder becomes less accurate rapidly, while the new decoding method retains low reconstruction errors up to around 2 kHz, a phenomenon that can be observed for most directions of incidence. At higher frequencies, both decoding methods become unreliable with no particular decoder outperforming the other in a significant way. Interaural phase difference is reconstructed to acceptable accuracy by both decoders, particularly at low frequencies where it is most significant. At higher frequencies, the proposed decoder produces more stable IPDs than the conventional one.

Our proposed decoder can also be seen to outperform the conventional method for sound sources radiating from non-zero elevations. An example of this, seen on Fig.10, shows that our proposal can achieve greater accuracy than the pseudo-inverse decoder even at low frequencies. Both decoders, however, are unreliable at frequencies beyond 2 kHz, with reconstruction errors for the ILD above 6 dB.

When the averages across all frequencies (up to 5 kHz) are considered, however, the proposed decoder shows a remarkably better performance than the conventional method. The results, shown on Fig. 11, show a pronounced difference in the ILD reconstruction error for azimuth angles between 30° and 55°. Not coincidentally, this region is also that for which the loudspeaker array exhibits greater irregularities in its layout, as it transitions from a densely sampled region at the front to a sparsely sampled one on the left side of the listener. The performance of the proposed decoder is also considerably better at elevation angles above 32°. This phenomenon is related to the hemispherical shape of the array, which cannot be properly handled by the pseudo-inverse decoder without some modifications to account for the lack of loudspeakers below the listener's level. Finally, a comparison of the IPD reconstruction error shows the proposed method to be more precise at practically all directions. The constrain on the radial derivative of the reconstruction error results in a drastic reduction of spatial alias within the listening region, which has a significant impact on the interaural phase difference.

4 CONCLUSION

Conventional ambisonic decoders work well with regular configurations; however, when an irregular loudspeaker array is used, our proposal shows a superior performance. We complemented our previous results, showing that our new decoder can re-create the physical variables over a wider region than conventional approaches, with an evaluation of perceptually meaningful parameters for human listeners. Our new decoder can accurately convey important interaural cues to the listener irrespective of the loudspeaker distribution.

The performance of our proposal is less dependent on the regularity of the loudspeaker array. Particularly promising results are observed when considering sources outside the horizontal plane and a hemispherical loudspeaker array.

Despite the improvement in both, physical reconstruction of the sound field and re-creation of interaural cues, reproduction of high frequency sources remains difficult due to the high amounts of information contained in their sound fields.

5 ACKNOWLEDGMENTS

This study was partly supported by the GCOE program (CERIES) of the Graduate School of Engineering, Tohoku University and Grant-in-Aid of JSPS for Specially Promoted Research (no. 19001004).

REFERENCES

- [1] D. Malham, "Higher order Ambisonic systems. Space in Music - Music in Space," *M.Phil. Thesis, University of York*, Apr. 2003.
- [2] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography, First Edition*, Academic Press, 1999.
- [3] J. Daniel, "Acoustic field representation, application to the transmission and the reproduction of complex sound environments in a multimedia context," *Ph.D. Thesis, Université Paris*, 2000.
- [4] J. Trevino, T. Okamoto, Y. Iwaya and Y. Suzuki, "High order Ambisonic decoding method for irregular loudspeaker arrays," *Proc. of the 20th International Congress on Acoustics*, 481 pp. 1-8, 2010.
- [5] J. Blauert, *Spatial Hearing*, MIT Press, Revised edition, 1996.
- [6] B. Wiggins, I. Paterson-Stephens, P. Schillebeeckx, "The analysis of multi-channel sound reproduction algorithms using HRTF data," *Proc. of the 19th AES International Conference*, 2001.
- [7] T. Okamoto, B.F.G. Katz, M. Noisternig, Y. Iwaya, Y. Suzuki, "Implementation of real-time room auralization using a surrounding loudspeaker array," *International Workshop on the Principles and Applications of Spatial Hearing (eProceedings)*, No. 10.1142-9789814299312-0049, Nov. 2009.
- [8] D. J. Wales, S. Ulker, "Structure and Dynamics of Spherical Crystals Characterized for the Thomson Problem," *Phys. Rev. B*, vol. 74, 212101, 2006
- [9] T. Okamoto, D. Cabrera, M. Noisternig, B.F.G. Katz, Y. Iwaya and Y. Suzuki, "Improving sound field reproduction in a small room based on higher-order Ambisonics with a 157-loudspeaker array," *Proc. 2nd Int. Symp. on Ambisonics and Spherical Acoust., Poster 5, May 2010*.