Evaluation of a new Ambisonic decoder for irregular loudspeaker arrays using interaural cues

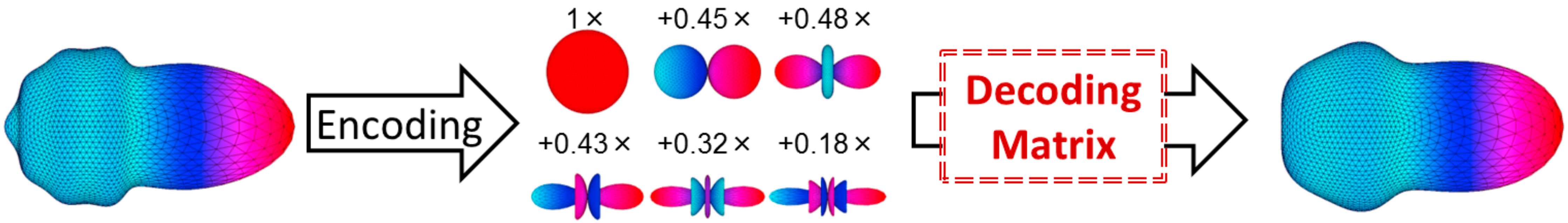
J. Treviño[1,2], T. Okamoto[1,3], Y. Iwaya[1,2] and Y. Suzuki[1,2]

1) Research Institute of Electrical Communication, Tohoku University, Japan
2) Graduate School of Information Sciences, Tohoku University, Japan
3) Graduate School of Engineering, Tohoku University, Japan

Motivation

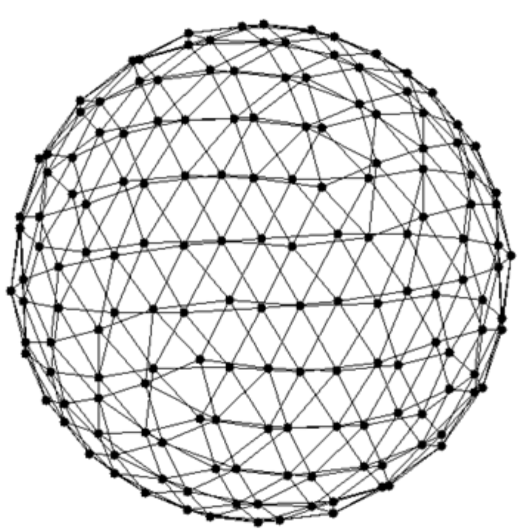
- Implementation of 3D audio need for systems that can convey the illusion of being immersed in a different environment
  - Telepresence, communications, entertainment, etc.
- Fast trend towards the adoption of 3D systems
- Mainstream use of multichannel audio
- Sound field reproduction systems can enhance the presentation of 3D multimedia content
- Ambisonic encodings of sound fields are ideal for applications where a specific loudspeaker distribution cannot be assumed

Decoding of Ambisonic data

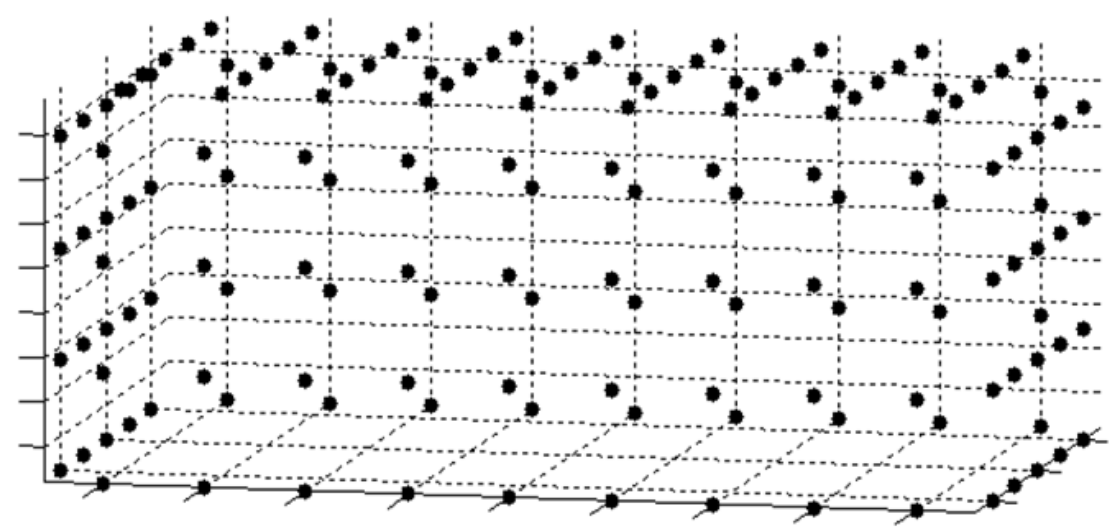


Sound field | Ambisonic data | Re-created field

- Decoding is usually based on the pseudo-inverse of a matrix of spherical harmonic functions
  - Requires a **uniform sampling** of the sphere



Regular array



Irregular array

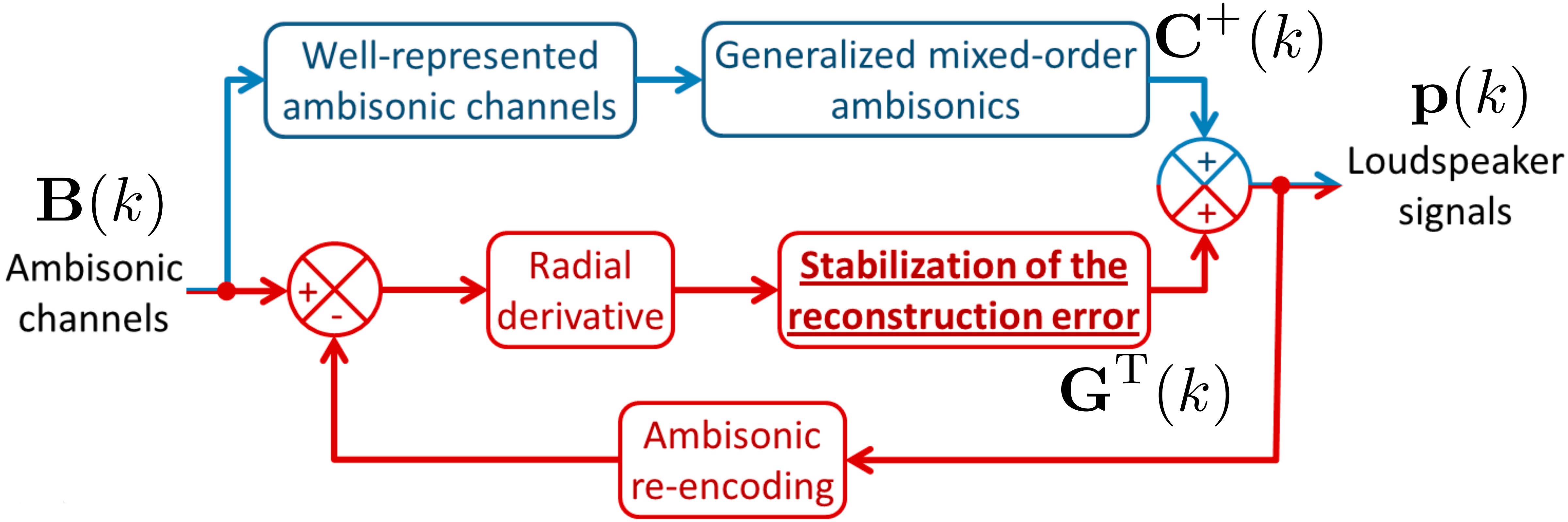
Drawbacks of standard decoders

- **Numerical instability** when using irregular arrays
  - The pseudo-inverse is not a continuous operation
  - Re-encoding matrices for irregular loudspeaker distributions tend to be ill-conditioned

- **Suboptimal solutions** when there are more loudspeakers than ambisonic channels
  - Decoding becomes an underdetermined problem
  - The pseudo-inverse picks the solution with minimum Euclidean norm (not a meaningful parameter for human listeners)

A new approach to Ambisonic decoding

- Can be used to decode ambisonic data for reproduction over irregular arrays
- Decodes the ambisonic data in two stages:
  - Decodes the non-problematic ambisonic channels
  - Stabilizes the reconstruction error by imposing a constraint on its radial derivative



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

Proposed decoder for irregular arrays

- Generalized mixed-order ambisonics
  - Decode the largest combination of channels that result in a well-conditioned re-encoding matrix

- Stabilization of the reconstruction error
  - Find the decoding gains that minimize the gradient of the reconstruction error

Target sound field

Mixed-order partial reconstruction

$$\mathbf{G} = \arg\min_{\mathbf{G}} \left| \frac{\partial}{\partial r} \left[\tilde{\psi}_k(r,\theta,\varphi) - \phi_k(r,\theta,\varphi) - \sum_{s} \sum_{m=0}^{N} \sum_{n=-m}^{m} G^{s}_{mn}(k) \frac{e^{-ik|\vec{r}-\vec{r}_s|}}{|\vec{r}-\vec{r}_s|} Y_{mn}(\theta_s,\varphi_s) \right] \right|$$

Decoding matrix

Gain for the s-th loudspeaker

Output of the s-th loudspeaker

The new decoder as a regularization scheme

- Constraining the radial derivative of the reconstruction error can also lead to a regularized least-squares problem:

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

Radial derivative of the reconstruction error: $\left|\frac{\partial}{\partial r}\epsilon(k,\vec{r})\right|$

Regularization term (difference operator): $\mathbf{L}$

Decoder gains: $\mathbf{G}$

Mixed-order reconstruction error: $\mathbf{d}$

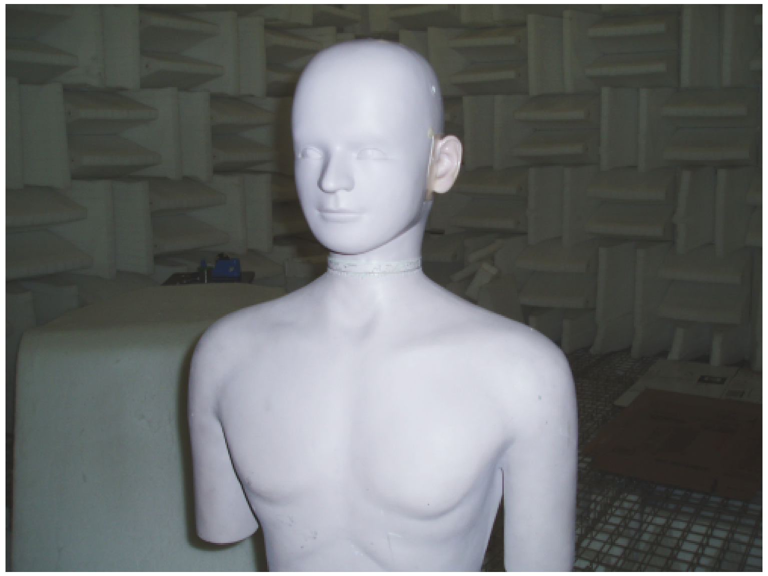
Target reconstruction error variation: ρ

Constant term: $\mathbf{d} \equiv \nabla[\tilde{\psi}(k,\vec{r}) - \phi(k,\vec{r})]\cdot\hat{r}$

Difference operator: $\mathbf{L}_s(k) \equiv \frac{|\vec{r}| - |\vec{r_s}|\cos(\vec{r},\vec{r_s})}{|\vec{r}-\vec{r_s}|}\left(\frac{1}{|\vec{r}-\vec{r_s}|} + ik\right)$

Evaluation

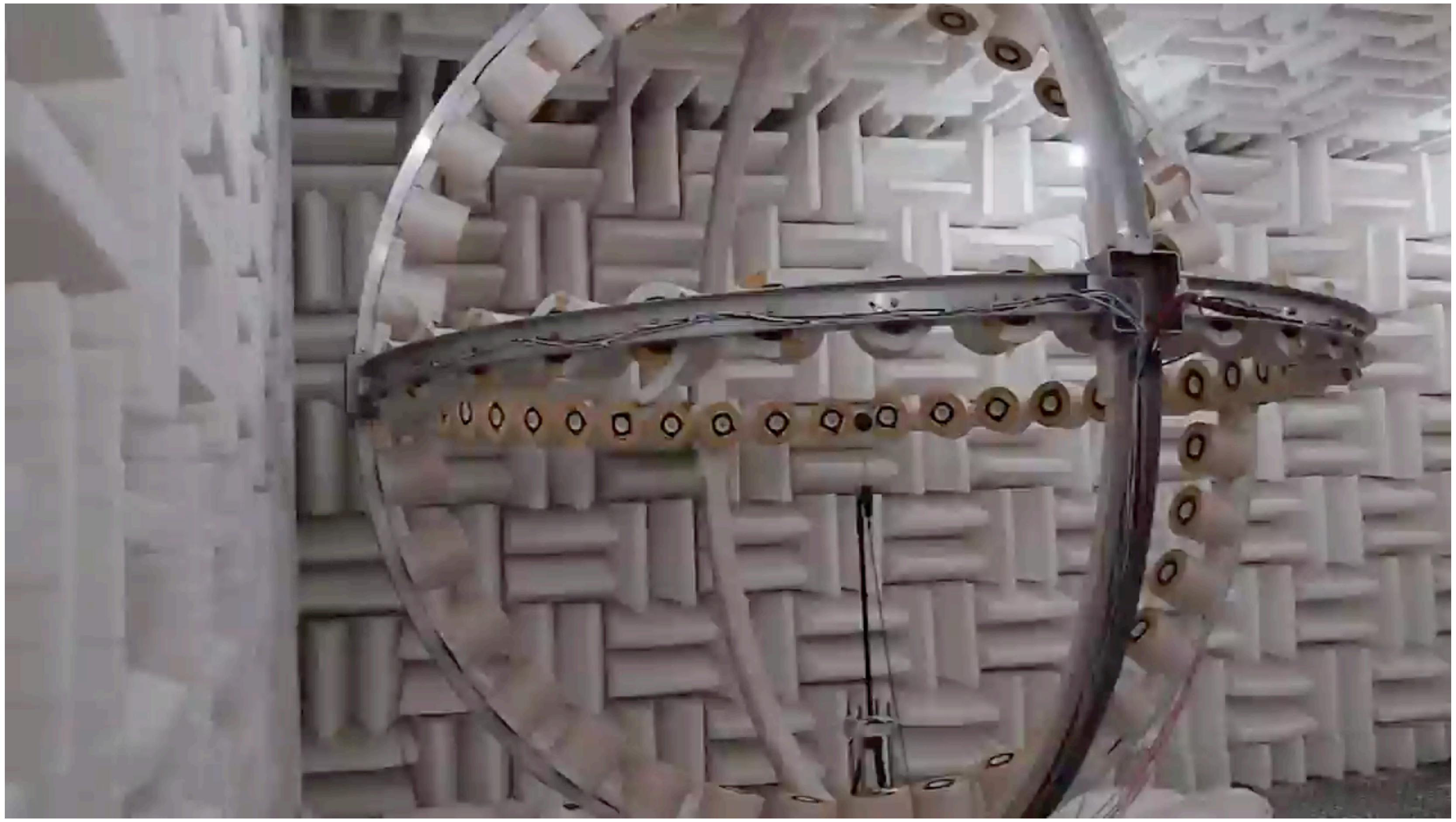
- Our previous research has focused on the physical accuracy with which the sound field is re-created
- Implementation of 3D audi need for systems that can convey the illusion of being immersed in a different environment
  - Interaural level difference
  - Interaural phase difference
- Interaural cues derived from the binaural rendering of ambisonic recordings
  - Three virtual loudspeaker arrays
  - SAMRAI's dummy head HRTF measurements



Dummy head SAMURAI
Koken Co., Ltd.

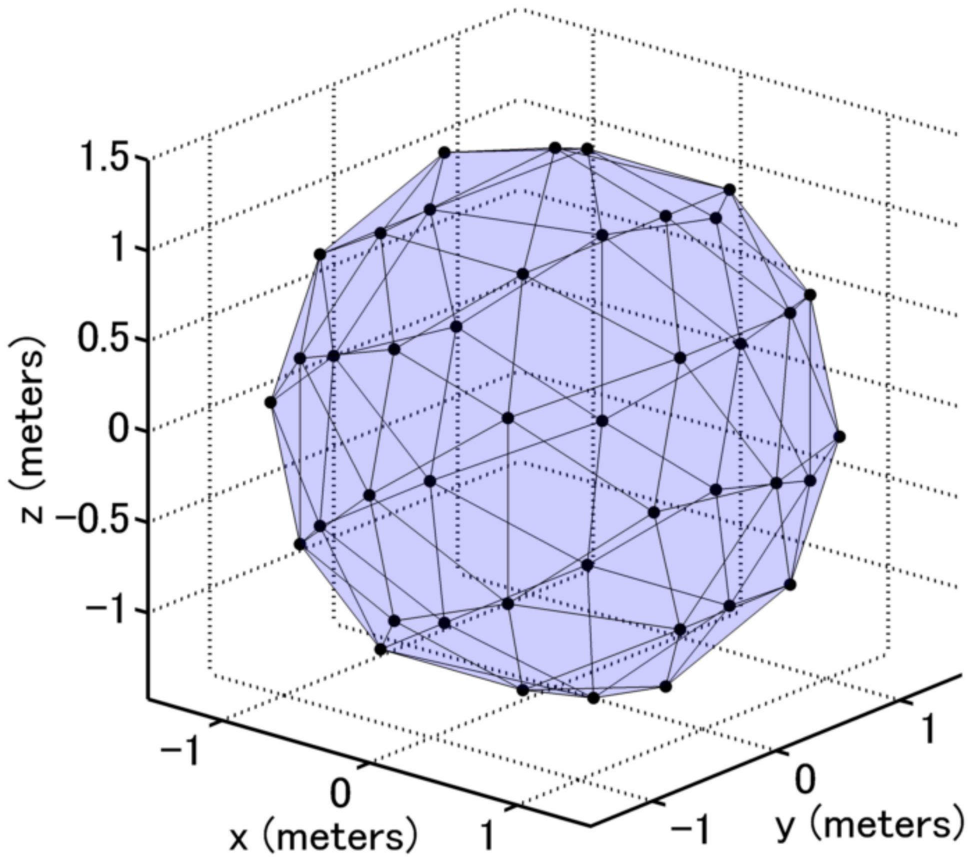
Measurement of HRTFs

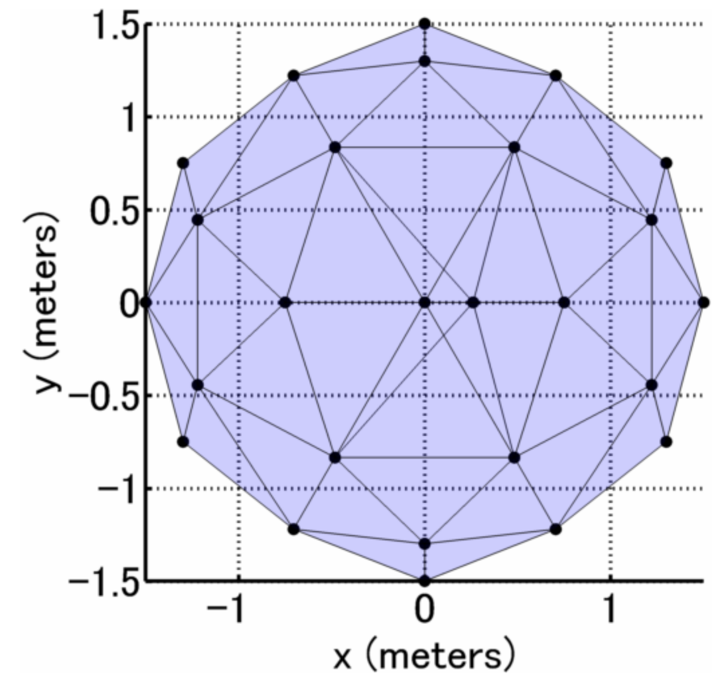
- Measurements were made using a spherical loudspeaker array housed in an anechoic chamber
  - All azimuth angles from −175 deg. to 180 deg. in increments of 5 degrees
  - Available for elevation angles between −80 deg. and 90 deg. in increments of 10 degrees

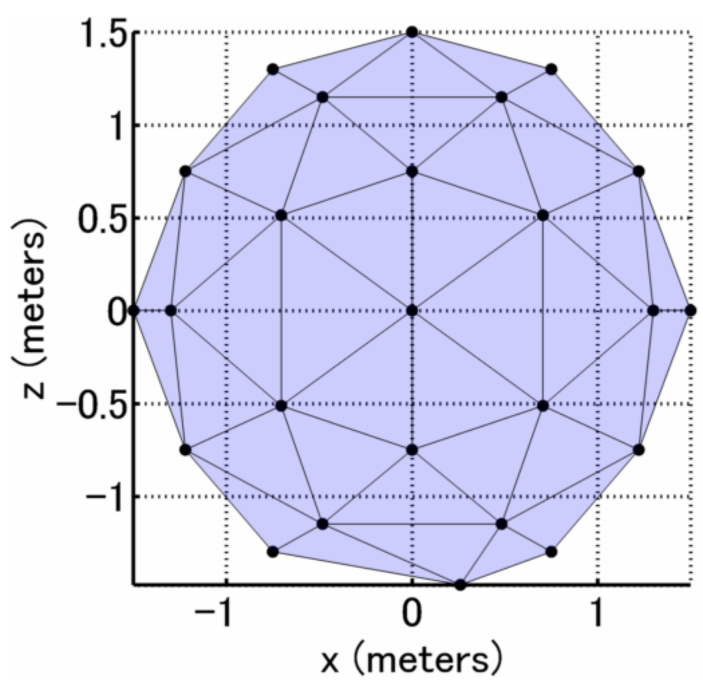


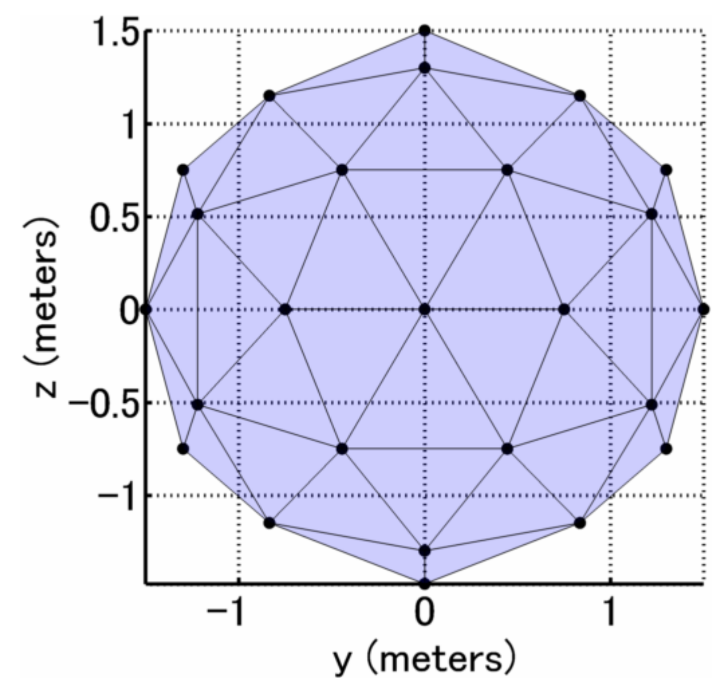
Spherical loudspeaker array in Tohoku University

Regular, 42-channel loudspeaker array

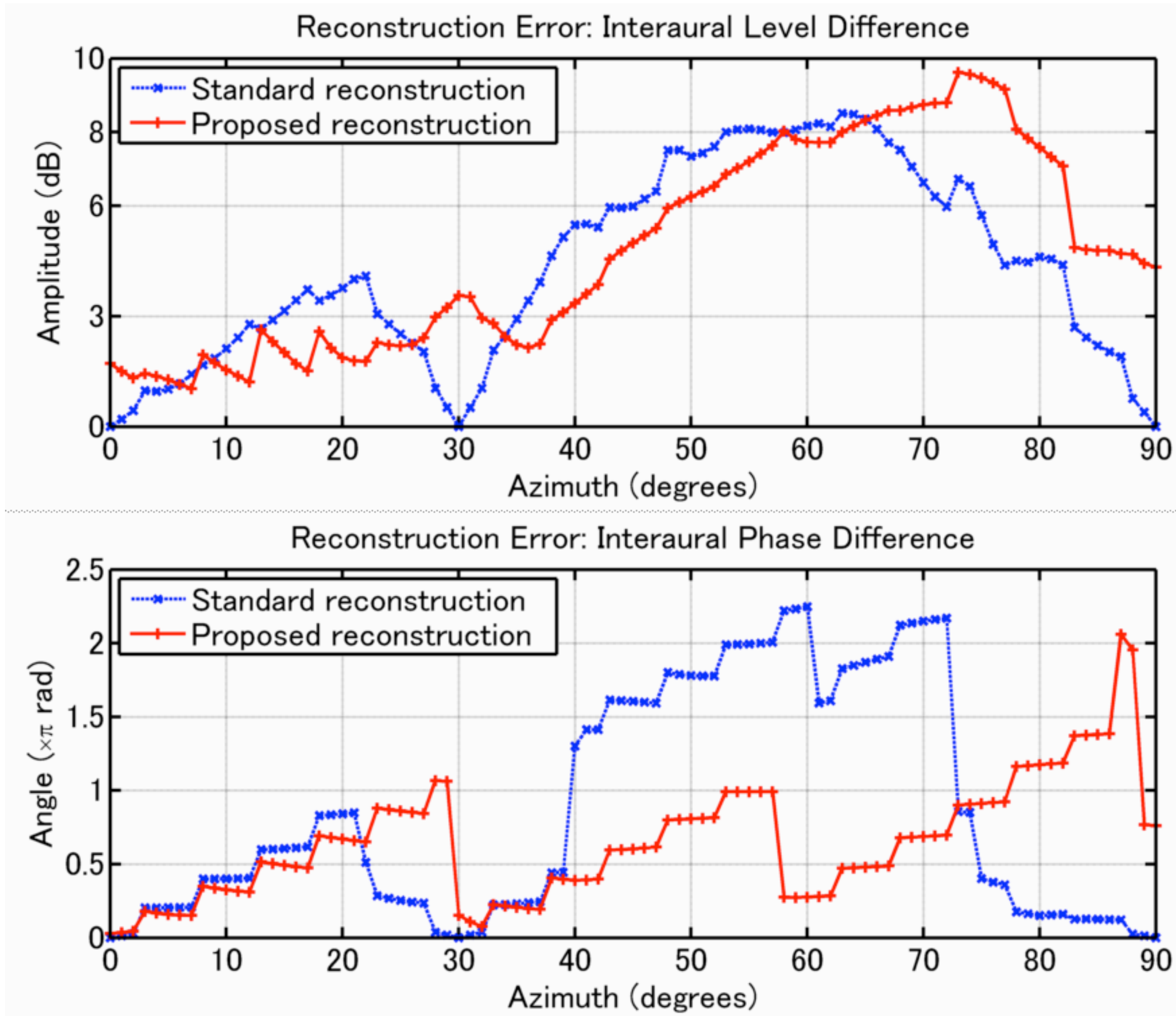






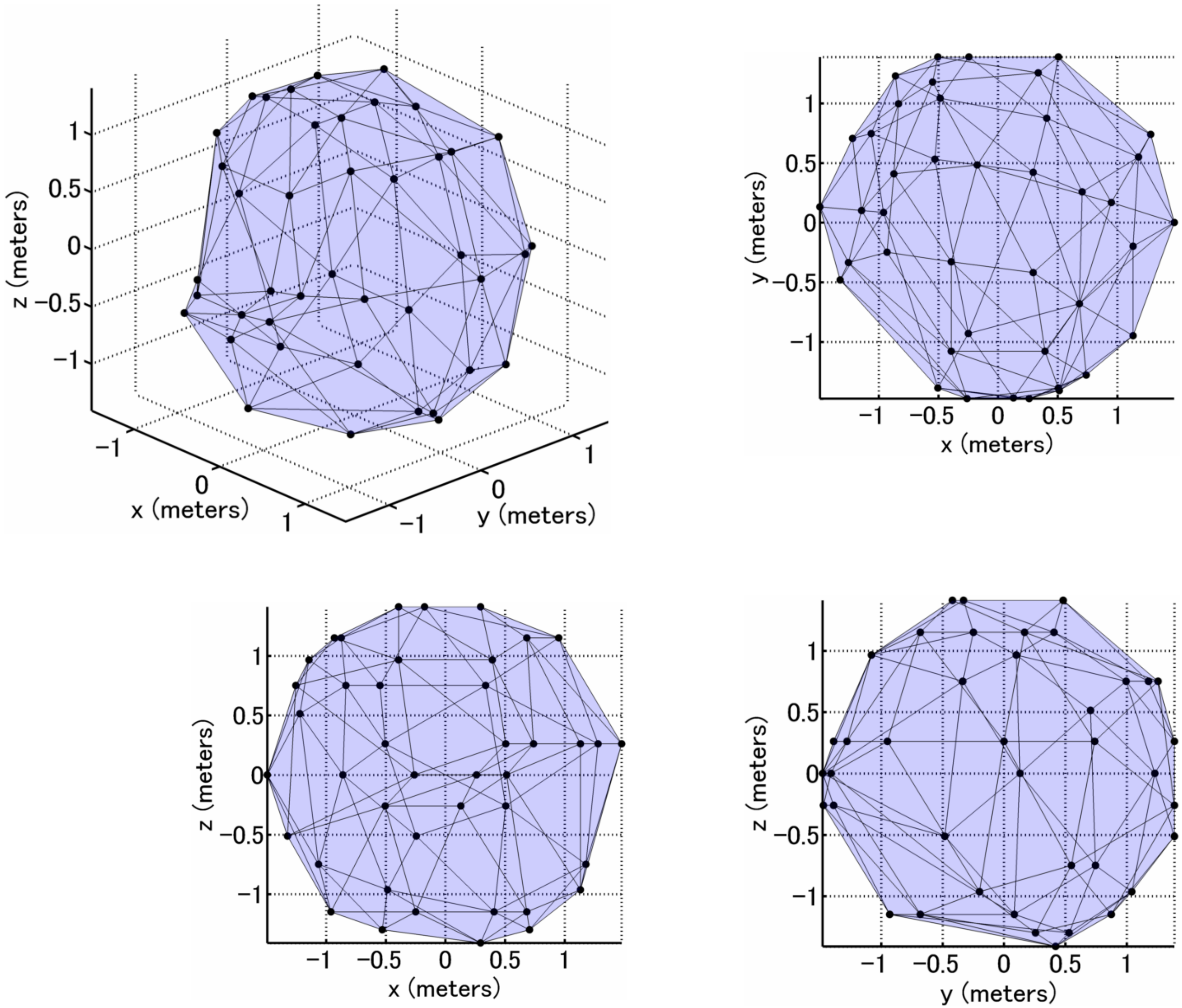


Results for the regular, 42-channel loudspeaker array (azimuth)

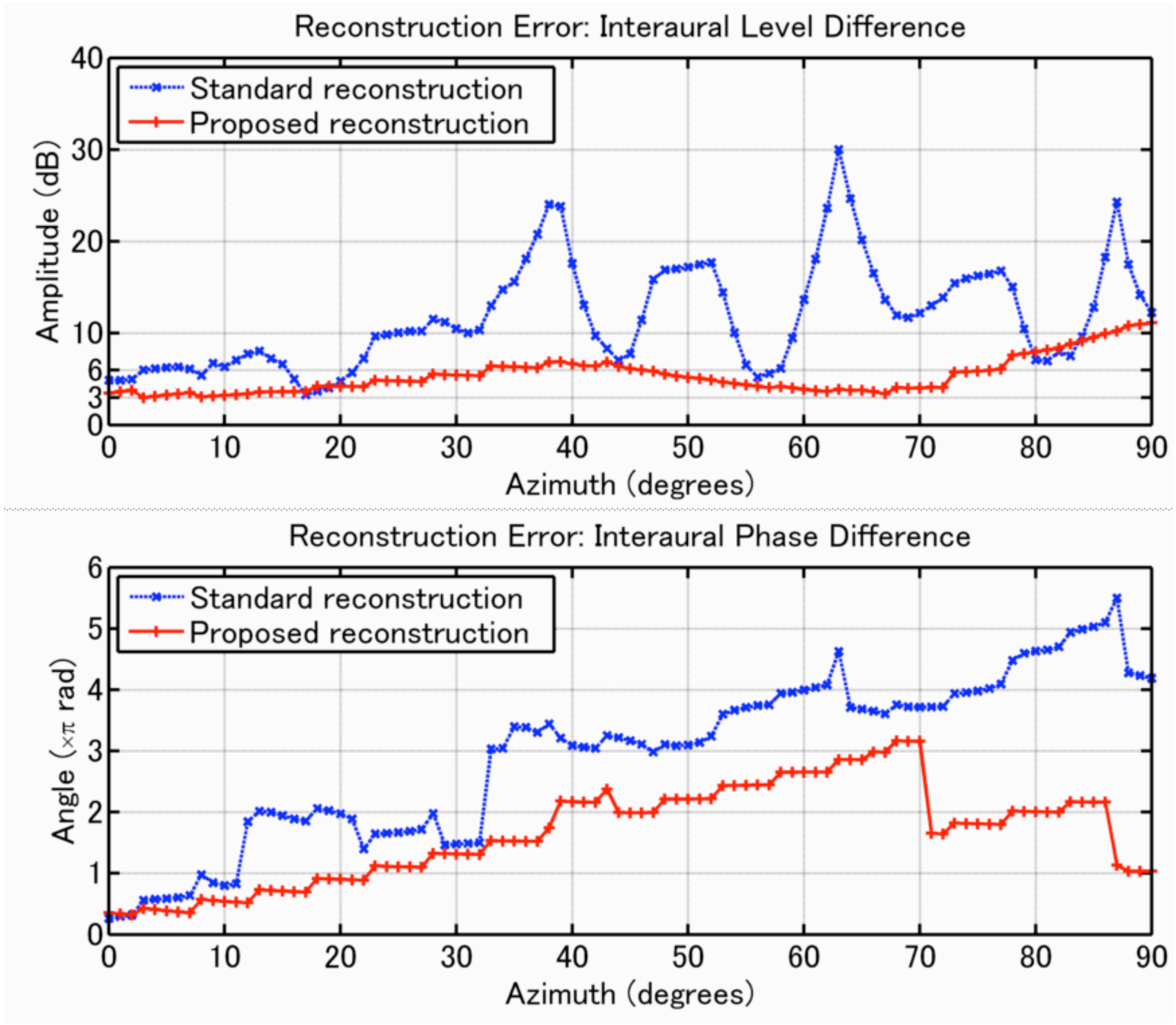


Varying azimuth angle (elevation: 0 deg.)

Irregular, 42-channel loudspeaker array



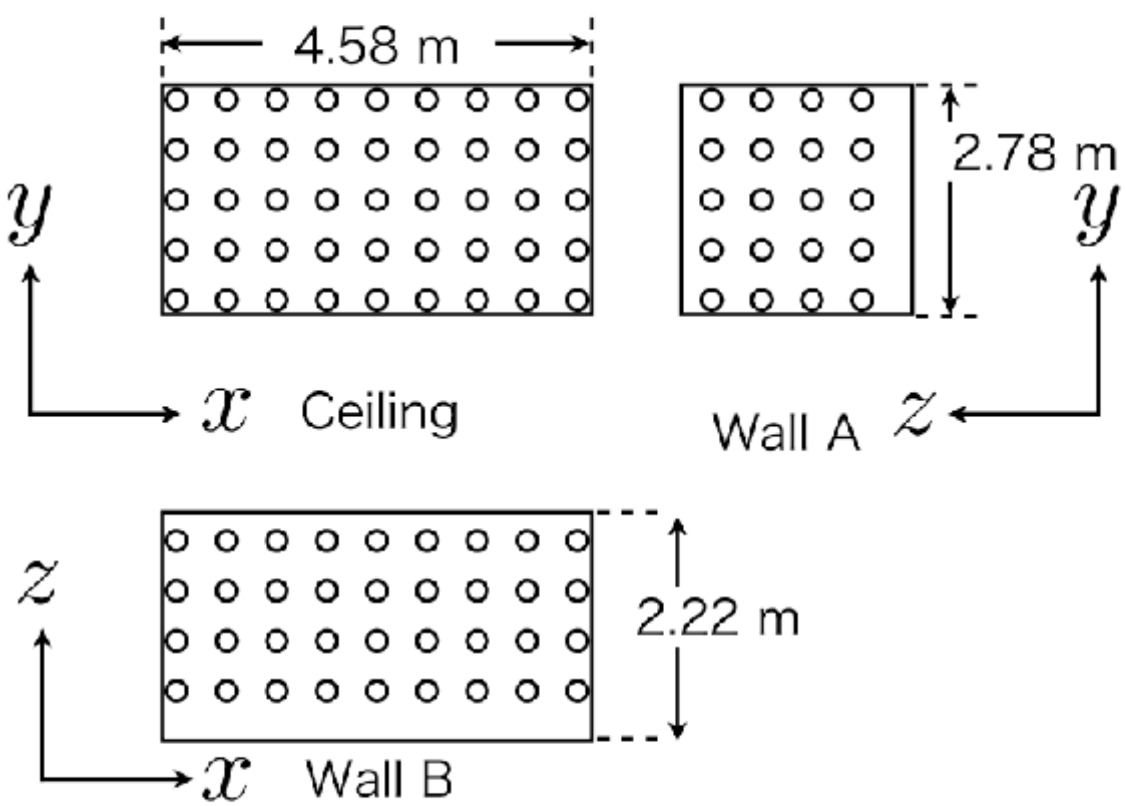
Results for the irregular, 42-channel loudspeaker array (azimuth)



Varying azimuth angle (elevation: 0 deg.)

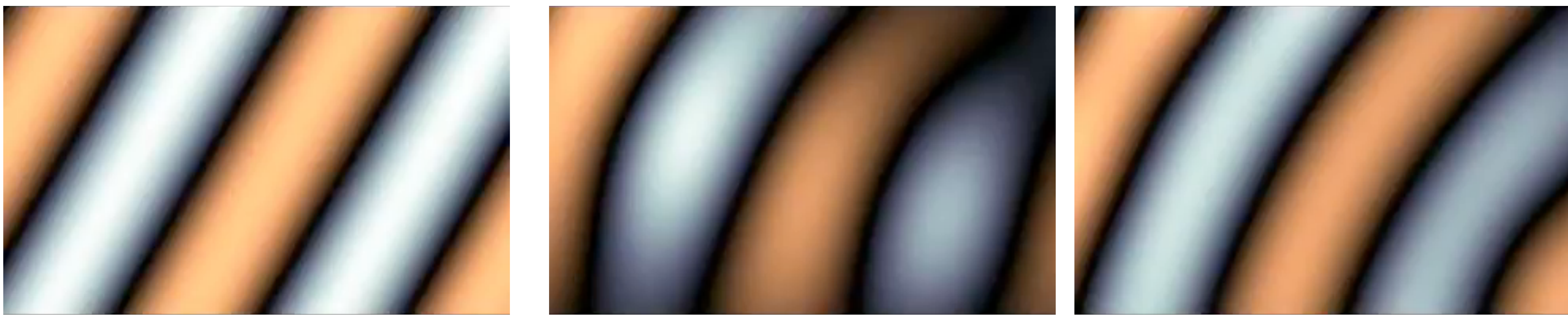
Irregular, 157-channel loudspeaker array

- The proposed decoder was tested through the **computer simulation** of a **157-channel, irregular** loudspeaker array



Results for the irregular, 157-channel loudspeaker array

- Simulation of a 500 Hz plane wave incident at an azimuth of 30 degrees



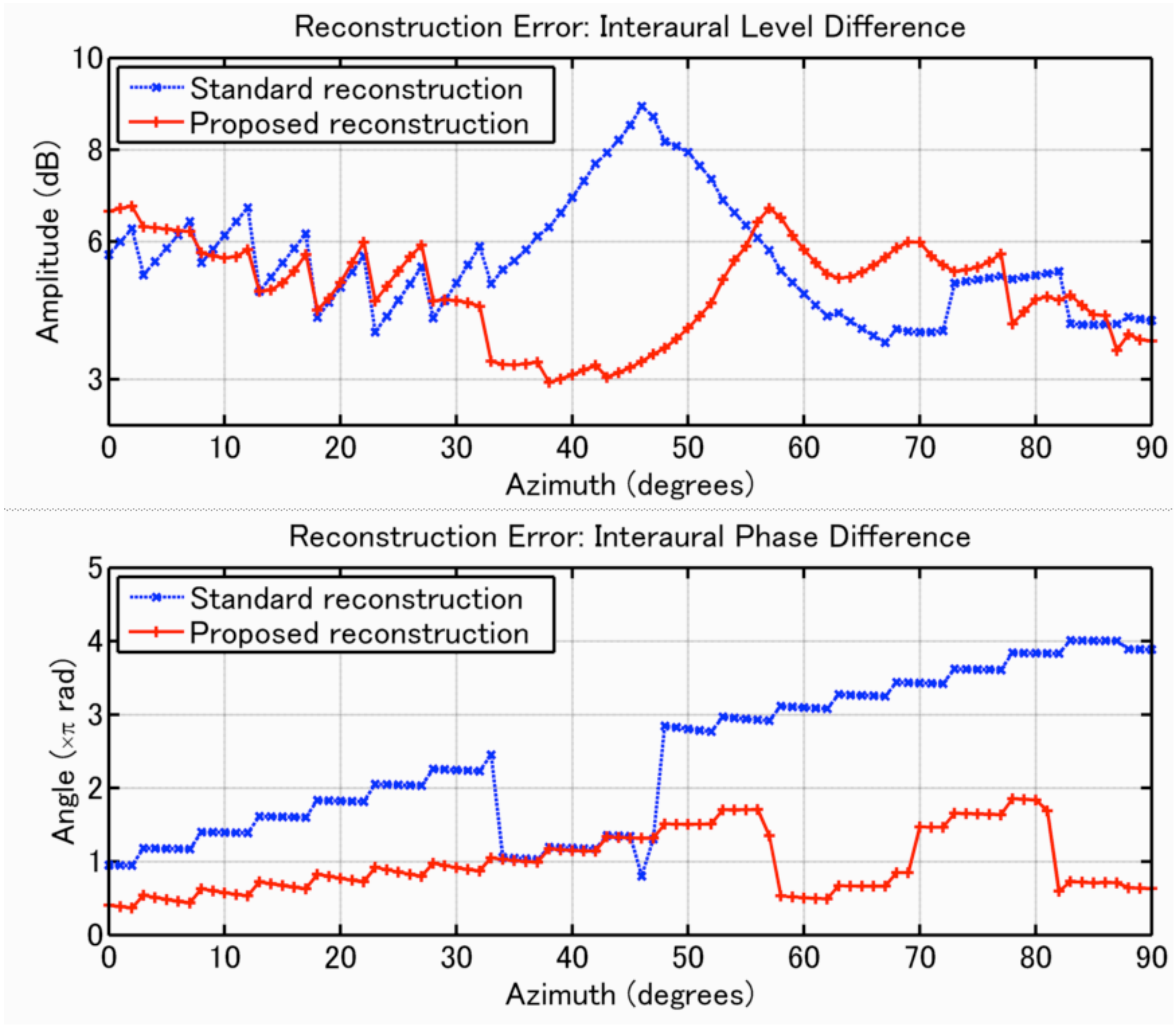
Ideal

Standard decoder

Proposed decoder

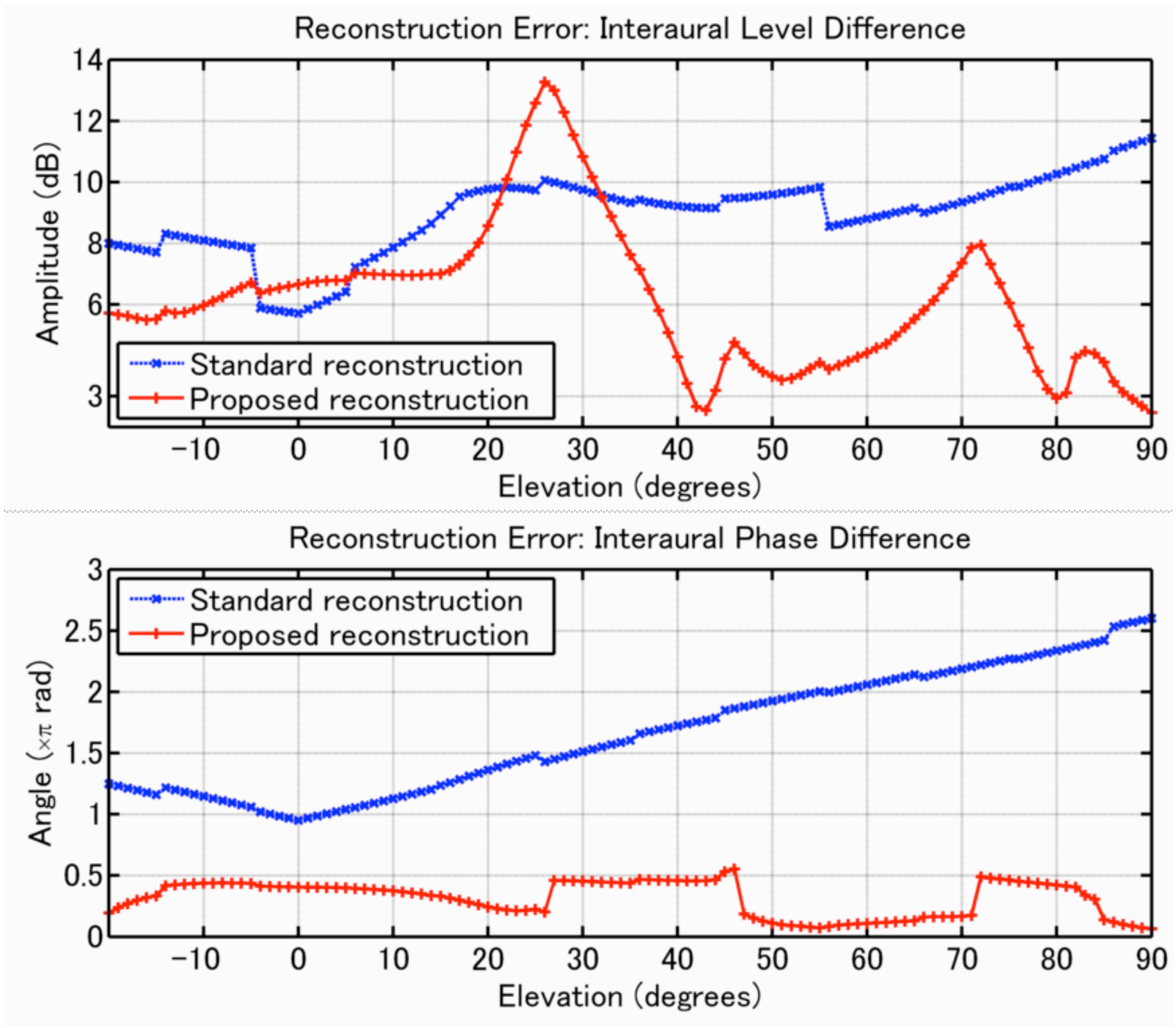
−1 Sound pressure (normalized) 1

Results for the irregular, 157-channel loudspeaker array (azimuth)



Varying azimuth angle (elevation: 0 deg.)

Results for the regular, 157-channel loudspeaker array (elevation)



Varying elevation angle (azimuth: 0 deg.)

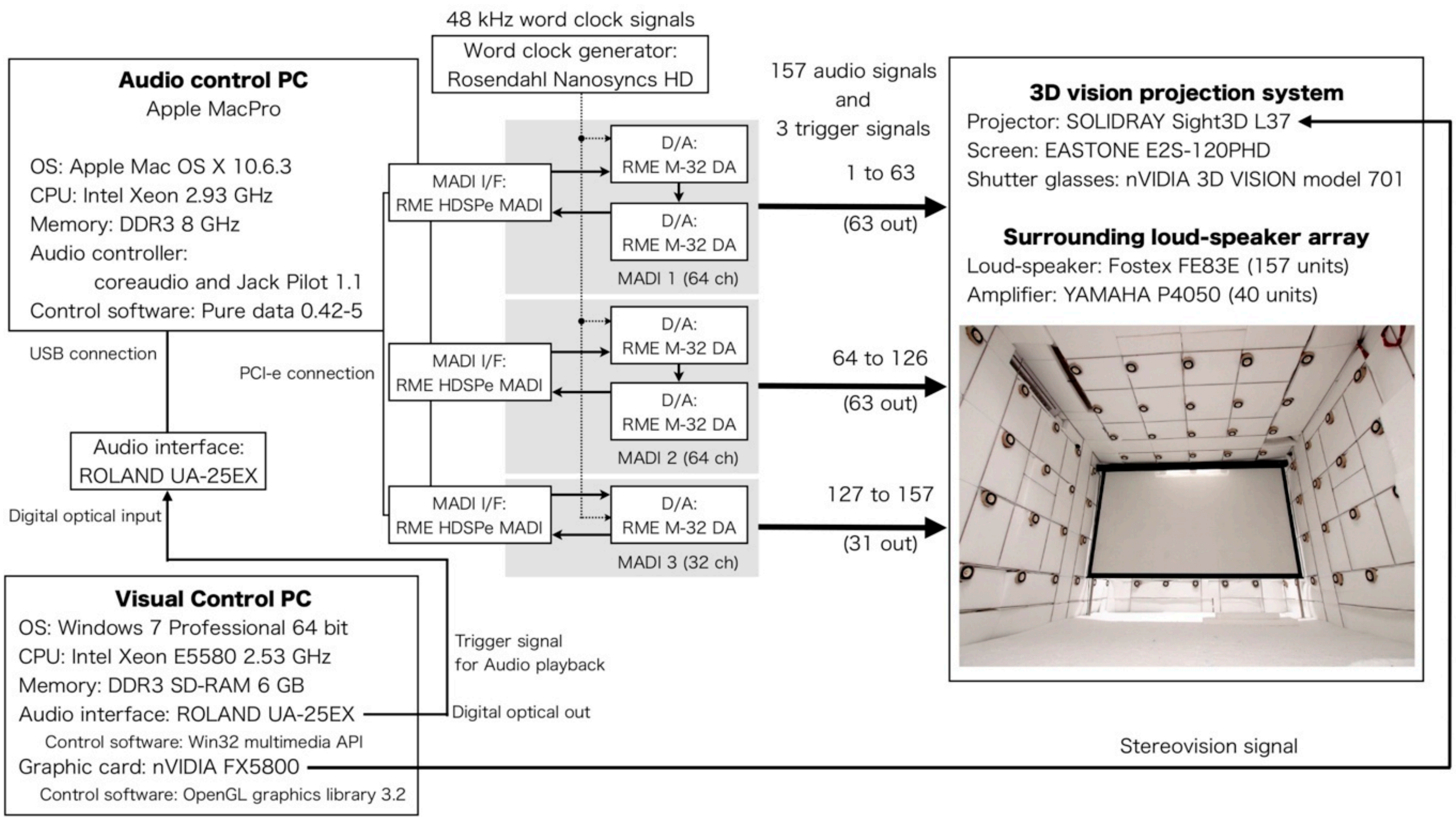
Concluding remarks

- Two drawbacks to standard ambisonic decoders
  - Numerical instability
  - Suboptimal solutions
- Introduced a new ambisonic decoding method
  - Weaker regularity constraints than standard decoders
  - Better solutions for human listeners
- Evaluation using a three virtual loudspeaker arrays
  - Improved reconstruction of interaural cues (ILD, IPD)
  - Reduced dependence on the specific distribution of the loudspeakers

Implementation of 3D audio-visual display

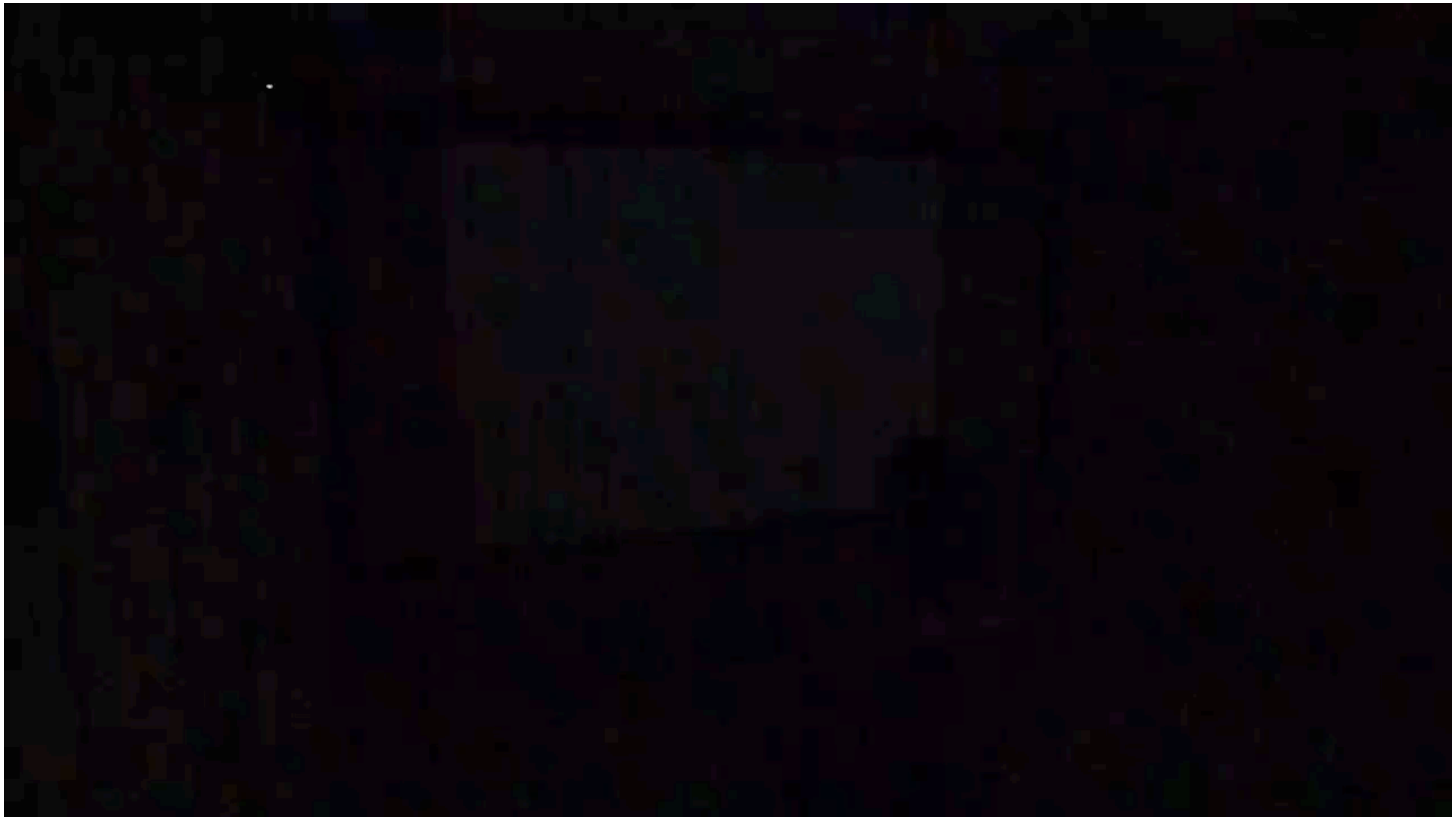
T. Okamoto *et al.*, *Proc. IEEE IC-NIDC 2010*

- Combining HOA system with 3D projection display
  - Audio system: Completely synchronous 157-loudspeaker array system
  - Visual system: Stereo shutter technique with acoustic transparent screen
    - 5th order decoding



Demonstration of 3D audio-visual display

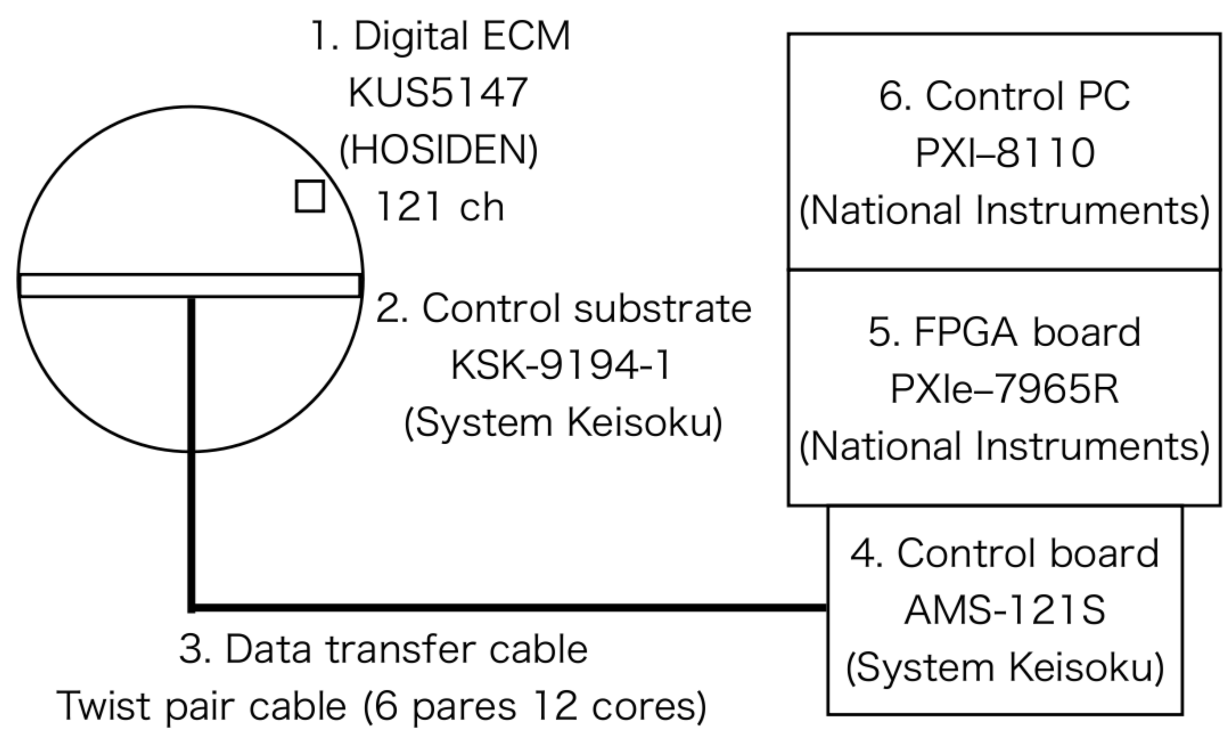
- Hit the drum in the gymnasium
  - Vision signals: Stereo video recording using FUJIFILM FINEPIX REAL 3D W1
  - sound signals: calculating by simulation and decoding by 5th order HOA



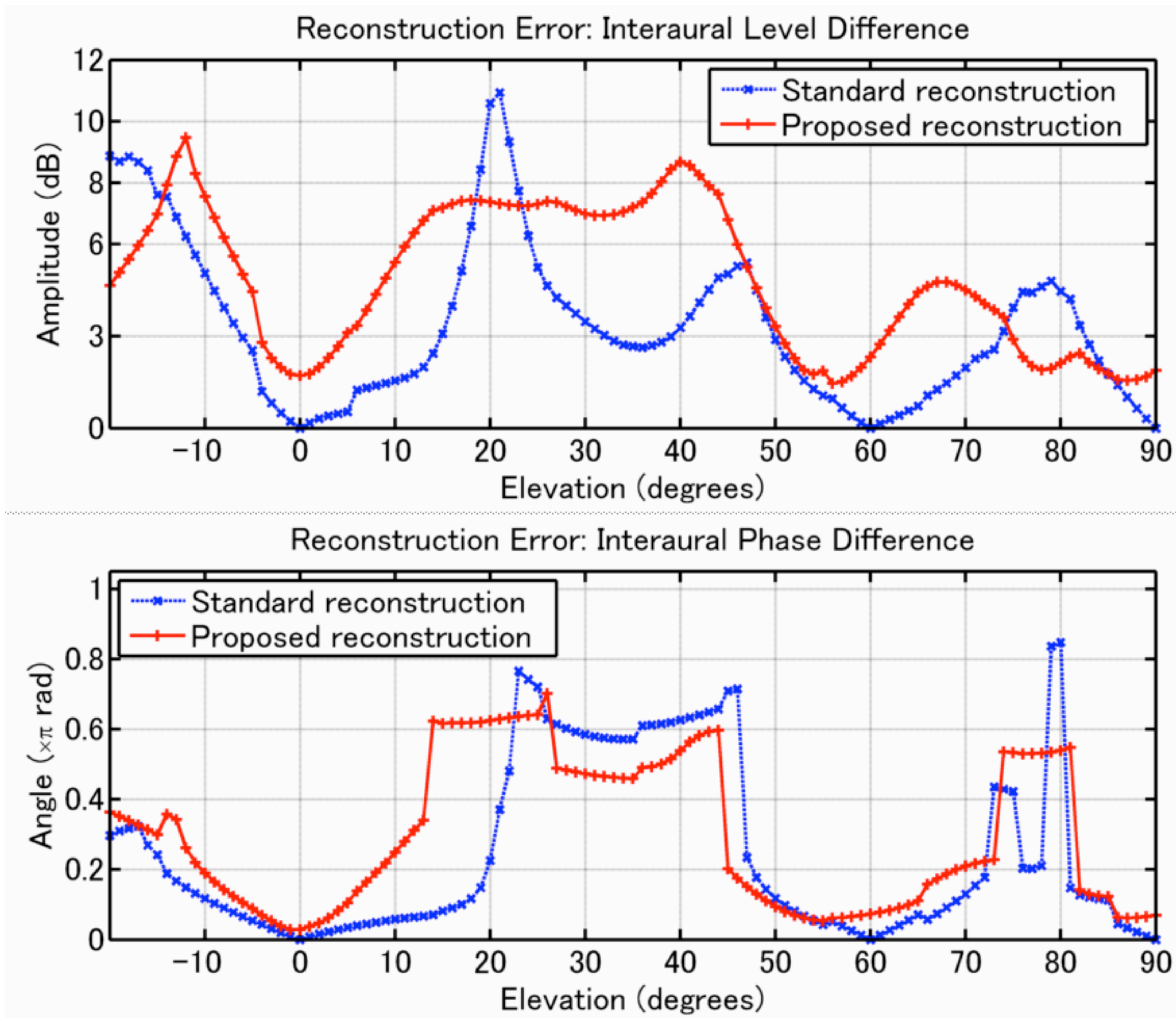
Implementation of sound field recording system based on HOA

T. Okamoto *et al.*, *Proc. SOIM-GCOE 2010*

- HOA recording system using a 121 spherical microphone array
  - 9th order decoding
  - Completely synchronous 121-audio recording system using Digital Electric Condenser Microphone (Digital ECM) and FPGA board
    - Sampling frequency is 48 kHz
    - 4th Delta-Sigma modulation is used for 1-bit signals
  - Controlled by LabVIEW (National Instruments) in Windows XP (SP3)

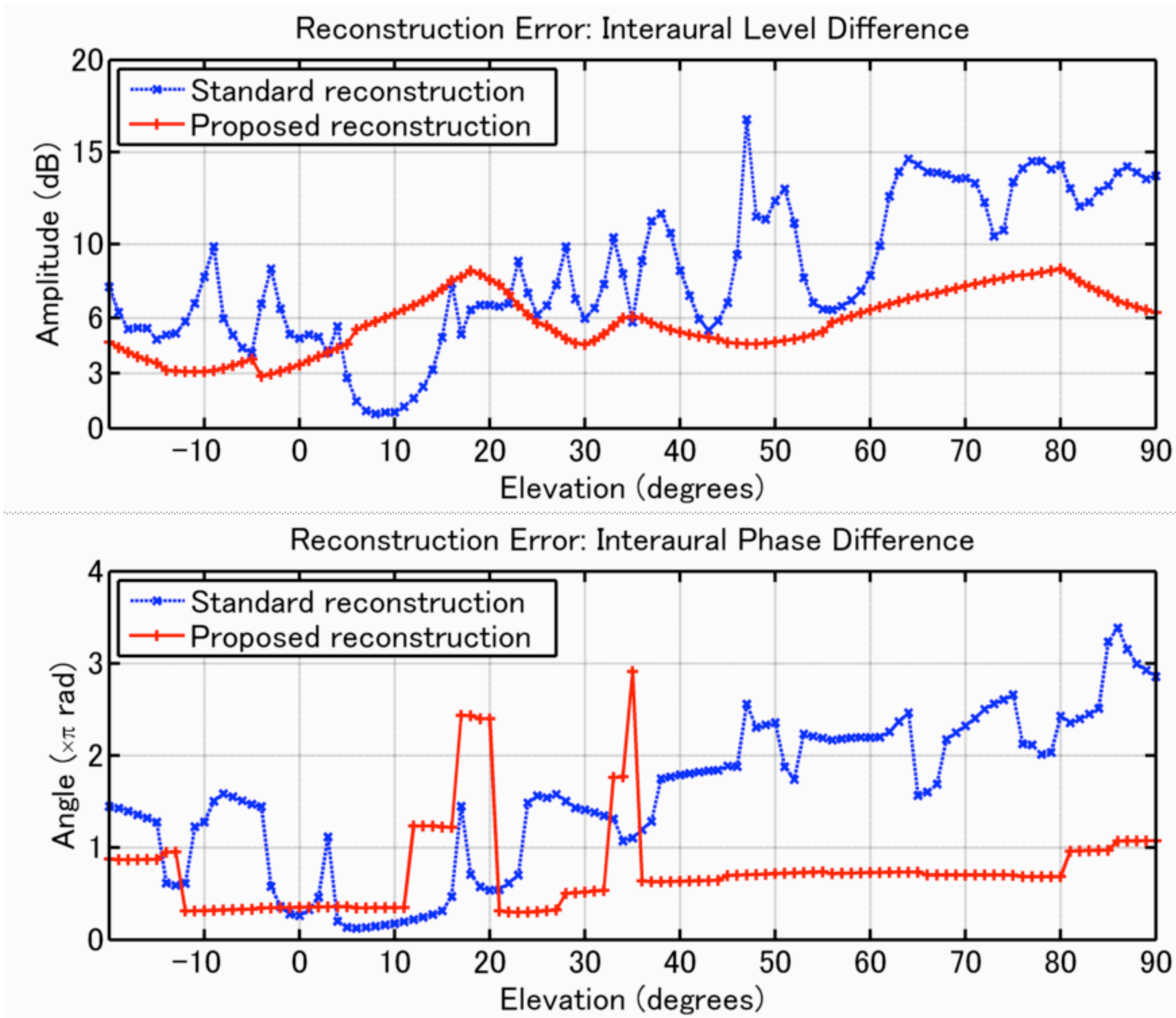


Results for the regular, 42-channel loudspeaker array (elevation)



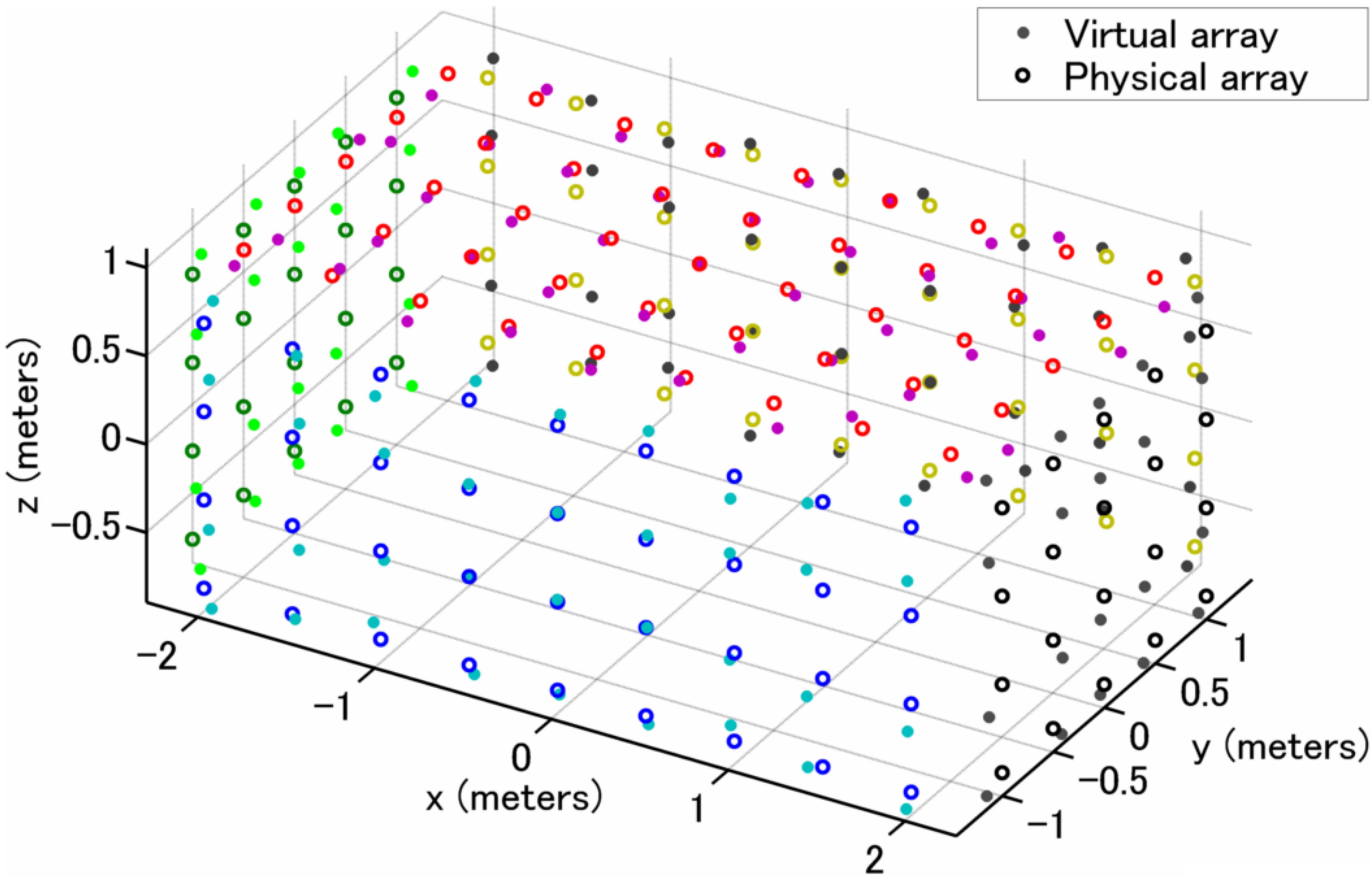
Varying elevation angle (azimuth: 0 deg.)

Results for the regular, 42-channel loudspeaker array (elevation)

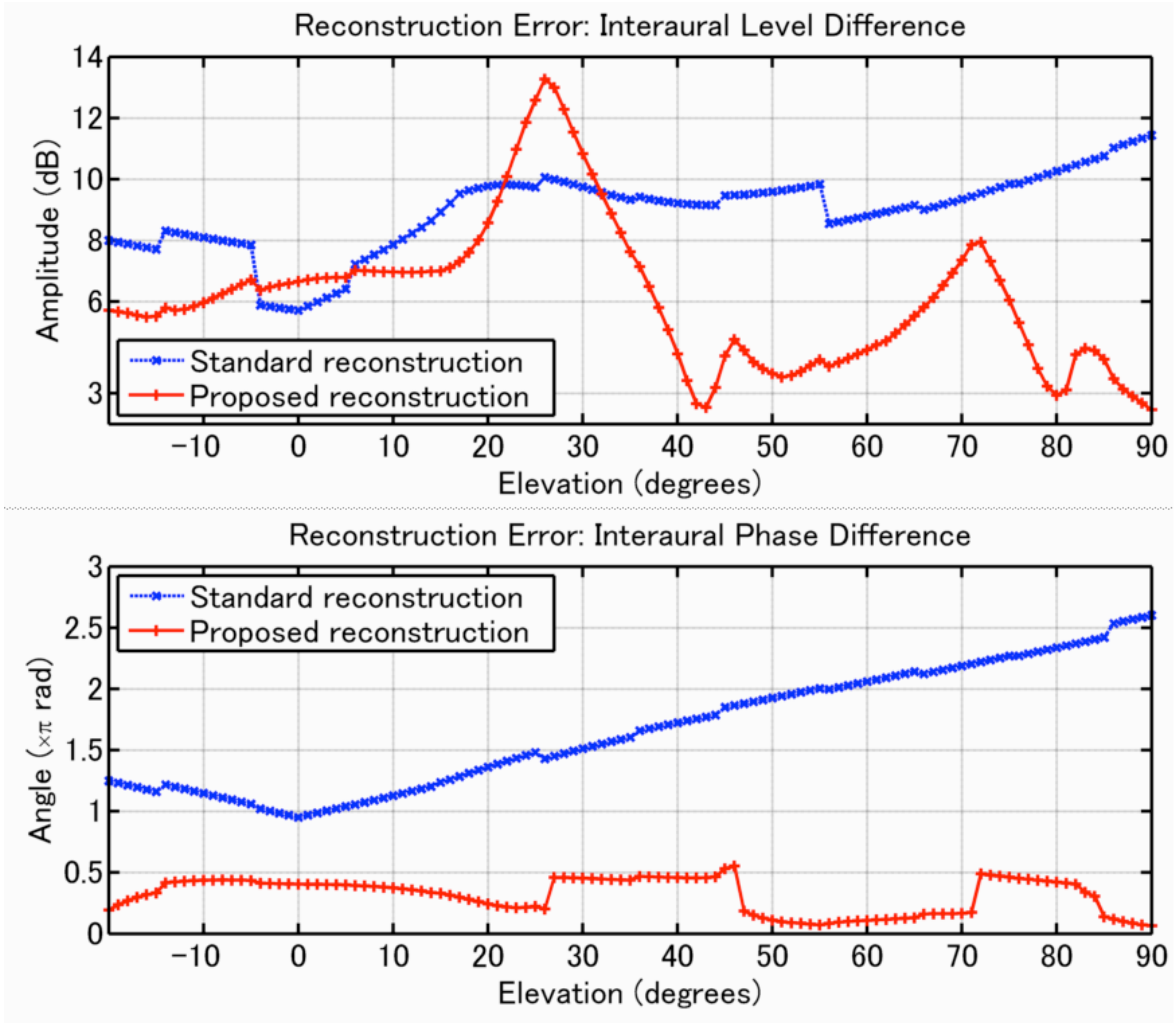


Varying elevation angle (azimuth: 0 deg.)

Irregular, 157-channel loudspeaker array



Results for the regular, 157-channel loudspeaker array (elevation)



Varying elevation angle (azimuth: 0 deg.)