Implementation of a high-definition 3D audio-visual display based on Higher-order Ambisonics using a 157-loudspeaker array combined with a 3D projection display

T. Okamoto[1,2)], Z. L. Cui[1)], Y. Iwaya[1,3)] and Y. Suzuki[1,3)]

岡本拓磨　　崔正烈　　岩谷幸雄　　鈴木陽一

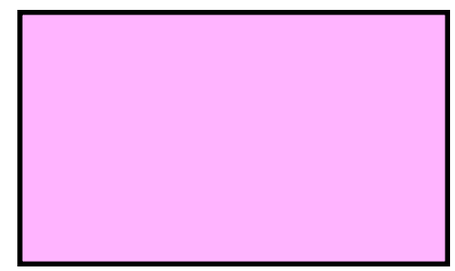
1) Research Institute of Electrical Communication, Tohoku University, Japan
2) Graduate School of Engineering, Tohoku University, Japan
3) Graduate School of Information Sciences, Tohoku University, Japan

東北大学
1) 電気通信研究所，2) 大学院工学研究科，3) 大学院情報科学研究科

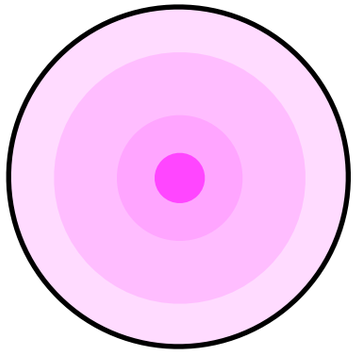
Introduction and research aim

- Implementation of 3D audio visual display
  - Realizing communications with high sense-of-presence in digital content
    - not only actual environment but also virtual and mixed environment
  - Clarifying human audio-visual and various multimodal perceptions including spatial information

- Today's topics
  - Implementation of 3D sound field reproduction system
    - using 157-loudspeaker array
    - sound field reproduction based on Higher-Order Ambisonics (HOA)
      - highly sound field reproduction around a center of array than Wave Field Synthesis (WFS) and Boundary Surface Control (BoSC)
    - Realization of highly synchronous all 157-audio playback system
  - Implementation 3D audio-visual display system
    - Combining HOA reproduction system with 3D projection display
    - System synchronous measurement between audio and visual signals



WFS or BoSC

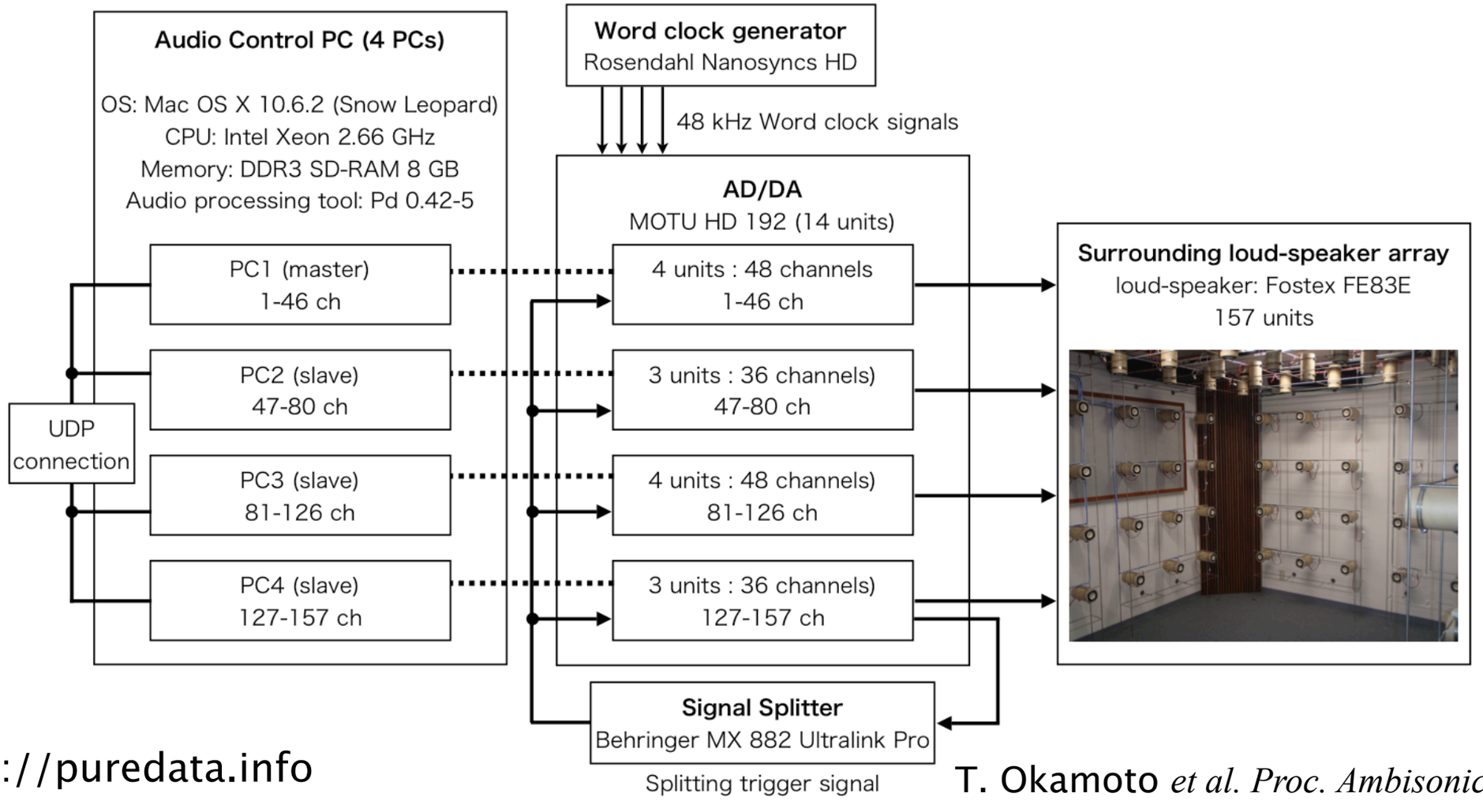


Ambisonics

Relationship between reproduction accuracy and frequency

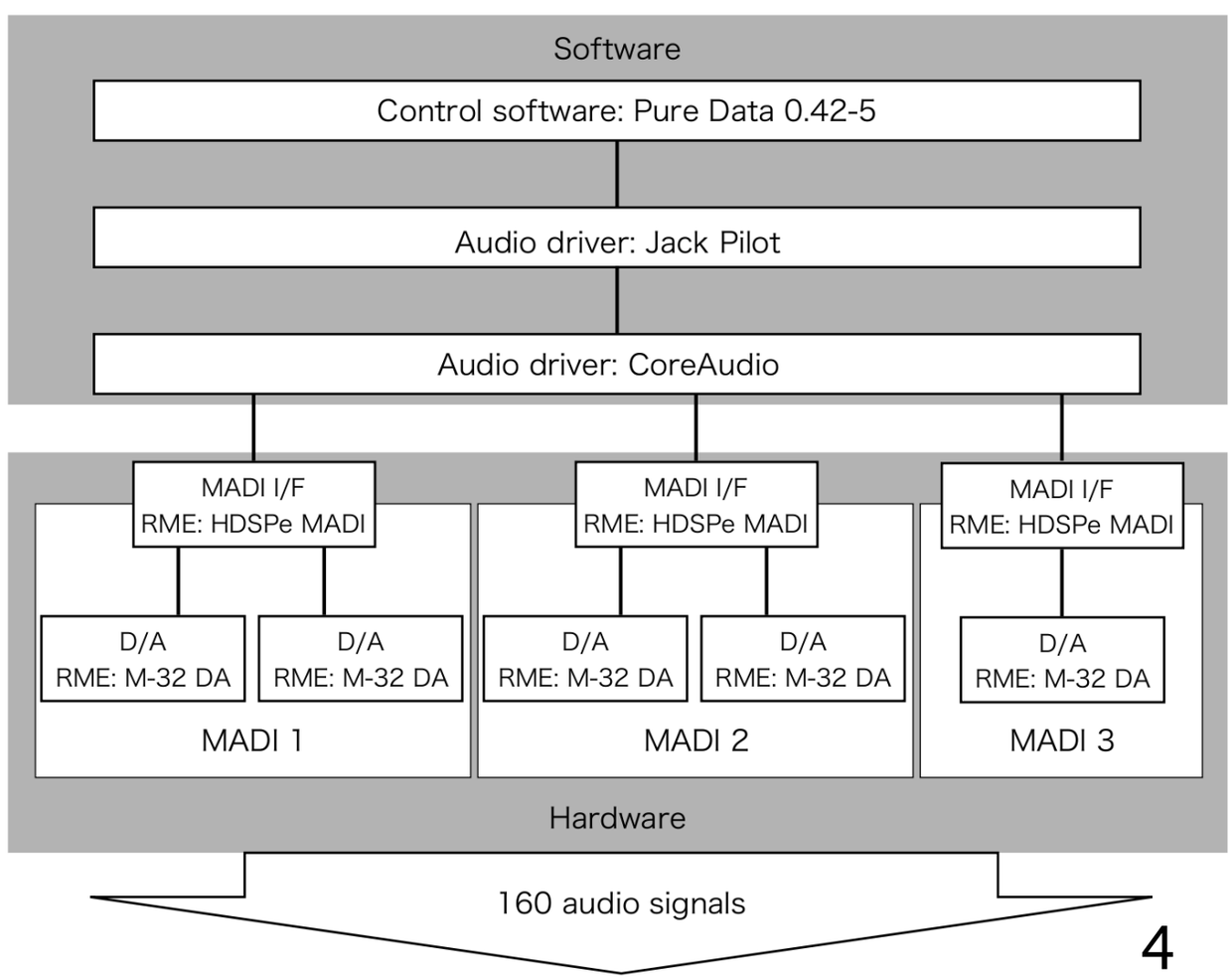
Our previous 3D sound field reproduction system based on HOA

- Sound field reproduction system using 157-loudspeaker array based on HOA
  - 5th order decoding, which is highest order in the world, was realized
  - Previous system was implemented by 4 PCs
    - Audio playback asynchronous between each PC was up to 51 samples (1.1 ms)



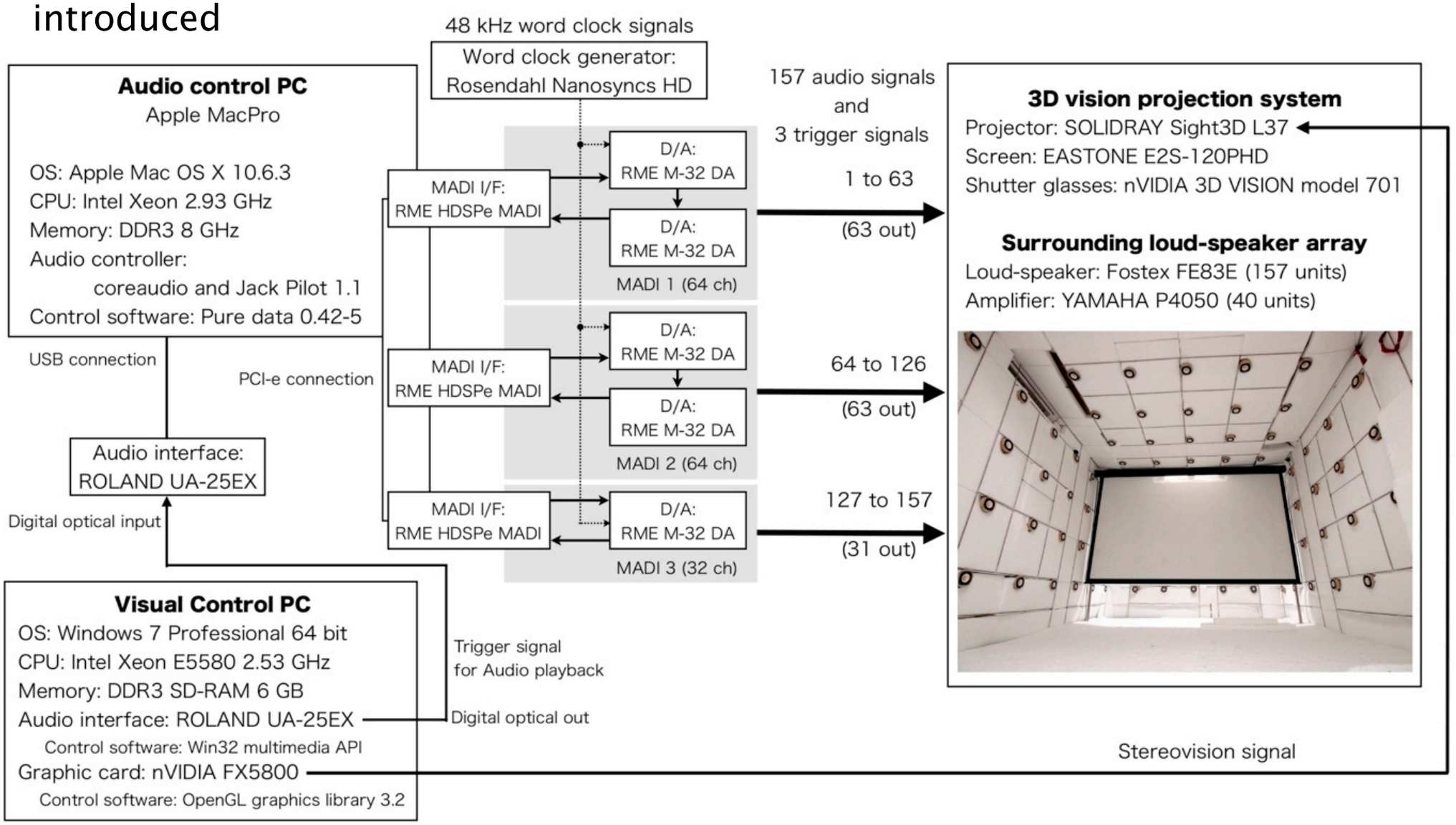
Implementation of highly synchronous all 157-audio playback system

- Completely synchronous all 157-audio playback system using 3 MADI system (audio sampling frequency: 48 kHz)
  - All 157 audio signals are controlled by 1 PC
  - Introduction of Multichannel Audio Digital Interface (MADI) system
    - A MADI system can control 64 audio input/output with completely synchronous
    - Using CoreAudio, which is MacOS X audio driver, for recognizing 3 MADI system as one unified audio interface
    - Moreover, another audio driver, Jack Pilot, is introduced for distributing audio in-out processes to cores of CPUs
  - Evaluation of interchannel synchrony using TSP (Time Stretched Pulse)
    - Jack Pilot off
      - 1 to 3 samples of asynchronous between each MADI
    - Jack Pilot on
      - Completely synchronous all 157 channels



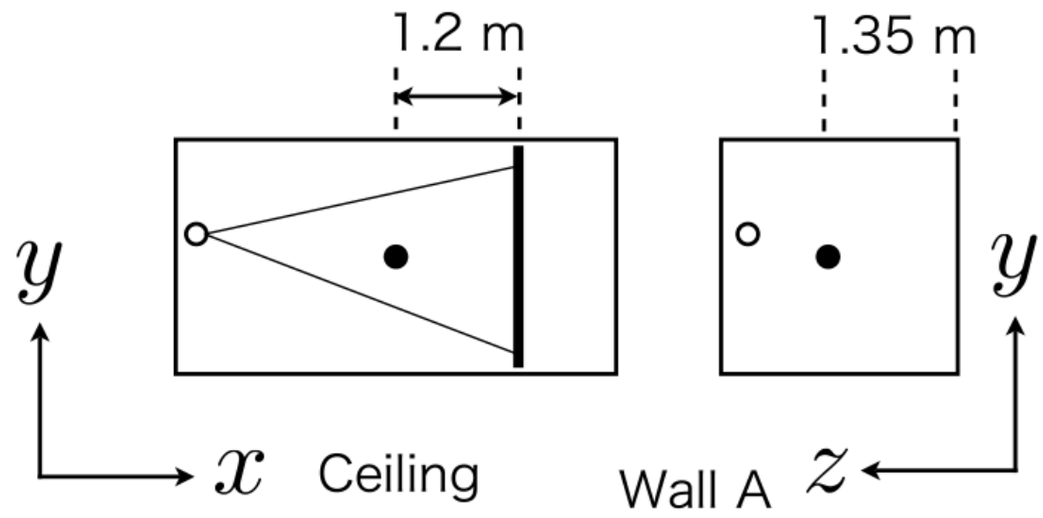
Proposed 3D audio-visual display system

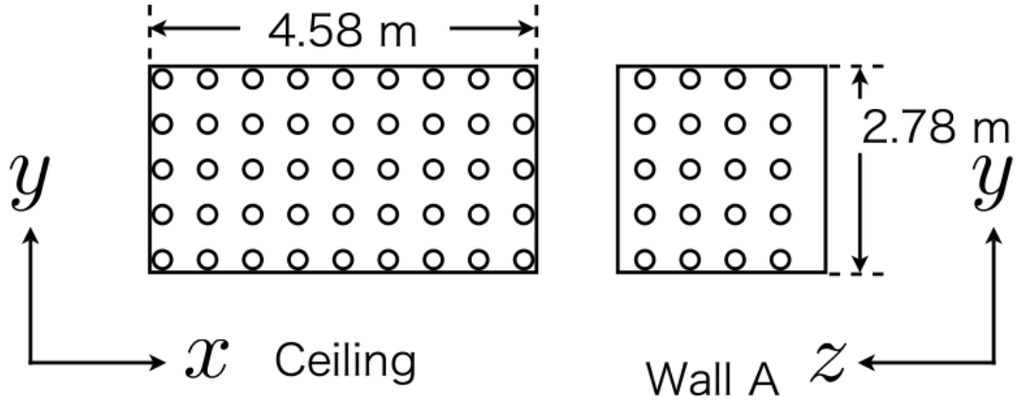
- Combining HOA system with 3D projection display
  - Audio system: Completely synchronous 157-loudspeaker array system
  - Visual system: Stereo shutter technique with acoustic transparent screen
    - For integrating the audio system and the visual system, the trigger signal was introduced

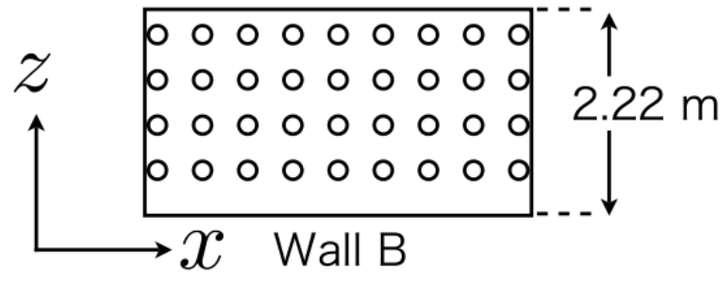


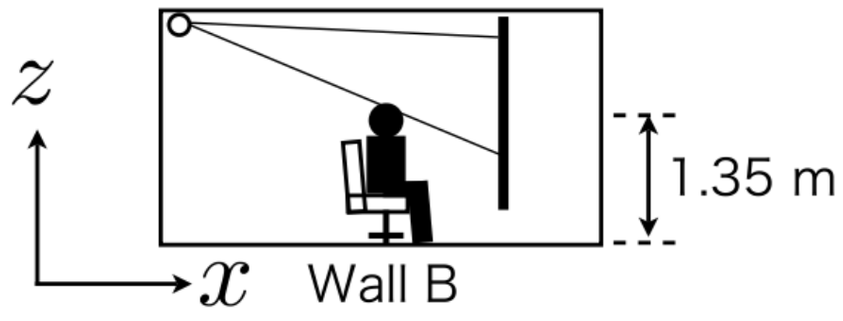
Arrangements

- Arrangements of loudspeakers, projector and screen
  - The sweet spot of for sound field reproduction by HOA decoding is the center of the array
  - The viewer's position for proper stereovision was set identical to the acoustical sweet spot









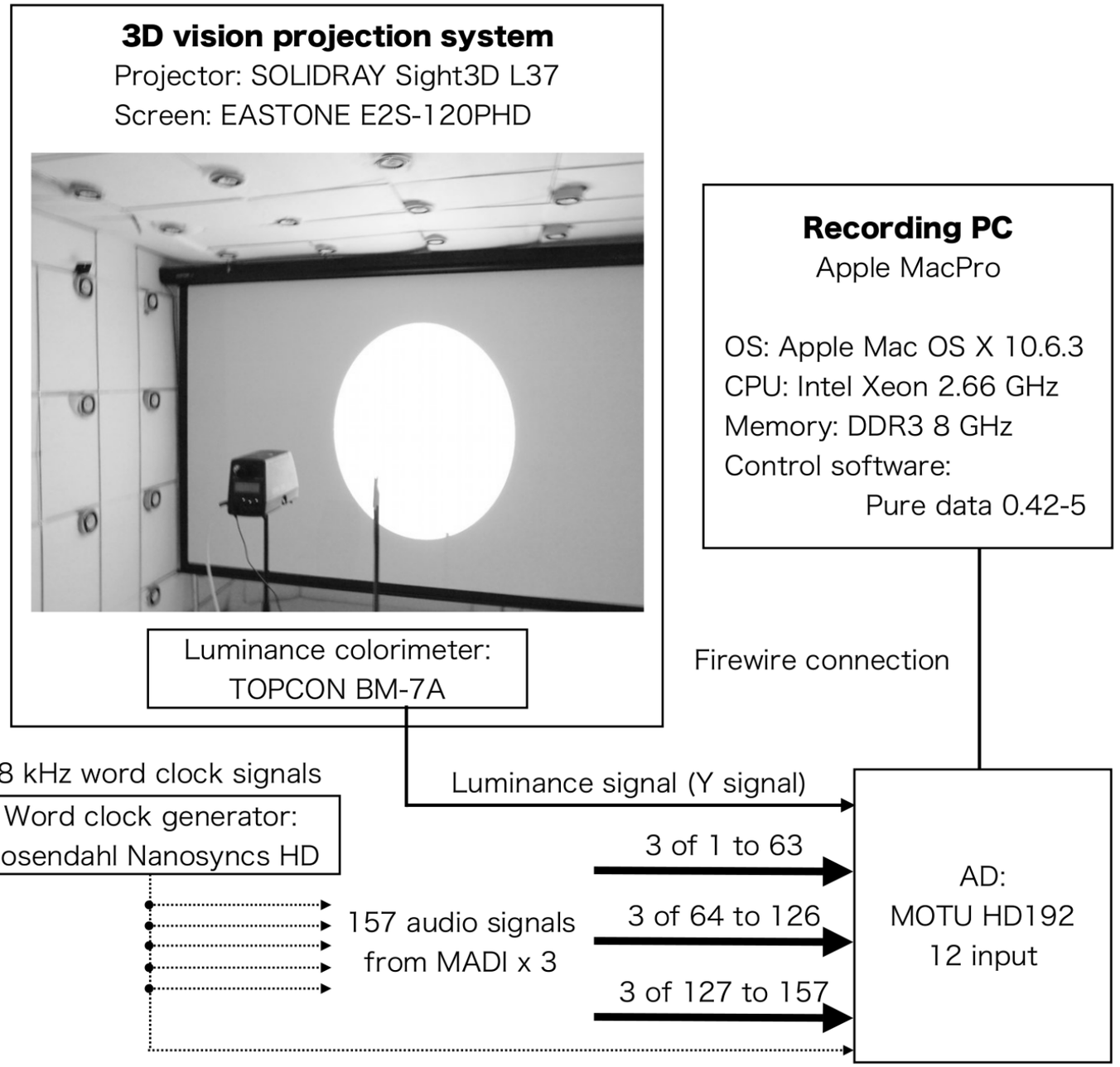
: Screen (x=3.49)

○ : Projector (x=0.15, y=1.56, z=2.15)

● : Viewing & Listening point
(x=2.29, y=1.39, z=1.35)

System synchronous measurement between audio and visual signals

- System synchronous measurement using luminance signal and TSP signals
- Result
  - Unsigned maximum asynchronous between the drawing on the screen and 157-audio stream was 51 samples (= 1.1 ms)
  - This value is much shorter than the detection threshold of audio-visual asynchrony by human observers



Concluding remarks

- Implementation of 3D audio-visual display
  - Realizing sound field reproduction system by 5th order Ambisonics using 157-loudspeaker array
  - Realizing completely synchronous 157 audio playback system introducing 3 MADI system and two audio drivers, CoreAudio and Jack Pilot
  - Implementation 3D audio-visual display system combining HOA system with 3D projection system
    - System synchronous between audio and visual signals was 1.1 ms and is much shorter than the detection threshold of human perception

Realizing not only actual environment but also virtual environment 3D audio-visual information reproduction

- Future works
  - Evaluation of the accuracy of reproduced sound field by proposed system
  - Experiments for clarifying human audio-visual and various multimodal perceptions
    - Control audio and visual information including spatial information

Concept of HOA

- Simple example of orthogonal expansion
  - Taylor expansion

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \cdots$$

- Orthogonal expansion of sound field information based on spherical harmonics in HOA

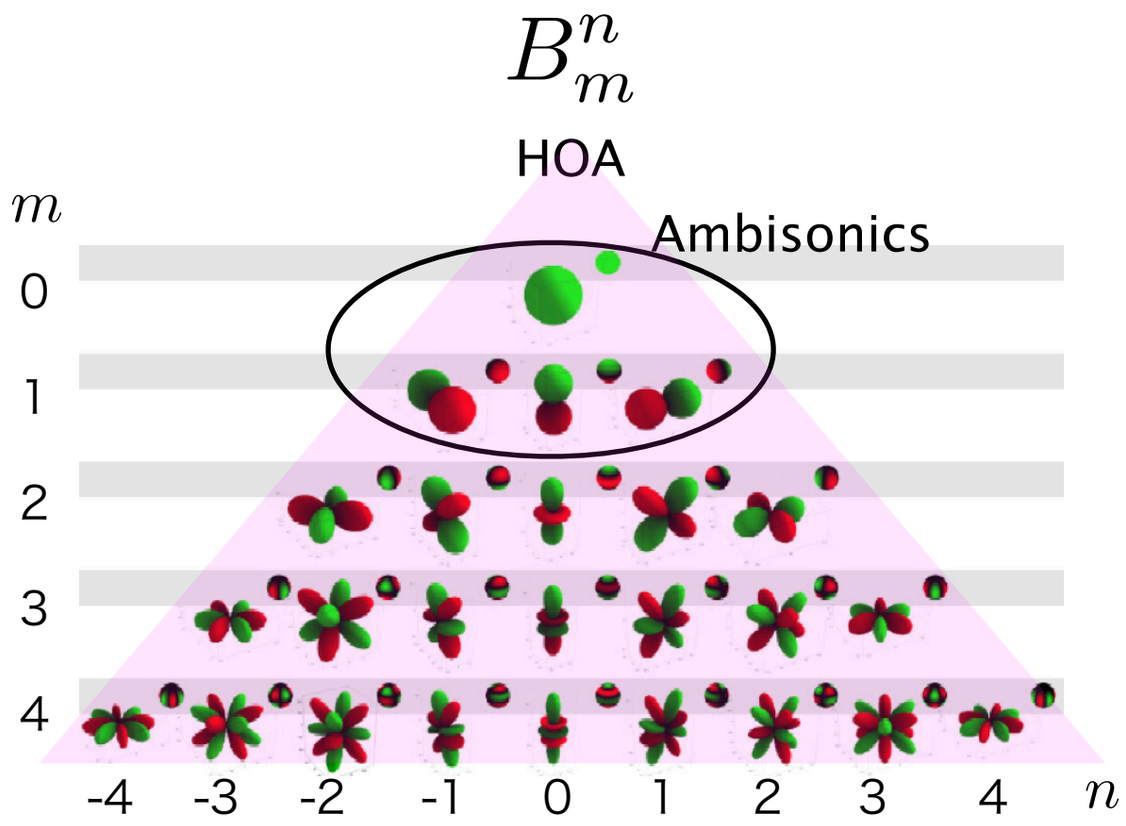
$$p(kr, \theta, \phi) = \sum_{m=0}^{+\infty} i^m j_m(kr) \sum_{n=-m}^{m} B_m^n Y_m^n(\theta, \phi)$$

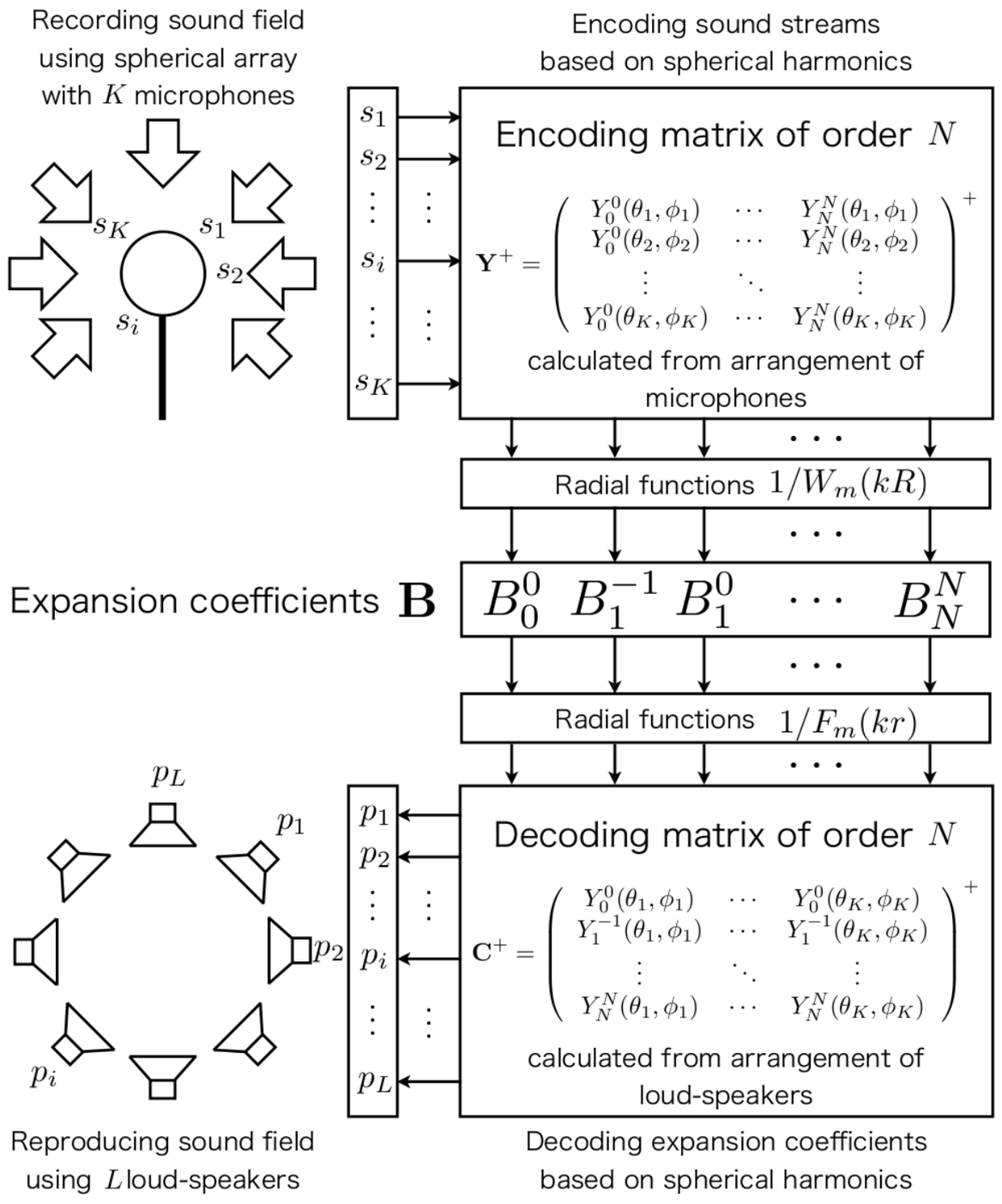
  - Sound information is decomposed to expansion coefficients B_m^n using spherical harmonics $Y_m^n(\theta, \phi)$ of each direction (θ, ϕ)
  - Using more higher order m, more accurate sound field is reproduced

$$\mathrm{Y}_{mn}(\theta, \varphi) = \begin{cases} N_m^0 P_m^0(\cos\varphi) & \text{if } n = 0, \\ \sqrt{2} N_m^n P_m^n(\cos\varphi)\cos n\theta & \text{if } n > 0, \\ \sqrt{2} N_m^n P_m^{-n}(\cos\varphi)\sin n\theta & \text{if } n < 0, \end{cases} \quad N_m^n = \sqrt{\frac{(2m+1)}{4\pi}\frac{(m-|n|)!}{(m+|n|)!}}$$

Sound field recording and reproduction based on HOA system

- Aspects of HOA system
  - Using more higher order m, more accurate sound field is reproduced
  - More than $(m+1)^2$ channels are needed to encode or decode order m
  - Recording system and reproduction system are independent each other

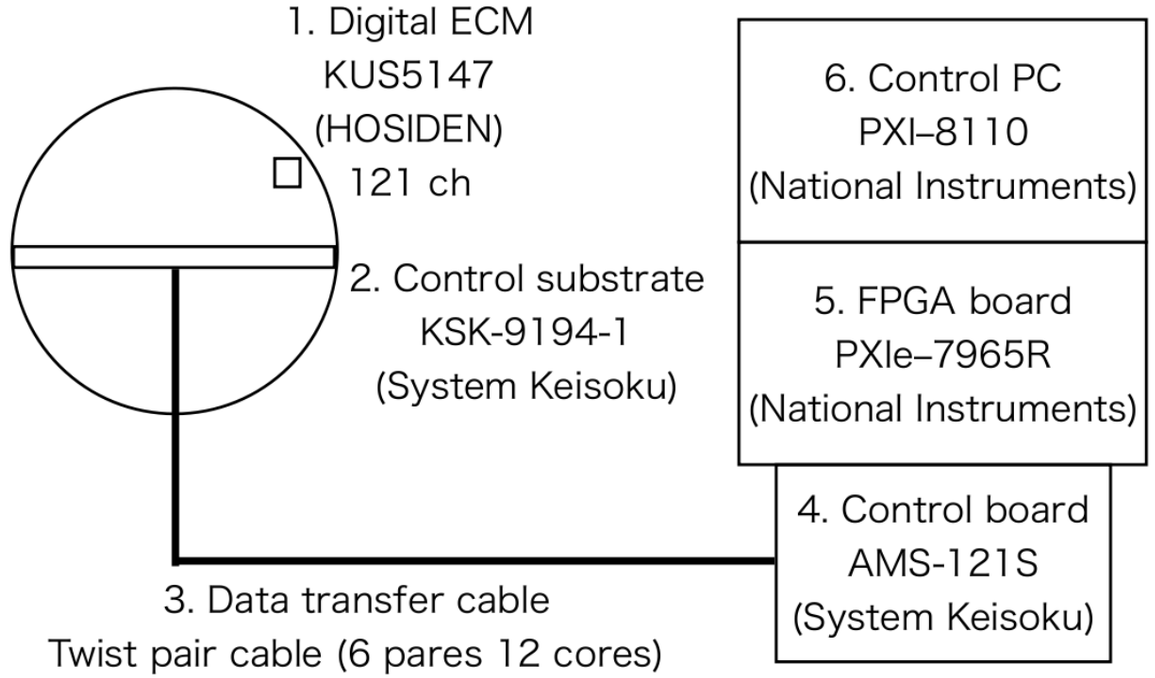




Implementation of sound field recording system based on HOA

T. Okamoto *et al.*, *Proc. SOIM-GCOE 2010* (in print)

- HOA recording system using a 121 spherical microphone array
  - 9th order decoding, which is highest order in the world, was realized
  - Completely synchronous 121-audio recording system using Digital Electric Condenser Microphone and FPGA board



Realizing actual sound field recording and reproduction system