ForumAcusticum
Budapest

RBA-VA 617-0

Sound Field Auralization System in Free Listening Positions Using Wave Field Synthesis and Head Related Transfer Function

Toshiyuki Kimura

Wataru Mizuno

Takanori Nishino

Katsunobu Itou

Kazuya Takeda

Introduction

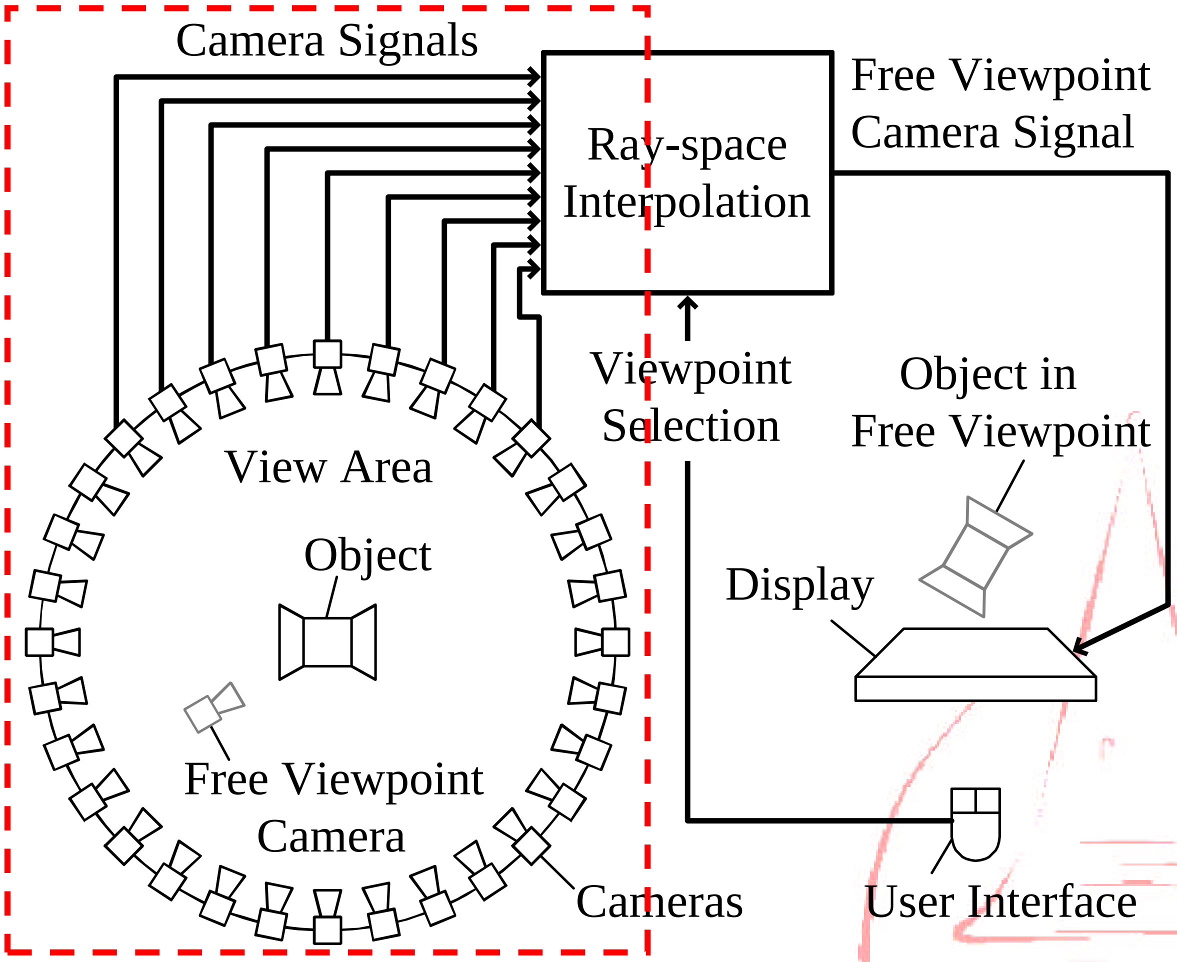
- More realistic communication system
  - Visual display technique
  - Sound field auralization technique
- Special visual display technique

Free Viewpoint Television (FTV) System

``Ultimate 3D TV''

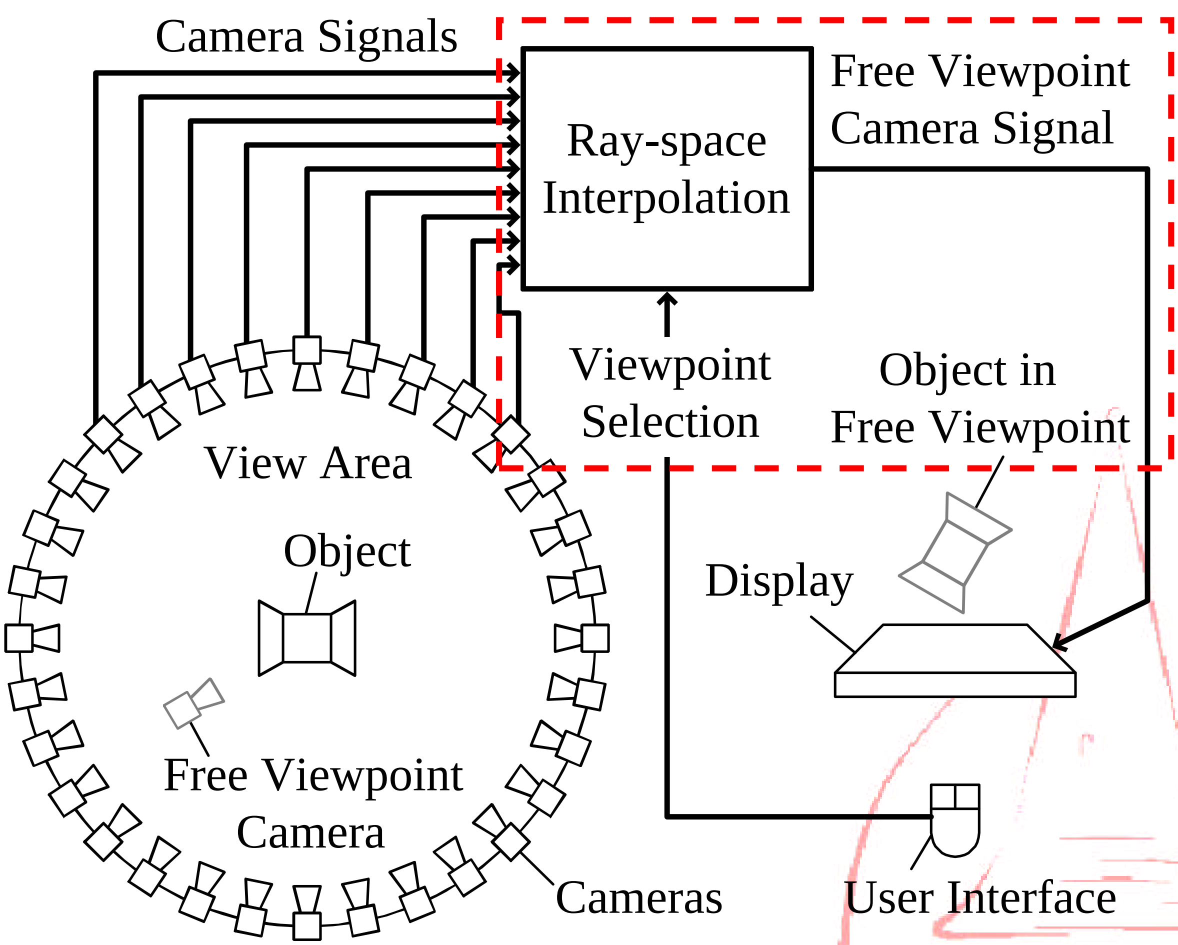
FTV System

- Images of the object are captured by cameras



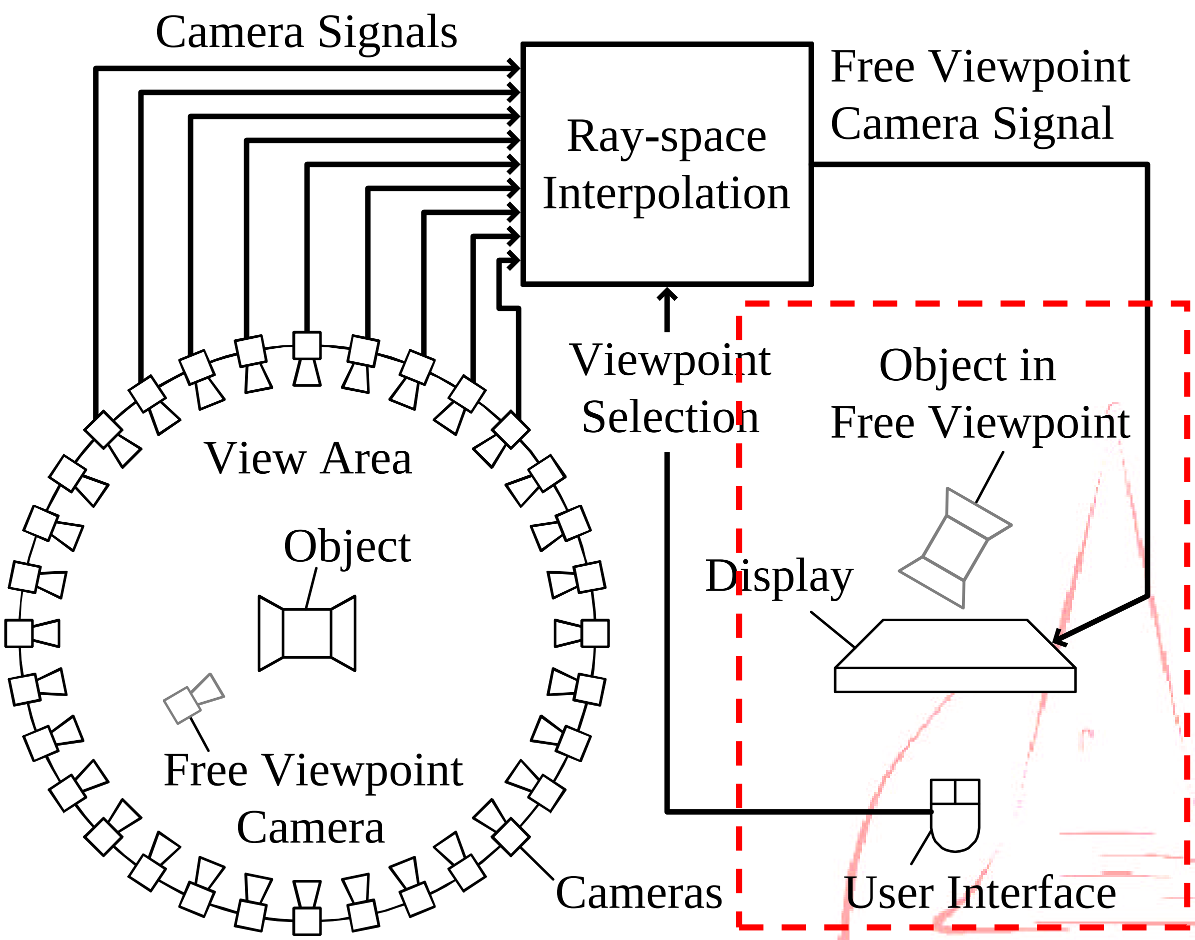
FTV System

- Free viewpoint camera signal is synthesized



FTV System

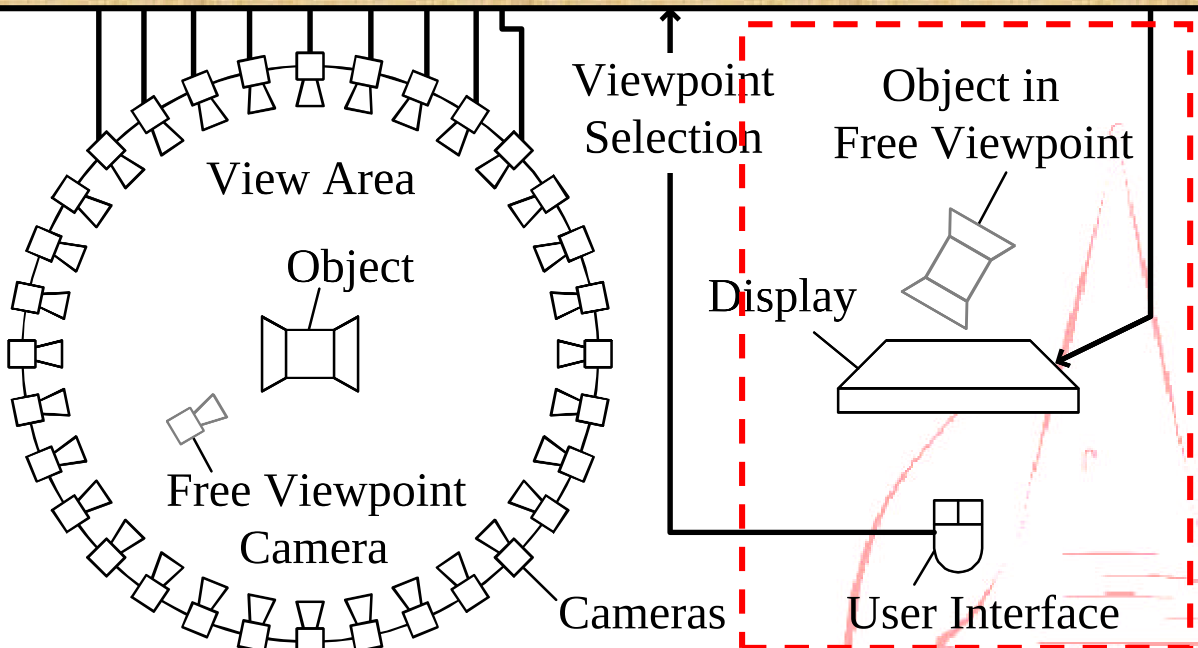
- Free Viewpoint image is displayed



FTV System

- Free Viewpoint image is displayed

User can freely view the object from virtually any viewpoint



Aim of Study

- Add sound information to FTV system
  - Develop a more realistic television system

↓

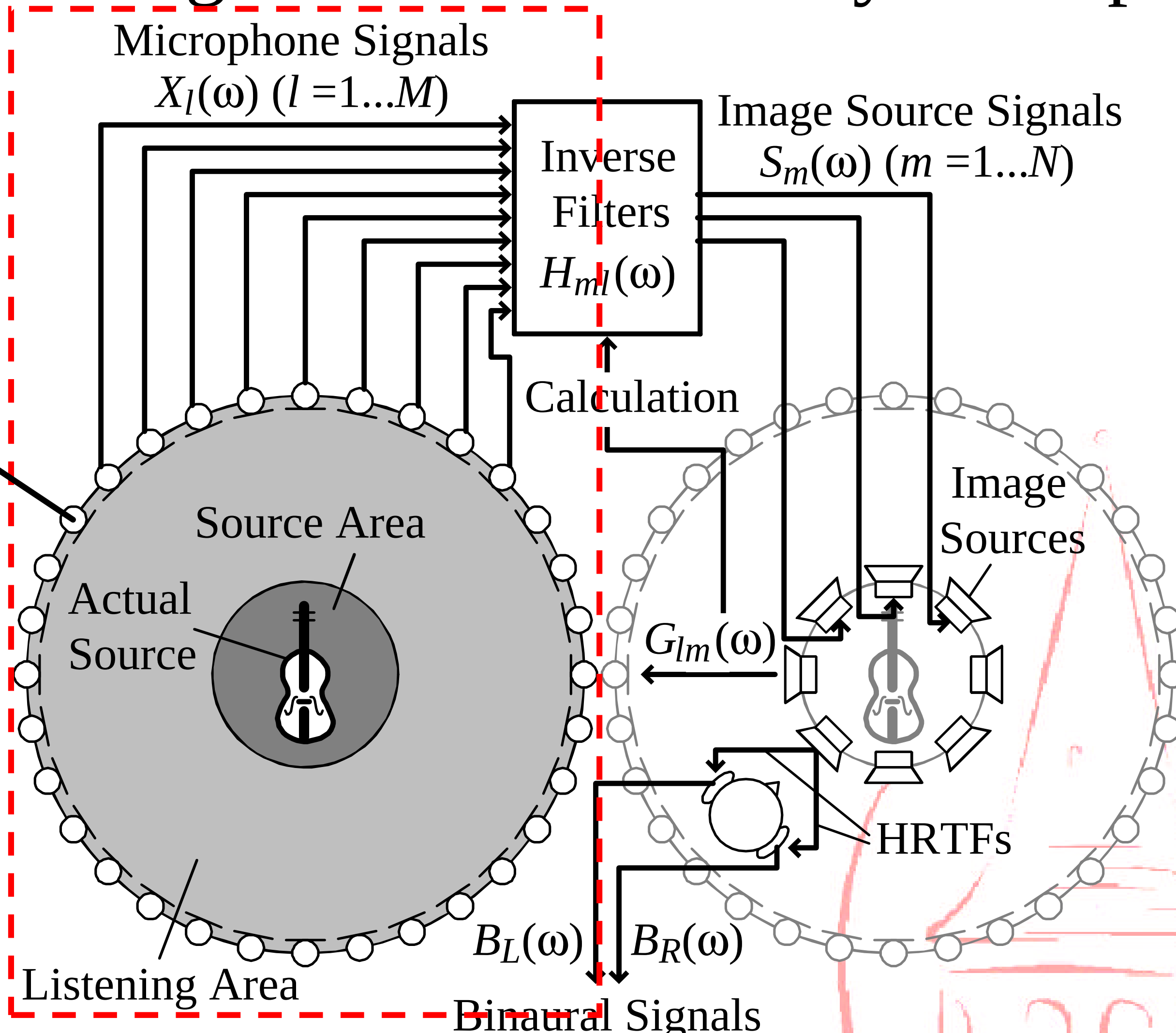
- Sound field auralization system in free listening positions
  - Head related transfer functions (HRTFs)
  - Wave field synthesis (WFS)

2. Sound Field Auralization System

Overview

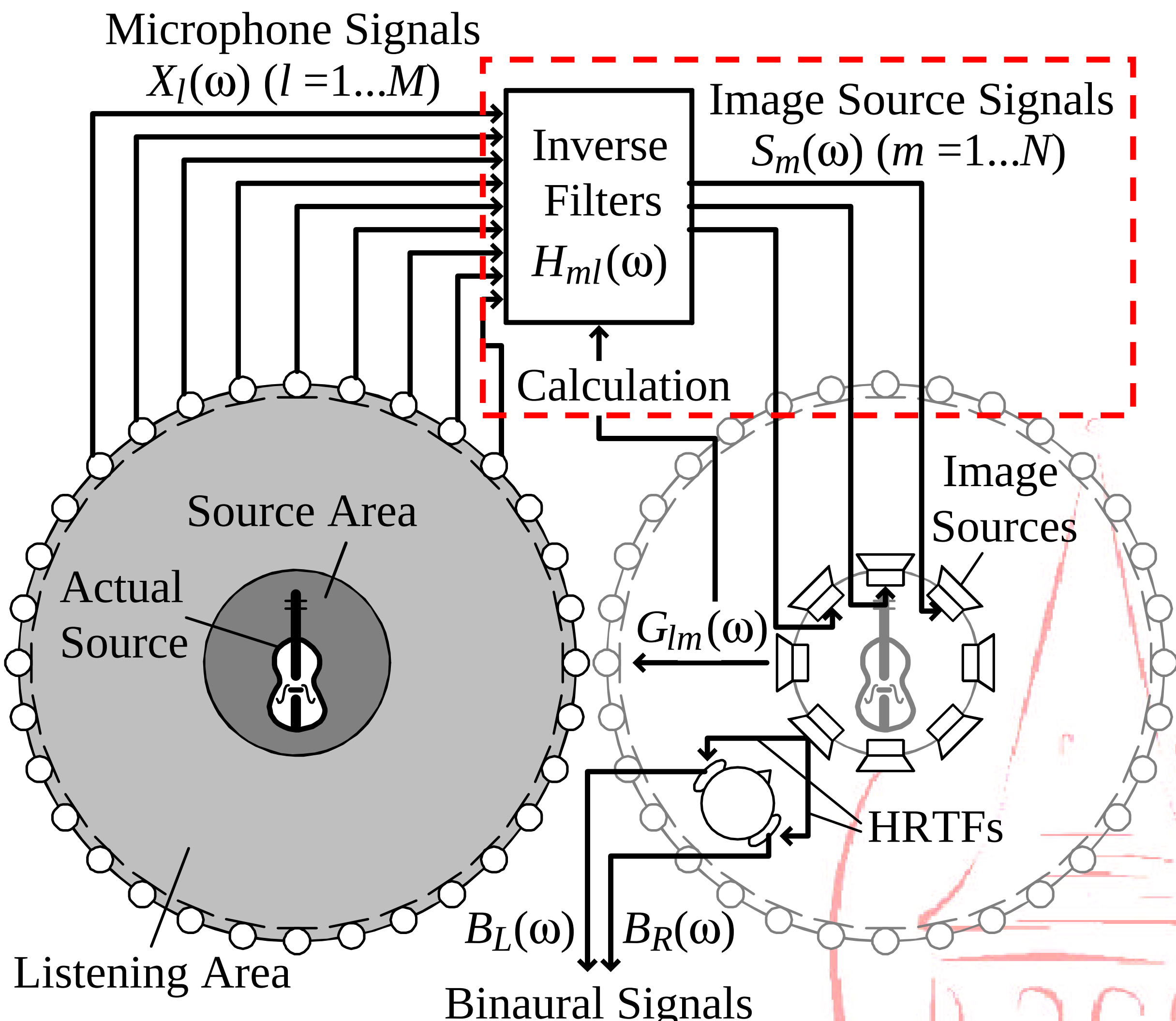
- Sound signals are recorded by microphones

Position …Same as cameras



Overview

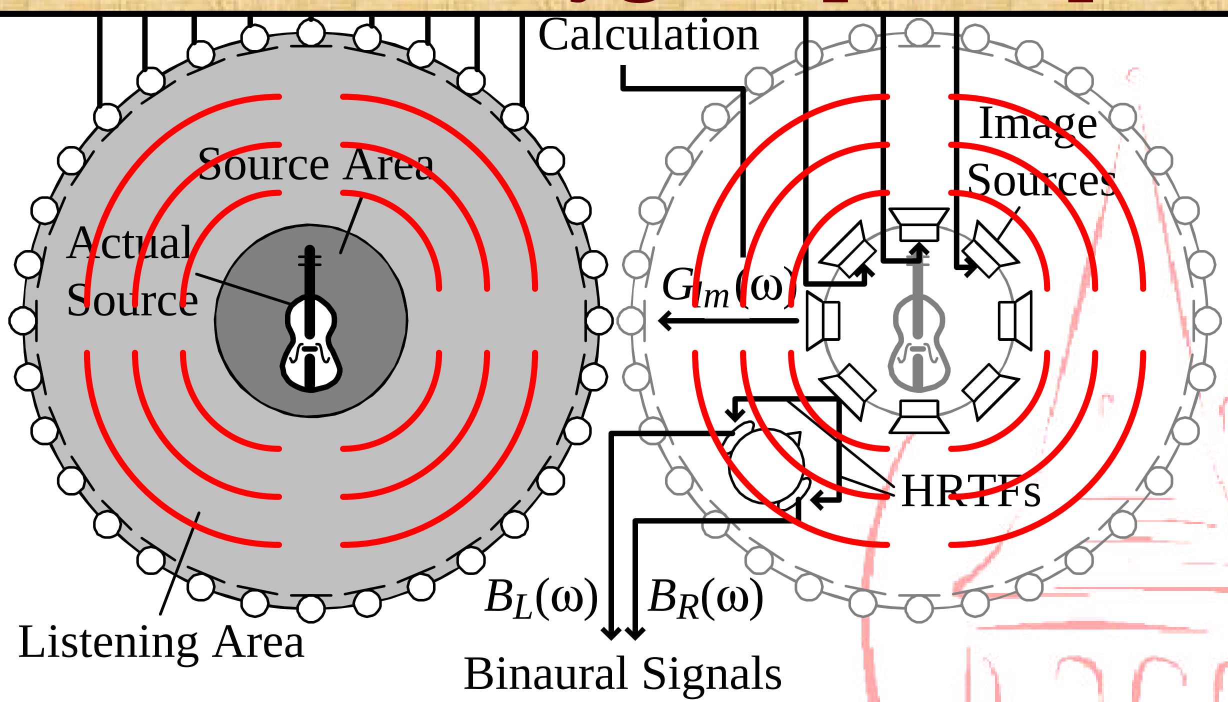
- Image source signals are estimated



Overview

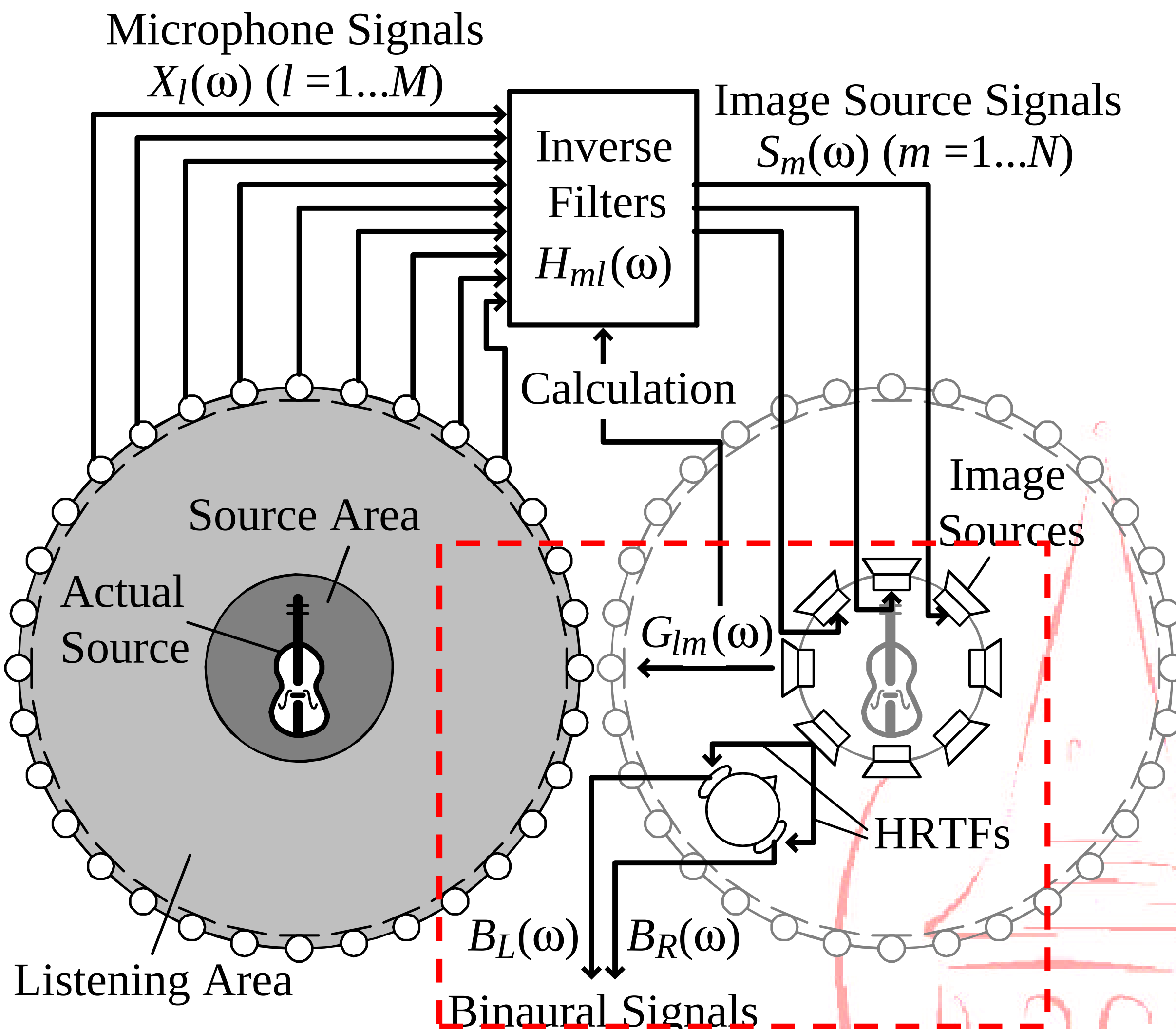
- Image source signals are estimated

Wave fronts of the listening area are synthesized by image sources based on Huygens principle



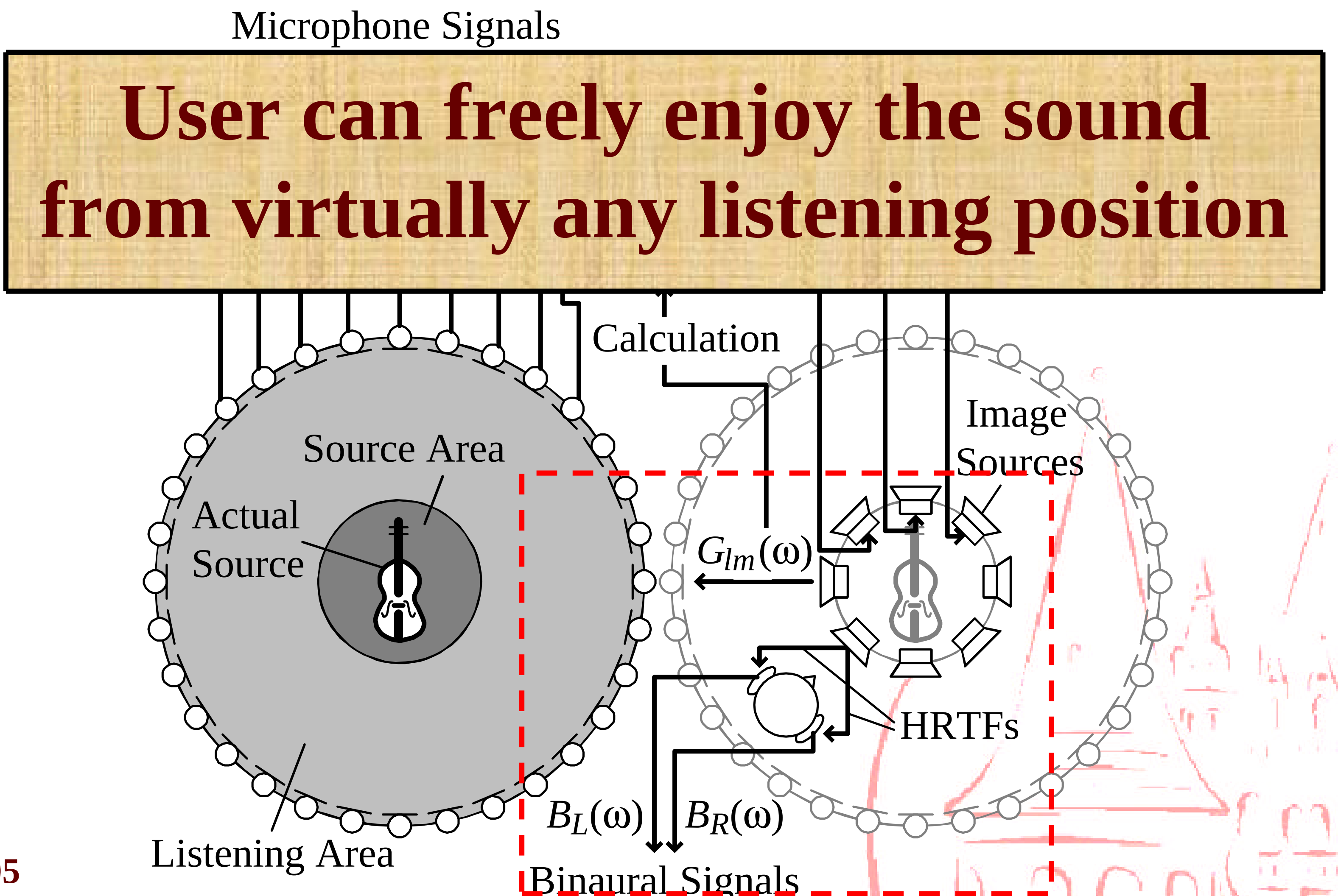
Overview

- Binaural signals are synthesized



Overview

- User listens to binaural signals by headphones



Estimation of Image Source Signals

- Microphone signals $X_l(w)$
  - Convolve room transfer functions (RTFs) $G_{lk}(w)$ to image source signals $S_k(w)$

$$X_l(w) = \sum_{k=1}^{N} G_{lk}(w) S_k(w)$$

N: The number of image sources

- Image source signals $S'_m(w)$
  - Convolve inverse transfer functions (ITFs) $H_{ml}(w)$ to microphone signals $X_l(w)$

$$S'_m(w) = \sum_{l=1}^{M} H_{ml}(w) X_l(w)$$

M: The number of microphones

Inverse Transfer Functions

- Be calculated from room transfer functions

$$\mathbf{G}(w)\mathbf{H}(w) = \mathbf{D}(w)$$

$$\mathbf{G}(w) = \begin{pmatrix} G_{11}(w) & \cdots & G_{M1}(w) \\ \vdots & \ddots & \vdots \\ G_{1N}(w) & \cdots & G_{MN}(w) \end{pmatrix} \quad \mathbf{H}(w) = \begin{pmatrix} H_{11}(w) & \cdots & H_{N1}(w) \\ \vdots & \ddots & \vdots \\ H_{1M}(w) & \cdots & H_{NM}(w) \end{pmatrix}$$

$$\mathbf{D}(w) = \begin{pmatrix} e^{-jw\frac{n_0}{F_s}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{-jw\frac{n_0}{F_s}} \end{pmatrix}$$

$\mathbf{G}(w)$: Room transfer function matrix
$\mathbf{H}(w)$: Inverse transfer function matrix
n_0: Delay samples
F_s: Sampling frequency

$$\mathbf{H}(w) = \mathbf{G}^{+}(w)\mathbf{D}(w)$$

$\mathbf{G}^{+}(w)$: Moore-Penrose pseudo inverse matrix of $\mathbf{G}(w)$

Synthesis of Binaural Signals

- Binaural signals $B_L(w)$, $B_R(w)$
  - Convolve HRTFs $I_L(d_m, f_m, w)$, $I_R(d_m, f_m, w)$ to image source signals $S_m(w)$

$$B_L(w) = \sum_{m=1}^{N} q(\Delta_m) I_L(d_m, f_m, w) S_m(w)$$

$$B_R(w) = \sum_{m=1}^{N} q(\Delta_m) I_R(d_m, f_m, w) S_m(w)$$

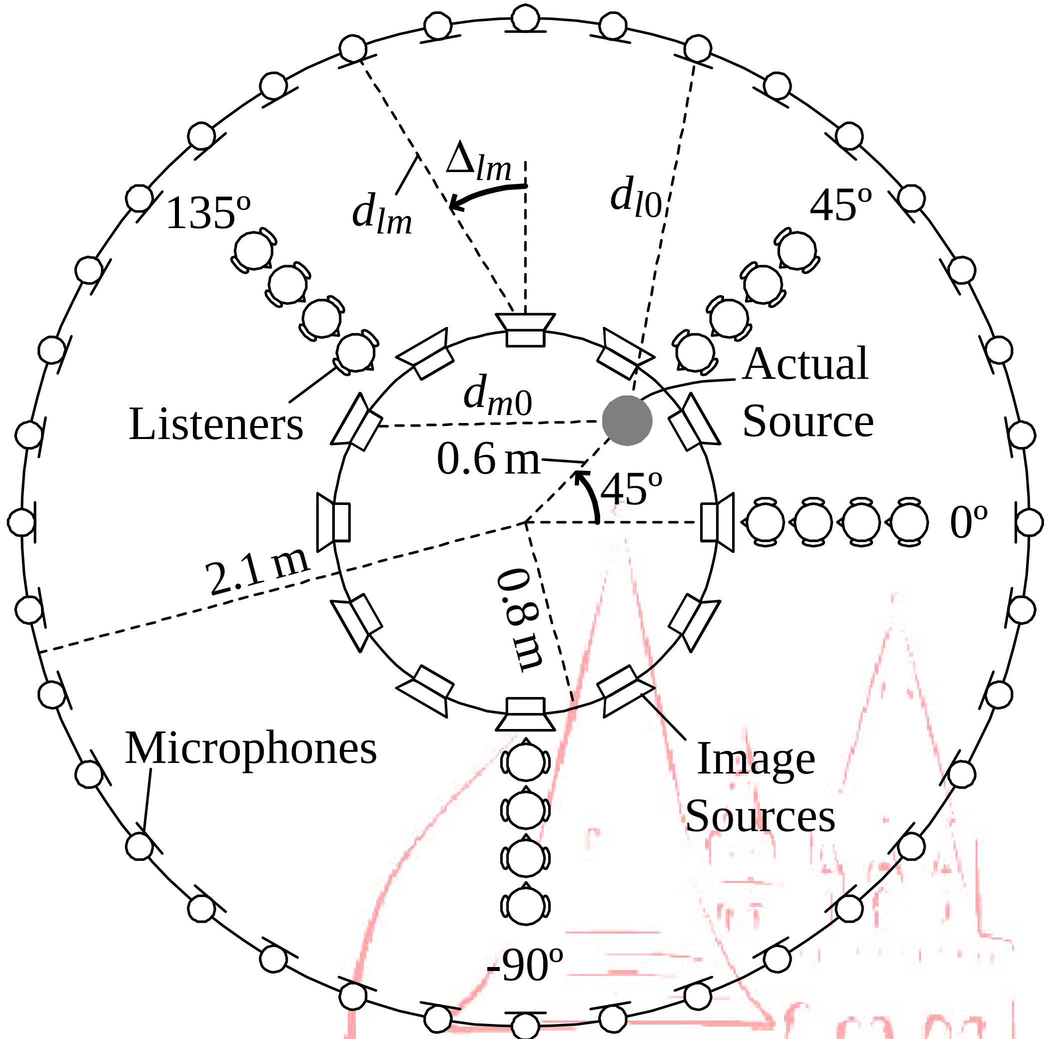
$$q(\Delta_m) = \begin{cases} \cos\Delta_m & |\Delta_m| \le 90^\circ \\ 0 & |\Delta_m| > 90^\circ \end{cases}$$

d_m: Distance between the *m*th image source and the listening position
f_m: Azimuth angle of the *m*th image source
Δ_m: Azimuth angle of the listening position in the *m*th image source
$q(\Delta_m)$: Directivity function of the *m*th image source

3. Evaluation of Performance

Experimental Arrangement

- Sound field
  - free space
- Actual source
  - 0.6 m distance
  - 45° azimuth angle
- Array's Radius
  - Microphones
    - 2.1 m
  - Image sources
    - 0.8 m



Synthesis of Microphone Signals

- Microphone signals $x_l(n)$
  - Be calculated from actual source signal

$$x_l(n) = \frac{1}{d_{l0}} s_0\left[n - \mathrm{round}\left(\frac{d_{l0}F_s}{c}\right)\right] \quad (l = 1 \ldots M)$$

$s_0(n)$: Actual source signal

Piano sound (sampling frequency…32 kHz, duration…5 s)

d_{l0}: Distance between the actual source and the lth microphone

Number of image sources (N)	12, 18, 24, 36, 48
Number of microphones (M)	N, N×2, N×3, N×4
Sampling Frequency (F_s)	32 kHz
Sound velocity (c)	340 m/s

Room Transfer Functions

- Room transfer functions between image sources and microphones $g_{lm}(n)$
  - Be calculated by computer

$$g_{lm}(n) = \frac{q(\Delta_{lm})}{d_{l0}} d\left[n - \mathrm{round}\left(\frac{d_{lm} F_s}{c} \right) \right] \quad (m = 1 \ldots N, l = 1 \ldots M)$$

$d\,(n)$: Dirac's delta function

d_{lm}: Distance between the *m*th image source and the *l*th microphone

Δ_{lm}: Azimuth angle of the *l*th microphone in the *m*th image source

$q(\Delta_{lm})$: Directivity function of the *m*th image source

Inverse Transfer Functions

- Inverse transfer functions $h_{ml}(n)$
  - Be calculated from room transfer functions

Calculation conditions of ITFs

FFT frame length	2048 samples
Calculated bandwidth	250 Hz–13333Hz
Delay samples (n_0)	512 samples
ITF length	1024 samples

- Image source signals $s_m(n)$
  - Convolve inverse transfer functions (ITFs) $h_{ml}(n)$ to microphone signals $x_l(n)$

Objective Evaluation

- Signal-to-Deviation Ratio (SDR)
  - Estimation accuracy of image source signals

$$\mathrm{SDR[dB]} = 10\log_{10} \frac{\sum_{m=1}^{N}\sum_{n}\{s'_m(n-n_0)\}^2}{\sum_{m=1}^{N}\sum_{n}\{s'_m(n-n_0)-s_m(n)\}^2}$$

$s_m(n)$: The *m*th estimated image source signal
$s'_m(n)$: The *m*th reference image source signal

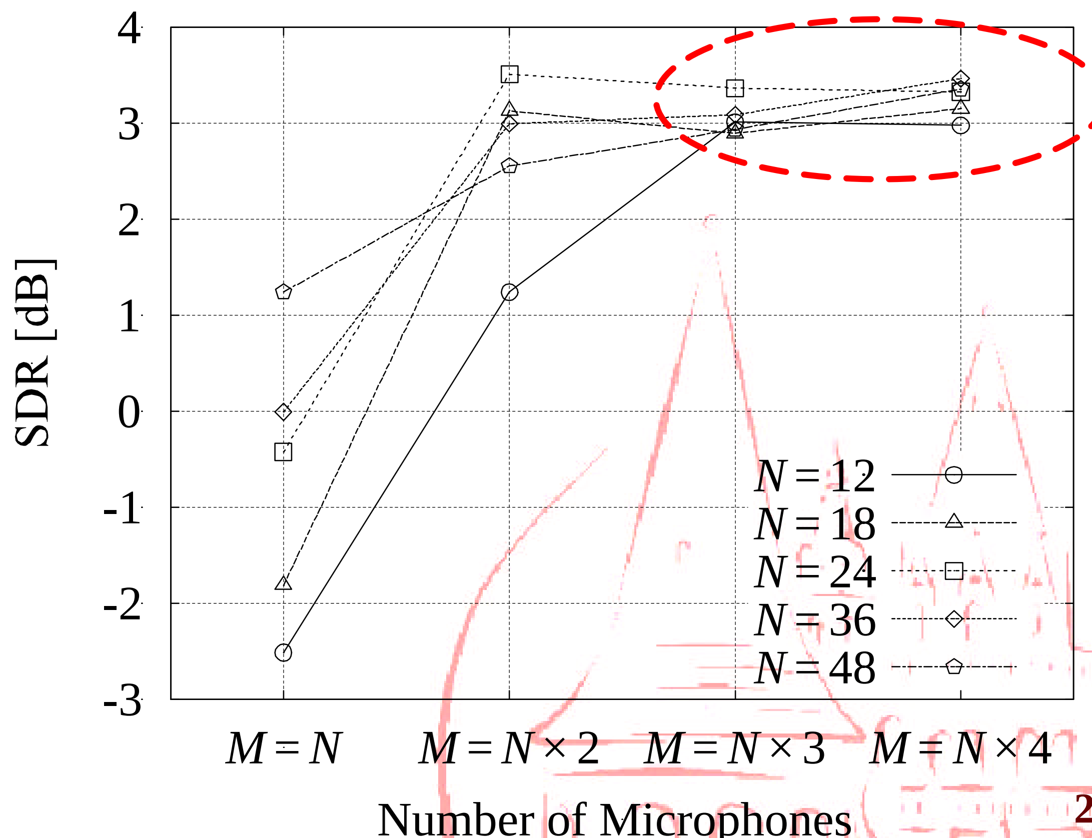
$$s'_m(n) = \frac{1}{d_{m0}} s_0\left[n - \mathrm{round}\left(\frac{d_{m0}F_s}{c}\right)\right]$$

d_{m0}: Distance between the actual source and the *m*th image source

SDR Results

- Number of microphones (*M*)
- Number of image sources (*N*)
  - $M > N \times 3$…SDRs are constant

M =N ´ 4



HRTFs of Close Distances

- Piezoelectric dodecahedral loudspeaker
- Distance
  - From 0.2 to 1 m
- Azimuth angle
  - 1° interval

Room temperature	24.0 °C
Background noise level	13.8 dB(A)
Sound pressure level	69.0 dB(A)
Sampling frequency	48 kHz
TSP signal length	32768 samples
HRTF length	512 samples

Synthesis of Binaural Signals

- Binaural signals $b_L(n)$, $b_R(n)$
  - Convolve measured HRTFs $i_L(d_m, f_m, n)$, $i_R(d_m, f_m, n)$ to image source signals $s_m(n)$

$$b_L(n) = \sum_{m=1}^{N} q(\Delta_m)[i_L(d_m, f_m, n) * s_m(n)]$$

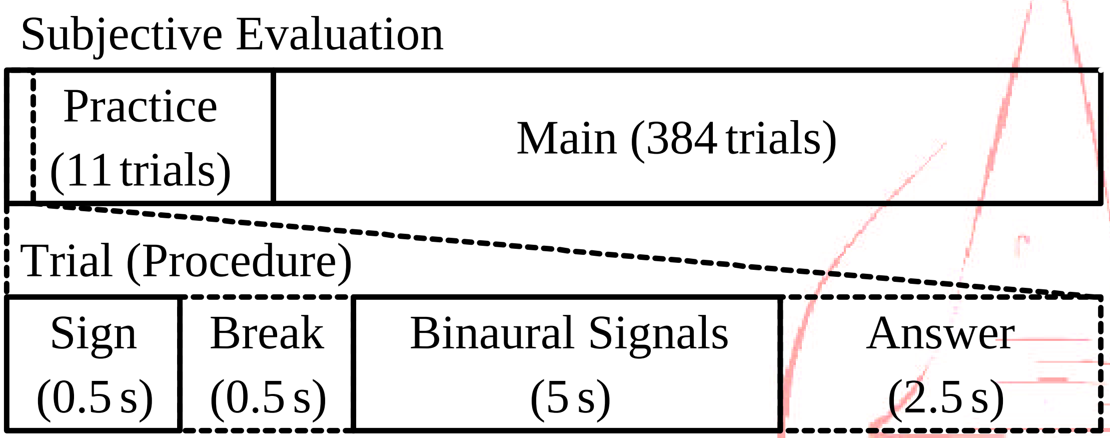
$$b_R(n) = \sum_{m=1}^{N} q(\Delta_m)[i_R(d_m, f_m, n) * s_m(n)]$$

$$q(\Delta_m) = \begin{cases} \cos\Delta_m & |\Delta_m| \le 90^\circ \\ 0 & |\Delta_m| > 90^\circ \end{cases}$$

d_m: Distance between the *m*th image source and the listening position
f_m: Azimuth angle of the *m*th image source
Δ_m: Azimuth angle of the listening position in the *m*th image source
$q(\Delta_m)$: Directivity function of the *m*th image source

Subjective Evaluation

- Localization test
  - Accuracy of the directional perception
  - Subject
    - 5 male students
  - Listening equipment
    - Headphone (Audio-Technica ATH-A1000)



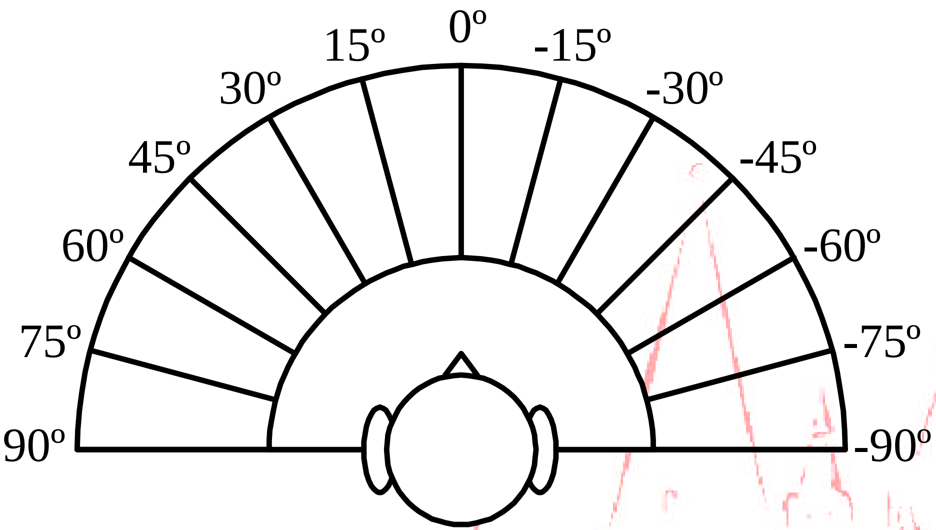
Design

- Comparison of localization results
  - 5 Image source conditions (*N*=12, 18, 24, 36, 48)
  - Actual Source condition

	Factor	Level
Practice (11)	= 1 condition × 11 directions	Actual Source -75°, -60°, …, 60°, & 75°
Main (384)	= 6 conditions × 4 distances × 4 azimuths × 4 repetitions	*N*=12, 18, 24, 36, 48, & AS 1.0, 1.2, 1.4, & 1.6 m 0, 45, 135, & -90°

Procedure

- Instruction
  - Identify the direction of sound
  - Mark on an answer sheet
- Answer sheet
  - 15^o interval

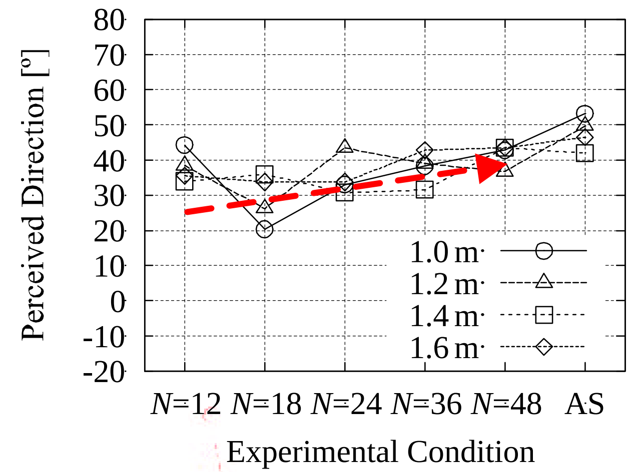


Localization Results

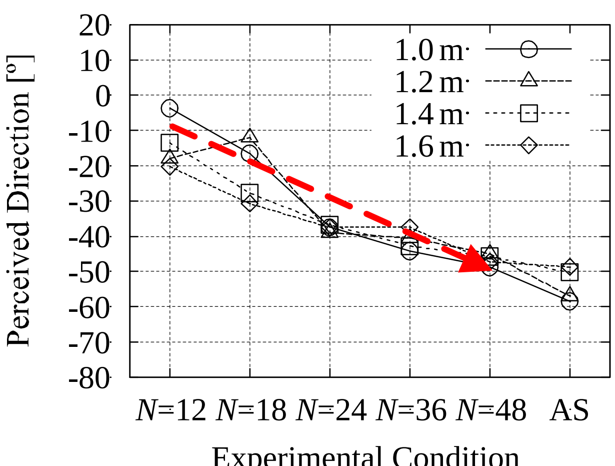
- Perceived direction approaches to that of the actual source

The directional perception is the same if the number of image sources is sufficient

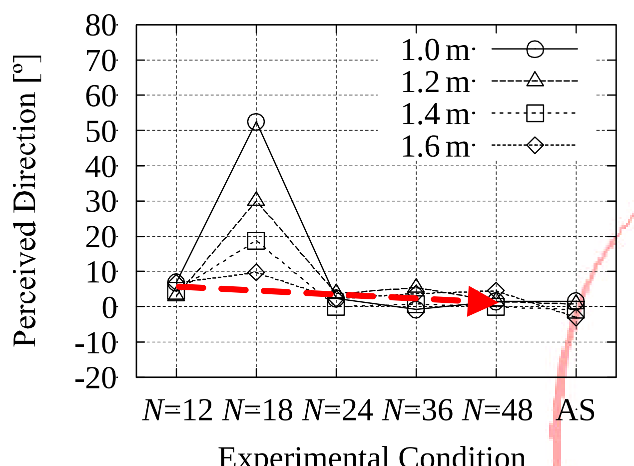
Azimuth = 135º



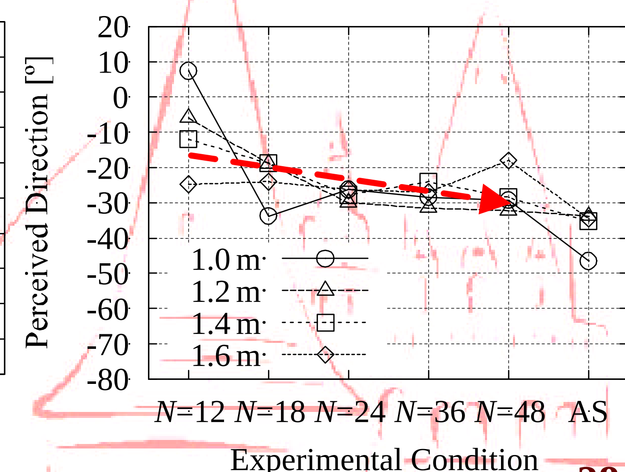
Azimuth = 0º



Azimuth = 45º



Azimuth = -90º



Conclusion

- Sound field auralization system in free listening positions was proposed
- SDR results
  - Image source signals can be estimated if the number of microphones is sufficient
- Localization results
  - The directional perception can be reproduced if the number of image sources is sufficient
- Future works
  - Actual environment, 3D sound field