Blind directivity estimation of a sound source in a room using a surrounding microphone array

Takuma Okamoto\textsuperscript{(1)(2)}, Yukio Iwaya\textsuperscript{(1)(3)} and Yôiti Suzuki\textsuperscript{(1)(3)}

\textsuperscript{(1)} Research Institute of Electrical Communication
\textsuperscript{(2)} Graduate School of Engineering
\textsuperscript{(3)} Graduate School of Information Sciences
Tohoku University, Japan
Introduction: editable sound field system

- Sensor network technologies in near future
  - multi-channel microphone array system
    - not only recording sound field
    - but also extracting sound field properties

- Contribution of Editable sound field system
  - Decomposing sound field properties accurately from recording sound information
  - Rendering extracted sound properties
    - not only original sound field
    - but also modified sound field

Surrounding 157 microphone array  

- 157 microphones is installed on all four walls and the ceiling
  - All microphones are installed 30 cm inside from all four walls and the ceiling
  - They are separated from each other by 50 cm
- Enable synchronous recording of 157 channels at the sampling frequency of 48 kHz with the linear PCM audio format using 4 PCs and 14 units of A/D controlled by PureData
Directivity of sound source

- Sound sources in actual environments have no omni-directional feature
- but they do have directivity in radiation
  - human voice, instrumental sound and so on...
- it is important to consider the directivity of a sound source according to a listening point when synthesizing a high-definition 3D sound field

Previous studies for estimating directivity of sound source

- In almost researches, directivity was recorded only in anechoic environment
- Nakadai et al. was estimated only front direction of the sound source in a reverberant environment (2005)

Our research aim

- Developing method for estimating and decomposing all-around directivity of a sound source in a reverberant environment along with information of the estimated source position and the original sound signal
Proposed 1: Definition of Directivity model of a sound source in a room

- **SIMO (Single-Input Multiple-Output) model including directivity of a sound source**
  - \( s(n) \): Source signal
  - \( d(\theta_i, n) \): each directivity of a direction \( \theta_i \)
  - \( h_{R_i}(n) \): each reverberant component
  - \( r(\theta_i) \): each distance between the source and each microphone \( i \)
  - \( x_i(n) \): each observed signal at microphone \( i \)

\[
x(n) = \sum_{k=0}^{\infty} s(n) \cdot d(\theta_k, n) \cdot h(\theta_k, n) = s(n) \cdot \sum_{k=0}^{\infty} \{d(\theta_k, n) \cdot h(\theta_k, n)\}
\]

\[
x(n) = s(n) \cdot \{d(\theta_0, n) \cdot h(\theta_0, n) + \sum_{k=1}^{\infty} d(\theta_k, n) \cdot h(\theta_k, n)\}
\]

\[
x(n) = s(n) \cdot \left\{ d(\theta_0, n) \cdot \frac{1}{r(\theta_0)} + d(\theta_0, n) \cdot h'(\theta_0, n) + \sum_{k=1}^{\infty} d(\theta_k, n) \cdot h(\theta_k, n) \right\}
\]

\[
= s(n) \cdot \{h_D(n) + h_R(n)\}
\]

\[
= s(n) \cdot h(n)
\]
Proposed 2: Decomposition of room transfer function of directivity and reverberant component

- Relationship between directivity and reverberant component
  - If sound source is omni-directional feature
    - direct sound component is only first response
  - When sound source has a directivity
    - directivity component is first to several samples responses

- Decomposition of directivity and reverberant component
  1. Estimating total impulse response
     - Blind identification or dereverberation method
  2. $h_{Di}(n)$ can be extracted as the early response from the first response to time $t = 2d/c$
     - cutting impulse response as the first $t$ response
  3. Amplitude correction corresponding to each distance $d_i(n) = r_i h_{Di}(n)$
     - each distance is estimated from estimated sound source position

![Diagram](image)
Estimating room impulse response using estimated sound source signal

- Estimating room impulse response
  - Blind identification is difficult problem
  - Estimating room impulse response using estimated sound source and received signals
    - Sound source signal can be estimated by a dereverberation method
    - We proposed a dereverberation algorithm, White-LIME

\[
\hat{h}_i(n) = (E\{\hat{s}(n)\hat{s}^T(n)\})^{-1}E\left[\begin{array}{c}
  x_i(n)s(n) \\
  x_i(n-1)s(n-1) \\
  \vdots \\
  x_i(n-M+1)s(n-M+1)
\end{array}\right]
\]

[Diagram showing the process of estimating the room impulse response]
Directivity measurement in an anechoic chamber

- Directivity measurement as impulse response
  - 1-way loudspeaker (Micropure; AP5001)
  - Measurements were taken from 0 deg to 180 deg in horizontal direction with a clockwise rotation by 15 deg in an anechoic chamber
  - Time Stretched Pulse (TSP) signal was used for measurement
    - The impulse responses for 13 directions were measured

![Diagram of anechoic chamber setup]

- Anechoic Room
- Loudspeaker: Micropure AP5001
- Loudspeaker Amp.: DENON PMA-1500AE
- Microphone: ONO SOKKI MI-1233
- Microphone Amp.: ONO SOKKI MI-3110
- Control PC:
  - CPU: Intel Core2Duo E8500 3.16 GHz
  - OS: Microsoft Windows XP SP3
  - Control Software: Degidesign Pro tools 7.4 M-Powered
- AD / DA: BEHRINGER ADA8000
- Audio Interface: M-Audio ProFire Lightbridge
- Word Clock Generator: Rosendahl Nanosyncs HD
- 48 kHz Word Clock
- Distance: 1.5 m
Room impulse response measurement in a room

- Measurement of room impulse response which includes both directivity of a sound source and reverberant component
- Sound signal and loudspeaker are the same as an anechoic measurement
- Measurement was taken by using surround 157-microphone array

Diagram:

- Control signal
  - UDP connection
  - PC1 (Master) Apple MacPro 46+1 ch
  - PC2 (Slave) Apple MacPro 34+1 ch
  - PC3 (Slave) Apple MacPro 46+1 ch
  - PC4 (Slave) Apple MacPro 31+1 ch
- A/D · D/A MOTU HD192
  - 4 units
  - 3 units
  - 4 units
  - 3 units
- Microphone amplifiers
  - B&K Type 2694
  - 10 units
- Control PC
  - OS: Microsoft Windows XP
  - Software: B&K BZ 5291
- Microphones
  - B&K Type 4951
  - 157 ch
- Loudspeaker
  - Micropure AP5001
- Surrounding microphone array room

Recording PCs spec
- Apple MacPro
  - OS: Apple Mac OS X 10.6.3
  - CPU: Intel Xeon 2.66 GHz
  - Memory: DDR3 8 GHz
  - Control software: Pure data 0.42-5
Recording conditions

- Two patterns of the arrangements of the loudspeaker
- 28 microphones \((z = 1.0\ m)\) in the room
- The reverberation time of this room was about 0.15 s
- The order of the room impulse responses was 7200

Pattern A

![Pattern A diagram]

Pattern B

![Pattern B diagram]
Calculating 157 received signals $x_i(n)$
- Convolving the source signal $s(n)$ to each measured impulse response $h_i(n)$
- The source signal was a musical piece (2.7 s) from RWC music database
- The measured impulse responses were downsampled 48 kHz to 44.1 kHz
  - The order of all responses was shortened from 7200 to 6615

Estimating impulse responses
- The estimated source signal $\hat{s}(n)$ was obtained using White–LIME
- Each estimated impulse response $\hat{h}_i(n)$ was calculated from each observed signal $x_i(n)$ and the estimated source signal $\hat{s}(n)$
  - The average of the score of the SDR of each original response $\hat{h}_i(n)$ was 62.6 dB
  - The estimated responses can be inferred accurately

Decomposing directivity from room impulse responses
- The distance between each microphone and the wall was 30 cm
- The clipping length of each estimated response was approximately $44,100 \times 2 \times 0.6/340 \approx 78$ taps
- Amplitude correction was taken from estimated sound source position and each microphones
Results of Amplitude–frequency characteristics

Original directivity

Measured impulse response (A)

Estimated directivity (A)

Measured impulse response (B)

Estimated directivity (B)
Results of 2 kHz in the 1/3 octave-band

Original directivity

Measured impulse response (A)

Estimated directivity (A)

Measured impulse response (B)

Estimated directivity (B)
Results of similarity analysis

Confirming the effectiveness of our proposed method

- Similarity based on the nearest neighbor method
  - Pattern vector \( P(f) = (x_{\theta_1}(f), x_{\theta_2}(f), \ldots, x_{\theta_N}(f)) \)
  - Similarity \( S(P_1(f), P_2(f)) = 1.0 - \frac{|P_1(f) - P_2(f)|}{|P_1(f)|} \)

Our proposed method can estimate directivity patterns that closely resemble real ones up to around 16 kHz
Concluding remarks

Conclusions

- Estimating directivity of a sound source in a reverberant environment
  - Using information of the estimated source position and the original sound signal
  - Cutting estimated impulse responses based on the distance between each microphone and nearest wall
  - The simulation results demonstrate the availability of our proposed method

Future works

- Our method presents a problem in estimating the sound source directivity when the walls and microphones are close to each other
  - Widening the applicability of our proposed method is development of a method extracting the early part from an impulse response with overlapped reflections
- Estimating directivity of a sound source not only 2-dimensional direction but also 3-dimensional direction for 3D sound field analysis and synthesis