SPATIAL CODING BASED ON THE EXTRACTION OF MOVING SOUND SOURCES IN WAVEFIELD SYNTHESIS

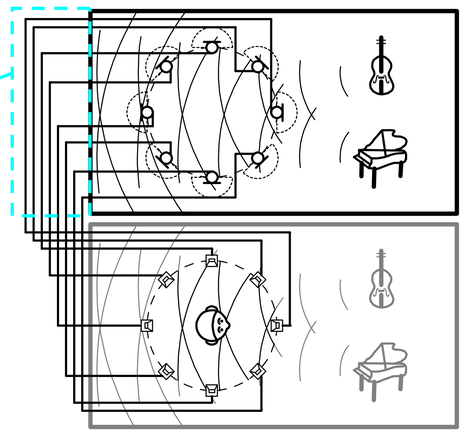
Toshiyuki Kimura[1] *Kazuhiko Kakehi*[2] *Kazuya Takeda*[1] *Fumitada Itakura*[3]
[1]Graduate School of Information Science, Nagoya University [2]School of Computer and Cognitive Sciences, Chukyo University
[3]Faculty of Science and Technology, Meijo University

1. INTRODUCTION

Sound Field Auralization Based on Wavefield Synthesis

- Synthesis of wave fronts at the listening area according to Huygens principle
- The number of channel signals is very large

The amount of data transmitted needs to be reduced

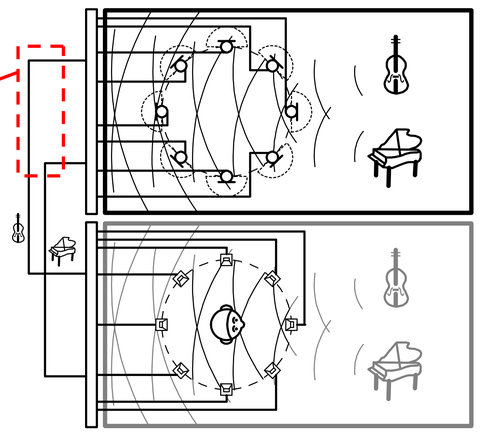


The amount of data transmitted
…The number of channel signals

Spatial Coding Method Based on the Extraction of Sound Sources

- The amount of data transmitted
  - The number of channel signals → The number of sound sources
- Conventional studies: Sound sources are not moving

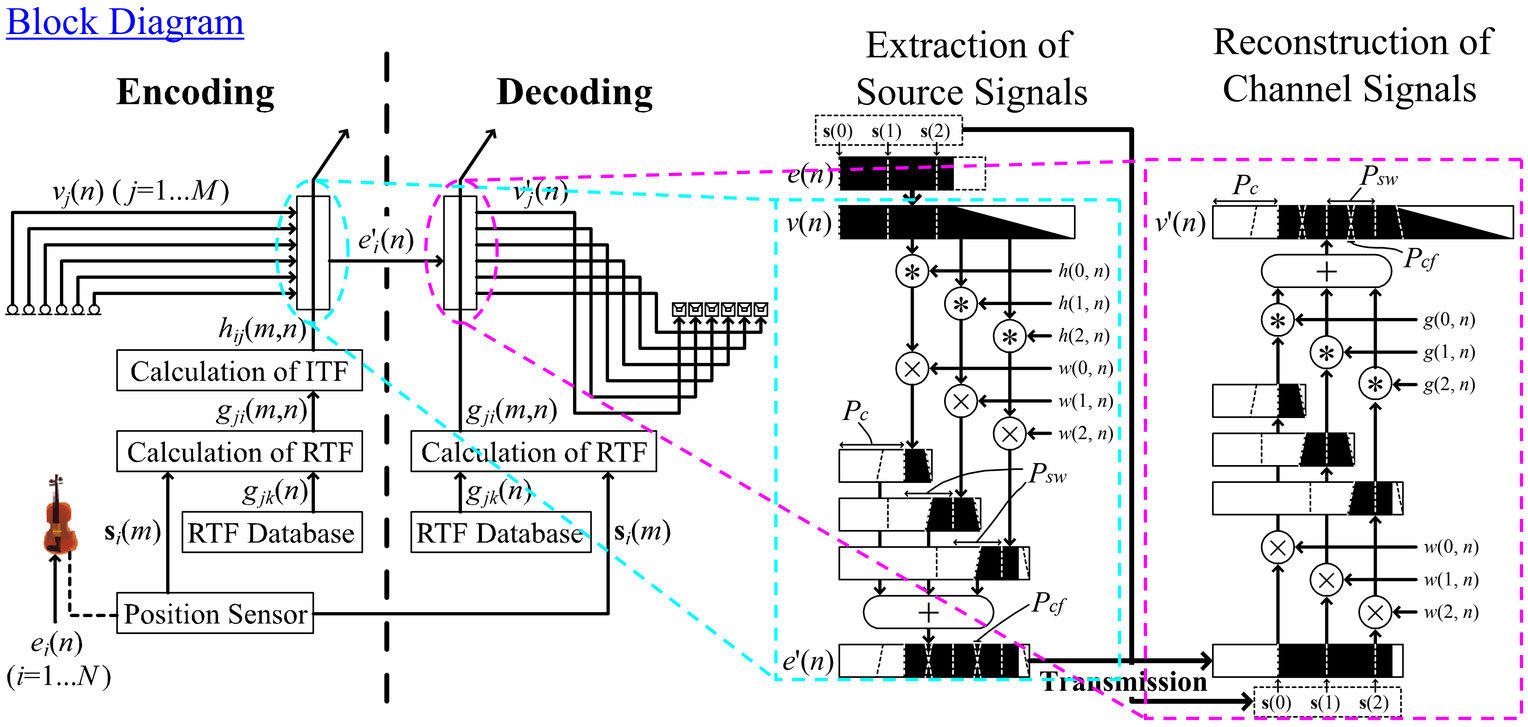
New spatial coding method for moving sound sources is proposed



The amount of data transmitted
…The number of sound sources

2. ALGORITHM

Block Diagram



- $e_i(n)$: ith source signal (i=1…N)
- $v_j(n)$: jth channel signal (j=1…M)
- $\mathbf{s}_i(m)$: Position information of ith source
- $g_{ji}(m,n)$: Room transfer function from ith source to jth channel signal in tine m
- $h_{ij}(m,n)$: Inverse transfer function from jth channel signal to ith source signal in time m
- m, n: Discrete time (Sampling frequency…F_p, F_s)

Extraction of Moving Sound Sources

$$e'_i(n) = \sum_m w(m,n) \sum_{j=1}^{M} h_{ij}(m,n) * v_j(n)$$

- $e'_i(n)$: Extracted ith source signal

Reconstruction of Channel Signals

$$v'_j(n) = \sum_m \sum_{i=1}^{N} g_{ji}(m,n) * [w(m,n) e'_i(n)]$$

- $v'_j(n)$: Reconstructed jth channel signal

Calculation of Inverse Transfer Function

$$\mathbf{H}(m,\omega) = \mathbf{G}^{+}(m,\omega)\mathbf{D}(\omega)$$

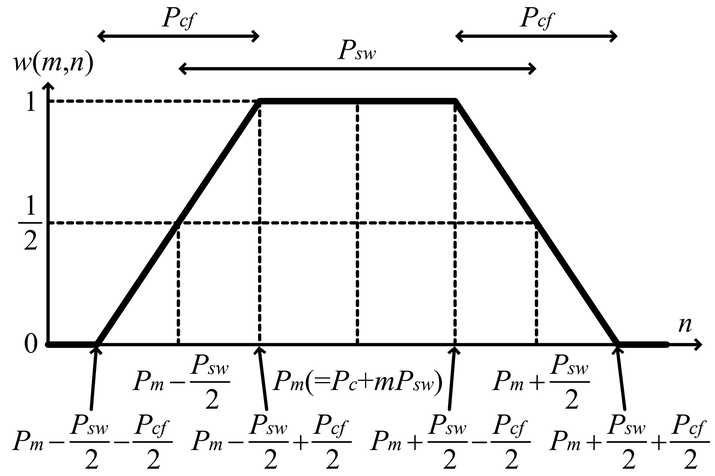
$$\mathbf{H}(m,\omega) = [H_{ij}(m,\omega)] = [FFT\{h_{ij}(m,\omega)\}]$$

$$\mathbf{G}(m,\omega) = [G_{ji}(m,\omega)] = [FFT\{g_{ji}(m,\omega)\}]$$

$$\mathbf{D}(\omega) = diag(e^{-j\omega T_c}, \ldots, e^{-j\omega T_c})$$

- $\mathbf{G}^{+}(m,\omega)$: Moore-Penrose pseudo inverse matrix of $\mathbf{G}(m,\omega)$
- $T_c(=P_c/F_s)$: Delay time in order to satisfy causality

Window Function

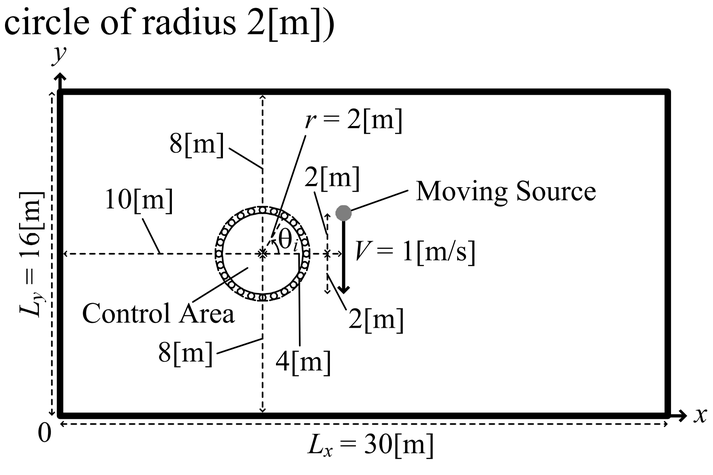


- $P_{sw}(=F_s/F_p)$: Switching sample of sound sources' position
- $P_m(=P_c+mP_{sw})$: Time of the sound sources' extraction
- $P_{cf}(=F_sT_{cf})$: Linear cross-fade sample

3. CODING EXPERIMENT

Original Sound Field

- 24 microphones (in the circle of radius 2[m])
- 1 moving sound source



Synthesis of Channel Signals

- Simulation by image method

Synthetic conditions of channel signals

Dry source	Speech	Flute
F_s (Sampling frequency)	48[kHz]	
Duration of sound source	4[second]	
Reflection coefficient	0.5	0.7
Maximum reflection order	6	10
Reverberation time	0.6[second]	1.0[second]
V (Velocity of sound source)	1[m/s]=3.6[km/h]	

Calculation of Room Transfer Function Database

- Calculation by image method
  - $\mathbf{s}_k$…Position vector of sound sources (k=1…481)
  - $\mathbf{r}_j$…Position vector of microphones (j=1…24)

$$\mathbf{s}_k = \begin{pmatrix} 14 \\ 10 - \frac{k-1}{120} \end{pmatrix}, \quad \mathbf{r}_j = \begin{pmatrix} 2\cos\left[\frac{\pi}{12}(j-12)\right] + 10 \\ 2\sin\left[\frac{\pi}{12}(j-12)\right] + 8 \end{pmatrix}$$

Calculation of Inverse Transfer Functions

Calculation conditions of inverse transfer functions

Reverberation time	0.6[second]	1.0[second]
FFT frame length [sample]	65536	131072
Coding delay time $T_c(=P_c=T_cF_s)$	20[ms](=960[sample])	
Inverse transfer function length [sample]	28800	48000

Convolution of Transfer Functions

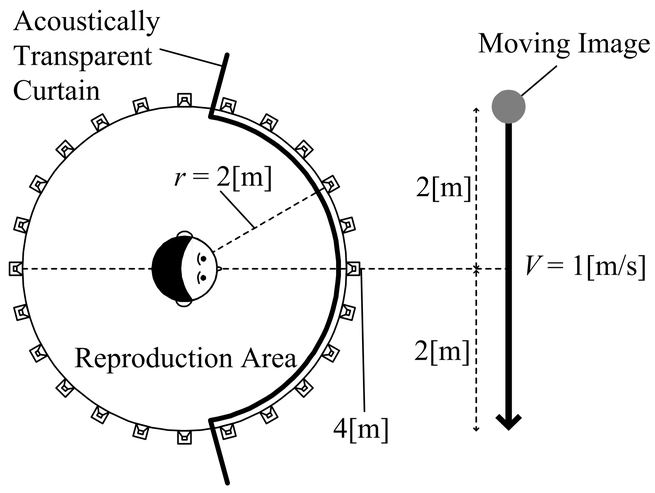
Conditions of window function

Sampling frequency of position information F_p (Switching sample of source position P_{sw})	30, 60, & 120 [Hz] (1600, 800, & 400[sample])
Linear cross-fade time T_{cf} (Linear cross-fade sample P_{cf})	1, & 4 [ms] (48, & 192 [sample])

4. SUBJECTIVE ASSESSMENT

Experimental Environment

- Room reverberation time…About 80ms
- Background noise level…25.0dB(A)
- Sound pressure level…About 70dB(A) (at the position of the subject)
- The affect of visual perception is avoided
  - The light in the room is dimmed
  - The loudspeakers are covered by an acoustically transparent curtain



Experimental Procedure

- Moving evaluation
  - "X" is the sound of reference movement
  - Either "A" or "B" is the same movement of sound as "X"
  - Grading the perceptual impairment of the sound movement
    - The stimulus of same movement…5.0
    - The other stimulus…1.0 to 4.9
- Total evaluation
  - "X" is the original sound
  - Either "A" or "B" is the same as "X"
  - Grading the perceptual impairment of the sound quality
    - The stimulus of same movement…5.0
    - The other stimulus…1.0 to 4.9

Scale table of impairment

Grade	Impairment	3.0	Slightly annoying
5.0	Imperceptible	2.0	Annoying
4.0	Perceptible, but not annoying	1.0	Very annoying

Experimental Design

- Subject…8 male students
- Protocol…Double-blind triple-stimulus with hidden reference
- Practice trials…12
  - 6 (Types of coding sound) × 2 (Either "A" or "B")
- Main trials…24
  - 6 (Types of coding sound) × 2 (Either "A" or "B") × 2 (Repetition)

Subjective Assessment

Moving Evaluation	Total Evaluation

Evaluation

Session 1	Session 2
Order...Randomized (Speech or Flute)	

Session

Practice (12 trials)	Main (24 trials)

Trial

X	Break	A	Break	B
Ref.	(0.5sec)	Ref./Test	(0.5sec)	Test/Ref.

Types of the coding sound

	F_p	T_{cf}
1	30Hz	1ms
2	30Hz	4ms
3	60Hz	1ms
4	60Hz	4ms
5	120Hz	1ms
6	120Hz	4ms

Selection of Subjects

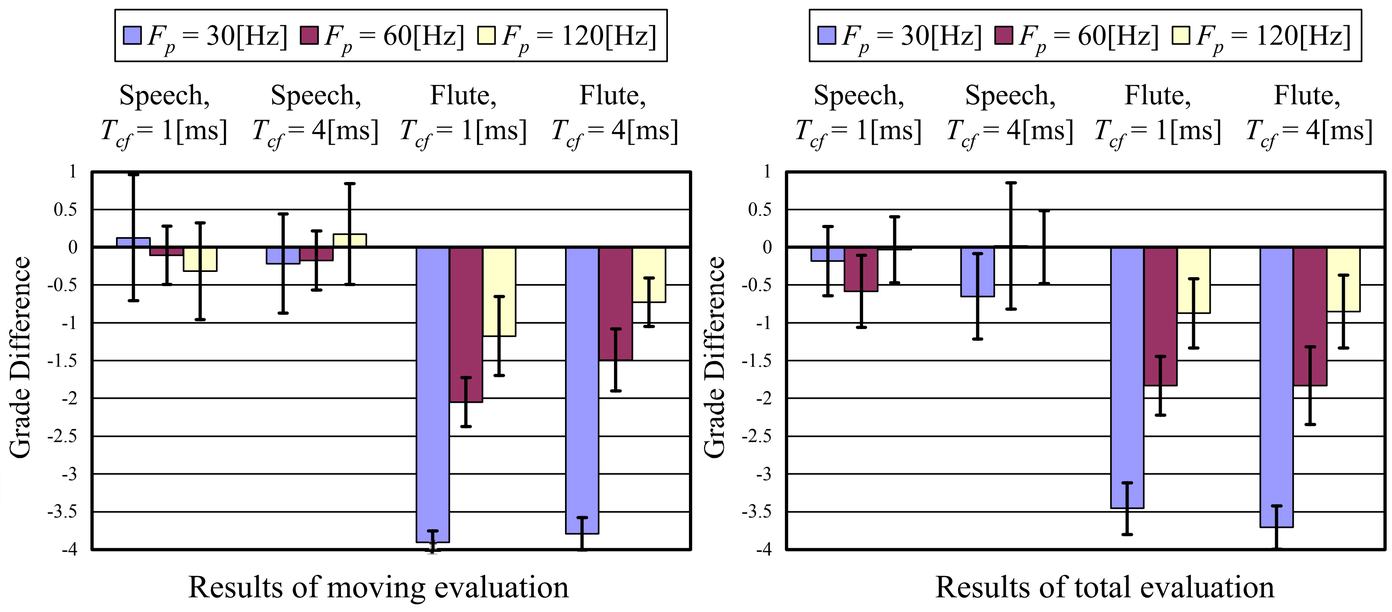
- The number of the correct response
  - Grading of the stimulus assigned the original sound as 5.0
- Further analysis data set
  - Top 3 subjects of each session (shown by color)

Discrimination results of each subject

Subject	Sample	Moving evaluation		Total evaluation	
		Speech	Flute	Speech	Flute
A	24	**15**	**23**	**15**	**22**
B	24	7	**23**	**16**	**24**
C	24	7	**24**	14	**24**
D	24	**14**	21	11	21
E	24	**14**	14	**15**	17
F	24	11	22	9	22
G	24	11	19	9	20
H	24	9	18	10	17

Experimental Result

- Grade difference
  - Grade of (the coding sound – the original sound)
- Flute (Case of the tele-ensemble system for music)
  - F_p=120[Hz]
    - The grade difference is about –1
  - The coding quality is preserved if F_p=120[Hz]
- Speech (Case of the tele-conference system)
  - F_p=30, 60, & 120[Hz]
    - The grade difference is almost 0
  - The coding quality is adequate even if F_p=30[Hz]



5. CONCLUSION

- Spatial coding method based on the extraction of moving sound sources was proposed
  - The amount of data to be transmitted
    - **The number of channel signals → The number of moving sound sources**
- A coding experiment with a reverberant sound field synthesized by an image method was performed
  - Reduction of the amount of data to be transmitted by the experiment
    - **24 channel signals → 1 moving sound source signal**
- The subjective assessment was performed to evaluate the performance of the proposed method
  - **The perceptual quality obtained with the proposed method was acceptable**
    when appropriate parameters for moving sound source were applied according to the type of sound sources
  - Parameters
    - Switching time of the position of sound sources
    - Cross-fade time to smooth the waveform of the extracted source signals
- Future works
  - Evaluation of recoding channel signals in a real environment
  - Interpolation method of the position of moving sound sources by the low-cost position detection system
  - Estimation method of room transfer functions from the shape of the room