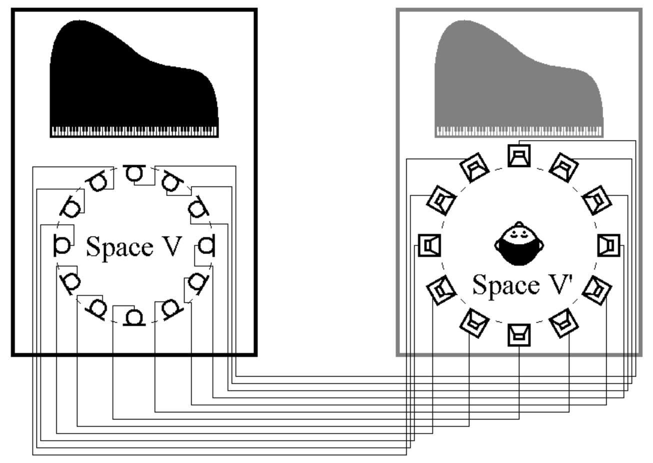
SPATIAL COMPRESSION OF MULTI-CHANNEL AUDIO SIGNALS USING INVERSE FILTERS

Toshiyuki KIMURA, *Kazuhiko KAKEHI* (Graduate School of Human Informatics, Nagoya University/CIAIR),
Kazuya TAKEDA and *Fumitada ITAKURA* (Graduate School of Engineering, Nagoya University/CIAIR)

1. INTRODUCTION

Sound Field Reproduction

- Sending the acoustic information from a space to the other space



Problem

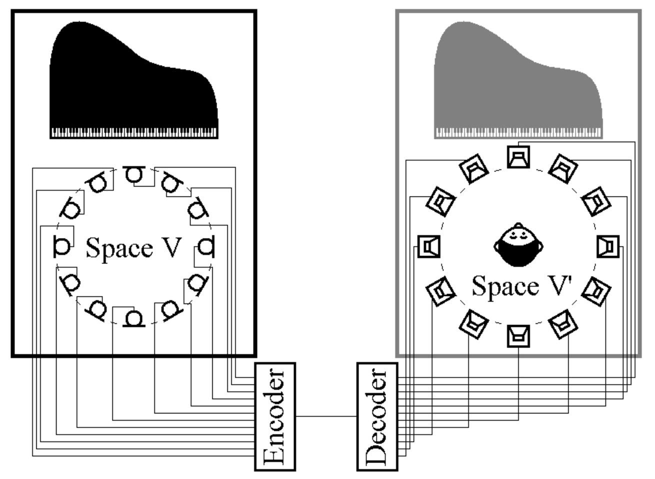
- It needs a great number of the channel signals
- The ordinary compression method is NOT enough to compress spatially (ex. AC-3 and MPEG2 AAC)

A New Approach of Spatial Compression

- The information of the sound sources is included in all channel signals

⇒ **Extract the source signals from the channel signals**

Transmission data reduction
The number of channels ⇒ The number of sound sources



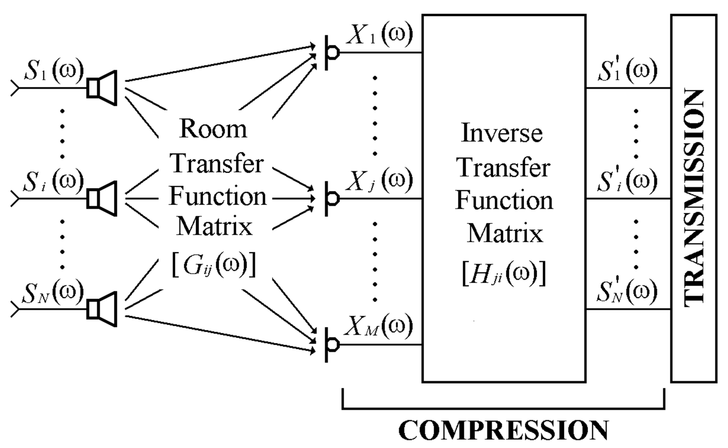
Applications

- Tele-conference system
- Tele-concert hall

2. COMPRESSION ALGORITHM

Compression (Extraction of the Sound Source Signals)

- Convolve the inverse transfer function matrix to the M channel signals



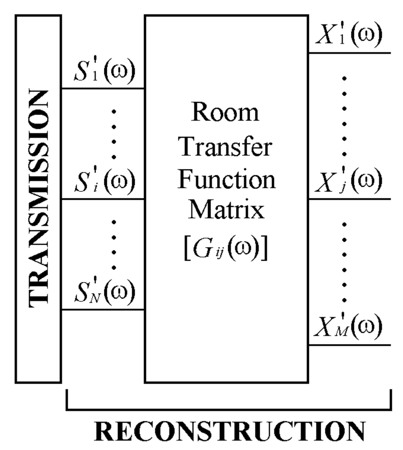
$S_i(\omega)$: ith source signal
$X_j(\omega)$: jth channel signal
$S'_i(\omega)$: ith extracted source signal
$G_{ij}(\omega)$: Room impulse response $[S_i(\omega) \rightarrow X_j(\omega)]$
$H_{ji}(\omega)$: Inverse transfer function $[X_j(\omega) \rightarrow S'_i(\omega)]$

$$\mathbf{GH} = \mathbf{D}$$
$$\Downarrow$$
$$\mathbf{H} = \mathbf{G}^{+}\mathbf{D}$$

$\mathbf{G}(=[G_{ij}(\omega)])$: Room impulse response matrix ($N\times M$)
$\mathbf{H}(=[H_{ji}(\omega)])$: Inverse transfer function matrix ($M\times N$)
$\mathbf{D}$: Unit matrix ($N\times N$)
Diagonal component : $D_i(\omega)(=S'_i(\omega)/S_i(\omega))$
$\mathbf{G}^{+}$: Moore-Penrose pseudo inverse matrix of $\mathbf{G}$

Reconstruction

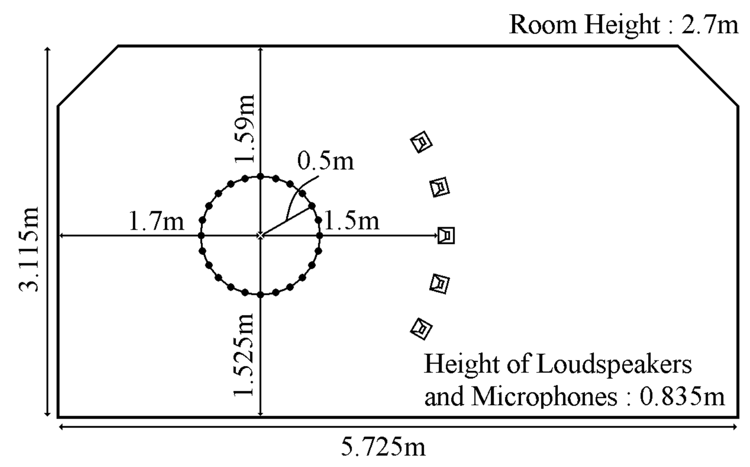
- Convolve the room transfer function matrix to the N extracted source signals



$S'_i(\omega)$: ith extracted source signal
$X'_j(\omega)$: jth reconstructed channel signal

3. EXPERIMENT

Measurement of Room Impulse Response



Measurement Conditions

Reverberation time	150ms	300ms
Room temperature	19.5ºC	19.2ºC
Noise level	20.0dB(A)	19.4dB(A)
Sound pressure level[*1]	90.0dB(A)	91.6dB(A)
Sampling frequency	48kHz	
Reference signal	TSP(65536 points)	
The number of repetitions	16	
FIR filter order	7200	14400

*1 One meter from loudspeaker

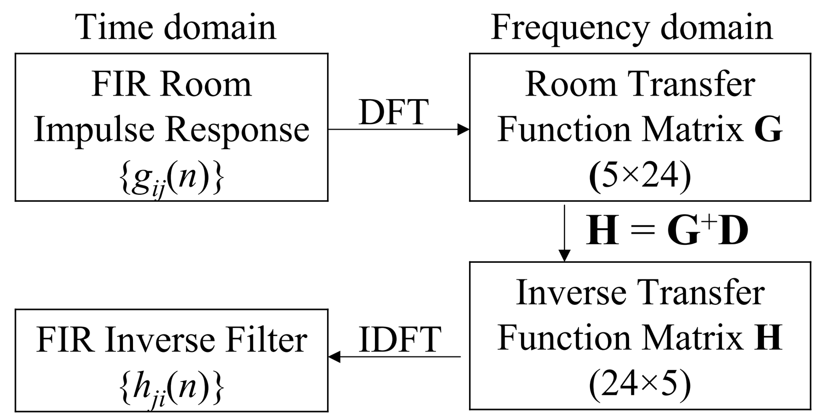
Synthesis of Channel Signals

Sound Source	Speech	Piano
The number of sources	1 (Azimuth angle = -15º, 0º, 15º)	
Duration	About 5sec	
Sampling Frequency	12kHz	
Bandwidth	50Hz-5kHz	

Reverberation time of channel signals

150ms → ×4 → 0.6sec
300ms → 1.2sec

Computation of Inverse Filter



Diagonal element of **D** : DFT of FIR band-pass filter
(Sampling frequency: 12kHz, Bandwidth: 50Hz-5kHz)

Conditions of Computation

Reverberation time	0.6sec	1.2sec
DFT points	16384	32768
Delay of BPF	10ms, 20ms, 40ms, 80ms, 160ms, 320ms	
Filter length	7200	14400

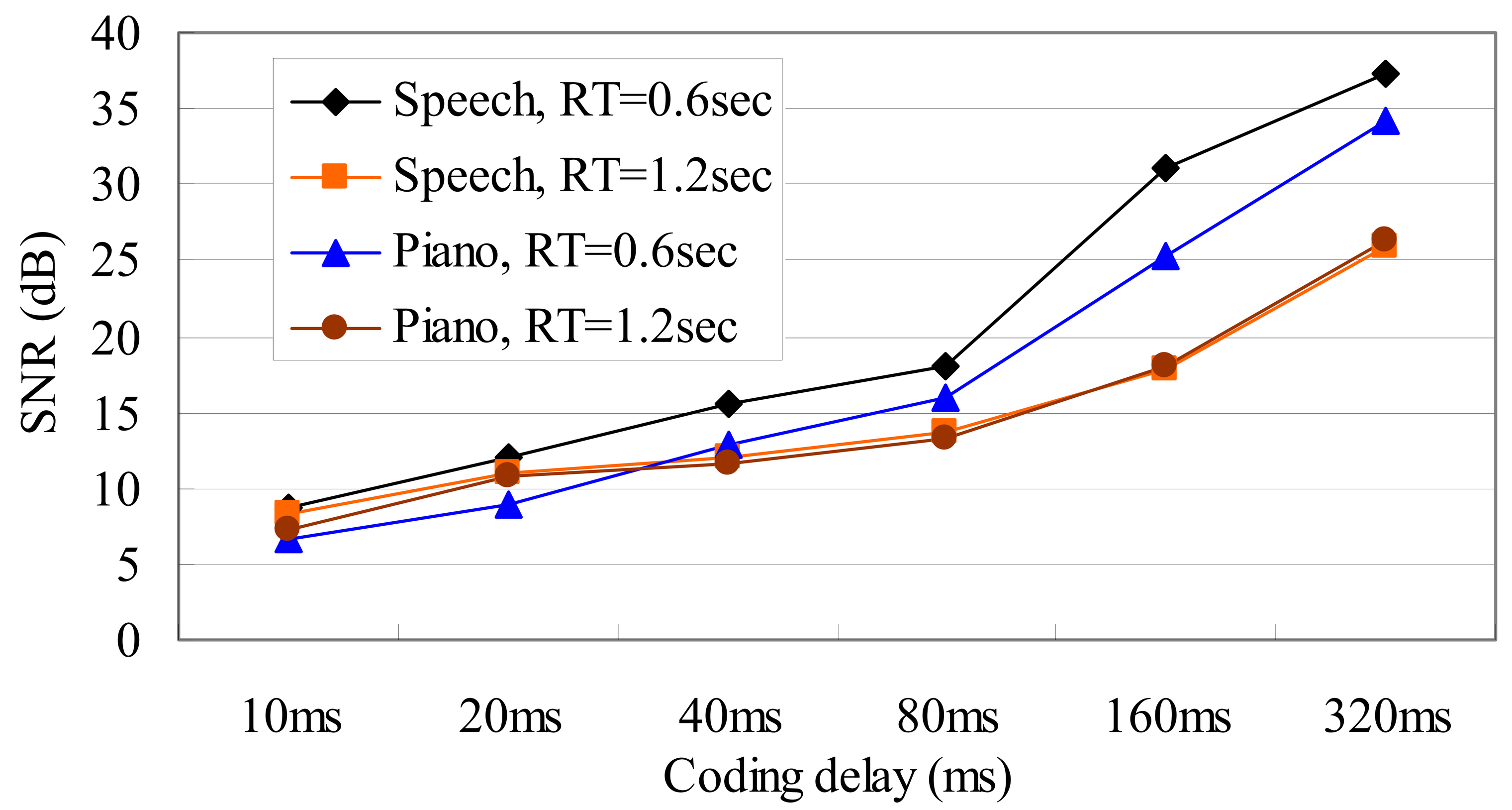
Delay of BPF : **Correspond to the coding delay**

4. EVALUATION OF EXTRACTION

Signal-to-Noise Ratio(SNR) of the Source Signals

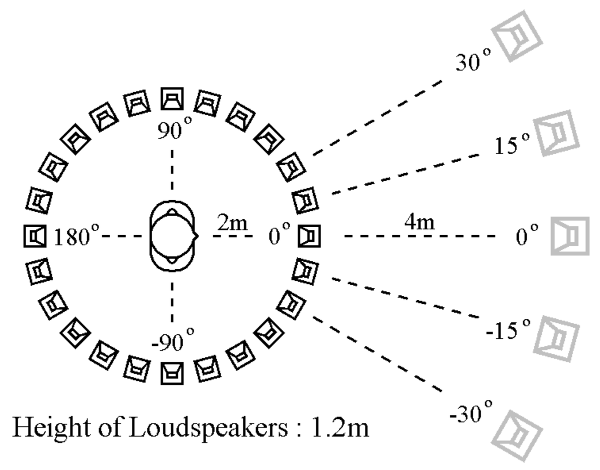
- The accuracy of extraction of the source signals

$$\mathrm{SNR(dB)} = 10\log_{10}\frac{\sum_i\sum_n\{S_i(n)\}^2}{\sum_i\sum_n\{S_i(n)-a_iS'_i(n)\}^2} \qquad a_i = \frac{\sum_n\{S_i(n)S'_i(n)\}}{\sum_n\{S_i(n)\}^2}$$



5. SUBJECTIVE ASSESSMENT

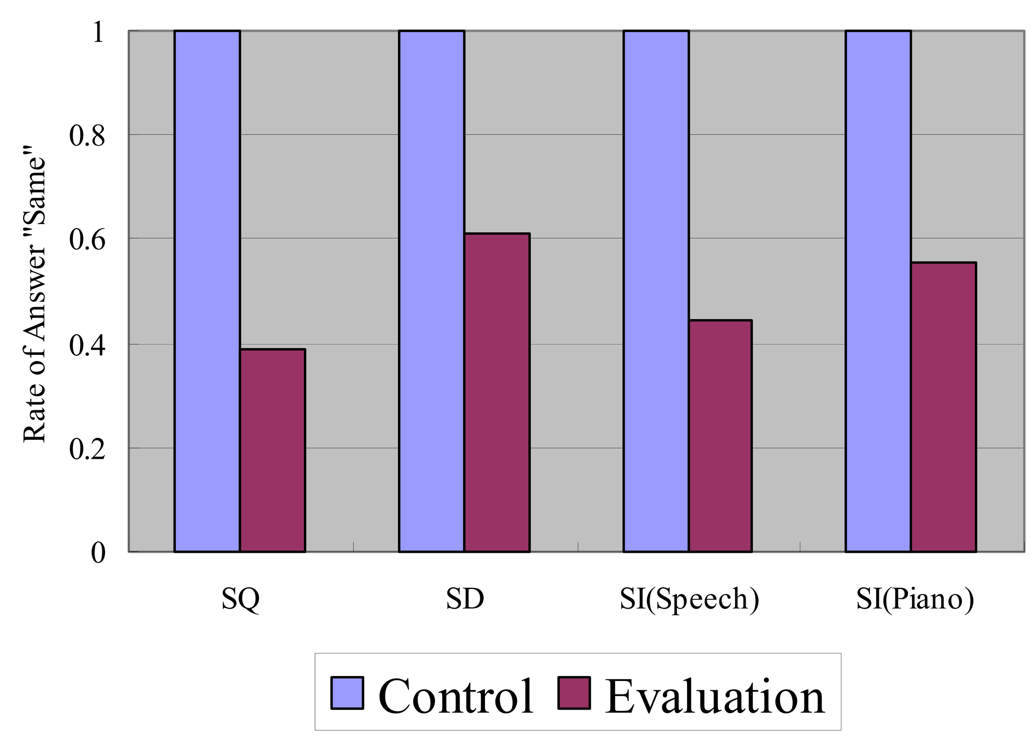
Experimental Environment



Experimental Conditions

Subjects	3 males (The person who pass the pre-test)
Sound pressure level	About 70dB(A) (at a center of circle)
Method	Pair test(Stimulus 1 vs Stimulus 2) Interval: 0.5sec
Answer	"Different" or "Same"
Note	A head is fixed

Result



Sessions

Sound Quality(SQ)

Source : Piano, RT=1.2sec, Azimuth angle=0º

Condition	Stimulus 1	Stimulus 2	Trials
Control	Original	Original	6
Evaluation	Original	Coding (Delay=10ms)	6

Spatial Impression(SI)

Source : Speech & Piano, Azimuth Angle=0º

Condition	Stimulus 1	Stimulus 2	Trials
Control	Original (0.6sec)	Original (0.6sec)	3
	Original (1.2sec)	Original (1.2sec)	3
Evaluation	Original (0.6sec)	Coding(0.6sec, Delay=10ms)	3
	Original (1.2sec)	Coding(1.2sec, Delay=10ms)	3

Sound Direction(SD)

Source : Speech, RT=0.6sec

Condition	Stimulus 1	Stimulus 2	Trials
Control	Original (-15º)	Original (-15º)	2
	Original (0º)	Original (0º)	2
	Original (15º)	Original (15º)	2
Evaluation	Original (-15º)	Coding(-15º , Delay=10ms)	2
	Original (0º)	Coding(0º , Delay=10ms)	2
	Original (15º)	Coding(15º , Delay=10ms)	2

6. CONCLUSION

A new spatial compression method is proposed

This method performs the extraction of the sound source signals using the inverse filters

The Extraction of the source signals is evaluated

⇒ **The source signals are extracted accurately in the condition of the longer coding delay**

The influence of the sound field perception is evaluated by the subjective assessment

⇒ **The sound field perception is affected by the coding in the condition of coding delay 10ms**

⇒ **It needs to evaluate in the condition of the longer coding delay**