

移動音源に対するマルチチャンネル 音場空間符号化の検討

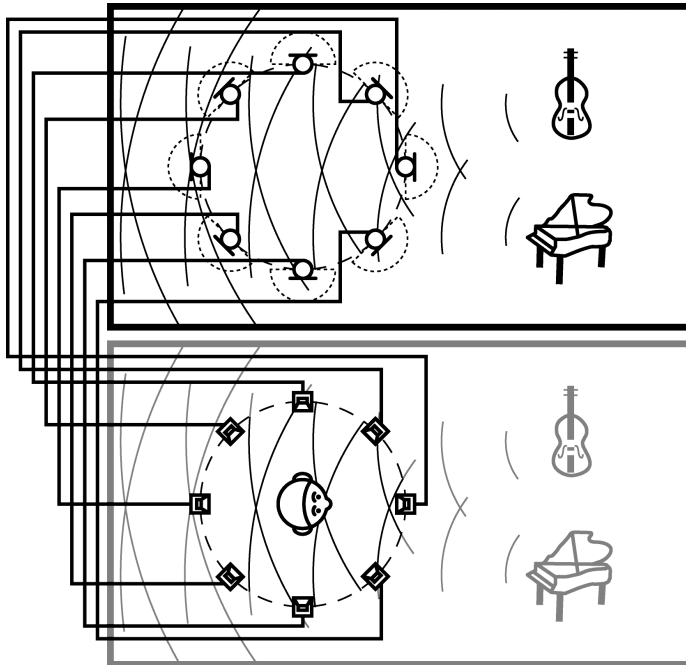
木村敏幸, 笈一彦 (名大・人情/CIAIR),

武田一哉 (名大・工/CIAIR),

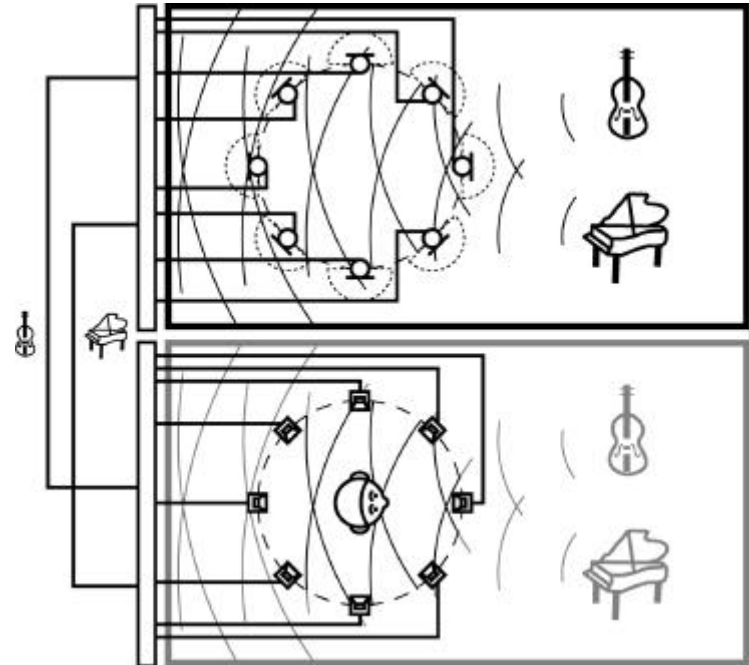
板倉文忠 (名大・情報メディア/CIAIR)

研究背景

- ホイヘンスの原理による波面合成
 - 多くのチャンネル信号の伝送が必要



- 空間符号化による伝送量の削減
 - チャンネル信号から音源信号を抽出



研究内容

- 今までの成果
 - 実環境でチャンネル信号を作成
 - 符号化
 - 24個のチャンネル信号 5個の静止音源
 - 主観評価
 - 符号化歪み 知覚できるが、気にならない
- 本報告の内容
 - 移動音源について検討
 - 提案手法の有効性をさらに検証

静止音源のチャンネル信号

- 音源に室内インパルス応答を畳み込み

$$\begin{aligned}x_j(n) &= \sum_{i=1}^N g_{ji}(n) * s_i(n) \quad (j = 1 \dots M) \\ &= \sum_{i=1}^N \sum_{l=0}^{L_g-1} g_{ji}(l) s_i(n-l)\end{aligned}$$

$s_i(n)$: i 番目の音源信号

$g_{ji}(n)$: 室内インパルス応答 [$s_i(n)$? $x_j(n)$]

$x_j(n)$: j 番目のチャンネル信号

N : 音源の数

M : チャンネル数

L_g : 室内インパルス応答の次数

移動音源のチャンネル信号

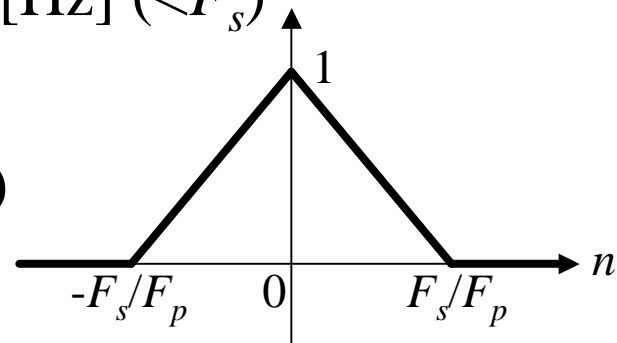
- p ごとに室内インパルス応答が異なる

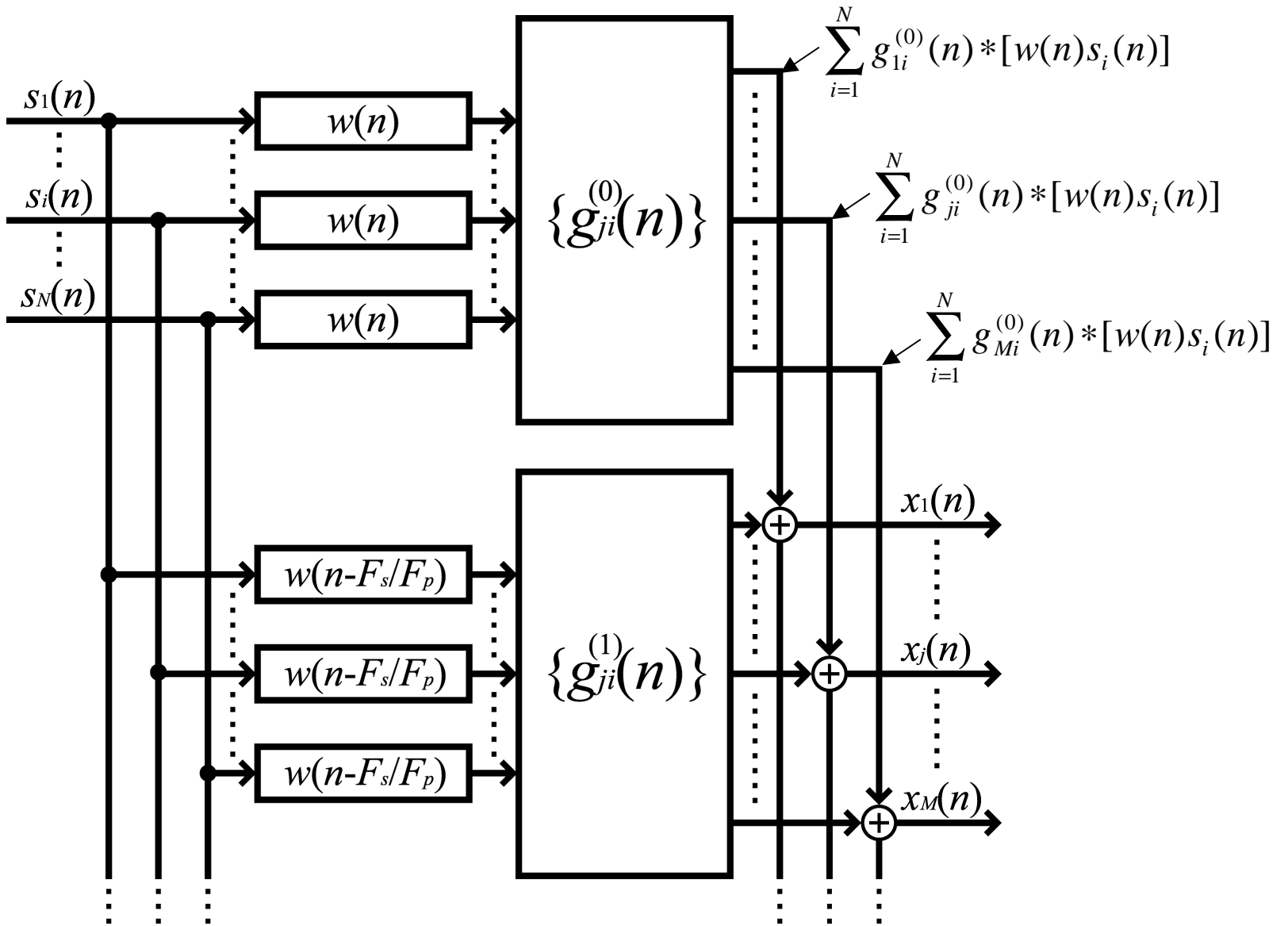
$$\begin{aligned}x_j(n) &= \sum_{p=-\infty}^{\text{floor}(nF_p / F_s)} \sum_{i=1}^N g_{ji}^{(p)}(n) * [w(n - pF_s / F_p) s_i(n)] \\ &= \sum_{p=-\infty}^{\text{floor}(nF_p / F_s)} \sum_{i=1}^N \sum_{l=0}^{L_g-1} g_{ji}^{(p)}(l) w(n - pF_s / F_p - l) s_i(n - l)\end{aligned}$$

F_s : 音声のサンプリング周波数[Hz]

F_p : 音源位置のサンプリング周波数[Hz] ($< F_s$)

$$w(n) = \begin{cases} 1 + nF_p / F_s & (-F_s / F_p \leq n \leq -1) \\ 1 - nF_p / F_s & (0 \leq n \leq F_s / F_p - 1) \\ 0 & (\text{otherwise}) \end{cases}$$





移動音源信号の抽出

- p ごとに逆フィルタを畳み込み

$$\begin{aligned}
 s'_i(n) &= \sum_{p=-\infty}^{\text{floor}(nF_p/F_s)} w(n - \tau_0 - pF_s/F_p) \sum_{j=1}^M h_{ij}^{(p)}(n) * x_j(n) \\
 &= \sum_{p=-\infty}^{\text{floor}(nF_p/F_s)} w(n - \tau_0 - pF_s/F_p) \sum_{j=1}^M \sum_{m=0}^{L_h-1} h_{ij}^{(p)}(m) x_j(n-m)
 \end{aligned}$$

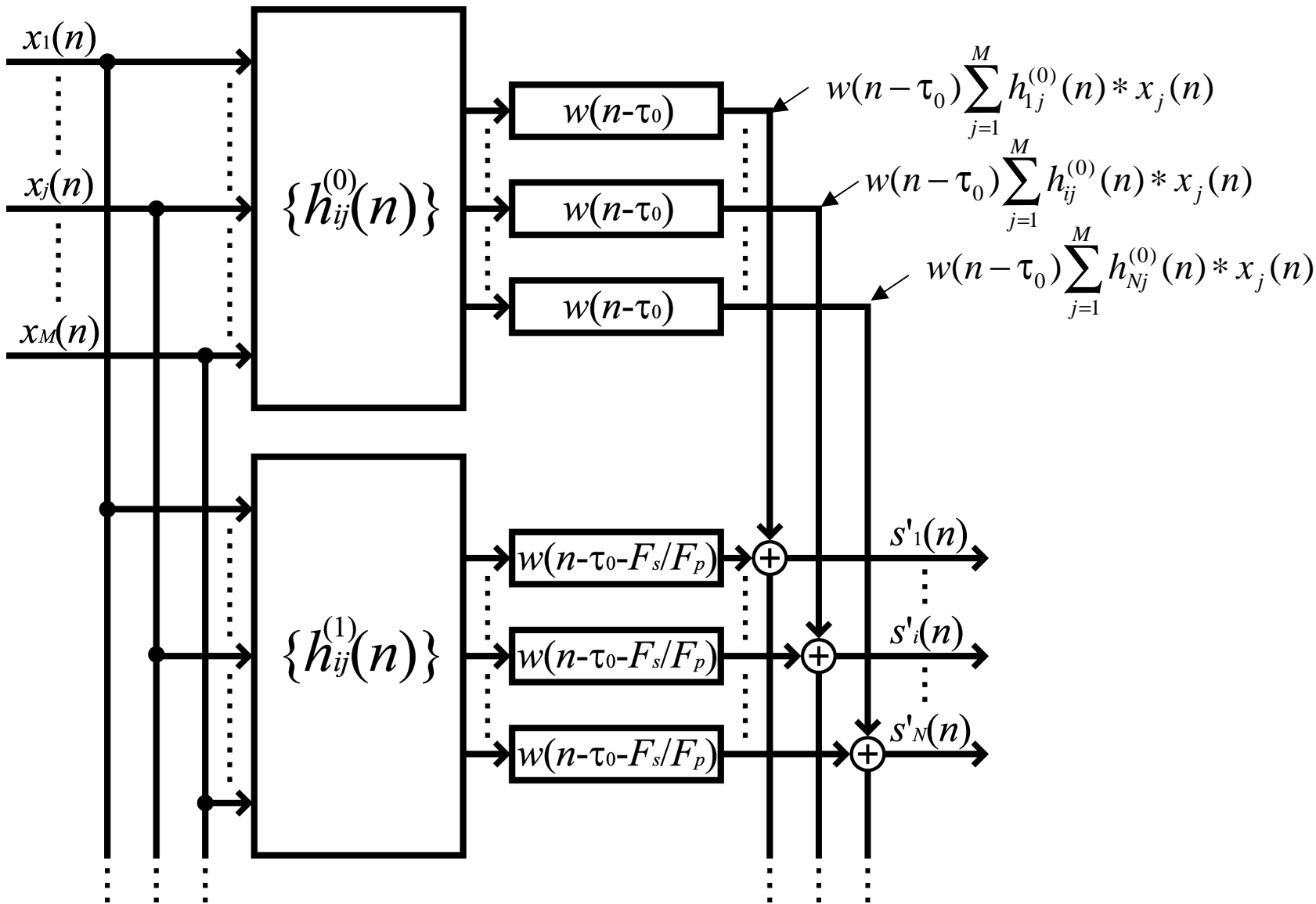
$$\begin{pmatrix} G_{11}^{(p)}(\omega) & \cdots & G_{M1}^{(p)}(\omega) \\ \vdots & \ddots & \vdots \\ G_{1N}^{(p)}(\omega) & \cdots & G_{MN}^{(p)}(\omega) \end{pmatrix} \begin{pmatrix} H_{11}^{(p)}(\omega) & \cdots & H_{N1}^{(p)}(\omega) \\ \vdots & \ddots & \vdots \\ H_{1M}^{(p)}(\omega) & \cdots & H_{NM}^{(p)}(\omega) \end{pmatrix} = \begin{pmatrix} e^{-j\omega\tau_0} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{-j\omega\tau_0} \end{pmatrix}$$

$$\mathbf{G}^{(p)}(\omega)\mathbf{H}^{(p)}(\omega) = \mathbf{D}(\omega) \quad \mathbf{H}^{(p)}(\omega) = [\mathbf{G}^{(p)}(\omega)]^+ \mathbf{D}(\omega)$$

L_h : 逆フィルタの次数

$[]^+$: Moore-Penrose 擬似逆行列

t_0 : 符号化遅延時間[points]



チャンネル信号の復元

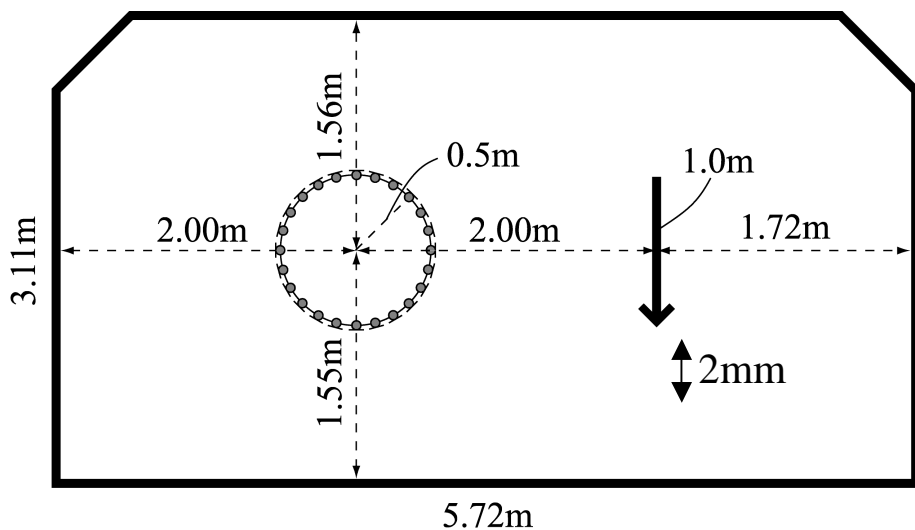
- p ごとに室内インパルス応答を畳み込み

$$\begin{aligned}x'_j(n) &= \sum_{p=-\infty}^{\text{floor}(nF_p/F_s)} \sum_{i=1}^N g_{ji}^{(p)}(n) * [w(n - \tau_0 - pF_s/F_p) s'_i(n)] \\ &= \sum_{p=-\infty}^{\text{floor}(nF_p/F_s)} \sum_{i=1}^N \sum_{l=0}^{L_g-1} g_{ji}^{(p)}(l) w(n - \tau_0 - pF_s/F_p - l) s'_i(n-l)\end{aligned}$$

- 伝送チャンネル数
 - M個 N個
 - $N \ll M$ の時には特に有利

室内インパルス応答の測定

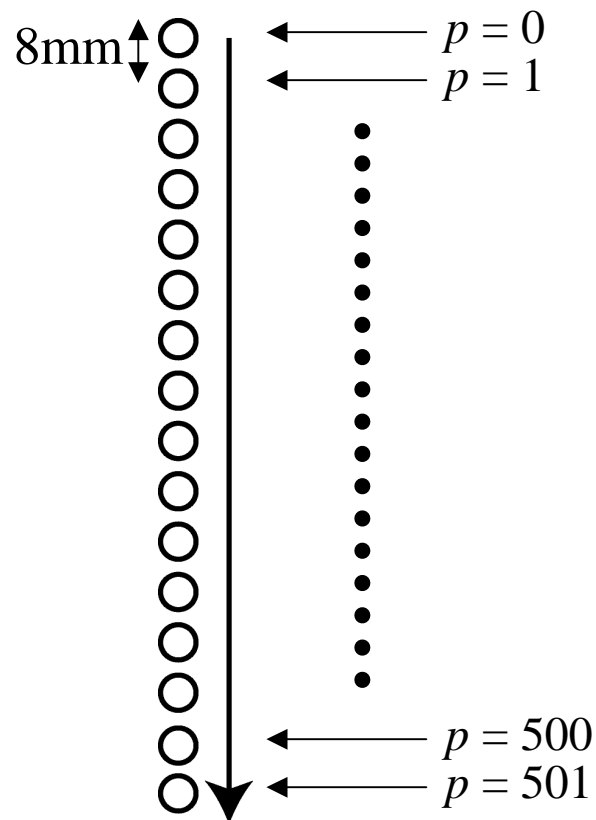
- 1.0mの範囲を2mm間隔で計501回測定



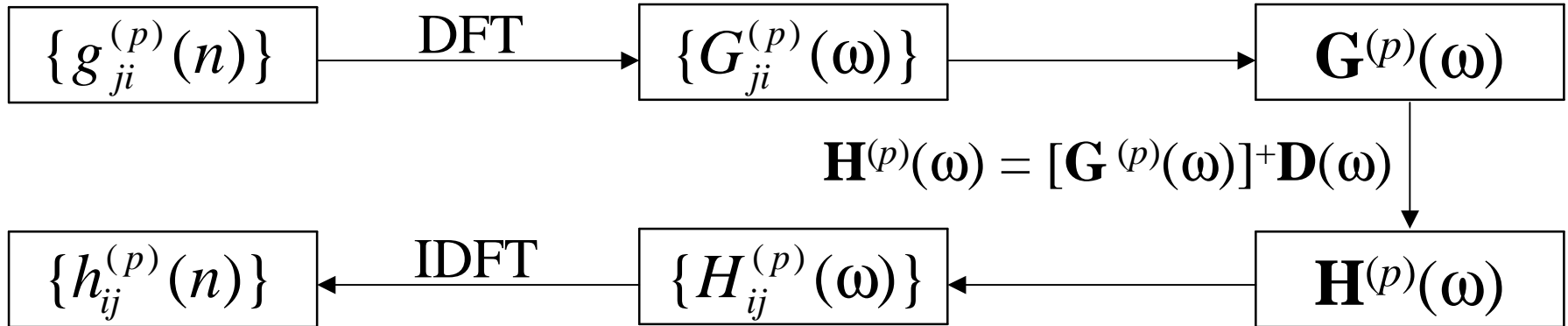
残響時間	300ms
室温	15.9°C
暗騒音レベル	19.3dB(A)
音圧レベル	85.2dB(A)
サンプリング周波数	48kHz
TSP信号の長さ	16384点
同期加算	8
フィルタ次数	14400

チャンネル信号の作成

音源信号	sin波	雑音	音声	楽音
音源数	1			
長さ	4sec			
F_s [Hz]	12kHz			
帯域[Hz]	1000	50-5000		
残響時間	1.2sec			
F_p [Hz]	125Hz			
移動速度	3.6km/h=1m/s			



逆フィルタの算出



$\mathbf{D}(\omega)$ の対角成分($=e^{-j\omega t_0}$): 線形位相のバンドパスフィルタ

因果律を満たす逆フィルタを作成

残響時間	1.2sec
DFT points	28800
標本化周波数	12kHz
BPFの帯域幅	50-5000Hz
BPFの遅延時間	10ms, 20ms, 40ms, 80ms, 160ms, 320ms, 640ms
フィルタ次数($=L_h$)	14400

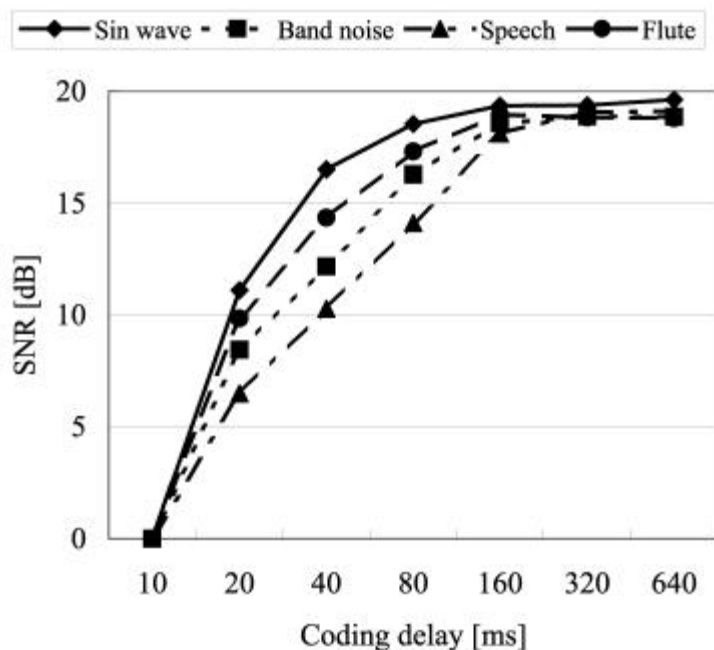
実験結果

• SNRで評価

t_0 : 符号化遅延 [points]

$$\text{SNR}[\text{dB}] = 10 \log_{10} \frac{\sum_n \{s(n)\}^2}{\sum_n \{s(n) - as'(n + \tau_0)\}^2}$$

$$a = \frac{\sum_n s(n)s'(n + \tau_0)}{\sum_n \{s'(n + \tau_0)\}^2}$$











• 考察

– 符号化遅延が20ms以上なら,チャンネル信号の聴感上の違いはほとんどない

– 波形では類似していいながら,主観的には評価が良い可能性がある

音声

source	channel	extracted source	reconstructed channel
		10ms 	10ms 
		20ms 	20ms 
		40ms 	40ms 

まとめ

- 移動音源の抽出に基づく空間符号化方式を提案
- SNRにより,抽出精度を検討
 - 符号化遅延が長ければ,波形的な歪みは小さくなる
- 聴感上で実験結果を考察
 - 主観評価では符号化遅延が短くても良い評価が得られることが示唆される

今後の予定

- 主観評価実験
 - 原信号...音源が滑らかに動いているか
 - 符号化による影響
 - 音場 , 音質 , 音源の動き
- 室内インパルス応答の測定の簡略化
 - いくつかの測定点から他の点を推定