

# 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, 箕一彦 (名大・人間情報/CIAIR),  
武田一哉 (名大・工/CIAIR), 板倉文忠 (名大・情報メディア/CIAIR)

## 1 はじめに

スピーカーアレーを用いた多点制御型音場再現技術は非常に多くのチャンネルを要するので、伝送するチャンネル数を空間符号化によって削減する必要がある。我々は今までに音源抽出に基づく空間符号化方式を提案し、静止音源に対して実験及び性能評価を行ってきた [1, 2, 3]。本報告では移動音源を対象とした空間符号化手法について研究する。符号化原理については従来の提案法を元に、移動音源の抽出に基づいた手法を検討する。

## 2 移動音源の抽出原理

先の研究 [1, 2, 3] では  $M$  個のチャンネル信号  $x_j(n)$  ( $j = 1 \dots M$ ) は  $N$  個の音源信号  $s_i(n)$  ( $i = 1 \dots N$ ) に室内インパルス応答  $g_{ji}(n)$  を畳み込んだものとして表していた。\* は畳み込み演算を表す。

$$x_j(n) = \sum_{i=1}^N g_{ji}(n) * s_i(n) \quad (1)$$

移動音源の場合は室内インパルス応答が音源位置によって変化していると考えられる。ゆえに、音源位置を  $F_p$  [Hz] でサンプリングしている場合、(1) 式は以下のように変形される。

$$x_j(n) = \sum_p \sum_{i=1}^N g_{ji}^{(p)}(n) * [w(n - \frac{F_s}{F_p} p) s_i(n)] \quad (2)$$

ここで、 $g_{ji}^{(p)}(n)$  は時刻  $p (= \text{floor}(\frac{F_p}{F_s} n))$  における室内インパルス応答、 $F_s$  は音声のサンプリング周波数 [Hz]、 $w(n)$  は (3) 式のように示される窓関数である。

$$w(n) = \begin{cases} 1 + \frac{F_p}{F_s} n & (-\frac{F_s}{F_p} \leq n \leq -1) \\ 1 - \frac{F_p}{F_s} n & (0 \leq n \leq \frac{F_s}{F_p} - 1) \\ 0 & (\text{otherwise}) \end{cases} \quad (3)$$

チャンネル信号  $x_j(n)$  に逆フィルタ  $h_{ij}^{(p)}(n)$  を畳み込み、 $p$  ごとに窓を掛けることによって、音源信号  $s'_i(n)$  が抽出される。

$$s'_i(n) = \sum_p w(n - \tau_0 - \frac{F_s}{F_p} p) \sum_{j=1}^M h_{ij}^{(p)}(n) * x_j(n) \quad (4)$$

逆フィルタは  $z$  領域において、室内インパルス応答の Moore-Penrose 擬似逆行列によって算出される。

\*Spatial coding of multi-channel audio signals in moving sound sources by T. Kimura, K. Kakehi, K. Takeda and F. Itakura (Nagoya University)

$$\mathbf{H}^{(p)}(\omega) = [\mathbf{G}^{(p)}(\omega)]^+ \mathbf{D}(\omega) \quad (5)$$

但し、

$$\mathbf{H}^{(p)}(\omega) = \begin{pmatrix} H_{11}^{(p)}(\omega) & \dots & H_{N1}^{(p)}(\omega) \\ \vdots & \ddots & \vdots \\ H_{1M}^{(p)}(\omega) & \dots & H_{NM}^{(p)}(\omega) \end{pmatrix} \quad (M \times N)$$

$$\mathbf{G}^{(p)}(\omega) = \begin{pmatrix} G_{11}^{(p)}(\omega) & \dots & G_{M1}^{(p)}(\omega) \\ \vdots & \ddots & \vdots \\ G_{1N}^{(p)}(\omega) & \dots & G_{MN}^{(p)}(\omega) \end{pmatrix} \quad (N \times M)$$

$$\mathbf{D}(\omega) = \begin{pmatrix} e^{-j\omega\tau_0} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & e^{-j\omega\tau_0} \end{pmatrix} \quad (N \times N)$$

で、 $[\ ]^+$  は Moore-Penrose 擬似逆行列、 $\tau_0$  は符号化遅延 [points] を示す。

一方、抽出した音源信号  $s'_i(n)$  に室内インパルス応答  $g_{ji}^{(p)}(n)$  を畳み込むと、チャンネル信号  $x'_j(n)$  が復元される。

$$x'_j(n) = \sum_p \sum_{i=1}^N g_{ji}^{(p)}(n) * [w(n - \tau_0 - \frac{F_s}{F_p} p) s'_i(n)] \quad (6)$$

抽出した音源信号を伝送することによって、伝送チャンネル数が  $M$  個から  $N$  個に削減される。これは、 $N \ll M$  の時には特に有利である。

## 3 符号化実験

### 3.1 室内インパルス応答の測定

名古屋大学工学部 7 号館 410 号室の可変残響室において室内インパルス応答の測定を行った。部屋の配置図を図 1 に示す。マイクロホンアレーとスピーカーの高さは 1.3m である。図 1 の太い矢印で示した範

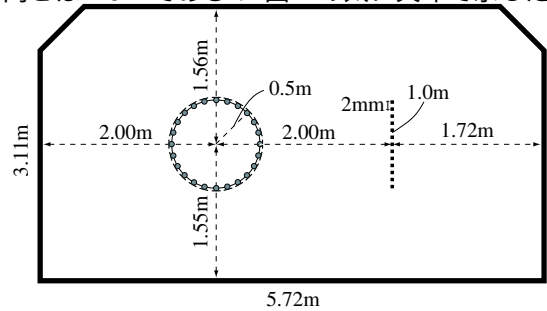


図 1: マイクロホンアレーとスピーカーの配置図

囲を音源の移動区間とした。移動距離 1.0m を 2mm ごとに区切った計 501 点の音源位置において、24 個のマイクロホンまでの室内インパルス応答の測定を実

行した．測定条件を表 1 に示す．音圧レベルはスピーカーから 1m での値である．

表 1: 測定条件

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

### 3.2 移動音源の作成

(2) 式に基づいて移動音源を模擬するチャンネル信号  $x_j(n)$  を作成した．作成条件を表 2 に示す．サン

表 2: 作成条件

$s_i(n)$	sin 波	雑音	音声	楽音
音源数 (= $N$ )	1			
音源の長さ	4sec			
$F_s$ [Hz]	12kHz			
帯域	1000Hz	50-5000Hz		
残響時間	1.2sec			
$F_p$ [Hz]	125Hz			
移動速度	3.6km/h			

プリング周波数が測定時の 1/4 なので，残響時間は測定時の 4 倍になる．上の移動条件では  $p$  が 1 増加すると，音源位置が 8mm 離れた隣の測定位置に移ることになる．

### 3.3 逆フィルタの算出

逆フィルタの算出は測定した音源位置ごとに計 501 回繰り返した．算出手順を以下に述べる．まず， $g_{ji}^{(p)}(n)$  を DFT 変換し， $G^{(p)}(\omega)$  を求めた．次に，(5) 式に従って  $H^{(p)}(\omega)$  を導出した．このとき， $e^{-j\omega\tau_0}$  には FIR バンドパスフィルタを DFT 変換したものをを用いた．最後に IDFT 変換をして  $h_{ij}^{(p)}(n)$  を算出した．算出条件を表 3 に示す．BPF の遅延時間が符号化によ

表 3: 算出条件

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	14400

る遅延時間に相当する．

### 3.4 実験結果

$s'_i(n)$  の抽出がどれだけ正確かを (7) 式に示す SN 比 [dB] によって評価する．

$$\text{SNR} = 10 \log_{10} \frac{\sum_i \sum_n \{s_i(n)\}^2}{\sum_i \sum_n \{s_i(n) - a_i s'_i(n + \tau_0)\}^2} \quad (7)$$

但し， $a_i$  は  $s_i(n)$  と  $s'_i(n)$  のゲインを等しくするための係数である．

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

ドライソースと符号化遅延時間ごとに SN 比を算出した結果を，図 2 に示す．音源に関わらず，符号化遅

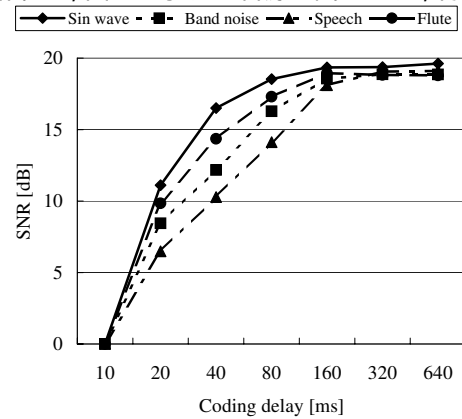


図 2: 音源及び符号化時間ごとの SN 比

延時間が 10ms のとき信号は全く抽出されておらず，長くなればなるほど抽出精度が向上している．一方，全体的に抽出精度が SNR20dB 程度で飽和している傾向がある．これは他の音源位置からの残響音が抽出の妨害をしているためであると考えられる．

## 4 まとめ

移動音源の抽出に基づく空間符号化方式を提案し，実験を行った．SN 比により波形歪みを評価したところ，符号化遅延が長ければ，ある程度の抽出精度を達成できることが分かった．

今後の課題として主観評価実験を行うことによって知覚的な歪みを評価する必要がある．また，提案手法では非常に多くの室内インパルス応答の測定を要するので，実用化に向けて測定の手間を省くための処理法を検討する必要がある．

### 参考文献

- [1] 木村， 寛， 武田， 板倉， ”逆フィルタを用いた音場空間符号化方式の主観評価”，音講論， 563-564， 2002.09
- [2] 木村， 寛， 武田， 板倉， ”逆フィルタを用いた空間音響圧縮技術の拡がり感”，音講論， 595-596， 2002.03
- [3] 木村， 寛， 武田， 板倉， ”空間情報を利用したマルチチャンネル音響信号の符号化と音源定位”，音講論， 711-712， 2001.10