

# 移動音源がある音場の空間符号化 —音源抽出による伝送量の削減—

木村 敏幸<sup>\*1</sup> 笥 一彦<sup>\*2</sup> 武田 一哉<sup>\*1</sup> 板倉 文忠<sup>\*3</sup>

Spatial Coding of the Sound Field by Moving Sound Sources  
- Bit Reduction Based on the Extraction of Sound Sources -

Toshiyuki Kimura<sup>\*1</sup> Kazuhiko Kakehi<sup>\*2</sup> Kazuya Takeda<sup>\*1</sup> Fumitada Itakura<sup>\*3</sup>

**Abstract** – The sound field reproduction technique is important to realize the more realistic communication system. However, a transmission data should be reduced for the remote system based on wave field synthesis since the technique usually needs a great number of channel signals. The spatial coding method for moving sound sources is proposed in this paper. In encoding side, the moving source signals are extracted by convolving channel signals to the inverse transfer functions, which are calculated from room transfer functions due to the location of sound sources. In decoding side, channel signals are reconstructed by convolving extracted source signals to the room transfer functions. As a result, the data to be transmitted is reduced from the number of channels to that of sound sources. To evaluate the coding method, a simulated reverberant sound field was synthesized by using an image method. The effect of the coding method on the perceptual quality was evaluated by subjective assessment.

**Keywords** : wave field synthesis, sound field reproduction, moving sound source, sound source extraction, image method

## 1 はじめに

近年、音場再現による聴覚バーチャリアリティ技術が盛んに研究されている。この技術が実用化されると、遠隔通信会議や遠隔アンサンブルのように離れた場所にいる人々があたかも同じ室内で会議を行ったり、一緒に演奏したりしているように思わせることが可能になる。これらのコミュニケーションシステムはテレビ電話のような従来のシステムとは比べ物にならないほどの臨場感を与えることができるため、遠隔コミュニケーションがより一層活発になっていくことが期待される。

波面合成法 [1, 2, 3, 4, 5, 6] は原音場内に設定した制御領域の境界面上にマイクロホンアレーを配置してチャンネル信号を収録し、制御領域と同じ形状をした再現領域の境界面上にスピーカアレーを配置してチャンネル信号を再生すると、Huygens の原理によって制御領域内の波面が再現領域内に再現される方法である。この方法は無響室などの音響的設備を必要とするが、バイノーラル方式 [7, 8] やトランスオーラル方式 [9, 10]

とは異なり、受聴者にとっての制約がほとんどないため、複数の受聴者が再現領域内のどこに居ても構わないし、頭部も自由に動かすことができる。

波面合成法による音場再現を遠隔地で実現するためにはチャンネル信号を伝送する必要がある。しかし、音場を再現するために必要なチャンネル信号の数は一般的に非常に多いため、映像と組み合わせた通信システムを構築する際には通信容量が足りなくなる可能性がある。このような問題を解決する一つの方法として音源のすぐそばにマイクロホンを配置し、収録した音源信号を伝送した上で室内伝達関数を畳み込んでチャンネル信号を復元する方法が考えられる。しかし、実際にはマイクロホンが映像に映り込んでしまうなどの理由でマイクロホンを音源のすぐそばに配置できないことが多い。このような事態を避けるためにはチャンネル信号を直接処理することによって伝送量を削減する(この処理のことを以降符号化と呼ぶ)必要が生じてくる。

従来提案されている符号化手法としてはDVD-video に応用されている AC-3 [11] やデジタル放送に応用されている MPEG2 AAC [12] などが挙げられ、これらの手法を適用すれば1チャンネルあたり  $48[\text{kHz}] \times 16[\text{bit}] = 768[\text{kbps}]$  を必要とする伝送量を  $64[\text{kbps}]$  にまで削減することができる。しかし、達成される伝送量は1チャンネルあたりの値なので、波面合成法による音場再

<sup>\*1</sup>名古屋大学大学院 情報科学研究科

<sup>\*2</sup>中京大学 情報科学部

<sup>\*3</sup>名城大学 理工学部

<sup>\*1</sup>Graduate School of Information Science, Nagoya Univ.

<sup>\*2</sup>School of Computer and Cognitive Sciences, Chukyo Univ.

<sup>\*3</sup>Faculty of Science and Technology, Meijo Univ.

現システムのようにチャンネルが多数存在する場合には伝送量がチャンネル数に比例して莫大な量になってしまう。一方、チャンネル信号間の冗長性を除去する空間符号化手法 [13, 14] も提案されているが、これらの手法を用いても伝送量はチャンネル数に比例したままである。この問題の解決のためには従来の手法とは異なる観点から伝送すべきチャンネル数を削減する空間符号化手法が必要となる。

我々は今までに音源からマイクロホンアレーまでの室内伝達関数を基にチャンネル信号から音源信号を抽出することによって伝送量をチャンネルの数から音源の数に削減する空間符号化手法を提案し、例えば伝送量を 24 個のチャンネル信号から 5 個の音源信号に削減してもその影響は知覚的に十分許容できるということを示してきた [15]。しかし、今までの検討は音源が静止している場合を扱っていた。音場再現システムの実際の応用例を考えてみると、演劇のように演者という音源が移動している場合がある。そこで、先に提案した空間符号化手法で移動音源も取り扱うことができるようにする必要がある。

本論文では移動音源信号を抽出することで伝送チャンネル数を削減する空間符号化手法を提案する。2 章では提案手法のアルゴリズムについて説明する。3 章では虚像法によって作成したチャンネル信号を対象とした符号化実験について述べる。4 章では主観評価実験によって提案手法の符号化品質について検討し、提案手法が遠隔通信システムを構築する際に有効であることを示す。本論文は筆者らの先行研究 [15] を拡張したものである。従って、アルゴリズムにおける数学的な詳細については先行研究 [15] を参照することをお勧めする。

## 2 アルゴリズム

図 1 に提案した空間符号化手法のブロック図を示す。入力チャンネル信号は原音場内に配置したマイクロホンアレーによって収録された信号である。その際、位置センサによって音源の位置も同時に収録する。送信側では収録した音源の位置情報を基に適切な室内伝達関数を求め、逆伝達関数を算出する。そして、チャンネル信号に逆伝達関数を畳み込んで音源信号を抽出し、音源の位置情報と共に伝送する。受信側ではまず受け取った音源の位置情報を基にそれに対応する室内伝達関数を求める。そして、受信した音源信号に室内伝達関数を畳み込んで出力チャンネル信号を復元する。この手法によって伝送量はチャンネル数から音源の数へ削減される。これは音源の数がチャンネル数よりも非常に少ない場合に大変効果的である。但し、音源の位置情報を付加情報として伝送する必要が生じてくるが、伝送すべ

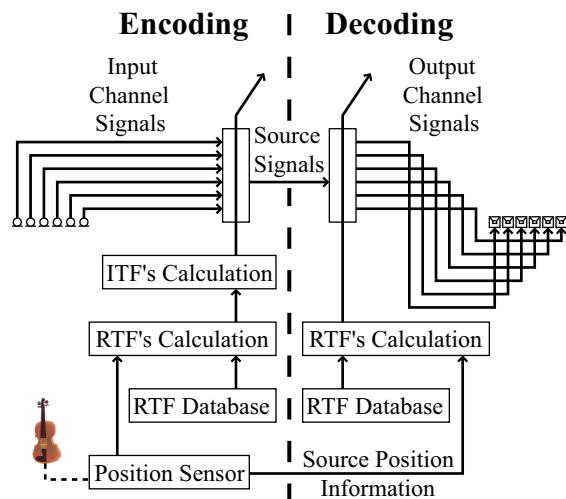


図 1 空間符号化手法のブロック図  
(RTF : 室内伝達関数, ITF : 逆伝達関数)  
Fig.1 Block diagram of spatial coding method (RTF: Room Transfer Function, ITF: Inverse Transfer Function).

き音源位置の情報量は音源信号の情報量と比べると非常に小さく、伝送量全体には影響を及ぼさないと考えられるので、以降では伝送すべき音源位置の情報量は考慮しないことにする。

### 2.1 移動音源信号の抽出

原音場中の  $N$  個の移動音源信号を  $e_i(n)$  ( $i=1\dots N$ ),  $M$  個 ( $M>N$ ) のマイクロホンで収録したチャンネル信号を  $v_j(n)$  ( $j=1\dots M$ ) とする。但し、信号のサンプリング周波数は  $F_s$  [Hz] である。それと同時に、位置センサによって音源の位置情報  $\mathbf{s}_i(m)=[s_{ix}(m), s_{iy}(m), s_{iz}(m)]^T$  もサンプリング周波数  $F_p$  [Hz] で収録されている。但し、 $m, n$  は離散時間を表わす整数で、連続時間  $t$  においてはそれぞれ  $\frac{m}{F_p}, \frac{n}{F_s}$  に換算される。

送信側ではまず得られた音源位置情報  $\mathbf{s}_i(m)$  を基に時間  $m$  における  $i$  番目の音源信号から  $j$  番目のチャンネル信号までの室内伝達関数  $g_{ji}(m, n)$  を以下の式によって求める。

$$g_{ji}(m, n) = g_{jk}(n) \quad (1)$$

但し、 $g_{jk}(n)$  は  $\mathbf{s}_k [=s_i(m)]$  に存在する音源から  $j$  番目のチャンネル信号までの室内伝達関数である。本論文ではこれらの室内伝達関数の情報は図 1 に示すようにデータベースとして送信側と受信側であらかじめ保存されているものとする。室内の形状から室内伝達関数を算出する手法 (例えば [16]) などを適用すれば、全ての  $\mathbf{s}_k$  に対する  $g_{jk}(n)$  を既知にすることは可能であるので、十分に実用的なシステムを構築できると考えられる。

得られた室内伝達関数  $g_{ji}(m, n)$  を基に時間  $m$  にお

ける  $j$  番目のチャンネル信号から  $i$  番目の音源信号までの逆伝達関数  $h_{ij}(m, n)$  を以下の式によって算出する [15].

$$\mathbf{H}(m, \omega) = \mathbf{G}^+(m, \omega) \mathbf{D}(\omega) \quad (2)$$

但し,  $\mathbf{G}^+(m, \omega)$  は  $\mathbf{G}(m, \omega)$  の Moore-Penrose 擬似逆行列で,  $\mathbf{H}(m, \omega)$ ,  $\mathbf{G}(m, \omega)$ ,  $\mathbf{D}(\omega)$  は以下のように定義される行列である.

$$\mathbf{H}(m, \omega) = \begin{pmatrix} H_{11}(m, \omega) & \dots & H_{N1}(m, \omega) \\ \vdots & \ddots & \vdots \\ H_{1M}(m, \omega) & \dots & H_{NM}(m, \omega) \end{pmatrix} \quad (3)$$

$$\mathbf{G}(m, \omega) = \begin{pmatrix} G_{11}(m, \omega) & \dots & G_{M1}(m, \omega) \\ \vdots & \ddots & \vdots \\ G_{1N}(m, \omega) & \dots & G_{MN}(m, \omega) \end{pmatrix} \quad (4)$$

$$\mathbf{D}(\omega) = \begin{pmatrix} e^{-j\omega T_c} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & e^{-j\omega T_c} \end{pmatrix} \quad (5)$$

また,  $H_{ij}(m, \omega)$ ,  $G_{ji}(m, \omega)$  は  $h_{ij}(m, n)$ ,  $g_{ji}(m, n)$  を Fourier 変換したもので,  $T_c (= \frac{P_c}{F_s})$  は因果律を満たす逆伝達関数を算出するために必要な遅延時間であり, この値が符号化遅延に相当する.

得られた逆伝達関数  $h_{ij}(m, n)$  を以下の式のようにチャンネル信号  $v_j(n)$  に畳み込むことによって移動音源信号  $e'_i(n)$  を抽出する.

$$\begin{aligned} e'_i(n) &= \sum_{m=0}^{\infty} w(m, n) \sum_{j=1}^M h_{ij}(m, n) * v_j(n) \\ &= \sum_{m=0}^{\infty} w(m, n) \sum_{j=1}^M \sum_{l=0}^{L_h-1} h_{ij}(m, l) v_j(n-l) \end{aligned} \quad (6)$$

但し,  $*$  は畳み込み演算を表わし,  $L_h$  は逆伝達関数の次数である. また,  $w(m, n)$  は以下の式のように定義される窓関数である.

$$w(m, n) = \begin{cases} \frac{1}{P_{cf}} [n - P_m + \frac{P_{sw}}{2} + \frac{P_{cf}}{2}] & (P_m - \frac{P_{sw}}{2} - \frac{P_{cf}}{2} \leq n < P_m - \frac{P_{sw}}{2} + \frac{P_{cf}}{2}) \\ 1 & (P_m - \frac{P_{sw}}{2} + \frac{P_{cf}}{2} \leq n \leq P_m + \frac{P_{sw}}{2} - \frac{P_{cf}}{2}) \\ -\frac{1}{P_{cf}} [n - P_m - \frac{P_{sw}}{2} - \frac{P_{cf}}{2}] & (P_m + \frac{P_{sw}}{2} - \frac{P_{cf}}{2} < n \leq P_m + \frac{P_{sw}}{2} + \frac{P_{cf}}{2}) \\ 0 & (\text{otherwise}) \end{cases} \quad (7)$$

但し,  $P_{sw} (= \frac{F_s}{F_p})$  は音源位置の切替サンプル数である. 窓関数の形状を図 2 に示す. この関数によって時間  $P_m (= P_c + mP_{sw})$  を中心に  $P_{sw}$  点の範囲で音源信号が抽出され, 隣接する音源位置との間で抽出した音源信号の波形が不連続にならないようにするために窓の両

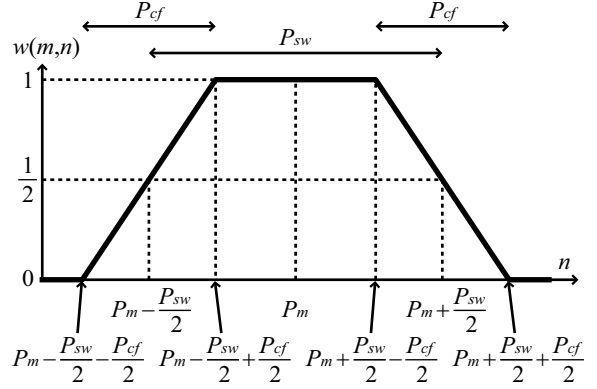


図 2 窓関数の形状  
Fig. 2 Shape of window function.

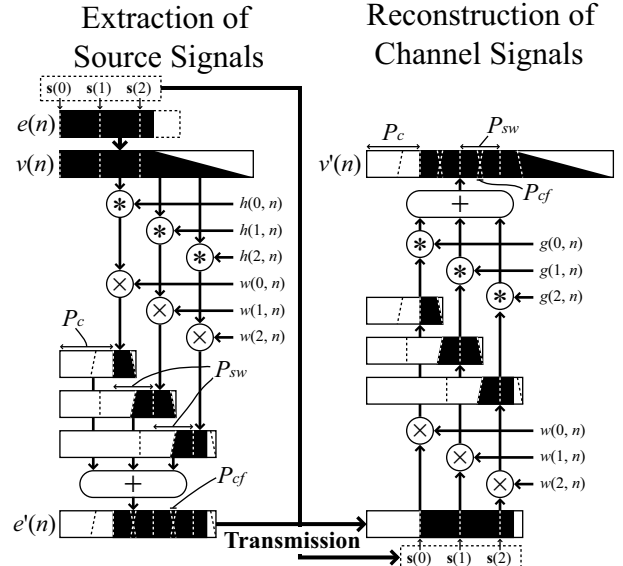


図 3 抽出及び復元のブロック図  
Fig. 3 Block diagram of extraction and reconstruction.

脇に  $P_{cf} (= F_s T_{cf})$  点のリニアなクロスフェードが設けられる. その際,  $P_c$ ,  $P_{sw}$ ,  $P_{cf}$  は以下の条件を満たす必要がある.

$$P_c \geq \frac{P_{sw} + P_{cf}}{2}, \quad P_{sw} \geq P_{cf} \quad (8)$$

音源信号が抽出される様子を図 3 の左側に示す. 図では 1 入力 1 出力なので,  $i, j$  は省略してある. この図のように音源位置が  $s(0), s(1), s(2), \dots$  と変化するに従って逆伝達関数  $h(0, n), h(1, n), h(2, n), \dots$  をチャンネル信号  $v(n)$  に畳み込めば, 抽出される音源  $e'(n)$  は移動音源のドライソースそのものになる.

## 2.2 チャンネル信号の復元

受信側ではまず受信した音源の位置情報  $s_i(m)$  を基に (1) 式によって室内伝達関数  $g_{ji}(m, n)$  を求める. そして, 得られた  $g_{ji}(m, n)$  を以下の式のように受信し

た音源信号  $e'_i(n)$  に畳み込むことによってチャンネル信号  $v'_j(n)$  を復元する。

$$\begin{aligned} v'_j(n) &= \sum_{m=0}^{\infty} \sum_{i=1}^N g_{ji}(m, n) * [w(m, n) e'_i(n)] \\ &= \sum_{m=0}^{\infty} \sum_{i=1}^N \sum_{l=0}^{L_g-1} g_{ji}(m, l) w(m, n-l) e'_i(n-l) \end{aligned} \quad (9)$$

但し,  $L_g$  は室内伝達関数の次数である。

チャンネル信号が復元される様子を図3の右側に示す。この図のように音源位置が  $\mathbf{s}(0), \mathbf{s}(1), \mathbf{s}(2), \dots$  と変化することによって室内伝達関数  $g(0, n), g(1, n), g(2, n), \dots$  を音源信号  $e'(n)$  に畳み込めば, 音源の動きまで再現されたチャンネル信号  $v'(n)$  が復元される。元の音の動きが連続的であるのに対して, 再現された音の動きはサンプリング周波数  $F_p$  [Hz] で離散化されている。  $F_p$  が小さすぎると音の動きが元の音の動きと異なって知覚されるので,  $F_p$  は音の動きが正しく知覚されるように設定する必要がある。一方, 符号化遅延時間  $T_c$  が大きすぎると双方向通信システムの構築の際に円滑なコミュニケーションの妨げとなる [17]。ゆえに,  $T_c$  も双方向通信システムの構築に支障をおきないように設定する必要がある。3章ではこれらのことをふまえて  $F_p$  や  $T_c$  の設定を行う。

### 3 符号化実験

この章では提案した空間符号化手法の有効性を検証した実験について述べる。

#### 3.1 チャンネル信号の作成

実験に用いるチャンネル信号は虚像法 [18] を用いて計算機上で作成した。本論文では2次元音場を対象とするため, 原音場は図4に示すような長方形の2次元音場とし, その中に1個の移動音源と24個の円状マイクロホンアレーが配置されているとした。

虚像法によると,  $x=0, x=L_x, y=0, y=L_y$  に壁面  $x_1, x_2, y_1, y_2$  を配置し,  $\mathbf{s}_i=(s_{ix}, s_{iy})^T$  に  $i$  番目の無指向性音源,  $\mathbf{r}_j=(r_{jx}, r_{jy})^T$  に  $j$  番目の指向性受音点を配置した場合の室内伝達関数  $g_{ji}(t)$  は以下の式のように表わされる。

$$\begin{aligned} g_{ji}(t) &= \sum_{p_x=0}^1 \sum_{p_y=0}^1 \sum_{q_x=-\infty}^{\infty} \sum_{q_y=-\infty}^{\infty} D[\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q})] \\ &\quad \beta_{x_1}^{|q_x-p_x|} \beta_{x_2}^{|q_x-p_x|} \beta_{y_1}^{|q_y-p_y|} \beta_{y_2}^{|q_y-p_y|} \frac{\delta[t - \frac{|\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q})|}{c}]}{4\pi|\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q})|} \end{aligned} \quad (10)$$

但し,  $\delta(t)$  はDiracのデルタ関数,  $c$  は音速 ( $=340$  [m/s]),  $\beta_{x_1}, \beta_{x_2}, \beta_{y_1}, \beta_{y_2}$  は壁面  $x_1, x_2, y_1, y_2$  の反射係数を表わす。以降では  $\beta_{x_1}^{|q_x-p_x|} \beta_{x_2}^{|q_x-p_x|} \beta_{y_1}^{|q_y-p_y|} \beta_{y_2}^{|q_y-p_y|} = B(\mathbf{p}, \mathbf{q})$ ,

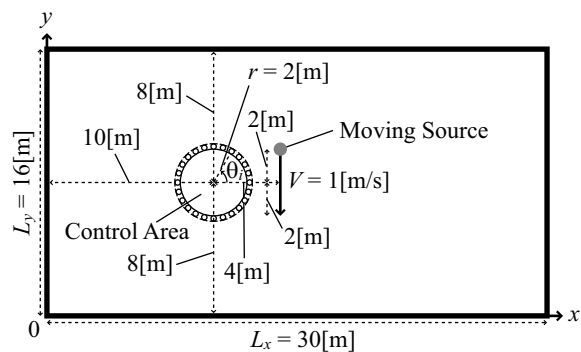


図4 符号化実験に用いた原音場  
Fig. 4 Original sound field used in coding experiment.

$\sum_{p_x=0}^1 \sum_{p_y=0}^1 = \sum_{\mathbf{p}}$ ,  $\sum_{q_x=-\infty}^{\infty} \sum_{q_y=-\infty}^{\infty} = \sum_{\mathbf{q}}$  と表記する。また,

$$\begin{aligned} \mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}) &= \mathbf{s}_i(\mathbf{p}, \mathbf{q}) - \mathbf{r}_j \\ &= \begin{pmatrix} (1-2p_x)s_{ix} + 2q_xL_x \\ (1-2p_y)s_{iy} + 2q_yL_y \end{pmatrix} - \begin{pmatrix} r_{jx} \\ r_{jy} \end{pmatrix} \end{aligned} \quad (11)$$

で,  $\mathbf{s}_i(\mathbf{p}, \mathbf{q})$  は虚像音源の位置ベクトルを表わす。一方,  $D[\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q})]$  は受音点の指向性を表わす関数であり, 本論文では以下の式に示されるようなカーゴイド型とする。

$$D[\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q})] = \frac{1}{2} \left( 1 + \frac{\mathbf{u}_j \cdot \mathbf{d}_{ji}(\mathbf{p}, \mathbf{q})}{|\mathbf{u}_j| |\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q})|} \right) \quad (12)$$

但し,  $\mathbf{u}_j=(u_{jx}, u_{jy})^T$  は  $j$  番目の受音点の指向ベクトルである。得られた  $g_{ji}(t)$  を音源信号  $e_i(t)$  に畳み込むことによって音源が静止している場合のチャンネル信号  $v_j(t)$  が得られる。

$$\begin{aligned} v_j(t) &= \sum_{i=1}^N g_{ji}(t) * e_i(t) \\ &= \sum_{i=1}^N \sum_{\mathbf{p}} \sum_{\mathbf{q}} D[\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q})] \end{aligned} \quad (13)$$

$$B(\mathbf{p}, \mathbf{q}) \frac{e_i[t - \frac{|\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q})|}{c}]}{4\pi|\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q})|}$$

音源が移動する場合には音源や虚像音源の位置が時間  $t$  によって変化する。本節で作成するチャンネル信号は符号化する前の信号であり, 音源の位置は全ての  $t$  において判明しているとする。従って, (13) 式は以下のように変形される。

$$\begin{aligned} v_j(t) &= \sum_{i=1}^N \sum_{\mathbf{p}} \sum_{\mathbf{q}} D[\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, t)] \\ &\quad B(\mathbf{p}, \mathbf{q}) \frac{e_i[t - \frac{|\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, t)|}{c}]}{4\pi|\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, t)|} \end{aligned} \quad (14)$$

計算機上で扱う音源信号やチャンネル信号は離散時間信号なので, 時間  $t$  を離散時間  $n(=t \cdot F_s)$  にすると,

$$v_j(n) = \sum_{i=1}^N \sum_{\mathbf{p}} \sum_{\mathbf{q}} D[\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n)] B(\mathbf{p}, \mathbf{q}) \frac{e_i[n - \frac{F_s |\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n)|}{c}]}{4\pi |\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n)|}. \quad (15)$$

(15) 式は時間  $n$  に音源信号を入力すると, 時間  $n + \frac{F_s |\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n)|}{c}$  にチャンネル信号が出力されるということを表わしている. しかし, 一般的には  $n$  に整数を代入しても  $n + \frac{F_s |\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n)|}{c}$  は整数にならないため, (15) 式をそのまま適用しても出力信号は  $n$  の離散時間信号にならない. そこで, 入力時間を  $n+a(0 \leq a < 1)$  とし, 得られる出力時間  $n+a + \frac{F_s |\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n+a)|}{c}$  を整数にすることによって, 出力信号が  $n$  の離散時間信号として出力されるようにする. 但し,

$$\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n+a) = (1-a)\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n) + a\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n+1). \quad (16)$$

その際, 出力時間を  $K = n+a + \frac{F_s |\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n+a)|}{c}$  と表記すると,  $a$  は  $i, j, \mathbf{p}, \mathbf{q}, n$  ごとに以下の二次方程式を解くことによって得られる.

$$\begin{aligned} & \{|\Delta \mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n)|^2 - \frac{c^2}{F_s^2}\} a^2 \\ & + 2\{\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n) \cdot \Delta \mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n) + \frac{c^2(K-n)}{F_s^2}\} a \\ & + \{|\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n)|^2 - \frac{c^2(K-n)^2}{F_s^2}\} = 0 \end{aligned} \quad (17)$$

但し,

$$\Delta \mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n) = \mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n+1) - \mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n). \quad (18)$$

また,  $K$  は以下の条件を満たす整数に設定する.

$$\begin{aligned} \text{ceil}(n + \frac{F_s |\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n)|}{c}) \leq K \leq \\ \text{floor}(n+1 + \frac{F_s |\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n+1)|}{c}) \end{aligned} \quad (19)$$

ゆえに, 音源が移動している場合のチャンネル信号は以下のようにして得られる.

$$v_j(n) = \sum_{i=1}^N \sum_{\mathbf{p}} \sum_{\mathbf{q}} D(\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n+a)) B(\mathbf{p}, \mathbf{q}) \frac{e_i[n+a - \frac{F_s |\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n+a)|}{c}]}{4\pi |\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n+a)|} \quad (20)$$

その際の入力信号  $e_i(n+a)$  は以下の式のように線形補間を行ってから用いる.

$$e_i(n+a) = (1-a)e_i(n) + ae_i(n+1) \quad (21)$$

表 1 チャンネル信号の作成条件  
Table 1 Synthetic conditions of channel signals

ドライソース	音声	フルート
$F_s$ (サンプリング周波数)	48[kHz]	
音源の長さ	4[second]	
$\beta$ (壁面の反射係数)	0.5	0.7
$R$ (最大の壁面反射回数)	6	10
部屋の残響時間 [second]	0.6	1.0
$V$ (音源の移動速度)	1[m/s]=3.6[km/h]	

チャンネル信号の作成条件を表 1 に示す. 2 種類の作成条件は会議室で歩きながら話をしたり, コンサートホールでフルートを演奏しながら歩いたりしている場合を想定した. この場合, 移動速度は音速に対して十分小さいため, ドップラー効果は発生しない. 音声は無響室で収録し, フルードは MIDI 音源から作成したものを用了. また, 4 つの壁面の反射係数は同一とし,  $p_x, p_y, q_x, q_y$  の値は以下の条件を満たす全ての組み合わせを用了.

$$|q_x - p_x| + |q_x| + |q_y - p_y| + |q_y| \leq R \quad (22)$$

但し,  $R$  は表 1 に示すように設定した壁面反射回数の最大値である. また, 部屋の残響時間は Schroeder 法 [19] により算出した.

### 3.2 室内伝達関数の算出

室内伝達関数の情報をデータベース化するために, 図 4 中の直線状の音源の軌跡を 8.33mm ごとに区切り, 得られた計 481 点の音源位置から 24 点の受音点位置までの室内伝達関数  $g_{jk}(n)(k=1 \dots 481, j=1 \dots 24)$  を以下の式によって算出した.

$$g_{jk}(n) = \sum_{\mathbf{p}} \sum_{\mathbf{q}} D[\mathbf{d}_{jk}(\mathbf{p}, \mathbf{q})] B(\mathbf{p}, \mathbf{q}) \frac{\delta[n - \text{round}(\frac{F_s |\mathbf{d}_{jk}(\mathbf{p}, \mathbf{q})|}{c})]}{4\pi |\mathbf{d}_{jk}(\mathbf{p}, \mathbf{q})|} \quad (23)$$

壁面の反射係数や最大の壁面反射回数は表 1 と同じ条件とした. また,  $D[\mathbf{d}_{jk}(\mathbf{p}, \mathbf{q})]$  を算出するために必要な  $\mathbf{u}_j$  や  $\mathbf{s}_k, \mathbf{r}_j$  は以下のように設定した.

$$\begin{aligned} \mathbf{u}_j &= \begin{pmatrix} \cos \frac{\pi(j-12)}{12} \\ \sin \frac{\pi(j-12)}{12} \end{pmatrix} \\ \mathbf{s}_k &= \begin{pmatrix} 14 \\ 10 - \frac{k-1}{120} \end{pmatrix} \\ \mathbf{r}_j &= 2\mathbf{u}_j + \begin{pmatrix} 10 \\ 8 \end{pmatrix} \end{aligned} \quad (24)$$

### 3.3 音源信号の抽出及びチャンネル信号の復元

まず, (1) 式によって  $g_{jk}(n)$  から移動音源を表わす室内伝達関数  $g_{ji}(m, n)$  を求めた. 本実験では  $m = \frac{F_p(k-1)}{120}$  とし,  $F_p$  には 30, 60, 120[Hz] の値を用了.

表 2 逆伝達関数の算出条件  
Table 2 Calculation conditions of inverse transfer functions

残響時間 [second]	0.6	1.0
FFT フレーム長 [sample]	65536	131072
符号化遅延時間 $T_c$ (=符号化遅延サンプル数 $P_c$ )	20[ms] (=960[sample])	
フィルタ長 [sample]	28800	48000

次に、 $g_{ji}(m, n)$  から (2) 式によって  $h_{ij}(m, n)$  を算出した。算出条件を表 2 に示す。我々が先に静止音源を対象に行った研究では符号化遅延時間  $T_c$  を 20ms 以上にすれば符号化歪みによる品質への影響は小さくなった [15]。ゆえに、 $T_c$  を 20ms にすれば逆伝達関数の算出による品質への影響は少ないと考え、本実験では符号化遅延を 20ms に設定した。また、2 章でも述べたように、設定した符号化遅延時間は双方向通信システムを構築する場合においても円滑なコミュニケーションの妨げにならない値である。

得られた逆伝達関数  $h_{ij}(m, n)$  を 3.1 節で得られたチャンネル信号  $v_j(n)$  に畳み込んで音源信号  $e'_i(n)$  を抽出した。その際、窓関数  $w(m, n)$  に用いた切替サンプル数  $P_{sw}$  の値 (=1600, 800, 400[sample]) は先ほど設定した  $F_p$  を換算した値で、クロスフェード時間  $T_{cf}$  には 1, 4[ms] の値を用いた (クロスフェードサンプル数  $P_{cf}$  に換算すると 48, 192[sample])。  $P_{sw}$  や  $P_{cf}$  の値は (8) 式で与えられた条件に従っている。また、抽出した音源信号  $e'_i(n)$  に室内伝達関数  $g_{ji}(m, n)$  を畳み込んでチャンネル信号  $v'_j(n)$  を復元した。

#### 4 主観評価実験

我々が先に行った研究では波形的な歪みが大きい場合においても音源定位や音像の大きさへの影響は必ずしも大きくなかった [15]。本論文における符号化実験に対しても SN 比のような尺度を用いて符号化性能を客観的に評価することは適切ではないと考えられる。そこで、主観評価実験によって提案した空間符号化手法が知覚に及ぼす影響を検討する。

##### 4.1 実験環境

実験条件は先の研究 [15] とほぼ同じ条件に設定した。Huygens の原理による音場再現においては主観評価実験は本来なら無響室で行う必要があるが、実験設備の都合上、残響時間が短い部屋 (約 80ms の低残響室) で行った。実験に用いたスピーカアレーと受聴者の位置を図 5 に示す。各スピーカからチャンネル信号を再生すると、スピーカアレーの中に音場が再現され、アレー内にいる受聴者は図 5 の灰色で示されるように音像が移動しているように感じる。暗騒音レベルは

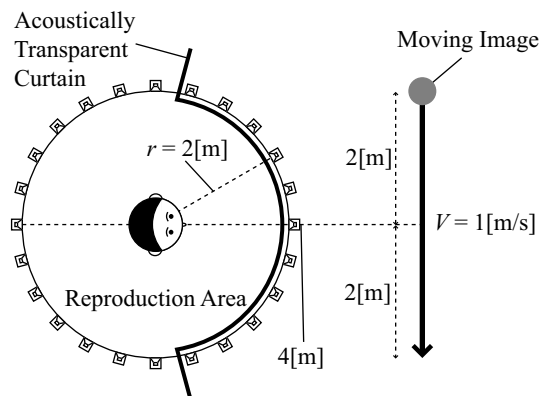


図 5 主観評価実験に用いたスピーカアレーと受聴者の位置

Fig.5 Positions of loudspeaker array and subject used in the subjective assessment.

##### Subjective Assessment

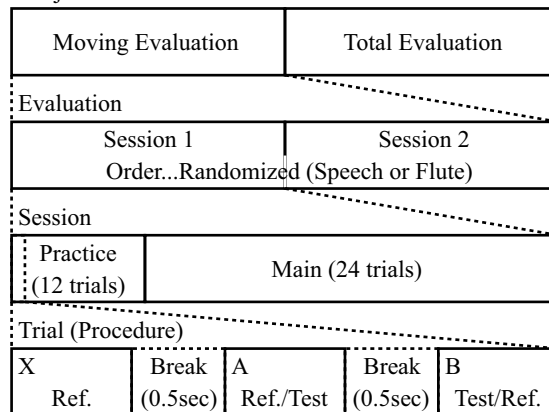


図 6 主観評価実験の流れ図

Fig.6 Flowchart of the subjective assessment

25.0dB(A) で、音圧レベルは受聴者の位置において約 70dB(A) とした。また、視覚の影響を低減するため、照明を薄暗くし、スピーカアレーの前方は音響透過カーテンで覆った。

##### 4.2 実験計画

被験者は日常生活において聴力に異常の見られない 8 名の男子学生である。ITU-R 勧告 [20] に従って三刺激二重盲験法を実験手法とした。主観評価実験の流れを図 6 に示す。まず、実験全体を 2 つに分割し、それぞれを“動き評価”、“全体評価”と名付けた。次に、それぞれの評価を 2 つのセッションに分割し、音源の異なる (音声とフルート) 2 種類のチャンネル信号を提示した。その際、音源の提示順序は被験者ごとにランダム化した。セッションごとに練習試行と本試行を順に行い、1 回の試行において 3 種類の刺激を順に提示した。1 つの刺激の長さは 4 秒である。最初に提示される刺激 (以下 “X”) には常に符号化しない原音を割

表 3 符号化音の種類  
Table 3 Series of coding sound

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

表 4 違いの評定表  
Table 4 Scale table of impairment

差の程度	評定
差が分からない	5.0
差は分かるが、気にならない	4.0
差がわずかに気になる	3.0
差が気になる	2.0
差が非常に気になる	1.0

り当て、2番目に提示される刺激(以下“A”)と3番目に提示される刺激(以下“B”)には原音もしくは表3に示されるような6種類の符号化音を割り当てた。試行の提示順序も被験者ごとにランダム化した。練習試行は6(符号化音の種類)×2(A or Bに割り当て)=12試行を行い、本試行は6(符号化音の種類)×2(A or Bに割り当て)×2(繰り返し)=24試行を行った。

### 4.3 実験手順

前半の動き評価では符号化による音像の動き知覚への影響を評価した。被験者に以下のように教示した。「X”は基準となる動き方をしている音です。“A”と“B”のどちらかは必ず基準の音と同じ動き方をする音です。“A”と“B”のうち、基準の音と同じ動き方をする音だと思った方を“5.0”と評定して下さい。もう一方の音は表4に従って、音の動き方の差がどれくらいあるかを小数点第1位単位で“1.0”から“4.9”の範囲で評定して下さい。」

後半の全体評価では先に評価した音の動き知覚への影響に加え、符号化による音質への影響をも含んだ全体的な知覚への影響を検討した。被験者に以下のように教示した。「X”は原音です。“A”と“B”のどちらかは必ず原音と同じ音です。“A”と“B”のうち、原音と思った方を“5.0”と評定して下さい。もう一つは表4に従って、原音との差がどれくらいかを小数点第1位単位で“1.0”から“4.9”までで評定して下さい。」

主観評価実験を行っている間、被験者は頭部を自由に動かすことができた。また、判断がつくまでは何度も“X”, “A”, “B”を繰り返し聞くことができた。

### 4.4 実験結果及び考察

実験結果を分析する前に、被験者が原音と符号化音を正しく識別し、原音を割り当てた刺激を「5.0」と評定しているかを検査する。各被験者が原音を「5.0」

表 5 各被験者の識別結果  
Table 5 Result of discrimination in each subject

被験者	標本数	動き評価		全体評価	
		音声	フルート	音声	フルート
A	24	<b>15</b>	<b>23</b>	<b>15</b>	<b>22</b>
B	24	7	<b>23</b>	<b>16</b>	<b>24</b>
C	24	7	<b>24</b>	14	<b>24</b>
D	24	<b>14</b>	21	11	21
E	24	<b>14</b>	14	<b>15</b>	17
F	24	11	22	9	22
G	24	11	19	9	20
H	24	9	18	10	17

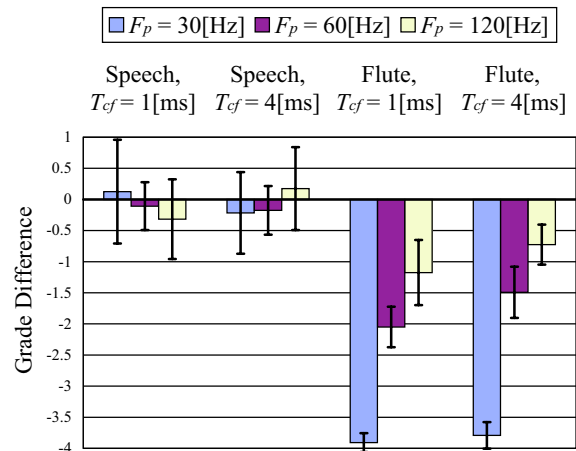


図 7 動き評価の結果

Fig.7 Results of moving evaluation in the subjective assessment.

と評定した回数をセッションごとに表5に示す。評定した回数が多いほど、被験者の判断は信頼性が高いと考えられるので、分析には各セッションにおける上位3名(表中で太字で示した箇所)の結果を用いることにした。

符号化音の評定値から原音の評定値の差を算出した結果を図7, 8に示す。エラーバーは95%信頼区間を表わす。動き評価の結果と全体評価の結果はほとんど同じである。これは提案した符号化手法による知覚的な歪みは2章で述べたように音の動きが離散的に再現されることで引き起こされる歪みに依存するというを示している。フルートの評定差は音声よりも全体的に低くなっている。これはフルートは調波複合音で、どちらかという純音に近い音であるため、音の動きを再現する際に窓関数を重ね合わせることで生じるスペクトル歪みが音声の場合よりも鋭敏に感知されてしまったと考えられる。ゆえに、提案した符号化手法は音源の種類によって知覚的な影響が大きく異なるということが言える。

これらの実験結果から遠隔通信システムが実用的に



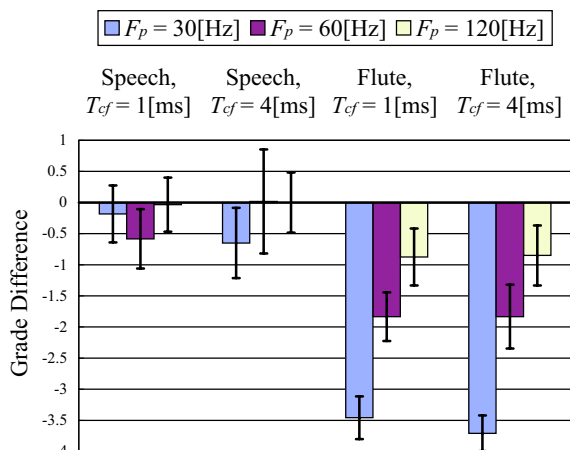


図8 全体評価の結果

Fig. 8 Results of total evaluation in the subjective assessment.

構築できるかどうかを考察する。遠隔アンサンブルシステムを構築する場合、対象となる音源は主に楽音であるため、フルートの実験結果を参考にすることがある。そのため、符号化による知覚的な影響を低減するためには  $F_p$  をなるべく大きい値に設定する必要がある。 $F_p$  が 120[Hz] の場合には音の動きや全体的な知覚への影響は「差は分かるが、気にならない」程度である。この値はカメラなどを用いた位置検出システムで十分に実現できる性能であるので、映像と組み合わせた遠隔アンサンブルシステムを構築する場合、提案した空間符号化手法を導入すれば十分実用的なシステムを構築できると考えられる。また、カメラなどを用いた位置検出システムを用いると音源の位置を収録するための位置センサを音源に直接配置する必要はなくなるので、音源収録用のマイクロホンや位置センサが映像に映り込むことのないよりマルチモーダルに適したシステムを構築することも可能であるとと言える。

一方、遠隔通信会議システムの場合は対象とする音源は主に音声である。音声に対する実験結果を見ると、 $F_p$  が 30[Hz] の場合でも、被験者は知覚的な影響を識別できていないことが多い。ゆえに、映像と組み合わせた遠隔通信会議システムを構築する場合、遠隔アンサンブルシステムを構築する場合に比べてより簡易な位置検出システムを導入しても知覚上問題ないシステムを実現できると考えられる。

## 5 まとめ

音場再現システムを遠隔で実現するには伝送量を削減する必要があるため、本論文では音源の位置情報を基に移動音源を抽出することで伝送量をチャンネル信号の数から音源の数に削減する空間符号化手法を提案

した。虚像法によって作成した残響空間におけるチャンネル信号を対象に符号化実験を行った。符号化性能を評価するために主観評価実験を行ったところ、システムを構築する際に対象となる音源に対して適切なパラメータを設定すれば、音の動き知覚や全体的な知覚への影響のほとんどないシステムを実用化できるということが分かった。

本論文で作成したチャンネル信号は虚像法を用いて仮想的に模擬したものであるが、今後は実環境で実際に収録したチャンネル信号を用いて検討する必要がある。また、実用化の際に高性能な位置検出システムを利用することができない場合も考えられるので、より簡易な位置検出システムによる音源位置情報から音源位置を補間によって予測し、符号化に用いる手法も検討する必要がある。また、本論文では送信側と受信側であらかじめ室内伝達関数が既知であると想定した場合のみを検討しているため、実用上室内伝達関数をあらかじめ測定しておく必要があり、実現上問題がある。特に、遠隔アンサンブルシステムを構築する場合には演奏者と楽器の位置が異なるためにそれぞれに対して室内伝達関数データベースが必要になる可能性がある。そこで、これらの問題を解決し、より実用的な遠隔通信システムを構築していくためには室内の形状などから室内伝達関数を推定する手法とその推定精度が品質に及ぼす影響を検討していく必要がある。

## 謝辞

本研究は文部科学省科学研究費補助金 COE 形成基礎研究費 (課題番号 11CE2005)、基礎研究 B(課題番号 11480075) 及び基盤研究 B(課題番号 14380159) の一部の補助を受けて行われた。

## 参考文献

- [1] M. Camras: "Approach to recreating a sound field", *J. Acoust. Soc. Am.*, **43**, 6, pp. 1425-1431 (1968).
- [2] A. J. Berkhout, D. de Vries and P. Vogel: "Acoustic control by wave field synthesis", *J. Acoust. Soc. Am.*, **93**, 5, pp. 2764-2778 (1993).
- [3] S. Ise: "A principle of sound field control based on the Kirchhoff-Helmholtz integral equation and the theory of inverse systems", *Acustica - Acta Acustica*, **85**, 1, pp. 78-87 (1999).
- [4] S. Takane, Y. Suzuki and T. Sone: "A new method for global sound reproduction based on Kirchhoff's integral equation", *Acustica - Acta Acustica*, **85**, 2, pp. 250-257 (1999).
- [5] 及川靖広, 馬場貞如, 山崎芳男: "人間の聴覚を考慮した波面合成法による音場制御と音場再現", *日本バーチャルリアリティ学会論文誌*, **5**, 3, pp. 919-926 (2000).
- [6] 平井徹, 本地由和, 池田雅弘, 清水寧, 川上福司, 東山三樹夫: "遠隔地音楽レッスン用 AVW システムとその原理", *応用音響研究会技術報告 EA2000-17*, 電子



[著者紹介]

- 情報通信学会 (2000).
- [7] 浜田晴夫: “バイノーラル音場再生系について”, 日本音響学会誌, **48**, 4, pp. 250–257 (1992).
- [8] S. Takane, Y. Suzuki, T. Miyajima and T. Sone: “A new theory for high definition virtual acoustic display named ADVISE”, *Acoust. Sci. and Tech.*, **24**, 5, pp. 276–283 (2003).
- [9] 安藤四一, 設楽貞樹, 前川純一, 城戸健一: “電子計算機による室内音響設計法に関する二, 三の基礎的考察”, 日本音響学会誌, **29**, 3, pp. 151–159 (1973).
- [10] J. Bauck and D. H. Cooper: “Generalized transaural stereo and applications”, *J. Audio Eng. Soc.*, **44**, 9, pp. 683–705 (1996).
- [11] S. Vernon: “Design and implementation of AC-3 coders”, *IEEE Trans. CE*, **41**, 3, pp. 754–759 (1995).
- [12] ISO/IEC 13818-7: “Information Technology - Genetic Coding of Moving Pictures and Associated Audio Information - Part 7 Advanced Audio Coding”.
- [13] D. Yang, H. Ai, C. Kyriakakis and C.-C. J. Kuo: “High-fidelity multichannel audio coding with Karhunen-Loève transform”, *IEEE Trans. SAP*, **11**, 4, pp. 365–380 (2003).
- [14] Y. Wang, M. Vilermo, M. Väänänen and L. Yaroslavsky: “A multichannel audio coding algorithm for inter-channel redundancy removal”, *AES 110th Convention Preprints*, No. 5295 (2001).
- [15] 木村敏幸, 寛一彦, 武田一哉, 板倉文忠: “音場再現におけるマルチチャンネル信号の空間符号化”, 日本バーチャルリアリティ学会論文誌, **8**, 4, pp. 433–442 (2003).
- [16] CATT-Acoustic: <http://www.catt.se/>.
- [17] 伊藤憲三, 北脇信彦: “双方向通信系における伝搬遅延の許容限に関する検討”, 聴覚研究会資料 H84-57, 日本音響学会 (1984).
- [18] J. B. Allen and D. A. Berkley: “Image method for efficiently simulating small-room acoustics”, *J. Acoust. Soc. Am.*, **65**, 4, pp. 943–950 (1979).
- [19] M. R. Schroeder: “New method of measuring reverberation time”, *J. Acoust. Soc. Am.*, **37**, 3, pp. 409–412 (1965).
- [20] ITU-R Recommendation BS.1116-1: “Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems”.

(2004年7月23日受付)

木村 敏幸 (正会員)



1998年名大・工・物化卒。2003年同大学院・人間情報・博士課程了。同年より、日本学術振興会特別研究員(特別推進研究COE)。2004年名大・情報科学・研究員。音場再生に関する研究に従事。日本音響学会, 電子情報通信学会各会員。

寛 一彦



1965年早大・理工・電気卒。1967年同大学院修士課程了。同年日本電信電話公社(現NTT)入社。通話品質, 聴知覚の基礎研究に従事。1994年名大・人間情報・教授。2003年名大・情報科学・教授。2004年中京大・情報科学・教授。工博。現在, 音声知覚・理解・コミュニケーションの研究を行う。マルチモーダルな音環境認知, 視聴覚情報の分離・統合の処理過程にも興味。1981年度日本ITU賞, 1984年度電子情報通信学会論文賞, 1985年度科学技術庁長官賞受賞。日本音響学会, 電子情報通信学会, 日本認知科学会, 日本音聲学会, 日本高次脳機能障害学会, 日本語学会等会員。

武田 一哉



1983年名大・工・電気卒。1985年同大学院博士(前期)課程了。同年国際電信電話株式会社入社。ATR自動翻訳電話研究所, KDD研究所において音声合成・認識システムの研究を行う。1994年名大・工・助教授。2003年名大・情報科学・教授。工博。日本音響学会, IEEE, 電子情報通信学会, 情報処理学会各会員。

板倉 文忠



1963年名大・工・電気卒。1968年同大学院博士課程満了。同年日本電信電話公社(現NTT)武蔵野通研入所。音声処理の研究に従事。1973–1975年ベル研究所にて音声認識・音声分析の研究を行う。1984年名大・工・教授。1998年名大・情報メディア教育センター教授。2004年名城大・理工・教授。工博。1970年, 1978年, 1981年度電子情報通信学会論文賞, 1972年, 1981年度同業績賞受賞。1996年IEEE Signal Processing Society Award受賞。2000年IEEE Third Millennium Medals受賞。電子情報通信学会フェロー。2003年紫綬褒章受章。日本音響学会, 電子情報通信学会, IEEE各会員。