

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*

木村敏幸, 箕一彦(名大・人間情報/CIAIR), 武田一哉, 板倉文忠(名大・工/CIAIR)

1. はじめに

マルチチャンネル音響信号を用いた音場再現は聴覚バーチャルリアリティに応用される技術として注目されている。しかし、その技術を通信に適用する際には AC-3 や MPEG2 AAC のような従来の符号化技術では不十分である。そこで、我々は現在の符号化技術では用いられていないチャンネル間の情報を利用して、空間に存在する音源情報を抽出して更なる圧縮を行うことを検討している [1]。

2. マルチチャンネル音響信号の作成

マルチチャンネル音響信号はドライソースに室内伝達関数を畳み込むことによって作成した。室内伝達関数を測定するために用いた部屋(名古屋大学工学部 7 号館 410 号室)の形状とスピーカーとマイクロホンアレーの配置図を図 1 に示す。

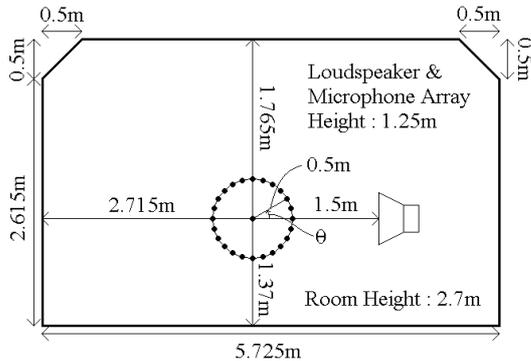


図 1 スピーカーと
マイクロホンアレーの配置図

部屋の残響時間は壁面の反射率を変化させて、151ms と 303ms に設定した。マイクロホンアレーは半径 0.5m の円状の 15 度間隔と円中心に計 25 点配置した。測定条件を表 1 に示す。

表 1 測定条件

| | |
|------------|----------------|
| 室温 | 21 |
| 暗騒音レベル | 19.6dB(A) |
| サンプリング周波数 | 48kHz |
| 信号 | TSP(1.37sec) |
| 同期加算 | 8 回 |
| FIR フィルタ次数 | 16384(0.34sec) |

ドライソースは 3 秒間の白色雑音および女性話者音声を用いた。共にサンプリング周波数は 12kHz で、帯域は 50Hz ~ 5kHz に制限されている。この場合、残響時間は測定時の 4 倍になるので、残響時間はそれぞれ 0.6sec、1.2sec となる。

3. 音源信号の抽出原理

図 2 に音源 N 個、受音点 M 個の音響系を示す。

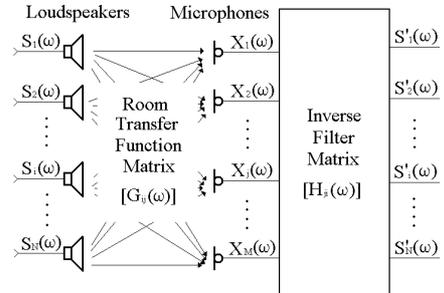


図 2 N 個の音源と M 個の受音点で
構成される音響符号化システム

スピーカー i からマイクロホン j への室内伝達関数 $G_{ij}(\omega)$ からなる行列を $\mathbf{G}([G_{ij}(\omega)]): N \times M$ 次元複素行列)、マイクロホン j からスピーカー i への逆フィルタ $H_{ji}(\omega)$ からなる行列を $\mathbf{H}([H_{ji}(\omega)]): M \times N$ 次元複素行列) とすると、次式の関係が成り立つ。

$$\mathbf{GH} = \mathbf{D}$$

但し、 \mathbf{D} は $D_{ii}(\omega) (=S'_i(\omega)/S_i(\omega))$ を対角成分にもつ N 次元正則複素行列である。

音源の数が受音点の数より少ない場合 ($N < M$)、連立方程式の解を一意に定めることができない。このような場合、最小ノルム解を用いることによって逆フィルタを設計できることが分かっている [2][3]。

$$\mathbf{H} = \mathbf{G}^T (\mathbf{GG}^T)^{-1} \mathbf{D}$$

\mathbf{G}^T は \mathbf{G} の共役転置行列である。

4. 符号化および復号化

3. で記述した原理を用いて 25 チャンネルの入力信号から 1 チャンネルの音源信号を抽出し、元に復元した。測定した FIR フィルタ(16384 点)を室内伝達関数 \mathbf{G} に使い、逆フィルタ \mathbf{H} は \mathbf{G} と \mathbf{D} から算出した。 \mathbf{D} には 8192 点の遅延を持った、50Hz ~ 5kHz のバンドパスフィルタを適用した。

逆フィルタ \mathbf{H} の算出は以下の手順に従って行った。まず、測定した室内伝達関数を 16384 点 FFT によって周波数領域に変換した。そして、最小ノルム解によって逆フィルタを周波数領域で求めた後、16384 点 IFFT によって時間領域の逆フィルタを算出した。

5. 主観評価実験

今回提案した符号化法によって再現される音場の方向感に関する再現性について調査した。

実験は低残響室にて行った。スピーカーを半径 2m の円状に 15 度間隔で計 24 個配置した。そのうち 30 度間隔の 12 個を再生に使用した。スピーカーの配置図を図 3 に示す。

*A study of multi-channel audio coding using spatial information and its sound localization, By T. Kimura, K. Kakehi, K. Takeda and F. Itakura (Nagoya University)

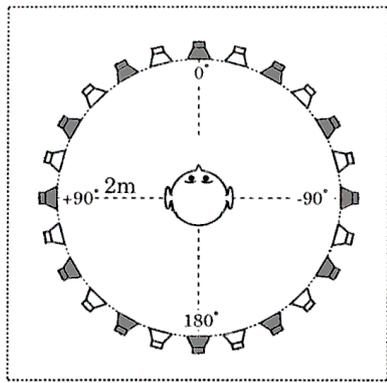


図3 スピーカー配置図[3]

実験条件を表2に示す。

表2 実験条件

| | |
|-------|---|
| 被験者数 | 13名(男8名、女5名) |
| 音圧レベル | 円中心で約70dB(A) |
| セッション | 4=2(ドライソース) ×2(残響時間) |
| 試行数 | 96 = 12(方向) ×2(符号化あり・なし) ×4(繰り返し) |

セッションの順序は被験者ごとにランダムに割り当て、試行の呈示順序はセッション内でランダムに割り当てた。

実験手順としてはまず試行前に正面方向に頭を向けてもらった。そして、音を聞いたら音像がどの方向から鳴っているかをスピーカー番号(24 択)で回答してもらった。その際に頭部運動は自由とした。そして、音が鳴り終わってから次の試行に入るまでに再び頭を正面方向に向けてもらった。

6. 結果および考察

得られた結果から前後誤りによる誤差を除去した後、呈示角度ごとにヒストグラムをプロットし、正答からの平均誤差を算出した。呈示角度ごとに平均誤差をプロットしたものを図4~7に示す。黒抜きが符号化を行わなかった場合、白抜きが符号化を行った場合の結果である。

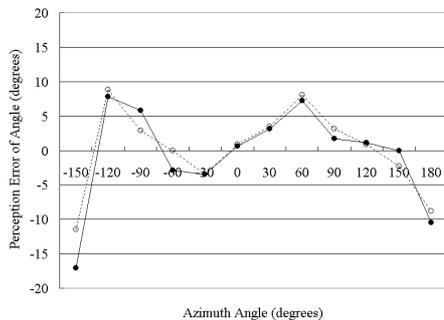


図4 白色雑音、残響時間0.6sec

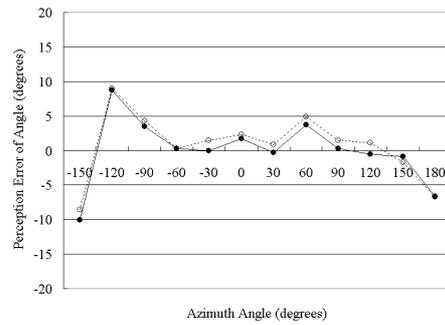


図5 女性話者音声、残響時間0.6sec

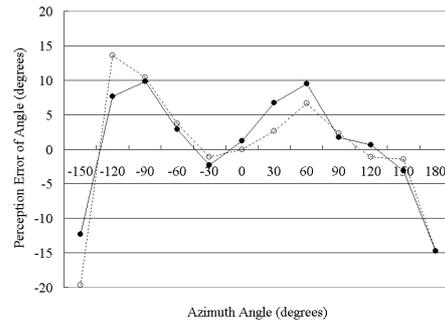


図6 白色雑音、残響時間1.2sec

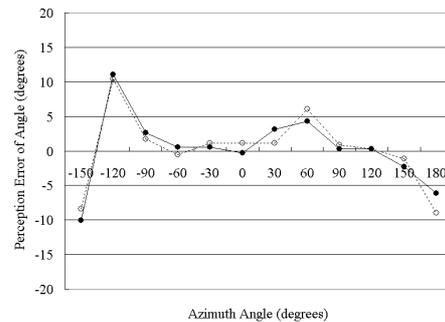


図7 女性話者音声、残響時間1.2sec

符号化による回答角度への影響はほとんど見られない。実際、符号化・呈示角度・サンプルの3要因被験者間分散分析を残響時間ごとに行ったところ、どちらの残響時間においても符号化の要因に関しては主効果及び交互作用には有意差が見られなかった(有意水準5%)。

8. まとめ及び今後の予定

本研究では音源分離技術を用いたマルチチャンネル音響符号化方式を提案した。主観評価を行ったところ、提案した方式が方向感には影響を及ぼさなかった。

今後の予定としては他の音響知覚への影響について検討する。また、今回は単音源の場合のみを検討したが、複数音源の場合の検討を考えている。

9. 参考文献

- [1] 木村, 箕, 武田, 板倉, "主成分分析によるマルチチャンネル音響符号化の検討," 音講論, 2-7-15, 547-548 (2001.3)
- [2] 神沼, 伊勢, 鹿野, "周波数領域における最小ノルム解を利用した多チャンネル音場再現システムにおける逆フィルタの設計," 音響学会誌, 57, 175-183 (2001)
- [3] 古見, 穂刈, 島田, "立体音像再生に関するスピーカ配置の検討," 信学論 A, Vol.J83-A, No.7, 920-923 (2000)