

主成分分析による マルチチャンネル音響符号化の検討*

木村敏幸(名大・人情), 箕一彦(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)

1. はじめに

最近のバーチャルリアリティ技術の発達により、ある空間上での音響信号を送信し、別の空間上で再現するといういわゆる音場伝送技術を構築することも不可能ではなくなってきた。

音場を再現するには多数のチャンネルの情報を必要とするが、そのままでは伝送に要するデータ量は膨大になってしまう。そこで、データを削減するために符号化技術を用いてデータを圧縮することが試みられている。現在確立されている符号化技術[1][2]ではチャンネル間の空間情報をあまり重視していないため、音場伝送技術に応用するためにはチャンネル間の空間情報の特性を利用した符号化技術を構築し、更なる情報の圧縮を行う必要がある。

2. マルチチャンネル音響信号の作成

符号化の検討の前に、対象として用いるマルチチャンネル音響信号を作成する必要がある。作成は名古屋大学工学部7号館410室の可変残響室内で行った。この部屋は室内の残響時間(RT)を150ms~300msに変化させることができる。マイクロホンアレーは半径0.5mの円周上に方位角 $\theta = -165^\circ \sim 180^\circ$ の15°間隔に24点、スピーカは0°の方向に配置した。ラウド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。

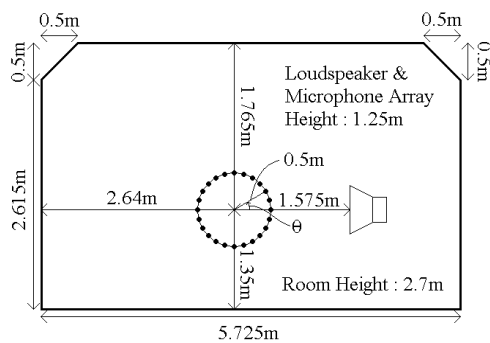


図1 ラウドスピーカーとマイクロホンアレーの配置図

残響時間が150ms, 300msの時にラウドスピーカーからマイクロホンアレーまでの室内インパルス応答をTSP信号(44.1kHz, 5.944sec)により測定し、得られたFIRフィルタ(16384点で打ち切り)をドライソースに畳み込んだ。ドライソースには200Hz~10kHzで帯域制限した白色雑音(5sec)、ガウス-マルコフ雑音(5sec)を使用した。

3. 符号化システムの概要

空間情報を利用して符号化する作業は、マルチチャンネル音響信号から音源に関する情報を抽出する作業に類似している。そこで、我々は音源分離技術を応用した符号化技術を用いた。具体的な符号化を以下に述べる。

(符号化手順)

各チャンネルの時間信号をフレームに区切る。 M チャンネルの信号 $x_i(n)(i=1, \dots, M)$ を N サンプルで区切ると、入力信号行列 \vec{x} は以下ようになる。

$$\vec{x}_i = [x_i(0) \quad \dots \quad x_i(N-1)]^T$$
$$\vec{x} = [\vec{x}_1 \quad \vec{x}_2 \quad \dots \quad \vec{x}_M]$$

区切った時間信号をチャンネルごとに周波数変換する。 \vec{x} を各列ベクトルごとに変換すると、

$$\vec{X}_i = \text{FT}\{\vec{x}_i\} = [X_i(0) \quad \dots \quad X_i(N-1)]^T$$
$$\vec{X} = [\vec{X}_1 \quad \vec{X}_2 \quad \dots \quad \vec{X}_M]$$

となり、各チャンネルの周波数成分が定まる。

逆畳み込みによって音源情報の周波数成分を抽出する。抽出する音源信号の周波数成分 $S_i(n)(i=1, \dots, M, n=0, \dots, N-1)$ は以下のようにして求められる。

$$S_i(n) = \frac{1}{M} \sum_{j=1}^M H_{ji}(n) X_j(n)$$

$$\vec{S}_i = [S_i(0) \quad \dots \quad S_i(N-1)]^T$$
$$\vec{S} = [\vec{S}_1 \quad \vec{S}_2 \quad \dots \quad \vec{S}_M]$$

$H_{ji}(n)$ は周波数 n におけるチャンネル j から音源 i までの伝達関数である。

音源情報の周波数成分をチャンネルごとに時間信号に戻す。周波数成分行列を各列ベクトルごとに逆周波数変換すると、

*Multi-channel audio coding using primary component analysis, By T. Kimura, K. Kakehi, K. Takeda and F. Itakura (Nagoya University)

$$\bar{s}_i = \text{IFT}\{\bar{S}_i\} = [s_i(0) \ \cdots \ s_i(N-1)]^T$$

$$\bar{s} = [\bar{s}_1 \ \bar{s}_2 \ \cdots \ \bar{s}_M]$$

となる。このとき、特定のチャンネルに情報が集約されるなら、そのチャンネルのみ送信することによって情報の圧縮が図られる。

復号化は符号化の手順を逆行することによって遂行される。一般に、復号化はドライソースにインパルス応答を畳み込むことに等しい。

4. 実験

2. で得られたマルチチャンネル音響信号の符号化及び復号化の実験を行った。フレーム長は 23ms (=1024sample)、フレーム間隔は 12ms (=512sample) で行った。窓関数には sin 窓を用いた。復号化に用いるインパルス応答は、2. で得られたインパルス応答の周波数応答を 1024 点に間引きし、200Hz ~ 10kHz に帯域制限したものをを用いた。符号化に用いる伝達関数はインパルス応答の逆フィルタに因果律を満たすために 512 点の遅延をかけたものをを用いた。そのようにして 1 チャンネルの音源情報を抽出し、復号化を行った。

5. 実験結果及び考察

得られた出力信号と入力信号を両信号の振幅成分の SN 比によって評価する。チャンネル i における SN 比(SNR) $_i$ を以下の式によって算出する。

$$(\text{SNR})_i = 10 \log_{10} \left[\frac{\sum_{\omega} \{ |X_i(\omega)| \}^2}{\sum_{\omega} \{ |X_i(\omega)| - |Y_i(\omega)| \}^2} \right]$$

$|X_i(\omega)|, |Y_i(\omega)|$ はそれぞれチャンネル i の入力信号と出力信号の 200Hz ~ 10kHz での振幅成分である。各チャンネルで符号化の際に生じる遅延は 512 点であるので、入力信号は 512 点シフトしたものをを用いた。フレーム長を 23ms (= 1024sample)、フレーム間隔を 12ms (=512sample)、窓関数を Hamming 窓にして分析を行い、方位角ごとにプロットしたものを図 2 に示す。

まず、方位角が 0°付近では SN 比が最も大きい。ゆえに、直接音の方向では音源が正しく再現されていることが分かる。0°から離れていくに従って SN 比は小さくなるが、実際はその先には音源は存在しないので、定位感に関しては影響を及ぼさないことが示唆される。

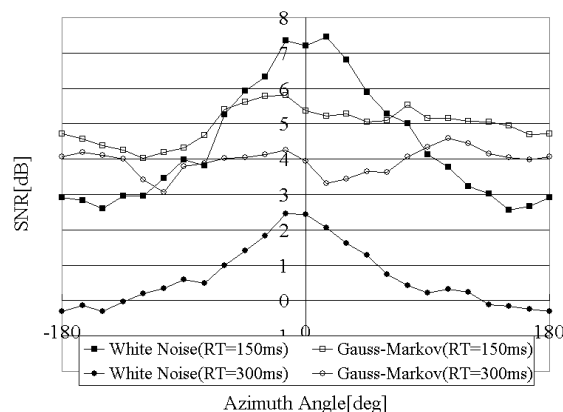


図 2 方位角ごとの SN 比

次に、残響時間が長くなると SN 比が低くなっている。これは一方向のみの音源情報を抽出したことによって他の方向からの残響による音源情報を無視したために再現性が劣化したと考えられる。そのため、他の方向の音源情報も考慮すれば再現性は改善されると考えられる。

6. まとめ

音源分離技術を応用したマルチチャンネル音響符号化システムの構築を行った。また、SN 比による評価を行ったところ、直接音に関しては音源の再現性が高いことが分かった。

今後の展望としては、再生装置を構築して、実際に主観評価を行う必要がある。また、複数音源や移動音源の場合を検討する必要がある。

7. 参考文献

- [1] Craig C. Todd, Grant A. Davidson, Mark F. Davis, Louis D. Fielder, Brian D. Link, Steve Vernon, "AC-3: Flexible Perceptual Coding for Audio Transmission and Storage," Proceeding of the 96th AES Convention, 1994
- [2] ISO/IEC 13818-7, "Information technology - Generic coding of moving pictures and associated audio information - Part 7 Advanced Audio Coding"