

## 非負値行列分解と位相復元に基づくオーディオ符号化の多チャンネル化\*

☆劉必翔(筑波大), 澤田宏(日本電信電話(株)), 宮部滋樹, 山田武志, 牧野昭二(筑波大)

## 1 はじめに

携帯型 MP3 プレイヤーまたは携帯電話による音楽鑑賞や、ネットワーク通信による音楽および動画の配信の普及により、オーディオ符号化の重要性はますます高まっている。オーディオ符号化技術は、信号の再生品質を損ねない可逆圧縮であるロスレス符号化と、非可逆圧縮であるがより低ビットレートの圧縮が可能なロッシー符号化に大別される。通信・記憶容量が限られたデジタルオーディオ再生においてより頻繁に用いられるのは、ビットレートの面で有利な後者のロッシー符号化であり、目的に合わせてより多様な圧縮技術や、さらなる圧縮の効率化が求められている。MP3などで実用化されている従来のロッシー符号化の多くは、人間の聴覚で感知しにくい情報を削減することにより低ビットレート化を実現しているが、長年の研究の中で最適化が進み、すでにその改善は飽和しているといえる。従って更なるオーディオ符号化の発展のためには、全く新しい符号化の枠組が必要であると考えられる。

近年、パワースペクトログラムの大域的な性質を分析し、頻繁に繰り返されるスペクトルパターンを基底分解の形で分析する非負値行列分解 (NMF) [1] が注目されており、ブラインド音源分離 (BSS) [2]、多重音解析 [3]、音声認識のための特徴抽出 [4] など、様々な音響信号処理への応用が研究されている。この枠組は符号化においても有効であり、振幅スペクトルに現れる繰り返しパターンの冗長性を NMF により削減し [5]、位相情報は低ビットレートに量子化して Griffin-Lim 位相復元アルゴリズム [6] により量子化誤差を低減する [7] 枠組が取り組まれている。また、NMF で分解された基底は単一音源に起因する傾向があるという性質は BSS に利用されているが、文献 [5] において指摘されているように、BSS を部分的に達成する NMF の基底分解をオーディオ符号化に利用することは、MPEG-SAOC [8] で実用化に向けて取り組まれているような、ユーザがオブジェクトを操作可能なオーディオ形態のための符号化のために非常に有効である。

一方、オーディオ CD などの一般によく利用されるオーディオ信号はステレオ形式であることが多いため、多チャンネル信号を効率的に圧縮できる符号化方式が望ましい。そこで本稿では、NMF と位相復元に基づく符号化 [5, 7] を、効率的な多チャンネル符号化が実現できるものへと拡張する。まず、NMF による基底分解に多チャンネル NMF [2] を用いることにより、全てのチャンネルを一つの基底分解で分析し、多チャンネルオーディオの振幅情報を、単一チャンネルの場合と同程度のビットレートに削減する。次に、多チャンネル NMF で符号化されるチャンネル間位相差の情報を利用して、チャンネル間位相差を維持したまま単一チャンネルの位相を最適化して多チャンネル位相復元を行うための、Griffin-Lim アルゴリズムの多チャンネル拡張を提案する。実験により、多チャンネルの NMF と Griffin-Lim アルゴリズムに基づく提案手法は、従来手法をチャンネルごとに適用するよりも、高品質かつ低ビットレートな多チャンネルオーディオ符号化が実現で

きることを確認した。

## 2 問題設定

ここでは本稿の多チャンネルオーディオ符号化の問題設定を述べる。本稿の目的は以下の  $M$  チャンネルオーディオ信号  $x_m(t)$ ,  $m = 1, \dots, M$  の良い近似  $y_m(t)$  を効率的に表現できるロッシー符号化を行うことである。

$$y_m(t) \approx x_m(t) \quad (1)$$

ここで  $t$  は時間サンプルを表す。本稿では、 $x_m(t)$  の短時間フーリエ変換 (STFT) 表現  $x_{ij,m}$  を符号化により近似した信号  $y_{ij,m}$  を、振幅と偏角を別々に符号化した表現で求めることについて議論する。

$$y_{ij,m} = \bar{y}_{ij,m} \exp(j\hat{y}_{ij,m}) \quad (2)$$

ここで  $\bar{y}_{ij,m}$  は  $y_{ij,m}$  の絶対値、 $\hat{y}_{ij,m}$  は  $y_{ij,m}$  の偏角、 $i = 1, \dots, I$  は周波数インデックス、 $j = 1, \dots, J$  はフレームインデックスを表し、 $j = \sqrt{-1}$  である。

複素スペクトログラム  $x_{ij,m}$  を表現するためには、実部と虚部に分けて  $2IJM$  の要素数の連続値実数が必要になる。その近似表現  $y_{ij,m}$  を振幅  $\bar{y}_{ij,m}$  と位相  $\hat{y}_{ij,m}$  に分けて表現すること自体は、信号表現の効率化にはならず、それぞれ  $IJM$  だけの要素数の連続値が必要になる。しかし振幅と位相は全く違う性質を持つため、それぞれに効果的な符号化方法を使い分けることにより、効率の高い信号符号化が達成できると考えられる。

## 3 NMF と位相復元を用いた符号化

従来手法である NMF を用いた振幅の符号化 [5] と位相の符号化 [7] を概説し、これらを組み合わせてチャンネルごとに適用することについて議論する。Fig. 1 に、NMF による符号化の流れを示す。符号化器では、時間領域のモノラル入力信号を時間周波数領域に変換し、振幅スペクトログラムから NMF の推定モデルに基づいて更新した基底パラメータとアクティベーションパラメータを量子化する。また、位相スペクトログラムから各時間周波数スロットにおける位相を量子化する。復号器では、量子化された符号から再構成して得た基底とアクティベーションから振幅スペクトログラムを復元する。更に位相を符号から再構成して複素スペクトログラムを求め、Griffin-Lim アルゴリズムを用いて位相の量子化誤差を補償したのちに、逆短時間フーリエ変換によってモノラル出力信号を得る。多チャンネル信号の符号化は、これをチャンネルごとに繰り返すことにより達成できる。

## 3.1 NMF を用いた振幅の符号化

Nikunen ら [5] が提案した NMF を用いた振幅スペクトログラムの符号化について述べる。NMF を用いることにより、チャンネルごとに振幅スペクトル  $\bar{x}_{ij,m}$  を以下のように非負係数の積和で表される  $\bar{y}_{ij,m}$  により近似する。

$$\bar{x}_{ij,m} \approx \bar{y}_{ij,m} = \sum_{k=1}^K t_{ik,m} v_{kj,m} \quad (3)$$

\* Multi-channel extension of audio coding based on non-negative matrix factorization and phase restoration. by Bixiang Liu (University of Tsukuba), Hiroshi SAWADA (NTT), Shigeki MIYABE, Takeshi YAMADA and Shoji MAKINO (University of Tsukuba)

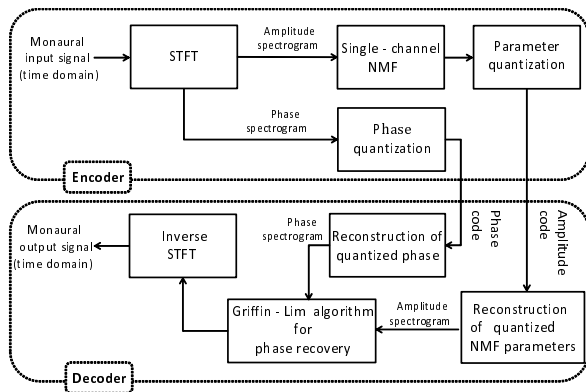


Fig. 1 Block diagram of coding scheme based on NMF.

ここで  $t_{ik,m}, v_{kj,m}$  はそれぞれ基底, アクティベーションとよばれる非負係数で, 基底  $t_{ik,m}$  は  $K$  個の代表的なスペクトルパターンを表し, アクティベーション  $v_{kj,m}$  によりフレーム毎に基底を別々に重みづけてスペクトログラムを表現する. この近似表現は, スペクトログラムが大域的に  $K$  個のパターンのみの重ね合わせでよく近似できるような, 繰り返しの多い信号に対して有効で,  $K \ll I$  かつ  $K \ll J$  という条件でこれらの係数を求めることにより, 要素数  $IJM$  の振幅スペクトル  $\bar{x}_{ij,m}$  を, はるかに小さい要素数  $K(I+J)M$  のパラメタ  $t_{ik,m}, v_{kj,m}$  で表すことができる. このうち要素数が  $KIM$  である基底  $t_{ik,m}$  は信号全体の符号化を通して使用されるため, 信号の時系列を表現する際のヘッダ情報として符号化することになる. それに対してアクティベーション  $t_{ik,m}$  は, 要素数  $KJM$  がフレーム数に依存し, 信号の時系列を表現するためにストリームとして扱うことになる. 振幅スペクトル  $\bar{x}_{ij,m}$  の要素数は  $IJM$  であり,  $K \ll I$  の条件で NMF を用いることにより, はるかに小さい要素数  $KJM$  のストリームに情報が縮約され, 低ビットレートで近似振幅スペクトログラム  $\bar{y}_{ij,m}$  を符号化することができる.

### 3.2 位相復元を用いた位相の量子化誤差低減

振幅スペクトル  $\bar{y}_{ij,m}$  は, NMF を用いた大域的な構造分析によって, 大幅に要素数を削減したパラメタで表現できることについて述べた. しかし位相  $\hat{y}_{ij,m}$  はフレームおよび周波数間の構造的な関係を仮定することはできず, 要素数を削減するのは困難である. そのため信号長 (フレーム数), 周波数ビン数とチャンネル数の全てに比例する要素数  $IJM$  の連続値である位相を効率的に量子化するためには, 非常に粗い精度での量子化が必要になる. 振幅スペクトルが得られた条件では, STFT 分析と逆 STFT 分析の繰り返しを用いた Griffin-Lim アルゴリズムにより, 振幅との整合性のとれた位相が推定できることが知られている. しかし, このアルゴリズムでは自然な位相が得られるものの, 符号化の目的である真の信号の近似とはならない. そこで, 量子化して誤差を含んだ位相を初期値として与え, Griffin-Lim アルゴリズムで修正することにより, 真値付近に収束する位相復元を行い, 精度の粗い量子化でも高い精度に復元する枠組みが提案されている [7]. アルゴリズムは以下のようになる.

1. 復元された振幅  $\bar{y}_{ij,m}$  に再構成した位相  $\hat{y}_{ij,m}$  を与え, 復号された近似複素スペクトログラム  $y_{ij,m}$  を求める.

$$y_{ij,m} \leftarrow \bar{y}_{ij,m} \exp(j\hat{y}_{ij,m}) \quad (4)$$

2. 複素スペクトログラム  $y_{ij,m}$  を逆 FFT, 合成窓

とオーバーラップアッドによる逆短時間フーリエ変換 (ISTFT) に施して時間領域に戻し, 更に短時間フーリエ変換 (STFT) に施す. これにより得られる複素スペクトログラムは元の  $y_{ij,m}$  と異なったものとなるが, その位相を新たな  $\hat{y}_{ij,m}$  とする.

$$\hat{y}_{ij,m} \leftarrow \arg[\text{STFT}[\text{ISTFT}[y_{ij,m}]]] \quad (5)$$

ここで  $\arg[\cdot]$  は引数の偏角を取り出す演算を表す.

3. 式 (4) により, Step. 2 で得られたスペクトログラムの位相  $\hat{y}_{ij,m}$  を振幅スペクトル  $\bar{y}_{ij,m}$  に与え, 複素スペクトログラム  $y_{ij,m}$  を求める.
4. Step. 2-3 を収束に十分な回数だけ繰り返す.

## 4 多チャンネル NMF を用いた符号化

ここでは前節で述べた NMF と位相復元に基づく従来の符号化手法を多チャンネルに拡張し, 全てのチャンネルの信号を単一チャンネルのパラメタに縮約することによる, 効率的な多チャンネルオーディオ符号化の枠組を提案する. まずこの枠組で用いる多チャンネル NMF [2] について述べ, 続いて効率的な多チャンネル位相復元を実現するための, Griffin-Lim アルゴリズムの多チャンネル拡張を提案する.

### 4.1 多チャンネル NMF に基づく符号化のパラメタ

Ozerov and Fevotte [9] によって最初に提案された多チャンネル NMF は, マイクロホンアレーを用いた BSS の振幅を NMF 基底により近似するものであり, 多くの BSS アルゴリズムで問題となるパーミュテーションが発生しない利点がある. Sawada *et al.* [2] は, 多チャンネル NMF をチャンネル間共分散行列の基底による近似表現として再定式化した. 特に多次元に拡張した板倉-斉藤距離尺度を用いたものは, 多変量複素正規分布の最尤推定として自然な確率モデルで定式化され, かつ高性能な BSS を達成できる. 本稿では Sawada 法の共分散行列を近似する枠組を用いて, NMF により振幅とチャンネル間位相差の両方を符号化するアルゴリズムを提案する.

Sawada 法では  $M$  チャンネルの複素スペクトログラム  $\mathbf{x}_{ij} = [x_{ij,1}, \dots, x_{ij,M}]^T$  の直積  $\mathbf{X}_{ij} = \mathbf{x}_{ij}\mathbf{x}_{ij}^H$  を, 以下のように  $L$  個の音源のパワーを表すための  $K$  個の基底の線形和で表した共分散行列  $\mathbf{Y}_{ij}$  として近似する.

$$\begin{aligned} \mathbf{X}_{ij} &\approx \mathbf{Y}_{ij} \\ &= \sum_{k=1}^K \left( \sum_{l=1}^L z_{lk} \mathbf{H}_{il} \right) t_{ik} v_{kj} \end{aligned} \quad (6)$$

ここで  $\{\cdot\}^T$  は転置,  $\{\cdot\}^H$  は複素共役転置,  $0 \leq z_{lk} \leq 1$  は第  $k$  番目の基底の第  $l$  番目の音源への所属を表す係数,  $\mathbf{H}_{il}$  は第  $i$  周波数における第  $l$  音源の空間相関を表す非負定値行列,  $t_{ik}$  および  $v_{kj}$  はそれぞれ NMF の基底とアクティベーションに相当する非負係数であり, 音源数  $L$  は既知であるものとする. この定式化は, 各  $l$  番目の音源の周波数ビン  $i$ , フレーム  $j$  における共分散行列が, 空間相関行列  $\mathbf{H}_{il}$  の基底とアクティベーションによる重み付けで表され, それらの和が観測信号の共分散行列  $\mathbf{Y}_{ij}$  になるというモデル化に相当する.

以上のように少ないパラメタに圧縮して表現される共分散行列  $\mathbf{Y}_{ij}$  から多チャンネル複素スペクトログラム  $\mathbf{x}_{ij}$  の近似  $\mathbf{y}_{ij} = [y_{ij,1}, \dots, y_{ij,M}]^T$  を復元するためには, 直積によって失われる 1 チャンネルのみの位相が必要となる. 二乗誤差を基準とすると, 最適な復元信号  $\mathbf{y}_{ij}$  は,  $\mathbf{y}_{ij}\mathbf{y}_{ij}^H$  が  $\mathbf{Y}_{ij}$  の最も良い近似と

Table 1 Quantization of NMF parameters

Parameter	Bit	Elements	Data
$t_{ik}$	16bit	$IK$	15390
$v_{kj}$	16bit	$KJ$	45870
$z_{lk}$	16bit	$KL$	90
$\mathbf{H}_{il}^{(\text{Re})}$	16bit	$IL \times \frac{M(M+1)}{2}$	4617
$\mathbf{H}_{il}^{(\text{Im})}$	16bit	$IL \times \frac{M(M-1)}{2}$	1539

$I$ : number of frequency bin.  $J$ : number of time frame.  
 $K$ : number of basis.  $L$ : number of signal.  $M$ : number of channel.

なる,  $\mathbf{Y}_{ij}$  の第一主成分ベクトルである. しかし主成分ベクトルには位相の任意性が残るため, 主成分ベクトル  $\tilde{\mathbf{y}}_{ij} = [\tilde{y}_{ij,1}, \dots, \tilde{y}_{ij,M}]^T$  を  $\arg[\tilde{y}_{ij,1}] = 0$  となるように求めた上で, 第1チャンネルの位相の近似値  $\hat{y}_{ij,1} \approx \arg[x_{ij,1}]$  を用いることにより,

$$\mathbf{y}_{ij} = \tilde{\mathbf{y}}_{ij} \exp(j\hat{y}_{ij,1}) \quad (7)$$

のように信号復元を行うことができる.

以上より, 多チャンネル NMF を用いて多チャンネル複素スペクトログラム  $\mathbf{x}_{ij}$  の近似  $\mathbf{y}_{ij}$  を表現するためのパラメータは, 複素 NMF のパラメータと単一チャンネルの位相  $\hat{y}_{ij,1}$  である. それらのうち, 要素数がフレーム数  $J$  に比例するストリーム情報はアクティベーション  $v_{kj}$  と第1チャンネルの位相  $\hat{y}_{ij,1}$  のみである. 単一チャンネル NMF に基づく符号化をチャンネルごとに行う場合は, ストリーム情報であるアクティベーション  $v_{kj,m}$  と位相  $\hat{y}_{ij,m}$  がチャンネル  $m = 1, \dots, M$  ごとに必要になるのに対し, 多チャンネル NMF では全てのチャンネルを表現するためのストリーム情報が, 単一チャンネルの場合と同等のパラメータ  $v_{kj}, \hat{y}_{ij,1}$  に縮約される. したがって, チャンネルごとの NMF 符号化に比べてヘッダ情報は要素数が少しくなるもの, ストリーム情報の要素数が  $1/M$  に減少し, 全チャンネルを符号化するためのビットレートを大幅に削減することができる.

## 4.2 パラメータの量子化

符号化の観点からは, 聴感上の音質の劣化を抑えながらできるだけ少ないビット和の割り当てで量子化することが求められる. したがって, パラメータ  $t_{ik}, v_{kj}, z_{lk}, \mathbf{H}_{il}$  の量子化する方法を適切に定める必要がある.

NMF パラメータの量子化では, 観測  $\mathbf{x}_{ij}$  から推定されたパラメータ  $t_{ik}, v_{kj}$  の最大値よりも2のべき乗が大きな値を取るような最小のビット数を選択してパラメータを量子化する. 他のパラメータ  $z_{lk}, \mathbf{H}_{il}$  は, 集合  $\{t_{ik}, v_{kj}\}$  の最大値を基準として定めた同一のビット数を用いて量子化する. そして,  $t_{ik}, v_{kj}, z_{lk}$  はそのまま量子化を行い, エルミート行列である  $\mathbf{H}_{il}$  は, 下三角要素の実部  $\mathbf{H}_{il}^{(\text{Re})}$  と, 対角要素を除いた下三角要素の虚部  $\mathbf{H}_{il}^{(\text{Im})}$  に分解し, 行列要素一つ一つについてこのビット数での量子化を行う.

位相  $\hat{y}_{ij,1}$  については, 第1チャンネルの位相  $\arg[x_{ij,1}]$  が一様分布に従う情報源であるとみなすことができるため, 任意のビット数を設定して,  $[-\pi, \pi]$  の範囲の等間隔量子化を適用する.

各符号の伝送では, 時間フレーム  $j$  によって変化しない基底パラメータ  $t_{ik}$ , クラスターインデックス  $z_{lk}$ , 音源の空間相関行列  $\mathbf{H}_{il}$  はヘッダ部として伝送し, アクティベーションパラメータ  $v_{kj}$  と位相  $\hat{y}_{ij,1}$  はフレーム毎に量子化してストリーム部として伝送する.

Table 1 に多チャンネル NMF のパラメータと位相の量子化ビット割り当てと, ビット数の例を示す.  $I = 513$ ,  $J = 1529$ ,  $K = 30$ ,  $L = 3$ ,  $M = 2$ , サンプリング周波数 16 kHz でオーバーラップが  $1/4$  フレームとなる分析でフレーム数が 62.5 という条件を例にとると, ヘッダ情報を除いたストリームであるアクティベーション  $v_{kj}$  と位相  $\hat{y}_{ij,1}$  の符号化のビットレートは, 約 286.5 kbps となる.

## 4.3 復号アルゴリズム

### 4.3.1 振幅・チャンネル間位相差の再構成

復号器では, まず NMF パラメータの符号から, 振幅とチャンネル間位相差を保持した複素スペクトログラム  $\tilde{\mathbf{y}}_{ij}$  を復元する. 量子化された NMF パラメータの符号  $t_{ik}, v_{kj}, z_{lk}, \mathbf{H}_{il}$  を再構成し, 式 (4) を用いて各時間周波数スロットの空間共分散行列の近似  $\mathbf{Y}_{ij}$  を復元する. さらにその第一主成分ベクトル  $\tilde{\mathbf{y}}_{ij}$  を, その第一要素の位相が 0 となる形で求める.

### 4.3.2 多チャンネル Griffin-Lim アルゴリズムによる信号復元

第1チャンネルの位相  $\hat{y}_{ij,1}$  は全ての時間周波数  $i, j$  に対して値を保持する必要があるため, 低ビットレート化のために量子化精度を低く設定する必要がある. そのため 4.3.1 で復元された位相は量子化誤差の影響を強く受けてしまう. ここでは Sato *et al.* によって提案された Griffin-Lim アルゴリズムによる位相の量子化誤差低減 [7] を, チャンネル間の位相差を固定した形で多チャンネルに拡張したものを提案する. アルゴリズムは以下ようになる.

1. 復元された第一主成分ベクトル  $\tilde{\mathbf{y}}_{ij}$  に, 再構成した第1チャンネルの位相  $\hat{y}_{ij,1}$  を与え, 復号された近似複素スペクトログラム  $\mathbf{y}_{ij}$  を求める.

$$\mathbf{y}_{ij} \leftarrow \tilde{\mathbf{y}}_{ij} \exp(j\hat{y}_{ij,1}) \quad (8)$$

2. 多チャンネル複素スペクトログラム  $\mathbf{y}_{ij}$  を ISTFT に施して時間領域に戻し, 再度 STFT 分析により複素スペクトログラムに変換して  $\mathbf{y}_{ij}$  を更新する. 更新により得られる  $\mathbf{y}_{ij}$  は振幅と位相の両方が元のものから変化するが, 新たな  $\mathbf{y}_{ij}$  への  $\tilde{\mathbf{y}}_{ij}$  の射影の位相を新たな第1チャンネルの位相  $\hat{y}_{ij,1}$  とする.

$$\mathbf{y}_{ij} \leftarrow \text{STFT}[\text{ISTFT}[\mathbf{y}_{ij}]] \quad (9)$$

$$\hat{y}_{ij,1} \leftarrow \arg[\tilde{\mathbf{y}}_{ij}^H \mathbf{y}_{ij}] \quad (10)$$

3. 多チャンネル振幅スペクトログラム  $\mathbf{y}_{ij}$  に Step. 2 で得られた位相  $\hat{y}_{ij,1}$  を式 (8) により与える.
4. Step. 2-3 を収束に十分な回数だけ繰り返す.

## 4.4 提案手法のアルゴリズムのまとめ

Fig 2 に, 多チャンネル NMF による符号化の流れを示す. 符号化器では, 時間領域の多チャンネル入力信号を短時間フーリエ変換し, 複素スペクトログラムを多チャンネル NMF 表現により要素数削減したのち, 多チャンネル NMF モデルで規定したパラメータにより量子化する. 第1チャンネルの位相はそのまま量子化する. 復号器では, NMF パラメータを再構成し, 式 (6) の多チャンネル NMF モデルに従って共分散行列  $\mathbf{Y}_{ij}$  を復元し, 第一主成分ベクトル  $\tilde{\mathbf{y}}_{ij}$  を求める. 位相符号を再構成し, 多チャンネル Griffin-Lim アルゴリズムによる位相復元を行って, 逆短時間フーリエ変換により多チャンネル出力信号を得る.

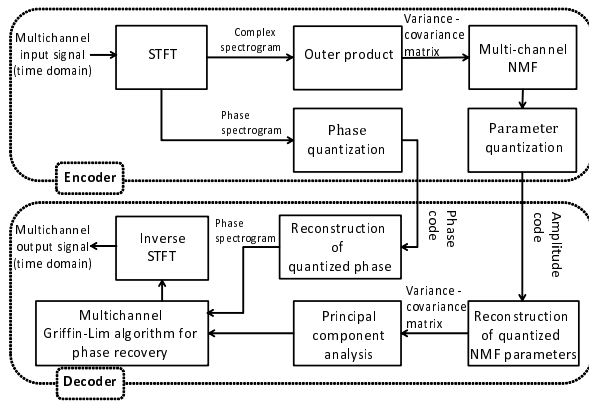


Fig. 2 Block diagram of coding scheme based on Multichannel NMF.

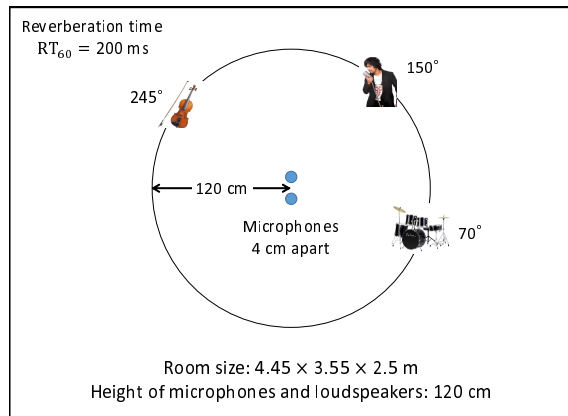


Fig. 3 Environment of signal observation.

## 5 実験

本章では、チャンネルごとに符号化する従来手法と、全てのチャンネルを同時に符号化する提案手法を実験により比較評価する。

### 5.1 実験条件

提案手法を評価するために、ステレオのオーディオデータを対象とする評価実験を行った。単一チャンネル、または多チャンネルのNMFに基づき、対象のステレオ信号を7種のビットレートで符号化して、PEAQ [10]により客観音質評価した。NMFのパラメタの更新は100回とし、更新後のパラメタは16ビットで量子化した。NMFの距離尺度は、単一チャンネルと多チャンネルのもの両方において板倉-斉藤距離を用いた。位相は、ビット数2, ..., 8で等間隔量子化し、Griffin-Limアルゴリズムの繰り返し回数は200回とした。符号化の対象とするステレオ信号は、SiSEC研究用音楽データベース [11]から選んだ男性ボーカル、ドラム、バイオリンの3つの音源をFig. 3に示す部屋で畳み込み混合し、16kHzにリサンプリングしたものを用いた。短時間フーリエ変換の窓関数および、Griffin-Lim手法における合成窓関数、分析窓関数はハンニング窓を用い、フレーム長は1024、フレームシフト256サンプルとした。なお、評価尺度である客観音質劣化度 Objective Difference Grade (ODG) スコアは、-4から0の範囲の値を取り、0に近いほど音質の劣化が少ない。

### 5.2 実験結果

Fig. 4にそれぞれのビットレートで符号化を行って得られたODGスコアを示す。位相量子化に同一の

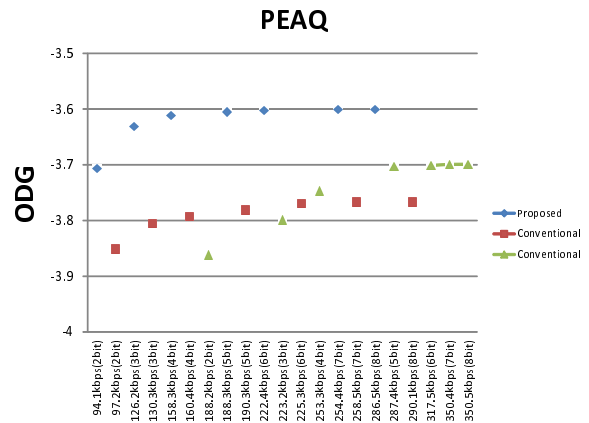


Fig. 4 PEAQ ODG scores.

ビット数を用いた場合、多チャンネルNMFを用いた提案手法は、ビットレートがチャンネルごとに符号化する従来手法よりも低いにもかかわらず、高いODGスコアとなっていることが確認できる。これにより、NMFと位相復元を多チャンネル化した提案手法が、NMFと位相復元をチャンネルごとに用いる提案手法よりも効率的な多チャンネル符号化を達成できることが確認された。

## 6 おわりに

本稿では、NMFと位相復元を用いたオーディオ符号化の、NMFと位相復元の両方を多チャンネル信号を扱えるものに拡張し、より効率的に多チャンネルオーディオ信号を符号化する手法を提案した。実験により、提案手法は従来のNMFと位相復元に基づく符号化をチャンネルごとに余起用するよりも、低ビットレートかつ高品質に多チャンネルオーディオ信号を符号化できることを確認した。

今後の課題はNMFに基づくオーディオ符号化の品質を底上げすることである。現状のこの枠組は、MP3などで実用化されている修正コサイン変換に基づく符号化よりも品質が悪く、提案した多チャンネル化もまた同様の問題点を抱えている。NMFを用いたオーディオ符号化の研究はまだ歴史が浅く、今後研究が進んで改善されていくものと期待される。

## 参考文献

- [1] Lee and Seung, *Nature*, 401, 788-791, 1999.
- [2] Sawada *et al.*, *IEEE Trans. ASLP*, 21(5), 971-982, 2013.
- [3] Nakano *et al.*, *Proc. MLSP*, 283-288, 2010.
- [4] Schuller *et al.*, *Proc. ICASSP*, 4562-4565, 2010.
- [5] Nikunen *et al.*, *Proc. AES Conv.*, 8083, 2010.
- [6] Griffin and Lim, *IEEE Trans. ASSP*, 32(2), 236-243, 1984.
- [7] Sato *et al.*, *Proc. DAFx*, 4 pages, 2012.
- [8] Breebaart *et al.*, *Proc. AES Conv.*, 124, 2008.
- [9] Ozerov and Fevotte, *IEEE Trans. ASLP*, 18(3), 550-563, 2010.
- [10] ITU-R Recommendation BS.1387-1.
- [11] Araki *et al.*, *Proc. LVA/ICA*, 414-422, 2012.