スペクトログラムの振幅・位相量子化と範囲制限位相復元に基づく音響信号符号化の検討*

☆佐藤 匠 (東大院・情報理工)，小野 順貴 (NII)，
鎌本 優 (NTT・CS 研)，嵯峨山 茂樹 (東大院・情報理工)

1 はじめに

MP3 や AAC などの音響信号の非可逆圧縮技術は，携帯音楽機器で大量の音楽を楽しむために必要不可欠な重要な技術である．これら最新の符号化技術は，修正離散コサイン変換と聴覚モデルに基づいて効率的な圧縮を行っているが，すでに多くのパラメータチューニングや最適化が行われ，性能は限界に達しつつあるとも考えられる．

これに対し我々は，短時間 Fourier 変換（STFT）の振幅・位相を符号化する新しいアプローチを試みている．この方式の狙いは，1) 位相復元技術 [1][2] によって位相へ割り当てるビット数を大幅に減らし，2) 振幅（スペクトログラム）の冗長性を利用してこれを効率的に符号化することにある．特に 2) に関しては，スペクトログラムの非負性を積極的に利用し，非負値分解行列（NMF）のような最新の信号処理技術を導入することを考えている．

これまで我々は最も極端な場合として，位相を全く符号化しない方式を検討したが，音質の低下は無視できないことがわかった [3][4]．よって本研究では，1) ビットレートを固定した場合の，振幅・位相への最適なビット割当と，2) 量子化された振幅・位相からの高精度な波形復元方法，について検討した結果を報告する．

2 本研究での検討事項

2.1 符号化手順

スペクトログラムに基づく音響信号符号化の手順を図 1 に示す．符号器では，入力音響信号（時間領域）を STFT し，スペクトログラムの振幅・位相を量子化する．復号器では，それらを逆量子化し，逆 STFT によって出力音響信号（時間領域）を得る．本研究では，スペクトログラムの振幅・位相の量子化方法に焦点を絞るため，エントロピー符号化は行わずにそのままの情報をビットストリームとして出力することとする．

2.2 振幅・位相スペクトログラムの量子化

n サンプルの音響信号 $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ から STFT により複素スペクトログラム $\boldsymbol{Y} = \{Y_{km}\}$ が得られる．ただし，k, m はそれぞれ時間，周波数のインデックスである．対数振幅 A_{km}，位相 ϕ_{km}(-π〜π) を用いて Y_{km} は式 (1) のように表わせる．

$$Y_{km} = \exp(A_{km} + j\phi_{km}) \tag{1}$$

本枠組みにおいて大きなポイントの 1 つは，複素スペクトログラムの振幅位相をどのように量子化するか，という点である．我々は、1) 人間の聴覚の音の大きさの知覚がほぼ対数的である，2) 位相は一般に一様分布である，ということから，対数振幅，位相に対して，最も単純な等間隔量子化を適用することを検討している．これは以下のように表される．

$$\hat{A}_{km} = \left\lfloor \frac{A_{km} - \bar{A}_k}{\Delta_{km}} + \frac{1}{2} \cdot 2^{a_{km}} \right\rfloor \tag{2}$$

$$\hat{\phi}_{km} = \left\lfloor \frac{\phi_{km}}{\pi/2^{p_{km}-1}} + \frac{1}{2} \cdot 2^{p_{km}} \right\rfloor \tag{3}$$

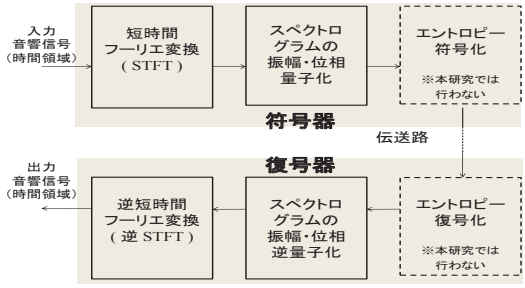

Fig. 1 スペクトログラムに基づく音響信号符号化の基本となるフローチャート

もし，式 (2) の $\hat{A}_{km}$ が 0 〜 $2^{a_{km}} - 1$ の範囲を超えたらクリッピングさせる．ここで，$\hat{A}_{km}, \hat{\phi}_{km}$ は量子化（整数化）された対数振幅と位相であり，$\bar{A}_k$ は k に関する A の項の平均である．$\lfloor \cdot \rfloor$ は床関数を表し，Δ_{km} は対数振幅の量子化ステップサイズである．また，a_{km} と p_{km} はそれぞれ時間周波数成分 km の対数振幅と位相に割り当てるビット数である．本研究で検討する課題の 1 つは，$a_{km} + p_{km} = C$ のように，固定ビットレートを前提とした際、a_{km}，p_{km} にどのようにビットを割り当てるのがよいか，という点である．人間の聴覚は位相に鈍感であること [5] より，$a_{km} \geq p_{km}$ が適当であると予想できる．

2.3 無矛盾性に基づく振幅・位相の逆量子化

対数振幅と位相は簡単のため次のように逆量子化する．

$$A'_{km} = \left(\hat{A}_{km} - \frac{1}{2} \cdot 2^{a_{km}} + \frac{1}{2}\right) \cdot \Delta_{km} + \bar{A}_k \tag{4}$$

$$\phi'_{km} = \left(\hat{\phi}_{km} - \frac{1}{2} \cdot 2^{a_{km}} + \frac{1}{2}\right) \cdot \frac{\pi}{2^{p_{km}-1}} \tag{5}$$

A'_{km}，ϕ'_{km} はそれぞれ逆量子化後の対数振幅と位相である．

本研究では，STFT のフレームのオーバーラップによる冗長性から得られたスペクトログラムの振幅と位相はそれぞれ独立したものではないということに着目し位相復元 [1] という技術を用いることにした．これは，STFT と逆 STFT を繰り返すことによって振幅スペクトログラムから出来るだけ矛盾しない位相スペクトログラムを求める技術である．本研究では，振幅と位相の無矛盾性を利用して復元された音質の向上を目指す．それを利用するためのひとつの方法として，式 (4) に式 (5) を初期値として用いた位相復元を行うことが考えられる．しかしこの場合，反復的に推定された位相が初期値とかけ離れた値になってしまうことも予想される．その場合には，量子化幅内に位相を引き戻す操作を加えることで推定精度の向上を図る．つまり，位相は制限された範囲の中で反復的に推定する．

最後に，範囲制限位相復元の有効性を検証するために 3 つの波形復元手法についての比較を行った．式 (4)，(5) より復号化された複素スペクトログラム $\boldsymbol{Y'}$ は式 (6) により求められ，以下にそれぞれの波形復元のアルゴリズムを示す．

---

*Spectrogram-based Audio Coding with Amplitude/Phase Quantization and Range-constrained Phase Reconstruction. by Sho SATO (The University of Tokyo), Nobutaka ONO (National Institute of Informatics), Yutaka KAMAMOTO (NTT Communication Science Laboratories) and Shigeki SAGAYAMA (The University of Tokyo)

$$Y'_{km} = \exp(A'_{km} + j\phi'_{km}) \quad (6)$$

ベースライン

$$\boldsymbol{x} = \mathrm{invSTFT}(\boldsymbol{Y}') \quad (7)$$

位相復元

逆量子化後の値を初期値とする ($\boldsymbol{Y} = \boldsymbol{Y}'$)．次の更新式は順次反復的に行われる．

$$\begin{aligned} \boldsymbol{x} &= \mathrm{invSTFT}(\boldsymbol{Y}) & (8)\\ \boldsymbol{Y} &= \mathrm{STFT}(\boldsymbol{x}) & (9)\\ \phi_{km} &= \angle Y_{km} & (10) \end{aligned}$$

範囲制限位相復元

逆量子化後の値を初期値とする ($\boldsymbol{Y} = \boldsymbol{Y}'$)．次の更新式は順次反復的に行われる．

$$\begin{aligned} \boldsymbol{x} &= \mathrm{invSTFT}(\boldsymbol{Y}) & (11)\\ \boldsymbol{Y} &= \mathrm{STFT}(\boldsymbol{x}) & (12) \end{aligned}$$

$$\phi_{km} = \begin{cases} \phi^{(u)}_{km} & (\angle Y_{km} > \phi^{(u)}_{km}) \\ \phi^{(l)}_{km} & (\angle Y_{km} < \phi^{(l)}_{km}) \\ \angle Y_{km} & (\text{otherwise}) \end{cases} \quad (13)$$

ここで

$$\phi^{(u)}_{km} = \left(\hat{\phi}_{km} - \frac{1}{2}\cdot 2^{a_{km}} + 1\right)\cdot \frac{\pi}{2^{p_{km}-1}} \quad (14)$$

$$\phi^{(l)}_{km} = \left(\hat{\phi}_{km} - \frac{1}{2}\cdot 2^{a_{km}}\right)\cdot \frac{\pi}{2^{p_{km}-1}} \quad (15)$$

3 客観評価実験

3.1 実験条件

本実験では，振幅と位相を合わせて 8 ビットで量子化するすることを考え，ビットレートを 128 kbps に固定した．表 1 は 8 ビットを振幅と位相に割り振る時に考えられる条件を示したものである．RWC 研究用音楽データベースからポップス，クラシックを，ATR 音声データベースから男声，女声を各 5 つずつ，計 20 個選び，16 kHz のモノラル信号に変換したものを用いた．STFT はハミング窓を用い，フレーム長 1024 点，ハーフオーバーラップで行った．ここでは，$\Delta_{km} = \alpha\sigma_k/2^{a_{km}-1}$, $\alpha = 3$ とし，位相復元回数は 200 回とした．σ_k は k に関する A の項の標準偏差である．音質の客観評価として PEAQ [6] を用い，それによって得られる客観品質劣化度合 ODG（Objective Difference Grade）のスコアは - 4 〜 0 の範囲の値となり，0 に近いほど劣化が少ない．

3.2 実験 1: 3 つの位相復元手法を用いた振幅位相の最適なビット割り当て

この実験は，振幅と位相への適切なビット割り当て比と波形合成手法を調べることが目的である．図 2 は表 1 のように 8 ビットをさまざまな比で振幅と位相に割り振ったときの ODG のスコアを求め，各ジャンルごとの平均値をプロットしたグラフである．この場合，(振幅, 位相) = (6 ビット, 2 ビット) のように割り振り，範囲制限位相復元を用いた時が最も適切であることが分かる．

Table 1 振幅，位相への 8 ビット割り当て方法

振幅 [bit]	0	1	2	3	4	5	6	7	8
位相 [bit]	8	7	6	5	4	3	2	1	0

Table 2 位相に平均 2 ビット割り当てる方法（振幅は 6 ビットに固定）

振幅 [bit]	6	6	6	6	6	6
位相 [bit]	2	5	8	10	20	40
0 ビットでない位相の割合 [%]	100	40	25	20	10	5

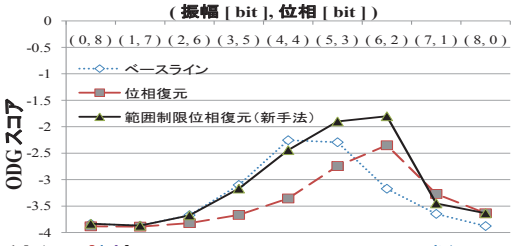

Fig. 2 PEAQ の ODG スコア (3 つの各位相復元手法を全曲［5 信号× 4 ジャンル］の平均値で比較)

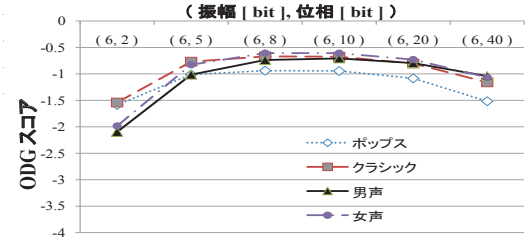

Fig. 3 PEAQ の ODG スコア (パワーに基づき量子化された位相に範囲制限位相復元を適用［振幅 : 6 bits, 位相 : 平均 2 bits］, それぞれのジャンルごとに 5 信号を用いて平均値で比較)

3.3 実験 2: パワーに基づく位相量子化

パワーの小さい時間周波数成分は，人間の聴覚には聴こえないので位相は無視することができる．したがって，パワーの小さい位相に対しては 0 ビットを割り当てても音質の劣化は小さいと考えられる．表 2 は振幅に 6 ビット割り当て，位相に平均 2 ビット割り当てるときの位相のさまざまなビット割り当ての方法示したものである．例えば，パワーが大きい上位 40 % の位相に対して 5 ビット，残りの 60 %に 0 ビットを割り当てる場合などが平均 2 ビットに相当する．図 3 は実験結果であり，(6 ビット, [上位 25 % に］8 ビット)，(6 ビット, [上位 20 % に］10 ビット) としたときに PEAQ の ODG スコアが最も高い．

4 おわりに

本研究では，スペクトログラムの振幅・位相量子化による符号化と範囲制限位相復元を用いた復号化について検討した．そして，8 ビットによる量子化を考える時には，(振幅, 位相) = (6 ビット, 2 ビット) で割り振り，量子化誤差の範囲内で位相を反復推定する範囲制限位相復元を用いた場合に最も PEAQ の ODG スコアが高くなることを確認した．また，位相の平均ビット数を保ちつつ，パワーの小さい位相を無視してパワーの大きい位相に平均ビット数以上を割り当てることでさらに PEAQ の ODG スコアが改善されることを確認した．今後は，聴覚心理モデルを用いること，効率的なコードブックを作成するために最新の信号処理技術（NMF など）を用いることでさらなる圧縮を図っていきたい．

謝辞 本研究は，日本学術振興会科学研究費補助金 (挑戦的萌芽研究：23650083) の助成を受けたものである．

参考文献

[1] D. W. Griffin and J. S. Lim, ASSP, vol. 32(2), pp. 236–243, 1984.
[2] J. Le Roux *et al.*, SAPA, Sep. 2008.
[3] 藤田 他，音講論 (秋)，pp. 1391–1392，2010.
[4] 佐藤 他，音講論 (春)，pp. 337–338，2011.
[5] M. Bosi and R. E. Goldberg, “Introduction to digital audio coding and standards, ” Kluwer Academic Publisher, 2003
[6] ITU-R Recommendation BS.1387-1