

## 音源分離技術を用いた

### Segmental Intensity Expanding (Sinex) 符号化方式

岩上直樹 守谷健弘 神明夫 森岳至 千喜良和明

(NTT研究所)

#### 概要

楽音符号化は、通信・放送・蓄積などの用途のために、デジタル・オーディオ信号の情報量を音質を劣化させずに圧縮する技術である。本報告では、新たに考案した音源分離技術 Foreground / Background Categorization (FBC) を用いた SinexAudio 符号化 (Segmental Intensity Expanding AUDIO coding) 方式を提案する。SinexAudio は、ベクトル量子化を利用した変換符号化方式で、このタイプの符号化方式に必須である平坦化操作のために、従来法である線形予測スペクトルとバークスケール・エンベロープに加えて FBC を導入する。このことにより、平坦化の能率が向上し、音質が改善する。SinexAudio と同タイプの代表的な従来法 TwinVQ とを比較評価し、提案方式は優れた結果を得た。

## Segmental Intensity Expanding (Sinex) Audio Coding using a Signal Categorization Technique

Naoki Iwakami, Takehiro Moriya, Akio Jin, Takeshi Mori, and Kazuaki Chikira  
(NTT Laboratories)

#### Abstract

Audio coding is a technique to compress digital audio signal without degradation of sound quality. In this paper a new coding method SinexAudio coding (Segmental Intensity Expanding AUDIO coding) is introduced. The SinexAudio includes a new technique called Foreground / Background Categorization (FBC). SinexAudio is a transform coder using MDCT with vector quantization at the bottom of the data flow. The FBC is used to flatten MDCT coefficients as well as conventional LPC spectrum and Bark-scale envelope method. This new technique works efficiently for this quality-determining procedure. Quality of SinexAudio reproduction was tested comparing with a typical transform/VQ coder, TwinVQ. The result showed superiority of SinexAudio.

#### 1. はじめに

放送・蓄積・ネットワーク音楽配信などのデジタル・オーディオでは、楽音の情報量を圧

縮することは伝送時間短縮、蓄積容量低減、コスト低減などのさまざまなメリットがある。特にネットワークオーディオの分野では、音楽デ

ータをフラッシュメモリに蓄積し、再生する携帯型の音楽再生装置が多く使われるが、高価なフラッシュメモリを有効に利用するために楽音の情報量を圧縮する意義は大きい。このようなことからオーディオ信号を、高音質を保ったまま圧縮する楽音符号化方式の利用価値は高く、さまざまな方式が提案されている。

楽音符号化の多くは、サブバンド化または周波数領域に変換された入力信号を量子化する方法をとっている[1]-[5]。量子化方法としてベクトル量子化を用いると、高圧縮率でも高い性能を得ることができる。TwinVQ（周波数重み付けインタリーブベクトル量子化, Transform-domain Weighted Interleave Vector Quantization）[6]はその一例であり、変換方法として変形離散コサイン変換(modified discrete cosine transformation, MDCT)[7]、量子化方法として、重み付けインタリーブベクトル量子化を行っている。また、LPC スペクトル平坦化とバークスケール・エンベロップ平坦化を用いてベクトル量子化の能率を上げていることが特徴である。

TwinVQ 方式は、特に圧縮率の高い領域でも高音質な優れた方式であるが、特殊な特性の入力音に対して音質を確保しにくい欠点を持つ。

この報告では、音源の特性の変化に良好に対応する SinexAudio (Segmental INTensity EXpanding AUDIO coding) 方式を提案する。

次節以降では、TwinVQ 楽音符号化方式の問題点を述べた後、その問題点の解決方法を示し、新方法を適用した SinexAudio 符号化方式の概要を述べる。最後に音質の評価について述べる。

## 2. ベクトル量子化タイプの変換符号化方式における従来法の問題点

### 2.1 TwinVQ 方式の概要

代表的なベクトル量子化タイプ変換符号化方式である TwinVQ の構成を図1に示す。符号化部では、入力信号を MDCT により周波数領

域に変換し、線形予測スペクトルで MDCT 係数を正規化することにより一旦平坦化した後、さらに高い周波数分解能で平坦化するために、バークスケール・エンベロップで正規化してもう一度平坦化を行った後に、重み付けインタリーブベクトル量子化する。

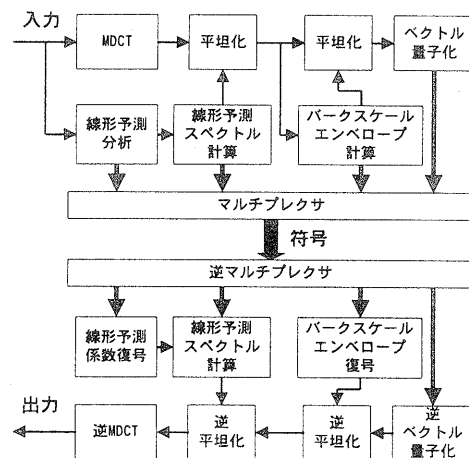


図1：TwinVQ 符号化方式の構成

### 2.2 平坦化の必要性

前節で述べたように、TwinVQ では、MDCT 係数をベクトル量子化する前に、2段にわたって係数を平坦化する。ベクトル量子化を用いた変換符号化では、平坦化は性能を決定する重要な操作である。本節では、ベクトル量子化の前処理として行われる平坦化の意義について述べる。

ベクトル量子化は、符号化すべきベクトル(目標ベクトル)を入力し、あらかじめコードブック中に用意してある複数のコードベクトルの中から、目標ベクトルに最近傍なものを1つ選出し、入力されたベクトルを、選出されたコードベクトルで表現する量子化方法である。

ベクトル量子化を高効率に行うためには、目標ベクトルの周辺にコードベクトルが高密度に分布している必要がある。目標ベクトルの平坦化は、この状態を得るための手法の一つである。

図2は、平坦化の効果の模式図である。問題

を単純化するために、ベクトルの次元数を2としている。図のように、コードベクトルの分布は、原点のまわりが最も高密度である。このようなコードブックに対して、点 $\alpha$ のように、ある軸の成分（図では次元0の成分）が極端に大きい目標ベクトルが与えられた場合、その近傍に最適なコードベクトルがある確率は小さい。そこで、目標ベクトルの極端に大きい成分を縮め、点 $\beta$ のようにコードベクトルが高密度な部分に移動させれば、高能率にベクトル量子化することができる。

目標ベクトルに対するこの操作は、目標ベクトルの値の凹凸を除去する操作になるので、平坦化と呼ばれる。

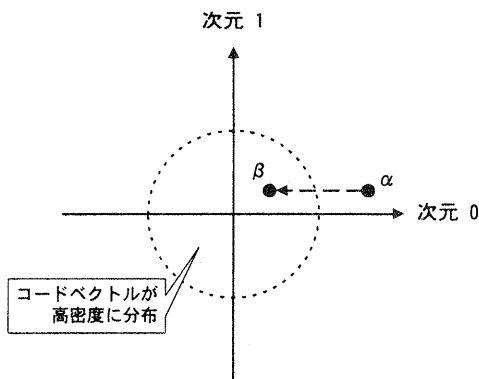


図2：目標ベクトルの平坦化

### 2. 3 従来法の平坦化における周波数分解能の限界

TwinVQ で2段目の平坦化に用いているバークスケール・エンベロープは、比較的高い周波数分解能を持つが、図3のように、高域になるほど周波数分解能が低くなる構造を持つ。そのため、細かいトーン成分が高域に入力されると、バークスケール・エンベロープの高域部の粗い周波数分解能では、MDCT係数の振幅の細かい凹凸を表現しきれなくなり、平坦化に不都合が生じる。このことが原因で、高域にトーン成分を含む音が入力されると、符号化歪みが増してしまう。

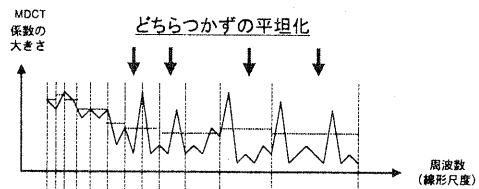


図3：従来法の問題点

## 3. Foreground / Background Categorization 音源分離技術を用いた平坦化

### 3. 1 原理

前項で述べた問題を解決するために、新たな手法 Foreground / Background Categorization (FBC) を提案する。この手法を用いると、少ない補助情報の追加で周波数分解能を大きく向上させることができる。

楽音の性質として、トーン成分は他の成分と比較して際立って振幅が大きく、トーン成分同士は似通った振幅を持つ傾向がある。本手法では、この傾向を利用し、係数を2系統に分離して処理している。

入力スペクトルに細かいトーン成分が存在する場合には、係数をその強弱によって前面側と背面側の2系統に分離する。分離された2系統の係数群は、各々に正規化のための代表値が与えられ、別々に平坦化を行う。

このような操作をすると、バークスケール・エンベロープの周波数分解能が十分でない高域についても、実際の係数の振幅をよりよく表現することができ、能率よく平坦化を行うことができる。

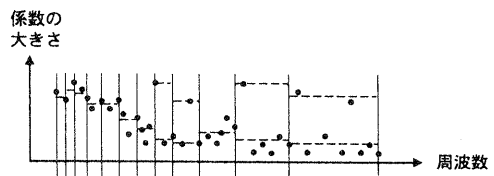


図4：スペクトル成分の強弱分離

### 3. 2 音源分離アルゴリズムの概要

図5に、FBC アルゴリズムの概要を示す。図6は、FBC アルゴリズムの動作の流れを示す。入力された MDCT 係数は、10 サンプル程度ずつまとめられ、小片群を形成する。これらの小片群を40個程度に帯域分割し、各帯域ごとに、小片の相対的な強弱を判断する。強いものは前面側、弱いものは背面側に分離して、それぞれ前面側係数と背面側係数を得る。補助情報として、小片の分離情報を符号化し復号器側に送る。

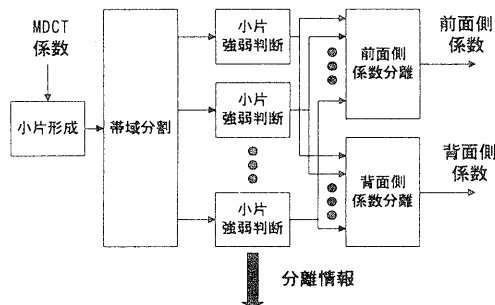


図5 FBC アルゴリズムの構成

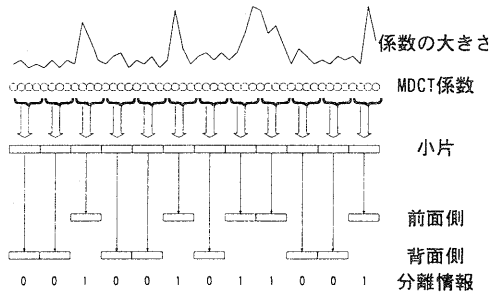


図6 FBC アルゴリズムの動作

### 3. 3 分離情報の符号化

FBC の特長の一つとして、必要とする補助情報が少ない点が挙げられる。

図6にあるように、分離情報は、1（前面側）と0（背面側）の値からなる。1の出現確率は0の出現確率と比べて小さいので、エントロピー符号化を利用して情報量を削減できる。

入力信号にトーン成分が含まれないときには、

平坦化の周波数分解能を上げる必要がないが、FBC では、トーン成分がない場合、前面側グループに小片を渡さないで、1と0の出現確率の格差はさらに高まり、エントロピー符号化の能率を向上させられる。

### 4. 提案符号化方式の構成

図7に新たに提案する小片強度伸長楽音符号化方式 (Segmental Intensity EXpanding Audio Coding, SinexAudio) の概要を示す。この方式では、入力信号を MDCT で周波数領域に変換した後、線形予測スペクトルにより平坦化し、その係数を FBC モジュールで2系統に分離した後、各々をバークスケール・エンベロープで平坦化する。このようにして得られた平坦化信号を再度合成し、ベクトル量子化する。

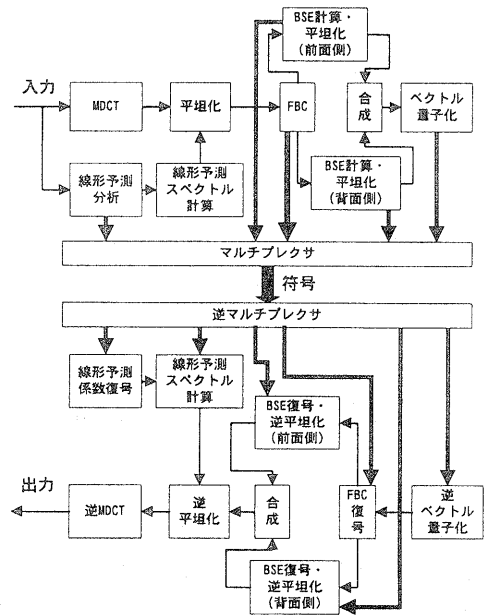


図7 : SinexAudio 符号化方式

### 5. 提案符号化方式の評価

#### 5. 1 客観評価

##### 5. 1. 1 平坦度

SinexAudio 符号化方式における FBC の効果を評価するために、ベクトル量子化器に入力さ

れる MDCT 係数の平坦度を TwinVQ と比較した。係数の平坦度は、次式で定義される。

$$S = \frac{\prod_{i=0}^{N-1} |x_i|^{1/N}}{\sum_{i=0}^{N-1} |x_i| / N}$$

ここで、 $S$  は平坦度、 $x_i$  はベクトル量子化にされる  $i$  番目の係数、 $N$  は係数の総数をあらわす。平坦度が高いことは、ベクトル量子化により適した形状の係数が得られていることを意味する。表 1 に比較の結果を示す。SinexAudio の方が平坦度が高く、FBC を使った平坦化が有効に働いていることがわかる。

表 1 : 様々な入力音に対する係数の平坦度

種類	SinexAudio	TwinVQ
ポップス	0.46	0.39
アコーディオン	0.40	0.37
バイオリン	0.54	0.52
トライアングル	0.39	0.36
ピアノ	0.56	0.54

### 5. 1. 2 セグメンタル SNR

表 2 は、復号化された信号の、原信号に対するセグメンタル WSNR の、SinexAudio と TwinVQ についての比較である。入力信号の条件は、44.1 kHz サンプリングのステレオ、符号化ビットレートは 48 kbit/s/ch である。SinexAudio 符号化方式の方が高い値を示し、ベクトル量子化が高効率に働いていることがわかる。

表 2 : 復元した音の原音に対する WSNR

種類	SinexAudio (dB)	TwinVQ (dB)
ポップス	33.2	31.5
アコーディオン	34.4	33.8
バイオリン	34.1	34.1
トライアングル	39.7	37.0
ピアノ	46.0	45.8

### 5. 1. 3 FBC モジュールの使用ビット

表 3 は、FBC モジュールと前面側のバークスケール・エンベロップが使用するビットの平均レートを示す。例えば 48 kbit/s/ch の符号化を行った場合、全ビットに占める FBC の割合は、高々全体の 10% 程度であり、少ないビット数で有効に働くことがわかる。

表 3 : FBC モジュールと前面側のバークスケール・エンベロップの使用ビットレート

種類	平均使用ビットレート (kbit/s/ch)
ポップス	4.07
アコーディオン	5.11
バイオリン	3.02
トライアングル	2.53
ピアノ	2.87

### 5. 2 主観評価

熟練した被験者 4 名により評価試験を行った。防音室においてスピーカ受聴し、SinexAudio と TwinVQ で作成した楽音を一対比較し、音質を -3 から +3 までの 7 段階で評価した。評価に使った信号は 44 kHz サンプリングのステレオ、符号化ビットレートは 48 kbit/s/ch とした。

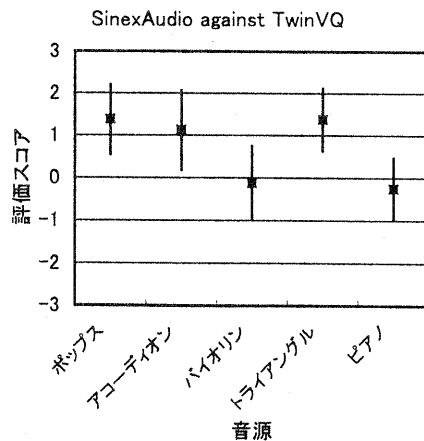


図 8 : 評価試験の結果

図8に評価試験の結果を示す。正の評価スコアは SinexAudio の音質が良いことを示す。全ての音源について、TwinVQ と同等かそれ以上のスコアを得た。

## 6. まとめ

本報告では、新たに考案した音源分離技術 FBC を用いた SinexAudio 符号化方式を提案した。SinexAudio は、ベクトル量子化を利用した変換符号化方式で、平坦化のために、従来法の線形予測スペクトルとバークスケール・エンベロープに加え FBC を使う。FBC は音源のトーン成分とそれ以外の成分を分離する技術で、トーン成分が含まれている音源について、局部的に周波数分解能を高めて平坦化することができる。FBC は分離情報の符号化のために使うビット数が少ないため、効果的にベクトル量子化の能率を高めることができる。

SinexAudio 符号化方式と、TwinVQ 符号化方式を比較評価した結果は、音質面で SinexAudio 方式のほうが優れていることを示した。

## 参考文献

- [1] Coding of Moving Pictures and Associated Audio for Digital Storage Media up to about 1.5 Mbit/s, ISO/IEC 11172, 1993.
- [2] R. Zelinski and P.Noll, "Adaptive transform coding of speech signals," IEEE Trans. ASSP, ASSP-25, pp.299-309, 1977
- [3] J. M. Tribolet and R.E. Crochiere: "Frequency Domain Coding of Speech Signals," IEEE Trans. ASSP, ASSP-27, pp.512-530.
- [4] 前田：ミニディスクシステム，音響学会誌，49, pp.277-283, 1993.
- [5] T. Moriya and M. Honda, "Transform coding of speech using a weighted vector quantizer," IEEE Trans. JSAC, vol. JSAC-6, pp. 425-631, 1988.

[6] N. Iwakami and T. Moriya, "Transform-domain weighted interleave vector quantization (TwinVQ)," AES 101<sup>st</sup> Convention, preprint, 1996

[7] J. Princen, A. Jhonson, and A. Bradley, "Adaptive transform coding incorporating time domain aliasing cancellation," Speech Commun., 6, p.299-308, 1987.