

# 音の音—NMFを用いた音の異音表現とそのエフェクト応用

柴田 理央<sup>†</sup> 小坂 直敏<sup>†</sup>

東京電機大学大学院 未来科学研究科 情報メディア学専攻<sup>†</sup>

## 1 はじめに

デジタル音合成特有のエフェクトとして、われわれは拡張音脈という概念を提唱し、ひとつの音(音脈)でありながらも他の音の要素を併せ持つ音合成手法を検討している。これらには、一つの音色から他の音色へ移りゆく音色モーフィング、数多くの音の要素を持つサウンドハイブリッド、一つの音が別の音から構成される「音の音」がある。これらを構造的音色とも呼んでいる。

「音の音」とは筆者らが新たに提唱するエフェクトで、一つの音を複数の異なる音で構造的に表現する枠組みである。例えば、水の攪拌音をフルートの音で表現するなどである。楽音、環境音を他の音により表現する検討は、茂出木による MIDI 音の楽音によるオート符[1]が知られている。しかし、その手法は MIDI 音源に限定された表現である。

本稿では、ひとつの音脈をより広範に、任意の音で表わす「音の音」に関して NMF 用いて音合成を行った検討結果について報告する。

ここでは人間の音声を複数の楽音で構成した音、つまり「楽音による音声の合成」を試みた。以後、表現される音(音声)を**目的音**、目的音を構成する音(複数の楽音)を**要素音**と呼ぶこととする。音声の音韻性を保ちつつこれを楽音により表現する目的で、その合成方法について検討した。

## 2 新たなエフェクトの合成

### 2.1 非負値行列因子分解[2]による音響信号分析

非負値行列因子分解 (Non-Negative Matrix Factorization; NMF) は、非負値のみからなる観測行列を、元のデータが持つ特徴パターンを示す基底行列と、その重み付けを示す係数行列の 2 つに近似的に分解するアルゴリズムである[2]。近年、音楽情報処理において自動採譜、雑音除去、音源分離など幅広く応用されている。

音響信号を扱う際、振幅スペクトログラムを観測行列とし、NMF により周波数スペクトルを表す基底と、その時変レベルを表すアクティベーションとに分解する。

本研究では分解時に乖離度を一般化 Kullback-Leibler ダイバージェンス (KL divergence) 基準で最小化する NMF アルゴリズム[3]を用いる。

### 2.2 再合成アルゴリズム

NMF による処理後の振幅スペクトルから音響信号に再合成する際、適切な位相復元をする必要がある。本研究では Griffin らの LSEE-MSTFTM[4]により振幅スペクトログラムからの位相復元を行った。

### 2.3 「音の音」の合成アルゴリズム

以下の学習、推定、再合成の 3 つの手順で音合成を行う。図 1 に学習と推定における NMF による分解の様子を示す。

**学習:** 要素音を NMF に適用し、要素音を構成する周波数スペクトルを抽出する。(図 1, 上) 1 音あたり 4 基底ですべての要素音に対し NMF 反復計算を行い、得られた周波数スペクトル行列を学習データとして保存する。

**推定:** 目的音である女性音声のパワースペクトログラムに対し、学習データを用いて NMF の適用をする。学習データで周波数スペクトル行列を固定し、アクティベーションのみ推定を行う。(図 1, 下)

**再合成:** 推定により得られたパワースペクトログラムを、LSEE-MSTFTM を用い音楽信号を合成する。

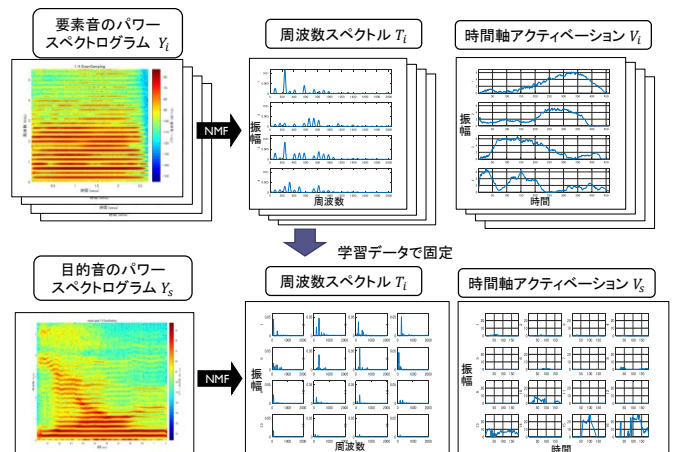


図 1 NMF を用いた学習と推定

表 1 使用楽器種と奏法

楽器名	奏法	ファイル名
バイオリン	ノーマル	151VNNOF
	ノンビブラート	151VNNVM
	Sul Ponticello	151VNPCF
ビオラ	ノーマル	161VLNOF
	Sul Ponticello	161VLPCF
チェロ	ノーマル	171VCNOF
	Sul Ponticello	171VCPCF
コントラバス	ノーマル	181CBNOF
	Sul Ponticello	181CBNOF
尺八	むら息	353SHMIF
	ノーマル	353SHNOF
	ビブラート(強)	353SHVIF
	ビブラート(弱)	353SHVIP
ビブラフォン	ハードマレット/ノーマル	041VIHNF
	ハードマレット/ペダル	041VIHPF
	ハードマレット/ビブラート	041VIHVF

sound by sounds—Sound representation in terms of different sounds using NMF and its application to an effect

Rio Shibata<sup>†</sup> Naotoshi Osaka<sup>†</sup>,

Tokyo Denki University<sup>†</sup>

2.4 合成条件

楽音による音声の合成にあたり，目的音は女声の母音「いえあおう」の音声データ，要素音には RWC 研究用音楽データベース：楽器音より，弦楽器(バイオリン，ビオラ，チェロ，コントラバス)，尺八，ビブラフォンの単独音，数奏法を用いた．使用した要素音の一覧を表 1 に，合成する際の要素音の組み合わせを表 2 に示す．

目的音・要素音のいずれもピッチは A2(220Hz 近辺)，すべてサンプリング周波数は 11025Hz とした．その他条件は表 3 に示す．

表 2 合成音 1~5 の要素音の組み合わせ

	要素音 楽器	要素音数	基底数
合成音①	弦楽器(Vn, Vl, Vc, Cb)	9	36
合成音②	弦楽器(Vn, Vl)	4	16
合成音③	尺八	4	16
合成音④	ビブラフォン	3	12
合成音⑤	尺八、ビブラフォン	7	28

表 3 「音の音」合成条件

学習反復計算	500回
推定反復計算	1000回
FFTポイント数	4096
分析窓	Modified hamming window[4]
窓幅	256
シフト幅	64
LSEE-MSTFTM更新回数	200回

2.5 結果と考察

目的音のソナグラムを図 3 に，合成結果例として合成音 1, 3 のソナグラムを図 4, 5 に示す．合成音 1, 5 のように，目的音を表す要素音を多く用意するほど音韻性(音色変化)が表れたが，その楽器らしさは損なわれた．一方，要素音が少ないと楽器らしさは保たれるものの，音韻性に劣る結果となり，音韻性と楽音性はトレードオフの関係といえる．

今後は音韻性があり，かつ楽音的な合成音を目指した検討を行う．

3 おわりに

「音の音」なる名称でひとつの音を別の楽音で表現する新たなエフェクトを合成するため，NMF を用い，音声の音韻を楽音で表わす方法を試みた．

この結果，一つの音脈を異なる音で階層的な表現をする，という目的を達成した．しかし，楽音としての印象はまだ弱いままであった．音韻性と楽音性はトレードオフの関係にあり，今後は楽音性を維持したまま音韻の変化(音色変化)を持たせた，より効果的なエフェクトとしての音合成を目指す．

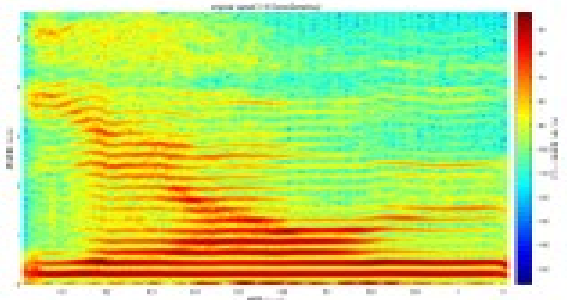


図 3 目的音 (女性音声) のソナグラム

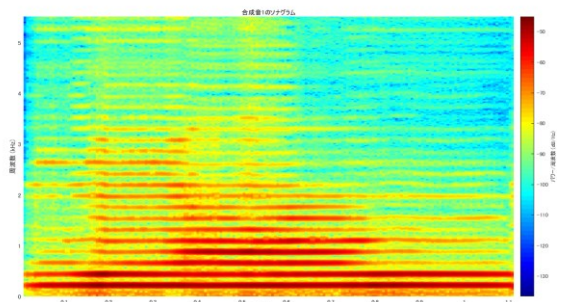


図 4 合成音 1 (要素音が弦楽器) のソナグラム

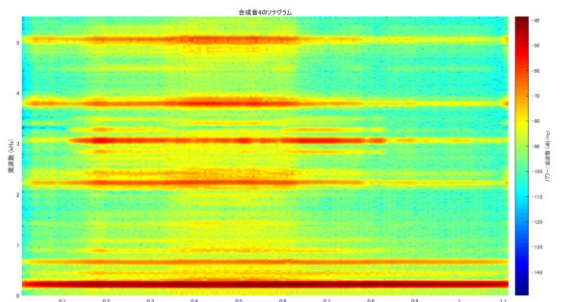


図 5 合成音 4 (要素音がビブラフォン) のソナグラム

参考文献

- [1] 茂出木敏雄, “音響信号の平均律音階に基づく汎用解析ツール「オート符」の開発”, 電気学会・電子情報システム部門誌, Vol.123-C, No.10, pp.1768-1775 (2003)
- [2] D.D. Lee and H.S. Seung, “Learning the parts of objects with nonnegative matrix factorization,” Nature, vol. 401, pp. 788-791 (1999)
- [3] P. Smaragdis and J. C. Brown, “Non-negative matrix factorization for polyphonic music transcription”, WASPAA, pp. 177-180 (2003)
- [4] D. W. Griffin and J. S. Lin, “Signal Estimation from modified Short-Time Fourier Transform”, ASSP-32, April 1984, pp. 236-242 (1984)