Vol.2020-MUS-126 No.8
2020/2/18

NMF を用いたサウンドコラージュの合成

池田将也 [1]　小坂直敏 [1]

概要：メディア表現に不可欠な音エフェクトの研究は，古くはギターエフェクトから始り，今でも研究，考案されている．我々は，「構造的音色」という概念を提唱し，その中の 1 つの音エフェクトとして，1 つの環境音(目的音)を，複数の楽器音(要素音)を用いて表現する音エフェクト．サウンドコラージュを提案した．
本稿では，NMF(非負値行列因子分解)による音源分離を応用したサウンドエフェクト，サウンドコラージュの合成手法を検討する．まず，一般的な NMF によるサウンドコラージュの合成手法を示す．次にその問題点とそれらを改善する手法を 2 つ示す．最後に，それらの手法の特徴と，合成結果を示す．

キーワード：NMF，サウンドエフェクト，音源分離

Synthesis of sound collage using NMF

MASAYA IKEDA[†1]　NAOTOSHI OSAKA[†1]

1. はじめに

芸術音楽，ゲーム，映像作品など，メディア作品の表現を豊かにするために, 音の表現は必要不可欠なものである．そのために，古くから音合成技術や音エフェクトが考案，研究されてきた．我々は，「構造的音色」(図 1)という概念を提唱し，その中の 1 カテゴリである「拡張音脈」について，サウンドモーフィング，サウンドハイブリッドのような音エフェクトを検討してきた[1].

構造的音色の中の 1 つの音エフェクトとして，1 つの環境音(目的音)を, 複数の楽器音(要素音)を用いて表現する音エフェクトを提案した[2]．このエフェクトはサウンドコラージュまたはサウンドモザイキングとも言われ，近年登場した新たなエフェクトである．1 つの環境音を目的音として設定し，それを複数の楽器音(要素音)を用いて階層的に表現する音エフェクトである．

これまでに，NMF を用いた手法を検討した．本稿ではサウンドコラージュの概念と条件を示し，その合成手法を更に詳細に検討する．

まず，基本的な NMF(Non-Negative Matrix Factorization;非負値行列因子分解)[4]による音源分離を応用した合成手法を示す．次にこの手法によるサウンドコラージュの問題点として，ごく短時間での同一音の繰り返しが起こる点や，多数の音の重なりにより，1 つ 1 つの要素音が個別の音色として知覚できない点を明らかにした．

それらを改善するために，NMF に 3 つの制約を加えた Driedger の手法[5]に更に制約を加えた手法と，Convolutive NMF[6]を用いた手法によってサウンドコラージュを合成し，アルゴリズム間の比較検討を行ったので，以下に報告する．

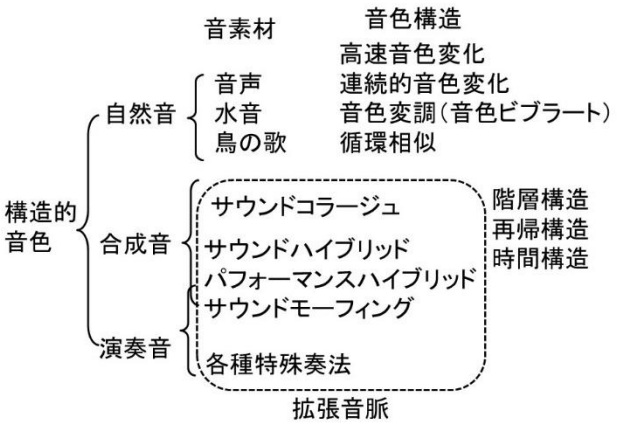

図 1 構造的音色

2. サウンドコラージュ

2.1 構造的音色の中のサウンドコラージュ

構造的音色は，楽音の演奏表現を含む自然音と合成音に大別される．音色が時間的あるいは空間的に，様々な要素が秩序立って組み合わされている音は，構造的な音色を持っていると言える．この分類にあてはまる自然音は構造的音色を持つ合成音の指標となるべき音色で，これを模倣して合成音を作ることが技術テーマである．サウンドコラージュは構造的音色の合成音の 1 つのカテゴリとして，筆者らは同エフェクトを，

1. 要素音の知覚(楽器性)を保つ
2. 目的音として知覚できる

という二重構造を保つことを条件として設定した．Arcimboldo が「花や葉を素材として，人の顔を画い

図 2 Arcimboldo 四季「春」

1 東京電機大学
Tokyo Denki University.

た」(図 2)ように，例えば「ヴァイオリンやフルートの音色を素材として，鳥の鳴き声を表現する」という音エフェクトである．

これは，要素音の組み合わせによって目的音に近似する最適化問題となっている．そして，目的音を要素音で分解する音源分離の応用として扱える．

ここでは，音源分離に NMF を利用し，その応用によってサウンドコラージュを合成する．

2.2 基本的な NMF によるサウンドコラージュ

NMF は，1 つの行列 V を，2 つの行列 W,H の積で近似できるような W と H を推定するアルゴリズムである．

$$V = W \times H \quad \cdots \quad (1)$$

式(1)が，基本的な NMF の式である．

ここで，$V_{M\times N}$，$W_{M\times R}$，$H_{R\times N}$ であり，V が観測行列，W が基底行列，H が係数行列となり，3 つの行列は全ての要素が非負値である．

しかし，実際には推定された基底行列と係数行列の積と，観測行列には誤差があり，式は式(2)のようになる．

$$V = W \times H + C \quad \cdots \quad (2)$$

(2)中の C は誤差であり，これを最小化するように反復演算する．NMF で用いられる代表的な誤差の基準として，ユークリッド距離，KL ダイバージェンス，板倉斎藤儀距離があり，それぞれの基準によって異なる更新式を反復演算することによって，誤差の最小化を図る．y と x を同じ大きさの行列として，本稿で使用した KL ダイバージェンスについて，y の x からの距離を式(3)で示す．

$$D_{KL}(y|x) = y\log\frac{y}{x} - y + x \quad \cdots \quad (3)$$

本稿で使用した KL ダイバージェンスの更新式を式(4)に示す．$v_{rn}, \hat{v}_{mn}, w_{mr}, h_{rn}$ はそれぞれ行列の要素である．

$$w_{mr} \leftarrow w_{mr}\frac{\sum_n \frac{v_{mn}}{\hat{v}_{mn}} v_{kj}}{\sum_n v_{rn}} \ , h_{rn} \leftarrow h_{rn}\frac{\sum_m \frac{v_{mn}}{\hat{v}_{mn}} w_{mr}}{\sum_m w_{mr}} \quad \cdots \quad (4)$$

音声信号処理に用いる際，V に対象音のスペクトログラムを入力する．W には対象音で使用されている音 R 個の周波数特性が推定され，H には係数行列として，それぞれの時変の振幅(時間軸アクティベーション)が推定される．(図 3)

音源分離に用いられる場合，分離対象音で使用されている音の周波数特性を，基底行列 W に予め与えた，教師あり学習が用いられる．例えば，フルート，ピアノ，ヴァイオリンからなる三重奏を楽器ごとに分離する場合，基底行列 W に，フルート単音，ピアノ単音，ヴァイオリン単音の周波数特性を予め入力し，時間軸アクティベーションのみを推定する．

サウンドコラージュに用いる場合，観測行列に目的音を入力する．そして，目的音中で使用されている音とは異なる音を要素音として，その周波数特性を基底行列に入力し，その時間軸アクティベーションを推定する．(図 4)基底行列

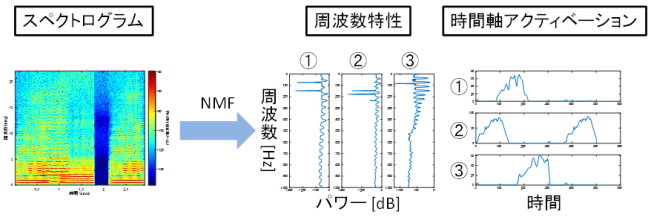

図 3 NMF による音響信号処理

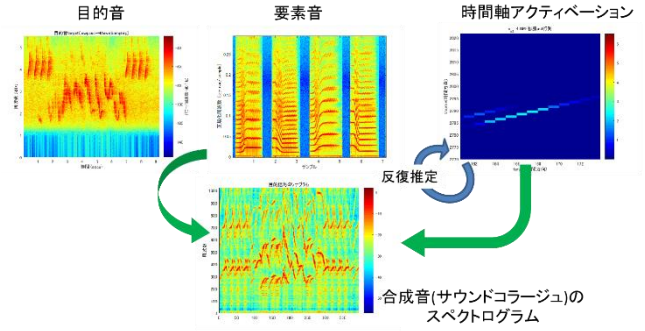

図 4 NMF によるサウンドコラージュ

と時間軸アクティベーションの積は，合成音声のスペクトログラムである．スペクトログラムから音響信号を合成する際，位相情報を復元する必要がある．

本研究では，位相情報の復元に，Griffin-Lim の手法(LSEE-MSTFTM)[7]を用いた．

2.3 NMF 制約モデル

前項の手法で，要素音による目的音のサウンドコラージュが合成できる．しかし，この方法では，要素音の音色が保持されないという問題点があった．J.Driedger らは，オーディオモザイキングとして先行研究を行い，その原因を以下のように述べている．

1. 要素音の 1 部分が短時間に何度も繰り返されるため，吃音となる
2. 同時刻に多数の要素音が重なり，1 つ 1 つの音色が埋もれる
3. 要素音のフレームを時間順序関係なく並べるため，時間構造が破壊される

以上の 3 つの理由により，サウンドコラージュに要素音の音色が保持されないとした．

Driedger らは，これらを改善するために，時間軸アクティベーションの推定に以下の 3 つの制約を加えた．

1. 水平方向隣接間隔制限
2. 垂直方向多重制限
3. 対角構造化

制約 1 は，要素音が時間的に隣接しすぎないための制約である．

制約 2 は，音楽の声部数に相当し，同時に発声する音の数を制限する制約である．

制約 3 は，時間軸アクティベーションが，基底と同じ時

間進行をすることを保証する制約であり，要素音の時間構造を保つための制約である．

以上 3 つの制約に従うように NMF の反復演算を収束させていた．

本論文ではこの手法を，NMF 制約モデルと呼ぶ．

2.4 NMF 制約モデルの問題点

要素音の音色を保つための NMF 制約モデルによるサウンドコラージュは，以下のような問題点がある．

1. ある要素音を基底に設定し，その要素音の時間構造を保ちながらも，その音の一部しか用いていない．
2. Griffin-Lim の手法による合成法を用いていることから，音質劣化が生じる

問題点 1 について，楽器音には，立ち上がり部分と立ち下り部分に大きな特徴を持つ音が存在する．例えば打楽器は立ち上がり部分に大きな特徴を持つ．そのような音を要素音として用いる場合，要素音の大きな特徴が失われる可能性がある．

また，問題点 2 について，同手法では最終的に，NMF 制約モデルによって推定したスペクトログラムの振幅情報に，Griffin-Lim の手法によって位相情報を推定して音波形を復元している．このため，音質の劣化が起こる．

これらの問題を解決するため，2 つの手法を提案する．

3. 提案手法 1

3.1 NMF 制約モデルの改良

Driedger らの NMF 制約モデルの，要素音の一部分のみを使用することによって，要素音の立ち上がりと立ち下がり部分の特徴が失われるという問題点を改善するために，NMF 制約モデルの制約 3 に関して，新たな制約を追加する．

従来手法では，一定時間の要素音の時間構造を保つように，NMF の反復演算を収束させていたが，提案手法 1 では，要素音の立ち上がりから消失までの全部分を使用するように時間軸アクティベーションの反復演算を収束させる．

3.2 位相推定を行わない合成

Driedger の手法では，時間軸アクティベーションを推定した後，Griffin-Lim の手法によって，位相推定を行い，音波形を合成する．しかし，一度失われた位相情報を推定することにより，合成音の品質が劣化する．これを避けるために，時間軸アクティベーションに対応した時間に，対応した振幅で要素音を直接貼り付けることで，品質の劣化を回避する．

提案手法 1 の概念図を図 5 に示す．

4. 提案手法 2

4.1 Convolutive NMF(CNMF)

提案手法 1 までの手法では，0.02 秒程度のごく短時間の要素音 1 フレームの周波数特性を基底 1 つとして扱っていた．

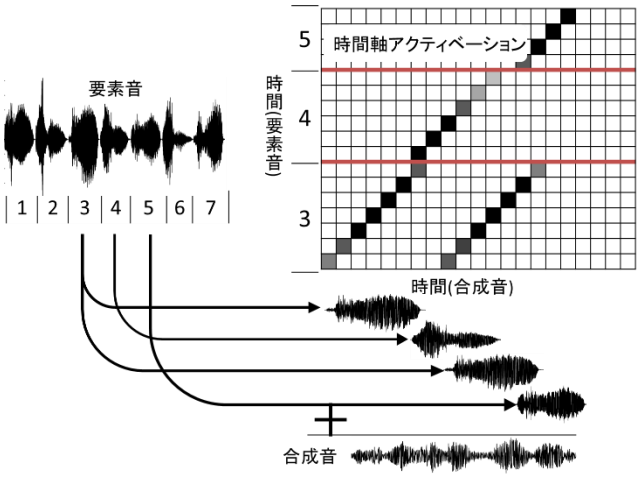

図 5 提案手法 1 によるサウンドコラージュ

長時間の要素音を，時間構造も含めて基底として保持するために，Convolutive NMF を使用する．図 6 に基本的な NMF との違いを示す．2 つの基底を扱う場合，NMF では 2 フレーム分の周波数特性を持つ．長時間の周波数特性を基底として持つこともできるが，時間的連続性を保持することはできない．Convolutive NMF の場合，時間構造も含めて長時間の周波数特性を 1 つの基底として扱える．

式(2)に Convolutive NMF の基本式を示す．

$$V = \int_{t=0}^{T} W(t) \cdot {}^{t}\vec{H} \quad \cdots \quad (5)$$

ここで，基底の数は R とし，$V(M \times N), W(t)(M \times R), H(R \times N)$である．

また，演算子 t→ は，行列を矢印方向に t 列ずらし，空いた列に 0 を入力する演算子である．

基本的な NMF 同様，誤差を最小化する更新式を反復演算することで，最適な基底と係数行列を求める．係数行列の更新式を式(6)，基底の更新式を式(7)に示す．

$$H = H \circ \frac{W(t)^T \cdot \left[\frac{V}{\hat{V}}\right]^{\overleftarrow{t}}}{W(t)^T \cdot 1} \quad \cdots \quad (6)$$

$$W(t) = W(t) \circ \frac{\frac{V}{\hat{V}} \cdot {}^{\overrightarrow{t}}H^T}{1 \cdot {}^{\overrightarrow{t}}H^T} \quad \cdots \quad (7)$$

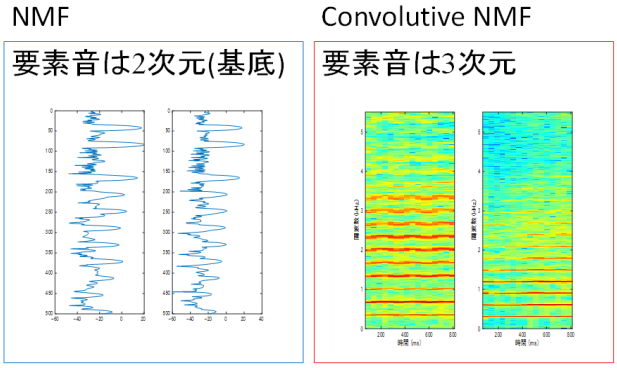

図 6 NMF と CNMF の基底

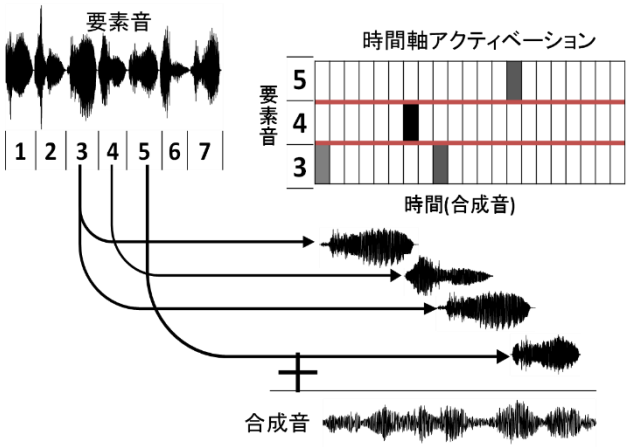

図 7 提案手法 2 によるサウンドコラージュ

4.2 CNMF によるサウンドコラージュ

Convolutive NMF を用いたサウンドコラージュの合成の手順を以下に示す．また，その概念図を図 7 に示す．

1. 目的音のスペクトログラムを求め，観測行列 V に与える
2. 要素音を複数用意し，それぞれのスペクトログラムを求める
3. 基底行列に，要素音のスペクトログラムを入力する
4. 重み付け行列 H の初期値として，乱数を入力する
5. Convolutive NMF の反復演算によって時間軸アクティベーションを推定する
6. 時間軸アクティベーションに従い要素音を当てはめる

また，今回 CNMF によってサウンドコラージュを合成するにあたり，nmf-toolbox[8] を用いた．

4.3 CNMF によるサウンドコラージュの問題点

CNMF によるサウンドコラージュでは，長時間の周波数特性をまとめて 1 つの基底として持つ性質上，確実に要素音の時間構造を保つことができる．しかし，同じ要素音の時間軸アクティベーションが，連続して大きな値を持った時, 合成音はごく短時間に, 同じ要素音が連続して起こる．その場合，吃音と呼ばれるノイズが発生する．特に，フルートやヴァイオリンなど，持続性の音ではそれが顕著に見られ，要素音の音色は知覚できなくなる．要素音 1 つ 1 つを知覚できるサウンドコラージュを合成するためには，この問題点を解決する必要がある．

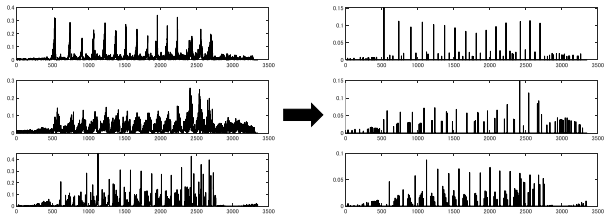

図 8 時間軸アクティベーションのピーク抽出

4.4 ピーク抽出 CNMF によるサウンドコラージュ

同一の要素音に対応する係数が，連続して大きな値を持つことで，要素音の音色が知覚できなくなる．そのため，時間軸アクティベーションはスパースな行列であることが理想的である．しかし，時間軸アクティベーションの疎密の差が大きくなると，一部分のみが強調された不自然な音が合成される．

つまり，時間軸アクティベーションは全体的に満遍なくパワーを持ち，かつ疎な行列であることが望ましい．

そこで，ここでは一定範囲内の最大値を抽出し，それ以外の要素を 0 にする処理を行った. その様子を図 8 に示す.

5. 2 手法の評価と考察

5.1 実験 1:単純なドラムサウンドの置き換え

サウンドコラージュは，目的音の中の音色を，要素音の音色に置き換えることが基本である．それがどれだけ達成できているかを確かめることによって，サウンドコラージュの基本性能を確かめる．

まずは，単純な目的音を，ヴァイオリンのピチカートに置き換える実験を行った．これは，減衰音であるドラムサウンドを, 同じ減衰音であるヴァイオリンのピチカートで, 全く同じリズムに置き換えられるか確かめるためである．

目的音はドラムサウンド 1 種類，要素音はヴァイオリン

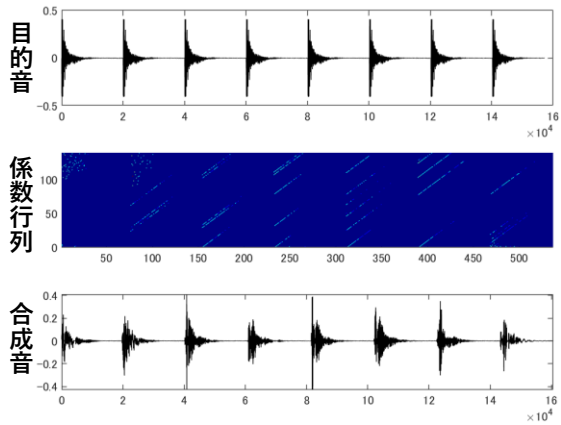

図 9 提案手法 1 の実験 1

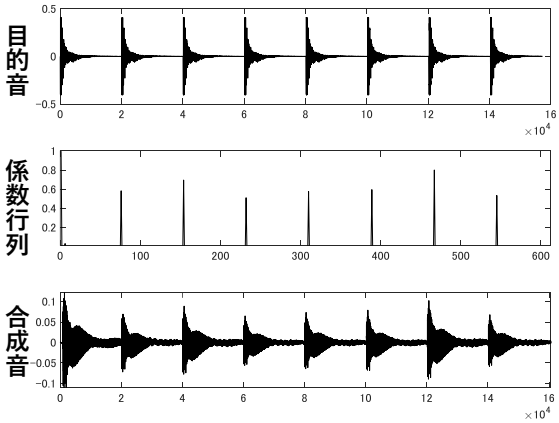

図 10 提案手法 2 の実験 1

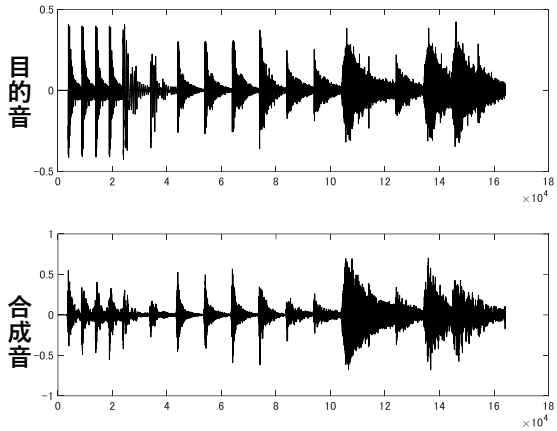

図 11 提案手法 1 の実験 2 音波形

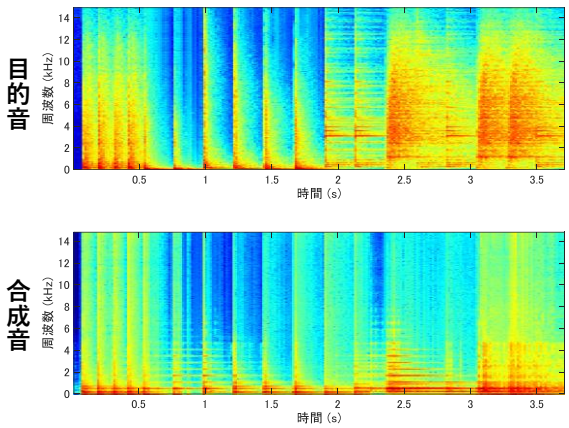

図 12 提案手法 1 の実験 2 スペクトログラム

のピチカート 1 ピッチである.

提案手法 1 の実験結果(図 9)を見ると，リズムの置き換えは行えている．しかし，要素音の音色が崩れていることが合成音の音波形が崩れている事から見て取れる．これは，時間軸アクティベーションが基本的には対角線を描いているが，パワーが一定ではなく，所々途切れている部分があるからである.

次に，提案手法 2(図 10)の実験結果を見てみる．リズムの置き換えは正確に出来ている．また，波形の大きさは違うものの，要素音の音色の保持も出来ていることが分かる.

5.2 実験 2:複雑なドラムサウンドの置き換え

次に，目的音のドラムサウンドを少し複雑にした例で，ヴァイオリンのピチカートに置き換える実験を行った.

ドラムサウンドは 5 種類で，要素音はヴァイオリンのピチカート 4 ピッチで合成を行う.

提案手法 1 の実験結果(図 11,12)を見ると，リズム及び，周波数特性は, 目的音を模倣できていると言える. しかし，聴感上要素音が歪む箇所が多く，やはり高品質な合成が出来ているとは言えなかった.

提案手法 2 の実験結果(図 13,14)を見ると，提案手法 1 同様, リズムとスペクトログラムの模倣は出来ている. また，提案手法 2 は要素音の音色は聴感上歪んでおらず，要素音の音色が保持できていると言える.

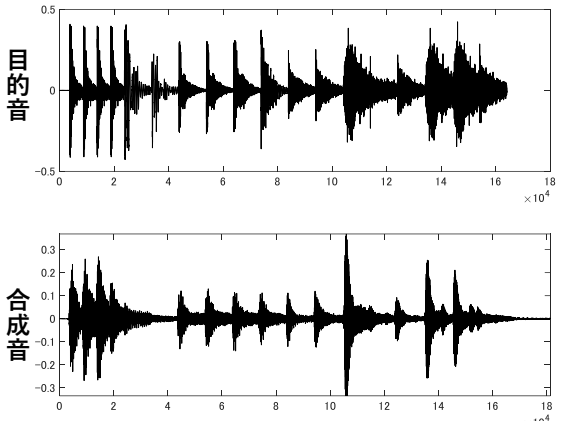

図 13 提案手法 2 の実験 2 音波形

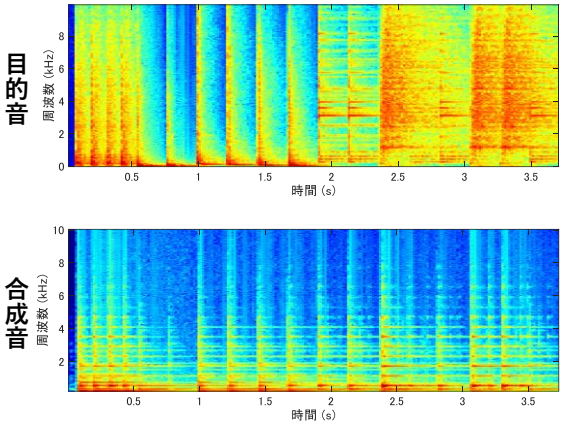

図 14 提案手法 2 の実験 2 スペクトログラム

5.3 スペクトル歪による物理評価

サウンドコラージュの基本性能を測るため，分析再合成実験によって，物理評価を行う.

目的音と合成結果が，周波数上でどれだけの差があるのかを測るため，評価尺度には，目的音と合成結果のスペクトル歪みを用いる. スペクトル歪の導出は, 式(8)で行った.

$$SD = 20log_{10}(\frac{\text{目的音のパワースペクトル}}{\text{合成音のパワースペクトル}}) \quad \cdots (8)$$

目的音には，チェロ，クラリネット，フルートによるブルガリア民謡の三重奏を入力．(図 15，クラリネットは記音) 要素音には上記 3 つの楽器の，目的音中で使われているピッチの単音をそれぞれ入力した.

合成結果と目的音をパワーで正規化し，そのスペクトル歪みを計測した結果が以下の表 1 である.

表 1 サウンドコラージュ合成手法とスペクトル歪み

手法	スペクトル歪み [dB]
NMF	5.89
改良 NMF 制約モデル	4.59
CNMF	2.71

5.4 2 手法の性質

手法 1 は，基本の NMF に，聴感上の品質を向上させるための制約を付与した手法である．要素音 1 部分でなく，全

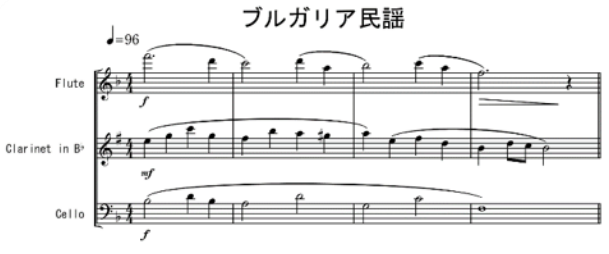

図 15 使用したブルガリア民謡の楽譜

部分を使うよう制約を加えた．しかし，目的音の振幅が小さくなった時，要素音の 1 部分の振幅を小さくして目的音に近似させる. このとき, 要素音の時間構造は保たれるが，振幅の整合性が失われ，音色を保つことが出来なくなる．

手法 2 は, 1 フレーム分の時間軸アクティベーションが，要素音全体に対して適応される．要素音全体が，確実にパワーの整合性を保ったまま用いられる．この点で，手法 1 よりも要素音の音色を保つことができると言える．しかし要素音の長さを変えることはできない．目的音に対応する要素音が過剰に長いとき，過剰に長いまま要素音が使われる．そのため，目的音に近似させることはできなくなる．

2 つの手法を用いて，いくつかの目的音に対し，サウンドコラージュを合成し，聴感上の音色の歪みや目的音の再限度から，合成手法の特徴を考察した．

セミの鳴き声を目的音にした場合，提案手法 1 による合成では要素音の音色が保たれており，目的音を正確に再現できていた．逆に提案手法 2 では，タイミングのずれが見られ，目的音の再限度は今一つだった．

ドラムサウンド，水音を目的音に設定した場合，提案手法 2 による合成は音色の歪みやタイミングのずれは見られなかった．しかし，提案手法 1 による合成では，要素音の音色の歪みが大きく，品質は悪かった．

このことから，振幅の激しい変化の少ない持続音の多い目的音を扱う場合，提案手法 1 による合成が適していると考えられる．打撃音が多い目的音を扱う場合，提案手法 2 による合成が適していると考えられる．

6. まとめ

近年生まれたデジタルサウンドエフェクトの 1 つとして，サウンドコラージュの概念を紹介した．われわれは同エフェクトの満たすべき条件として，1)目的音の知覚と 2)楽器性を保つことを挙げた．そしてその合成手法を 2 つ提案した．

提案手法 1 は，Driedger の NMF 制約モデルを，より要素音の音色を保つために，要素音全部分を扱うよう，制約を加えたものである．

提案手法 2 は，Convolutive NMF を用いた手法である．Convolutive NMF の, 基底を長時間保持できる性質により，確実に要素音の時間構造を保てる．

また，Griffin-Lim の手法による位相情報の復元を行わずに直接波形を貼り付ける事によって，品質の劣化を回避する．

2 つの手法は，要素音の時間構造を保つための手法であり，それによって要素音の音色を保ったサウンドコラージュの合成が可能になった．

しかし両手法ともに周波数特性で近似させるため，目的音と近いピッチの要素音を選択する必要がある．また，提案手法 2 は要素音の長さを変えることができないため，時間的にも相性の良い要素音を選択する必要がある．

今後は，より詳細な評価を行い，更なる品質の向上を目指す．また，目的音に適した要素音を自動的に選び出すようなシステムの構築も検討したい．

7. 参考文献

[1] 小坂直敏，“構造的音色とその電子音響音楽への応用，” JSSA 会報 Vol.9 No.1， pp.7-12, 2017.

[2] 池田将也，小坂直敏，”新たな音エフェクト―サウンドコラージュの合成，”音講論 秋季 2-3-10, 2018

[3] Jonathan Harvey: Speakings (2007/2008) https://www.youtube.com/watch?v=6UJ2RXIEXa4 [2018/1/26]

[4] 亀岡弘和，“非負値行列因子分解とその音響信号処理への応用”, 日本統計学会誌第 44 巻第 2 号, pp.383－407 (2015).

[5] J. Driedger, et al.，“LET IT BEE‐Towards NMF-Inspired audio mosaicking”, Proc. of the 16th ISMIR, M’alaga, Spain, 2005.

[6] Paris Smaragdis, “Non-negative Matrix Factor Deconvolution; Extraction of Multiple Sound Sources from Monophonic Inputs,” ICA 2004: Independent Component Analysis and Blind Signal Separationpp 494-499

[7] D. W. Griffin and J. S Lin, “Signal Estimation from modified Short-Time Fourier Transform”, ASSP-32, April 1984, pp. 236-242 (1984)

[8] “nmf-toolbox”. https://github.com/colinvaz/nmf-toolbox, (参照 2020-01-29)