Vol.2018-MUS-118 No.20
Vol.2018-SLP-120 No.20
2018/2/21

記号と信号処理の相互作用フレームワークの構築に向けたGTTMの大域的構造を考慮した音響信号の分節の調整

澤田 隼[1,a)]　竹川 佳成[2,b)]　平田 圭二[2,c)]

概要：本論文では，記号処理と信号処理の相互作用フレームワークの構築に向けて，Generative Theory of Tonal Music（GTTM）の大域的構造を考慮したスペクトログラムの分割位置を調節する実験を行った．従来の GTTM を音楽のスペクトログラムに適用する試みでは，スペクトログラムをビートで分割し（bin）その bin 毎に特徴量を抽出し，階層的クラスタリングを行うことによってタイムスパン・セグメンテーションを生成した．しかし，スペクトログラムの分割する位置によって bin が持つ特徴量の値が変化するが，スペクトログラムの分割の段階まで戻って修正する枠組みが無かった．本論文ではスペクトログラムを分割する段階に GTTM の大域的な構造をフィードバックすることで適切な分割位置を獲得する枠組みを提案し，その有用性を検証する．その結果，スペクトログラムの分割位置によって精度が変わり，適切な分割位置では期待するタイムスパン・セグメンテーションが得られ，大域的構造を考慮した分割位置の修正が有用であることが示された．

1. はじめに

音楽情報処理の研究において，楽曲構造分析やカバーソングの楽曲同定，コード推定などの分野では，ビート毎に特徴量を抽出し処理をする研究が多数存在する [3]．これらのビート同期した特徴量を用いた手法の精度はビートトラッキングの精度に依存する．また，伴奏システムにおいては，人間の演奏に同期するように伴奏のテンポを変化させるためにビートトラッキングが必要となる．音楽音響信号を対象とした伴奏システムとして足立ら [1] のシステムがある．足立らは独奏と伴奏のずれと，伴奏のテンポ変化の過去の履歴から次の伴奏のテンポ変化量を決めるモデルを構築した．中村ら [10] や鈴木ら [14] の伴奏システムは，時刻 t までの演奏者の音響演奏系列を持つ隠れマルコフモデルを用いて人間の演奏をモデル化し，楽譜追跡問題は観測系列が与えられた時の事後確率を最大にする拍位置系列を求める問題とした．これらのシステムはそれより前の情報からのみ次の位置を予想するが，適切な分析を行うためには楽曲の大域的構造を考慮する必要がある．

通常，音響信号の情報から拍の位置や和音などの記号を生成するという意味で，音響から信号への処理は一方向であるが，適切な記号接地を実現させるためには音響信号と記号が相互に作用する枠組みが必要であり，記号から音響信号へのフィードバックが必要になる．例えば，大域的構造がわかればその構造に従って，あるいはその構造に当てはまる様に音響信号の処理に修正を加える事ができる．本論文ではスペクトログラムを分割する段階に音楽理論に基づいた大域的な構造をフィードバックすることで適切な分割位置を獲得する枠組みを提案し，その有用性を検証する．

2. スペクトログラムのタイムスパンセグメンテーション

楽譜に書かれた楽曲の構造や意味を分析する手法として Generative Theory of Tonal Music (GTTM) がある [8]．これはグルーピング構造分析と拍節構造分析を経て，人間の認知過程を踏まえた音楽の階層的な構造を抽出する分析手法である．グルーピング構造分析は楽曲全体を音楽的にまとまりのあるグループに分割し，グループの階層構造を抽出する分析であり，拍節構造分析は楽曲から各階層毎に強拍と弱拍の位置を示す階層構造を抽出する分析である．各分析は構成規則 (well-formedness rules) と選好規則 (preference rules) の二種類の規則からなる．構成規則は満たすべき基本的な構造の特性を示す規則であり，選好規則は経験豊かな聴衆の聴取によって好ましい構造を示す規則である．

[1] 公立はこだて未来大学大学院
Graduate School of Future University Hakodate
[2] 公立はこだて未来大学
Future University Hakodate
[a)] g2116022@fun.ac.jp
[b)] yoshi@fun.ac.jp
[c)] hirata@fun.ac.jp

Vol.2018-MUS-118 No.20
Vol.2018-SLP-120 No.20
2018/2/21

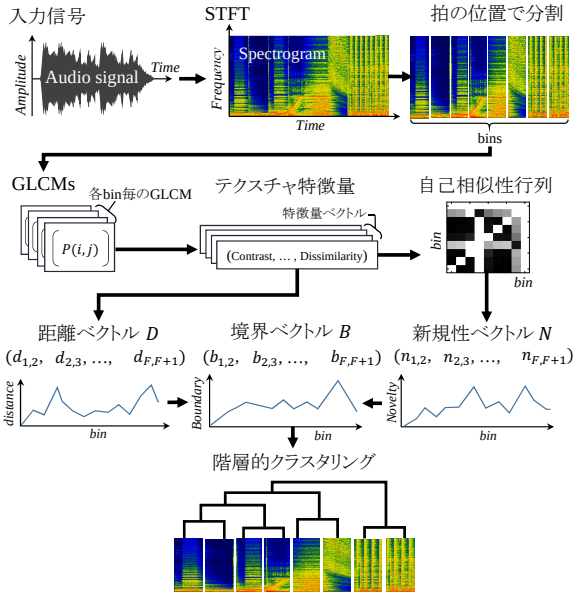

図 1　システム構成

タイムスパン・セグメンテーションとは，Lerdahl と Jackedoff によって導入された基本的な音楽構造の 1 つで，グルーピング選好規則によるグルーピングの結果と，周期構造やリズムといった拍節構造の結果を統合した構造であり，タイムスパン簡約が行われる領域として定義されている．人間は旋律の特徴による上位のグループ境界の情報とリズムによる下位のグループ境界の情報を使って，人間の認知と整合する認知的リアリティのある境界を同定している．この 2 つの異なる境界の認知を統合する構造としてタイムスパン・セグメンテーションが導入された．

GTTM のタイムスパン簡約は，グルーピング構造分析と拍節構造分析をもとに構造的に重要な音の選出を繰り返すことで旋律中の各音符の重要度を二分木（タイムスパン木）で表すことができる．それは楽曲の構造の記述にとどまらず，楽曲の構造の操作を可能にするものであった．タイムスパン木は認知的リアリティーを持つことが Dibben によって確認されており [2]，人間が音楽を聴取した際の認知過程を踏まえた音楽的に信頼できる分析が可能となる．

2.1　従来手法とその限界

GTTM を音楽音響信号に適用する従来の手法を述べる．GTTM のグルーピングの選好規則によると，グループの境界はピッチイベント間の時間軸方向の近接性及び，音高や音量などの変化に基づいて形成される．ピッチイベントの近接性や変化は，スペクトログラム上ではテクスチャのパターンとしてあらわれる．そこで我々はパターン認識技術を使用し，スペクトログラム内の隣接するセグメント間の特徴量の距離を計算し，これを近接及び変化の尺度として使用した [12], [13]．以下に処理手順を示す（図 1）．初めに，入力された音楽音響信号を短時間フーリエ変換し，グレースケールのスペクトログラムを周波数軸はメルスケールで描画する．Ullrich ら [15] や中鹿ら [11] に倣って，スペクトログラムの周波数軸はメルスケールを採用した．次に，スペクトログラムを時間軸方向にビートの位置で分割し，ビートの長さ分の短冊状のデータ (bin) の集合として考える．ビート同期技術は McFee と Ellis[9] と同じ技術を採用した．Costa ら [16] の研究ではスペクトログラムをいくつかの bin に分割する方が，全体的に処理するよりも優れていることを示している．

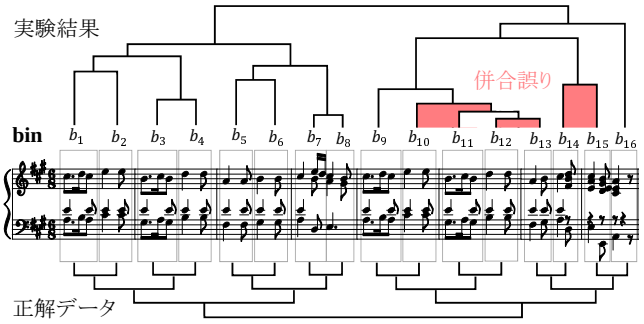

図 2　従来手法のタイムスパンセグメンテーション

次に各 bin 毎にスペクトログラムのテクスチャ特徴量 [7] を抽出する．その後，楽曲内の繰り返し構造を抽出するためにこのテクスチャ特徴量を用いて自己相似性行列を計算する．これは GTTM のグルーピングの選好規則（GPR6）による繰り返し現れるフレーズは同じ構造になることが望ましいという思想に基づいており，繰り返し構造の始点と終点でグループの境界が強くひかれるような設計をした．最後に，隣接する bin 間のテクスチャの特徴量の距離と，楽曲内の繰り返し構造の情報を用いて時間軸方向に制約を持つ階層的クラスタリングを行う．テクスチャの変化が小さいものから併合されていき，テクスチャの変化が大きい場合は上位の境界として抽出される．

従来の我々の手法を用いて，楽曲を分析した結果を図 2 上に，期待する結果を図 2 下に示す．本来，$b_1 - b_4$ 間と $b_9 - b_{12}$ 間は譜面上は同じであるが，そのセグメンテーションは同じになっていない．音響信号を対象とした場合，音量の変化やテンポの揺らぎなどの演奏の表情付けによる影響もあるが，GPR6 によると，繰り返し現れるフレーズは同じ構造になることが望ましい．この原因として，スペクトログラムの分割する位置が適切でないことが挙げられる．従来の我々の手法は，スペクトログラムの分割する位置によってその bin が持つ特徴量の値が変化するが，適切な構造になる様にスペクトログラムの分割の段階まで戻って分割位置を修正する枠組みが無かった．図 2 の様な楽曲を適切に分析するためには，GPR6 の制約を強くする必要がある．つまり，同じフレーズの場所は同じ特徴量を持つようなスペクトログラムの分割になるようにスペクトログラムの分割位置を適切に修正する必要がある．

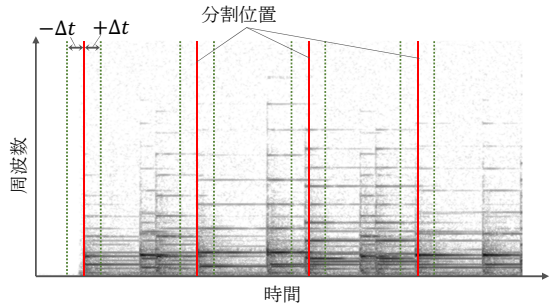

図 3 スペクトログラムの分割位置とその変域

2.2 大域的構造を考慮した分割位置の変更

GTTM のグルーピングの選好規則には GPR6 がある．GPR6 は繰り返し現れるフレーズは同じ構造になるのが望ましいという規則である．何らかの方法で大域的な繰り返し構造が獲得できれば，トップダウンにスペクトログラムの分割位置や誤った併合を修正することができる．繰り返し構造を持ったペアの距離が小さくなるようにスペクトログラムの分割位置を最適化する．

i 番目の bin を b_i，その bin の持つ特徴量を f_i とする．また，修正前の分割位置ベクトル X の要素である i 番目の bin の先頭の分割位置を $x_i \in X$，修正後の分割位置ベクトル X' の要素である分割位置を $x'_i \in X'$ とし，i 番目と j 番目の 2 つの bin 間の特徴量の距離を $d(f_i, f_j)$ と表す．GPR6 を考慮した分割位置の変更は，$|x_i - x'_i| \leqq \Delta t$ の制約のもとで，同じフレーズの全ての bin 同士のペアにおける $D = \sum d(f_i, f_j)$ を最小化することで得られる x'_i のベクトル X'_{opt} を求める事と定義できる（図 3）．

3. 実験内容と結果

3.1 実験 1：大域的構造を考慮した bin 間の距離の最小化によるタイムスパンセグメンテーション

大域的構造を考慮した分割位置の変更が有用であるかを調べるために，以下の実験を行った．今回は大域的な繰り返し構造を既知とし，楽曲内で同じフレーズの bin 同士の類似度が高くなるような分割位置を目指した．Mozart の piano sonata in A major（K.331）の第一楽章のテーマの実験結果を示す（図 5）．ここで，使用した K.331 は Maria João Pires によるピアノ演奏のホモフォニー楽曲であった．正解データはオリジナルの GTTM のルールを楽譜に適用した場合の結果を正解とみなし，GTTM の原本に書かれているものと，浜中によって公開されている GTTM database [6] を使用した．

まず，オンセット検出によって抽出された分割位置ベクトル X の各分割位置を，$\pm\Delta t$ ミリ秒の範囲内でランダムにずらしながら分割位置ベクトル X' を生成した．今回 Δt は ±100 ミリ秒とし，2432 個のデータの分割位置ベクトル X' を生成した（図 4，表 1）．今回の実験では，$|x_i - x'_i| \leqq 100$ の制約のもとで，$D = d(f_1, f_9) + d(f_2, f_{10}) + d(f_3, f_{11}) + d(f_4, f_{12})$ を最小化することで得られる x'_i のベクトル X'_{opt} を求める．

修正前の分割位置ベクトル X と，距離 D が最小になる様に修正をした分割位置ベクトル X'_{opt} を図 4 と表 1 に示す．修正前の分割位置を破線で，修正後の分割位置を実線で示した．この修正後の分割位置を用いて，従来手法と同じようにタイムスパン・セグメンテーションを抽出した結果を図 5 に示す．従来手法では後半部分（b_9 から b_{16}）に初期の段階で併合誤りがあったが，適切に分割位置を調節すると，期待する結果が得られた．

3.2 実験 2：類似度の上位群と下位群の比較

生成された分割位置ベクトル X' のうち，距離 D が小さい（類似度が高い）群を上位群，距離 D が大きい（類似度が低い）群を下位群とした．分割位置を修正する前の bin 間の距離（表 2）と，上位群の bin 間の距離（表 3），下位群の bin 間の距離（表 4）をそれぞれ示す．また，上位群と下位群で bin の分割位置に有意な差があるかを検証するためにウェルチの t 検定を行った．さらに，特定の分割位置同士の相関を求めその関係を調べた．同じ構造を持つ bin の分割位置 x'_i，x'_j と 2 つの bin の特徴量の距離 $d(f_i, f_j)$ を 3 次元空間上に全てのデータ分プロットしたものを図 7，8，9，10 に示す．

分割位置を修正する前の bin 間の距離を表 2 に，上位群から上位 5 つのデータを抜粋したものを表 3 に，下位群から下位 5 つのデータを抜粋したものを表 4 に示す．上位群の bin 間の距離は 2 番目の $d(f_3, f_{11})$ と 4 番目の $d(f_3, f_{11})$ を除いてどれも修正前よりも小さくなっている．また，下位群の bin 間の距離は $d(f_3, f_{11})$ のみ小さくなっている．

上位群と下位群の分割位置のヒストグラムを図 6 上に示す．図 6 下には分割位置 x'_1 から x'_4 のヒストグラムを拡大表示した．青が上位群の分割位置のヒストグラムであり，赤が下位群の分割位置のヒストグラムである．上位群と下位群で分割位置に関して t 検定を行った結果を表 5 に示す．両群での分割位置をウェルチの t 検定により比較したところ，分割位置 x'_1，x'_3，x'_4，x'_5，x'_{11}，x'_{13} において有意差が認められた（$p<0.05$）．

特定のペアの分割位置の上位群のデータの相関，下位群のデータの相関をそれぞれ表 6 に示す．上位群では x'_2 と x'_{10} 間と x'_3 と x'_{11} 間にやや正の相関が認められた（$r = .42$，$r = .52$）が，下位群では x'_2 と x'_{10} 間にやや負の相関が認められた（$r=.49$）．全ての分割位置ベクトル X' において，同じ構造を持つ bin の分割位置 x'_i，x'_j と 2 つの bin の特徴量の距離 $d(f_i, f_j)$ を 3 次元空間上にプロットしたものを図 7，8，9，10 に示す．

4. 考察

以下に実験 1 に対する考察を述べる．修正前の分割位置では b_{12} と b_{13} が初期の段階で併合されてしまったために，

Vol.2018-MUS-118 No.20
Vol.2018-SLP-120 No.20
2018/2/21

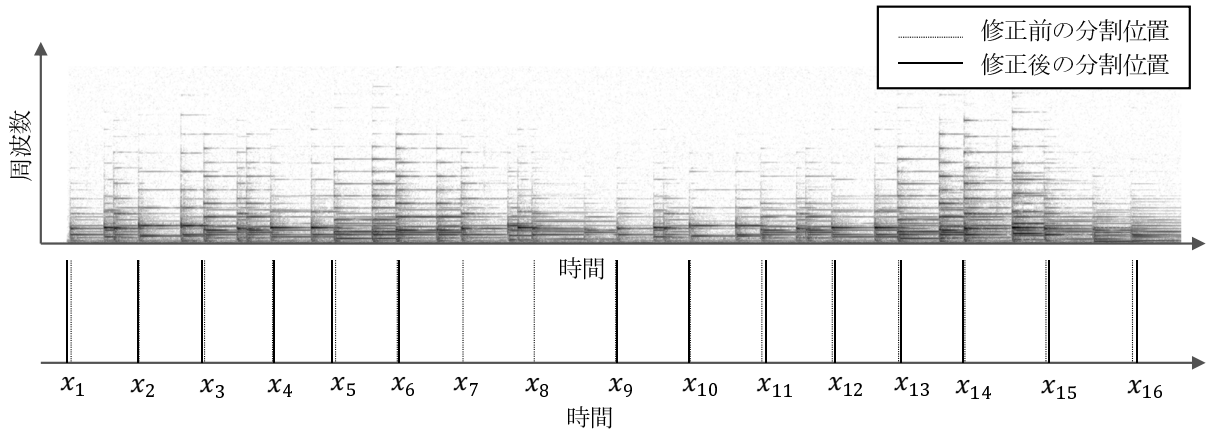

図 4 スペクトログラムの分割位置の修正前と修正後

表 1 スペクトログラムの分割位置の修正前と修正後の比較

分割位置	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
修正前 [ms]	650	2130	3550	5010	6380	7730	9140	10680	12450	14070	15620	17150	18590	20030	21780	23660
修正後 [ms]	560	2110	3490	5050	6310	7750	9150	10680	12480	14060	15700	17210	18640	19990	21840	23750
Δt[ms]	-90	-20	-60	+40	-70	+20	+10	0	+30	-10	+80	+60	+50	-40	+60	+90

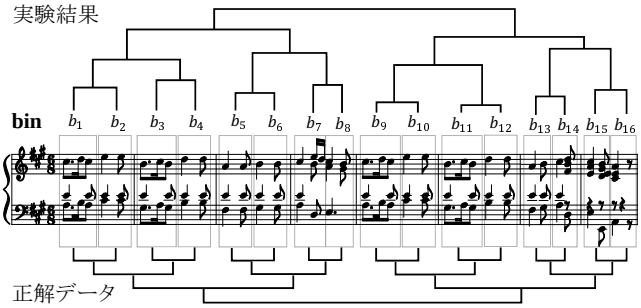

図 5 bin の分割位置を修正したタイムスパンセグメンテーション

表 2 修正前の分割位置での bin 間の距離

$d(f_1$ と $f_9)$	$d(f_2$ と $f_{10})$	$d(f_3$ と $f_{11})$	$d(f_4$ と $f_{12})$	合計（D）
1.83	2.43	2.91	0.90	8.08

表 3 修正後の分割位置での bin 間の距離（上位群の上位 5 個）

$d(f_1$ と $f_9)$	$d(f_2$ と $f_{10})$	$d(f_3$ と $f_{11})$	$d(f_4$ と $f_{12})$	合計（D）
1.28	2.22	2.84	0.68	7.03
1.29	2.26	3.00	0.57	7.13
1.24	2.34	2.88	0.67	7.13
1.38	2.28	3.03	0.44	7.14
1.46	2.36	2.87	0.46	7.14

表 4 修正後の分割位置での bin 間の距離（下位群の下位 5 個）

$d(f_1$ と $f_9)$	$d(f_2$ と $f_{10})$	$d(f_3$ と $f_{11})$	$d(f_4$ と $f_{12})$	合計（D）
1.96	2.77	2.71	1.51	8.95
2.08	2.68	2.76	1.39	8.91
2.06	2.48	2.70	1.67	8.90
2.02	2.74	2.67	1.44	8.87
1.77	2.87	2.69	1.43	8.77

その誤りが上位のレベルまで伝播してしまった．今回 b_4 と b_{12} の類似度が高くなるように分割位置を変更した際に，条件には入っていなかったにも関わらず $d(f_5,\ f_6)$ や $d(f_{13},\ f_{14})$ が適切に小さくなっていた．

表 5 t 検定

bin	t 値	検定
x'_1	-8.09	p<.01
x'_2	-1.27	n.s.
x'_3	-7.45	p<.01
x'_4	5.90	p<.01
x'_5	-13.18	p<.01
x'_9	1.82	n.s.
x'_{10}	-0.65	n.s.
x'_{11}	2.19	p<.05
x'_{12}	-0.23	n.s.
x'_{13}	13.13	p<.01

表 6 分割位置の相関

	上位 15 の相関	下位 15 の相関
x'_1 と x'_9 間	0.10	-0.08
x'_2 と x'_{10} 間	0.42	-0.49
x'_3 と x'_{11} 間	0.52	0.26
x'_4 と x'_{12} 間	0.16	0.06

以下に実験 2 に対する考察を述べる．上位群と下位群で有意な差が得られた．分割位置 x'_2 と x'_{10} は有意な差が得られなかったが，上位群と下位群のそれぞれの相関を見てみると，上位群は正の相関が認められ，下位群は負の相関が認められた．これは分割位置の絶対的な位置が重要なのではなく，他の分割位置との相対的な関係が重要であることを示唆している．また，上位群の bin 間の距離は $d(f_3, f_{11})$ の 2 番目と 4 番目が修正前よりも大きくなっている．それに対応して $d(f_4, f_{12})$ が小さくなっている．b_3 と b_4 間の分割位置 x'_4 が動くことで，片方の類似度が上がるともう片方の類似度が下がるトレードオフの関係になっていると考えられる．x'_{12} についても同様だと考えられる．これは分割位置の調節の限界であるともいえる．

Vol.2018-MUS-118 No.20
Vol.2018-SLP-120 No.20
2018/2/21

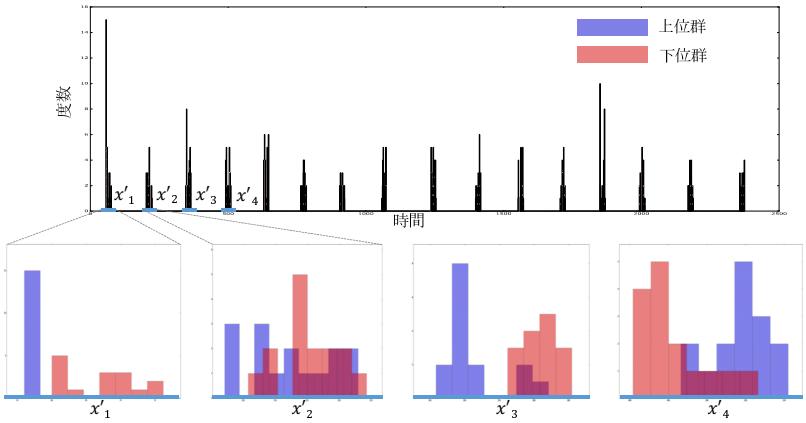

図 6 上位群（青）と下位群（赤）の分割位置のヒストグラム

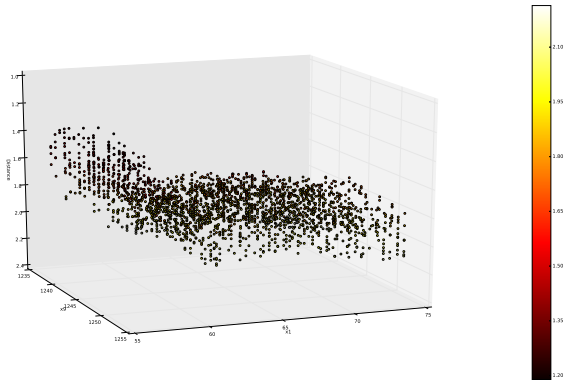

図 7 分割位置 x_1，x_9 と $d(f_1,\ f_9)$ の関係

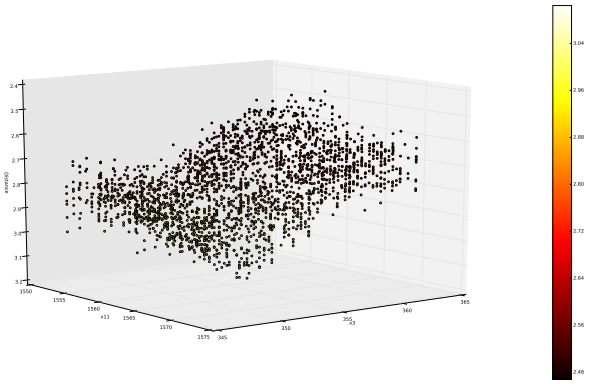

図 9 分割位置 x_3，x_{11} と $d(f_3,\ f_{11})$ の関係

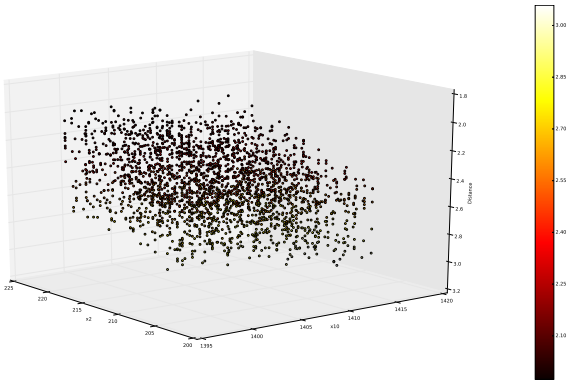

図 8 分割位置 x_2，x_{10} と $d(f_2,\ f_{10})$ の関係

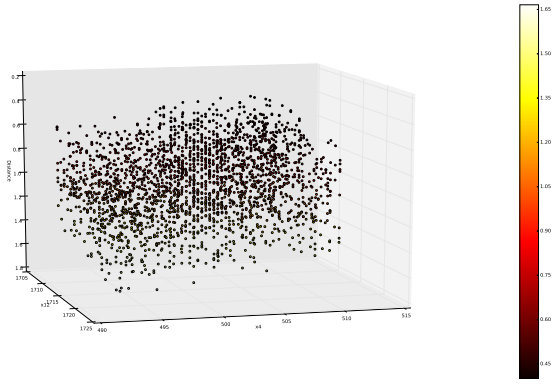

図 10 分割位置 x_4，x_{12} と $d(f_4,\ f_{12})$ の関係

5. おわりに

GTTM を音楽のスペクトログラムに直接適用してタイムスパン・セグメンテーションを生成する従来の方法には，スペクトログラムの分割する位置が適切でないことによって期待する結果が得られない問題があった．本稿では，スペクトログラムを分割する段階に大域的な構造をフィードバックすることで適切な分割位置を獲得する枠組みを提案し，その有用性を検証した．その結果，スペクトログラムの分割位置によって精度が変わり，適切な分割位置では期待するタイムスパン・セグメンテーションが得られ，大域的構造を考慮した分割位置の変更が有用であることが示された．これは記号から音響信号へのフィードバックの枠組みであると言える．今後は分割位置を自動で調節する枠組みを構築する必要がある．

謝辞 研究を通じて議論をしていただいた寺井あすか先生（公立はこだて未来大学），浜中雅俊先生（理化学研究所）に感謝いたします．本研究は JSPS 科研費 16H01744，26280089 の助成を受けたものです．

参考文献

[1] 足立亜里紗，堀内靖雄，黒岩眞吾：独奏認識誤りに頑健な音響入力伴奏システム，研究報告音楽情報科学 (MUS)，Vol.2017-MUS-114 No.1，pp. 1-5 (2017).
[2] Dibben, N.: Cognitive Reality of Hierarchic Structure in Tonal and Atonal Music, *Music Perception: An Interdisciplinary Journal*, vol.12 No.1, pp.1–25 (1994).
[3] Ellis, D. P., Poliner, G. E.: Identifyingcover songs' with chroma features and dynamic programming beat tracking. In Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on (Vol. 4, pp. IV-1429). IEEE.
[4] Hamanaka, M., Hirata, K. and Tojo, S.: Implementing "A Generative Theory of Tonal Music", *Journal of New Music Research*, 35:4, pp.249–277 (2007).
[5] Hamanaka, M., Hirata, K. and Tojo, S.: Implementing Methods for Analysing Music Based on Lerdahl and Jackendoff's Generative Theory of Tonal Music, David Meredith (Ed), *Computational Music Analysis*, Chapter 9, pp.221-249, Springer (2016).
[6] 浜中雅俊: Interactive GTTM Analyzer / GTTM Database Download Page, http://gttm.jp/gttm/ja/database/, (2017).
[7] Haralick, M. R.: Statistical and structural approaches to texture, In *Proc. of the IEEE*, vol.67, No.5, pp.786–804 (1979).
[8] Lerdahl, F. and Jackendoff, R.: *A Generative Theory of Tonal Music*, The MIT Press (1983).
[9] McFee, B. and Ellis, P. W. D.: Learning to Segment Songs with Ordinal Linear Discriminant Analysis, In *Proc. of ICASSP* (2014).
[10] 中村友彦，水野 優，鈴木孝輔，中村栄太，樋口祐介，深山 覚，嵯峨山茂樹：音楽演奏の誤りや反復に頑健な音響入力自動伴奏，日本音響学会 2012 年秋季研究発表会，pp.931-934，2012
[11] Nakashika, T., Garcia, C. and Takiguchi, T.: Local-feature-map Integration Using Convolutional Neural Networks for Music Genre Classification, In *Proc. of Interspeech*, pp.1752–1755, ISCA (2012).
[12] 澤田隼，竹川佳成，平田圭二：音楽音響信号を対象とするGTTM 的アプローチによるグルーピング構造の抽出について，研究報告音楽情報科学 (MUS) Vol.2016-MUS-111，No.23，pp.1-6（2016）.
[13] 澤田隼，竹川佳成，平田圭二：スペクトログラムの階層的クラスタリングを用いたグルーピング構造分析について，研究報告音楽情報科学 (MUS) Vol.2017-MUS-114，No.7，pp.1-8（2017）.
[14] 鈴木孝輔，上田雄，齋藤康之，小野順貴，嵯峨山茂樹：HMM を用いた音響演奏の楽譜追跡による弾き直しに追従可能な自動伴奏，研究報告音楽情報科学 (MUS)，Vol.2011-MUS-89 No.29, pp. 1-6 (2011).
[15] Ullrich, K., Schlüter, J. and Grill, T.: Boundary Detection in Music Structure Analysis using Convolutional Neural Networks, In *Proc. of ISMIR*, pp.417–422 (2014).
[16] Costa, Y. M., Oliveira, L. S. and Koerich, A. L., et al. : Comparing textural features for music genre classification, The 2012 International Joint Conference on Neural Networks，pp.1-6 (2012).