

Instrogramを用いた楽器構成に基づく類似楽曲検索*

北原 鉄朗 (京大), 後藤 真孝 (産総研),
駒谷 和範, 尾形 哲也, 奥乃 博 (京大)

1 はじめに

類似楽曲検索は、内容に基づく音楽推薦やプレイリスト自動生成などの礎となる技術である。類似楽曲検索の研究において重要なことは、楽曲間の類似度をどのような音楽要素に着目して設計するかである。本研究では、楽器構成に基づいた楽曲間類似度を検討する「ピアノソナタ」「弦楽四重奏」などのような分類が用いられることから示唆されるように、「どのような楽器で演奏されているか」は、その楽曲の雰囲気の特徴づける重要な要素である。そのため、似た雰囲気の楽曲を集めたプレイリストを生成するといった場面で、楽器構成に基づく類似楽曲検索は重要な役割を担うと考えられる。

楽器構成に基づく類似楽曲検索を実現する方法として、2つの方法が考えられる。1つは、楽器同定手法を用いて楽器構成を同定した後、同定結果同士のマッチングを取る方法である。しかし、混合音から楽器同定を正しく行うのは難しく、実演奏に対して十分な精度で楽器同定が実現された事例はほとんどない。もう1つは、楽器構成に関連の高い特徴量を抽出し、この特徴空間上で距離を測る方法である。実際、従来研究において、メル周波数ケプストラム係数 (MFCC) などが音色類似度として用いられた [1]。このような特徴量は比較的簡単な処理で求めることができ、一定の成功を収めているが、必ずしも楽器の音色を鮮明に反映するものではなく、和音のポインティングを始めとする編曲などの他の音楽的要素からも影響を受ける。

本研究では、上のどちらでもなく、我々が以前提案した Instrogram [2, 3] の類似度尺度を設計することで、楽器構成に基づく類似楽曲検索を実現する。Instrogram は、非決定論的に楽器構成を表現する方法で、楽器存在確率の時間・周波数表現として与えられる。Instrogram では楽器構成をシンボル化して表さないで、一致する/しないといったマッチングにとどまらない連続的な類似度尺度を提供できるとともに、MFCC などに比べて、楽器構成を確率として直接的に表すため、実際の楽器構成をより反映した類似度計算が期待できる。つまり、本アプローチは上記2つの中間を狙ったものとも考えることもできる。

2 Instrogram

Instrogram [2, 3] は、スペクトログラムに似た楽器存在確率の視覚表現である。解析対象となる楽器ごとに1つの画像が存在し、画像の色の強さによってその楽器が存在する確率を表す。各画像は横軸が時刻、縦軸が基本周波数 (F0) を表し、対象楽器を $\Omega = \{\omega_1, \dots, \omega_m\}$ とすると、 i 番目の画像の各ピクセル (t, f) の色の強さが、時刻 t において周波数 f を F0 とする楽器 ω_i の音が存在する確率 $p(\omega_i; t, f)$ を表す。ただし、F0 が同じ楽器音が同時に複数あることはないとは仮定している。Fig. 1 に例を示す。これは、ピアノ、

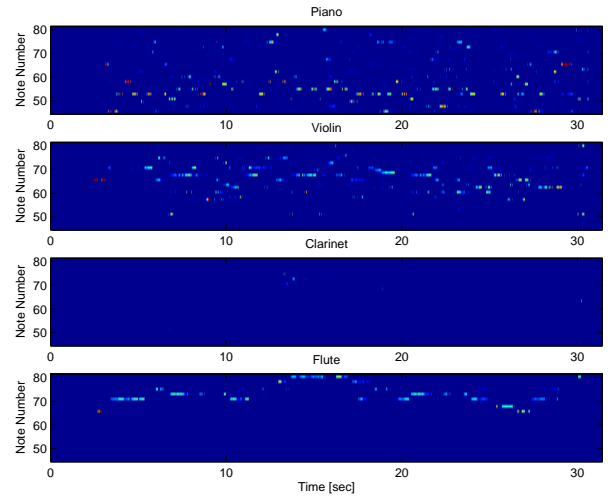


Fig. 1 Instrogram の例 (ピアノ, バイオリン, フルートによる「蛍の光」の三重奏)

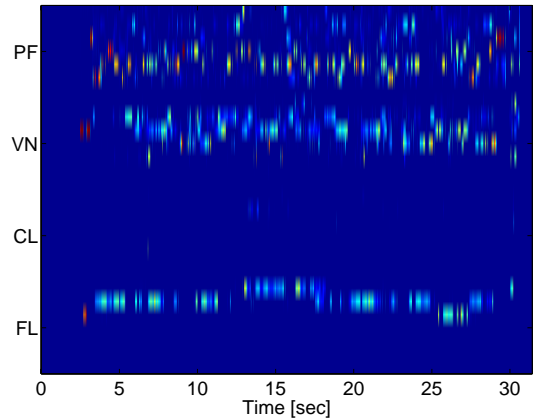


Fig. 2 Fig. 1 の簡略版 (低周波数分解版)

バイオリン, フルートによる「蛍の光」の三重奏を、ピアノ, バイオリン, クラリネット, フルートを対象に Instrogram を作成したものである。ここで、時間分解能は 10ms, 周波数分解能は 100cent とした。高い周波数分解能が要らない場合は、周波数軸をいくつかの区間に分割して区間内の値をマージすることで周波数分解能を粗くすることもできる。全周波数区間を N 個の区間 I_1, \dots, I_N に分割したとき、 k 番目の周波数区間 I_k の楽器存在確率 $p(\omega_i; t, I_k)$ は $p(\omega_i; t, \bigcup_{f \in I_k} f)$ と定義する。簡略化された Instrogram を Fig. 2 に示す。Fig. 1 あるいは Fig. 2 より、この楽曲は高音部はフルート、中音部はバイオリン、低音部はピアノによる演奏であることがわかる。

3 Instrogram を用いた楽器構成類似度

楽器存在確率を用いて楽曲間の類似度を計算する方法を述べる。ここでは、類似度の代わりに、Instrogram 間の距離 (非類似度) を DP マッチング (DTW) で求める。

* Music Information Retrieval based on Instrumentation Similarity Using Instrogram.
by Tetsuro Kitahara (Kyoto University), Masataka Goto (AIST), Kazunori Komatani, Tetsuya Ogata,
and Hiroshi G. Okuno (Kyoto University)

Table 1 楽曲間の非類似度計算結果．C12 などの C/J は Classical/Jazz，番号は楽曲番号．
(a) Instrogram (楽器存在確率)を用いた場合

	(i)				(ii)		(iii)	(iv)			3-best-similarity pieces
	C12	C14	C21	C38	C19	C40	C43	J01	J02	J03	
C12	0										C21, C14, C38
C14	6429	0									C21, C12, C38
C21	5756	5734	0								C14, C12, C38
C38	7073	6553	6411	0							C21, C14, C38
C19	7320	8181	7274	7993	0						C21, C12, C38
C40	8650	8353	8430	8290	8430	0					J02, J01, C43
C43	8910	9635	9495	9729	8148	8235	0				J01, J02, J03
J01	9711	10226	10252	10324	8305	8214	6934	0			J02, J03, C43
J02	9856	10125	10033	10610	8228	8139	7216	6397	0		J01, C43, J03
J03	9134	9136	8894	9376	8058	8327	7480	6911	7223	0	J01, J02, C43

(b) MFCC を用いた場合

	(i)				(ii)		(iii)	(iv)			3-best-similarity pieces
	C12	C14	C21	C38	C19	C40	C43	J01	J02	J03	
C12	0										C21, C40, J02
C14	17733	0									C43, C12, J02
C21	17194	18134	0								C12, J01, J02
C38	18500	18426	18061	0							J01, J02, C21
C19	17510	18759	18222	19009	0						J02, C12, J03
C40	17417	19011	18189	19099	18100	0					C12, J02, J01
C43	18338	17459	17728	18098	18746	18456	0				J01, C14, J02
J01	17657	17791	17284	17834	18133	17983	16762	0			J02, C43, J03
J02	17484	17776	17359	18009	17415	17524	17585	15870	0		J01, J03, C21
J03	17799	18063	17591	18135	17814	18038	17792	16828	16987	0	J01, J02, C21

1. 各時刻 t に対して，ベクトル p_t を全楽器存在確率の連結として求める：

$$p_t = (p(\omega_1; t, I_1), p(\omega_1; t, I_2), \dots, p(\omega_m; t, I_N))'$$

ここで，' は転置を表す．

2. 2 つのベクトル p, q 間の距離をコサイン距離として定義する：

$$\text{dist}(p, q) = 1 - (p, q) / (\|p\| \cdot \|q\|)$$

ただし， $(p, q) = p' R q$ ， $\|p\| = \sqrt{(p, p)}$ ． $R = (r_{ij})$ は正定値行列で，要素間の関係を定義するものである．たとえば，ある 2 曲において同じ楽器が異なる音域で演奏されている場合，あるいは同じ楽器族だけ異なる楽器で演奏されている場合，この 2 曲には高い類似度を与えたい場合があるであろう．そのような場合，対応する r_{ij} に 0 より大きい値を与えることで，こういった要素間の関係を考慮することができる． R が単位行列のとき， (p, q) および $\|p\|$ は通常の内積およびノルムとなる．

3. 上で定義した距離尺度を利用して， $\{p_t\}, \{q_t\}$ 間の距離（非類似度）を DTW で計算する．

4 実験

RWC 研究用音楽データベース（クラシック，ジャズ）[4] から選んだ比較的小編成の 10 曲の冒頭 1 分に対して，楽器構成の類似度計算の実験を行った．使用した 10 曲の楽器構成は Table 2 に示す． R は単位行列を用いた．また，比較のために楽器存在確率の代わりに MFCC（12 次元）を用いた場合も行った．

結果を Table 1 に示す．Instrogram を用いた場合，グループ内の非類似度は多くが 7000 以下となり，グループ (i)-(iv) 間の非類似度はほとんどが 9000 以上となるなど，おおむね実際の楽器構成を反映した非類似度を得ることができた．

Instrogram と MFCC とでは，前者の方がグループ内/グループ間の非類似度にはっきりとした差があっ

Table 2 実験で用いた楽曲およびその楽器構成

	(i) No. 12, 14, 21, 38	Strings
Classical	(ii) No. 19, 40	Piano+Strings
	(iii) No. 43	Piano+Flute
Jazz	(iv) No. 1, 2, 3	Piano solo

た．実際，弦楽器による楽曲（グループ (i)）と弦楽器を含まない楽曲（グループ (iii) および (iv)）に対して類似楽曲ベスト 3 を求めると，Instrogram の場合では，前者はすべてグループ (i) の楽曲が選ばれ，後者はすべて弦楽器を含まない楽曲が選ばれたのに対し，MFCC ではそうでないことがあった．

以上の結果に基づき，楽器構成に基づいて類似楽曲検索をするプロトタイプシステムを構築した¹．ユーザが指定した楽曲との類似度が高い順にデータベース中の楽曲を提示し，その中から 1 曲選ぶと，楽器存在確率をリアルタイムに表示しながら再生する．

5 おわりに

本稿では，楽器構成の確率的表現法 Instrogram 間の類似度を定義し，楽器構成に基づく類似楽曲検索を実現した．今後は，メロディやリズムなどの他の音楽要素を取り入れ，総合的な類似楽曲検索を目指す．

謝辞 本研究の一部は，日本学術振興会科学研究費補助金，21 世紀 COE プログラム，CREST の支援を受けた．

参考文献

- [1] J.-J. Aucouturier and F. Pachet: Music Similarity Measure: What's the Use? Proc. ISMIR, pp.157-163, 2002.
- [2] 北原 他: Instrogram: 楽器存在確率に基づく音楽視覚表現法, 音講論集(春), 2-2-12, 2006.
- [3] T. Kitahara et al.: Instrogram: A New Musical Instrument Recognition Technique Without Using Onset Detection Nor F0 Estimation, Proc. ICASSP, V, pp.229-232, 2006.
- [4] 後藤 他: RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース, 情報学論, 45, 3, pp.728-738, 2004.

¹<http://winnie.kuis.kyoto-u.ac.jp/~kitahara/instrogram/>