

混合音からの特徴量テンプレート作成と 音楽的文脈の利用による多重奏の音源同定*

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

楽器音の音源同定は、自動採譜や音楽情報検索などにおいて重要なタスクである。しかし、これまでの音源同定研究の多くは単一音を対象としており (e.g. [1]), 多重奏への取り組みが始まったのは最近である [2, 3]。

我々は、多重奏に対して高精度な音源同定を実現する上での課題として、「音の重なりによる特徴変動」「音高による音色変化」「音楽的文脈」を指摘し、これらを解決する手法を提案した [4, 5]。本稿では、この手法を述べるとともに、文献 [4, 5] では紙面の制約から報告できなかった、より大規模な実験結果について報告する。

2 多重奏に対する音源同定手法

2.1 全体像

本研究で提案する多重奏に対する音源同定手法の全体像を Fig. 1 に示す。音楽の音響信号が与えられると、まず短時間フーリエ変換を行ってスペクトログラムを作成する。その後、各単音 n_k の発音時刻・音高・音長を推定し (ただし本実験では正解を付与), それに基づいて各単音の調波構造を抽出する。次に、調波構造から音響的特徴量を抽出する。抽出する特徴量は、我々が以前提案した 129 個の特徴量 [6] から混合音では抽出が難しいものを除いた 43 個である。そして、次元圧縮により 4 次元特徴ベクトル x_k に変換した後、各楽器 ω_i に対する事後確率 $p(\omega_i | x_k)$ (特徴ベクトル x_k が楽器 ω_i である確率) を求める。事後確率の計算は、あらかじめ収集した各楽器の特徴ベクトルのデータベースとの比較に基づいて行われる。この比較用データベースを特徴量テンプレートと呼ぶ。最後に事後確率が最大になる楽器名を同定結果として出力する。

2.2 音の重なりによる特徴変動

音の重なりによる特徴変動は、多重奏の音源同定において最も重要な問題である。複数の楽器が同時に発音すると、いくつかの周波数成分が重複し、干渉を引き起こす。そのため、干渉する周波数成分から抽出した特徴量が大きく変動する。

本研究では、学習に用いる特徴量テンプレートを混合音から作成することでこの問題を解決する。これを混合音テンプレート [4, 5] と呼ぶ。特徴量テンプレートを混合音から作ることで、すでに音の重なりの影響を受けているデータで学習することになり、混合音に対するロバストな同定を期待できる。

しかし、混合音の組み合わせは非常に多いため、すべての組み合わせを網羅的に収集するのは現実的には不可能である。そこで本研究では、実際の楽曲の楽譜から混合音を作成することで、現実の楽曲で出現され得る混合音の組み合わせのみを重点的に収集する。

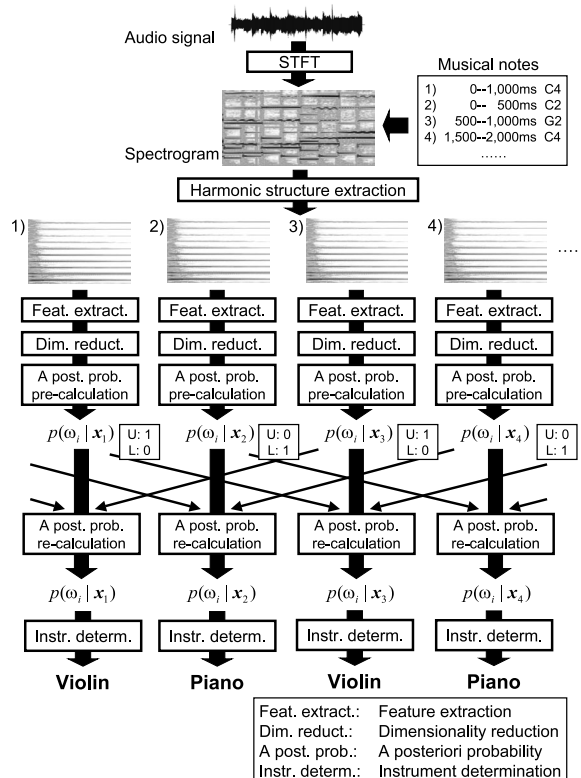


Fig. 1 多重奏に対する音源同定手法の全体像

2.3 音高による音色変化

楽器音は音声などの他の音に比べ広音域のため、音高による音色の違いが顕著に現れる。そのため、特にピアノなどの広音域の楽器では、音高による変化を考慮した音色のモデル化が重要である。本研究では、音高による変化を考慮した音色のモデル化として、我々が以前提案した「F0 依存多次元正規分布」 [6, 7] を用いる。これは、楽器の音色を表す音響的特徴の各々が音高 (基本周波数, F0) によって変化する様子を F0 の関数として近似するものである。

2.4 音楽的文脈の利用

実楽曲に対して音源同定を行う際には、音楽的文脈 (前後の単音に関する情報) を利用することが重要である。なぜなら、個々の単音に対して個別に同定処理を行うと、たとえばフルートの旋律のなかで 1 音だけクラリネットが現れるといった音楽的に不自然な誤りが発生し、音楽的文脈を利用することでこれを防ぐことが可能だからである。本研究では、単音 n_k の事後確率 $p(\omega_i | x_k) = p(x_k | \omega_i) p(\omega_i) / p(x)$ の計算時に、事前確率 $p(\omega_i)$ に前後のいくつかの単音の事後確率を反映させることで、音楽的文脈を考慮した同定を実現する [4, 5]。

* Instrument Identification in Polyphonic Music by Constructing Feature Vector Template from Polyphonic Sounds and Using Musical Context. by Tetsuro Kitahara (Kyoto University), Masataka Goto (AIST), Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno (Kyoto University)

Table 1 実験結果 (認識率) [単位: %] (Leave-one-out 法利用. 太字は 75%以上を表す)

テンプレート 音高依存 文脈		単一音のみ				単旋律+二重奏				単旋律+二重奏+三重奏			
		×		○		×		○		×		○	
		×	○	×	○	×	○	×	○	×	○	×	○
二重奏	PF	53.2	63.3	71.7	85.1	62.8	65.7	68.9	79.1	70.3	72.6	70.9	82.0
	CG	51.4	50.8	58.1	58.8	50.1	67.2	68.8	84.7	43.3	56.4	71.3	82.7
	VN	58.9	77.5	63.3	76.5	68.1	85.5	66.6	87.7	65.4	84.2	67.7	88.1
	CL	61.5	69.7	52.9	55.7	81.8	92.1	80.4	89.9	84.6	95.1	82.9	92.6
	FL	33.2	48.5	37.2	42.4	67.6	84.9	68.9	78.8	56.8	70.5	61.5	74.3
	平均	51.6	62.0	56.6	63.7	66.1	79.1	70.7	84.0	64.1	75.8	70.9	84.0
三重奏	PF	41.9	49.3	63.3	74.1	44.1	43.8	57.0	61.4	52.4	53.6	61.5	68.3
	CG	45.3	44.5	47.4	46.1	52.1	66.8	68.3	82.0	47.2	62.8	68.3	82.8
	VN	57.2	72.6	63.1	75.0	67.0	81.8	70.8	83.5	60.5	80.6	68.1	82.5
	CL	50.9	53.4	44.4	43.5	69.5	77.1	72.2	78.3	71.0	82.8	76.2	82.8
	FL	39.4	51.4	42.1	48.3	68.4	77.9	68.1	76.9	59.1	69.3	64.0	71.5
	平均	46.9	54.3	52.1	57.4	60.2	69.5	67.3	76.4	58.0	69.8	67.6	77.6
四重奏	PF	37.7	45.8	53.8	63.9	38.7	38.6	50.3	53.1	46.1	46.6	53.3	57.2
	CG	39.0	37.5	41.7	40.8	51.2	62.7	64.8	75.3	51.2	64.5	65.0	79.1
	VN	57.3	71.2	64.0	75.5	70.0	81.2	72.7	82.3	67.4	79.2	69.7	79.9
	CL	43.8	42.4	38.7	35.3	62.6	66.8	65.4	69.3	68.6	74.4	70.9	74.5
	FL	42.2	55.0	45.9	54.0	69.8	76.1	69.9	76.2	61.7	69.4	64.5	70.9
	平均	44.0	50.4	48.8	53.9	58.5	65.1	64.6	71.2	59.0	66.8	64.7	72.3

3 評価実験

RWC 研究用音楽データベース(楽器音)[8]の音響信号(ピアノ(PF),クラシックギター(CG),バイオリン(VN),クラリネット(CL),フルート(FL))を標準MIDIファイル(SMF)に従って切り貼りして作成した二重奏・三重奏・四重奏の音響信号に対して実験を行った。SMFにはRWC研究用音楽データベース(クラシック)[8]のNo.13, 16, 17から2~4パートを抜粋して使用し,楽器の組み合わせは音域的に可能な範囲で総当たりとした。混合音テンプレートは,認識対象楽曲以外の2曲を用いて作成した(leave-one-out法)。

実験結果をTable 1に示す!混合音テンプレート,「F0依存多次元正規分布」「音楽的文脈」により,二重奏は51.6%から84.0%へ,三重奏は46.9%から77.6%へ,四重奏は44.0%から72.3%へ認識率を改善することができた。また「単旋律+二重奏」のテンプレート使用時と「単旋律+二重奏+三重奏」のテンプレート使用時では結果に大きな差はなかった。特に四重奏に対しても「単旋律+二重奏」のテンプレートだけで単一音のテンプレートに比べて高い認識率を示した。これは,必ずしも認識対象と同程度の複雑さの楽曲でなくても,同時発音する他の音からの影響を受けているデータであれば,認識率の改善に貢献できることを示唆している。5楽器のうちPFはあまり認識率が高くない場合が多かった。これは,CGと音色が似ているからと考えられる。実際,両者の調波構造を抽出し再合成したものを試聴したところ,区別の難しいものが多かった。

次に,混合音テンプレートを1曲のみから作成した場合の実験結果をTable 2に示す。1曲のみからの作成でもCG, VN, CLは比較的高い認識率であった。FLは作成用楽曲によっては30%代の認識率であったが,2曲用いて作成することで低い認識率が解消された。

4 おわりに

本稿では,多重奏に対する高精度な音源同定実現のため,混合音テンプレート, F0依存多次元正規分布,音楽的文脈の利用を検討し,実験により認識率向上を確認した。

Table 2 混合音テンプレートを1曲のみから作成した場合の実験結果(認識率,四重奏)[単位: %]

		単旋律+二重奏				単旋律+二重奏+三重奏			
		13	16	17	*	13	16	17	*
No. 13	PF	(57.8)	32.3	38.4	36.6	(67.2)	33.2	45.1	39.7
	CG	(73.3)	78.1	76.2	76.7	(76.8)	84.3	80.3	82.1
	VN	(89.5)	59.4	87.5	86.2	(87.2)	58.0	85.2	83.1
	CL	(68.5)	70.8	62.2	73.8	(72.3)	72.3	68.6	75.9
	FL	(85.5)	40.2	74.9	82.7	(86.0)	38.9	68.8	80.8
No. 16	PF	74.1	(64.8)	61.1	71.2	79.6	(67.1)	73.0	78.3
	CG	79.2	(77.9)	78.9	74.3	70.4	(82.6)	74.0	75.2
	VN	89.2	(85.5)	87.0	87.0	86.0	(83.5)	84.7	85.0
	CL	68.1	(78.9)	68.9	76.1	72.4	(82.8)	76.3	82.1
	FL	82.0	(75.9)	72.5	77.3	77.9	(72.3)	35.7	69.2
No. 17	PF	53.0	39.4	(51.2)	51.6	52.2	40.6	(55.7)	53.7
	CG	73.7	69.0	(75.8)	75.0	76.0	74.3	(78.4)	80.0
	VN	79.5	61.2	(78.3)	73.6	77.4	58.0	(78.7)	71.7
	CL	51.3	60.5	(57.1)	57.9	61.1	62.6	(66.9)	65.4
	FL	65.0	35.0	(73.1)	68.7	58.6	34.7	(70.9)	62.6

*Leave-one-out 法の場合(参考) 括弧付は closed 実験 F0 依存多次元正規分布,音楽的文脈はともに使用 「単旋律+二重奏 13」は,楽曲 No.13 の SMF から抜粋した単旋律と二重奏のみを用いたテンプレート作成の意。

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プログラムの支援を受けた。

参考文献

- [1] K. D. Martin: Sound-Source Recognition: A Theory and Computational Model, PhD Thesis, MIT, 1999.
- [2] K. Kashino et al.: A Sound Source Identification System for Ensemble Music based on Template Adaptation and Music Stream Extraction, Speech Comm., **27**, pp.337-349, 1999.
- [3] J. Eggink et al.: A Missing Feature Approach to Instrument Identification in Polyphonic Music, Proc. ICASSP, **V**, pp.553-556, 2003.
- [4] 北原 他: 多重奏の音源同定のための混合音からのテンプレート作成法, 第 67 回情処全大, 3G-4, 2005.
- [5] T. Kitahara et al.: Instrument Identification in Polyphonic Music: Feature Weighting with Mixed Sounds, Pitch-dependent Timbre Modeling, and Use of Musical Context, Proc. ISMIR, 2005.
- [6] 北原 他: 音高による音色変化に着目した楽器音の音源同定: F0 依存多次元正規分布に基づく識別手法, 情処学論, **44**, 10, pp.2448-2458, 2003.
- [7] 北原 他: 音色空間の音高依存性を考慮した楽器音の音源同定, 音講論集(秋), 1-1-4, pp.643-644, 2002.
- [8] 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, 情処学論, **45**, 3, pp.728-738, 2004.