

# 音楽的ストリーム構造を用いたパート分離\*

大津 幹弘, 坂井 修一, 田中 英彦

{mikihiro,sakai,tanaka}@mtl.t.u-tokyo.ac.jp

東京大学大学院 情報理工学系研究科

## 1 はじめに

多種多様な楽曲に対して人間が採譜を行う場合、絶対音感などの稀有な聴覚能力や、音楽に対する豊富な知識とセンスが要求される。よって採譜は、限られたスペシャリストによる特殊技能というのが現状である。一方、スコア作成だけでなく、近年需要が増加している通信カラオケの MIDI データや携帯電話の着信メロディ作成にも採譜技術が必要であり、短時間で効率的に採譜を行う技術に対する需要は高い。そこで、採譜を計算機によって短時間で自動的に行う自動採譜に関して、これまでいくつかの研究が行われてきたが、音の倍音構造や音の重なりによる干渉などの信号的な問題や、アートを論理的に記述することの難しさなど様々な問題があり、現在のところどれも実用レベルには達していない。従来の研究の主な問題点として、以下の点が挙げられる。

- 多くの研究は信号処理に重みが置かれており、音楽学的なアプローチが少ない。
- 音を短いタイムスパンで処理する研究が多く、音の流れの特性を解析に用いていない。
- 採譜の完全な自動化を目的とするシステムが多く、マンマシン協調の可能性が試されていない。
- 評価に使われる曲がシンプルなものが多く、実際に自動採譜が活躍出来る分野での評価が難しい。

以上の問題点を踏まえて、本研究では、以下の4つのコンセプトから解析を行う。

- アンサンブルの特性を分析し、より音楽的なアプローチを行う。
- 従来のように1音ずつ処理するのではなく、楽曲のストリーム構造を用いて、音の流れを処理する。
- ユーザが簡単なヒントを入力することで、マンマシン協調を実現する。
- システムの評価に邦楽、洋楽、ジャンルを問わずに多種多様な楽曲を用いる。

なお本研究では信号的な処理は行わず、MIDI データから得られる単音列を、各パートに分離する精度、すなわち音源分離精度を向上させることを目的とする。

## 2 音楽的ストリーム構造を用いたパート分離

### 2.1 楽曲のストリーム構造

音楽的ストリームとは、ある特徴を持った音の一連の流れのことである [1]。人間がひとまとまりと知覚する音の流れをストリームとした研究 [2] もあるが、本研究では、アンサンブルにおいて1つの楽器が担当するパートをストリームと呼ぶことにする。すなわち楽曲は、ある役割を持ったストリームの集合で構成されていることになり、その役割や特徴をパート分離に用いる。

### 2.2 ストリーム判定特徴量

ストリーム判定に、Pitch(音高変化量), Velocity(音量変化量), Duration(音長変化量), Deviation(音の揺らぎ), Expression(音の抑揚) の5つの特徴量を用いる。12,846 曲の MIDI データの統計 [3][4] から、連続する2音間の特徴量変化量の傾向を分析し、アンサンブルの特性 [5] と併せて考慮することで、ある音のあるストリームに対する適性値である Similarity を得る。

### 2.3 提案手法の概要

本研究で提案するストリーム分離手法の概要を図1に示す。

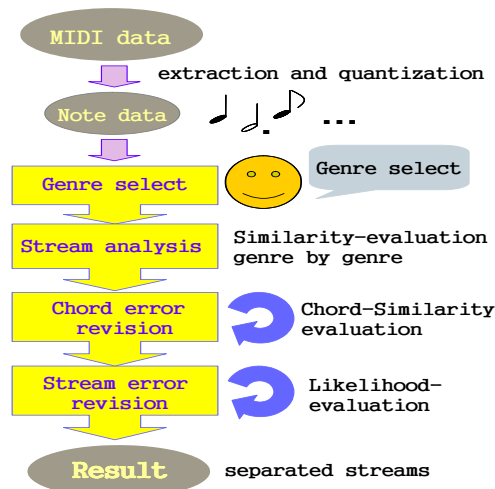


図 1: ストリーム分離システムの概要

\* "Part separation based on musical stream structure"

### 2.3.1 単音データの抽出と量子化

まず入力されたMIDIデータから、ドラム、パーカッションなどのリズムトラックを除き、トラック内の全ての音のGate timeの合計上位Nトラックを抽出する。今回の実験ではN=4とした。そしてその中からNote dataとして(Note name:音名,Velocity:音量,Gate time:実際の音長,Step time:論理的な音長,Deviation:揺らぎ,Expression:抑揚,Time:発音時刻,Track#:トラックナンバー)を抽出する。Step timeはGate timeを60tickで量子化した値とし、Track#は解析率を出す時のみに用いる。

### 2.3.2 ジャンルの選択

最初に、ユーザが大まかな楽曲のジャンルをPOPS,CLASSIC,JAZZ,DANCE,民謡その他の中から選択する。これは楽曲のジャンルによって、5つの特徴量の最適重み比が異なり、評価関数が変わってくるからである。なお最適重み比は、分散遺伝的アルゴリズムを用いて事前に各実験曲の最適重み比を求め、それを傾向ごとにクラスタリングし、ジャンルごとに与える。

### 2.3.3 ストリーム分離

時間的に最初に最大同時発音している場所から処理を始める。開始地点から、時間軸正の向きに順次追加される音に対して、5つの特徴量を統合したSimilarity評価関数に基づいて、各ストリームに所属させていく。曲の最後まで処理が終わったら、開始地点から時間軸負の向きに曲頭まで遡って処理を行う。

### 2.3.4 ストリーム誤り補正

時間軸に沿って処理を行う為、分離誤りが起きた箇所から誤りが伝播してしまう。そこで音をストリームに所属させる際にSimdiffという、ストリームへの所属の確からしさを表すパラメータ(ある音のSimilarity最大のストリームと次に大きなストリームのSimilarityの差)を与え、その値が最小の音を、最も誤っている可能性が高いとして、その音以降のストリームを入れ替える。その際、Likelihood(曲全体の確からしさ)が最も高くなるような入れ替え方をする。Likelihoodは、統計的なデータとアンサンブルの特性から、各ストリームの音高、音量、音長、音長揺らぎ、抑揚の平均及び標準偏差からなる10つのパラメータに重みをつけてストリーム間で組み合わせることによって求めた。もしLikelihoodが上がるような入れ替え方がなかったら、その音は正しいとして補正せず、同様の処理をSimdiffが閾値以下の全ての音について行う。

## 3 評価実験

入力曲として、YAMAHAが作成したMIDIデータ115曲(POPS,ROCK,R&B,HOUSE,TECHNO,民族音楽)、RWC音楽データベースのMIDIデータ200曲(POPS,CLASSIC,JAZZ)及び、Web上で収集したMIDIデータ20曲(主に洋楽)を用いた。この結果、図2のような実験結果が得られた。なおPSAはパート分離精度、STAはストリーム追跡精度を表す。

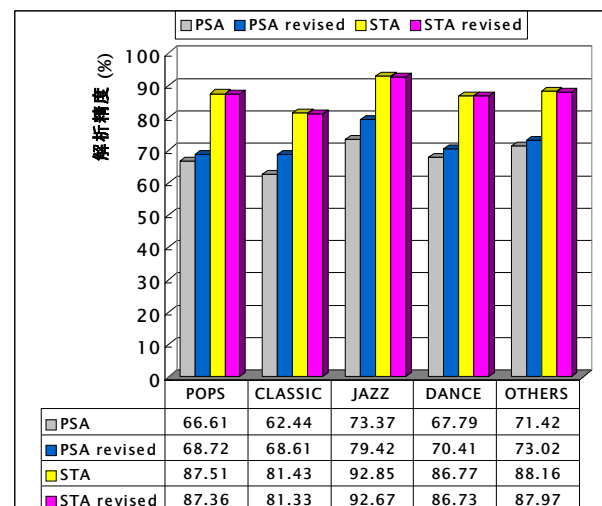


図2: システムの解析精度

## 4 おわりに

本研究では、楽曲のストリーム構造を用いて、音楽的なアプローチから音源分離を行うことを試みた。今後の課題としては、誤り補正機能の強化、既存のストリームに音を属させるだけでなく、新しいストリームを検出する機能の追加、曲全体の統計ではなく、パートや楽器ごとの音列の統計をとって統計データの信頼性を向上させることなどが挙げられる。

## 参考文献

- [1] A.S.Bregman: "Auditory scene analysis", MIT Press,1990
- [2] 木下 智義, 半田 伊吹, 武藤 誠, 坂井 修一, 田中英彦: "自動採譜処理における知覚的階層に着目したパート分離処理", 信学論 (D-II), Vol.J85-D-II, No.3, pp.373-381, 2002
- [3] 松田 稔, 秋山 好一: "日本の楽曲の基本的特徴 (その2)", 音響学会誌 51,pp.271-281,1995
- [4] 小杉 尚子, 小島 明, 片岡 良治, 串間 和彦: "大規模音楽データベースのハミング検索システム, 情報処理学会論文誌", Vol43, No.2, pp.287-298, 2002
- [5] 篠田元一: "新しい音楽理論", Rittor Music, 2001