

音楽情報処理 ・ 擬音語認識 音源追跡

奥 乃 博

京都大学 大学院情報学研究科
知能情報学専攻
知能メディア講座 音声メディア分野
<http://winnie.kuis.kyoto-u.ac.jp/~okuno/>
okuno@i.kyoto-u.ac.jp, okuno@nue.org

1

目次

1. 音楽情報処理とMPEG-7
2. 自動採譜
3. 音高による音色変化に着目した楽器音の音源同定
4. 未知の楽器を考慮した楽器音の音源同定
5. ドラム音認識
6. 発達論的コミュニケーションのうち、擬音語認識
7. レポート課題(再掲)

2

音楽情報処理とAI

- 作曲
- 演奏
- 音楽理論
 - ・ 音楽心理学
 - ・ 音楽知覚
- デジタル音響処理
- 信号・記号・変換

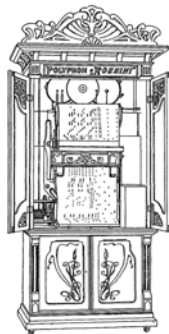


Figure 1. Automatic string instrument with piano-roll, shown in a perforated paper used (1904).

3

自動採譜の例

- Guitar duet
- 長い音の認識
- 雑音
- ギターではチューニングの変化の影響



Figure 5. (a) Original score (120 beats per minute); (b) computer-printed score produced by means of automatic music transcription.

4

音楽信号認識

1. デジタル信号
 - アタックの検出
 - ピッチの決定(あれば)
 - ラウドネスの決定
 - 音源分離(音源分離)
 - 音源同定
 - ニュアンス(ビブラート・トレモロ・スタッカート)検出
 - 休止検出
2. 音響マップ
 - 調律(チューニング)同定
 - 調同定
 - テンポの揺らぎを追跡
 - 拍子決定
 - 音符と休符の割り当て
 - 音声の分離
 - スラー・タイの決定
3. 音楽マップ

5

テンポ追跡アルゴリズム



Figure 7. Tempo tracking algorithm.

6

音楽構造の表現法

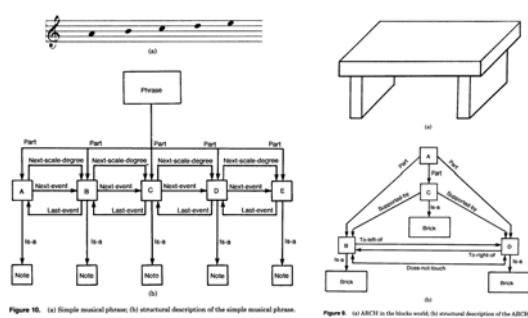


Figure 10. (a) Simple musical phrase; (b) structural description of the simple musical phrase

Figure 3. (a) $\Delta\text{H}^\circ_{\text{f}}(\text{kJ mol}^{-1})$ in the black model; (b) measured formation enthalpies in the grey model.

音声情報処理特論 2005年12月7日

音楽情報処理とMPEG-7

MPEG-7

- Moving Picture Experts Group (MPEG):'88
- Multimedia Content Description Interface:'97
- MPEG-7の構成
 - ・ Audio part, Video part → 音響, 映像だけ規定
 - ・ Multimedia Description Schemes (MDS) part → audio・visual descriptorsを含む記述法を規定
 - ・ Description Definition Language (DDL) → description schemeを表現する言語の標準化
 - ・ System part → 実環境での使用に対する糊
 - ・ Reference Software → オープンソースコード

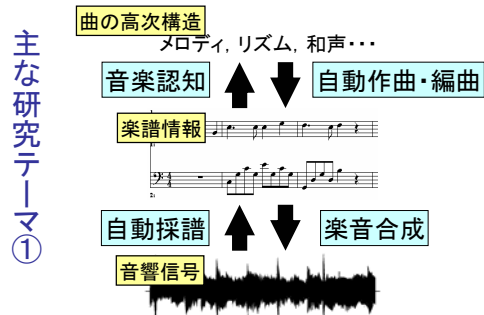
MPEG-7

- MPEG-1, MPEG-2, MPEG-4: audiovisual data に対するツールを提供
- MPEG-7: contents についてのnavigation するための手段を提供
- メタデータの種類: Dublin Core, GDA (Global Data Annotation), UNL (Universal Network Language)
- 設計方針:
 - Wide application base, Relation with content, Wide array of data type, Media independence, Object-base, Format independence, Abstract level, Extensibility

10

音楽情報処理とは(その1)

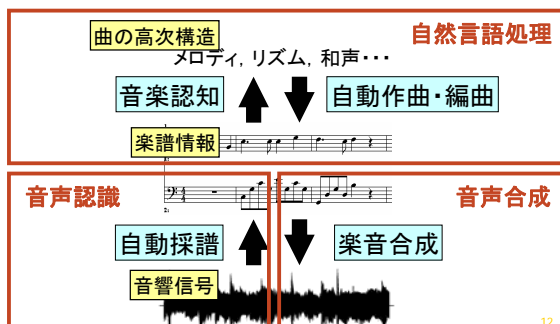
音楽情報処理＝音楽を扱う情報処理技術全般



11

音楽情報処理の音声でのアナロジー

主な研究テーマ① 「音声」にたとえると...



12

音楽情報処理とは(その2)

主な研究テーマ②＝音楽情報検索

- デジタル音楽配信の普及
⇒入手可能な音楽データの急増
- 音は一覧性のないメディア
⇒目的の曲を探すには一曲一曲試聴が必要
- 代表的な3つのアプローチ
①鼻歌検索 ②印象語検索 ③類似楽曲検索
- 検索に適した高次アーカイブの構築
⇒音楽へのMPEG-7タグ付けが重要な課題
- <http://www.ismir.net/>

13

音楽情報処理とは(その3)

主な研究テーマ③＝演奏の自動表情付け

- 入力: 楽譜or表情のないMIDIファイル
出力: 表情のついたMIDIファイル
 - 2002年より, コンクール(Rencon)開始
 - <http://shouchan.ei.tuat.ac.jp/~rencon/>
- その他の研究テーマ
- ジャムセッションシステム
 - 音楽表現のための新インターフェース
 - 感性情報処理, メディアアート, etc.

14

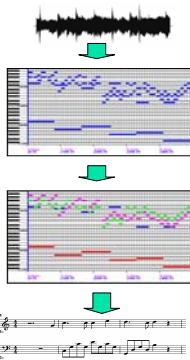
研究事例紹介① 自動採譜

15

自動採譜とは

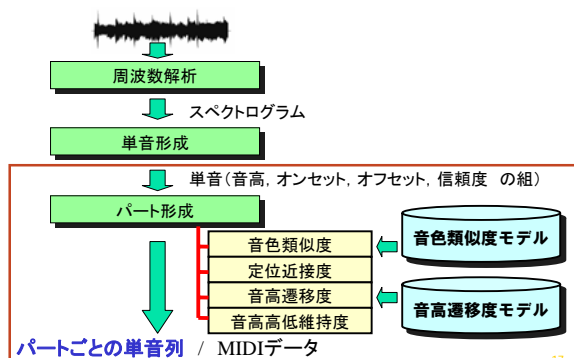
- 自動採譜＝音響信号から楽譜へ
- 自動採譜の中心となる処理

1. 音楽音響信号から個々の単音 (音高, オンセット, オフセット) を推定 ⇒ 単音形成
2. 単音 (音符) をパートごとに分類 ⇒ パート形成
3. 音価推定, テンポ推定, 調認識, ...



16

標準的なシステムの概念図

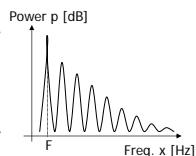


17

単音形成 (混合音からの音高推定)

- 基本周波数Fの調波構造 $p(x|F)$ をモデル化 (音モデル)
- 混合音のスペクトル $p(x)$ を, あらゆる基本周波数Fの音モデルの加重混合とみなす

$$p(x) = \int_{F_l}^{F_h} w(F) p(x|F) dF$$

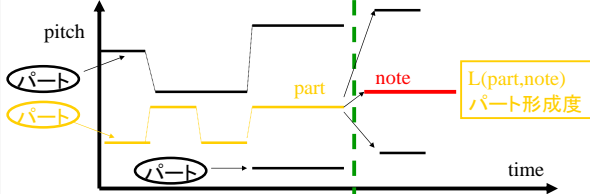


- 重み $w(F)$ をEMアルゴリズムで推定
- $w(F)$ の値が大きい iff Fに基本周波数成分がある
- 特長①音源数仮定せず②Missing fundamental対応

18

パート形成

入力: 単音列 (音高, オンセット, オフセット, 信頼度の組)
出力: パート (追跡した単音の集合)



- パート形成度 $L(\text{part}, \text{note})$ / パートpartと単音候補noteとの結びつきの強さ
- パート形成度が最大となる単音候補とパートを形成

19

パート形成

- パート形成 $L(\text{part}, \text{note})$ をどのように算出するか
⇒ 以下の4つの手がかりを利用

1. 一つの旋律は類似した音色の系列を持つ
⇒ 音色類似度
2. 一つの旋律は近接した定位の系列を持つ
⇒ 定位近接度
3. 一つの旋律に出現する音高の遷移には傾向がある
⇒ 音高遷移度
4. 旋律同士の音高の高低関係は維持する傾向にある
⇒ 音高高低維持度

20

音色類似度 $L_t(\text{part}, \text{note})$

- 音色類似度 $L(\text{part}, \text{note})$ を2単音間の音色類似度の平均で求める

$$L_t(\text{part}, \text{note}) = \frac{1}{c} \sum_{\text{note}_j \in \text{part}} L_t(\text{note}_j, \text{note})$$

- 二単音が同じ/異なる楽器である群 Π_0 / Π_1
- 二単音 $\text{note}_j, \text{note}_k$ の音色特徴ベクトルの差 x_{jk}

$$\begin{aligned} L_t(\text{note}_j, \text{note}_k) &= \frac{p(\Pi_0 | x_{jk})}{p(x_{jk} | \Pi_0)} \\ &= \frac{p(\Pi_0)}{p(x_{jk} | \Pi_0) + p(x_{jk} | \Pi_1)} \end{aligned}$$

$$p(x | \Pi_i) = \frac{1}{(2\pi)^d |\Sigma|} \exp\left(-\frac{D_i^2(x, \mu_i)}{2}\right)$$

d : 正規分布の次元数, Σ : 共分散行列

$D_i^2(x, \mu_i)$: x と群 Π_i の平均 μ_i とのマハラノビス距離
各群の事前確率は等しいと仮定

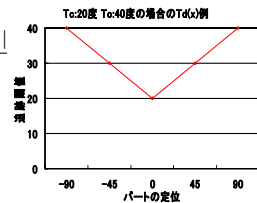
21

定位近接度 $L_d(part, note)$

- T_c : 0度(中央)における定位閾値
- T_o : 左右90度(真横)における定位閾値
- D : 定位 (パートの定位はIPDから求める)

$$L_d(part, note) = 1 - \frac{|D(part) - D(note)|}{T_d(D(part))}$$

$$T_d(x) = T_c + (T_o - T_c) \cdot \frac{|x|}{90}$$



22

音高推移度 $L_p(part, note)$

- 音高のトライグラムモデルを利用
- 学習データ
 - RWC研究用音楽データベース: クラシック(50曲)
 - 付属MIDIデータの単旋律のトラックを利用
 - 総音符数(167179個)
 - 長調と短調それぞれ統計をとる
 - 調性で正規化(調は楽譜から得る)

23

音高高低維持度 $L_r(part, note)$

- パートpartが単音noteとパート形成する
⇒ 他のパートより音高が高くなる

$$part_i (1 \leq i \leq N)$$

$$L_r(part, note) = \frac{1}{N} \sum_i \left(\frac{1}{2} + (q_i - \frac{1}{2}) \cdot M(time_i) \right)$$

$$M(t) = 1 - \exp(-C \cdot t)$$

q_i : part の音高が part より高かった割合, C : 定数

$time_i$: part と part の同時発音時間 (1小節1として正規化)



time が十分小さい: $L_r(part, note) = \frac{1}{2}$
 time が十分大きい: $L_r(part, note) = \frac{1}{N} \sum_i q_i$

24

パート形成実験

- 評価データは無響室で録音した
 - ・ パッヘルベルのカノン(四重奏)
 - ・ 蛍の光(三重奏)
- 各手がかりの組み合わせによる性能差を評価
- 正解条件(楽譜と比較)
 - ・ 音高が正しい
 - ・ 発音時刻の誤差が32分音符以内
 - ・ 出力パートに含まれる単音が入力の楽譜で同一音源に由来

pf	pf	pf	vl	pf	pf	pf	fl	pf
vl	vl	fl	pf	vl	vl	vl		

➡
 75%
(12/16)

28

パート形成 実験結果

音色類似度	定位近接度	音高遷移度	音高高低維持度	精度(%) 「カノン」	精度(%) 「蛍の光」
○	—	—	—	63.32	79.87
—	○	—	—	77.39	84.90
—	—	○	—	66.50	75.17
—	—	—	○	57.03	63.76
○	○	○	—	85.01	90.27
○	○	—	○	77.39	84.90
○	—	○	○	66.50	79.19
—	○	○	○	84.60	91.61
○	○	○	○	84.86	90.27

29

研究事例紹介② 音高による音色変化に着目した 楽器音の音源同定

31

楽器音の音源同定とは

楽器音の音源同定
(音からの楽器名の同定)



自動採譜・音楽アーカイブ構築
などにおいて重要な課題

```
<inst begin=00:00:00  
      end=02:10:00>  
  piano  
</inst>  
  .....
```

関連研究:

- 音楽認識関連の多くは、音高推定を指向(音源同定は少数)
- 現状の性能は、
単独音: 70-80% / 10-30クラス
混合音: 60-70% / 3-5クラス

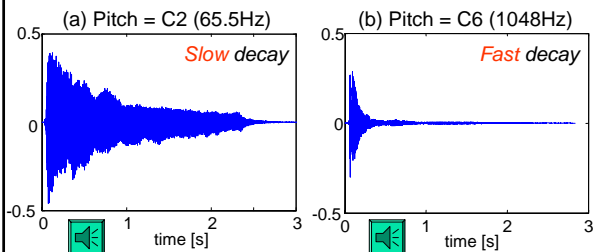
32

特徴量抽出での問題点の所在

同一楽器でも音高によって音色が変化する

(∵ 楽器音は、他の音に比べて音域広)

例 Piano



33

音高による音色変化(音高依存性)への対応

[課題] 音高による音色変化の確率分布としての
モデル化(∵ Bayes-basedの様々な識別器利用可)

[こんな方法は?]

半音毎に異なる多次元正規分布を学習し、
同定時に適切な多次元正規分布を選択する。

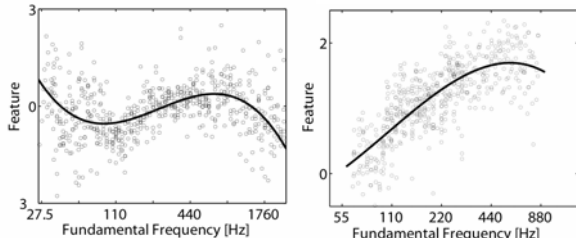
十分な学習データを用意するのは非現実的
(多次元正規分布学習には大量の学習データ)

少ない学習データのためには、確率分布そのもの
に音高による音色変化を表現する機能が必要

34

音高依存性の表現法

- 音色を表す各特徴量の音高依存性を
基本周波数(F0)の関数として近似
- このF0の関数をパラメータとする確率分布を使用



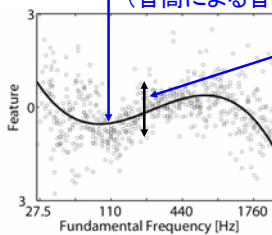
左はピアノの第4軸, 右はチェロの第1軸(いずれも次元圧縮後)

F0依存多次元正規分布

多次元正規分布の拡張
= 平均ベクトルが基本周波数の関数

代表値関数

各特徴量のF0依存性を近似するF0の関数
(音高による音色変化を表現)



F0正規化共分散行列

各特徴量を代表値関数で
正規化したうえで算出
(音高以外の要因による
音色変化を表現)

音源同定アルゴリズム

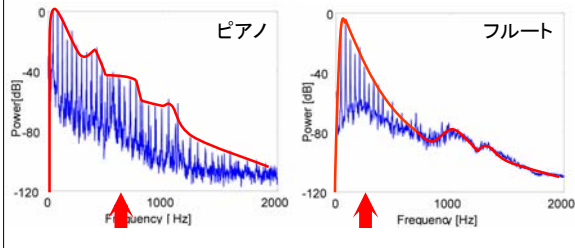
- 特徴抽出(129個)
- 主成分分析で次元圧縮
(累積寄与率99%で79次元に圧縮)
- 線形判別分析でさらに次元圧縮
(19楽器なので18次元に圧縮)
- F0依存多次元正規分布のパラメータ推定
- ベイズ決定規則に基づいて楽器名を同定
(事後確率が最大になる楽器名を見つける)

$$g_i(x; f) = \log p(x | \omega_i; f) + \log p(\omega_i; f)$$

使用している特徴量の例

1. 特徴抽出(129個)

例: 周波数重心

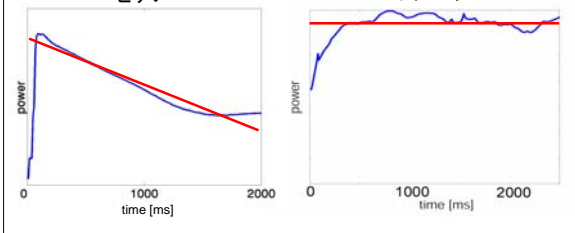


38

使用している特徴量の例(続き)

1. 特徴抽出(129個)

例: パワー包絡線の最小二乗法による近似直線の傾き



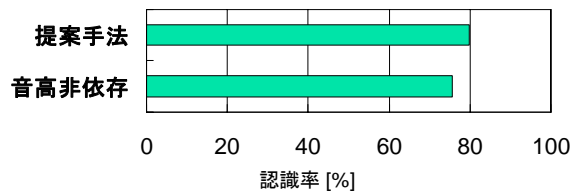
39

実験方法

- 使用データベース: RWC-MDB-I-2001
 - 実楽器の単独発音を半音ごとに収録
 - 今回は19種類の楽器を使用
 - 各楽器に, 3楽器個体, 3種類の音の強さ
 - 今回は, 通常の奏法のみ使用
 - 使用したデータ総数: 6,247個
- 上記のデータを無作為に10等分し, クロスバリデーション
- 音高は既知

40

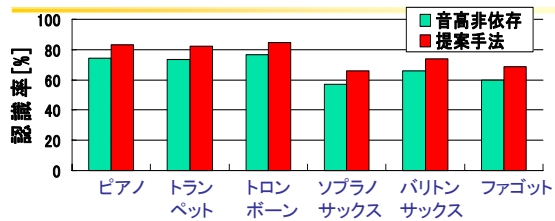
実験結果



提案手法: 79.73%の認識率を実現.
音高非依存に比べ, 4.00%認識率向上.
誤り削減率は, 16.48%.

41

実験結果: 認識率が7%以上改善された楽器



- ・**ピアノ**: 最も性能改善
(認識率9.06%改善, 誤り削減35.13%)
∵ 音域が広く音高による音色変化が顕著
- ・PF, TR, TBで約33~35%の認識誤りを削減
- ・SS, BS, FGでも20%以上の認識誤りを削減⁴²

研究事例紹介③ 未知の楽器を考慮した 楽器音の音源同定

43

未知楽器の問題

- 学習データにない楽器(未知楽器)が入力されたときに、それをどう扱うかという問題
- 実際のタグ付けでは不可避な問題

実際の音楽では、オーケストラ用楽器、民族楽器、シンセサイザーによる合成音など**多種多様な楽音**が使用され、これらの学習データを**網羅的に収集するのは困難**




- 従来研究では指摘されてこなかった

44

未知楽器の問題

人間ならどうするか

たとえば、バイオリンとビオラの音をシンセサイザー上で合成して作った音を初めて聴いたとき、人はその音をどう理解するか

- バイオリン 
- ビオラ 
- 合成音 

45

未知楽器の問題

人間ならどうするか

たとえば、バイオリンとビオラの音をシンセサイザー上で合成して作った音を初めて聴いたとき、人はその音をどう理解するか

我々の予測:

「楽器名はわからないけど、弦楽器系」
⇒ 一段**抽象度の高いレベル**で音をとらえる



今回の目的=これを計算機上で実現する

46

未知楽器の問題

- 本研究における解決策:
既知楽器⇒**楽器名レベル**で認識
未知楽器⇒**カテゴリーレベル**で認識
- 具体的なアルゴリズム
 - (1) **楽器名レベル**で認識
 - (2) (1)の結果が本当に正しいかどうか判定
(「既知」か「未知」かの判定に相当)
 - (3) (2)でFalse(=未知)と判定されたら,
カテゴリーレベルで再認識

47

楽器カテゴリーの設計

- 音源同定に適した楽器カテゴリーとは?
⇒楽器の**音響的類似性**を総合的にとらえた
楽器カテゴリー

従来からある楽器の発音機構に基づく階層表現が
使えるのでは？

48

楽器カテゴリーの設計

楽器の発音機構に基づく階層表現

大分類	中分類	小分類	属する楽器
弦楽器	—	打弦楽器	PF
		撥弦楽器	CG, UK, AG
		擦弦楽器	VN, VL, VC
管楽器	木管楽器	無簧楽器	PC, FL, RC
		単簧楽器	SS, AS, TS, BS, CL
		複簧楽器	OB, FG
	金管楽器	—	TR, TB
打楽器	(省略)	(省略)	(省略)

楽器カテゴリーの設計

- 音源同定に適した楽器カテゴリーとは？
⇒楽器の**音響的類似性**を総合的にとらえた楽器カテゴリー
⇒楽器の**音響的類似性に基づく階層表現**を自動獲得し、そこから楽器カテゴリーを作成

従来からある楽器の発音機構に基づく階層表現が使えるのでは？

No! 楽器の発音機構に基づく階層表現は、必ずしも**音響的類似性をとらえていない**.
e.g. バイオリンとギターはともに弦楽器だが音響的には大きく異なる

50

楽器階層の獲得における課題と解決策

課題1 使用する特徴空間によって結果が変化
音源同定で用いるものと同じ特徴空間を使用
⇒任意の音源同定システムに対して、適切な階層表現を自動的に獲得

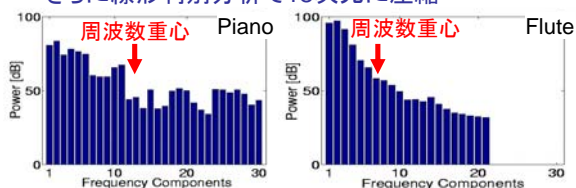
課題2 音高などにより特徴空間上の位置が変化
各楽器**多数の音響信号**を用意し、各楽器の**分布**に対して階層的クラスタリング
⇒各楽器1音のみに比べ、各楽器の**特徴空間上の位置関係を適切に把握可能**

51

使用する特徴空間

音源同定に用いるものと同じ特徴空間を使用
⇒[北原, 音情研2002]で用いたものを使用

- 「**周波数重心**」, 「**パワー包絡線の近似直線の傾き**」など, 129個の特徴量を抽出
- 主成分分析で79次元(累積寄与率:99%)に圧縮し, さらに線形判別分析で18次元に圧縮



使用する楽器音の音響信号

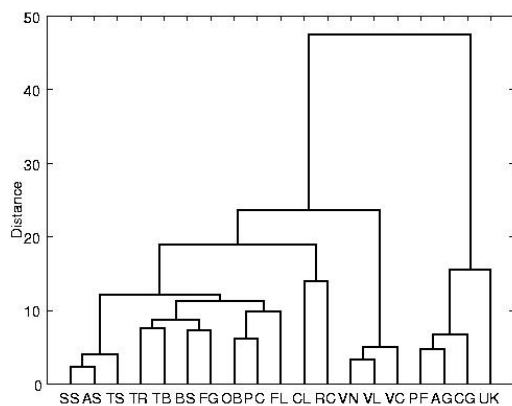
1楽器あたり130～700個, 計6,247個を使用

- 19種類のオーケストラ楽器の実楽器音データを「RWC-MDB-I-2001」から抜粋
- 半音ごとに全音域収録
- 各楽器, 3楽器個体, 3種類の音の強さ
- 通常の奏法のみ使用



以上のデータから得られる各楽器の特徴空間上の分布を**多次元正規分布で近似**し, 各楽器間の**マハラノビス汎距離**を使って階層的クラスタリング

53



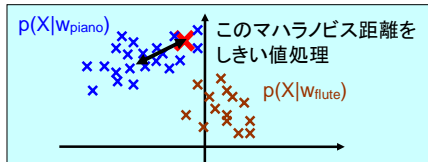
楽器カテゴリー設計結果

大分類	中分類	小分類	属する楽器
減衰系楽器	——	ウクレレ以外	PF, CG, AG
		ウクレレ	UK
持続系楽器	弦楽器	——	VN, VL, VC
	管楽器	サククス	SS, AS, TS
		クラリネット	CL
		リコーダー	RC
		低音系 + α	TR, TB, BS, FG
		高音系	OB, PC, FL

55

楽器音同定アルゴリズム

- (1) **楽器名レベル**で認識
- (2) (1)の結果が本当に正しいか判定
(Falseなら「未知楽器である」とみなす)
認識対象音から学習データ(分布)までのマハラノビス距離がしきい値以内ならTrue
- (3) (2)の結果がFalseなら**カテゴリーレベル**で再認識



56

評価実験条件

- 既知楽器なら**楽器名レベル**で,
未知楽器なら**カテゴリーレベル**で認識
- 学習データ, 評価用データともに
単音を1つ1つ個別に収録したものを使用
- 認識(楽器名・カテゴリーともに)では,
129次元の特徴空間を**PCAで79次元**に,
LDAでさらに18次元に圧縮したものを使用
- 既知／未知の判定では, 129次元の特徴空間
を**PCAで23次元**に圧縮したものを使用
- 既知／未知の判定で用いる**しきい値は40**

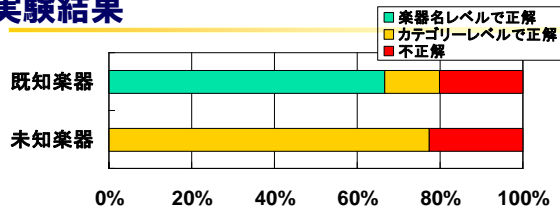
57

使用データベースの詳細

- **学習データ: 自然楽器音**
(RWC-MDB-I-2001から抜粋した
19楽器6,247音のうち, ランダムに半分を選択)
- **既知楽器の評価データ: 自然楽器音**
(上記の残り半分)
- **未知楽器の評価データ: 電子楽器音**
(ヤマハ製MU2000Iに収録されている
- エレクトリックピアノ(ElecPf),
- シンセストリングス(SynStr),
- シンセブラス(SynBrs).
※各々2バリエーションずつ使用)

58

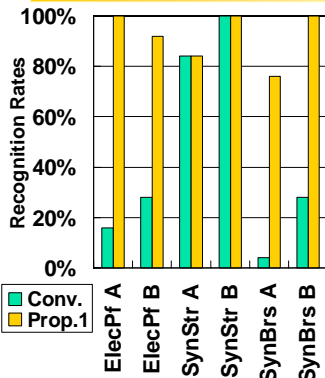
実験結果



- 誤り率は、既知楽器で約20%, 未知楽器で約23%.
- このような楽器音理解は、情報統合においても有用
e.g. 音から「楽器名はわからないが弦楽器」と同定
画像から「ある民族楽器」
⇒弦楽器に属する新たな楽器として再学習

59

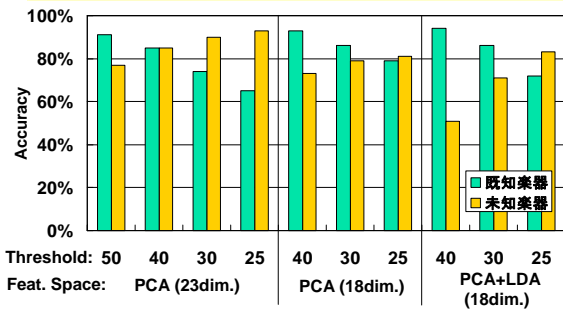
未知楽器のカテゴリーレベルの認識



- 提案手法によって得られた楽器カテゴリーを用いることで、未知楽器を精度良く認識
- 楽器の発音機構に基づく分類は、(機械的発音機構の持たない) 電子楽器には有効でないことを示唆

60

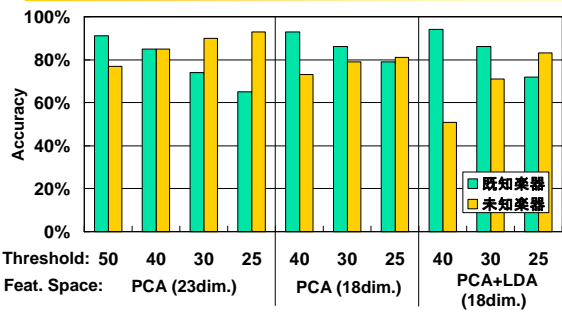
楽器音の「既知」か「未知」かの判定



PCA (23dim) において、約85%の正解率

61

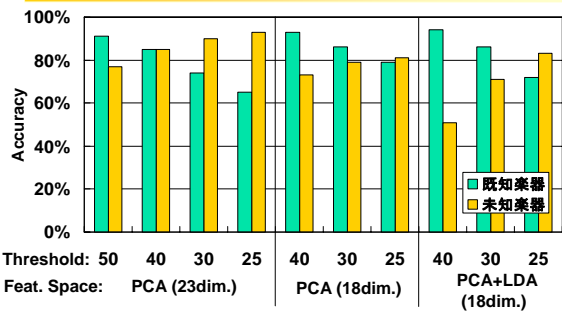
楽器音の「既知」か「未知」かの判定



「既知を正しく既知」と「未知を正しく未知」はトレードオフ
⇒ 応用に応じた適切なしきい値設定の必要性

62

楽器音の「既知」か「未知」かの判定



PCA+LDAは精度低

∴ LDAは学習データの分離を良くする次元圧縮法

63

音声情報処理特論 2005年12月7日

ドラムパターン推定による ドラム音認識誤り補正手法

吉井 和佳† 後藤 真孝‡

駒谷 和範† 尾形 哲也† 奥乃 博†

†京都大学情報学研究科 知能情報学専攻

‡産業技術総合研究所

64

研究の背景

■ 音楽データの爆発的な増加

- デジタル音楽配信 (WEB・携帯電話)
- 書誌情報を「知らない」楽曲の増加

■ 書誌情報に基づく楽曲検索

- タイトル・アーティストに着目したテキスト検索
- 書誌情報を「知っている」ことが前提
 - データベース制作者サイド
 - ユーザサイド



書誌情報を「知らなくてもいい」検索システムが必要

研究の背景

■ Music Information Retrieval (MIR)

- 音楽コンテンツを記述して利用
 - リズム: ビート, テンポ, 拍子
 - メロディ: 楽器・歌手名, 区間
 - ハーモニー: キー, コード

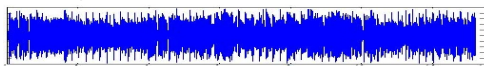
ポピュラー音楽では、
ドラムパートが楽曲のリズム形成に寄与している
と考え...

ドラム音認識手法を開発: 音楽音響信号中から
バスドラムとスネアドラムの発音時刻を検出する

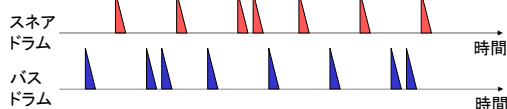
66

ドラム音認識

音楽音響信号



発音時刻検出

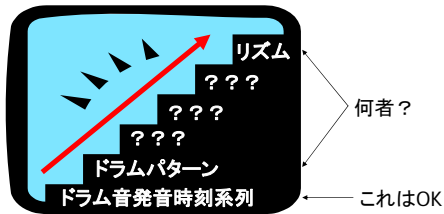


67

リズム記述へ向けて

■ ドラム音の発音時刻系列からドラムパターンへ

- ・ 着目する音楽コンテンツの **高次化**
- ・ リズム記述へ一歩進める



68

リズムとは [Gouyon2004]

- The word “rhythm” is used to refer to **all the temporal aspects** of a musical work.
 - ・ テンポ推定
 - ・ 拍子推定
 - ・ ビートトラッキング
- リズム記述時に考慮すべき3要素
 - ・ **Metrical structure**
 - ・ **Tempo**
 - ・ **Timing**

69

リズム記述の要件

- **Metrical structure** **周期的構造**
 - ・ Objective regular temporal structure underlying musical event occurrences and organizing them into a hierarchical metrical structure. **階層的構造**
- **Tempo**
 - ・ bpm は四分音符レベルにおける **ビートの割合**
- **Timing**
 - ・ 厳密な楽譜上のタイミングからの **逸脱**
 - ・ 演奏表現の“豊かさ”

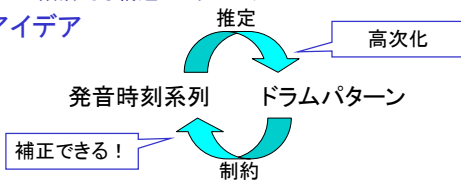
70

ドラムパターンとは

■ ドラムパターンとは

- Objective regular temporal structure underlying **drum-onset sequences** and organizing them into a hierarchical metrical structure.
 - 規則的な構造＝発音時刻系列の周期性
 - 階層的な構造＝ドラムパターン

■ アイデア

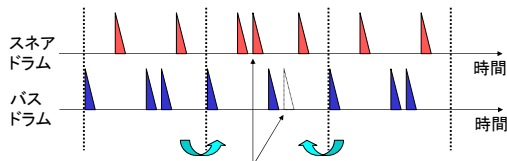


71

ドラムパターン推定による誤り補正

同じドラムパターンは連続反復されやすいという仮定のもとで...

① ドラムパターン推定



② 認識誤りの可能性が高い時刻の検出と再検証

72

まとめ

- MIRシステム構築にはリズム記述が必要
 - ドラム音認識手法を提案
- 着目するコンテンツの高次化
 - ドラム音発音時刻系列からドラムパターンへ
- 高次コンテンツを制約に利用
 - ドラムパターン推定に基づく認識誤り補正

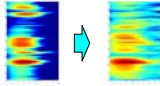
ボトムアップ・トップダウン処理

フィードフォワード・フィードバック処理

73

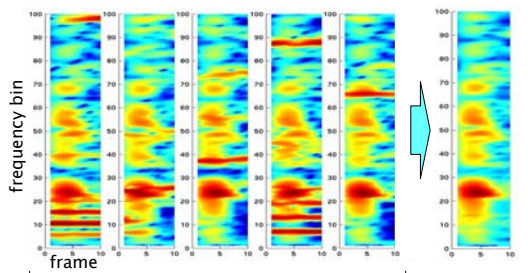
音楽音響信号中のドラム音認識

- 目的
 - 音楽音響信号中のドラム音発音時刻の検出
- 問題設定
 - RWC研究用音楽データベース ポピュラー音楽
 - BD, SDの発音時刻を個別に検出
- 課題と解決手法
 - スペクトログラムテンプレートに基づく認識
 - 個体差問題:
 - テンプレート適応
 - 混合音問題:
 - テンプレートマッチング (特別な距離尺度)



74

テンプレート適応

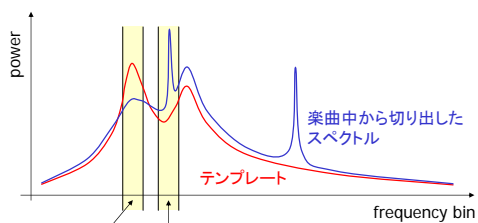


テンプレートを手がかりに発音しているような時刻を楽曲中から複数見つける → 精度が高い方がよい
適応後テンプレート

75

テンプレートマッチング

- スペクトログラムの重畳を考慮した距離尺度



局所的距離1 局所的距離0 → 周波数平面全体で積分
距離の閾値で発音しているかどうかを2値判定

76

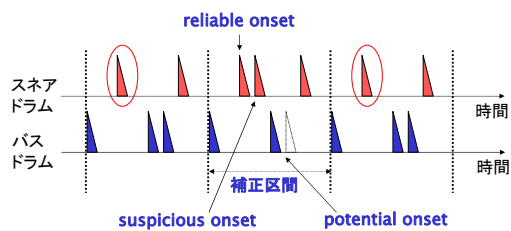
まとめ

- テンプレート適応
 - ・ スペクトログラムの選択精度が重要
- テンプレートマッチング
 - ・ 距離の閾値が重要

77

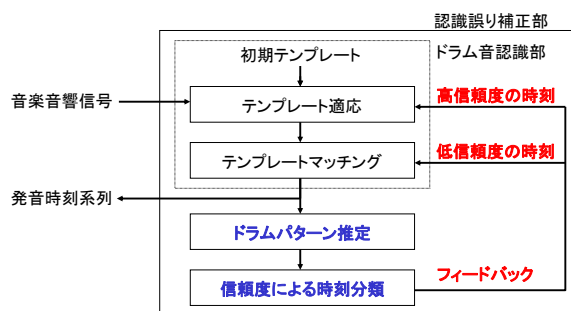
認識誤り補正法

- 前後のドラムパターンに着目した信頼度付与
 - ・ **Reliable onset** → 適応に利用
 - ・ **Suspicious onset** → 閾値を下げて再マッチング
 - ・ **Potential onset** → 閾値を上げて再マッチング



78

システム構成



79

ドラムパターン推定の課題

- 入力音響信号に関して以下を仮定する

- 拍子 : 2/4 か 4/4

- bpm : 60から200

安直な発想では
ドラムパターン=小節



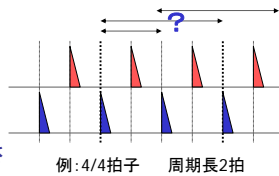
推定がかなり難しい

- 開始時刻の決定

- ・ 1拍目推定の難しさ

- 時間長の決定

- ・ 小節長推定の難しさ



80

解決方法

- 1拍目開始、4拍長のドラムパターンを切り出す

- 切り出し誤りの影響を小さくする

- リファレンスパターンの利用

- 1拍目開始、4拍長の平均的なドラムパターン

- 相関値が高いパターンを切り出す

- 同一ドラムパターンが反復されやすいという仮定

- 周期性が高いパターンを切り出す

- 切り出し誤りの影響の小さい補正法

- 切り出されたパターンの前後のみを補正に利用する

81

周期長の算出

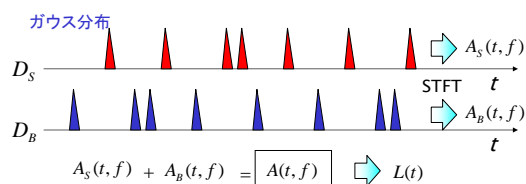
- 発音時刻分布 D の生成

- ・ 各発音時刻に誤差モデル配置

- STFTによる振幅スペクトル A の計算

- ・ 窓幅2048[frames], シフト長1[frame] (10[ms])

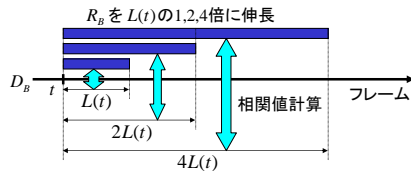
- 自己相関法により周期長 L を算出



82

ドラムパターンの推定

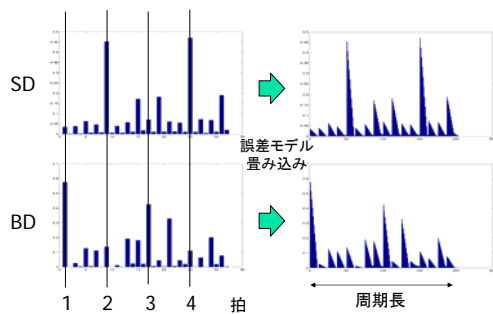
- 4拍長のリファレンスパターン R の準備
- リファレンスパターン R との相関値計算
 - ・ パターン長を 120~400[frames] に限定
 - ・ 周期長の2の累乗倍のリファレンスパターンを生成し、
相関値が閾値以上のパターン長・開始時刻を決定



83

リファレンスパターン

- 7819パターンの平均



84

予備実験(bpm推定)

- 評価対象
 - ・ ポピュラー音楽50曲から認識した発音時刻系列
- 推定方法
 - ・ パターン長が4拍だと仮定してbpmを算出・多数決

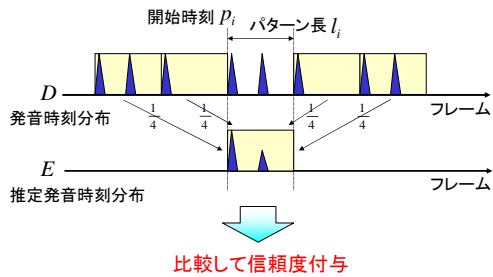
正解	倍速に誤り	誤り
40曲	9曲	1曲
80%	18%	2%

- 実際にはパターン長が2拍だと倍速に誤る
- パターン内の発音時刻数に着目すると精度向上の可能性
- 誤って推定してもパターン推定での問題は小さい

85

発音時刻分布推定

■ ドラムパターン区間の前後区間の平均



86

信頼度付与

■ 検出された発音時刻の検証

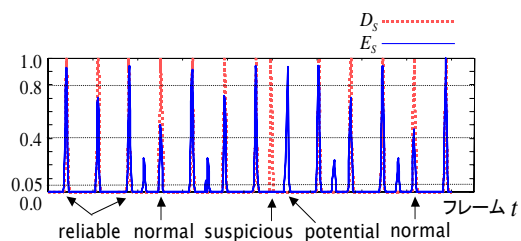
- **Reliable onset**
 - $D(t) = 1.0, E(t) \geq 0.8$
- **Normal onset**
 - $D(t) = 1.0, 0.8 > E(t) \geq 0.05$
- **Suspicious onset**
 - $D(t) = 1.0, 0.05 > E(t)$

■ 検出されなかった時刻の検証

- **Potential onset**
 - $D(t) = 0.0, E(t) > 0.4$

87

推定例 (No.35 snare drum)



88

適合率（50曲平均）

- バスドラム音の発音時刻：77%
 - Reliable onset：91%
 - Normal onset：77%
 - Suspicious onset：66%
 - Potential onset：23%
- スネアドラム音の発音時刻：78%
 - Reliable onset：92%
 - Normal onset：37%
 - Suspicious onset：51%
 - Potential onset：6%

89

まとめ

- テンプレート適応
 - スペクトログラムの選択精度が重要
- テンプレートマッチング
 - 距離の閾値が重要
- ドラムパターン推定に基づく信頼度付与
 - reliable onset：適応のやり直し
 - normal onset：マッチングによる検証（閾値↓）
 - suspicious onset：マッチングによる検証（閾値↓↓）
 - potential onset：マッチングによる検証（閾値↑）

90

評価実験

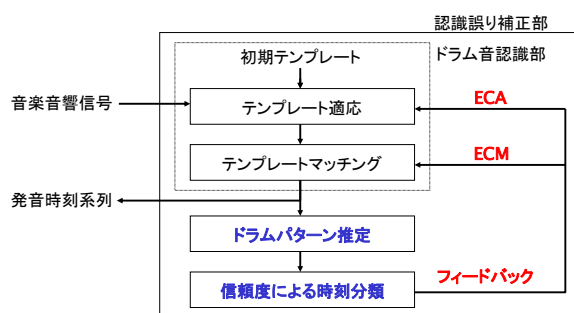
- 入力
 - RWC-MDB-P-2001 ポピュラー音楽50曲
- 実験条件

TM	テンプレートマッチング	認識手法
TA	テンプレート適応	
ECM	マッチングによる発音検証	補正手法
ECA	適応によるテンプレート再構成	

1つずつ有効にして評価

91

システム構成



92

評価尺度

$$\text{再現率 } R = \frac{\text{正解した発音時刻数}}{\text{実際の発音時刻数}}$$

$$\text{適合率 } P = \frac{\text{正解した発音時刻数}}{\text{出力された発音時刻数}}$$

$$F \text{ 値} = \frac{2RP}{R+P}$$

実際の発音時刻数は
バスドラム音20818、スネアドラム音15228

93

実験結果

■ F値の比較(単位は%)

	バスドラム	スネアドラム
TM	70.115	67.997
TM+TA (baseline)	76.786	77.996
baseline+ECA	77.510	79.404
baseline+TCM	80.567	79.340
baseline+ECA+ECM	81.121	80.319

- 誤り補正により認識率が改善
 - バスドラム: マッチングの効果が高い
 - スネアドラム: 適応の効果がバスドラムと比べて高い

94

まとめ

- ドラム音発音時刻系列からドラムパターンへ
 - ・ 周期性
 - ・ 階層性
- ドラムパターン推定に基づく認識誤り補正
 - ・ ボトムアップ・トップダウン処理
 - ・ フィードフォワード・フィードバック処理
 - ・ 副次的な効果: tempo推定

今後の課題としては・・・
より高度なドラムパターン遷移のモデル化が必要

95

音声情報処理特論 2005年12月7日

INTER:D: A Drum Sound Equalizer for Controlling Volume and Timbre of Drums

Kazuyoshi Yoshii*
Masataka Goto**
Hiroshi G. Okuno*

*Kyoto University, Japan
**AIST, Japan

96

New-Concept Equalizer

- The difference of equalizing targets.
 - Conventional graphic equalizers are frequency-based equalizers.



e.g. graphic equalizer of
Windows Media Player

- INTER:D is an instrument-based equalizer.



instrument names

100

Equalizing Functions of INTER:D

- Equalizing targets
 - Bass drum sounds
 - Snare drum sounds
- Real-time playback interface
 - **Volume Control**
 - Boost or cut the volume of the drum sounds
 - **Timbre Change**
 - Change the timbre of the drum sounds
 - **Drum Pattern Change** (new function!) 101

Easy-to-use Interface

- GUI similar to graphic equalizers
 - Volume control via horizontal sliders.
 - Timbre change via dropdown lists.



Demonstration

- Please watch a demo video!
<http://www.winnie.kuis.kyoto-u.ac.jp/~yoshii/interd/>

Conclusion and Future Work

- INTER:D is an **instrument-based equalizer** focusing on drum types, not on frequency bands.
- We achieved the **real-time interactive interface** in music appreciation environments.
- We integrated **low-level processing, semantic annotation and content sharing**.
- Future work includes...
 - improvements of drum-sound recognition accuracy and equalizing quality.
 - equalizing of other instrument sounds (e.g., vocals).

109

Drum Pattern Change

- You can edit drum patterns as you like!
 - MIDI-sequencer-like interface.
 - Automatic estimation of bar lines.
 - Beginner support by clustering drum patterns with a self-organizing map.



113

Music Information Retrieval

- Low-level audio processing methods for music content annotation.
 - Melody extraction methods.
 - Tempo estimation methods.
 - Drum-sound detection methods.
- Music content analysis for achieving music information retrieval (MIR).
 - Conventional retrieval systems focus only on text information such as singer names and titles.
 - MIR systems focus on music contents such as rhythms and melodies.

114

目次

1. 音楽情報処理とMPEG-7
2. 自動採譜
3. 音高による音色変化に着目した楽器音の音源同定
4. 未知の楽器を考慮した楽器音の音源同定
5. ドラム音認識
6. 発達論的コミュニケーションのうち、擬音語認識
7. レポート課題(再掲)

116


発達論的コミュニケーション

- 脳モデル化による対話能力自己獲得仮説原理
- 社会的対話能力獲得原理、ソーシャルインタラクション
- 対話能力を獲得する上での視覚認識機能
- 未発達レベルでの、曖昧語の発話・音声認識機能
- 音素認識、擬音語認識、真似発話(鸚鵡返し)
- 対話を通じた運動と知覚の双方向発達スキーム、アクティブパーセプション
- 対話による視覚、聴覚の発達の融合機能
- その他感覚モダリティの融合

117

環境音の擬音語認識とは

- 発達論的コミュニケーションの立場から、聴覚機構の工学的実現の第一歩として、「環境音を擬音語として認識」

- 環境音とは？『にゃあお』
音声・楽音以外の音全般
 - ・ 自然音(風の音、動物の声など)
 - ・ 人工音(機械の駆動音など)
 - ・ 楽器音(楽器の単発音)

118

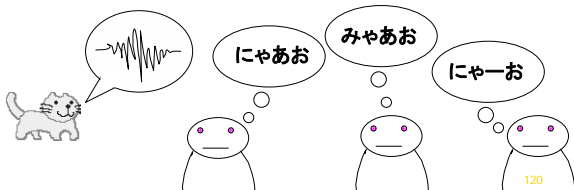
環境音の擬音語認識の背景

- 擬音語使用の利点
 - ・ 自然で人間らしい表現を実現
 - 日本語では日常生活で擬音語が多用される
 - ・ 音の詳細な表現が可能 [Wake01]
- 擬音語認識の期待される効用
 - ・ マンマシンインターフェースの高度化
 - ・ 聴覚障害者の補助(擬音語を用いた字幕)
 - ・ 音声認識の未定義語問題を処理
 - ・ 音響信号記号変換による情報圧縮

119

環境音の擬音語表現の問題

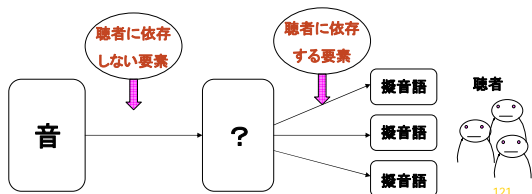
- 擬音語変換の問題点
 - 環境音を表す擬音語は聴者により異なる
 - 「にゃあお」「みゃあお」「にゃーお」・・・
 - ⇒『正しい』認識を定義することが難しい



120

擬音語表現の曖昧性の解決策

- 解決策
 - ・ 共通項⇒聴者に依存しない要素
 - ・ その他⇒聴者に依存する要素
- これらを分けて段階的に処理を行う
- * 音節レベル/モーラレベル/音素レベル



121

音節とモーラ

- 音節 例: 漢字(音読み)、など
- モーラ 例: かな文字、など
 - ・ 日本語はモーラ言語
- 音素 例: アルファベット、など

	新聞紙	こけこっこー
音節	しんぶんし	こけこっこー
モーラ	しんぶんし	こけこっこー
音素	sh i N b u N sh i	k o k e k o Q k o R

122

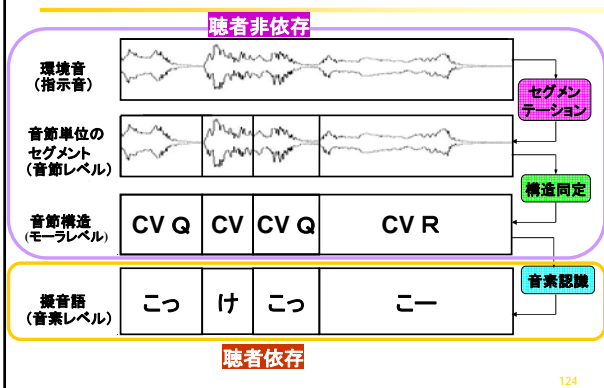
各レベルの聴者非依存性

- 音節レベル (の聴者非依存性)
音節数やアラインメントが聴者に依らず一致
- モーラレベル(の聴者非依存性)
右表の4種類のモーラ記号による表現が聴者に依らず一致
- 音素レベル (の聴者非依存性)
モーラレベルが一致し、対応する各CVの音素もまた一致

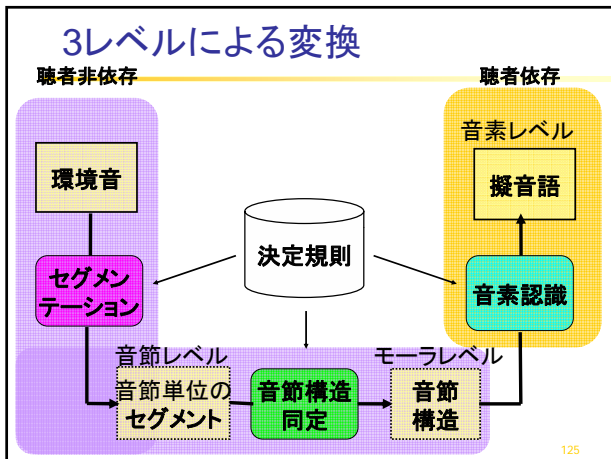
モーラ記号
CV:かな R:長音
Q:促音 N:撥音

123

擬音語認識の階層的処理



124



各レベルにおける聴者非依存性

- 音節レベル (聴者非依存)
Sonority Theory (聞こえ度理論)
「音節 = Sonority の一山」
- モーラレベル (聴者非依存)
日本人の感覚では一般に非依存 [比屋根ら]
⇒ 聴取実験で妥当性を検証
- 音素レベル (聴者依存)
聴者依存であるが何らかの傾向はある

126

参考文献

- ・“Sound Retrieval with Intuitive Verbal Descriptions”
和氣早苗 旭敏之 IEICE TRANS. INF. & SYST. vol.E84-D, No.11, 2001
- ・「犬は「びよ」と鳴いていた ―日本語は擬音語・擬態語が面白い―」
山口仲美 光文社 2002
- ・「オノマトペ ―形態と意味―」 田守育啓 くろしお出版 1999
- ・“Industrial Sound Design” 前田修 (<http://member.nifty.ne.jp/~ISD/>)
- ・異音の表現における擬音語の検討
田中基八郎 日本機械学会論文集C編 Vol.61 No.592, 1995
- ・単発音のスペクトル構造とその擬音語表現に関する検討
比屋根一雄 他, 信学技法 SP97-125, 1998
- ・“Sound Retrieval with Intuitive Verbal Expressions”
Sanae Wake, Toshiyuki Asahi, Proc. Of ICAD '98 (International Conference on Auditory Display), The proceedings have been published as an eWiC, 1998

127

擬音語変換と音高・音長情報を 統合した環境音認識

2005.10.26
B4 田口明裕

128

研究の背景

- マンマシンインタラクションの多様化
 - ・ 環境音: 人間の状況知覚で重要な役割
 - ・ 計算機が環境音情報を理解する必要性
- 環境音検索
 - ・ DTMなど、大量の効果音データを扱う機会の増加
 - ・ 音には一覧性が無く、検索には多大な手間



人間の感覚に即した環境音の表現手法が必要

129

環境音を扱う従来研究

- 音声・楽音からは遅れつつも、環境音に主眼を置いた研究は増えてきている
- 環境音を扱う従来研究は、主に以下の3種類に大別できる
 - ・ 環境音の音源同定に関する研究
 - ・ 環境音の表現に関する研究
 - ・ 環境音の検索に関する研究

130

従来研究（音源同定）

- 鳴き声による鳥の種類の認識システム[芦谷]
- コンピュータによる牛音声の理解[池田]
- 環境音の音響信号から、音源（音響事象）を識別
- 音のドメインは限定される
- 音源が既知・明確な音の分類を目的とする
→ 音の「聞こえ方」を記述するという観点ではない

131

従来研究（表現）

- 単発音のスペクトル構造とその擬音語表現に関する検討[比屋根]
- 環境音を対象とした擬音語自動認識[石原]
- 環境音を直感的・効率的に記す表現方法（およびその変換方法）を検討する
- 対象となる音は単発音、もしくはその反復
→ 複雑な構造を有する様々な音にそのまま適用できるには至っていない

132

従来研究（検索）

- 直感的な音データ検索/編集システムの開発[和氣]
- 映像コンテンツの音響製作を効率化する為の効果音検索[林]
- 直感的なメタデータ（印象、擬音語、状況、etc.）を付与し、一覧性を持たない音を効率よく検索する
- メタデータを手動で付与する
→ 大規模な音データベースへの適用は困難



構造を持った環境音一般を直感的に扱えるようなメタデータの自動付与は従来、実現されていない

133

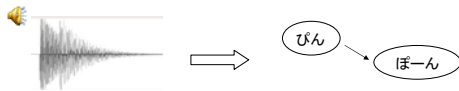
本研究の考え・目的

- 計算機によって、環境音の「聞こえ方」を認識する
- 人間の実際の感覚に即した表現を行う
 - ・ ヒトが、聞こえた音を表現する一般的な手段：
「韻律」に乗せた「擬音語」表現
(韻律：音長(リズム)および音高)
 - ・ これらの情報を併せて記述することで、元の音をより忠実に再現する
- その表現を利用して、大規模な音データベースに対する効率的な検索を実現する

134

アプローチ

- 対象とする音：人間が擬音語で模倣しうる環境音
 - ・ 高々1フレーズに収まる程度
 - ・ 複雑な複合音は含まないものとする
- 1音節の擬音語で表せる部分 → 1セグメント
- 環境音を、セグメントの系列として表現
 - ・ 各セグメントに、擬音語と音高情報を持たせる



135

課題

- セグメンテーション
 - ・ 擬音語1音節で表現できる部分を正しく決定する
- 擬音語認識
 - ・ 各セグメントに妥当な擬音語表現を与える
- 音高認識
 - ・ 明確な音高推移があれば、それを正しく取得する
- 検索システムの設定
 - ・ 類似度の評価方法・検索条件の入力方法

136

1. セグメンテーション

- パワー包絡の1山が擬音語の1音節に相当する[石原]
- パワー包絡の変動が少ない場合でも、音色や音高の急激な変化によって擬音語の変化が起こることがある



- パワー包絡の山、谷を用いてセグメント分割
→更に、スペクトル包絡の変形度によって再分割

137

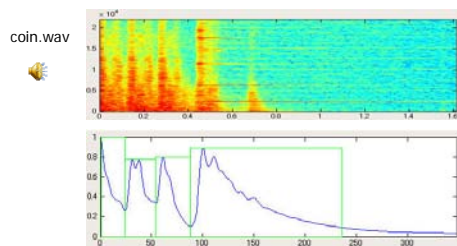
パワー包絡による分割

- 音響波形のパワー包絡を計算
- 包絡のピークを抽出
- 隣接ピーク間の距離・山谷比を計算
 - ・ 閾値以下のものは1ピークとしてまとめる
- 繰返し、最終結果からセグメントを得る
- 断続音らしき部位は1セグメントとして纏められる

138

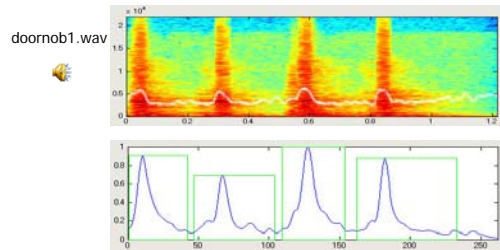
実験

- 環境音を、パワー包絡のみに基づいて分割
- 殆どの場合、人間の感覚と同じ分割が行われた



139

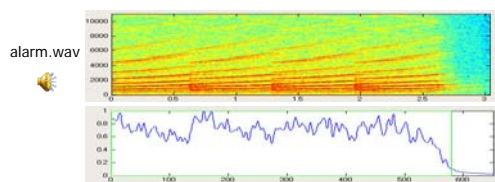
実験



140

実験

- 音高・音色の急激な変化が含まれる音には有効ではない



- スペクトル包絡の変形度も考慮すべき

141

スペクトル変形量の取得

- スペクトル包絡を、荒い周波数分解能でリサンプリング
 - AM変調の影響除去のため、各時刻でパワーの正規化
 - 各周波数について、過去一定期間のパワーの変化を最小2乗法で直線近似
 - 近似直線の傾きと誤差から、変形量を算出
-
- 有意な結果はまだ得られていない
 - 変形量の算出法など、更に検討したい

142

リズム認識について

- 環境音は、セグメントの系列として表現される
- セグメントのin, out をそのまま記述してよいのか
- 等間隔で鳴っている、等のリズム推定が必要か？
(音符のような表現を用いる？)

143

2. 擬音語認識

- 音声認識と同様、HMMを用いる
- セグメンテーション済みの波形を入力
→セグメンテーション問題は発生しない
- 現状は、石原の手法を踏襲
 - HMM(16混合モノフォンモデル)
 - 特徴量
 $\text{MFCC}(16) + \text{Pow} + \Delta\text{MFCC}(16) + \Delta\text{Pow}$
 - 擬音語1音節(モーラ構造)の出力
子音 + 母音 + 促音 or 撥音
 - 環境音用の音素を新たに設計する
e.g.) blにもdlにも聞こえる音を音素b-dと表現

144

3. 音高認識

- セグメント内、及びセグメント間での音高推移の様子を取得する
- 一般環境音が対象(調波構造を持たない可能性)
 - F0が無い場合の音高知覚：様々な要因の影響
e.g.) 音の強さ、先行音・後続音の有無、周波数の組合せ
 - 計量法は明らかになっていない
- 綺麗なピークを持つものを判別し、それだけを扱う(信頼度のようなパラメータを設定しておく)

145

取得すべき音高情報

- 絶対的な指標を取ることは困難
→ 相対的・大まかな表記で充分
- 現時点では次のようなものを想定している
 - セグメントの音高
(数段階の離散的なもの)
 - 隣り合うセグメント間の相対的上下関係
 - セグメント内での音高推移
 - 一定・上昇・下降・山なり・谷型の推移パターン
 - 振動の性質を持つかどうか

146

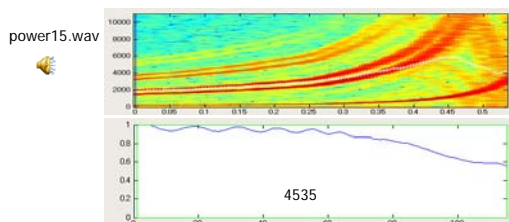
周波数重心

- 1つのセグメント内では音色はあまり変化しない
(スペクトルの形状が保たれる)と仮定
↓
- 周波数重心を求めることで、セグメント内の音高推移が求められるのではないか？

147

実験

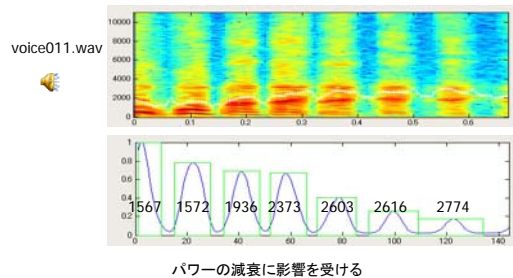
- 周波数重心を計算
- それに基づき、音高推移パターンを取得できるか



148

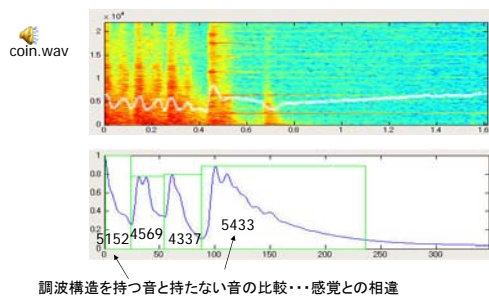
実験

- 上手くいかない例



149

実験



150

実験から

- 綺麗な調波構造を持つ or 鋭いピークを持つ音
 - 音高推移については
おおよそ感覚に合った結果が得られる
- 特定のピークを持たない音
 - 周波数重心の数値自体は意味を持たない
(他のセグメントとの音高比較には用いられない)
 - パワーの大小など、他の要因の影響が強い
- 調波構造らしさを見分ける指標がやはり重要

調波構造らしさについて

- ログスケールで倍音パターンとの相関を取る
 - ・ 調波構造のあるものは出力結果が大きくなる傾向が見られる、程度
 - ・ これだけで判別するのは困難
- 有声無声の判定で用いられている手法が利用できないか検討している段階

152

4. 検索システムの設定

- 人間の環境音認識は曖昧性が高い
- 曖昧なクエリに対して、適切な結果を返すには、
 - 検索条件の入力方法
 - 類似度の評価基準をよく考慮する必要がある

153

検索システム入力法(イメージ)

- 直感的なインターフェース
 - ・ ピアノロールのようにセグメントを配置する、等
- 音によって、重要となる指標は異なるはず
e.g.) リズムが明快な音、音高推移が特徴的な音
 - ・ 擬音語・音高・リズムの重要度を設定させる



154

まとめ・残された課題

- スペクトル変形量に基づくセグメンテーションの必要性
- リズム認識(セグメント系列→音符的表現)の検討
- 擬音語変換手法の改善点の検討
- 調波構造らしさの算出法を定める
- それを用いた検索システムの構築

155

音声情報処理特論 2005年12月7日

音声対話システムにおける 音韻的類似表現の混同を 防ぐための確認の生成

京都大学

浜辺良二、駒谷和範、尾形哲也、奥乃博

156

音声対話システムの課題

- システムが聞き間違ふ
- 人が聞き間違ふ

プチ告白! こんなんやっちゃいました

私の失敗談ではないんですが...

「金閣寺」と「銀閣寺」は聞き間違えが多い。独自の対策としては、相手が日本人でも「ゴールド7シルバー?」と確認すること。

メーターを起こし忘れ、780円が2800円になるまで気づかなかった! 果しそなお客さんがみんななぜ手を挙げないのか不思議に思いながら走った。

お客さんをドアサービスで後部座席に乗せた後、自分も隣に座ってしまった!!

157

背景・目的

- 音声対話システムの実用化
- ユーザがシステムの確認応答を聞き間違えると対話が破綻
- 特に、音韻的に紛らわしい内容語は聞き間違いやすい

本研究の目的：
音韻的に紛らわしい内容語の聞き間違いを
なくすための確認応答を行う

158

本手法で実現する対話例

(京都市バス運行情報案内システム)

S: 乗車場所をおっしゃってください。

U: 金閣寺道から。

(そのまま聞き返す確認)

S: 金閣寺道からですか？

「金閣寺道からですか？」
か「銀閣寺道からで
すか？」かユーザが混同する

(本手法による確認)

S: 金閣寺道か銀閣寺道か分からないのですが、
先頭に英語で **ゴールド** の金の字が付きますか？

音韻的に紛らわしい内容語を聞き取りやすくする
新たな単語(付加表現)を加えた確認を行う

159

本手法の概要

- 対話システムの語彙(ドメイン)に依存しない
手法を提案⇒既存の辞書から付加表現を取得

「先頭に英語で **ゴールド** の金の字が付きますか？」

⇒英語辞書を利用する

「先頭の音は、切手の『き』ですか？」

⇒簡単な単語を付加して一音を強調する(和文通話表)

160

問題点(1)－付加表現の取得

■ 問題点(1)

単一の辞書から適切な付加表現が得られるとは限らない

「先頭に英語でゴールドの金の字が付きますか？」

適用範囲が広い

より分かりやすい

「先頭の音は、切手の『き』ですか？」

複数の辞書を用いて付加表現の候補を取得する

161

問題点(2)－付加表現の選択

■ 問題点(2)

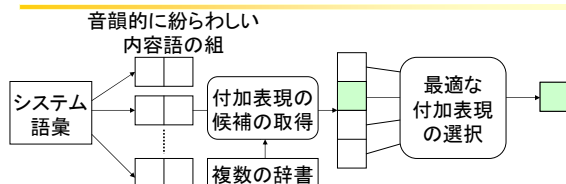
どのようにして最適な付加表現を選択するか

- ・ 付加表現の適切性の定義が必要

適切性を「理解のしやすさ」および「聞き取りやすさ」によって表現する

162

本手法の構成



1. 音韻的に紛らわしい内容語の組をシステムの語彙から抽出する
2. 複数の辞書から付加表現の候補を取得する
3. 候補の中から最適な付加表現を選択する

163

1. 音韻的に紛らわしい語の抽出

- 単語間の音韻的距離を定義
 - DPマッチングで音素列の編集距離を計算
 - モーラ数が増える場合のコストを2とする
 - 単語の長さで正規化
- 音韻的距離の小さい単語の例
 - 「二条駅」、「四条駅」
 - 「東福寺」、「長福寺」
 - 「高野」、「高雄」

164

2. 付加表現候補の取得法

- 付加表現の候補を以下の4つの方法で辞書から取得

- 英訳(英語辞書を利用)
「英語でゴールドの金の字が付きますか？」
- 一音を強調(和文通話表を利用)
「先頭の音は、切手の『き』ですか？」
- 熟語を取得(単語辞書:JC1 を利用)
「金曜日の金(きん)の字が付きますか？」
- 訓読み(漢字辞書を利用)
「金(かね)の字が付きますか？」

165

3. 最適な付加表現の選択

付加表現の適切性を「理解のしやすさ」と「聞き取りやすさ」で表現

- 「理解のしやすさ」
 - ユーザがある単語を聞いたとき、それを容易に想像できるか
- 「聞き取りやすさ」
 - ユーザがある単語を聞いたとき、他の単語と混同しないか

166

付加表現の理解のしやすさ

- 辞書に与えられた単語の難易度によって、付加表現の理解のしやすさを表現
 - 単語辞書: JC1、英語辞書では単語の難易度が5段階に分類

難易度の高い付加表現を用いた確認



ユーザが何度聞いても理解できない可能性が高い



難易度が最小の付加表現のみを候補に残す

167

付加表現の聞き取りやすさ

- 付加表現の聞き取りやすさを以下の2つで表現

◆一意性: 他の単語と混同しにくい

一意性が小さい例:

「先頭の音は汽車の『き』ですか？」

- 「汽車」と「医者」とが混同しやすい

◆相違性: 元の単語がどれだけ変化したか

相違性が小さい例:

「金額の金(きん)の字が付きますか？」

- 「金閣」と「金額」との違いが少ない

168

聞き取りやすさの尺度

一意性:

付加表現との音韻的距離の小さい、辞書内の他の単語上位10個の平均距離

相違性:

元の単語と付加表現との音韻的距離



聞き取りやすさをこれらの重み付き和で定義

• $\text{score} = W_1 \cdot (\text{一意性}) + W_2 \cdot (\text{相違性})$

(予備実験により $W_1 = 4$, $W_2 = 1$ と定めた)

169

付加表現の選択の評価

- 評価の高い候補が選択されているかを検証
- 候補に主観的評価をA(良), B, C(悪)で与える

《選択された付加表現の評価の分布》

	京都市バス運行情報案内システム(全58語)		ホテル検索システム(全50語)	
	選択前の候補	選択後の候補	選択前の候補	選択後の候補
A	113個	49個	102個	45個
B	102個	9個	60個	5個
C	13個	0個	9個	0個

- 評価の高い候補が多く選択されている

170

評価実験

- 実験目的
 - ・ 確認をユーザが聞き取れるかを検証
- 実験内容
 - ・ 京都市バス運行情報案内システム、ホテル検索システムの語彙内の単語54組を対象
 - ・ システムがそのまま聞き返して確認してもユーザが聞き取れなかった単語に対し、本手法による確認を行う
 - ・ 確認は3回まで繰り返せる
 - ・ 実用場面を想定し背景雑音(SN比0)を加える
 - ・ 被験者は5人

171

実験結果

そのまま聞き返す確認

正しく聞き取れた単語	27.8組	52%
誤って聞き取った単語	1.8組	3%
聞き取れなかった単語	24.4組	45%



本手法による確認

正しく聞き取れた単語	16.6組	68%
誤って聞き取った単語	1.2組	5%
聞き取れなかった単語	6.6組	27%

172

考察

- そのまま聞き返す確認で聞き取れなかった単語の68%が本手法により正しく聞き取れるようになった
⇒本手法の有効性が示された
- 本手法を用いてもなお聞き取れない単語が27%
⇒対話内の話題と関係のない単語が突然現れるため、予測して聞くことが困難

173

本研究のまとめ

- 研究の目的:システムの確認応答において音韻的に紛らわしい語を、新たな単語を付加することで聞き取りやすくする
- 本手法は特定のドメインに依存しない
- 単一の辞書から適切な付加表現が得られるとは限らない⇒複数の辞書から候補を取得
- 最適な付加表現をどう選択するか⇒付加表現の理解のしやすさ・聞き取りやすさを定義
- 聴取実験によって本手法の有効性を確認

174

レポート課題（再掲）

- レポートは3題のうち、2題選択回答
- 1. 柏野さんの講演(情報学展望)のまとめと感想(10ptで A4 5ページ以上)
- 2. 音源定位・音源分離・分離音認識について2つ以上の技法を詳細に報告(10ptで A4 5ページ以上)
- 3. 音声に限らず一般の音に対して、他のメディアとの情報統合によりどのような機能が実現できるのか、よりロバストな処理が可能になるか、について論ぜよ。(10ptで A4 5ページ以上)。
- レポートの締切は12月21日(水)
- 提出先は10号館レポートボックス

177