楽器音を対象とした音源同定：音高による音色変化を考慮する識別手法の検討

北原 鉄朗[†] 後藤 真孝[††] 奥乃　博[†]

[†]京都大学大学院情報学研究科知能情報学専攻

[††]科技団さきがけ21／産業技術総合研究所

July 2002

発表の流れ

1. 音源同定とは
2. 音高による音色変化に着目した音源同定 [北原, MUS-40-2, 2001]
3. 本発表で提案する手法
4. 処理の流れ
5. 評価実験
6. まとめ

1. 音源同定とは

楽器音の同定

---入力された音は，piano? flute? …

- **パターン認識**の一分野
- **自動採譜・メディア検索**などで有用
- 研究対象として，広く扱われるようになったのは最近（**1990年代**に入ってから）

1. 音源同定とは

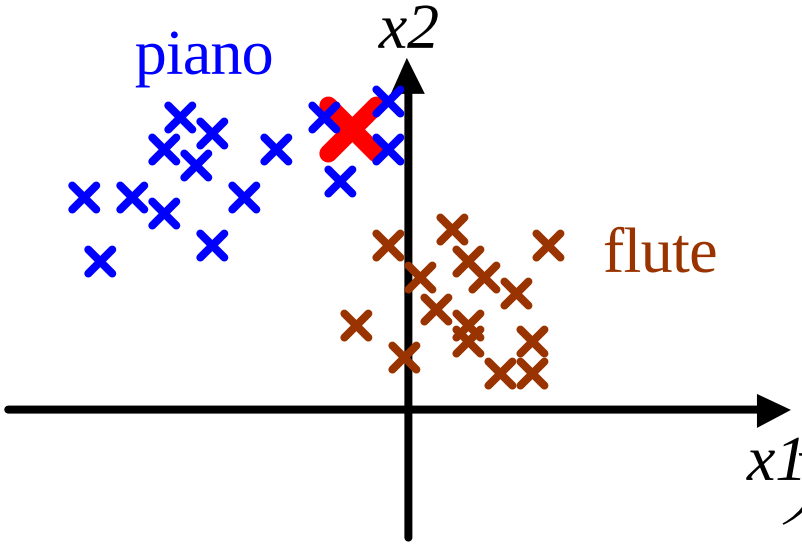
処理の概要

音響信号

特徴抽出

x1 :パワー包絡線の傾きの中央値
x2 :周波数重心　など

特徴空間



あらかじめ用意された
各楽器の音響信号と比較

入力された楽器音は**piano**

1. 音源同定とは

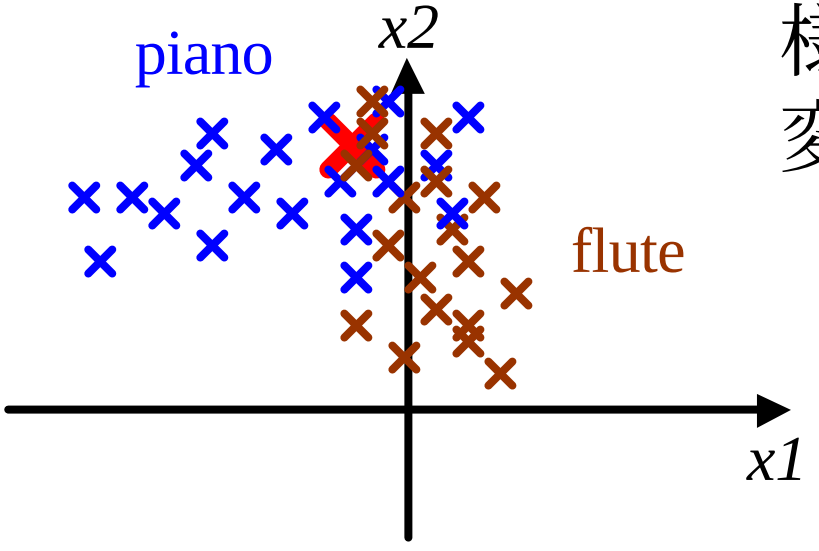
処理の概要　実際には…）

音響信号

特徴抽出

x1 :パワー包絡線の傾きの中央値
x2 :周波数重心　など

特徴空間



様々な要因により特徴量が変動するため，同定が困難

1. 音源同定とは

楽器音における特徴変動

楽器音における特徴変動の要因：

音高 ・音の強さ ・楽器の個体差 ・奏法など

→　これらの特徴変動をどのように考慮するか

この問題を扱った従来研究は少ない

たとえば, 楽器の個体差に着目

→ 「適応型混合テンプレート法」(柏野ら, '98)など

1. 音源同定とは

楽器音における特徴変動

楽器音における特徴変動の要因：

音高 **音の強さ** **楽器の個体差** **奏法**など

→　これらの特徴変動をどのように考慮するか

この

た

→

これらのうち，**音高**は物理量（**基本周波数**）として抽出可能

’98)など

1. 音源同定とは

楽器音における特徴変動

楽器音における特徴変動の要因：

音高 **音の強さ** **楽器の個体差** **奏法**など

→　これらの特徴変動をどのように考慮するか

これらのうち，**音高**は物理量（**基本周波数**）として抽出可能

↓

音高による特徴変動を

基本周波数の関数として近似

この

た

→

’98)など

2. 音高による音色変化に着目した音源同定 [北原, 2001]

1. 音高による変化の仕方で**特徴量を3つに分類**
   （∵ 特徴量によって音高による変化の仕方は様々）
2. 特徴量の分布を表現する**基本周波数の関数**を導入
   代表値関数 :各音高における分布の平均に相当
   変動値関数 :各音高における分布の分散に相当
3. これらの関数を用いて識別するため，
   木下の識別関数（類似度）を拡張

2. 音高による音色変化に着目した[原, 2001]

特徴量を手動で
分類する必要がある

1. 音高による変化の仕方で**特徴量を3つに分類**
   （∵ 特徴量によって音高による変化の仕方は様々）
2. 特徴量の分布を表現する**基本周波数の関数**を導入
   代表値関数 :各音高における分布の平均に相当
   変動値関[illegible]相当
3. これらの関数を用いて識別するため,
   木下の識別関数（類似度）を拡張

一般的な識別関数ではない

2. 音高による音色変化に着目した [原, 2001]

特徴量を手動で
分類する必要がある

1. 音高による変化の仕方で**特徴量を3つに分類**
   （ ）

⇒ **音色変化をより高次の関数で近似**

2. 特徴量の分布を表現する**基本周波数の関数**を導入
   代表値関数：各音高における分布の平均に相当
   変動値関　　相当

一般的な識別関数ではない

3. これらの関数を用いて識別するため，
   木下の識別関数（類似度）を拡張

⇒ **多次元正規分布を拡張**

3. 本発表における提案手法

多次元正規分布の拡張

音高による音色変化を扱えるように
　　多次元正規分布を拡張

3. 本発表における提案手法

多次元正規分布の拡張

音高による音色変化を扱えるように
　多次元正規分布を拡張.

「音高ごとに学習すればよいのでは？」
　（たとえば音高C4用の多次元正規分布をC4のデータだけで学習する）

3. 本発表における提案手法

多次元正規分布の拡張

音高による音色変化を扱えるように
　多次元正規分布を拡張

「音高ごとに学習すればよいのでは？」
　（たとえば音高C4用の多次元正規分布をC4のデータだけで学習する）

この方法では，より多くの学習データが必要
　（88鍵のピアノであれば，学習データが1/88に減ったのと同じ）

3. 本発表における提案手法

多次元正規分布の拡張

音高による音色変化を扱えるように
　多次元正規分布を拡張

「音高ごとに学習すればよいのでは？」
　（たとえば音高C4用の多次元正規分布をC4のデータだけで学習する）

この方法では，より多くの学習データが必要
　（88鍵のピアノであれば，学習データが1/88に減ったのと同じ）

⇒　**平均** :**音高によって連続的に変化**すると仮定
　共分散 :**音高に依存しない**と仮定

3. 本発表における提案手法

多次元正規分布の拡張

音高による音色変化を扱えるように
　多次元正規分布を拡張

「音高ごとに学習すればよいのでは？」
　（たとえば音高C4用の多次元正規分布をC4のデータだけで学習する）

この方法では，より多くの学習データが必要
　（88鍵のピアノであれば，学習データが1/88に減ったのと同）

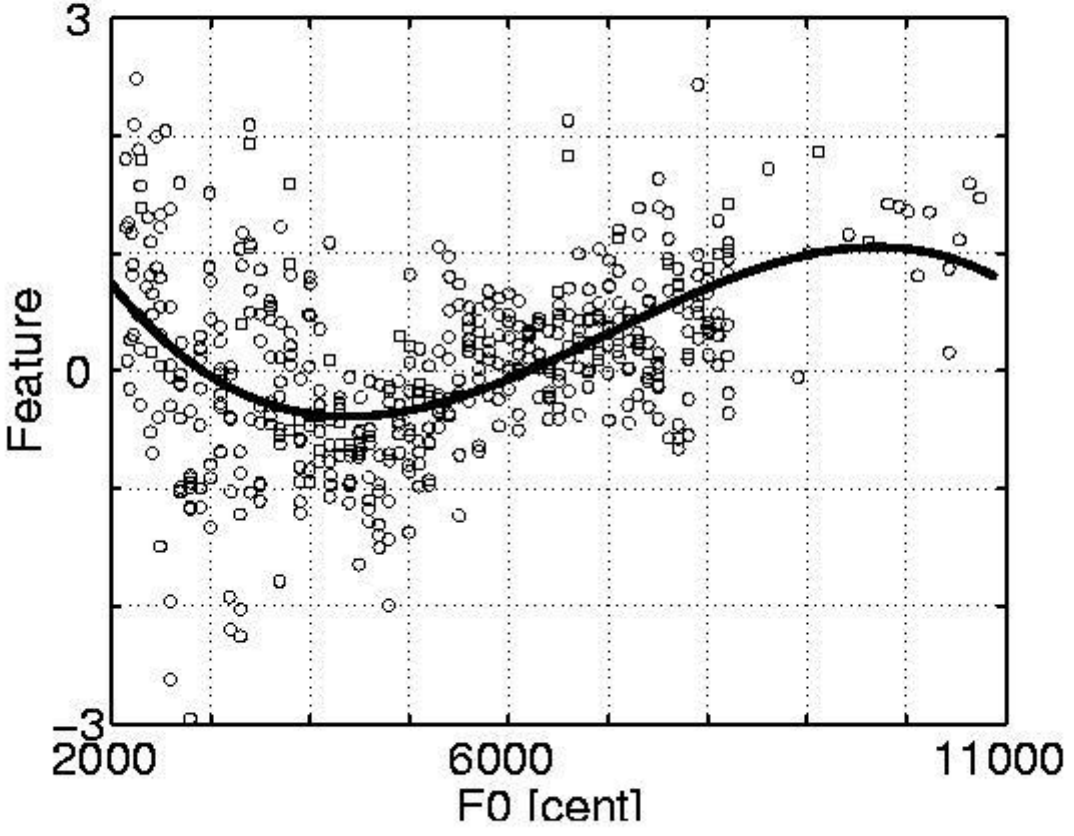
F0依存多次元正規分布

⇒ **平均**：**音高によって連続的に変化**すると仮定
　共分散：**音高に依存しない**と仮定

3. 本発表における提案手法

代表値関数

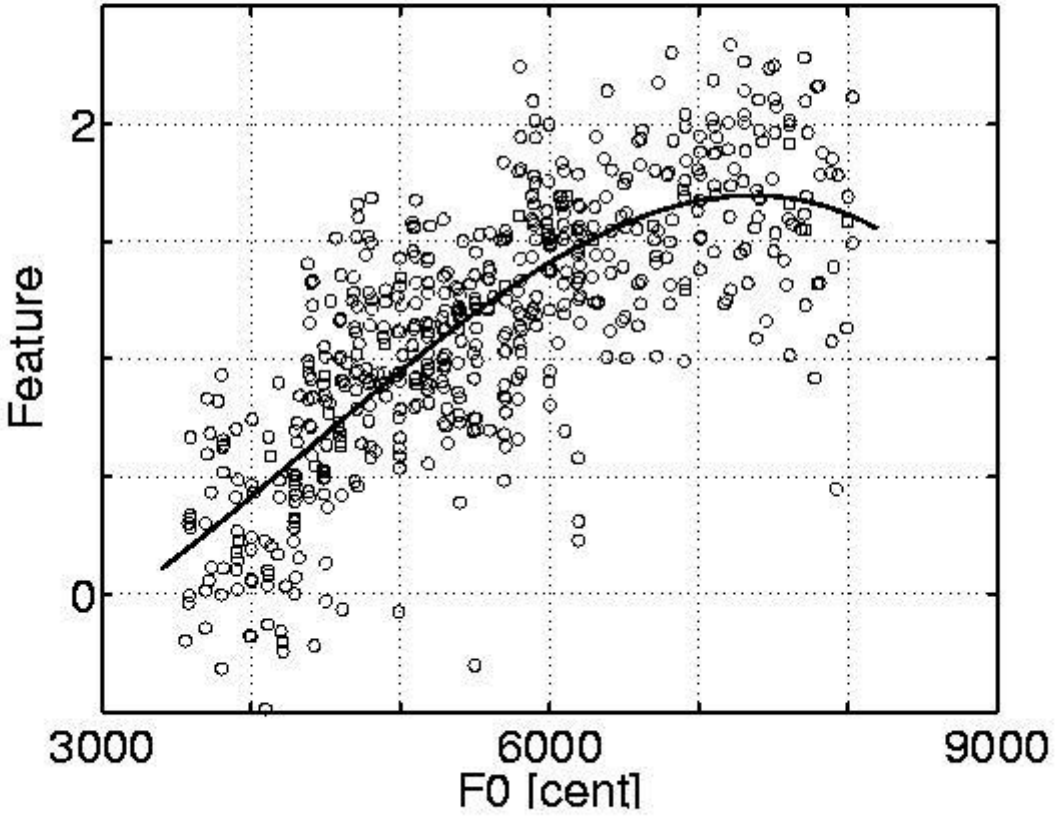
音高によって変化する分布の平均を
関数近似（3次関数）により推定



3. 本発表における提案手法

代表値関数

音高によって変化する分布の平均を
関数近似（3次関数）により推定



3. 本発表における提案手法

F0正規化共分散行列

代表値関数からのちらばりの程度を表す

↓

音高による音色変化を表現

3. 本発表における提案手法

F0正規化共分散行列

代表値関数からのちらばりの程度を表す

↓

音高による音色変化を表現

⇒**音高以外の要因による音色変化**を表す

3. 本発表における提案手法

F0正規化共分散行列

代表値関数からのちらばりの程度を表す

↓

音高による音色変化を表現

⇒**音高以外の要因による音色変化**を表す

↓

音色空間を**代表値関数で正規化**してから，
共分散行列を求める

↓

音高による音色変化を除去

3. 本発表における提案手法

ベイズ決定規則による識別

各楽器がF0依存多次元正規分布に従うと仮定

→　事後確率$p(\omega_i|x)$を最大にするω_iを見つける

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

$\log p(x|w_i;f)$：F0依存多次元正規分布の確率密度関数

$\log p(w_i;f)$：事前確率

$$-\frac{1}{2}D^2(x,m_i(f)) - \frac{1}{2}\log|\Sigma_i| + (\text{定数})$$

このgを最大にするω_iが同定結果

4. 処理の流れ

1. 特徴抽出（129個）
2. 主成分分析で次元圧縮
   （累積寄与率99%で79次元に圧縮）
3. 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）
4. F0依存多次元正規分布のパラメータ推定
5. ベイズ決定規則に基づいて楽器名を同定
6. 出力は楽器名だけでなくカテゴリーも

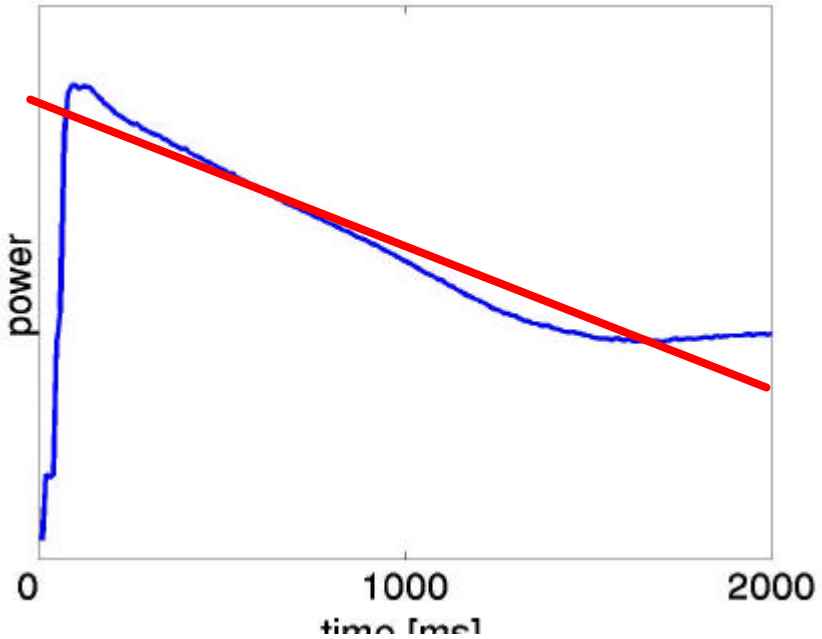
4. 処理の流れ

1. 特徴抽出（129個）

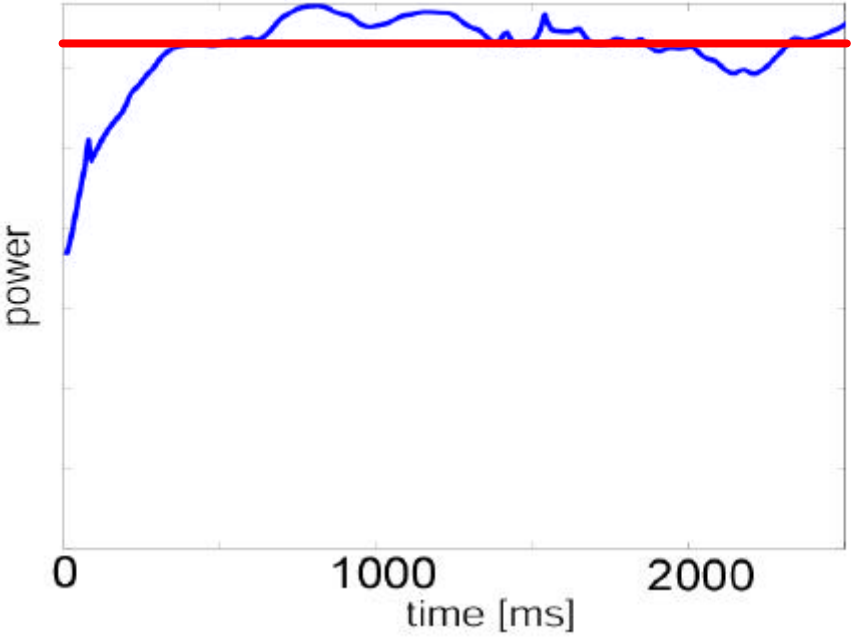
(1) スペクトルに関する定常的特徴（40個）
　　周波数重心, etc
(2) パワーの時間変化に関する特徴（35個）
　　パワー包絡線の線形最小二乗法による
　　近似直線の傾き, etc
(3) 各種変調の振幅／振動数（32個）
　　振幅変調, 周波数変調,
　　周波数重心の時間変化, MFCCの時間変化
(4) 発音開始直後のピーク尖度に関する特徴（22個）

パワー包絡線の線形最小二乗法による近似直線

ピアノ



フルート

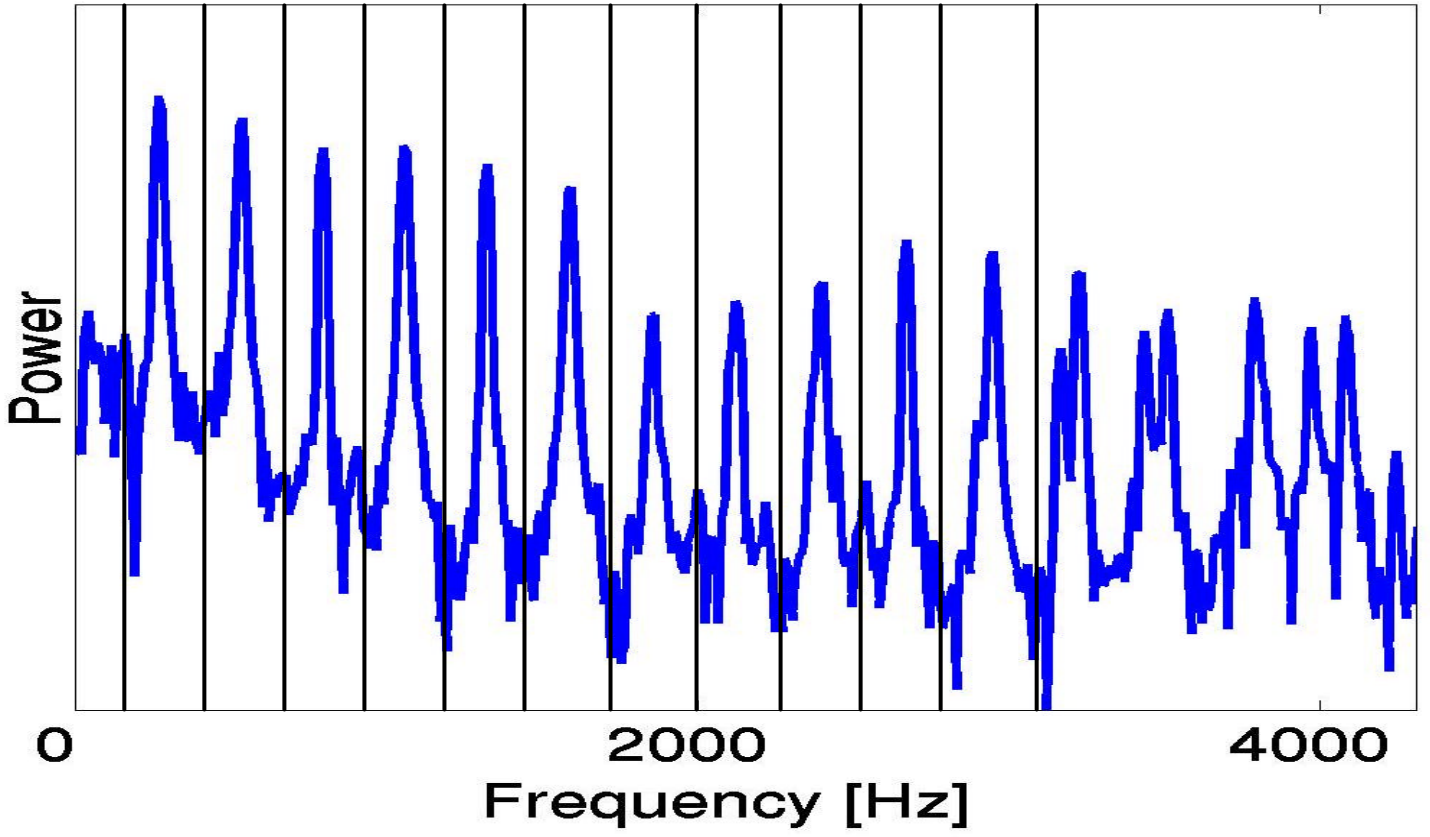


4. 処理の流れ

1. 特徴抽出（129個）

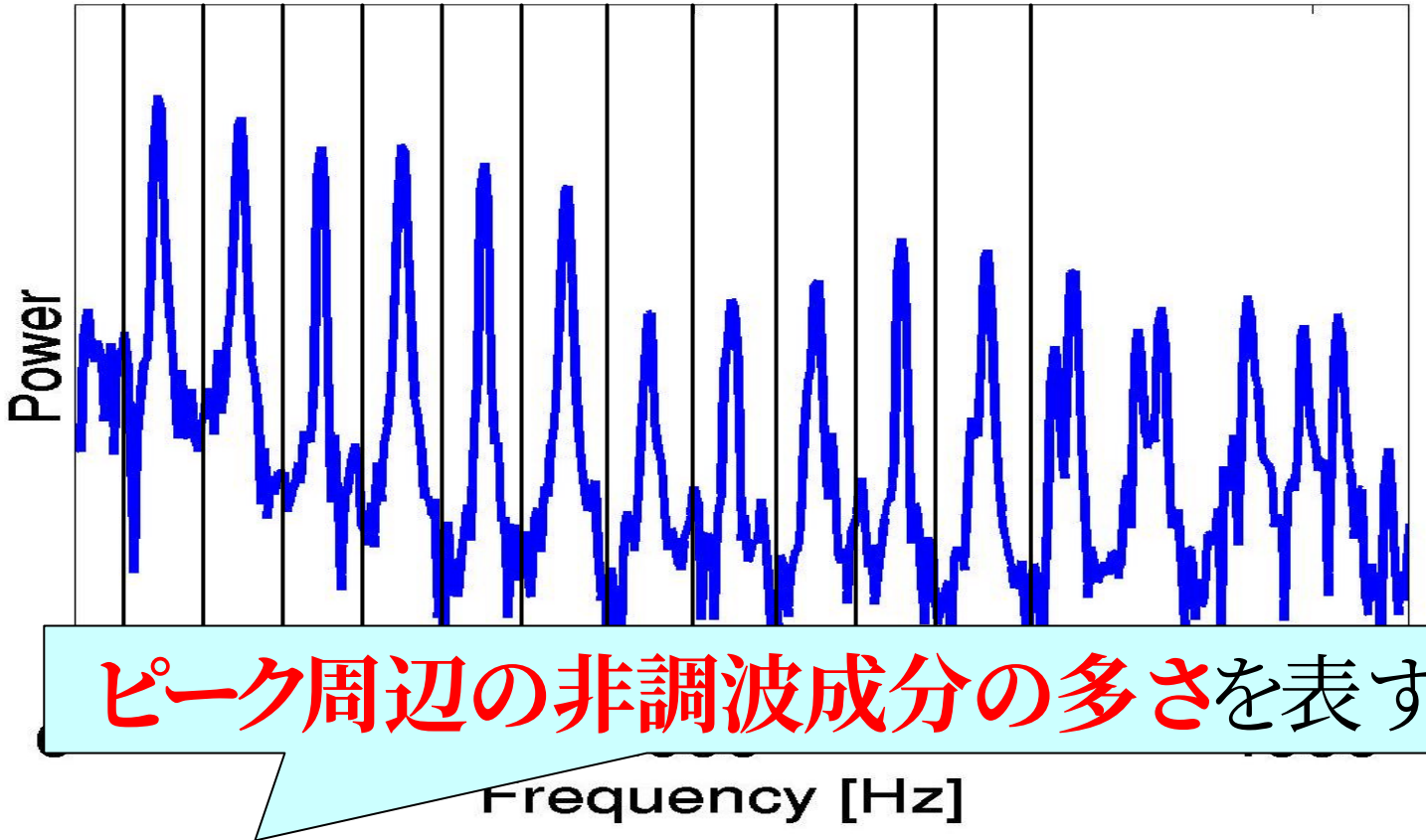
(1) スペクトルに関する定常的特徴（40個）
　　周波数重心, etc
(2) パワーの時間変化に関する特徴（35個）
　　パワー包絡線の線形最小二乗法による
　　近似直線の傾き, etc
(3) 各種変調の振幅／振動数（32個）
　　振幅変調, 周波数変調,
　　周波数重心の時間変化, MFCCの時間変化
(4) 発音開始直後のピーク尖度に関する特徴（22個）

発音開始直後のピーク尖度に関する特徴



各周波数成分（1〜11次倍音まで）を取り出し，
各ピークの尖度（とんがり度）を
4次モーメントから算出

発音開始直後のピーク尖度に関する特徴



ピーク周辺の非調波成分の多さを表す

各周波数成分（1～11次倍音まで）を取り出し，
各ピークの尖度（とんがり度）を
4次モーメントから算出

4. 処理の流れ

1. 特徴抽出（129個）
2. 主成分分析で次元圧縮
   （累積寄与率99%で79次元に圧縮）
3. 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）
4. F0依存多次元正規分布のパラメータ推定
5. ベイズ決定規則に基づいて楽器名を同定
6. 出力は楽器名だけでなくカテゴリーも

ピアノ	ピアノ(PF)	
ギター	クラシックギター(CG) ウクレレ(UK)	アコースティック ギター(AG)
弦楽器	バイオリン(VN) ビオラ(VL)	チェロ(VC)
金管楽器	トランペット(TR)	トロンボーン(TB)
サックス	ソプラノサックス(SS) アルトサックス(AS)	テナーサックス(TS) バリトンサックス(BS)
複簧楽器	オーボエ(OB)	ファゴット(FG)
クラリネット	クラリネット(CL)	
無簧楽器	ピッコロ(PC) フルート(FL)	リコーダー(RC)

4. 処理の流れ

1. 特徴抽出（129個）
2. 主成分分析で次元圧縮
   （累積寄与率99%で79次元に圧縮）
3. 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）
4. F0依存多次元正規分布のパラメータ推定
5. ベイズ決定規則に基づいて楽器名を同定
6. 出力は楽器名だけでなくカテゴリーも

5. 評価実験

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: 6247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.

5. 評価実験

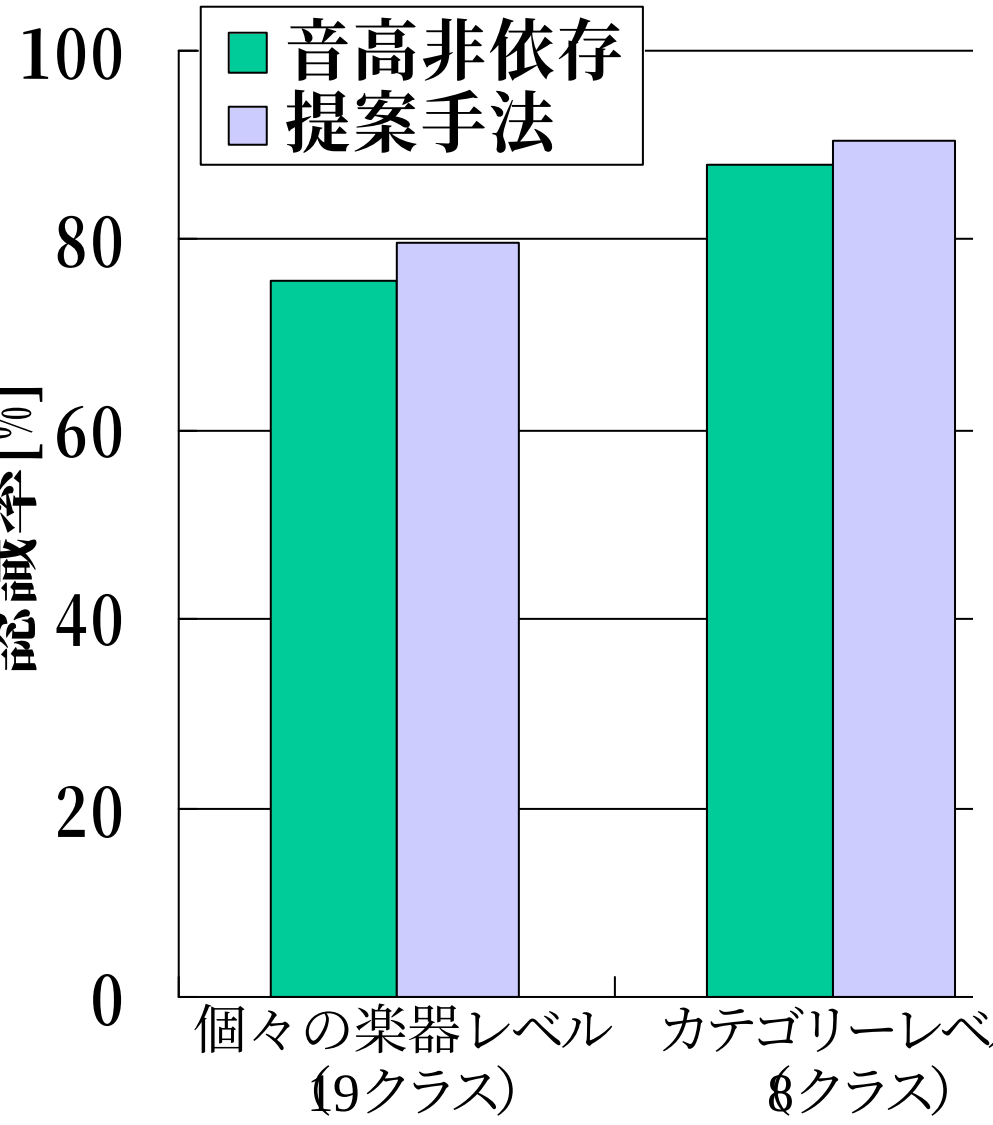
実 験 方 法

- 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
  - 使用したデータ総数: 6
- 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.

各グループχ_k(k=1,…,10)に対して,
「グループχ_k以外のデータで学習して
グループχ_kのデータで評価」を繰り返す.

5. 評価実験

実 験 結 果



個々の楽器レベルで**約80%**,
カテゴリーレベルで**約90%**の
認識率を実現

音高非依存に比べて,
個々の楽器レベルで4.00%,
カテゴリーレベルで2.45%,
認識率向上

誤り削減率は,
個々の楽器レベルで
16.48%,
カテゴリーレベルで20.67%

5. 評価実験

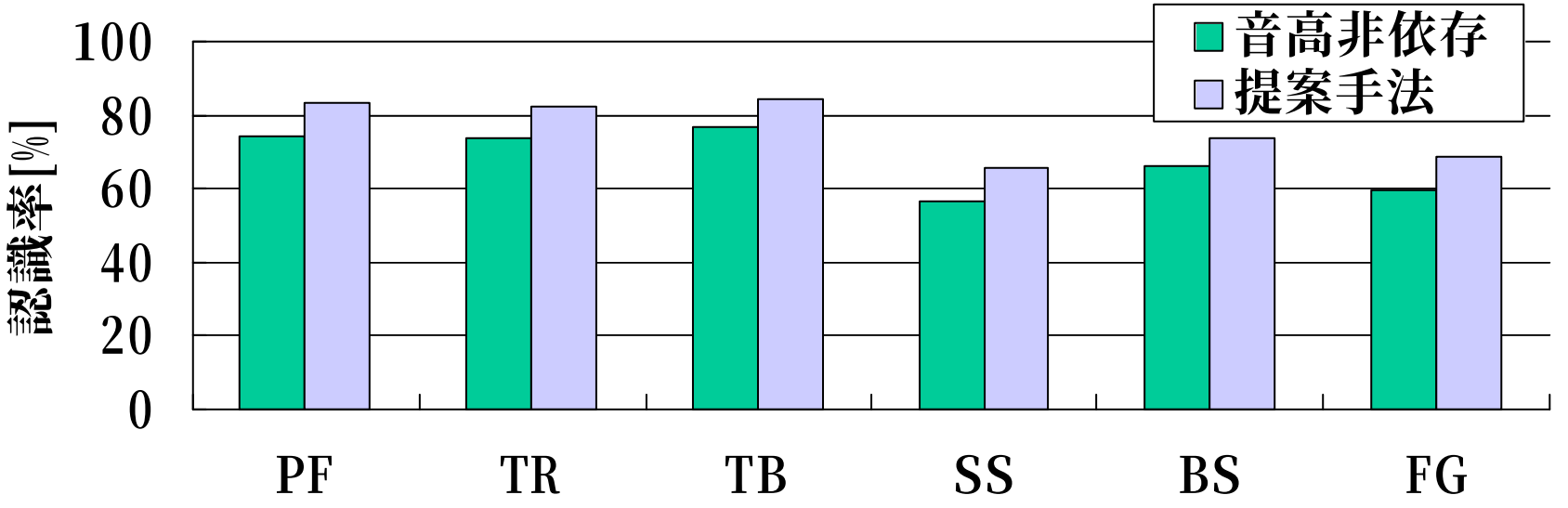
実験結果（個々の楽器レベル）

認識率 7%以上向上	ピアノ(PF) トランペット(TR) トロンボーン(TB)	ソプラノサックス(SS) バリトンサックス(BS) ファゴット(FG)
認識率 3%以上向上	バイオリン(VN) チェロ(VC) アルトサックス(AS)	ピッコロ(PC) フルート(FL)
認識率向上	アコースティックギター(AG) ビオラ(VL) テナーサックス(TS)	オーボエ(OB) クラリネット(CL)
変化なし	クラシックギター(CG)	ウクレレ(UK)
認識率低下	リコーダー(RC)	

5. 評価実験

実 験 結 果

認識率が7%以上改善された楽器（個々の楽器レベル）

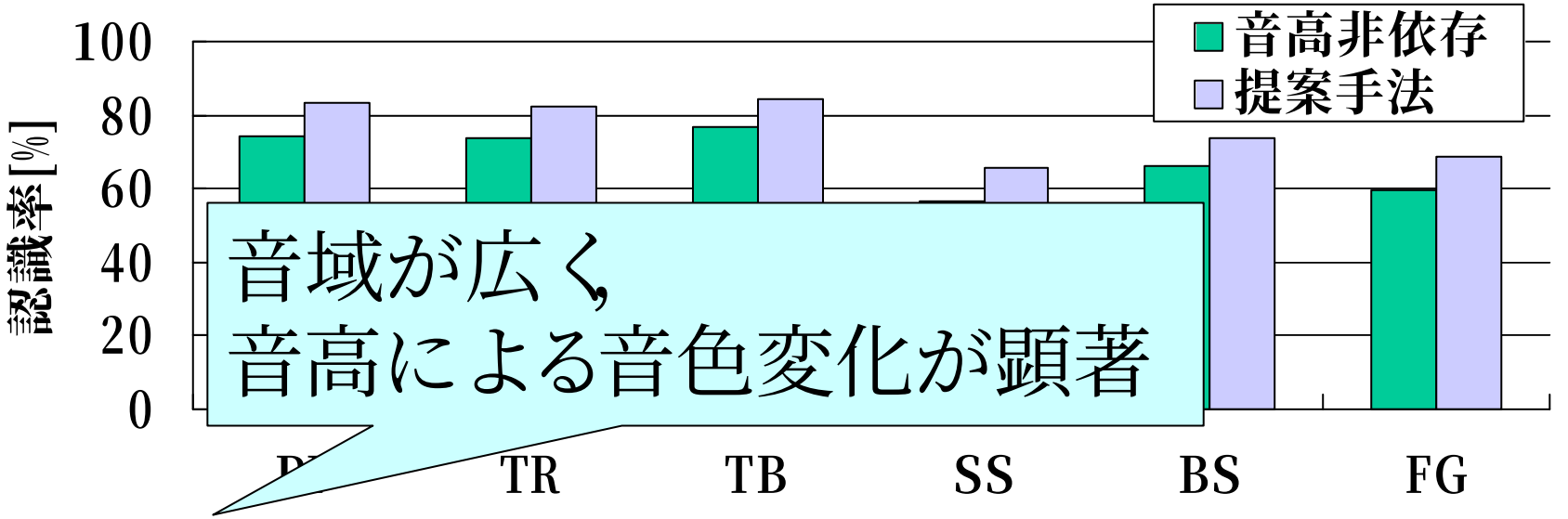


- ピアノ：最も性能改善
  （認識率9.06%改善，誤り削減35.13%）
- PF, TR, TBで約33～35%の認識誤りを削減
- SS, BS, FGでも20%以上の認識誤りを削減

5. 評価実験

実 験 結 果

認識率が7%以上改善された楽器（個々の楽器レベル）

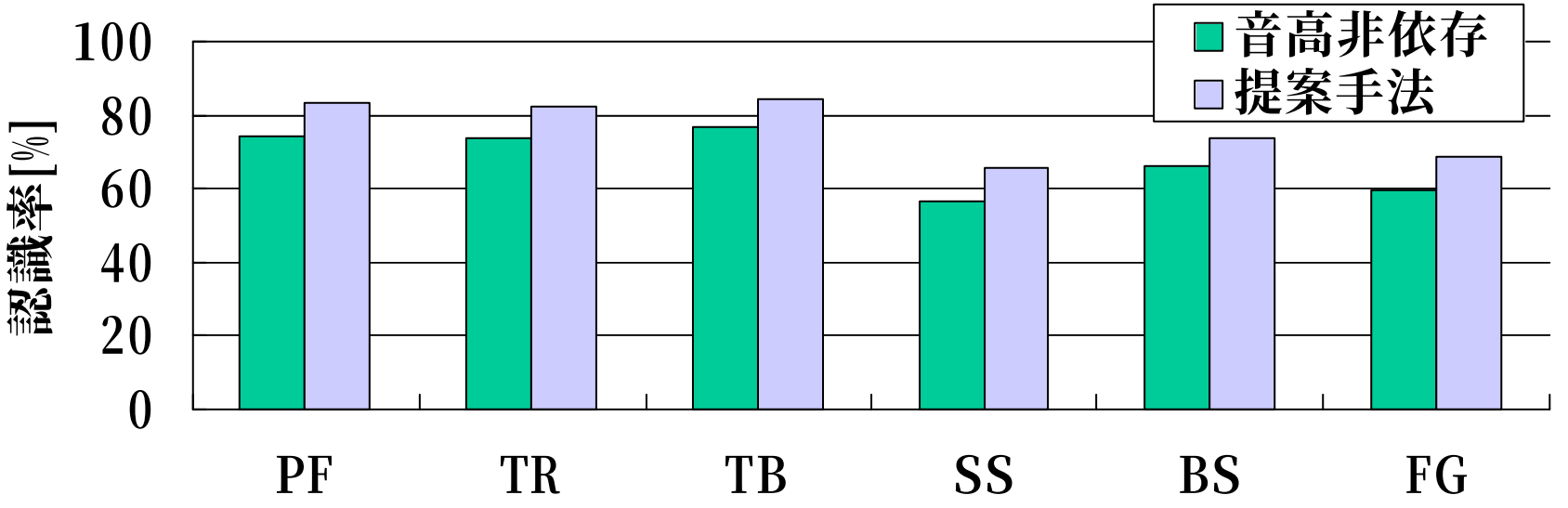


- **ピアノ**：最も性能改善
  （認識率9.06%改善，誤り削減35.13%）
- PF, TR, TBで**約33～35%**の認識誤りを削減
- SS, BS, FGでも**20%以上**の認識誤りを削減

5. 評価実験

実 験 結 果

認識率が7%以上改善された楽器（個々の楽器レベル）



- ピアノ：最も性能改善
  （認識率9.06%改善，誤り削減35.13%）
- 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

5. 評価実験

実験結果（個々の楽器レベル）

認識率 7%以上向上	ピアノ(PF) トランペット(TR) トロンボーン(TB)	ソプラノサックス(SS) バリトンサックス(BS) ファゴット(FG)
認識率 3%以上向上	バイオリン(VN) チェロ(VC) アルトサックス(AS)	ピッコロ(PC) フルート(FL)
認識率向上	アコースティックギター(AG) ビオラ(VL) テナーサックス(TS)	オーボエ(OB) クラリネット(CL)
変化なし	クラシックギター(CG)	ウクレレ(UK)
認識率低下	リコーダー(RC)	

5. 評価実験

実験結果 個々の楽器レベル）

認識率 7%以上向上	ピアノ(PF)	ソプラノサックス(SS)
	トランペット(TR)	バリトンサックス(BS)
	トロンボーン(TB)	ファゴット(FG)
認識率 3%以上向上	バイオリン(VN)	ピッコロ(PC)
	チェロ(VC)	フルート(FL)
	アルトサックス(AS)	
認識率向上	アコー	ボエ(OB)
	ビオ	リネット(CL)
	テナーサックス(TS)	
変化なし	クラシックギター(CG)	ウクレレ(UK)
認識率低下	リコーダー(RC)	

音高非依存」でも
90%以上の認識率

5. 評価実験

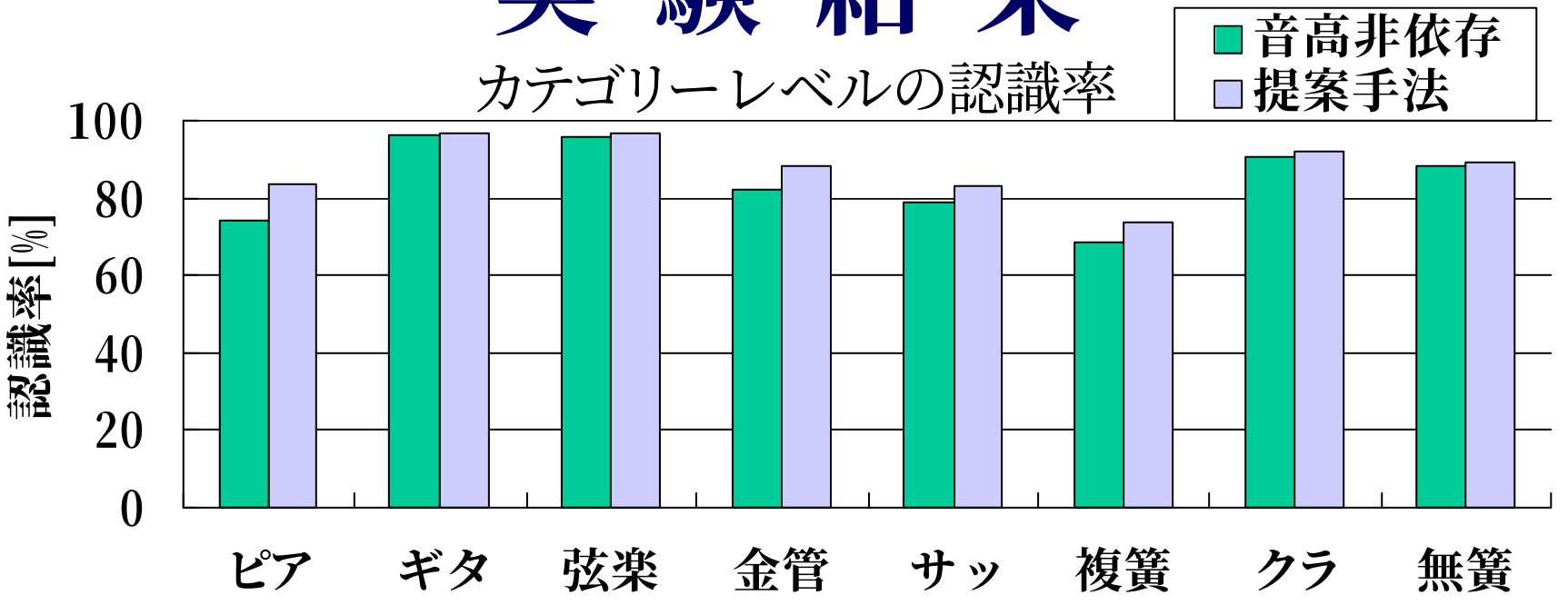
実験結果 個々の楽器レベル）

認識率 7%以上向上	ピアノ(PF) トランペット(TR) トロンボーン(TB)	ソプラノサックス(SS) バリトンサックス(BS) ファゴット(FG)
認識率 3%以上向上	バイオリン(VN) チェロ(VC) アルトサックス(AS)	ピッコロ(PC) フルート(FL)
認識率向上	アコースティックギター(AG) ビオラ(VL)	オーボエ(OB) (CL)
変化なし		K)
認識率低下	リコーダー(RC)	

160個のデータのうち,
誤認識が1個増えたに過ぎない.

5. 評価実験

実 験 結 果

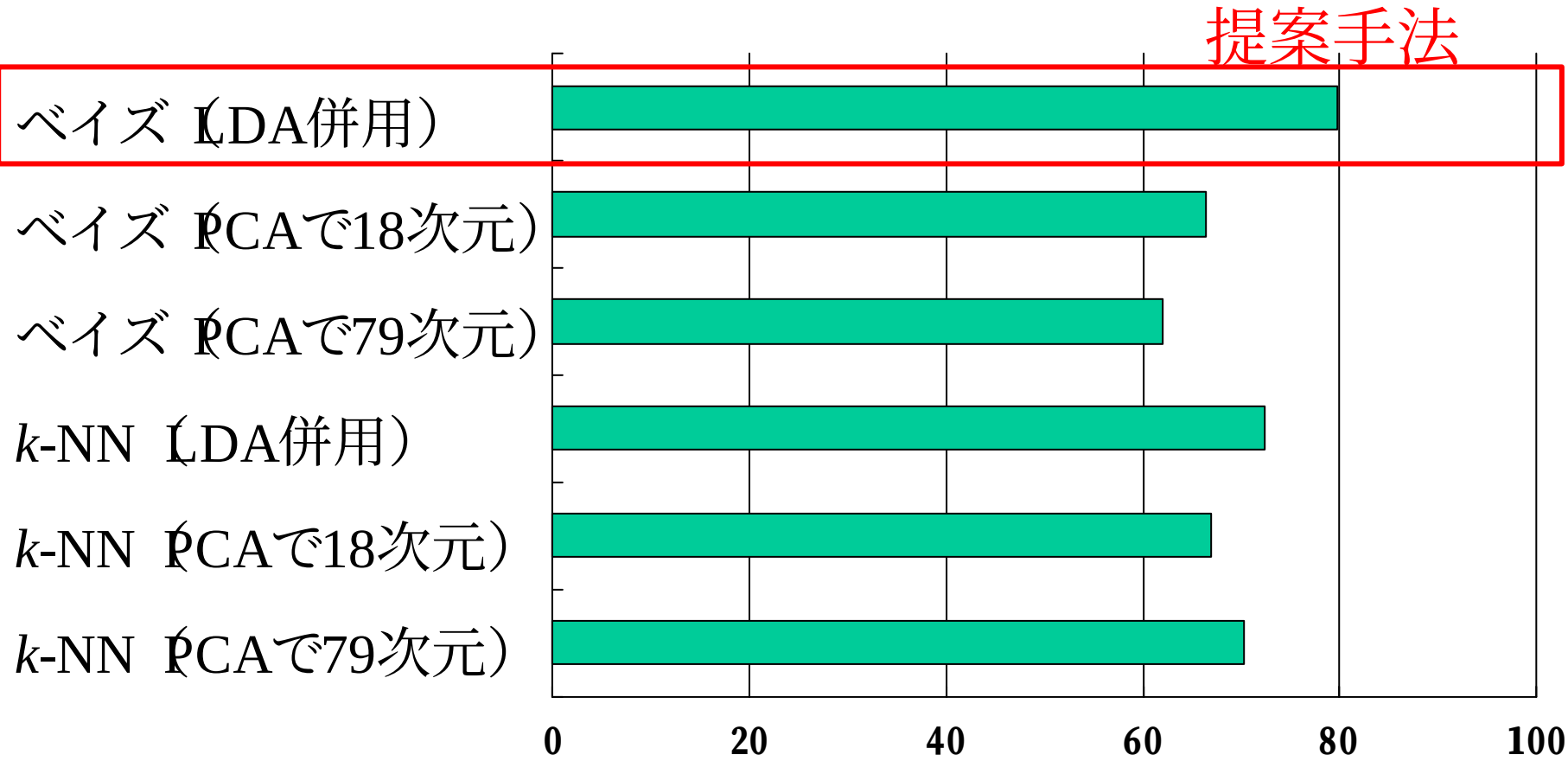


誤り削減 35% 8% 23% 33% 20% 13% 15% 8%

- すべてのカテゴリーで認識率改善
- ギター，弦楽器の認識率（提案手法）：96.7%
- 最も低いカテゴリーでも72%の認識率（提案手法）

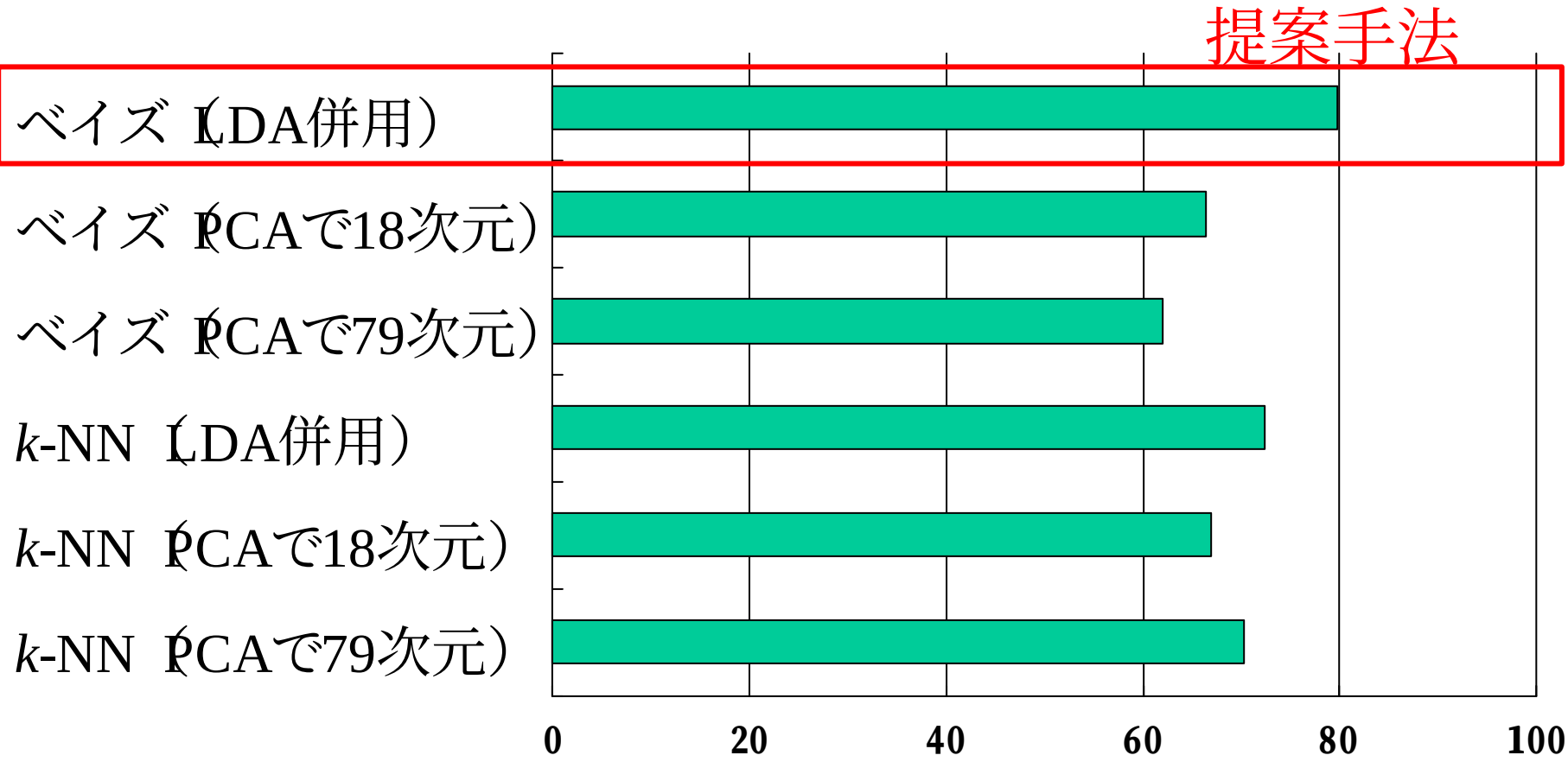
5. 評価実験

k-NN法との比較



提案手法が最も認識率が高い

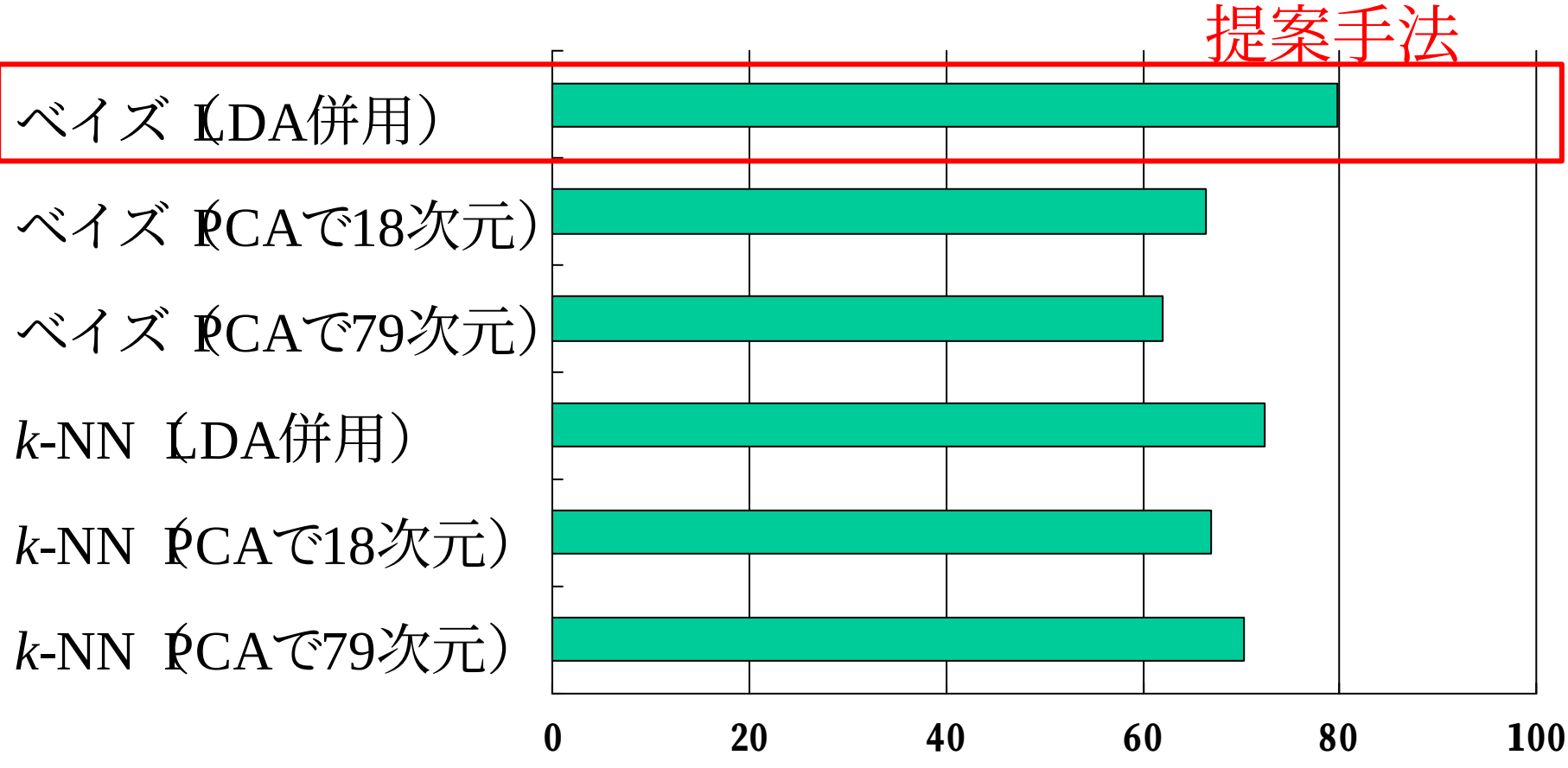
5. 評価実験
k-NN法との比較



79次元でのベイズ決定規則が最も認識率低い

∵　データ数に対して次元が高すぎる

5. 評価実験
k-NN法との比較



LDA併用により認識率向上

∵ LDAはクラス間分離を考慮した次元圧縮法

6. まとめ

- 音高による音色変化を考慮するため，
  F0依存多次元正規分布を提案
- F0依存多次元正規分布のための
  識別関数をベイズ決定規則から定式化
  ⇒**音源同定の性能向上に貢献**
  個々の楽器で16.48%，
  カテゴリーレベルで20.67%認識誤りを削減）
- 今後の課題
  - ベイズ決定規則以外への応用
  - より大規模な実験，混合音への適用など