

## 歌声情報処理: 歌声を対象とした音楽情報処理

産業技術総合研究所  
後藤 真孝 齋藤 毅 中野 倫靖 藤原 弘将



2010/07/28 情報処理学会 音楽情報科学研究会

### 歌声情報処理

- 「歌声情報処理」
  - 歌声を対象とした音楽情報処理
- 音楽は産業・文化の面で主要なコンテンツ
  - 歌声は音楽の最も重要な要素の一つ
  - 歌声を中心に音楽を聴く人達も多い

↓

- 歌声情報処理の研究成果は社会的にも大きなインパクト
- 様々な歌声関連技術が社会的に関心を集める



### 歌声情報処理

- 「歌声情報処理」
  - 学術的な観点からだけでなく、産業応用的な観点からも注目
    - 歌声の音高を信号処理で自動補正 (例: Auto-Tune)
    - ハミング検索 (例: midomi)
    - 歌声合成 (例: VOCALOID)
    - カラオケでの歌声評価(採点)機能
  - 研究対象は多岐にわたり拡大中

↓

産総研での研究事例を紹介



### 歌声情報処理システム

- 三つのカテゴリ
  1. 歌声を聴いて理解するシステム
    - LyricSynchronizer
    - Singer ID
    - MiruSinger
    - Hyperlinking Lyrics
    - Breath Detection
  2. 歌声に基づく音楽情報検索システム
    - VocalFinder
    - Voice Drummer
  3. 歌声合成システム
    - SingBySpeaking
    - VocalListener



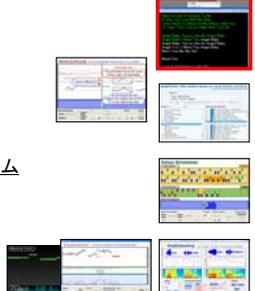
### 歌声情報処理システムを可能にする技術

- 共通して用いた主要技術

	混合音中の 歌声抽出	歌詞・音調 の認識	歌声の音質 の認識	歌声の音高 のモデル化
1. 歌声理解				
LyricSynchronizer	V	V(アラインメント)		
Singer ID	V		V(歌手の声質)	
MiruSinger	V			V(ビブラート)
Hyperlinking Lyrics	V	V(スポッティング)		
Breath Detection				
2. 音楽情報検索				
VocalFinder	V		V(声質類似度)	V(ΔF0)
Voice Drummer		V(ロドラマ)		
3. 歌声合成				
SingBySpeaking		V(音声)		V(F0軌跡)
VocalListener		V(アラインメント)		V(ビブラート)

### 歌声情報処理システム

- 三つのカテゴリ
  1. 歌声を聴いて理解するシステム
    - LyricSynchronizer
    - Singer ID
    - MiruSinger
    - Hyperlinking Lyrics
    - Breath Detection
  2. 歌声に基づく音楽情報検索システム
    - VocalFinder
    - Voice Drummer
  3. 歌声合成システム
    - SingBySpeaking
    - VocalListener



### LyricSynchronizer [Fujihara, Goto, Okuno, 2006-]

□ 混合音中の歌声と歌詞の時間的対応付けシステム

- 楽曲の再生と同期して色が変わる歌詞を見る
- 歌詞上の任意の単語をクリックするとそこから再生

現在の再生位置

クリックした場所から再生可能

### 歌声情報処理システムを可能にする技術

□ 共通して用いた主要技術

	混合音中の歌声抽出	歌詞・音韻の認識	歌声の音質の認識	歌声の音質のモデル化
1. 歌声理解				
LyricSynchronizer	V	V(アライメント)		

自動ボーカル抽出手法 [Goto, 1999-]  
メロディー音高推定手法 PreFEst によって  
様々な楽器音が含まれる混合音からボーカルを抜き出す

混合音 → ボーカル

### 歌声情報処理システムを可能にする技術

□ 共通して用いた主要技術

	混合音中の歌声抽出	歌詞・音韻の認識	歌声の音質の認識	歌声の音質のモデル化
1. 歌声理解				
LyricSynchronizer	V	V(アライメント)		

自動歌詞同期手法 [Fujihara, Goto, Okuno, 2006-]  
抽出したボーカル中の各音素の位置を  
歌声用音響モデル(HMM)を用いてViterbiアライメント

### 歌声情報処理システム

□ 三つのカテゴリ

- 歌声を聴いて理解するシステム
  - LyricSynchronizer
  - Singer ID
  - MiruSinger
  - Hyperlinking Lyrics
  - Breath Detection
- 歌声に基づく音楽情報検索システム
  - VocalFinder
  - Voice Drummer
- 歌声合成システム
  - SingBySpeaking
  - VocalListener

### Singer ID [Fujihara, et al., 2005-]

□ 混合音中の歌手名同定(ID: identification)するシステム

- 入力した楽曲中の歌声の歌手名を自動的に同定  
話者認識に近い
- 歌手名がメタデータに記録されていなくても楽曲検索可能

楽曲: "Open Arms" アーティスト: Journey → 歌手名: Steve Perry

楽曲: "With or Without You" アーティスト: U2 → 歌手名: Bono

### 歌声情報処理システムを可能にする技術

□ 共通して用いた主要技術

	混合音中の歌声抽出	歌詞・音韻の認識	歌声の音質の認識	歌声の音質のモデル化
1. 歌声理解				
LyricSynchronizer	V	V(アライメント)		
Singer ID	V		V(歌手の音質)	

自動ボーカル抽出手法 [Goto, 1999-]  
メロディー音高推定手法 PreFEst によって  
様々な楽器音が含まれる混合音からボーカルを抜き出す

混合音 → ボーカル

### 歌声情報処理システムを可能にする技術

共通して用いた主要技術

	混合音中の 歌声抽出	歌詞・音調 の認識	歌声の音質 の認識	歌声の音質 のモデル化
1. 歌声理解				
LyricSynchronizer	V	V(アライメント)		
Singer ID	V		V(歌手の音質)	
MiruSinger	V			V(ピブラート)

**歌声の音質モデリング手法** [Fujihara, et al., 2005-]  
歌声らしさの高いフレームの特徴量(LPMCC)を用いて  
各歌手ごとに音質をモデリングしたGMMを学習

### 歌声情報処理システム

三つのカテゴリ

- 歌声を聴いて理解するシステム
  - LyricSynchronizer
  - Singer ID
  - MiruSinger**
  - Hyperlinking Lyrics
  - Breath Detection
- 歌声に基づく音楽情報検索システム
  - VocalFinder
  - Voice Drummer
- 歌声合成システム
  - SingBySpeaking
  - VocalListener

### MiruSinger

[Nakano, Goto, Hiraga, 2007-]

歌唱力向上支援システム

- 既存の楽曲のボーカルの歌い方に忠実に歌いたい!
- 混合音中のボーカルを分析して可視化
- それに合わせてユーザの歌声も比較表示
- リアルタイムに音高(F0)が可視化され、ピブラート区間も表示

### 歌声情報処理システムを可能にする技術

共通して用いた主要技術

	混合音中の 歌声抽出	歌詞・音調 の認識	歌声の音質 の認識	歌声の音質 のモデル化
1. 歌声理解				
LyricSynchronizer	V	V(アライメント)		
Singer ID	V		V(歌手の音質)	
MiruSinger	V			V(ピブラート)

**自動ボーカル抽出手法** [Goto, 1999-]  
メロディー音高推定手法 PreFEst によって  
様々な楽器音が含まれる混合音からボーカルを抜き出す

### 歌声情報処理システムを可能にする技術

共通して用いた主要技術

	混合音中の 歌声抽出	歌詞・音調 の認識	歌声の音質 の認識	歌声の音質 のモデル化
1. 歌声理解				
LyricSynchronizer	V	V(アライメント)		
Singer ID	V		V(歌手の音質)	
MiruSinger	V			V(ピブラート)

**自動ピブラート検出手法** [Nakano, Goto, Hiraga, 2006-]  
ΔF0の短時間フーリエ変換(STFT)に基づいて  
「ピブラートらしさ」を求める

### 歌声情報処理システム

三つのカテゴリ

- 歌声を聴いて理解するシステム
  - LyricSynchronizer
  - Singer ID
  - MiruSinger
  - Hyperlinking Lyrics**
  - Breath Detection
- 歌声に基づく音楽情報検索システム
  - VocalFinder
  - Voice Drummer
- 歌声合成システム
  - SingBySpeaking
  - VocalListener

### Hyperlinking Lyrics [Fujihara, Goto, Ogata, 2008-]

□ 歌詞中の共通するフレーズ間へのリンク作成システム

- ある曲の歌詞中のフレーズから別の曲の歌詞中の同一フレーズへとリンクを作成

歌詞が既知の楽曲      歌詞が未知の楽曲

### 歌声情報処理システムを可能にする技術

□ 共通して用いた主要技術

	混合音中の歌声抽出	歌詞・音韻の認識	歌声の音質の認識	歌声の音質のモデル化
1. 歌声理解				
LyricSynchronizer	V	V(アライメント)		
Singer ID	V		V(歌手の音質)	
MiruSinger	V			V(ピブラート)
Hyperlinking Lyrics	V	V(スポッティング)		

自動ボーカル抽出手法 [Goto, 1999-]  
メロディー音高推定手法 PreFEST によって様々な楽器音が含まれる混合音からボーカルを抜き出す

混合音 → ボーカル

### 歌声情報処理システムを可能にする技術

□ 共通して用いた主要技術

	混合音中の歌声抽出	歌詞・音韻の認識	歌声の音質の認識	歌声の音質のモデル化
1. 歌声理解				
LyricSynchronizer	V	V(アライメント)		
Singer ID	V		V(歌手の音質)	
MiruSinger	V			V(ピブラート)
Hyperlinking Lyrics	V	V(スポッティング)		

キーワードスポッティング手法 [Fujihara, Goto, Ogata, 2008-]  
抽出したボーカル中の各キーフレーズの位置を歌声用音響モデル(HMM)を用いてViterbiアライメント

### 歌声情報処理システム

□ 三つのカテゴリ

1. 歌声を聴いて理解するシステム
  - LyricSynchronizer
  - Singer ID
  - MiruSinger
  - Hyperlinking Lyrics
  - Breath Detection
2. 歌声に基づく音楽情報検索システム
  - VocalFinder
  - Voice Drummer
3. 歌声合成システム
  - SingBySpeaking
  - VocalListener

### Breath Detection [Nakano, Ogata, Goto, Hiraga, 2008-]

□ 無伴奏の単独歌唱中のブレス音を自動検出するシステム

- 検出したブレス音の用途
  - (1) 歌声の収録においてブレスを削除、強調
  - (2) メロディーのフレーズ境界の検出、楽曲の構造分析
  - (3) 歌唱力の自動評価

□ 三種類のHMM: ブレス音、有声区間、無音

ブレス音

### 歌声情報処理システム

□ 三つのカテゴリ

1. 歌声を聴いて理解するシステム
  - LyricSynchronizer
  - Singer ID
  - MiruSinger
  - Hyperlinking Lyrics
  - Breath Detection
2. 歌声に基づく音楽情報検索システム
  - VocalFinder
  - Voice Drummer
3. 歌声合成システム
  - SingBySpeaking
  - VocalListener

### VocalFinder [Fujihara, Goto, 2007-]

□ 歌声の**声質の類似度**に基づく楽曲検索システム

- ユーザの好みの楽曲の歌声に  
類似した歌声を持つ楽曲を**検索・発見**

元の楽曲 → 検索された楽曲

### 歌声情報処理システムを可能にする技術

□ 共通して用いた**主要技術**

1. **自動ボーカル抽出手法** [Goto, 1999-]  
メロディー音高推定手法 **PreFEst** によって  
様々な楽器音が含まれる**混合音**から**ボーカル**を抜き出す

2. **音楽情報検索**

Breath Detection				
VocalFinder	V		V(声質類似度)	V( $\Delta F0$ )
Voice Drummer		V(ドラム)		
SingBySpeaking		V(音高)		V(音高)
VocalListener		V(歌詞)		V(歌詞)

### 歌声情報処理システムを可能にする技術

□ 共通して用いた**主要技術**

1. **歌声の**声質モデリング**手法** [Fujihara, et al., 2005-]  
歌声らしさの**高いフレーム**の特徴量 (**LPMCC**,  $\Delta F0$ ) を用いて  
各曲ごとに**声質をモデリング**した**GMM**を学習

2. **音楽情報検索**

Breath Detection				
VocalFinder	V		V(声質類似度)	V( $\Delta F0$ )
Voice Drummer		V(ドラム)		
SingBySpeaking		V(音高)		V(音高)
VocalListener		V(歌詞)		V(歌詞)

### 歌声情報処理システム

□ **三つのカテゴリ**

- 歌声を聴いて理解するシステム  
LyricSynchronizer  
Singer ID  
MiruSinger  
Hyperlinking Lyrics  
Breath Detection
- 歌声に基づく**音楽情報検索システム**  
VocalFinder  
Voice Drummer
- 歌声**合成システム**  
SingBySpeaking  
VocalListener

### Voice Drummer [Nakano, Goto, Ogata, Hiraga, 2005-]

□ **口(くち)ドラム**による**ドラム譜入力システム**

- 口ドラム(ボイスパーカッション)によって**ドラムパターン**を検索  
ドラム音を真似た「ドンタンドタン」のような発声
- 既存の楽曲のドラムパートだけを差し替えて**編曲**

認識結果 → 口ドラム入力「ドンタンドタン」

### 歌声情報処理システムを可能にする技術

□ 共通して用いた**主要技術**

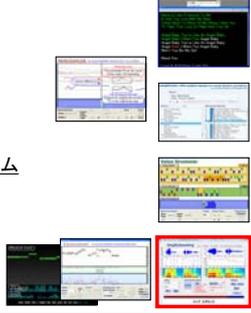
	混合音中の 歌声抽出	歌詞・音高 の認識	歌声の <b>声質</b> の認識	歌声の <b>音高</b> のモデル化
1. 歌声理解				
LyricSynchronizer	V	V(アラインメント)		
Singer ID	V		V(歌手の <b>声質</b> )	
MiruSinger	V			V(ビブラート)
Hyperlinking Lyrics	V	V(スポットニング)		
Breath Detection				
2. 音楽情報検索				
VocalFinder	V		V(声質類似度)	V( $\Delta F0$ )
Voice Drummer		V(口ドラム)		

**HMMI**に基づく**口ドラム認識手法**  
[Nakano, Goto, Ogata, Hiraga, 2005-]

### 歌声情報処理システム

□ 三つのカテゴリ

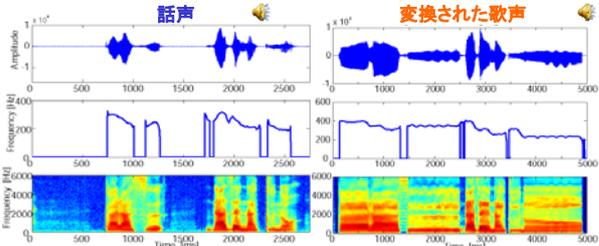
1. 歌声を聴いて理解するシステム
  - LyricSynchronizer
  - Singer ID
  - MiruSinger
  - Hyperlinking Lyrics
  - Breath Detection
2. 歌声に基づく音楽情報検索システム
  - VocalFinder
  - Voice Drummer
3. 歌声合成システム
  - SingBySpeaking
  - VocalListener



### SingBySpeaking [Saitou, Goto, 2007-]

□ 話しを歌声に変換する歌声合成システム

- 歌詞の朗読音声と楽譜情報を入力すると  
音高、音韻長、歌唱フォルマントを加工して歌声に変換



### 歌声情報処理システムを可能にする技術

□ 共通して用いた主要技術

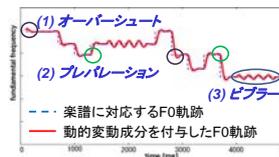
	混成音中の 数字抽出	歌聲・音韻 の認識	歌聲の音高 の認識	歌聲の音高 のモデル化
1. 歌声理解				
LyricSynchronizer	V	V(アラインメント)		
Singer ID	V		V(歌手の声質)	
MiruSinger				
Hyperlinking Lyrics				
Breath Detection				
2. 歌声に基づく音楽情報検索システム				
VocalFinder				
Voice Drummer		V(ロドラム)		
3. 歌声合成システム				
→ SingBySpeaking		V(音声)		V(F0軌跡)
VocalListener				

自動歌詞同期手法 [Fujihara, Goto, Okuno, 2006-]  
2. 話声中の各音素の位置を  
音響モデル(HMM)を用いてViterbiアラインメント

### 歌声情報処理システムを可能にする技術

□ 共通して用いた主要技術

歌声の音高(F0)軌跡生成手法  
楽譜情報に対して4種類のF0動的変動成分を付与



1. 歌声理解			
LyricSynchronizer			
Singer ID			
MiruSinger			
Hyperlinking Lyrics			
Breath Detection			
2. 歌声に基づく音楽情報検索システム			
VocalFinder			
Voice Drummer		V(ロドラム)	
3. 歌声合成システム			
→ SingBySpeaking		V(音声)	V(F0軌跡)
VocalListener			

### 歌声情報処理システム

□ 三つのカテゴリ

1. 歌声を聴いて理解するシステム
  - LyricSynchronizer
  - Singer ID
  - MiruSinger
  - Hyperlinking Lyrics
  - Breath Detection
2. 歌声に基づく音楽情報検索システム
  - VocalFinder
  - Voice Drummer
3. 歌声合成システム
  - SingBySpeaking
  - VocalListener

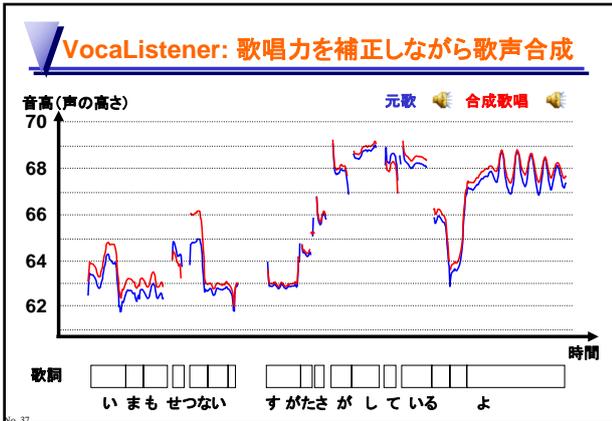


### VocalListener [Nakano, Goto, 2008-]

□ 歌声を歌声に変換する歌声合成システム

- ユーザの歌声の音高(F0)と音量を真似るように  
歌声合成ソフトウェアVOCALOIDのパラメータを推定





### 歌声情報処理システムを可能にする技術

共通して用いた主要技術

	混合音中の 歌声抽出	歌詞・音調 の認識	歌声の音質 の認識	歌声の音質 のモデル化
<b>1. 歌声理解</b>				
LyricSynchronizer	V	V(アライメント)		
Singer ID	V		V(歌手の音質)	
MiruSinger	V			V(ビブラート)
Hyperlinking Lyrics	V	V(スポッティング)		
Breath Detection				
<b>2. 音楽情報検索</b>				
VocalFinder	V		V(声質類似度)	V(ΔF0)
Voice Drummer		V(ロドラマ)		
<b>3. 歌声合成</b>				
SingBySpeaking		V(音声)		V(F0軌跡)
VocaListener		V(アライメント)		V(ビブラート)

### 歌声情報処理システム

三つのカテゴリ

- 歌声を聴いて理解するシステム**  
LyricSynchronizer  
Singer ID  
MiruSinger  
Hyperlinking Lyrics  
Breath Detection
- 歌声に基づく音楽情報検索システム**  
VocalFinder  
Voice Drummer
- 歌声合成システム**  
SingBySpeaking  
VocaListener

### 歌声情報処理システム

	後藤真孝
<b>1. 歌声を聴いて理解するシステム</b>	
LyricSynchronizer	藤原弘将, 後藤真孝, 奥乃博
Singer ID	藤原弘将, 北原鉄朗, 後藤真孝, 奥乃博
MiruSinger	中野倫博, 後藤真孝, 平賀謙
Hyperlinking Lyrics	藤原弘将, 後藤真孝, 緒方洋
Breath Detection	中野倫博, 後藤真孝, 緒方洋, 平賀謙
<b>2. 歌声に基づく音楽情報検索システム</b>	
VocalFinder	藤原弘将, 後藤真孝
Voice Drummer	中野倫博, 後藤真孝, 緒方洋, 平賀謙
<b>3. 歌声合成システム</b>	
SingBySpeaking	齋藤毅, 後藤真孝, 鶴木祐史, 赤木正人

### 歌声情報処理システムを可能にする技術

共通して用いた主要技術

	混合音中の 歌声抽出	歌詞・音調 の認識	歌声の音質 の認識	歌声の音質 のモデル化
<b>1. 歌声理解</b>				
LyricSynchronizer	V	V(アライメント)		
Singer ID	V		V(歌手の音質)	
MiruSinger	V			V(ビブラート)
Hyperlinking Lyrics	V	V(スポッティング)		
Breath Detection				
<b>2. 音楽情報検索</b>				
VocalFinder	V		V(声質類似度)	V(ΔF0)
Voice Drummer		V(ロドラマ)		
<b>3. 歌声合成</b>				
SingBySpeaking		V(音声)		V(F0軌跡)
VocaListener		V(アライメント)		V(ビブラート)

### おわりに

今後の展望

- 歌声は音声と音楽の両方の側面を持つ  
いずれの分野の観点からも未解決の研究課題は多い
- 音声認識の観点から  
歌声の自動認識: 技術的に最も難しいクラスの音声認識問題  
歌声は音声よりも概して変動が大きい  
歌声と関連し合う伴奏音が多い  
実際、伴奏を伴う歌声の歌詞の自動認識は実現できていない

**おわりに**

□ 今後の展望

- 歌声は音声と音楽の両方の側面を持つ  
いずれの分野の観点からも未解決の研究課題は多い
- 音声認識の観点から  
歌声の自動認識: 技術的に最も難しいクラスの音声認識問題
- 音楽の認識・理解の観点から  
歌声の自動認識: 技術的に難しくかつ興味深い課題が多い  
従来主に研究されてきた楽器音に比べ  
歌声の変動の大きさは並外れている
- 歌声の合成: まだまだ研究途上で課題が多い  
話声のように言語として内容が伝わる必要がある  
声の高さや強さ、声色の動的で複雑な変化  
歌声としての表現力が求められる

No. 43

**おわりに**

□ 今後の展望

- 歌声は音声と音楽の両方の側面を持つ  
いずれの分野の観点からも未解決の研究課題は多い
- 音声認識の観点から  
歌声の自動認識: 技術的に最も難しいクラスの音声認識問題
- 音楽の認識・理解の観点から  
歌声の自動認識: 技術的に難しくかつ興味深い課題が多い  
従来主に研究されてきた楽器音に比べ  
歌声の変動の大きさは並外れている
- 歌声の合成: まだまだ研究途上で課題が多い  
話声のように言語として内容が伝わる必要がある  
声の高さや強さ、声色の動的で複雑な変化  
歌声としての表現力が求められる

**歌声情報処理の研究は、  
学術的にもまさにフロンティア！**

No. 44

歌声情報処理を  
研究しませんか？

No. 45