



## 「あの人に歌ってほしい」を かなえるHMM歌声合成システム

◎大浦圭一郎（名工大）

間瀬絢美（名工大）

山田知彦（名工大）

◎徳田恵一（名工大）

後藤真孝（産総研）

# はじめに

HMM-based **Singing Voice Synthesis System**








- 歌声合成
  - VOCALOID, UTAU, Cadenciiなどの普及
  - 認知度の向上
- 要望：好きな声で好きな歌を歌わせたい
  - VOCALOIDシリーズ
  - UTAU（様々なUTAU音源）
- 課題：自然に歌わせるための調整作業が必要
  - 人手に基づく調整
  - 一般ユーザーには敷居が高い
  - ユーザー歌唱に基づく調整（VocaListenerなど）

# HMMに基づく音声合成手法

HMM-based **Singing Voice Synthesis System**



- HMMテキスト音声合成
  - 波形を保持しない
    - ⇒ **小さなメモリ容量で動作**
  - 統計モデルに基づく手法
    - ⇒ **適切なパラメータ変換で多様な声質が合成可能**
  - 全自動学習
    - ⇒ **話者の声質を再現**
- 合成音声サンプル
  - 読み上げ合成音声：日本語  英語  中国語 
  - 感情合成音声：平成  怒り 

# Sinsyプロジェクト

HMM-based **S**inging Voice **S**ynthesis System



## • 経緯

- 2002年 HMM歌声合成システム構築 [酒向他]
- 2004年 RENCON歌声部門参加 [大浦他]
- 2007年 VOCALOID2発表
- 2008年 VocaListener発表
- 2009年 5月 Sinsyプロジェクト発案 [後藤他]
- 2009年 8月 Sinsyプロジェクト開始 [徳田他]
- 2009年12月 [Sinsyオンラインデモページ](#)・[デモ動画公開](#)
- 2010年 4月 DTMマガジン「音楽情報処理最前線」

## • コンセプト

- CGM的・オープンソース的アクティビティの活性化
- 2次創作・N次創作への貢献

# 関連コミュニティにおける反響

HMM-based **Singing Voice Synthesis System**



- MusicXML関連ツールの開発・公開
  - ファイルフォーマットコンバータ型
  - スコアエディタ型
- 多くのMusicXMLのアップロード
  - 公開から昨日(2010/07/27)までで2059回
  - 利用規約 ⇒ 合成波形の自由な利用を保証
- ユーザー作品一覧
- キャラクタ案 : A, B, C
- twitter, blog等の情報
  - Sinsyオンラインデモページへのフィードバック

# Sinsyとは？

HMM-based **S**inging Voice **S**ynthesis System



- 名工大で開発されたHMM歌声合成システム  
HMM-based **S**inging Voice **S**ynthesis System
- ユーザーはオンラインデモページに楽譜をアップロードすることにより，歌声を合成可能
- HMMテキスト音声合成を歌声に応用
  - 多様な合成歌声
  - 調整作業が不要（ベタ打ちのみでOK）
  - 全自動学習
  - 少量のメモリ容量で動作（数Mbytes）
- オープンソースのソフトウェアを利用して構築

# 構築に利用したソフトウェア

HMM-based **Singing Voice Synthesis System**



- 波形の分析
  - [Speech Signal Processing Toolkit \(SPTK\)](#)
  - [STRAIGHT](#)
- HMMの学習
  - [HMM-based Speech Synthesis System \(HTS\)](#)
- 楽譜(MusicXML)の変換
  - [CrestMuseXML \(CMX\)](#)
- 波形の合成
  - [HMM-based Speech Synthesis Engine \(hts\\_engine API\)](#)

# 楽譜の入力ツール

HMM-based **Singing Voice Synthesis System**



楽譜(MusicXML)を出力できるツールを利用

## – 商用

- Cubase
- Finaleシリーズ

## – フリー

- Cadencii
- sinsyconv
- utau2sinsy
- VSQ to MusicXML Conversion
- MuseScore
- ust2musicxml
- vocalips2musicxml

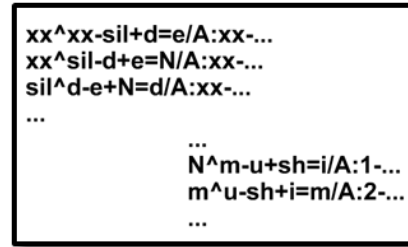
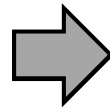


# HMM歌声合成システムの概要

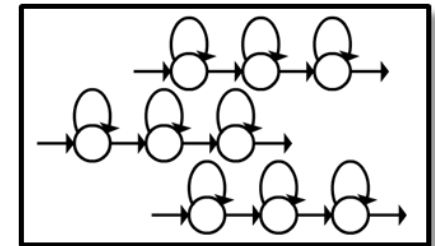
HMM-based **Singing Voice Synthesis System**



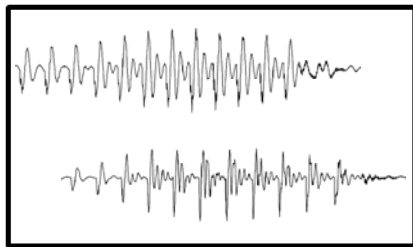
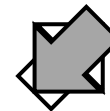
歌詞付き楽譜(MusicXML)



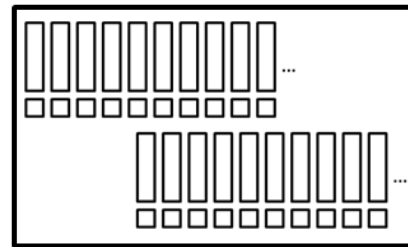
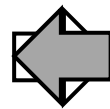
ラベル



HMM



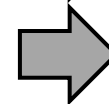
楽譜に対応した歌声波形



スペクトルと基本周波数  
パラメータ



学習ステップ



合成ステップ

# 導入手法1/5：コンテキスト

HMM-based **Singing Voice Synthesis System**



- テキスト音声合成  
音素, 品詞, アクセントなど
- **歌声合成のためのコンテキストを新たに定義**  
音素, モーラ, 音符, フレーズ, 曲  
⇒ 階層毎に整理, 対称性を考慮

音素	当該音素	o	o	k	i	n	a	k	u	r	i	n	o	k	i	n	o	s	i	t	a	d	e
モーラ	音符内モーラ位置	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	音符内モーラ長	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
音符	音高	C4	C4	D4	E4	E4	G4	E4	E4	D4	D4	C4											
	音長	4	8	8	8	8	4	8	8	8	8	4											
	相対音高	0	0	2	4	4	5	4	4	2	2	0											

⋮

# 導入手法2/5 : ビブラート

HMM-based **Singing Voice Synthesis System**



- ビブラート


基本周波数, 音量等を周期的に揺らす歌唱表現

Waveform



Log  $F_0$



 : Vibrato

- 対数基本周波数系列からビブラートを抽出

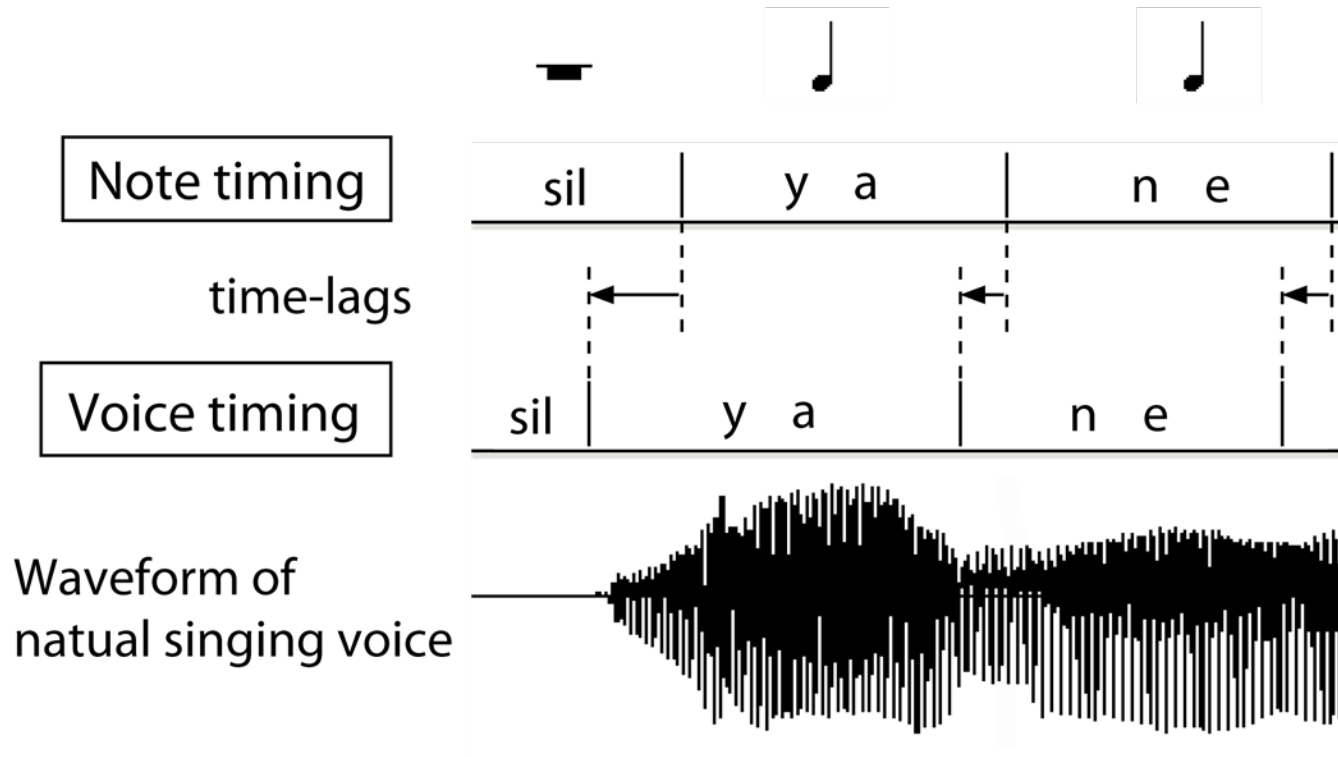
⇒ **スペクトル, 基本周波数と同様にモデル化**

# 導入手法3/5 : タイミング

HMM-based **Singing Voice Synthesis System**



音符と実際の発声開始時間のずれを学習



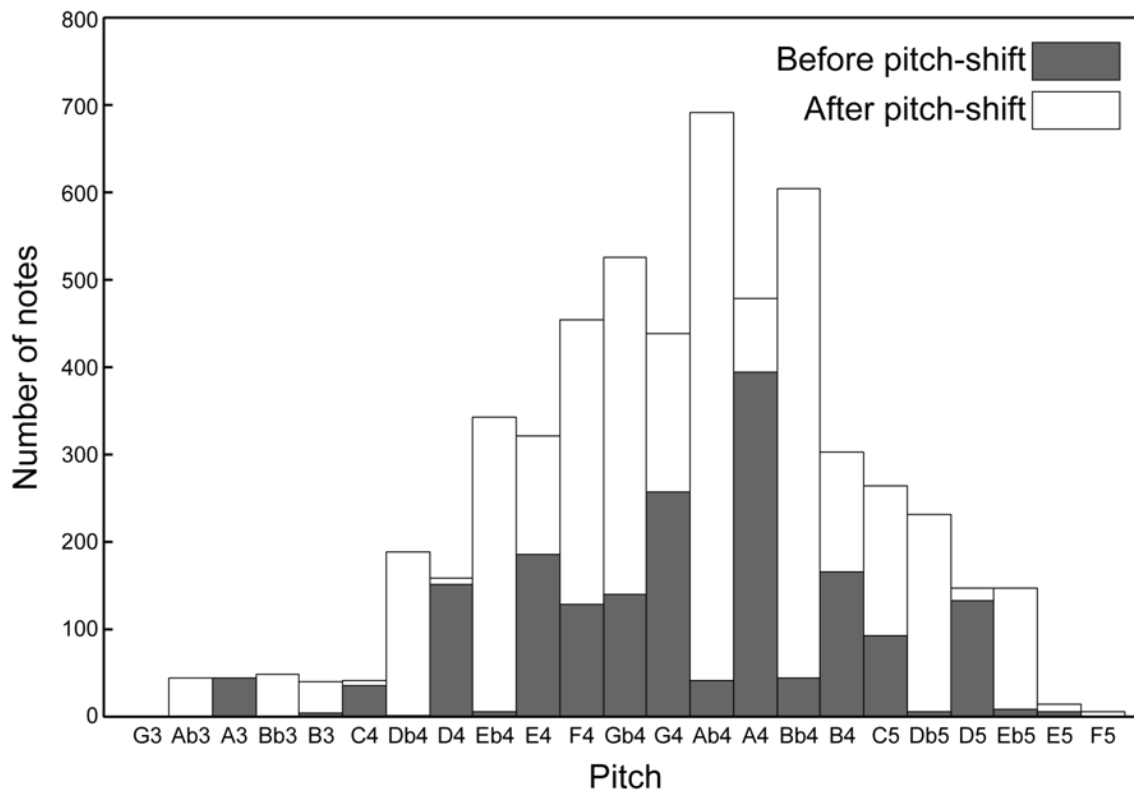
⇒ 継続長モデルと同様にモデル化

# 導入手法4/5：音高シフト

HMM-based **Singing Voice Synthesis System**



## 音高シフトによる疑似学習データ



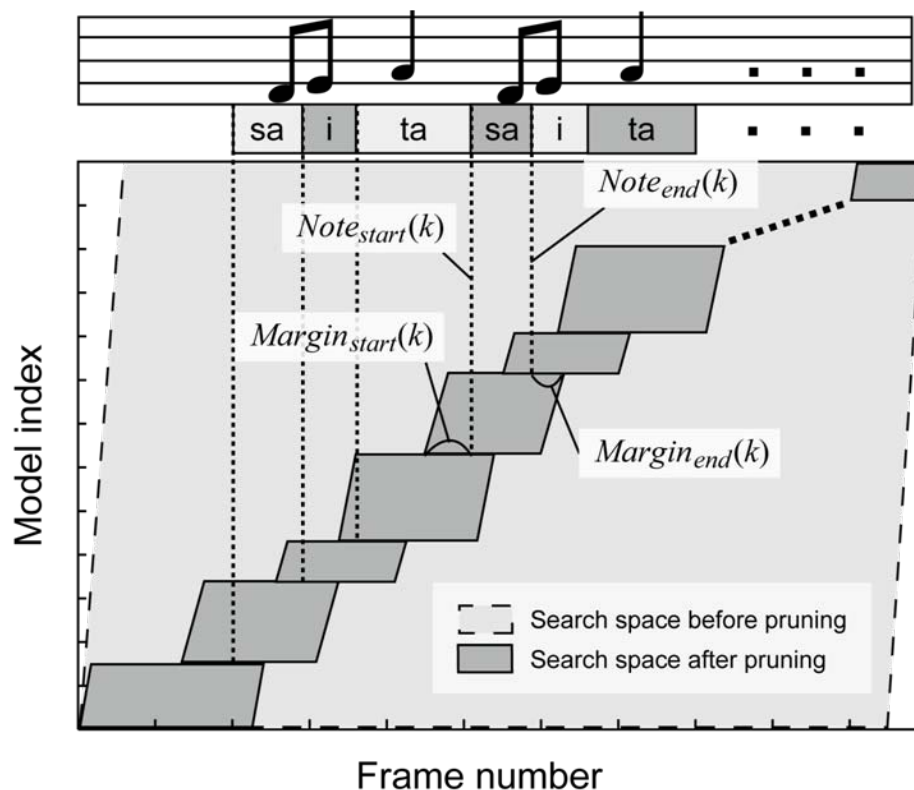
⇒ 学習データ内に少ない音高を適切にモデル化

# 導入手法5/5 : 学習の効率化

HMM-based Singing Voice Synthesis System



## 時間情報を用いた効率的な学習



⇒ 推定誤りの削減, 学習の高速化

# 話者適応

HMM-based **Singing Voice Synthesis** System



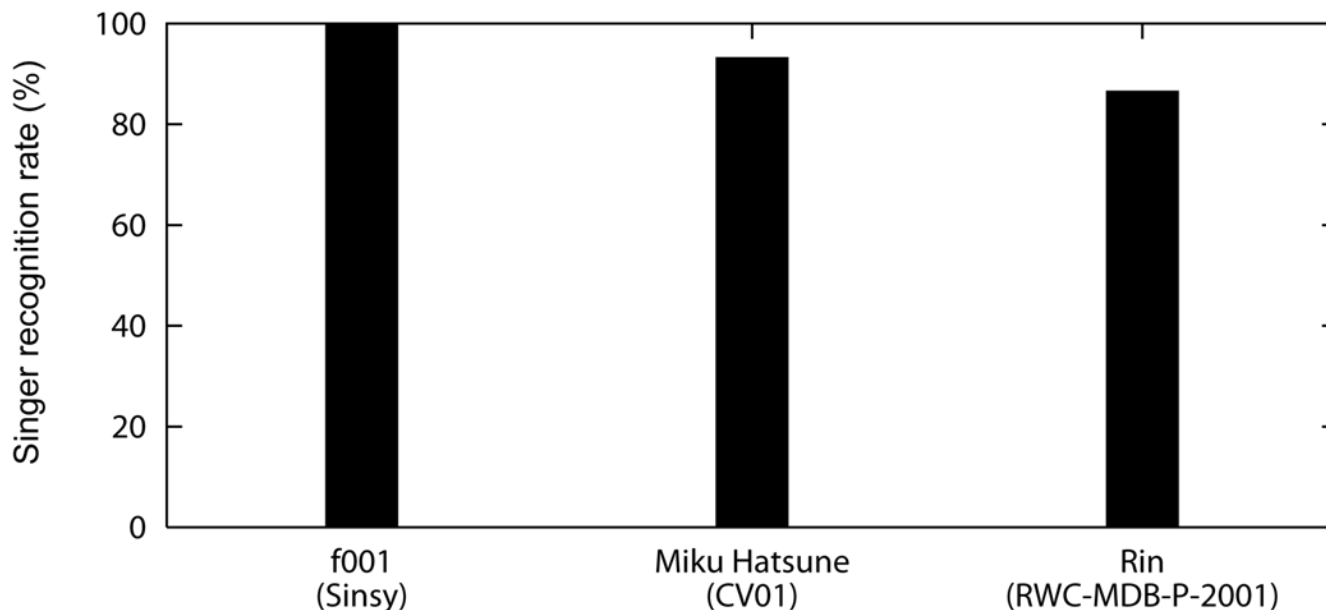
- 話者適応とは目標の声質にモデルを近づける手法
  - 声質・歌い方を真似る
  - 少量のデータでも動作するので低コスト
- 主観評価実験（歌い手を当てる3択クイズ）
  - 元話者
    - f001 (Sinsy) : 70曲
  - 適応話者
    - 初音ミク (CV01) : 5曲
    - 凜 (RWC-MDB-P-2001) : 7曲

# 実験結果

HMM-based Singing Voice Synthesis System



- 話者正解精度



- 高い正解精度

- ターゲットの歌い手の特徴をもった歌声
- 特定話者モデルでは話者性を高精度に再現可能



# デモンストレーション

HMM-based Singing Voice Synthesis System



## メリーさんの羊 📢

1. f001 (Sinsy)
2. 初音ミク (CV01)
3. 凛 (RWC-MDB-P-2001)

$\text{♩} = 100$  *mf*

め り さ ん の ひ つ じ ひ つ じ ひ つ じ

5

め り さ ん の ひ つ じ ま ー つ し ろ ね

# 今後の予定

HMM-based **Singing Voice Synthesis System**



- 新たな歌声モデルのリリース
- 多言語化（日英中）
- ユーザー独自の歌声・楽譜アップロードによる歌声モデルの自動作成
- CEATEC

お手持ちの歌声データ・MusicXMLで  
歌声モデルを作ってみたい方は  
#Sinsyでつぶやいてみてください