Sinsy

HMM-based Singing Voice Synthesis System

自動学習により 人間のように歌う音声合成システム ― Sinsy ―

名古屋工業大学 国際音声技術研究所
徳田恵一・大浦圭一郎

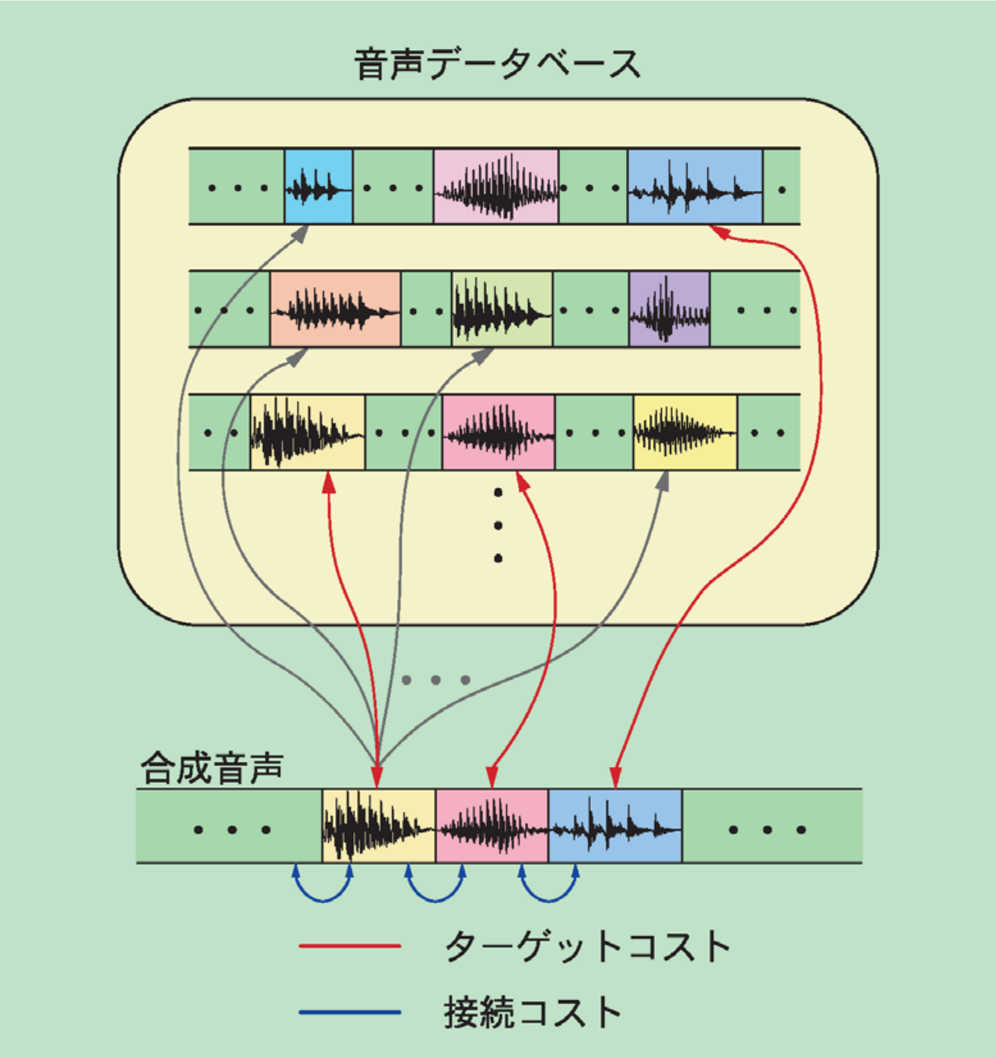
はじめに

- **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）

背景: 最近の音声合成方式

- **素片接続型音声合成**
  - 波形接続アルゴリズム(PSOLA 他) により高品質な合成音
  - 多様な声質，発話スタイルを得るためには，大量の音声データベースが必要
- **HMM音声合成（HMM: 隠れマルコフモデル）**
  - 統計量に基づいて音声パラメータを生成
  - 統計モデルパラメータを変換することにより，多様な声質・発話スタイルを容易に得ることが可能

背景：素片接続型音声合成

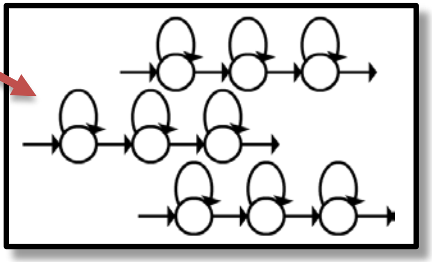


素片は音素・ダイフォン・
トライフォン等に対応

背景: HMM音声合成

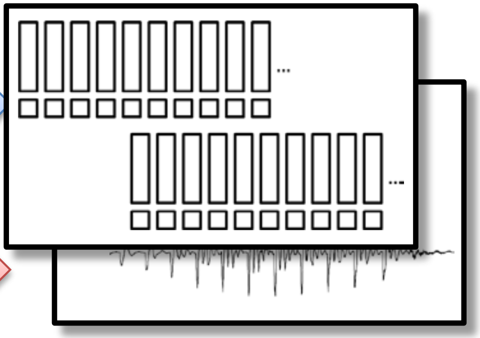
ひとつのモデルは
音素等に対応

「ひとつ前がaで次がiのｋで
文の先頭から10番目にあって
当該単語は名詞で
アクセントは・・・」



HMM

スペクトルと基本周波数
パラメータ



あらゆる現実を・・・
救急車が動けず・・・
・・・
・・・

テキスト

① 対応関係をHMMにより学習

② 与えられたテキストから
波形を推定

対応した音声波形

背景: HMM音声合成の利点

- **HMM音声合成の特徴**
  - 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
    ⇒ 話者の声質および発話スタイルを再現
- **応用**
  - 感情音声合成（喜んだり悲しんだり）
  - 話者補間（声を混ぜる）
  - 話者適応（声を真似る）
  - 固有声（声をつくる）
  - 多言語音声合成（地球上に数百の言語あり）

デモ: 感情音声合成

テキスト	平静	怒り
「授業中に携帯いじってんじゃねえよ！電源切っとけ！」		
「ミーティングには毎週参加しなさい！」		

200文により学習

デモ: 発声スタイルの補間

HMM-based Singing Voice Synthesis System

「平静」モデル

「ハイテンション」モデル

補間

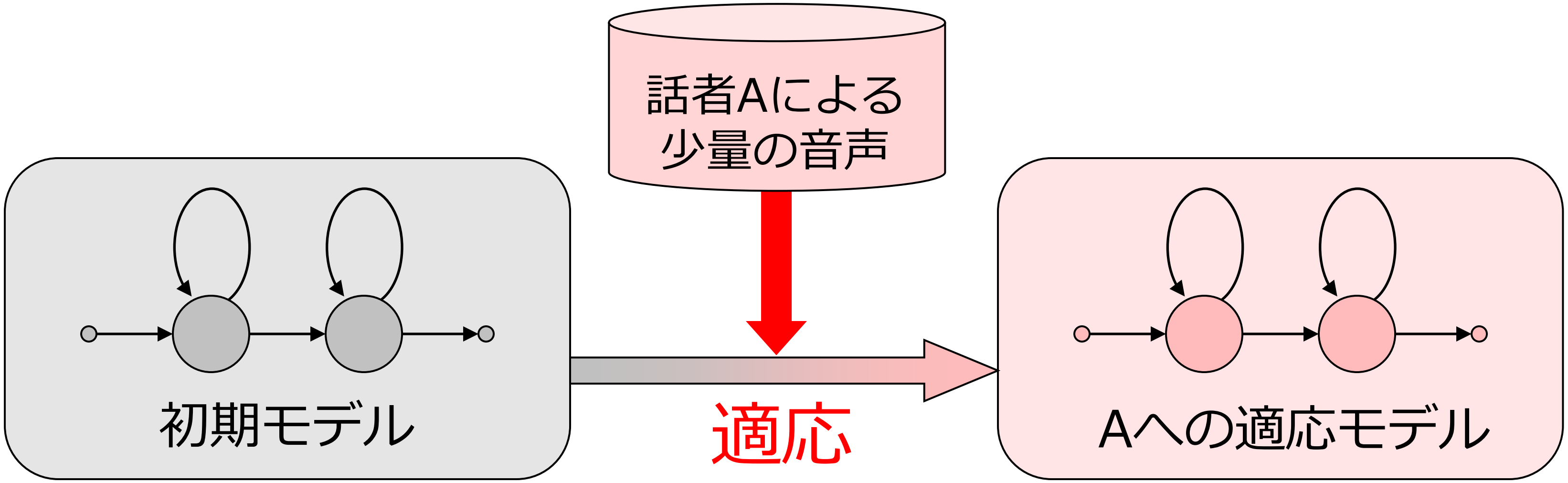
補外

平静

ハイテンション

デモ：話者適応（声を真似る）

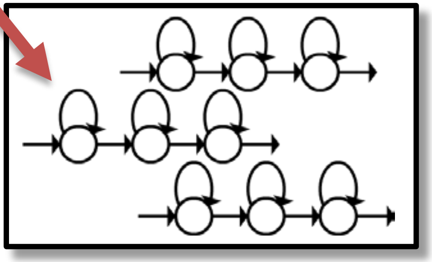
HMM-based Singing Voice Synthesis System



HMM歌声合成の概要

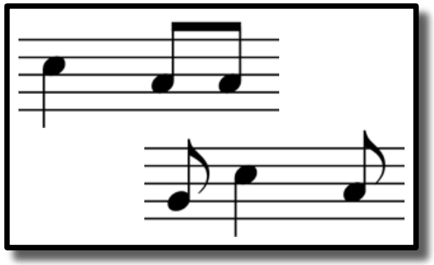
ひとつのモデルは
音素等に対応

「音高がC3で音名がラで
歌詞に対応する音素がaで
音長が0.1秒で
小節内の2番目にあり
ひとつ前の音符は音高がD3で
・・・」



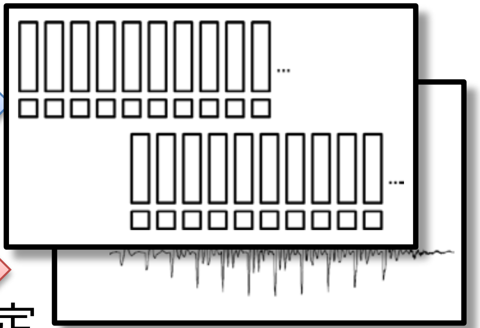
HMM

スペクトルと基本周波数
パラメータ



① 対応関係をHMMにより学習

② 与えられた楽譜から波形を推定



歌詞付き楽譜(MusicXML)

対応した音声波形

Sinsy（www.sinsy.jp）とは？

HMM-based Singing Voice Synthesis System

- HMM歌声合成システムのオンラインデモページ (HMM-based Singing Voice Synthesis System)
- 楽譜 (MusicXML)をアップロードすることにより，歌声を合成可能
  - 現在までに1万件以上のアップロード
- 特徴
  - 調整作業が不要（ベタ打ちのみでOK）
  - OSに依存せず利用可能
- オープンソースのソフトウェアを利用して構築
- 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

Sinsy - HMM-based Singi
www.sinsy.jp
Sinsy
HMM-based Singing Voice Synthesis System
アップロードされた楽譜(MusicXML)に基づいて自由な歌声を生成するHMM歌声合成システム，Sinsy(しぃんしぃ)です.
ボーカル
f001j : 謡子 (Yoko)
声質
0.55
(-0.8~0.8, 標準: 0.55)
ビブラート強度
1
(0.0~2.0, 標準: 1.0)
ピッチシフト
0
(-24~24, 標準: 0)
楽譜(.xml)
ファイルを選択
選択されていません
送信
合成結果
ボーカル f001j
声質 0.55
ビブラート強度 1
ピッチシフト 0
wav, mgc, lf0, log
About Sinsy
サンプル
げんこつ山のタヌキさん (f001j 声質: 0.55 ピッチシフト: 0) xml wav
春が来た (f001j 声質: 0.45 ピッチシフト: 4) xml wav
利用規約
デモ動画
ユーザー作品
ニコニコ動画
YouTube
PIAPRO
MUSIC TRACK
使い方動画
キャラクター案
関連情報
お知らせ
2011/12/25 [Ver. 3.0]
音質を向上させました.
ボーカルf001をf001jに，f002j_aをf002jに変更しました.
ユーザー作品にPIAPRO, MUSIC TRACKを追加しました.
2011/06/22
関連情報を追加しました.

ユーザー作品（抜粋）

HMM-based Singing Voice Synthesis System

作者の皆様，ありがとうございます！

ユーザーによる使い方動画他

HMM-based Singing Voice Synthesis System

- **CadenciiをインストールしてSinsyに歌わせるまでの録画**

- **MuseScoreで楽譜作成し、Sinsyに歌わせる手順(2010/09/02版)**

- **MusicXML関連ツールの開発・公開**
  - スコアエディタ型
  - ファイルフォーマットコンバータ型

耳ロボＰ様，作者の皆様，ありがとうございます！

ユーザーによるキャラクタ案

作者の皆様，ありがとうございます！

ユーザーからのフィードバック

HMM-based Singing Voice Synthesis System

- 多くのバグ修正
- 合成可能音高の制限の取りやめ
- ビブラート強度の調整
- ログ（＋継続長）の出力
- MusicXML解析エラーの出力
- 基本周波数，スペクトルパラメータの出力
- 曲中のテンポ変更に対応
- 利用規約の設定および改定
- カタカナの歌詞への対応

#Sinsy でバグレポート・
改善提案お願いします！

調整（所謂「調教」）できない？

HMM-based Singing Voice Synthesis System

ログ（＋継続長）

基本周波数

スペクトルパラメータ

波形

SPTKによる波形生成

基本周波数

スペクトルパラメータ

```
sopr < xxxxxx.lf0 -magic -1e10 -EXP -INV -m 48000 -MAGIC 0 |¥
excite -p 240 |¥
mlsadf xxxxxx.mgc -m 49 -p 240 -a 0.55 >! xxxxxx.raw
raw2wav xxxxxx.raw +f +s -s 48000
```

周期への変換

励振源の生成

合成フィルタ

ヘッダの付加

波形

SPTK

http://sp-tk.sourceforge.net/

Sinsyは音痴？

HMM-based Singing Voice Synthesis System

- **歌声データ提供者の声質・歌い方をコピー**
  - 歌の上手い人 → 歌の上手いモデル
  - 歌の下手な人 → 歌の下手なモデル

- **むしろ音痴すら再現可能**

補正することは容易だが
敢えてやっていない

デモ: 歌声モデル補間

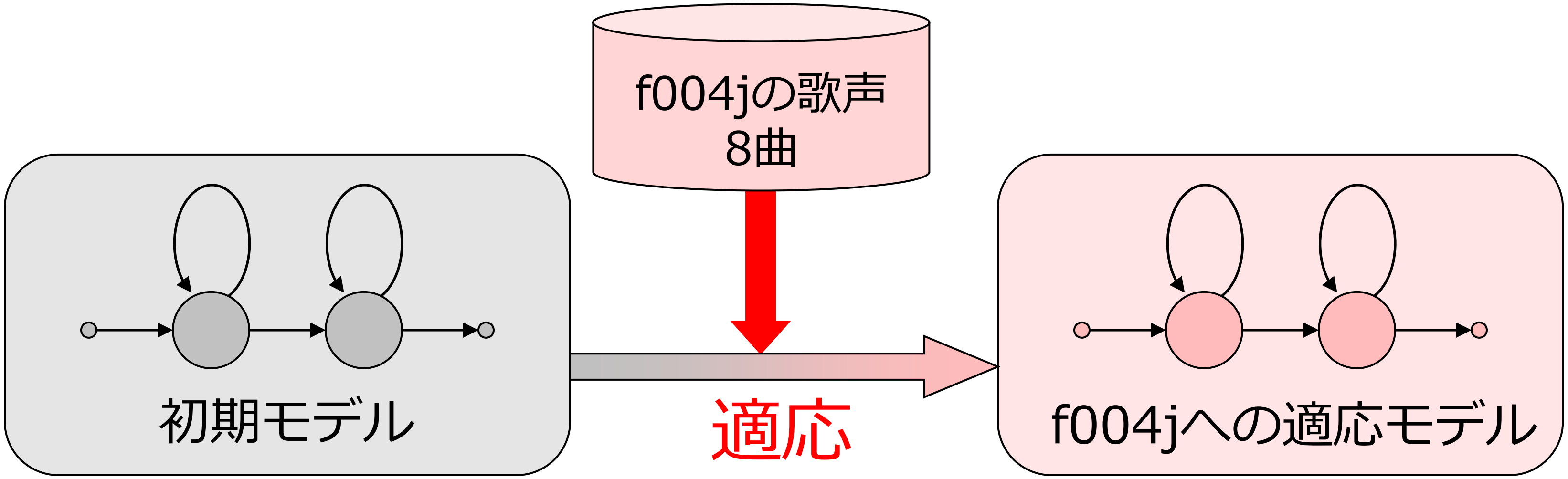
HMM-based Singing Voice Synthesis System

f001j
(童謡，歌劇)

f002j
(J-POP, バラード)

補間

デモ：歌声適応（歌声を真似る）



f004j_a
（波音リツＳ）

その他のデモは以下でもどうぞ

HMM-based Singing Voice Synthesis System

本日（2月3日）

(7) ラップスタイル歌声合成の検討

才野 慶二郎（ヤマハ），大浦 圭一郎（名工大），橘 誠, 剣持 秀紀（ヤマハ），徳田 恵一（名工大）

明日（2月4日）

(15) HMM歌声合成における音高正規化学習の検討

大浦 圭一郎，間瀬 絢美，南角 吉彦，徳田 恵一（名工大）

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）**
- **言語間話者適応**
- **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**

※ 2025年4月現在、下記の募集は終了していますので、
あしからずご了承ください。

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でつぶやくか
徳田 (http://sp.nitech.ac.jp/~tokuda/) まで
ご連絡ください