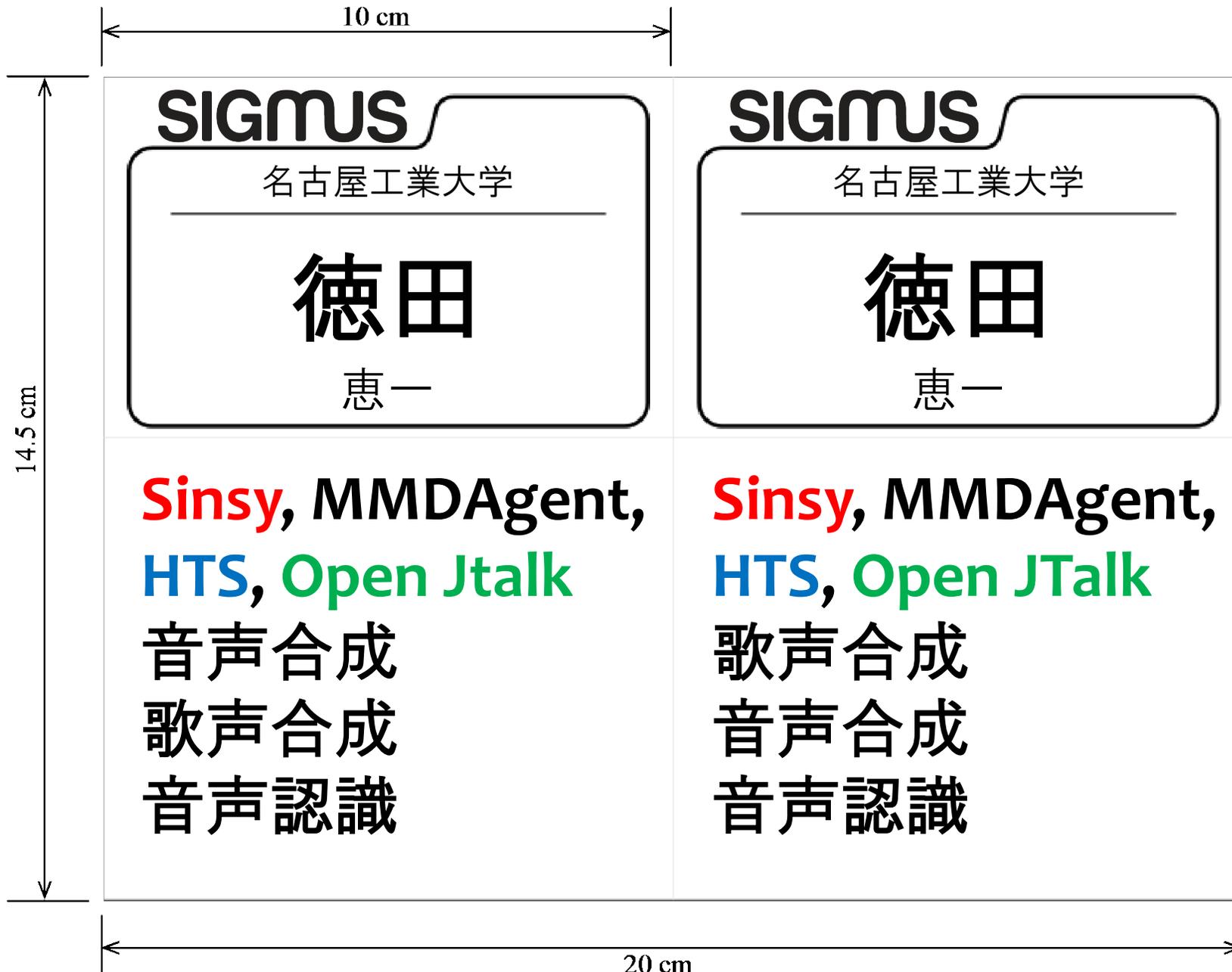


「これからの歌声合成」

名古屋工業大学
徳田恵一





音声合成の歴史（超簡略版）

- **ルールベース**: フォルマント音声合成 (～'90)
- **コーパスベース**: **波形**接続型音声合成 ('90s～)
 - 単一インベントリ: ダイフォン音声合成
 - 複数インベントリ: 単位選択型音声合成
- **コーパスベース**: **統計**的パラメトリック音声合成 ('95～)
 - 隠れマルコフモデルによる音声合成 (HMM音声合成)
 - DNN, LSTM等に基づく音声合成

音声技術に関する名言たち (1980s)

“There is no data like more data”
(Robert Mercer)

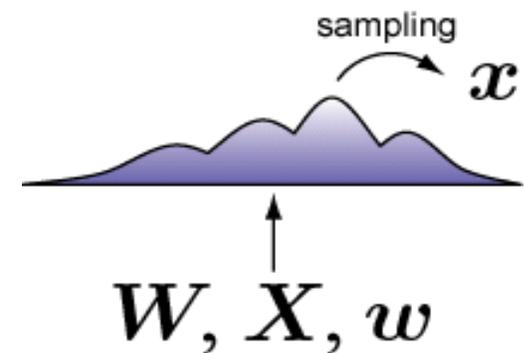
“Every time I fire a linguist,
the performance of the speech recognizer goes up”
(Frederick Jelinek)

「ビッグデータ」 ・ 「機械学習」 など、昨今のトレンドのはしり

私の取り組む歌声合成の基本問題 (1/2)

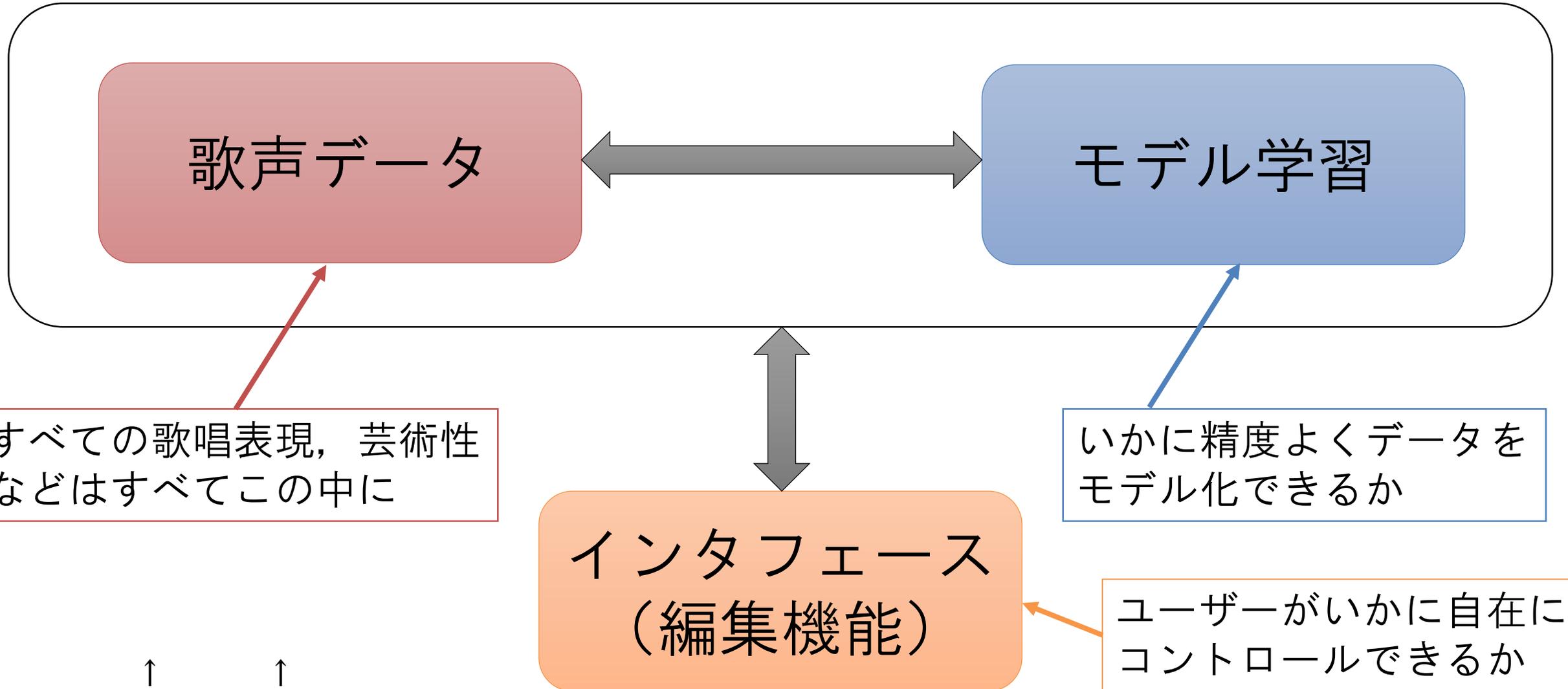
楽譜とそれに対応する歌声波形の集合があるとき、
任意に与えられた楽譜に対応する音声波形を求めよ。

- W : 楽譜
 - X : 歌声波形
- } 歌声データベース
- } 既知
- w : 任意の楽譜 ($w \notin W$)
 - x : 合成歌声波形 ← ?



$$x \sim p(x|w, X, W)$$

私の取り組む歌声合成の基本問題 (2/2)



Open JTalk

HMM-based Text-to-Speech System Open JTalk Demonstration Page

入力された日本語テキストに基づいて自由な音声を生成するHMMテキスト音声合成システム, Open JTalkのデモンストレーションです.

話者	<input type="text" value="男性:M001"/>
声質	<input type="text" value="0.55"/> (-0.8~0.8, 標準: 0.55)
ピッチシフト	<input type="text" value="0"/> (-24~24, 標準: 0)
話速	<input type="text" value="1"/> (0.5~2.0, 標準: 1.0)
合成テキスト (最大200字)	<input type="text"/>
	<input type="button" value="合成"/>

合成結果



利用規約

お知らせ

- 2012/12/25 [Ver. 1.8]
Open JTalkのバージョンを1.06に更新しました.
女性話者「Mei (Happy)」
「Mei (Bashful)」
「Mei (Angry)」
「Mei (Sad)」を追加しました.
音質を安定させました.
- 2011/12/25 [Ver. 1.7]
Open JTalkのバージョンを1.05に更新しました.
- 2011/07/07 [Ver. 1.6]
Open JTalkのバージョンを1.04に更新しました.
- 2011/05/01 [Ver. 1.5]
女性話者を追加しました.

Sinsy

HMM-based **Singing Voice Synthesis System**

アップロードされた楽譜(MusicXML)に基づいて任意の歌声を生成するHMM歌声合成システム, Sinsy(しいんしい)です.

言語	<input type="text" value="日本語"/>
ボーカル	<input type="text" value="f001j : 謡子 Yoko (日本語)"/>
声質	<input type="text" value="0.55"/> (-0.8~0.8, 標準: 0.55)
ビブラート強度	<input type="text" value="1"/> (0.0~2.0, 標準: 1.0)
ピッチシフト	<input type="text" value="0"/> (-24~24, 標準: 0)
楽譜 (.xml)	<input type="button" value="ファイルを選択"/> 選択されていません <input type="button" value="送信"/>



[利用規約](#)

デモ動画

- [ニコニコ動画](#)
- [YouTube](#)

ユーザー作品

- [ニコニコ動画](#)
- [YouTube](#)
- [SoundCloud](#)
- [PIAPRO](#)
- [MUSIC TRACK](#)

使い方動画

音声創作ソフトウェア 「CeVIO Creative Studio」

- 感情音声合成
+
- 歌声合成

- コンテンツ制作を意識した編集機能



ユーザー生成コンテンツ



マニュアルチューニングなし



CeVIO Creative Studioによるマニュアルチューニング
(Auto-Tune, Melodyne等の外部ツール不使用)

カラオケでの利用

JOYSOUND 「ボーカルアシスト」

- 知らない曲・うろ覚えの曲を歌いたいとき
- 合成音声で代わりに歌ってくれる！
- 日本語楽曲ほとんどすべてに対応



★ ★
ボーカルシンセが歌唱サポート！お手本ボーカルと一緒に歌おう！
★ ★

ボーカルアシスト

新登場!

対応機種
JOYSOUND F1 JOYSOUND FR

こんなときに超便利!

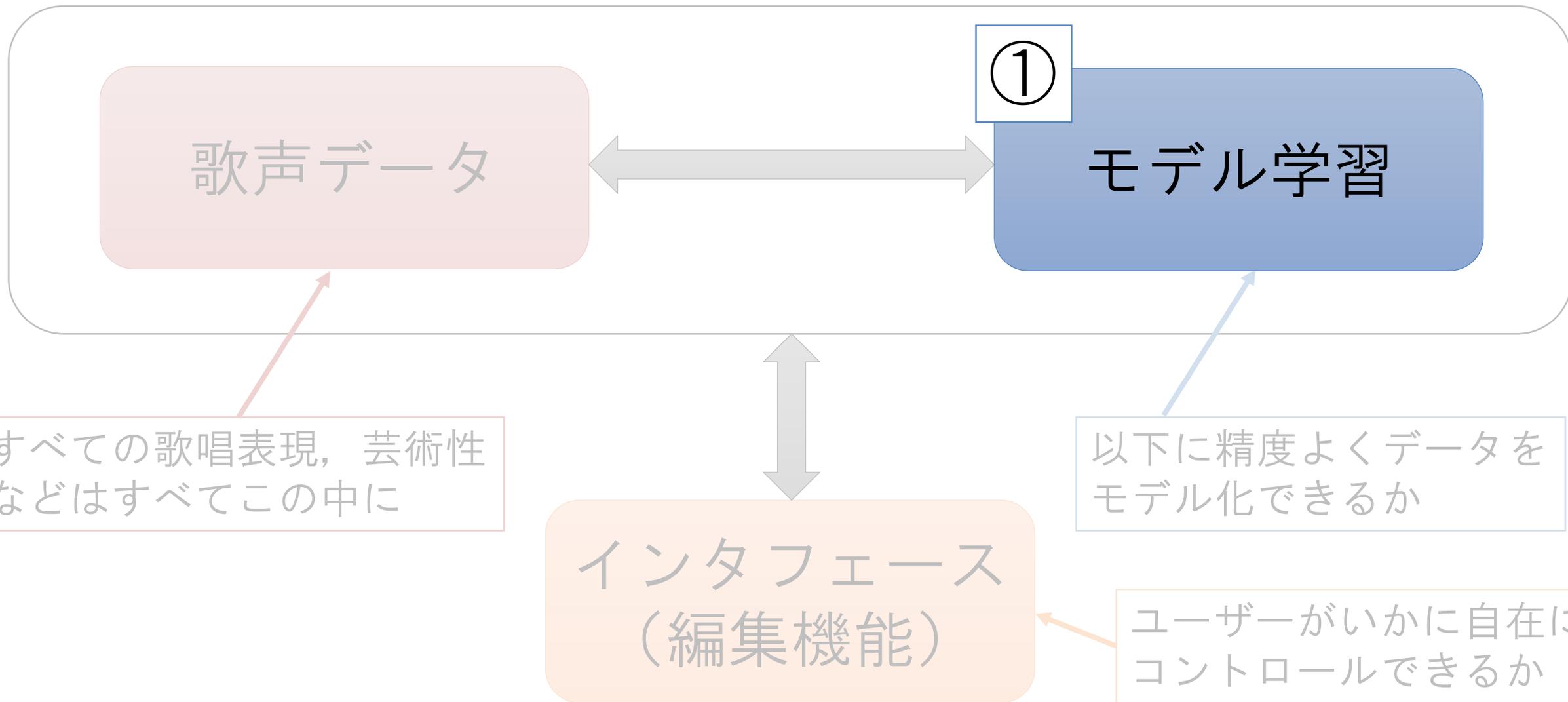
- 高得点を目指して練習する時**
▶詳しくはこちら
- ハモリ曲を歌う時やちょっと自信のない曲を歌う時に**
▶詳しくはこちら
- ギター演奏のお供に**
▶詳しくはこちら
- どんな曲だったか忘れちゃった時に**
▶詳しくはこちら
- ご高齢者、小さいお子様と一緒に歌う時**
▶詳しくはこちら

VOCALOID4 Library Sachiko

- Job Plugin Sachikobushi により、
小林幸子本人による歌唱表現
を再現
- Job Plugin にHMM音声合成方
式を採用



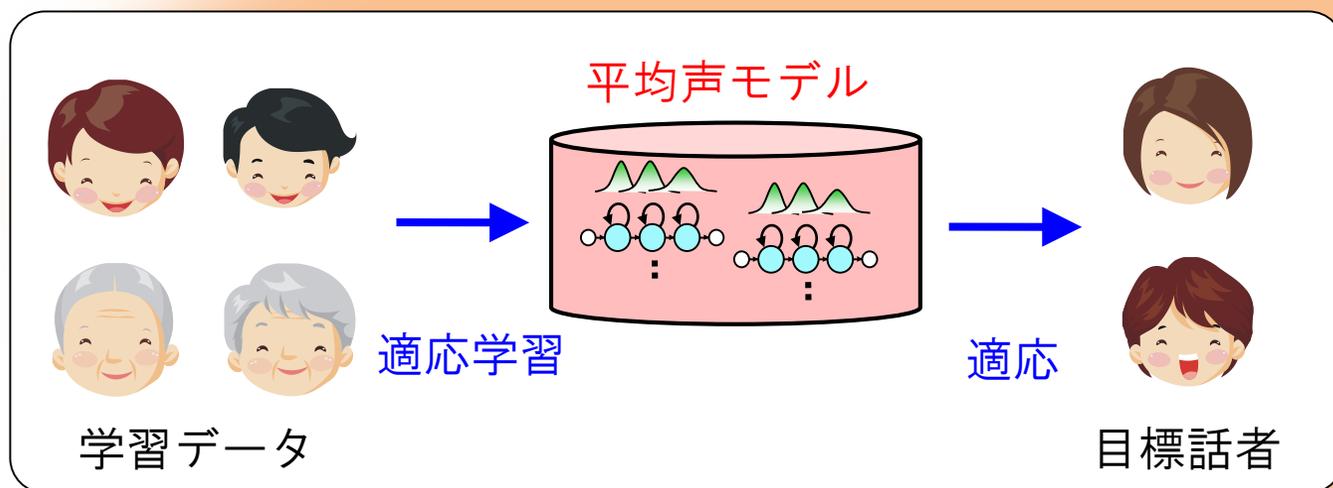
これから取り組む問題



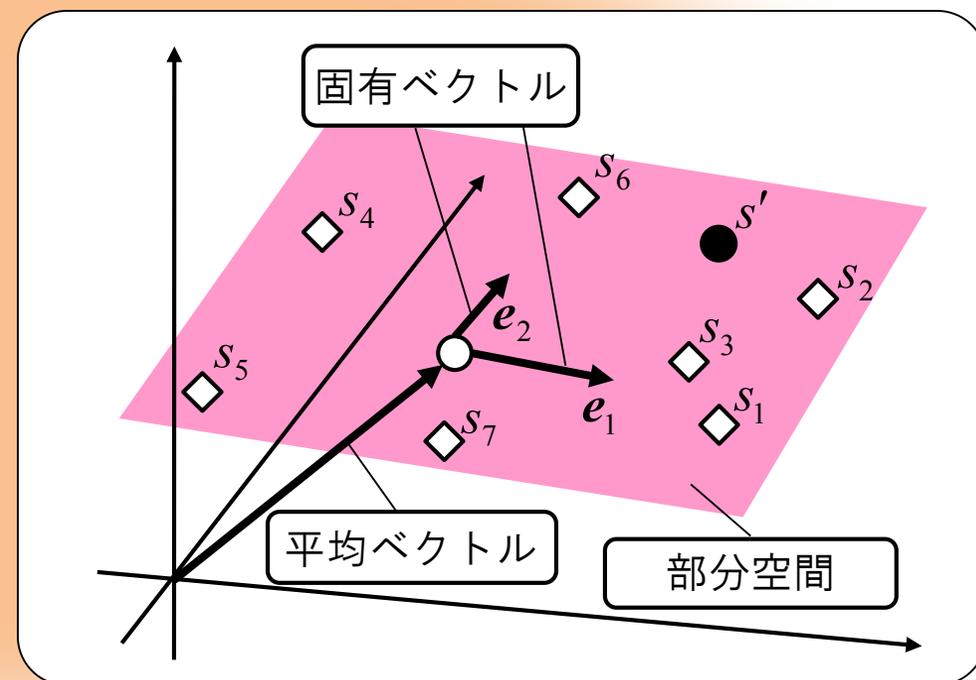
ユニバーサル歌声モデル

あらゆる声質，発話スタイル，歌唱スタイル，感情表現，言語等を
自在にモデル化・制御可能な音声モデル

「平均声」



「固有声」



これから取り組む問題

②

歌声データ

モデル学習

すべての歌唱表現，芸術性
などはすべてこの中に

以下に精度よくデータを
モデル化できるか

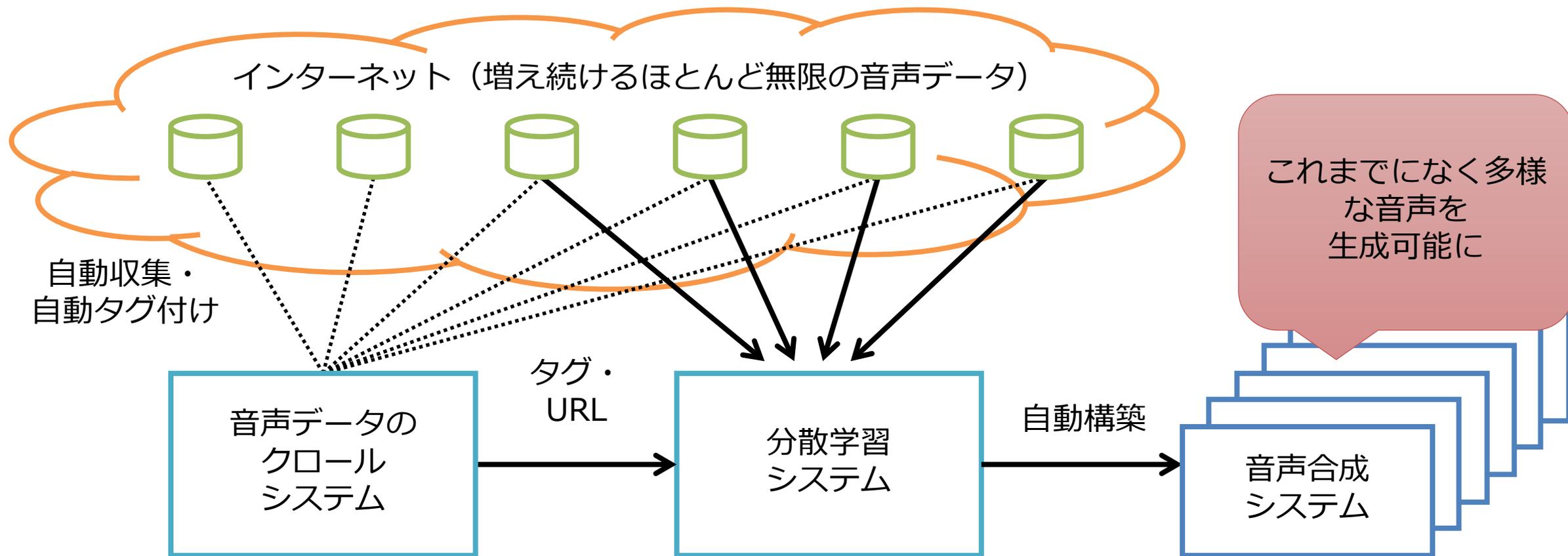
インタフェース
(編集機能)

ユーザーがいかに自由に
コントロールできるか

音声データを集積・共有する仕組み



学習用データの分散共有と分散型モデル学習



これから取り組む問題

歌声データ

モデル学習

すべての歌唱表現，芸術性
などはすべてこの中に

③

インターフェース
(編集機能)

以下に精度よくデータを
モデル化できるか

ユーザーがいかに自由に
コントロールできるか

モデル構造

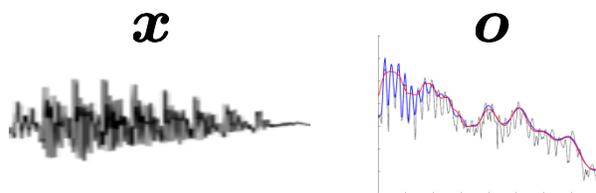


音響モデル
(HMMなど)

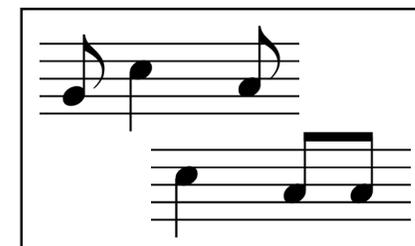


$$p(x|w, \lambda) = p(x|o) p(o|l, \lambda_A) P(l|w, \lambda_L)$$

音声分析・波形生成
(Vocoding)



歌詞・楽譜
分析



様々な階層での編集機能



- 言語
 - 日本語, 英語, 中国語, 日本語英語, ...
- 発音変形
 - 「スター」 (/u/の脱落), 「want you」 (ワンチュー), ...
 - ブレス, しゃくり
- 歌い方
 - ロック調, オペラ調, ...
- 声質
 - 男性, 女性, 大人, 子供
- 音声パラメータ
 - 基本周波数パターン, 音量変化パターン, ...

高次

歌詞・楽譜
分析



音響モデル

低次

波形生成

まとめ

歌声空間を埋め尽くす超巨大分散データにより，あらゆる歌声を自在に生成することの可能な歌声合成システムを構築する

- 超巨大データを用いたユニバーサル音声モデルの構築
 - 「平均声」および「固有声」の手法をベースに融合
 - 分散・並列処理，荒れたデータへの対応
- データの分散共有
 - インセンティブ・ライセンス
- 編集機能の多元化・多層化
 - 高次から低次まで，多様な操作機能