

音声合成モジュール

山下 洋一

yama@cs.ritsumei.ac.jp

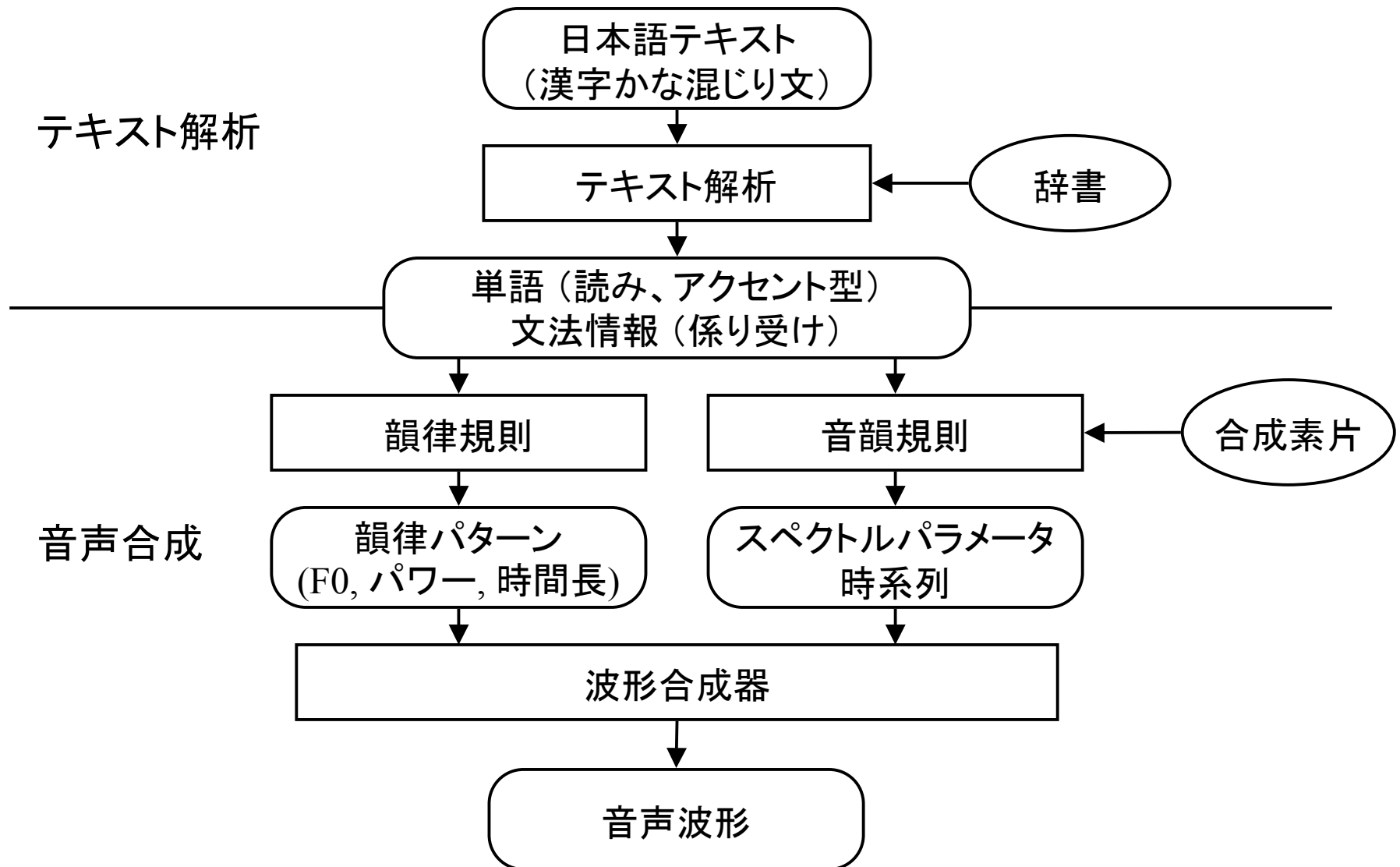
立命館大学 理工学部

(4月から 情報理工学部)

2004年2月27日

- Galatea における音声合成モジュール
 - 単体の日本語テキスト音声合成システムとしても動作
 - 無償で利用可能、ソースコードの公開
 - 顔画像との同期
 - テキスト解析部、韻律生成部、波形生成器のモジュール化
 - 対話システムにおける音声合成

音声合成における処理の流れ



GalateaTalk の構成

■ テキスト解析

- chasen の利用
- 辞書の整備: アクセント型、アクセント結合様式
- 辞書のエントリ数: 約23,000

■ 韻律生成

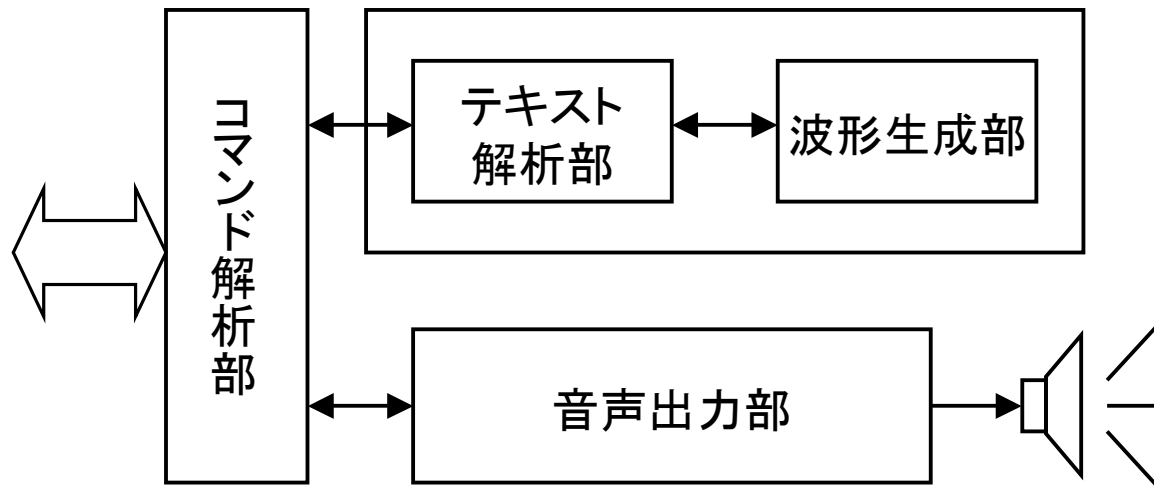
- アクセント結合: 先行研究を実装
- 韻律パラメータ: コーパスベース

■ 波形生成

- HMMに基づく音声合成

■ 構成

- chasen: 形態素解析
- chaone: 複合語処理、音韻交替処理
- gtalk: コマンド解析、韻律・波形生成、音声出力



■ 基本的な日本語テキスト音声合成

■ プレーンテキストからの音声合成

■ タグ付きテキストからの音声合成

- JEIDA-62-2000 日本語テキスト音声合成用記号の規格

- 記述例:

<SPEECH>

<VOICE REQUIRED="male01">

これは<PRON SYM="アイピーエー">IPA</PRON>の
プロジェクトで開発された<EMPH>対話</EMPH>
音声合成システムです。

</VOICE>

<SPEECH>

- VOICE タグ
 - 話者の変更
- RATE タグ
 - 話速の変更
- VOLUME タグ
 - 音量の変更
- PITCH タグ
 - 基本周波数の変更
- SILENCE タグ
 - ポーズ挿入
- EMPH タグ
 - 強調
- PRON タグ
 - 読みの指定
- SPELL タグ
 - 英字の綴り読みの指定
- CONTEXT
 - 数字の読み方の指定
 - NUMBER, DIGITS, DATE, TIME, PHONE

- 音声出力の途中中断
 - 中断までに音声出力した音素列の出力
- 指定した時刻での発話
- 合成音声における音素列の時間長の出力
 - 音素の継続時間を共有することにより顔画像出力との同期
- 話者の変更
 - 一発話内においても話者の変更が可能
 - 話者数: 男声2名、女性1名

■ 韻律の制御

- 対話における多様な韻律の生成
- 「JEIDA-62-2000日本語テキスト音声合成用記号の規格」に準拠
 - 基本周波数、パワー、発話速度の部分的変更
 - 「強調」の指定

3年間の予定

1. 辞書の拡充

- 50,000語程度に
 - 現在、約23,000語

2. 音声出力処理の改良

- 合成音声出力開始までの時間短縮
 - 波形生成と音声出力の並行処理
 - ファイルにセーブされた音声データの利用

3. テキスト解析精度の向上

- 読み決定の精度向上
 - 読みわけ、単漢字の読み付与
- アクセント句決定の精度向上

4. 合成音声のカスタマイズ機能の実現

- 声質のカスタマイズ
- 韻律のカスタマイズ

1年目の予定

1. 辞書の拡充

- サ変名詞の追加 (約6000語)

2. アクセント句生成、アクセント型決定処理の改良

- chaone の改良
- gtalk 内部処理の変更

3. ファイルにセーブされた音声の再生

- gtalk で合成しセーブしたファイルの利用

4. 韻律パラメータのファイルからの読み込み

- 音素時間長と基本周波数(F0)

5. XML 形式によるアクセント句情報の出力