

音声合成モジュール

山下 洋一

yama@media.ritsumeai.ac.jp

立命館大学 情報理工学部

2009年3月12日

日本語音声合成システムの開発

- IPA 擬人化音声対話基本ソフト開発プロジェクト: Galatea Project (2000～2003年度)
 - 最初のフリーの日本語音声合成(TTS)システム
 - HMM音声合成
 - 多様な音声合成の可能性
 - 辞書の開発
 - 発音, アクセント関連情報
- ISTC では
 - 対話
 - カスタマイズ

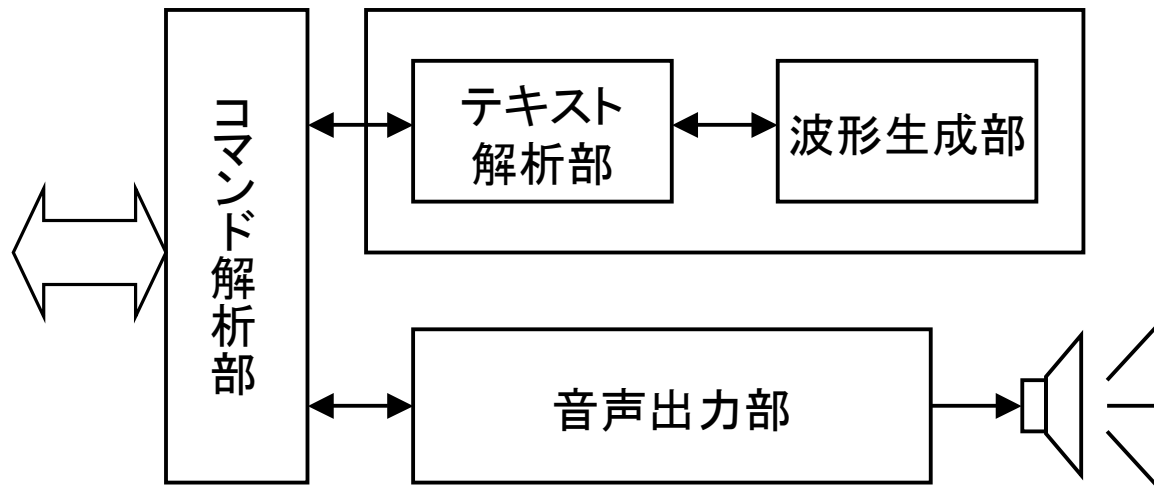
初年度総会 (2004年2月27日) から

- 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
 - IPAのプロジェクトから継続

GalateaTalk の構成

■ 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
 - HMMに基づいた音声合成



- 基本的な日本語テキスト音声合成
 - JEIDA-62-2000 日本語テキスト音声合成用記号の規格
- 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
- 合成音声における音素列の時間長の出力
 - 顔画像出力との同期
- 話者の変更
- 韻律の制御
 - 基本周波数, パワー, 発話速度の部分的変更

初年度総会 (2004年2月27日) から

1. 辞書の拡充

- 50,000語程度に
 - 当初、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

活動成果 (1)

■ 辞書の整備

■ ISTC当初

- 約23,000語, サ変名詞が入っていない

■ UniDic プロジェクトからの辞書の提供

- UniDic1.2.0pre (2006年4月)

- 約46,000語

- UniDic1.3.0 (2007年4月)

- 約106,000語

- UniDic-chasen1.3.5 (2007年10月)

- 約110,600語

- 連体形と終止形を区別

- 名称を UniDic-chasen に変更 (UniDic-mecab の開発を開始)

活動成果 (2)

■ ChaOne の開発

- チャンキング
- 音韻交替処理
- 読み選択
- アクセント結合処理
 - gatlk から chaone での処理へ。
 - 規則を宣言的に記述可能に。

アクセント型決定規則の記述例

```
<!-- 接頭辞(pre) + 自立語(current) -->
<rule id="P1">
  <!-- 一体化型 -->
  <if>
    <pre aConType="P1" />
  </if>
  <then>
    <if>
      <current no_nucleus="yes" />
    </if>
    <then aType="0" />
    <else aType="pre/@mora + current/@aType" />
  </then>
</rule>
...
```

■ GalateaTalk の改良

■ 音声出力

- ファイルにセーブされた音声の再生
- 波形生成と合成音声出力の並行処理

■ 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)
- 韻律修正ツール (prosedit) との連携

■ ユーザによる辞書登録

■ アクセント句を強制分割するタグの導入

- 例) 山下洋一 → 山下<ARB/>洋一

■ カスタマイズ機能

■ 声質のカスタマイズ

- VoiceMaker: 話者モデル作成ツール
- 「音声認識・音声対話技術講習会」で利用
- サンプル

学習データ:   合成音(50文で学習):  

■ 韻律のカスタマイズ

- Prosedit: 韻律修正ツール
- GUI を利用した手修正
- GalateaTalk の呼び出し
- 参照音声からの韻律のコピー

■ 韻律修正ツール Prosedit

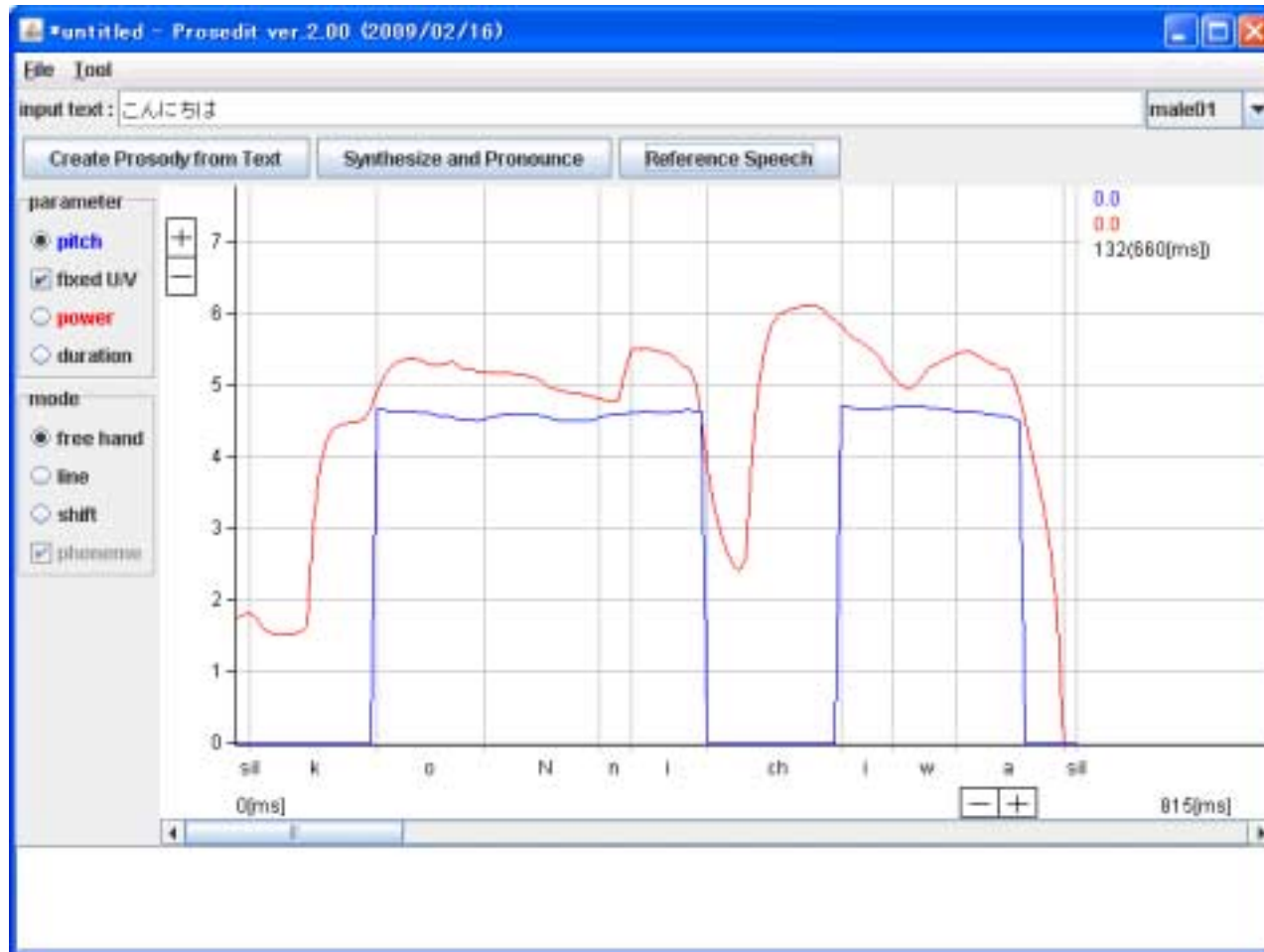
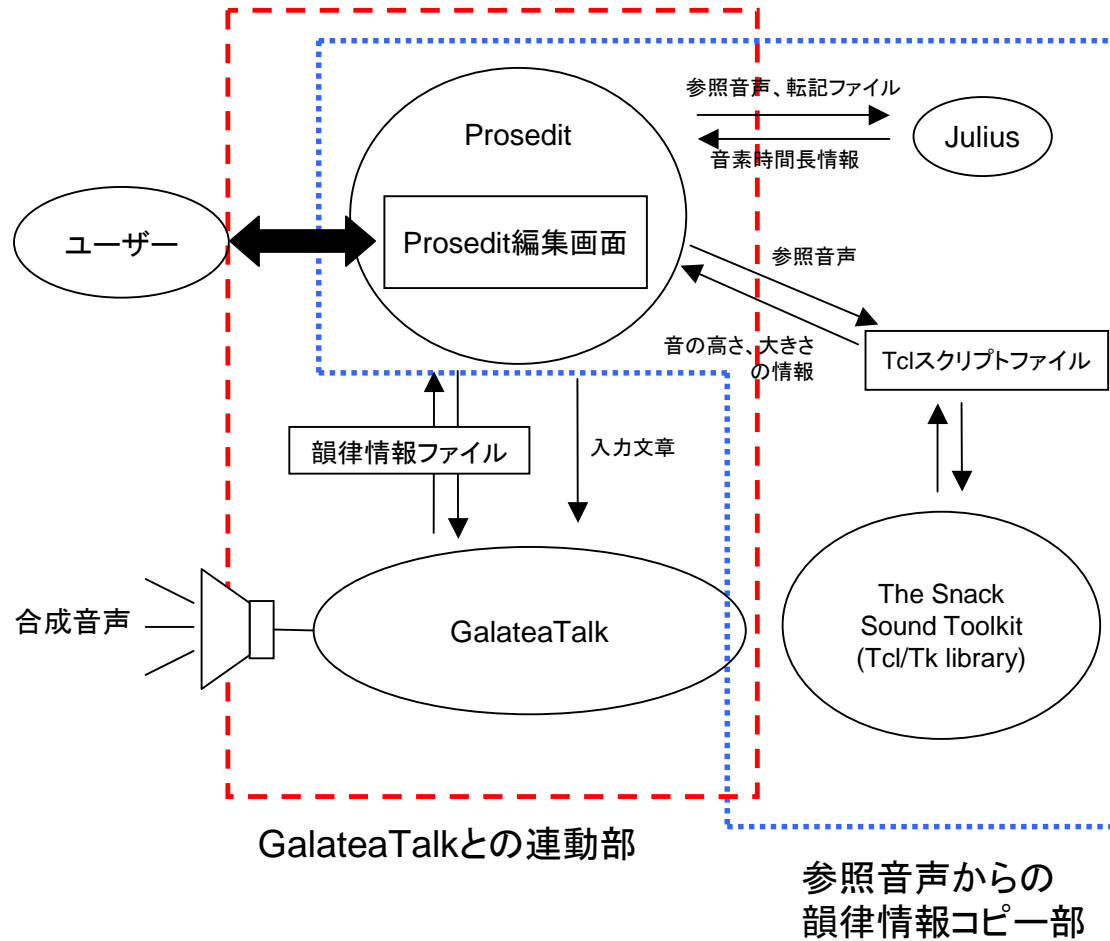


図 Prosedit の動作画面

Prosedit システム構成



```

input_text: こんにちは
spoken_text: こんにちは
number_of_phonemes: 11
total_duration: 680
-----
sil [10]
k [100]
o [85]
(中略)
w [50]
a [95]
sil [10]
-----
total_frame: 136
-----
0: 0.000000 1.723706
1: 0.000000 1.779953
2: 0.000000 1.819527
3: 0.000000 1.737764
...
    
```

韻律情報ファイルの例

今後の課題

1. 読み・アクセント型の精度向上
 - 実用性において重要
2. 多様な声質での音声合成
 - 声質変換
 - 韻律ラベリング
3. 韻律修正ツール(prosedit)の改良・機能追加
 - GalateaTalk のフロントエンドに
4. 辞書の整備
 - UniDicプロジェクトとの連携