

ISTCにおける 音声認識ソフトウェアの開発状況

河原達也（京都大）
李 晃伸（名工大）

これまでの経過

- 1995～1997 IPSJ/SLP傘下 WG
 - JNASコーパスの設計
- 1997～2000 IPAプロジェクト
 - 「日本語ディクテーション基本ソフトウェア」の開発
- 2000～2003 連続音声認識コンソーシアム CSRC
 - 認識ソフトウェアの改善
 - 音響・言語モデル等の充実
- 2003～2008 e-Society基盤ソフトウェア
- 2003～2006 音声対話技術コンソーシアム ISTC
 - 対話システムを指向した音声認識の改善

ISTCでの主要開発目標

- 音声認識エンジンJuliusの性能改善・機能追加
- 音声認識エンジンJuliusのSAPI/SALT対応
→ 2003年度済
- 音声認識ソフトウェアのカスタマイズツール
- 音声合成エンジンgtalkのSAPI対応

用語の説明

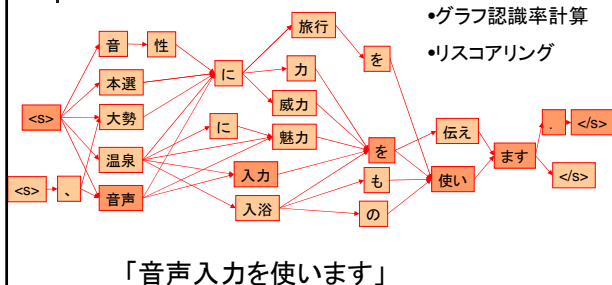
- Julius: 大語彙連続音声認識のフリーソフトウェア
 - 当初はディクテーション向けのN-gram言語モデル対応のものだったが、現在は下記Julianも統合
- Julian:
 - (人手による)記述文法対応の認識プログラム
- SAPI: Speech API
 - マイクロソフト社策定のAPI、Windows XPに標準搭載
- SALT: Speech Application Language Tags
 - マイクロソフト社などが策定している、HTMLブラウザで音声認識・合成を行うためのタグ
 - Speech Application SDKに含まれる
- SRM: Speech Recognition Module
 - JuliusのGalatea Toolkitのためのインタフェース

Julius 3.5

新機能

- GMM による環境音識別
 - 音声対話システムにおける誤動作防止
- 単語グラフ出力
 - 外部知識によるリスコアリング
- 単語信頼度による枝刈り
- Julianで複数文法認識時に、文法ごとの認識結果を出せるようにした
- Julianで起動時に複数の文法を一度に指定できるようにした
- (Linux) Esound オーディオサーバー入力対応

単語グラフの例



メモリ使用量の削減

- ビームのワークエリアおよび木構造化辞書の不要部分を削除
- 3-gramをもたない 2-gram の不要領域を削除
- 2-gramのN-gramインデックスを32bit→24bitに縮小



20kディクテーションでプロセスサイズが
74MB→63MB

ドキュメンテーション & バグフィックス

- Doxygen対応
 - ソース内のほぼすべての宣言・関数にコメントを追加
 - 構造化されたソース解説文書を HTMLやLaTeXで出力可能。
(ライブラリ: 英語 Julius: 英語と日本語併記)
- 実装改善
 - Windows での開発: minGW へ対応
 - Linux版とWindowsコンソール版のソースを統合
- バグ修正
 - スペクトルサブトラクションが動かないバグを修正
 - (Win/Mac) 改行コード対応(文法ファイル・filelist)
 - (Win) cygwin環境で mkdafa.pl が動かないバグを修正
 - (Linux) USBオーディオ対応、複数音声デバイス対応
 - その他細かいバグフィックス

今後の課題(?)

- 標準的な文法フォーマットSRGF(W3C)のサポート
- 発話検出・音声切り出し部分の改善
- 耐雑音の前処理ETSI Advanced FEの統合

音声認識ソフトのカスタマイズ ツール

Julius音声認識実行キット

- ディクテーション実行キット(2004年8月)
 - 実行バイナリと標準モデル
- Julian音声認識キット(2004年11月)
 - 実行バイナリとサンプル文法
- 人工知能学会誌1月号に解説記事
「連続音声認識ソフトウェアJulius」

音声認識ソフトウェアの カスタマイズツール

- Webベース
- 使用環境・タスクに応じて、
 - 音響モデル・言語モデルの組合せ
 - Juliusの自動コンフィグレーション
 - Juliusの標準jconfファイル作成
- 近日中にJuliusのWebサイトからリンク

デモンストレーション

オプション(要望があれば)

- 対話用音響モデル...効果なし(?)
- 対話用言語モデル
- 英語用ディクテーションキット
- 話者認識ツールキット

関連Webページ

- 日本語ディクテーションツールキット
 - 最終版は「音声認識システム」(オーム社)の付録CD-ROM
 - <http://winnie.kuis.kyoto-u.ac.jp/dictation/>
- 連続音声認識コンソーシアム(CSRC)
 - 最終版を一般頒布中(有償)
 - <http://www.lang.astem.or.jp/CSRC/>
- Julius
 - 最新版(SAPI版含む)フリーダウンロード
 - <http://julius.sourceforge.jp/>
- Microsoft Speech Application SDK
 - SALTには必要
 - <http://www.microsoft.com/speech/>
- リーディングプロジェクト e-Society基盤ソフトウェア
 - <http://cif.iis.u-tokyo.ac.jp/e-society/>

講習会

- 日時:8月22日(月)~26日(金)
- 場所:京都大学学術情報メディアセンター
- ISTC一般会員は無料(3名まで)
- 内容:
 - (前半)音声認識の基礎からディクテーションまで
 - (後半)音声合成、顔画像合成、音声・MM対話
- 一般の申込みは締め切りましたが、ISTC会員の方はあと数名受け付けますので、**本日中午**に河原まで連絡して下さい

プログラムの概要

1. 概要と音声の入力・分析: 河原(京大)
2. HMMIによる音響モデル: 篠田(東工大)
3. 記述文法による認識: 伊藤(名大)
4. 形態素解析・言語モデル: 伊藤(東北大)
5. ディクテーション: 李(名工大)
6. 音声合成: 山下(立命大)
7. 顔画像合成: 森島(早大)
8. 音声対話: 荒木(京都工繊大)
9. マルチモーダル対話: 新田(豊橋技科大)