

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
 - Julius 3.5 2005年11月リリース
- 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を容易に
- 音声合成エンジンtalkのSAPI対応

Julius 3.5

2005年11月11日リリース

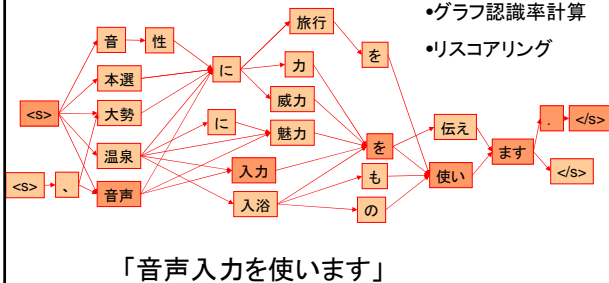
主な更新点

- 多くのバグ修正
- 使用メモリ量の削減
- 単語グラフ出力のサポート
- GMM による環境音識別・不要入力棄却のサポート
- Windows コンソール版の改善 (mingw対応, 文字コード変換等)
- ソースのドキュメンテーション
- ソースの統合 (Linux版, Windows版, マルチパス版が1つに)

新機能

- 単語グラフ出力
 - 外部知識によるリスコアリング
- 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
- 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

Julius 3.5.1 (近日リリース)

- MFCC計算の拡張
 - 二次差分係数(Accel: _A)の抽出に対応。
 - 任意の型指定の組み合わせに対応(_O, _E, _N, _D, _A, _N)
 - 任意の次元数に対応
 - 抽出される特徴量のタイプと次元数は、音響モデルのヘッダから自動判断
 - 新オプション "-zmeanframe": フレームごとの DC offset 除去 (HTK互換)
- モジュールモードの安定化
 - 中断・再開のタイミングに関するバグを複数修正
 - pause/resume による認識の停止・再開時に "<STARTPROC/>" "<STOPPROC/>"を出力
- バグ修正
 - 要素数が24bitに収まらない巨大なN-gramを扱えないバグを修正
 - マルチバス版で3状態(出力1状態)の音響モデルが読み込めないバグを修正
 - 文法の最後の状態がカウントされないことがあるバグを修正

音声認識パッケージ

- 雑音対応版
- 英語版

音声認識ソフトのカスタマイズ

- 認識ソフト開発者向け
 - ソースコード... LinuxとWindowsの統合
 - ドキュメント
 - 一般利用者向け
 - メニュー選択によるカスタマイズ
 - モデルは既存のものから選択
 - 多様なパッケージ
 - 雑音対応版
 - 英語版
 - アプリ開発者
 - 音響モデル
 - 言語モデル
- 前回 → 今回 → 検討・研究中

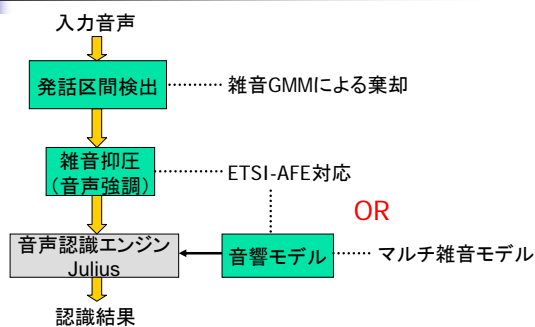
音声認識ソフトウェアのカスタマイズツール

- 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
- JuliusのWebサイトからリンク
 - <http://julius.sourceforge.jp/>
 - <http://htk.ar.media.kyoto-u.ac.jp/julicus/>

雑音対応版パッケージ

- 発話検出・音声切り出し部分の改善
 - GMMによる音声・非音声判別
- 雑音抑圧の前処理ETSI Advanced FEの統合
- 音響モデル
 - ETSI AFE済みのクリーンモデル
 - マルチコンディション学習モデル

耐雑音音声認識の処理の流れ



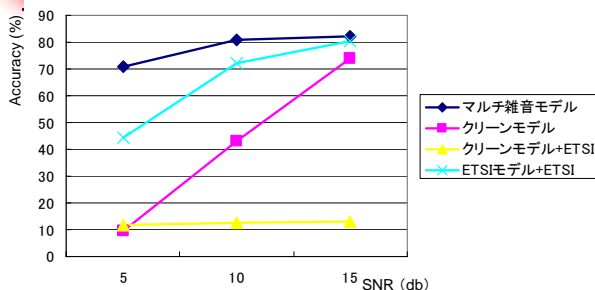
マルチ雑音モデル vs. ETSI AFE

ロボットとの会話を想定した文法による認識
(語彙サイズ: 865)

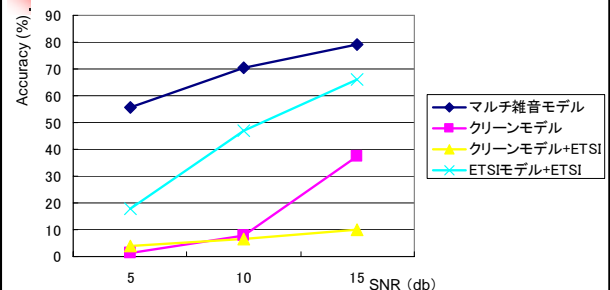
()内は人手切り出し

Accuracy(%)	5dB	10dB	15dB	平均
マルチ雑音モデル	31.3 (62.9)	57.5 (76.8)	68.7 (81.6)	52.5 (73.8)
ETSI-AFEモデル	16.0 (33.6)	46.3 (61.2)	63.7 (74.6)	42.0 (56.5)
IPAクリーンモデル	-0.7 (6.1)	20.2 (28.6)	50.1 (58.8)	23.2 (31.2)

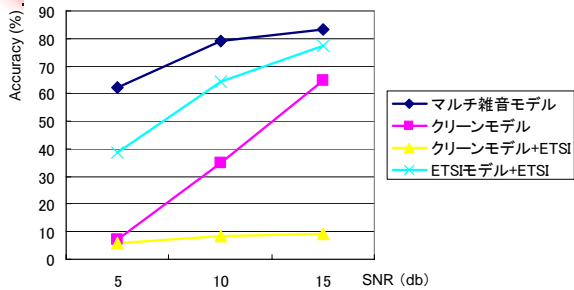
雑音ごとの比較(空調)



雑音ごとの比較(工作機械)



雑音ごとの比較(話し声)



英語版ディクテーションキット

- WSJコーパスでモデル学習
 - 音響モデル...EDAZ, 7000x16
 - Si284set (37K発話, 62時間)
 - 単語辞書...39(CMU phoneset)+sp+sil
 - 言語モデル...5K, 20K trigram
- Juliusの設定
 - マルチパス音響モデル
 - ショートポーズの処理

WSJ評価結果

- ARPA Hub2, Nov. 1993
 - 5K: 215発話、7.3秒
 - 20K: 213発話、7.0秒

	Julius	HTKデコーダ
5K	91.4 /5RT	90.8 /10RT
	88.7 /RT	84.3 /RT
20K	82.4 /5RT	NA
	80.0 /RT	

その他(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
- 話者認識ツールキット

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

関連Webページ

- Julius
 - 最新版(SAPI版含む)フリーダウンロード
 - <http://julius.sourceforge.jp/>
- 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/>
- 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/>
- 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/>
- Microsoft Speech Application SDK
 - SALTには必要
 - <http://www.microsoft.com/speech/>