



## 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

## 音声対話システム用 音響モデルの比較・評価

- 評価データ:
  - 「音声版ダイアログナビ」への入力
  - ソフトウェアサポート知識ベースの検索  
<http://www.ar.media.kyoto-u.ac.jp/msnavi/>
  - 語彙サイズ:18K、言語モデル:クラス3-gramモデル
  - 359発話 / 22名(男性)
  - できるだけ人間相手のように話すよう指示
- 音響モデル:
  - いずれも京都大学とASTEMで作成
  - ATR対話音声モデルはATRの技術を反映するものではありません

## 音声対話システム用 音響モデルの比較・評価

単語認識精度

IPA新聞記事モデル (98時間・264名)	tri2000x16 PTM64	81.1% 79.6%
CSRC音韻バランス文モデル (260時間・4130名)	tri2000x16 PTM64	82.0% 80.0%
	PTM128	81.1%
	PTM256	82.1%
ATR対話音声モデル (17時間・197名男性)	tri2000x16 PTM64	78.6% 75.2%
CSJ講演音声モデル (486時間・2496名)	tri2000x16 PTM64	72.8% 66.4%

- 人間どうしの対話や講演はかえってミスマッチする傾向
- 読上げ音声データベースの方が大規模

## (参考)CSRC音響モデルの 各世代に対する認識性能

評価データ: 新聞記事読上げ(JNAS)

	成人 モデル 128mix	高齢者 モデル 64mix	小児 モ デル 64mix	全世代 モデル 128mix
成人音声 (20K連続)	93.2	-	-	91.5
高齢者音声 (20K連続)	82.1	87.5	-	86.6
小児音声 (20K単語)	54.1	-	75.8	77.1

## 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) (予定)
- 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
- 6月頃に募集予定
  - <http://www.ar.media.kyoto-u.ac.jp/seminar/>

## プログラムの概要(昨年;参考)

1. 音声の入力と分析: 中村(ATR)
2. HMMによる音響モデル: 武田(名大)
3. 記述文法による認識: 甲斐(静岡大)
4. 形態素解析・言語モデル: 河原(京大)
5. ディクテーション: 李(奈良先端大)
6. 音声合成: 山下(立命大)、徳田(名工大)
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)