

# 音声認識理解システムとその応用

篠崎 隆宏 研究室

- 専門分野：音声認識、音声情報処理、機械学習
- Home Page : <http://www.ts.ip.fitech.ac.jp>



## 研究目的

工学の立場から人間の音声認識・理解・学習機能を解明し、コンピュータ上に実現することを目的としています。さらに、それらの機能を備えたシステムの応用を図ります。

## 研究テーマ

### 1. 音声のモデル化とモデル学習法

音声認識ではマイクから取り込んだ音声信号をコンピュータにより解析し、単語列など音声が表示する記号的内容を抽出します。様々な可能性を考慮し最適な認識結果を探索する為には、音声の音響的・言語的な知識を計算機処理に適した形で表現したモデルが必要となります。この音声モデルの善し悪しが、システムの認識性能を決める重要な要素となります。普段私たちは耳から入った音声を当たり前のように認識していますが、実際には発話内容の認識に必要な情報は音声信号中に希薄で曖昧な形で散らばっています。また発話の内容も、必ずしも文法に従ったものばかりではありません。それら曖昧な手掛かりを元に認識を行い、さらには理解を行うため、隠れマルコフモデルやベイジアンネット、ニューラルネットワークなどを用いた音声のモデル化方法について研究しています。

音声モデルは多数のパラメータから構成されています。これらパラメータは事前に学習用データから学習しますが、有限のデータから精度よく推定することが必要です。また既存のモデルを認識対象音声の特性に適合させ認識性能を向上させるための、モデル適応も重要となります。これら学習や適応の問題に対して、アンサンブル法を用いた手法等を提案し研究を行っています（図1）。

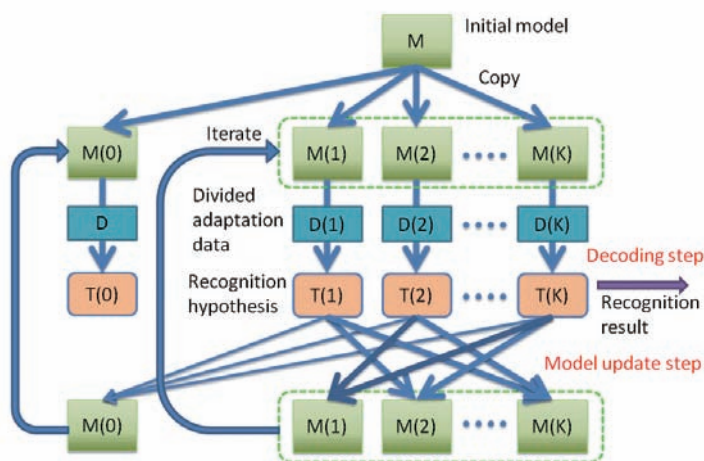


図1 Unsupervised CV adaptation.

### 2. リアクティブ音声認識・理解・応答システムの構成法

音声認識には大規模な音声モデルを用いた効率的な探索が要求されるうえ、音声信号はコンテキスト依存の時系列データであることなどから、既存の認識システムのソフトウェアは複雑で理解が難しいという問題

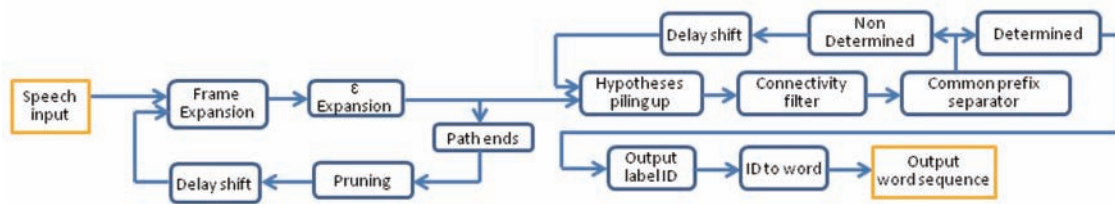


図2 Pipeline based design of a speech decoder.

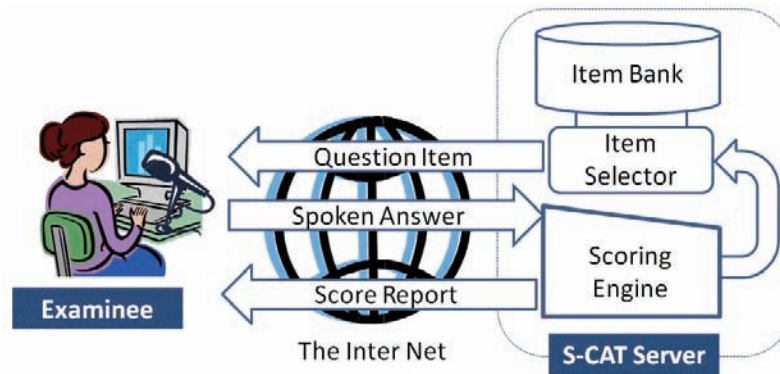


図3 S-CAT automated speaking test system.

があります。また人間と同じタイムスケールで動作するリアクティブ対話ロボットを実現しようとする、システムはさらに複雑になります。そこでこれらのシステムをより容易に実現する方法として、パイプラインに基づいた抽象度の高いシステム記述方法について研究を行っています（図2）。

### 3. 音声情報処理の応用

音声情報処理の応用として、日本語学習者向けに音声発話能力を自動で評価するスピーキングテストシステムの開発を他大学と共同して行っています（図3）。また音声認識で用いられる技術の応用として、目の動きを入力として音声を合成する情報保障のためのシステムや、タンパク質の立体構造予測アルゴリズムの開発などにも取り組んでいます。

## ● 教員からのメッセージ

コンピュータを用いて「新しく面白そうなこと」に挑戦したい学生を歓迎します。

## ● 参考文献

1. Takahiro Shinozaki, Toshinao Iwaki, Shiqiao Du, Masakazu Sekijima and Sadaoki Furui, "Distance-based Factor Graph Linearization and Sampled Max-sum Algorithm for Efficient 3D Potential Decoding of Macromolecules," IPSJ Transaction on Bioinformatics, Vol. 4, pp. 34-44, 2011.12.
2. Takahiro Shinozaki, Yu Kubota and Sadaoki Furui, "Unsupervised Acoustic Model Adaptation Based on Ensemble Methods," IEEE Journal of Selected Topics in Signal Processing, Vol. 4, No. 6, pp. 1007-1015, 2010.12.
3. Takahiro Shinozaki and Mari Ostendorf, "Cross-validation and Aggregated EM Training for Robust Parameter Estimation," Computer Speech and Language, Vol. 22/2, pp. 185-195, 2008.4.
4. Takahiro Shinozaki and Sadaoki Furui, "Dynamic Bayesian Network-based Acoustic Models Incorporating Speaking Rate Effects," Trans of the IEICE, Vol. E87-D, No. 10, pp. 2339-2347, 2004.10.