

音声認識理解とその応用

准教授　篠崎 隆宏

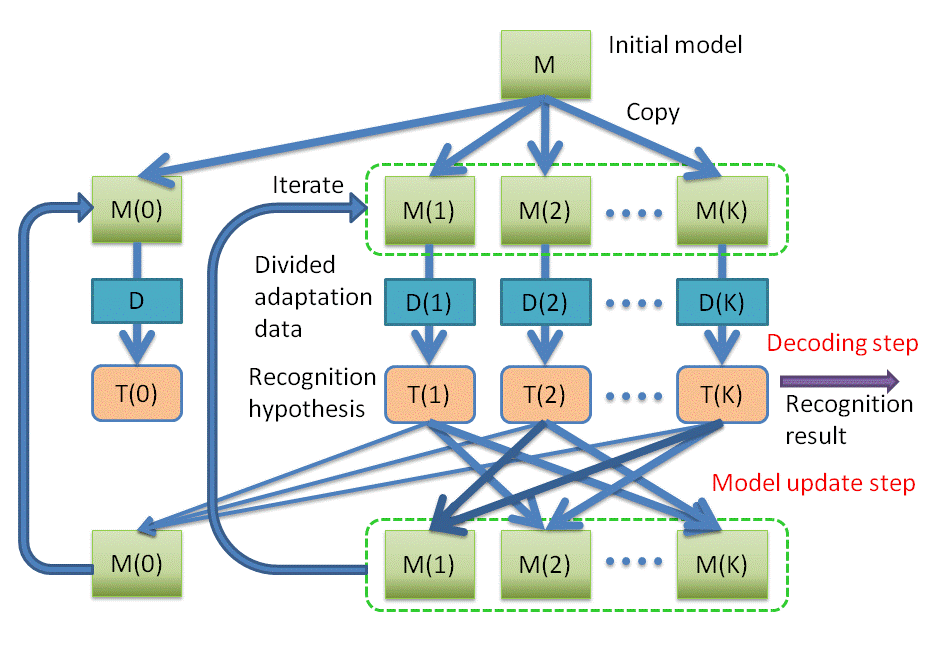
研究分野：音声認識、音声情報処理、機械学習

ホームページ: http:www.ts.ip.titech.ac.jp

●**研究目的・内容**

　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はかります。

我々が音声を認識理解する能力は生まれながらのものではなく、学習により後天的に獲得したものです。音声認識システムでも同様で、音声を認識するたには認識対象言語の音響的・言語的知識をコンピュータ上に取り込み音声モデルとして蓄える必要があります。音声認識システムの認識性能の大部分は、音声モデルの性能で決まります。そのために、隠れマルコフモデルやディープニューラルネットワークなどを駆使した様々な音声モデルが提案されており、それらを用いてより高い認識性能を得るための研究をしています。さらに、システムに音声をより賢く学習し理解させるための仕組みの実現を目指した研究を行っています。

●**研究テーマ**

**１．進化的アルゴリズムを用いた音声認識システムや機械翻訳システムの自動最適化**

性能の高い音声認識システムや機械翻訳システムを構築するためには、ニューラルネットワークの構造や、学習時の各種条件などを最適化する必要があります。従来このような最適化は専門家のノウハウに頼って行われてきましたが、システムの複雑さが増すにつれて次第に実行が困難になってきています。そこで進化的アルゴリズムを応用し、スーパーコンピューター上で少しずつ違った遺伝子を持った多数の認識システムを学習し評価することを繰り返すことで、システムの最適化を自動化する方法について提案し研究を行っています。

図 1 Unsupervised CV adaptation.

**２．音声認識理解システムの教師なし学習・強化学習**

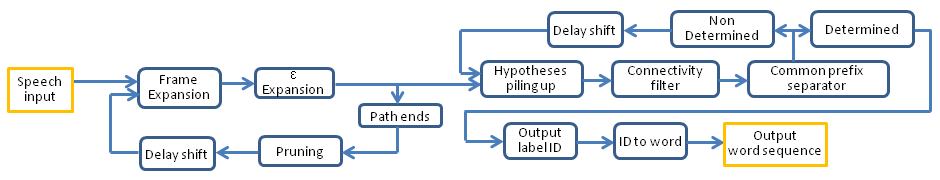
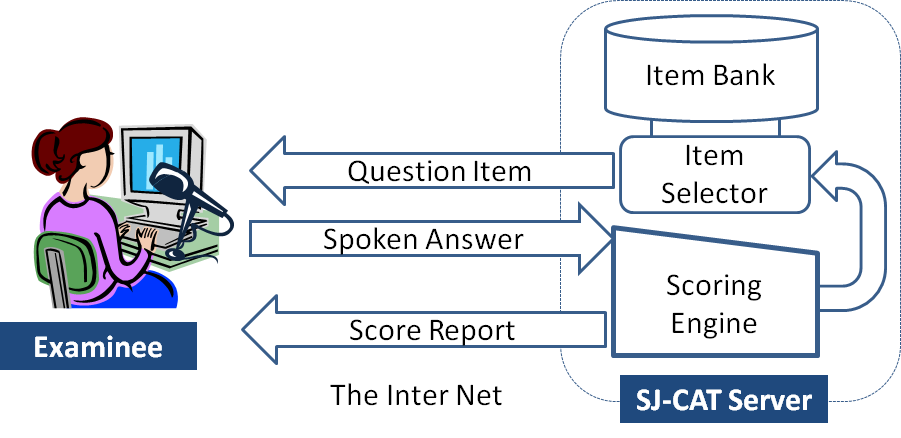
　現在の音声認識技術の限界は、音声モデルの学習を大きなコストのかかる教師あり学習に頼っており、汎化性にも欠ける点です。そこでラベル付き音声を全くあるいは殆んど使用せずに、人間のように音声を学習し理解することの出来るシステムの実現を目指して、音声モデルの教師なし学習や強化学習について取り組んでいます。Sequence-to-Sequence型のニューラルネット等を用いたシステムでは、音声とともに画像など他のモダリティを組み合わせたり、システムに「欲求」に相当するような目的を持たせたりするなどの、フレキシブルな構成が可能です。

図 2 Pipeline based design of a speech decoder.

**３．リアクティブ音声認識・理解・応答システムの構成法**

音声信号はコンテキスト依存の時系列データであるとともに、認識処理は効率的に行う必要があります。このため、人間と同じタイムスケールで動作するリアクティブな対話システムを実現しようとすると、システム実装は複雑になります。そこで高度なシステムを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。

**４．音声情報処理の応用**

　当研究室で開発し一般公開している高性能日本語話し言葉音声認識システム(Kaldi CSJレシピ)は、国内外の多くの企業や大学で使用されています。また音声情報処理の応用として、日本語や英語学習者の音声発話能力を自動評価するスピーキングテストシステムや、機械の動作音を監視し異常を自動検出する仕組み、低消費電力で動作する音声センサーなどにも取り組んでいます。

図 3 SJ-CAT automated speaking test system.

●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。企業や海外の研究機関との協力も積極的に行っています。

●関連する業績、プロジェクトなど

１．T. Moriya, T. Tanaka, T. Shinozaki, S. Watanabe, K. Duh, "Evolution-Strategy-Based Automation of System Development for High-Performance Speech Recognition," IEEE TASLP, pp. 77-88, 2019.

2. B. Zhuang, W. Wang, T. Shinozaki, "Investigation of Attention-Based Multimodal Fusion and Maximum Mutual Information Objective for DSTC7 Track3", Proc. Dialog System Technology Challenges (DSTC7), 2019.

3. T. Kato, T. Shinozaki, "Reinforcement Learning of Speech Recognition System Based on Policy Gradient and Hypothesis Selection," Proc. IEEE ICASSP pp.5759-5763, 2018.