

# 表現豊かな音声合成の実現

教授 小林 隆夫

研究分野：音声情報処理、デジタル信号処理、機械学習

ホームページ: <http://www.kbys.kbys.ip.titech.ac.jp>



## ● 研究内容・目的

音声合成、音声認識、音声分析、音声変換、音声符号化など、音声情報処理のための基盤技術を対象として、信号処理、統計的枠組み、機械学習に基づいた新たな手法の開発を行い、表現豊かで気の利いた音声言語インタフェースの実現をめざす。

## ● 研究テーマ

### 1. 平均声からの多様な音声の合成

与えられた任意のテキストに対応する音声を生成する技術はテキスト音声合成 (TTS) と呼ばれ、自然なコンピュータインタラクションを実現する上で必要不可欠な要素技術となっている。最近では TTS により自然性の高い合成音声が可能になりつつあるが、様々な声質や発話様式・感情表現を持った表現豊かな音声を合成することが課題として残されている。この問題に対し「平均声」からの多様な音声の合成という新たな枠組みを提案している。提案手法では、複数話者の音声から隠れマルコフモデル (HMM) を利用してそのスペクトルや声の高さ(F0)など複数話者の平均的な特徴を表す音声単位モデル (平均声モデル) を求める。音声合成時には、ターゲットとなる話者の数文章の音声(適応データ)を用いて平均声モデルを話者適応したモデルを作成し、これに基づいてターゲット話者の声質、韻律特徴を持った任意文章の音声合成を実現している。

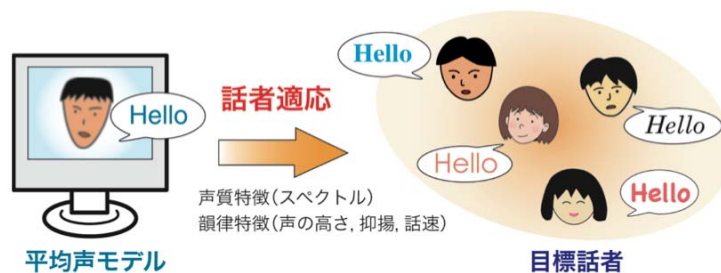


図 1 平均声からの任意話者音声の合成: 少量の目標話者の音声を基にその人の喋りを合成する

### 2. 多様な感情表現・発話様式による音声合成

人間の音声に存在する個人性には、その人独特の話し方や喜び・悲しみといった様々な感情表現が含まれる。ここでは、各話者の様々な感情表現・発話様式を音声のスタイルと呼び、多様なスタイルによる音声合成の実現をめざした研究を行っている。これまでに、HMM 音声合成に基づいたスタイルモデル化、スタイル補間、スタイル適応、スタイル制御などの新たな手法を提案し、多様な話者性と共に多様

なスタイル音声合成が可能であることを示した。図 2 に提案したスタイル制御法の概念図を示す。この手法では、スタイル空間の各軸がある一つのスタイルに相当しており、スタイル空間の座標を指定するだけで、所望のスタイル及びその表出度合を制御することができる。

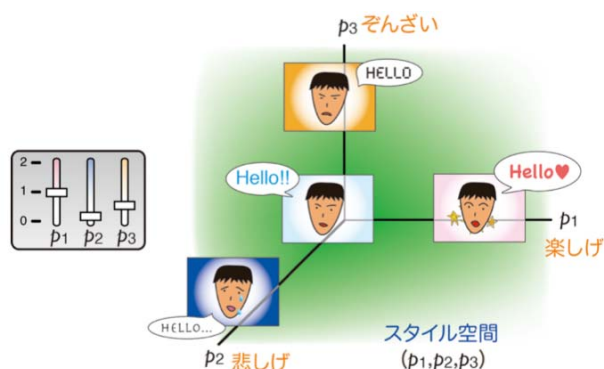


図 2 スタイル制御: 任意のスタイルを指定して音声合成が可能

### 3. 声質変換とその応用

アニメや映画などで主人公の声を別の人の声に変えるいわゆる

「ボイスチェンジャー」は音声処理分野では声質変換と呼ばれ、合成音声に多様な話者性を付与したり、自分の声をプロの声優やナレーターの声に変換できる技術として注目されている。ここでは、新たなアプローチに基づいて、話者の個人性に対しロバストでかつ自然性の高い変換音声の生成をめざすと共に、異なる言語間の声質変換への応用を検討している。

### 4. 音声合成の新たな枠組みの開拓

合成音声の品質を向上させる取り組みとして、従来の HMM に代わりディープラーニングやガウス過程回帰 (GPR) などの機械学習手法を取り入れた新たな音声合成の枠組みが注目されている。我々も、GPR に基づく音声合成手法を世界に先駆けて提案し、合成音声の品質が改善されることを示した。現在、GPR に基づいた多様な話者性やスタイルによる音声合成の実現をめざして研究を進めている。

## ● 教員からのメッセージ

コンピュータが喋ったり人間の言葉を理解する能力がどれだけ人間に近付いたか知っていますか？ 現実にはまだまだなのです。簡単そうなのになぜできないか？ そう思った人は、実はすでに研究に一歩足を踏みいれているのです。一緒に研究をしましょう。

## ● 関連する業績、プロジェクトなど

1. J. Yamagishi, T. Kobayashi, Y. Nakano, K. Ogata, J. Isogai : Analysis of speaker adaptation algorithms for HMM-based speech synthesis and a constrained SMAPLR adaptation algorithm; IEEE Trans. Audio, Speech, and Language Processing, vol.17, 1, pp.66-83 (2009.1)
2. T. Nose, M. Tachibana, T. Kobayashi : HMM-based style control for expressive speech synthesis with arbitrary speaker's voice using model adaptation; IEICE Trans. Information and Systems, vol.E92-D, 3, pp.489-497(2009.3)
3. T. Koriyama, T. Nose, T. Kobayashi : Statistical parametric speech synthesis based on Gaussian process regression; IEEE Journal of Selected Topics in Signal Processing, vol.8, 2, pp.173-183 (2014.4)
4. 小林隆夫 : 音声合成技術; 小特集 2020 年, 言葉の壁を超える音声翻訳 —新しい技術と研究の可能性—, 電子情報通信学会誌, vol.98, 8, pp.726-733 (2015.8)