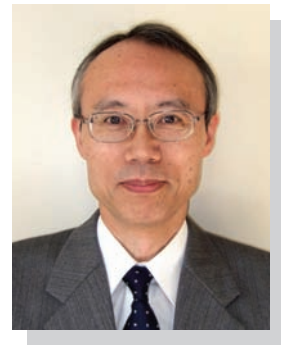


# 音声を用いた ヒューマンインタフェースの実現

小林 隆夫 研究室

- 専門分野：音声情報処理、マルチモーダルインタフェース、デジタル信号処理
- Home Page： <http://www.kbys.ip.titech.ac.jp/>



## 研究目的

視覚、聴覚など複数の感覚・知覚情報を統合した新たな情報メディアと信号処理・音声情報処理手法の開発を行い、人間に優しい音声言語・マルチモーダルインタフェースの実現をめざしている。

## 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。最近のテキスト音声合成システムでは、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、これを音声単位とし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からその人の声で話す音声合成システムを実現

### 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、多様な話者性ととも、多様なスタイル音声が可能であることを示した。図2に提案する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。

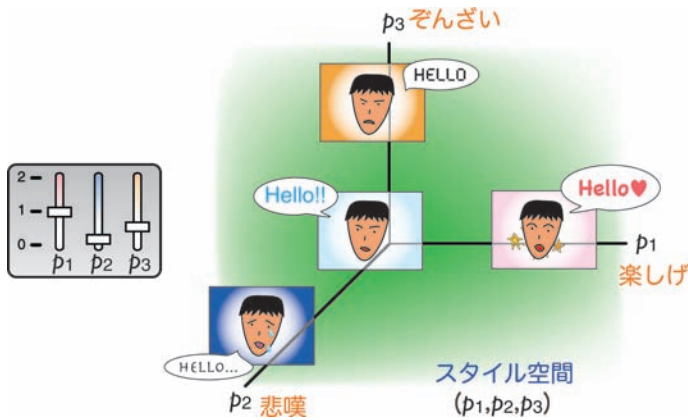


図2 スタイル制御：任意スタイルを指定して音声合成が可能

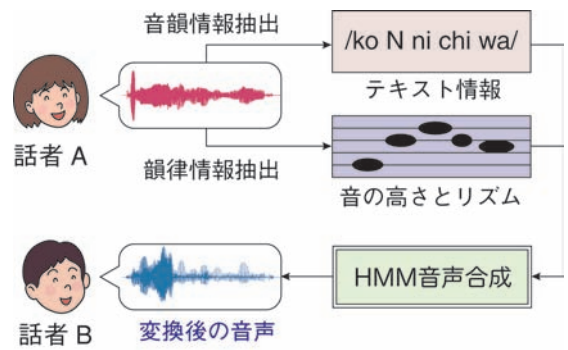


図3 声質変換システム：発話内容はそのまま話者性のみを変換することが可能

### 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」は音声信号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、新たなアプローチとして、HMM音声合成に基づく手法を提案している。これにより話者の個性に対しロバストでかつ自然性の高い変換音声の生成が可能となる。

### 4. 動作プリミティブからのヒューマンモーション生成

自然言語による動作の記述からそれに対応する手や腕の動作のアニメーションを自動生成するためのテキスト→モーション変換の枠組みを提案し、多様な音声合成の研究と併せて、新たなマルチモーダルインタラクションの開拓をめざしている。提案手法では、あらかじめ決められた動作だけでなく、最低限の動作プリミティブが定義されていれば、任意の動きの生成が可能となる。動作プリミティブのモデル化にはHMMを利用している。

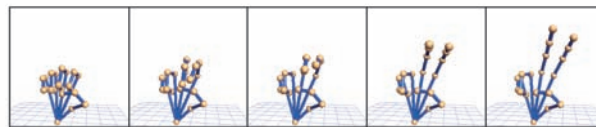


図4 じゃんけんにおける指の動きの生成例。動作記述「グーからチョキ」

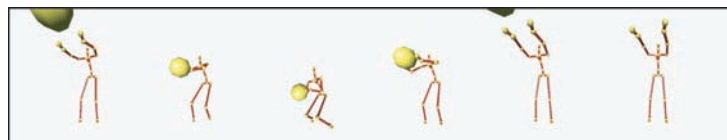


図5 ボールの位置・速度を考慮した捕球動作の生成

## ● 教員からのメッセージ

コンピュータが話し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入れているのです。一緒に研究してみませんか。

### ● 参考文献

1. Junichi Yamagishi, Takao Kobayashi, Yuji Nakano, Katsumi Ogata, Juri Isogai : Analysis of speaker adaptation algorithms for HMM-based speech synthesis and a constrained SMAPLR adaptation algorithm; IEEE Trans. Audio, Speech, and Language Processing, 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, E92-D, 3, pp.489-497 (2009.3)
3. T. Nose, T. Kobayashi: Speaker-independent HMM-based voice conversion using adaptive quantization of the fundamental frequency; Speech Communication, vol. 53, 7, pp.973-985(2011.9)
4. N. Niwase, J. Yamagishi, T. Kobayashi : Human walking motion synthesis with desired pace and stride length based on HSMM; IEICE Trans. Information and Systems, E88-D, 11, pp.2492-2499 (2005.11)
5. <http://www.kbys.ip.titech.ac.jp/demo/index-j.html>.