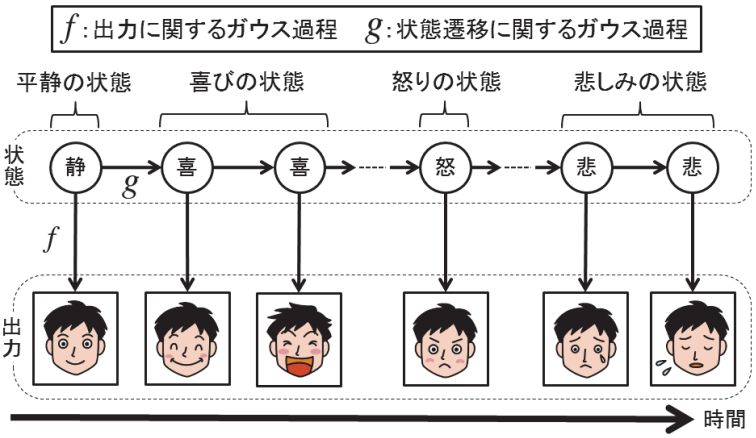
K13 非言語的情報を活用したマルチモーダル音声対話システム

情報科学部・情報科学科・准教授・玉森　聡
akira-tamamori@aitech.ac.jp

キーワード　音声対話システム、非言語的情報、マルチモーダル

概要

近年の音声認識・音声合成・画像処理技術の進歩を背景に、音声情報と映像情報を利用したマルチモーダル対話システムの研究が行われている。先行研究では人間と対話システムの間のコミュニケーションにおいて、対話中におけるユーザの表情や、同意を表すうなずきの仕草、否定を表す首振りの仕草などの「非言語的情報」を活用することの有効性が示されている。しかしながら、これまでの先行研究においては、対話中における表情の動的な変化をモデル化する試みはいくつかあるが、その方法は未だ確立されていない。そこで本システムでは、ユーザの表情や感情の変化を予測・検出する方法を音声対話システムへ組み込むことで、ユーザにとってよりフレンドリーとなる新たなマルチモーダル音声対話システムの実現を目指す。



セールスポイント

1．人間の表情から感情を高速に推定できる
2．表情変化の裏に潜む複雑な感情変化を予測・理解できる

企業等での活用例、今後の展望等

1．対話型インタラクティブ教育支援システムの実現
2．スマートスピーカーとの連携による新たなIoTサービスの創出
3．表情理解に基づく対話型医療支援システムの実現

参考資料

Akira Tamamori and Tomoko Matsui: “A sequential prediction method of quasi-periodicity based on Gaussian process state space model,” 2019 APSIPA ASC, pp. 255-261.