

音声中の非言語情報の生成・知覚の特性解析と多言語間コミュニケーションへの応用 (071705001)

Analysis of production and perception mechanisms of non-linguistic information in speech and its application to inter-language communication

研究代表者

赤木正人 北陸先端科学技術大学院大学 情報科学研究科

Masato Akagi School of Information Science Japan Advanced Institute of Science and Technology

研究分担者

党建武[†] 鶴木祐史[†] 徳田功[†] Li Junfeng[†] Lu Xugang[†] 末光厚夫[†]

Donna Erickson^{††} 榊原健一^{†††} 北村達也^{††††}

Jianwu Dang[†] Masashi Unoki[†] Isao Tokuda[†] Li Junfeng[†] Lu Xugang[†] Atsuo Suemitsu[†]

Donna Erickson^{††} Ken'ichi Sakakibara^{†††} Tatsuya Kitamura^{††††}

[†]北陸先端科学技術大学院大学 情報科学研究科 ^{††}昭和音楽大学音楽学部

^{†††}北海道医療大学 心理科学部 ^{††††}甲南大学知能情報学部

[†]Japan Advanced Institute of Science and Technology ^{††}Showa Music University

^{†††}Health Sciences University of Hokkaido ^{††††}Konan University

研究期間 平成 19 年度～平成 21 年度

概要

本研究では、音声における非言語情報のコミュニケーションにおいて、言語・民族・文化を越えたユニバーサルなコミュニケーション環境を構築するために、(1) 非言語情報の知覚・生成について「何が本質なのか」を考察し、(2) 言語・民族・文化によらないヒトの生物学的「共通要素」の存在を明らかにする、そして、(3) この「共通要素」を音声認識・合成などに工学的に応用すること、について研究を行う。これにより、非言語情報によるコミュニケーションの解明が進み、グローバルでユニバーサルなコミュニケーション環境構築の一助となる。また、工学的応用として、多言語への非言語情報の付加（音声合成への応用）、非言語情報の認識（音声認識・対話理解への応用）への道が開け、多言語間の円滑なコミュニケーション、また、幼児や老人、音声・聴覚障害者も含めたユニバーサルなコミュニケーションへの応用も可能となる。

Abstract

This project aims at constructing global and universal communication environments beyond languages, nations and cultures based on non-linguistic information in speech. Towards this purpose, the issues include (1) what is essential in production and perception of non-linguistic information; (2) what are biological common features among persons that are independent of languages, nations and cultures; (3) How to apply the discovered common features to speech recognition and speech synthesis applications. This project will contribute a lot to build up global and universal communication based on non-linguistic information, and to expressive speech synthesis and non-linguistic information recognition with further application to speech recognition among multiple languages. Moreover, it will also help to realize the universal smooth communication among elders, infants, handicapped persons, as well as those in different languages, nations and cultures.

1. まえがき

近年一層の国際化が進むにあたり、言語・民族・文化を越えた（＝グローバルな）、また、言語・民族・文化のみならず老人、幼児、あるいは障害者との障壁のない（＝ユニバーサルな）コミュニケーションの重要性が増している。

本研究では、音声における非言語情報のコミュニケーションにおいて、言語・民族・文化を越えたユニバーサルなコミュニケーション環境を構築するために、

1. 非言語情報の知覚・生成について「何が本質なのか」を考察する、
2. 言語・民族・文化によらないヒトの生物学的「共通要素」の存在を明らかにする、そして、
3. 「共通要素」を多言語間での非言語情報の付加（音声合成への応用）、非言語情報の認識（音声認識、音声対話理解への応用）に応用する

2. 研究内容及び成果

2.1. 研究概要

音声コミュニケーションにおいて音声の生成と知覚は不可分であり、しかも図 1 に示すように環構造となっている。このため、非言語情報の生成・知覚を総合的に探求し、調査・研究するためには、環構造の中でのそれぞれの関係を考慮しながら研究を進める必要がある。本研究では、上記目的を実現するために、図 1 中の脳と音声生成の相互作用、脳と音声知覚の相互作用、生成から音声特徴を経て知覚への経路、それぞれの中で、非言語情報の生成・知覚機構の解明を目指し、言語・民族・文化によらない共通要素とは何かについて検討した。そして、共通要素を核として、非言語情報（感情音声）の合成・認識を試みた。なお、研究を遂行するための方略として、ヒトの観測→モデル化→工学システムの実現という 3 つのフェーズを設定した。

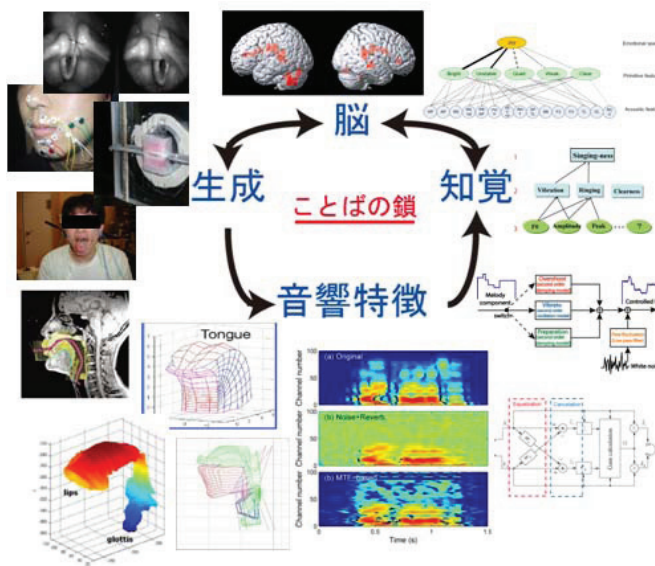


図1 音声生成・知覚の環構造と研究成果（ヒトの観測・モデル化・工学システム）

2.2. 研究成果

【観測】感情音声生成時の音声および生理指標を同時に観測したデータの作成、および、音声生成・知覚の環構造におけるそれぞれの関係の記述を行った。

1. 声帯振動の3次元計測：喉頭音源の生成メカニズムの解明のために、従来の撮影速度の2倍に相当する9000コマ/秒でのステレオ視による高速度デジタル撮影
2. 舌運動の計測：人間の音声生成に関するメカニズムを究明するために、3次元実時間磁気センサシステム（EMA）および核磁気共鳴画像法（MRI）により計測
3. この他に、喉頭針筋電、身体他の部位の表面筋電、声門下圧、流量計測、声帯振動の高速度撮影、骨導音声測定しこれらを統合したデータベースを作成
4. 音声特徴抽出の精緻化：雑音・残響等の外乱に左右されない高精度の分析方法の検討
5. 感情知覚に関する心理物理実験：感情知覚について多言語間での比較を行い共通性および違いについて議論
6. fMRIによる情動に関係した脳活動計測、情動系からの調音器官への影響の調査

【モデル化】感情音声生成・知覚における計算機上での実働モデルの作成を行った。

1. 声区（レジスタ）のモデル：胸声およびファルセットを生成する3質量声帯振動モデルを構築
2. 生理学的発話機構モデルの開発・応用：EMAおよびMRIにより計測した舌運動にもとづいて3次元生理学的発話機構モデルを開発。感情音声発話時の声道形状および音響特徴をシミュレーション
3. 感情音声知覚の心理モデル：非言語情報の知覚と音響特徴を結び付ける知覚モデルとして、形容詞を第2層に含む聴覚印象の三層モデルを提案

【工学システム】生理学的生成モデルにもとづいた音声合成システム、および、非言語情報知覚モデルにもとづいた非言語情報の合成・認識システムの実現を行った。

1. 音声中の感情認識システム（プロトタイプ）の構築
2. 表現豊かな音声の合成・音声変形システムの開発：感情音声、歌声の合成システムの開発。歌声合成システムについては、Interspeech 歌声合成コンテストで第1位を獲得した。

3. むすび

本研究では、音声生成・知覚の環構造の中で、ヒトの観測、モデル化を通して、非言語情報の生成・知覚機構の解明を目指し、言語・民族・文化によらない共通要素とは何かについて検討した。そして共通要素を核として、非言語情報（感情音声）の合成・認識のための工学システムの実現を試みた。その結果、音声中の非言語情報の生成と知覚に関する環構造を裏付ける結果が得られ、最終的な目標が達成された。また、認識システム、および、合成システムのプロトタイプ構築を行った。さらに、これらの研究について多くの発表を行い、発表した論文のいくつかに対して賞を授与された。

感情などの非言語情報を認識し、また、音声に付加し制御できるようになれば；

1. 非言語情報によるコミュニケーションの解明が進み、グローバルでユニバーサルなコミュニケーション環境構築の一助となる
2. 工学的応用として、多言語への非言語情報の付加（音声合成への応用）、非言語情報の認識（音声認識・対話理解への応用）への道が開け、多言語間の円滑なコミュニケーション、幼児や老人、音声・聴覚障害者も含めたユニバーサルなコミュニケーションへの応用も可能となる

【誌上発表リスト】

- [1] Sawamura K., Dang J., Akagi M., Erickson D., Li, A., Sakuraba, K., Minematsu, N., and Hirose, K., “Common factors in emotion perception among different cultures”, Proc. ICPhS2007 2113-2116 (2007年8月7日)
- [2] Huang, C-F., and Akagi, M., “A three-layered model for expressive speech perception”, Speech Communication 50, 810-828 (2008年10月1日)
- [3] Tokuda, I., Zemke, M., Kob, M., and Herzel, H., “Biomechanical modeling of register transitions and the role of vocal tract resonators”, Journal of the Acoustical Society of America, Vol. 127, Issue 3, 1528-1536 (2010年3月10日)

【受賞リスト】

- [1] Vu, Thang, Unoki, Masashi, and Akagi, Masato, 信号処理学会 Best Paper Award, “A Study on Restoration of Bone-Conducted Speech with MTF-Based and LP-based Models”, 2009年3月2日
- [2] 齋藤毅、後藤真孝、鶴木祐史、赤木正人、情報処理学会インタラクティブ2009 インタラクティブ発表賞、“SingBySpeaking：ユーザの話を歌声に変換する歌声合成インタフェース”、2009年3月6日
- [3] 齋藤毅、辻直也、鶴木祐史、赤木正人、日本音響学会佐藤論文賞、“歌声らしさの知覚モデルに基づいた歌声特有の音響特徴量の分析”、2010年3月9日

【報道発表リスト】

- [1] “言語を超えて意思疎通 先端大の赤木教授ら研究”、北國新聞、2007年5月10日
- [2] “歌声はなぜ心に響くのか”、東京MX-TV ガリレオチャンネル、2008年11月15日
- [3] “七色の声の謎を追う”、神戸新聞、2009年7月28日

【本研究開発課題を掲載したホームページ】

<http://www.jaist.ac.jp/~akagi/>