

音声の個人性知覚における話者内変動への頑健性に関する研究 (042107002)

A Study on Robustness for Intra-Speaker Variation on Perception of Speaker Characteristics

研究代表者

北村 達也 株式会社国際電気通信基礎技術研究所 認知情報科学研究所 (現: 甲南大学 理工学部)
Tatsuya Kitamura ATR Cognitive Information Science Laboratories (Currently with Faculty of Science
and Systems Engineering, Konan University)

研究分担者

パーハム・モクタリ[†] 齋藤 毅^{††}

Parham Mokhtari[†] Takeshi Saitou^{††}

[†]株式会社国際電気通信基礎技術研究所 認知情報科学研究所 (現: 独立行政法人情報通信研究機構 ユニバーサルメディア研究センター) ^{††}株式会社国際電気通信基礎技術研究所 認知情報科学研究所 (現: 独立行政法人産業技術総合研究所 情報技術研究部門)

[†]ATR Cognitive Information Science Laboratories (Currently with Universal Media Research Center, NICT) ^{††}ATR Cognitive Information Science Laboratories (Currently with Information Technology Research Institute, AIST)

研究期間 平成 16 年度～平成 18 年度

本研究開発の概要

音声の中の種々の特徴量は発話内容や話者の状態などによって様々に変化するにもかかわらず、人間は音声から個人性(話者の特徴)情報を抽出し話者を識別することができる。本研究開発では聴取実験を通して人間のこの能力の解明を試みるとともに、得られた知見を利用して個人性の再現性が高い音声合成システムを開発した。個人性が顕著に現れる母音を中心に検討を行い、(1) 音声には母音間に共通する個人性情報が存在すること、(2) 発話機構の制約から話者が意識的に制御しにくく、話者内変動の小さい特徴量ほど個人性知覚に寄与すること、(3) 連続音声においても(2)のような特徴量が個人性知覚に寄与すること、(4) 音声スペクトルの高周波数成分に加え、従来個人性知覚との関連が議論されてこなかった声帯音源波の非周期性が個人性知覚に寄与することなどを明らかにした。これらの結果は、人間は、話者内変動が小さい特徴量を選択的に利用することによって話者内変動に対する頑健性を実現している可能性が高いことを示している。

Abstract

The aim of this study is to explore mechanisms of the humans' ability to identify speakers from their voices adapting to intra-speaker variations through psycho-acoustic experiments. The results showed that humans perceive speaker characteristics by focusing on less arbitrary and more invariant acoustic parameters. It is thus probable that humans possess robustness against intra-speaker variations on speaker identification by using such acoustic parameters. Based on the results, we also developed an articulatory speech synthesizer, which can provide high-fidelity reproduction of speaker characteristics.

1. まえがき

音声の中の種々の特徴量は、発話内容や話者の状態などによって様々に変化する。それにもかかわらず、人間は音声から一貫した個人性(話者の特徴)情報を抽出し、話者を識別することができる。つまり、人間は話者内変動に対して頑健に個人性を知覚する能力を持っているのである。人間がどのような特徴量とストラテジーを用いてこの頑健性を実現しているのかを明らかにできれば、様々な技術的応用が可能である。本研究開発では、以上の考えにもとづいて人間が音声の個人性知覚の際に用いている特徴量を明らかにした。さらに、得られた知見を応用して個人性の再現性が高い音声合成システムを開発した。

2. 研究内容及び成果

2.1 音韻間に共通する個人性情報^[1]

音声には声帯音源や声道音響特性などの生得的な身体特徴に起因する情報が含まれている。従って、音声には音韻に依存しない個人性情報が含まれていると考えられる。そこで、「母音間に共通する個人性情報(個人性に関する不変項)が存在する」との仮説を立て、これを検証する聴取実験を行った。実験では、実験参加者が声を知らない話者

(未知話者)の音声を用い、ある母音から知覚した個人性情報によって別の母音の話者を識別できるか否かを調査した。さらに、このタスクにおけるピッチ周波数の影響についても調査した。

実験の結果から次のことが明らかになった。(1) 実験参加者は任意の母音から抽出した個人性情報を用いて別の母音の話者を識別することができる。(2) 未知話者の識別においてピッチ周波数が重要であるが、ピッチ周波数を話者間で正規化してもある程度話者を識別できる。以上の結果は、母音間に共通する個人性情報が存在し、その情報は音声スペクトルに含まれていることを示している。

2.2 持続母音に含まれる個人性情報^[2]

上記の実験により、母音間に共通する個人性情報の存在が確認された。そこで、「話者が意識的に制御しにくく、話者内変動の小さい特徴量ほど個人性知覚に寄与する」という仮説を立て、聴取実験によってこれを検証した。

この実験では未知話者10名の持続母音/a/を対象にして、ピッチ周波数と振幅の時間パターン、平均ピッチ周波数、声帯音源波の周波数特性、音声スペクトルの高周波数成分の個人性知覚に対する寄与を調査した。これらの特徴量を

変化させた刺激音を作成し、サーストンの一対比較法により聴取実験を行った。そして、これらの特徴量の個人性知覚に関する間隔尺度を求めた。上述の仮説にもとづけば、声帯音源波の周波数特性と音声スペクトルの高周波数成分の寄与が大きいと予想された。

この実験によって、(1) 音声スペクトルの高周波数成分、(2) 声帯音源波の周波数特性、(3) 平均ピッチ周波数、(4) ピッチ周波数と振幅の時間変動パターン、の順序で個人性知覚に対する寄与が大きいという結果が得られた。このリストは、より上位の特徴量ほど発話機構の生得的な特徴に依存し、話者が意識的に制御しにくいものであると見ることが出来る。例えば、声道下部は母音の通常発声時の形状変化が小さく、その形状は音声スペクトルの高周波数成分に大きな影響を与えることが研究代表者らにより明らかにされている (Kitamura et al., 2005)。従って、人間が音声の個人性を知覚する際には、上記の仮説の通り音声中で変動しにくい特徴量を重視していると結論される。そして、このことによって、人間の個人性知覚は音声の中の特徴量の話者内変動に対して頑健性を有しているものと考えられる。

2.3 連続音声に含まれる個人性情報

持続母音を対象にして得られた上記の結果が連続音声にもあてはまるか否かを検討した。連続音声の場合は時間変動成分にも個人性が含まれる。そのため、各種特徴量の個人性知覚に及ぼす影響が持続母音とは異なる可能性がある。そこで、未知話者10名の3連続母音 (/aea/, /ouo/ など) を対象にして聴取実験を行った。

この実験でも、声道や声帯音源波の音響特性に起因する特徴量に着目した。すなわち、平均ピッチ周波数、音声スペクトルの高周波数成分、声帯音源波の周波数特性と非周期性を対象にして、これらの特徴量を操作した合成音声を用いて聴取実験を行った。A-X法(二肢強制選択法)による話者識別実験の結果、持続母音の場合と同様に、音声スペクトルの高周波数成分の寄与が大きいこと、声帯音源特性が話者識別に影響を与えていることが明らかとなった。

さらに、サーストンの一対比較実験によって各種特徴量の個人性知覚に与える影響を比較し、(1) 音声スペクトルの高周波数成分、(2) 声帯音源の周波数特性、(3) 平均ピッチ周波数の順で影響の大きいことを明らかにした。この結果も持続母音を対象にした実験結果と同じ傾向を示している。以上の結果から、連続音声を対象にした個人性知覚においても、音声中で変動しにくい特徴量を重視していることが明らかになった。

加えて、これまで個人性知覚との関連が議論されてこなかった声帯音源の非周期性が個人性知覚に寄与することが明らかになった。これは音声による生体認証技術のための新たな特徴量が発見できたことを意味する。今後、その有効性について検討する必要がある。

2.4 声道下部の音響特性を反映させた音声合成システムの開発^[3]

これまでの検討によって、音声スペクトルの高周波数成分が音声の個人性知覚に大きく寄与することが明らかになった。上記のように音声スペクトルの高周波数成分の個人差は、声道下部の形状の個人差に由来することが知られている。従って、声道下部の音響特性を反映させられる音声合成システムを作ることができれば、従来法よりも個人性の再現性が高い音声を合成できると考えられる。

そこで、Maeda(1982)により提案された、声道内の音波伝搬を時間領域でシミュレートする手法にもとづいて、声

道下部に含まれる分岐管など(喉頭腔や梨状窩)の音響特性を正確に反映させられる音声合成システムを開発した。この手法では、図1に示すように声道の音響特性を等価回路で模擬し、声帯音源波を入力することによって音声を合成する。これは、人間の音声発話原理をそのままモデル化した手法であり、次世代の音声合成方式としての期待が大きい。

本研究では、Maeda(1982)が提案した数学モデルを変形し、声道に任意の数の分岐管を接続できるようにした。これによって、声道下部のみならず鼻腔や副鼻腔も含めた声道の正確な再現を可能にした。得られた数学モデルにもとづいてプログラムを作成し、磁気共鳴画像法(MRI)により得られた声道形状を用いて音声を合成した。その結果、従来法と比較して合成音声の個人性の再現性が高く、さらに自然性も向上していることが確認できた。

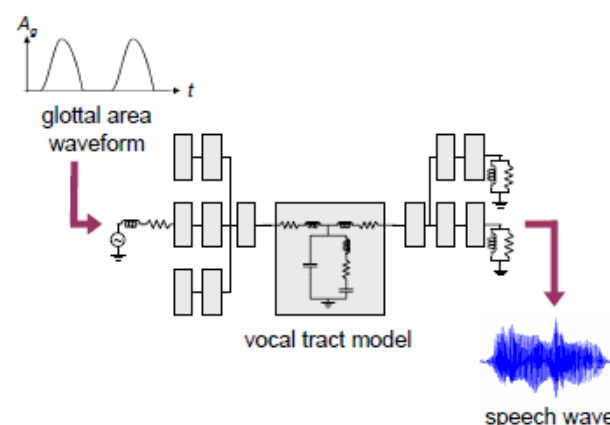


図1 開発した音声合成システムの概要

3. むすび

本研究開発では、人間が音声の話者内変動に対して頑健に個人性を知覚することができる能力の解明を目指した。音声の個人性が顕著に現れるとされる母音を主な対象として聴取実験を行った。そして、人間は個人性知覚の際に音声中で変動しにくい特徴量、すなわち話者が意識的に制御しにくい特徴量を重視していることを明らかにした。人間はこの戦略によって話者内変動に対する頑健性を実現している可能性が高い。今後は、得られた知見にもとづく技術開発を進めていく予定である。

【誌上発表リスト】

- [1] T. Kitamura and P. Mokhtari, "Effects of vowel types on perception of speaker characteristics of unknown speakers," Proceedings of the 2006 RISP International Workshop on Nonlinear Circuits and Processing (NCSP'06), pp. 45-48 (2006年3月3日)
- [2] T. Kitamura and T. Saitou, "Effects of acoustic modification on perception of speaker characteristics for sustained vowels," Acoustical Science and Technology (投稿中)
- [3] P. Mokhtari, H. Takemoto, and T. Kitamura, "Single-matrix formulation of a time-domain acoustic model of the vocal tract with side-branches," Speech Communication (投稿中)

【本研究開発課題を掲載したホームページ】

<http://basil.is.konan-u.ac.jp/>