

## 科学研究費助成事業 研究成果報告書

平成 26 年 6 月 28 日現在

機関番号：31303

研究種目：挑戦的萌芽研究

研究期間：2012～2013

課題番号：24650137

研究課題名(和文)フォルマント抑圧音声を用いた母音の個人性知覚に関する研究

研究課題名(英文)A speaker identity perception for formant-suppressed vowels

研究代表者

伊藤 仁(Ito, Masashi)

東北工業大学・工学部・准教授

研究者番号：00436164

交付決定額(研究期間全体)：(直接経費) 2,300,000円、(間接経費) 690,000円

研究成果の概要(和文)：本研究課題ではフォルマント抑圧刺激を用いた知覚実験により、音声から話者の個人性を知覚する認知科学的メカニズムの解明を目指した。様々な年齢と性別の成人話者300名が発話した母音の音響分析を行い、話者知覚に関してスペクトル全体形状に基づく特徴量が既知のフォルマント周波数と同等以上の情報量を持つことを明らかにした。またフォルマント抑圧実験を用いた母音の知覚実験から、少なくとも声道長と基本周波数が個人性知覚において重要な手掛かりであることを示した。しかし実験結果から得られた知覚特性は聴者により大きく異なっており、すべての被験者の知覚特性を説明できる一般的な知覚モデルを構築することはできなかった。

研究成果の概要(英文)：In the present study, perceptual mechanism for speaker identity was investigated using formant-suppressed vowel stimuli. At first, Japanese vowels were acoustically analyzed which uttered by 300 adult male and female speakers. The results indicated that whole-spectral characteristics had more information than well-known formant frequencies. Based on the analysis, formant-suppressed vowel stimuli were synthesized. Using the stimuli, perceptual experiments were carried out. In the experiments, three vowels (ABX) were successively presented to listeners. The listeners were required to answer vowel A or B, which was uttered by the same speaker for vowel X. The results indicated that both fundamental and formant frequencies to be effective cues for speaker identity perception. However, the results also showed crucial listener dependencies. Thus, it was difficult to generalize perceptual mechanism for speaker identity.

研究分野：情報学

科研費の分科・細目：人間情報学・認知科学

キーワード：音声 認知科学 話者

1. 研究開始当初の背景

我々の音声は、調音器官の個体差や発話の癖などの違いにより、同じ言葉を発話した場合でも音響特性が大きく異なる。例えば母音は、声道伝達特性の共振ピークであるフォルマント周波数により特徴付けられ、単一話者の母音なら低次の2つのフォルマント周波数(F1,F2)だけで概ね識別できる。だが、例えば男性と女性ではF1やF2の範囲が異なるため、男性の/a/と女性の/o/など音韻性が異なる母音が、良く似たフォルマント周波数を持つ場合がある。音声知覚系には、このような個人差を正規化し、発話者が意図した音韻を正確に知覚する仕組みが存在する。また一方で、我々は音声から話者の性別や年齢等の個人性を知覚することも可能であり、この音韻性と個人性の知覚メカニズムは互いに密接に関係していると考えられる。

だが、個人性知覚に関するこれまでの研究は、ひとつの音韻を対象として個人性に影響する音響パラメータを調べたものや、基本周波数やスペクトル包絡など多くの音韻に共通する一般的な知覚の手がかりについて調べたものが殆どであり、音声知覚における個人性と音韻性の関係は十分に明らかにされていない。例えば、我々は2つの母音/a/と/i/を聞いて、それらの話者が同一であることをどの程度正確に判断できるのか、またこの判断の正確さは音韻によりどの様に変化するのかなど、知覚のメカニズムを検討する際に必要となる基礎的なデータは、これまで報告されていない。

2. 研究の目的

我々は音声を聞いて、その音韻・言語的な意味だけでなく、話者の年齢や性別など個人性を知覚することができる。これまでの研究では、この音韻性と個人性の知覚特性が個別に調べられてきたため、これらの認知科学的な関係が十分に明らかになっていない。本申請課題の目的は、(1)単独発話された母音に対する個人性の知覚能力を定量化し、(2)この個人性の知覚が音韻の違いによりどの様に影響されるか調べ、(3)これらの関係と認知科学的メカニズムを明らかにすることである。これを実現するため、多数の話者が発話した母音信号と、そのフォルマントピークを抑圧した信号を刺激として、4種類の音声聴取実験を行う。

3. 研究の方法

本課題研究では、以下の4つの心理物理実験により、母音の音韻性と個人性の知覚特性を調べる。まず各実験で刺激として用いる音声信号を選定するために、多数の話者が発話した母音について音響分析を行い、代表的な話者を選定する。次に選定した話者の母音を刺激として、音韻性の知覚実験を行う(実験)。この実験では、被験者に単一の母音信号を呈示して、日本語5母音のいずれが知覚

されたか回答させる。また同様の実験を、フォルマント抑圧母音を刺激として行う(実験)。その後、標準母音3つをひとつの刺激として、ABX法により話者の個人性の知覚特性を調べる実験を行う(実験)。最後に、フォルマント抑圧母音を用いて同様の実験を行い(実験)、4つの実験結果を比較することで、母音の個人性知覚における音韻性の影響を定量的に評価し、あり得る知覚のメカニズムについて検討する。

4. 研究成果

まず様々な年齢の成人男女計300名が発話した日本語5母音の音響特性を分析した。図1に第1~3フォルマント周波数の平均と分散を示す。図から明らかなように、母音のフォルマント周波数は男性より女性の方が高いが、この差は対数軸上での平行移動として表される。この結果は、母音の個人性が主に話者の声道長で特徴付けられることを意味している。また声の高さに対応する基本周波数についても男女間で明確な差が存在することが確認できた。図2に話者ごとの平均基本周波数と平均フォルマント周波数のヒストグラムを示す。

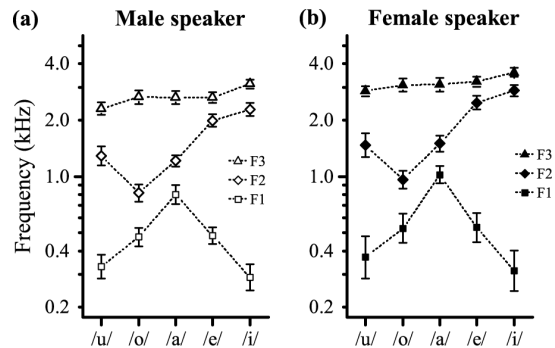


図1. 日本語5母音のフォルマント周波数

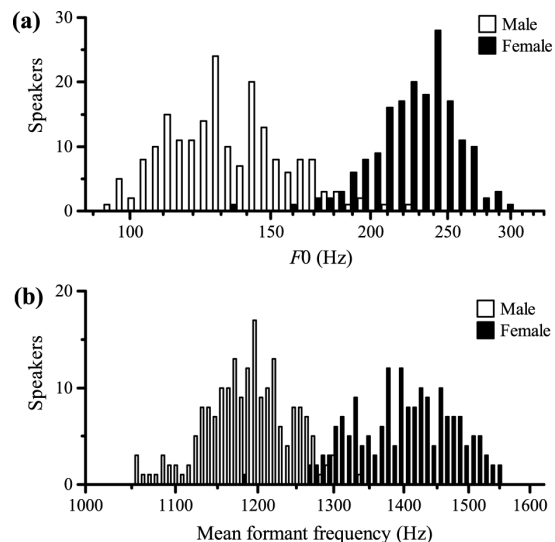


図2. 基本周波数(上段)と平均フォルマント周波数(下段)

この結果に基づき、知覚実験で用いる話者を選定した。図2の分布全体をカバーするため、基本周波数と平均フォルマント周波数ができる限り等間隔となる話者を選んだ(図3)。これらの話者が発話した母音3つをABX法で呈示し、Xの話者がAとBどちらと同一であると知覚されるか聴取実験を行った。

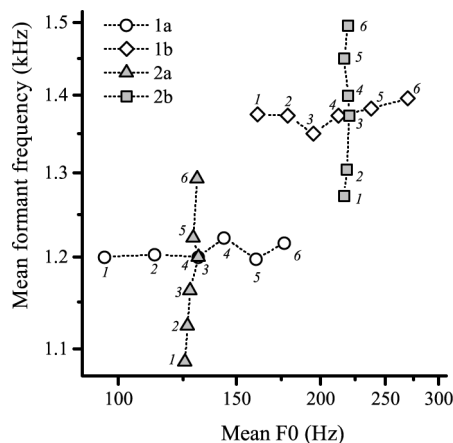


図3. 選定した話者

図4に結果の一部を示す。基本周波数が異なる話者により構成された刺激を用いた場合は、話者間の基本周波数の差が大きいほど正答率が高かった。これは基本周波数が話者知覚の重要な手掛かりであることを示唆する結果であった。またABXに用いる母音を全て同じ音韻とした場合と、ABとXの母音を変えた場合を比較すると、前者が後者より正答率が高かった。これはスペクトル形状が同一の方が、話者に関する情報を正確に知覚できることを示すものだが、異なる母音を用いた場合でも正答率はチャンスレベルより有意に高く、話者知覚が基本周波数だけで定まるものではないことを示す結果であった。この結果は、研究成果[2]で報告している。また、これらの音響分析と知覚実験に基づいて、高品質の歌声合成アルゴリズムを開発し、その成果を国際会議で発表した[1]。

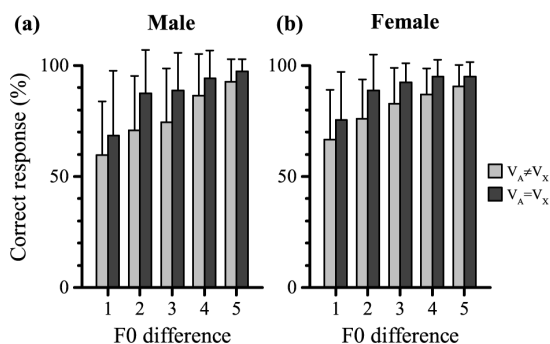


図4. 聴取実験の平均正答率

上記の話者を用いてフォルマント抑圧刺激を作成し、同様の実験を行った。その結果、フォルマントを抑圧することで個人性の知

覚は不正確になるが、音韻の種類と周波数帯域によっては影響が殆ど見られない場合もあることが分かった。だが、この傾向は聴取実験の被験者ごとに大きく異なっており、一般的な知覚特性として整理することが困難であった。

この被験者間のばらつきの要因のひとつは、自然発話音声から刺激を作成していることにあると考えられる。即ち、自然発話のパラメータは発話ごとに異なるため、これを刺激とした場合に調べたいパラメータ以外の要因が影響する可能性がある。また合成音声を刺激とした場合には、その結果から話者の個人性を議論することが難しいという問題がある。

そこで本研究課題の後半では、自然発話音声の基本周波数と声道フィルタ特性を高精度で分離できる新たな音声分析手法の開発に取り組んだ。この結果は、研究成果[3,4]で報告しており、基本アイデアの有効性については検証ができています。だが、現在までのところ聴取実験の刺激を作成できるレベルには達しておらず、今後さらに研究を進める必要がある。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 0件)

[学会発表](計 4件)

[1] Ito, M. "A singing voice synthesizer controlled by arm motions," 9th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IHMSP-2013-IS20-04., 2013.10.

[2] 相澤佑輔, 小笠原直也, 伊藤仁 (2014). "母音の個人性知覚研究," 平成26年東北地区若手研究者研究発表会 p.203-204.

[3] 伊藤仁, 伊藤 彰則(2014). "複素分析窓を用いた正弦波モデルに基づく混合音声の分析," 日本音響学会2013年春季研究発表会講演論文集 2-6-12.

[4] 伊藤仁, 伊藤彰則 (2013). "複素時間窓を利用した混合音声の分析手法の検討," 信学技報, vol. 113, no. 177, EA2013-53, pp. 1-6, 2013年8月

[図書](計 0件)

[産業財産権]  
出願状況(計 0件)

名称:

発明者：  
権利者：  
種類：  
番号：  
出願年月日：  
国内外の別：

取得状況（計 0 件）

名称：  
発明者：  
権利者：  
種類：  
番号：  
取得年月日：  
国内外の別：

〔その他〕  
ホームページ等

## 6. 研究組織

### (1) 研究代表者

伊藤 仁 (ITO, Masashi)  
東北工業大学・工学部・知能エレクトロニクス学科・准教授  
研究者番号： 00436164