

母音間に共通する個人性情報の知覚的要因*

北村達也, パーハム・モクタリ (ATR 人間情報科学研究所)

1 はじめに

我々は、基本周波数や音韻などの変化が音声の個人性情報に与える影響について調査している。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ため、母音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、我々は、人間が母音間に共通する個人性情報を用いて話者を識別できるか否かを調査するため、未知話者の単母音を用いた聴取実験を行い、1つの母音から抽出した個人性情報で他の母音の話者を識別できることを示した [1]。

母音のスペクトル包絡における個人性は約 1.7 kHz 以上の周波数帯域に顕著に現れると報告されている [2]。また、音声の立ち上がり、立ち下がり、ゆらぎなどの時間構造も個人性情報に寄与することが経験的に知られている。そこで、本研究では、これらの特徴量が母音間に共通する個人性情報として用いられるか否かを聴取実験により調査する。

2 実験条件

2.1 刺激音

日本語を母語とする成人男性 6 名が持続発声した母音 /a/、/e/、/o/ をマイクロホン (Sony ECM-77B) および IC レコーダ (Marantz PMD-670) を用い標準化周波数 16 kHz、量子化 16 bit で防音室にて収録した。話者にヘッドホンで 1 sec の白色雑音を呈示し、音声の継続時間長を合わせるよう指示した。得られた音声データから、母音毎に 3 トークンを選び、以下の刺激音を作成した。なお、音声データの基本周波数の平均値は 114.7 Hz、標準偏差は 15.6 Hz であった。

刺激音 1 振幅を話者間で正規化した音声。

刺激音 2 刺激音 1 の平均基本周波数を全音声データの平均値 (114.7 Hz) にした音声。基本周波数の時間変化パターンは保持。

刺激音 3 刺激音 2 を標準化周波数 5 kHz にダウンサンプリングした音声。

刺激音 4 刺激音 1 の基本周波数を一定 (114.7 Hz) にし、さらに母音中心の 10 フレームのスペクトルをランダムに並べたものので有声区間のスペクトルを置換した音声。

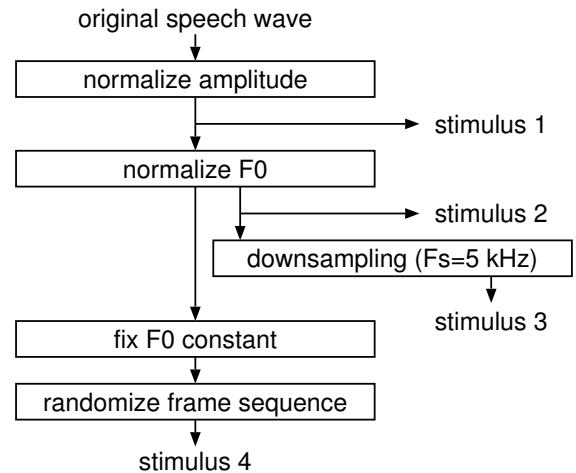


Fig. 1 各刺激音の作成処理過程。

これらの刺激音の作成における処理過程を図 1 に示す。刺激音 2, 3, 4 は STRAIGHT 分析合成系 [3] にて作成した。これらの刺激音はそれぞれ、話者識別における基本周波数の影響、高周波数領域の影響、時間構造の影響を調査するためのものである。

2.2 実験参加者

正常聴力を有する成人女性 5 名。

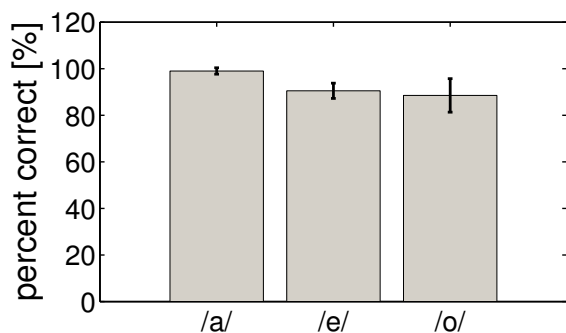
2.3 実験方法

ABX 法により行った。A, B, X の順で刺激音を呈示し、X の話者が A と B のどちらの話者であるかを強制選択させた。A と B は母音 /a/ を呈示し、X では母音 /a/、/e/、/o/ のいずれかを呈示した。本研究では A, B で呈示する母音を参照母音と呼ぶ。A の話者は任意に選択した 2 名のみ、B および X の話者は全話者とした。以上の条件で ABX の組合せを作成した。その際、A, B, X に現れる刺激音は 3 トークンの中からランダムに選択し、かつ X のトークンは A または B のトークンとは異なるようにした。

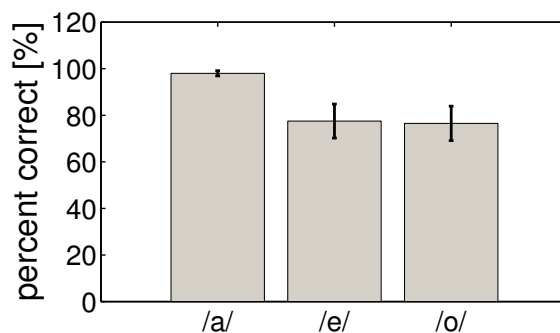
そして、4 種の刺激音が混在するようにランダムに並べ替え、実験参加者に呈示した。1 つの ABX の組合せは実験中に 1 回呈示される。なお、順序効果を打ち消すために BAX の組合せについても実験を行った。全試行数は 480 である。

実験参加者はヘッドホン (Sennheiser HDA200) により刺激音を両耳受聴し、パーソナルコンピュータの

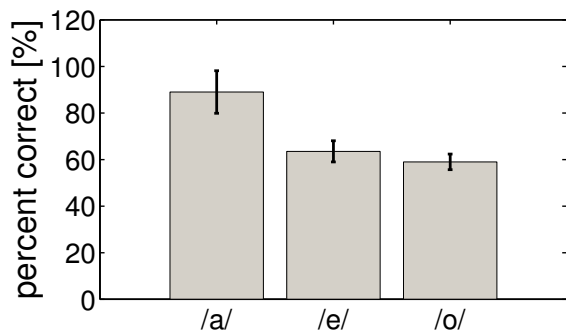
*Vowel independent factors contributing to perception of speaker characteristics. by KITAMURA Tatsuya and MOKHTARI Parham (ATR Human Information Science Laboratories)



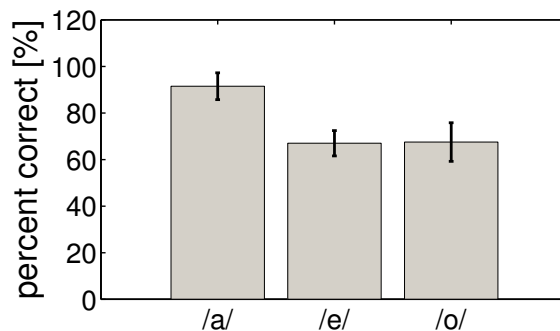
(a) 刺激音 1 の話者識別率



(b) 刺激音 2 の話者識別率



(c) 刺激音 3 の話者識別率



(d) 刺激音 4 の話者識別率

Fig. 2 各刺激音の平均話者識別率および標準偏差。

ディスプレイ上の A または B のボタンをマウスで選択することにより回答を行った。聴き直しは 3 回まで許した。各実験参加者は聴き易いレベルにて受聴した。

3 結果と考察

各刺激音の話者識別率を図 2 に示す。参照母音 (/a/) に関する刺激音間の話者識別率の差異は、刺激音 1 と 2 の間で-1.0 ポイント、刺激音 2 と 3 の間で-9.0 ポイント、刺激音 2 と 4 の間で-6.5 ポイントである。参照母音以外 (/e/ と /o/) に関する刺激音間の話者識別率の差異は、刺激音 1 と 2 の間で-13.0 および-12.0 ポイント、刺激音 2 と 3 の間で-17.5 および-14.0 ポイント、刺激音 2 と 4 の間で-10.5 および-9.0 ポイントである。

これらの結果から、未知話者の知覚には平均基本周波数が重要であり、高周波数帯域および時間構造にも母音間に共通する個人性情報が含まれるといえる。また、参照母音の話者識別に関して、平均基本周波数はほとんど影響を与えないが(刺激音 2)、高周波数帯域(刺激音 3) および時間構造(刺激音 4) は影響を与えることがわかる。これは刺激音 3 および 4 に加えた操作が話者識別を著しく困難にすることを示唆している。さらに、高周波数帯域と時間構造では、前

者のほうが話者識別への寄与が大きい。この結果は、Kitamura & Akagi[4] が既知話者の単母音を対象にして行った話者識別実験の結果と一致する。

4 おわりに

本研究では、母音間に共通する個人性情報を明らかにするために、平均基本周波数、高周波数帯域、時間構造を変化させた音声を用いて聴取実験を行った。その結果、これらはいずれも個人性情報に影響を与えることを示した。今後、音声の高周波数帯域および時間構造から個人性情報に対応する特徴量を抽出することを試みる。

謝辞 本研究は平成 17 年度総務省 SCOPE-R により実施したものである。

参考文献

- [1] Kitamura & Mokhtari, NCSP'06, 2006.
- [2] 北村, 赤木, 音響誌, 53(3), 185-191, 1997.
- [3] Kawahara, *et al.*, Speech Commun., 105, 3497-3508, 1999.
- [4] Kitamura & Akagi, J. Acoust. Soc. Jpn. (E), 16(5) 283-289, 1995.