

文音声の基本周波数のシフトが個人性知覚に及ぼす影響*

○北村達也, 肥田友紀子, 川元広樹 (甲南大)

1 はじめに

音声中的特徴量は発話のたびに変動するにもかかわらず, その個人性は安定して知覚される。従って, 音声の個人性に関する脳内表現においては, 確率的な変動や若干の差異を許容する形で音響的特徴量と話者とが対応付けられているはずである。この許容範囲を把握できれば, 知覚上の話者間距離を定義することができる。

出水田と赤木 [1] は, 基本周波数の変化の傾きとスペクトル傾斜を変化させると, ある程度までは個人性に変化がないが, ある点を超えると急に大きな変化が生じることを示した。これは, 音響的特徴量のある程度の変動に対しては感度が低いことを示している。そこで, 本研究では基本周波数のシフトを対象にして, 個人性知覚に同様の特性が見られるか否かを調査する。

2 実験 1

2.1 実験条件

原音声は, ATR 音声データベースセット C に含まれる関東出身男性話者 5 名 (M318, M509, M601, M603, M710) の文音声である。これらの話者の年齢は 21 歳から 37 歳で, 本研究の実験協力者にとって未知話者である。刺激音としては, 文 A 「あの坂を上れば海が見える」, 文 B 「飛ぶ自由を得ることは人類の夢だった」の 2 つを用いた。

上記の原音声の基本周波数全体を対数軸上でシフトさせたものを刺激音とした。刺激音の作成には STRAIGHT[2] を用いた。文全体の基本周波数を十二平均律に基づき -6 半音から +6 半音まで 2 半音ずつシフトさせ, 基本周波数に変化していないものも含めて 7 種類の刺激音を作成した。刺激音の標準化周波数は 20 kHz, 量子化ビット数は 16 bit である。また, 刺激音の最大振幅を話者間で正規化した。

実験には, 21 歳から 23 歳の聴覚に異常のない男性 16 名, 女性 4 名の計 20 名が参加した。

実験では, 原音声の基本周波数をもつ刺激音と上記の 7 種の刺激音のうちの 1 つから成る刺激対を実験協力者に提示した。2 つの刺激音の話者と文は同一である。順序効果を排除するために順序を入れ替えた刺激音対も提示し, 1 つの刺激音対は 2 度提示した。

実験は防音室にて行った。実験協力者は, 刺激対の話者が同一話者か異なる話者かを PC の画面上に表示されたボタンをクリックすることにより回答した。刺激音は 1 度だけ聞き直すことを許した。実験は刺激音の文ごとに 2 セットに分けて実施し, 間に数分間の休憩を入れた。

刺激音は, PC から出力された音声をヘッドフォンアンプ (Fostex HP-A3) にて D/A 変換し, 密閉型ヘッドフォン (Sennheiser HDA200) にて提示した。実験協力者は各自の聴きやすいレベルで聴取した。

2.2 結果

文 A, 文 B に関して同一話者と回答された割合を図 1 に示す。これらの結果は実験協力者間で平均した値である。エラーバーは標準偏差を表す。

基本周波数のシフトにより結果に差異があるか否かを評価するため有意水準 5 % の分散分析を行った。その結果, 文 A で $F(6, 133) = 68.38$, 文 B で $F(6, 133) = 52.38$ となり, いずれの文についても有意差があることがわかった。さらに, Tukey の HSD 検定により多重比較を行ったところ, 文 A, 文 B ともすべての群の間に有意差があった。

図 1 から, 基本周波数の ± 2 半音のシフトは個人性知覚への影響が小さく, それ以上のシフト量は個人性知覚への影響が大きいがわかる。基本周波数を ± 2 半音シフトさせても同一話者と回答された割合は 90 % 以上を保っている。それに対して, ± 4 半音のシフトにより同一話者と回答された割合はほぼチャンスレベルとなり, ± 6 半音のシフトにより 30 % 程度まで減少する。

また, 基本周波数の上昇と下降による影響はほぼ等しく, シフトによる影響は変化方向に対して対称性がある。

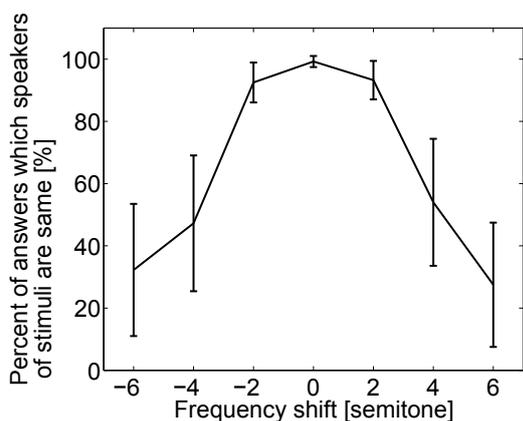
3 実験 2

実験 1 の結果が刺激対のピッチを弁別できないことに起因するの否かを調べるための実験を行った。

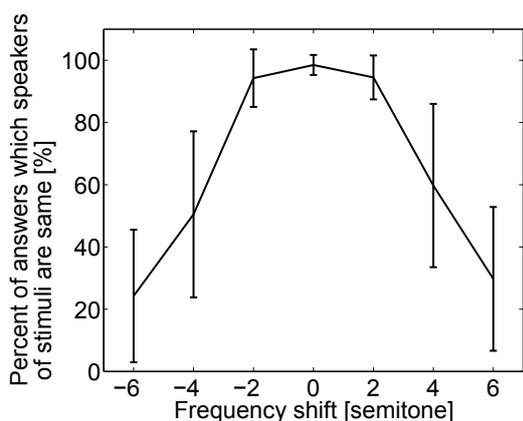
3.1 実験条件

実験協力者に刺激対のピッチが同じか否かを回答させた以外の方法は実験 1 と同様である。実験には, 21 歳から 25 歳の聴覚に異常のない男性 13 名, 女性 5 名の計 18 名が参加した。

*Effects of shift of pitch frequency of sentence speech on perception of speaker individualities. by KITA-MURA, Tatsuya, HIDA, Yukiko, KAWAMOTO, Hiroki (Konan University)



(a) Sentence A



(b) Sentence B

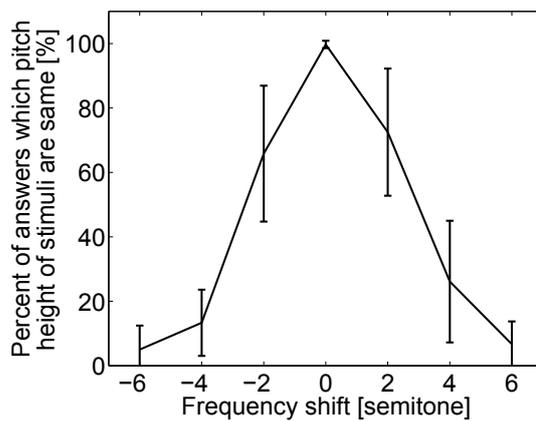
Fig. 1 Percent of answers which speakers of stimuli are same for the sentences A (a) and B (b).

3.2 結果

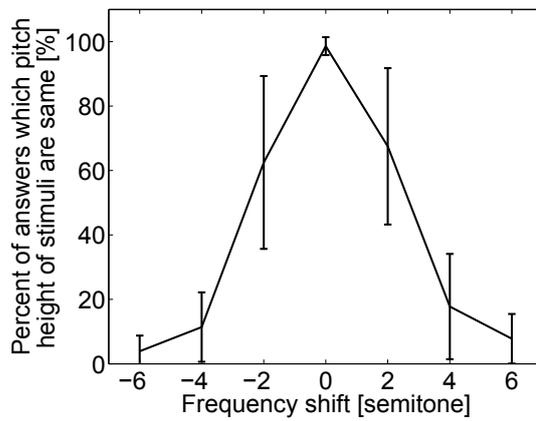
実験結果を図2に示す。これらの結果は実験協力者間で平均した値である。エラーバーは標準偏差を表す。この結果から、全ての実験協力者が基本周波数の ± 2 半音シフトによるピッチの違いを検出できたとは言えない。しかし、実験1で示された個人性知覚特性とは明らかに異なる特性を示している。

4 考察

実験1により本研究の条件下では文音声の基本周波数の ± 2 半音のシフトは個人性知覚に影響しないことが明らかになった。一方、 ± 4 半音以上のシフトは個人性知覚への影響が大きいことが示された。この ± 2 半音という範囲は比較対象の音声に対して相対的に決まることが明らかとなっている [3]。以上の結果は、個人性知覚において、一定範囲内の平均基本周波数の変動は許容され、それ以上は話者の特徴として知覚されることを示している。さらに、実験1の結果は文音声に対するピッチ知覚特性とは異なること



(a) Sentence A



(b) Sentence B

Fig. 2 Percent of answers which the pitch height of stimuli are same for the sentences A (a) and B (b).

が実験2の結果から明らかになった。これは、個人性知覚にはピッチ以外にもスペクトル包絡などが寄与していることによるものと考えられる。

5 おわりに

本研究では、文音声の平均基本周波数の ± 2 半音シフトは個人性知覚にほとんど影響を与えないことを明らかにした。今後、基本周波数以外の音響的特徴量に関しても変動に対する特性の調査を行う。

謝辞 本研究は平成24年度科研費基盤(B) (21300071), 平成25年度科研費基盤(A) (25240026) および私立大学等経常費補助金の支援を受けた。

参考文献

- [1] 出水田, 赤木, 音講論 (春), 423–426 (2012).
- [2] H. Kawahara, et al., *Speech Commun.*, 27, 187–207 (1999).
- [3] 北村, 川元, 信学技報 (EA), 113, 134, 33–37 (2013).