

個人性の類似性に着目した個人性知覚要因の検討*

☆出水田剛志 (甲南大・理工), 北村達也 (甲南大・知能情報)

1 はじめに

本研究は個人性の類似性に着目し、連続音声において個人性を表す特徴量を特定することを目的とする。類似している音声のペアを比較分析することによって、個人性知覚に必要な特徴量とそうでない特徴量とを区別できる。

連続音声の個人性知覚に関して、伊藤ら [1] は発話者を熟知した実験協力者による聴取実験を行い、基本周波数 (F0) と継続時間長よりもスペクトルが支配的であると報告している。橋本ら [2] はこれら 3 つの特徴量の変化が個人性知覚に及ぼす影響を比較した。また、大野ら [3] は F0 パターンが多くの個人性情報を含むという結果を示している。

本研究では、まず予備実験により類似性がある話者のペアを見つけ、その音声を分析して、類似した特徴量を抽出する。そして、その特徴量が個人性知覚に与える影響を聴取実験により調査する。

2 音声データ

話者を混同されることがある成人男性 2 名 (MT, TK) および成人男性 3 名 (HF, HH, HT) の文章音声を用いた。防音室にて ATR 音声データベース [4] 中の 5 文章を 5 回ずつ読み上げさせ、標準化周波数 48 kHz、量子化 16 bit で DAT に録音した。そしてパーソナルコンピュータ (PC) にて 16 kHz にダウンサンプリングした。聴取実験には 5 文章のうち「あらゆる現実を全て自分の方へねじ曲げたのだ」を用いた。

話者 TK と MT の間に類似性があるか確認するために予備実験を行った。10 名の実験協力者に上記 5 名の音声をランダムに 2 名ずつ呈示し、「似ていない」「あまり似ていない」「やや似ている」「似ている」の 4 段階で評価させた。その結果、TK-MT 間で高い類似性があると回答された他、HF-HT 間でも類似性があると判断された。実験後に類似性の判断材料を尋ねたところ、韻律との回答が多数であった。

3 基本周波数パターンの比較

実験協力者の上記の内観報告に基づき、音声データの F0 を求め比較した。F0 抽出には WaveSurfer を用いた。予備実験で類似性が認められたペア (TK-MT, HF-HT) の F0 を図 1 に示す。前者の F0 パターンが

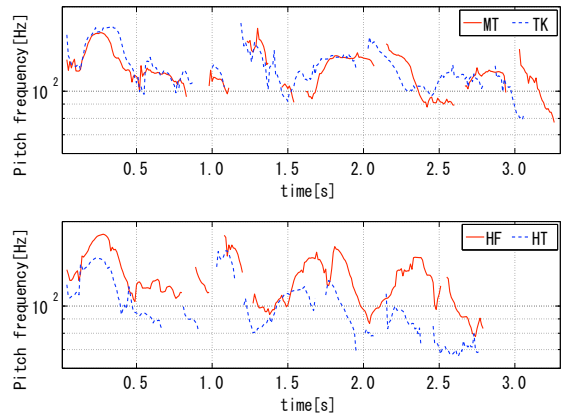


Fig. 1 (上) 話者 MT, TK と, (下) 話者 HF, HT の F0 パターン (周波数方向は対数軸)

類似していることから、F0 パターンが類似性に寄与する可能性がある。

4 聴取実験

上記の結果から、F0 パターンが個人性知覚に与える影響を調べるための聴取実験を行った。

4.1 刺激音

刺激音は話者 TK, HF, HH の音声データをもとに STRAIGHT 分析合成系 [5] により作成した。実験に用いた刺激音は以下の 3 種類である。

刺激音 1 STRAIGHT 分析合成音声。

刺激音 2 話者 S_1 の音声の継続時間長を話者 S_2 のもので置換した合成音声。

刺激音 3 刺激 2 の処理に加え、F0 も話者 S_2 のもので置換した合成音声。

F0 抽出誤りは、目視により修正した。時間軸の対応付けは、DP マッチングにより行った。その際、特徴量として STRAIGHT スペクトルから求めた、1 次から 30 次の STRAIGHT ケプストラムを用いた。この実験では、3 名の話者の音声を用いたため、話者 S_1 , S_2 の組み合わせは、6 通り (${}_3P_2$) である。

4.2 実験協力者

正常聴力を有する成人 20 名 (男性 10 名, 女性 10 名) である。このうち、男性 8 名が話者 TK に接したことがあり、その声を知っている。

*A study on perceptual factors for speaker individualities focusing on their similarity. by IZUMIDA, Tsuyoshi, KITAMURA, Tatsuya (Kounan Univrsity)

4.3 実験方法

XAB法により実験を行った。実験協力者に刺激音Xの話者が刺激音A, Bの話者のどちらかに似ているかを強制選択させた。刺激音Xは刺激音2または3であり、その合成に用いた話者の刺激音1を刺激音A, Bとして呈示した。順序効果を打ち消すためにXBAの順序でも呈示した。XABとXBAを各3回呈示し全試行回数は72回であった。X, A, Bの時間間隔は0.5sである。

実験協力者はヘッドホン(Sennheiser HDA200)により刺激音を両耳受聴し、PCのディスプレイ上のボタンをマウスで選択することにより回答した。PCから出力された刺激音はオーディオインターフェース(Roland UA-25)でD/A変換し呈示した。聴き直しは何度でも許した。実験協力者は各々の聴き易いレベルで受聴した。

4.4 実験結果と考察

実験協力者が刺激音2または3を聴いて、話者 S_2 に似ていると回答した割合を図2に示す。値は実験協力者間で平均した。刺激音2での話者 S_2 と回答された割合は10.8%、刺激音3では25.3%であった。このことから、継続時間長とF0パターンよりも他の特徴量(スペクトル)の方が個人性情報に寄与することがわかる。この結果は、先行研究[1][2]と一致する。また、刺激音2と3の間で分散分析($F(1,238;0.05)=3.881$)を行った結果、有意差が見られた($F(1,238)=14.353$)。この結果は、F0パターンに個人性情報が含まれるという報告[3]を支持するものである。

図2の結果のうち、話者HHを S_1 として合成した刺激音の結果を図3に示す。 S_2 を話者TK, HFとしたときの間に有意差が見られた($F(1,38;0.05)=4.098$, 刺激音2: $F(1,38)=5.439$, 刺激音3: $F(1,38)=12.148$)。従って、継続時間長とF0パターンの個人性情報への寄与は話者の組み合わせによって異なるといえる。

また、 S_2 を話者HFにした際の刺激音3の値は52.5%で、ほぼチャンスレベルであったことから、話者HFとHHの間ではスペクトルに類似性があることが示唆される。一方で、予備実験ではこの2名に類似性が認められなかった。以上のことから、スペクトルが類似しているような場合には継続時間長とF0パターンが話者を区別する主要な要因になることが示唆された。

5 おわりに

本研究では、連続音声の類似性に着目し、個人性情報となる特徴量の検討を行った。その結果、継続時間長とF0パターンよりもスペクトルが個人性情報

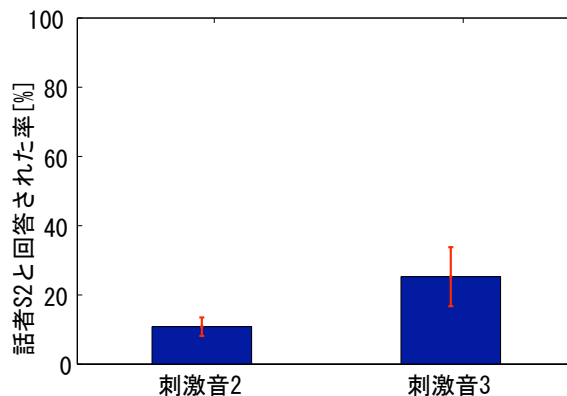


Fig. 2 刺激音Xの話者が S_2 に似ていると回答された割合

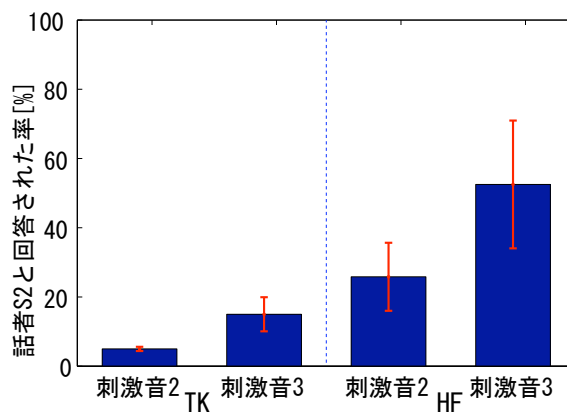


Fig. 3 話者HHを S_2 として合成した刺激音の結果。 S_1 を話者TKとした場合(左), HFとした場合(右)

に寄与することが示された。しかし、話者の組み合わせによっては、継続時間長とF0パターンが個人性情報に大きく寄与することが分かった。

謝辞 STRAIGHTの使用をご快諾いただいた、和歌山大学システム工学部の河原英紀先生に感謝いたします。本研究の一部は、2009年度総務省SCOPE(071705001)、2009年度科研費基盤(B)(21300071)の援助を受けた。

参考文献

- [1] 伊藤, 齋藤, “音声の音響的特徴パラメータが個人性情報に及ぼす影響,” 信学論, J65-A(1), 101-108 (1982).
- [2] 橋本ら, “音声の個人性情報に及ぼす音響的特徴の定量的分析,” 音響誌, 54(3), 169-178 (1998).
- [3] 大野, 赤木, “文音声中の基本周波数パターンに含まれる個人性情報の検討,” 信学技報, SP97-128 (1998).
- [4] 阿部ら, “研究用日本音声データベース利用解説書,” Tech. Rep. ATR, TR-I-0166 (1990).
- [5] 河原, “STRAIGHTとGUIで体験する音声知覚の物理的基盤,” 聴覚研資, 39(6), 407-412 (2009).