

ATR 音声データベースの文音声の話者間類似度：実験結果の検証*

☆中間隆正 △大村宙 北村達也 (甲南大・知能情報)

1 はじめに

個人性の知覚要因に関する研究はこれまで数多く行われてきたが、多くの場合は研究者もしくは研究組織独自の音声データが用いられている。そのため、得られた結果はその話者セットに多少なりとも依存することが否定できない。この問題を解消するには、共通の音声データを用いて研究し、その結果を相互に比較検証できるようにしていく必要がある。

一般向けに販売、公開されている ATR 音声データベースセット C には、男女各 140 名による単語、数字、および音素バランス文の音声が含まれており、それらには音声セグメントラベルが付与されている。この規模の話者数が収録された日本語音声データベースはほとんどなく、現段階で個人性研究の共通基盤とするには十分と考えられる。

そこで、我々はこの音声データベースを対象として、聴取実験に基づき個人性の類似度評価を行う研究を行っている。これまでに男性話者 20 名を対象とした調査 [1]、女性話者 20 名を対象とした調査 [2] を行ってきた。本研究では、別の実験参加者群でこれらの研究と同じ実験を行い、同様の結果が得られるのかを調査した。

2 方法

2.1 刺激音

ATR 音声データベースのセット C から、男性話者として 19 歳から 39 歳の関東 (東京、神奈川、千葉、埼玉) 出身の 20 名 (M109, M113, M211, M214, M318, M409, M504, M508, M509, M510, M517, M519, M520, M601, M603, M614, M705, M710, M714, M718)、女性話者として 20 歳から 29 歳の関東 (東京および神奈川) 出身の 20 名 (F213, F214, F306, F308, F406, F407, F409, F418, F507, F509, F605, F606, F609, F611, F614, F702, F704, F709, F714, F720) を選択した。対象とした文は、「冷房では冷えすぎが問題になる」である。標本化周波数 20 kHz、量子化ビット数 16 bit で、話者間で振幅を正規化した。

2.2 実験参加者

実験には、日本語を母語とする聴覚に異常のない者が参加した。男声を対象とした実験には 18 歳から 24 歳の男性 20 名、女性 6 名の計 26 名が参加し、女声を対象とした実験には 18 歳から 24 歳の男性 18 名、女性 6 名の計 24 名が参加した。これらの実験参加者には文献 [1] および [2] の実験に参加した者もいるが、同性の話者の実験には参加させないようにした。すなわち、本研究とこれらの文献の実験における参加者は別人である。

2.3 実験手続き

2 つの刺激音を 1 組として、話者 20 名のすべての組み合わせで刺激音を提示した。刺激音間の無音区間は 0.3 s である。順序効果を排除するため提示順を入れ替えた刺激対も提示したので、刺激対の数は 400 である。この刺激対を各 1 回提示した。刺激対はランダムに提示した。100 試行を 1 セットとし、実験参加者 1 名に対して計 4 セットを実施した。

実験参加者は、各刺激対に対し、5 段階 (似ていない、あまり似ていない、やや似ている、似ている、同一人物) [3] で類似性を評定した。聴き直しは 1 度だけ許した。

実験は防音室にて実施した。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。実験参加者は各自の聴きやすいレベルで聴取した。

2.4 分析方法

まず、2 つの刺激音の話者が同じ刺激対に対して、「同一人物」と評定した回数が 60 を超える実験参加者、「似ていない」と評定した回数が 250 回を超える実験参加者のデータは除外した。その結果、最終的に男声を対象とした実験では 20 名 (男性 16 名、女性 4 名)、女声を対象とした実験では 20 名 (男性 16 名、女性 4 名) の実験結果を分析対象とした。

そして、上記の 5 段階評定に 1 から 5 の数字を割り当て (1: 似ていない, 2: あまり似ていない,

* Perceptual speaker similarity of sentence speech in ATR speech database: Verification of results. by NAKAMA, Takamasa, OHMURA, Hiroki, KITAMURA, Tatsuya (Konan University)

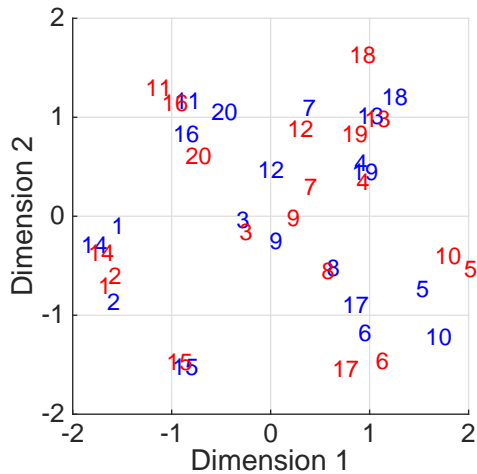


Fig. 1 Spatial layout of perceptual speaker similarity for 20 male speakers in ATR speech database set C. The results of the present study and Kawamoto and Kitamura [1] are indicated by blue and red characters, respectively.

3: やや似ている, 4: 似ている, 5: 同一人物), 実験参加者間で平均した評定値を求めた. この評定値はすべての話者の組み合わせに対して得られるので, 話者間類似度は距離行列として表現される. 得られた距離行列に関して, 男声に関する結果は文献 [1], 女声に関する結果は文献 [2] で得られた話者間類似度との差異の平均値を求めた. この計算の際, 対称となる要素は除外した.

また, 本研究で得られた距離行列を非計量多次元尺度構成法にて分析した. この分析には R の isoMDS [4] を用いた.

3 結果と考察

男声に関して本研究と文献 [1] にて得られた話者間類似度のユークリッド距離の平均値は 0.2, 女声に関して本研究と文献 [2] にて得られた話者間類似度の差異の平均値は 0.3 であった. この結果は, 異なる実験参加者群により得られた話者間類似度がおおむね類似していることを示している.

また, 男声, 女声の結果に対する非計量多次元尺度構成法の分析結果をそれぞれ図 1, 2 に示す. 図中の数字は 2.1 節の話者の掲載順に対応している. これらの図は第 1, 第 2 次元に関する話者の布置で, 青字で示されている結果が本研究の実験結果であり, 赤字で示されているのが文献 [1],

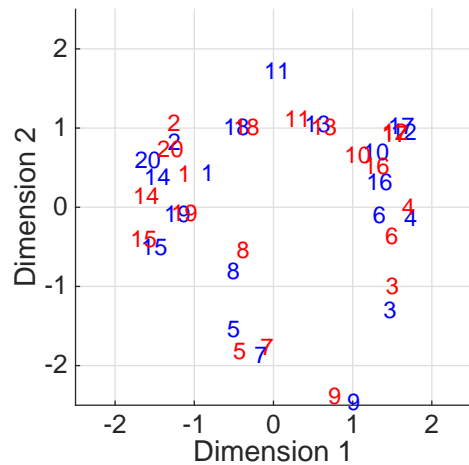


Fig. 2 Spatial layout of perceptual speaker similarity for 20 male speakers in ATR speech database set C. The results of the present study and Ohmura and Kitamura [2] are indicated by blue and red characters, respectively.

[2] の結果である. すべての分析において, ストレス値が 5 % 以下となる次元を採用した. ストレス値 5 % の適合度は「良い適合」といわれている [4]. これらの図において, 話者の相対的位置関係はおおむね一致している. つまり, 異なる実験参加者群により得られた結果が同様の傾向を示した.

4 おわりに

本研究では, ATR 音声データベースセット C 内の男性話者 20 名, 女性話者 20 名の文音声を対象として行った話者間類似度計測実験 [1][2] と同様の実験を再度実施し, 得られる結果に安定性があることを確認した. 今後の課題は, 異なる文でも同様の傾向が得られるかを調査することである.

謝辞 本研究の一部は平成 26 年度科研費 (25280026, 25280066) の支援により行われた.

参考文献

- [1] 川元, 北村, 信学技報 (SP), 112(450), 33-34 (2013).
- [2] 大村, 北村, 音講論 (春), 659-660 (2014).
- [3] 出水田, 北陸先端大 修士論文 (2012).
- [4] 中村, 多次元データ解析法, 共立出版 (2009).