ATR 音声データベースの文音声の話者間類似度*

☆大村 宙, 北村達也 (甲南大)

1 はじめに

個人性の知覚要因に関する研究は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 [1].

このような折，高度言語情報融合フォーラム (ALAGIN) から ATR 音声データベースが公開された．この中にはセット C と呼ばれる男女各 120 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．川元, 北村 [2] では上記データベースの男性話者 20 名を対象として個人性類似度を調査した．引き続き，本研究では女性話者 20 名に関して同様の実験を行う．

2 方法

2.1 刺激音

ATR 音声データベースのセット C から，20 歳より 29 歳の関東 (東京および神奈川) 出身の女性話者 20 名 (213，214，306，308，406，407，409，418，507，509，605，606，609，611，614，702，704，709，714，720) 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

日本語を母語とする 19 歳から 25 歳の聴覚に異常のない男性 26 名，女性 16 名の計 42 名が参加した．

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，実験協力者 1 名に対して計 4 セットを実施した．

実験協力者は，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．実験協力者は各自の聴きやすいレベルで聴取した．

2.4 分析方法

本研究では，上記の 5 段階評定に 1 から 5 の数字を割り当て，実験協力者間で平均した評定値を非計量多次元尺度構成法にて分析した．それぞれ，1：似ていない，2：あまり似ていない，3：やや似ている，4：似ている，5：同一人物，とした．非計量多次元尺度構成法の分析は R の isoMDS [4] を用いた．

なお，2 つの刺激音の話者が同じ刺激対に対して,「同一人物」と評定した回数が他の実験協力者と比較して著しく少ない実験協力者のデータは除外した．本研究では最終的に 40 名 (男性 25 名，女性 15 名) の実験結果を分析対象とした．

3 結果

各話者対の評定値の標準偏差を計算したところ，全 400 対のうち 310 対で 1 未満であった．男性話者で同様の実験を行った際 [2] には，全 400 対中 399 対で 1 未満であったので，今回の実験では実験参加者間の回答のばらつきが若干大きくなったといえる．なお，評定値の標準偏差の最大値は 1.32 であった．

---

*Perceptual speaker similarity of sentence speech of ATR speech database. by OHMURA, Hiroki, KITAMURA, Tatsuya (Konan Univ.)

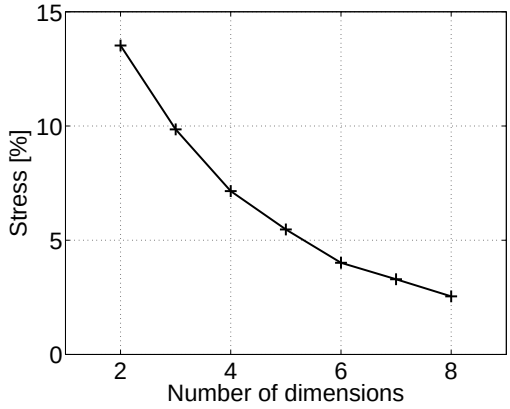

Fig. 1　Stress

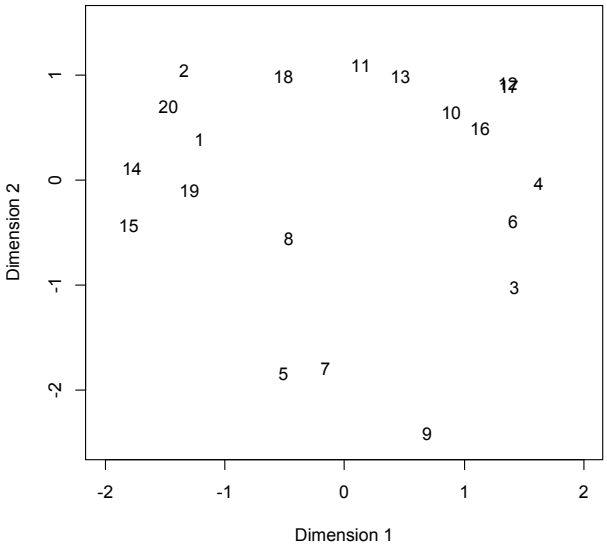

Fig. 2　Spatial layout of perceptual speaker similarity for 20 female speakers in ATR speech database set C.

非計量多次元尺度構成法におけるストレス値を図 1 に示す．ストレス値とは，n 次元におけるデータ間の心理距離の適合度を示すものであり，その値が小さいほど適合度は高くなる．本研究では，ストレス値が 5 %以下となる 6 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].

非計量多次元尺度構成法により求めた個人性類似度の空間配置を図 2 に示す．図中の数字と話者は表 1 の通り対応している．この空間配置においては，データ間の距離が近いほど話者の知覚的類似度が高いと評価される．

第 1 次元の値と音声の平均基本周波数 (F0) の間の相関を図 3 に示す．文音声全体の F0 を STRAIGHT [5] により求め，有声区間で平均した．分析の結果，$R^2 = 0.85$ と相関が高く，第 1 次元は平均 F0 と対応すると考えられる．

Table 1　Speaker index in Fig. 2 and speaker ID.

No.	ID	No.	ID	No.	ID
1	213	8	418	15	614
2	214	9	507	16	702
3	306	10	509	17	704
4	308	11	605	18	709
5	406	12	606	19	714
6	407	13	609	20	720
7	409	14	611		

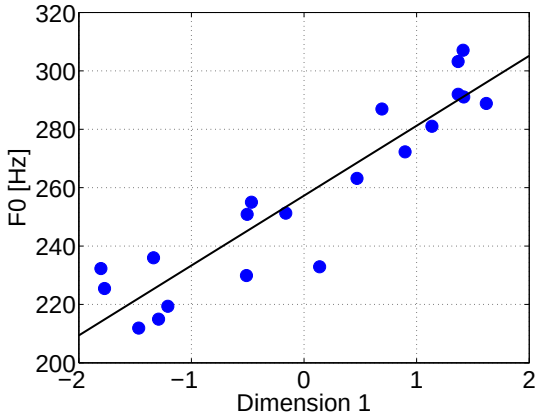

Fig. 3　Correlation between dimension 1 and F0. $R^2 = 0.85$.

4　おわりに

ATR 音声データベースセット C 内の関東出身女性話者 20 名の文音声を対象にして，個人性類似度の空間配置を行った．この結果は web ページにて公開し，音声の個人性に関する研究に役立てていただく予定である．

謝辞　本研究の一部は平成 24 年度科学研究費 (25280066, 25240026) にて実施された．ATR 音声データベース C セットの開発および公開に関わられた皆様に感謝いたします．

参考文献

[1] 北村, 出水田, 橘, 音講論 (秋), 253–256, 2011.
[2] 川元, 北村, 信学技報 (SP), 112, 450, 33–34, 2013.
[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.
[5] 河原, 聴覚研資, 39, 6, 407–412, 2009.