

## 文音声における知覚的話し者間類似度の計測：前報の追試\*

☆西河史也, 北村達也 (甲南大)

## 1 はじめに

音声には個人性が含まれており、我々は音声からその情報を得て話し者を識別することができる。この処理過程とそれに用いられる特徴量の説明は、音声科学の重要な課題の一つである。

北村ら [1] は ATR 音声データベースセット C の男女各 20 名の文音声の知覚的話し者間類似度を計測した。彼らの研究では、1 つの文を対象に実験が行われ、得られた話し者間類似度が他の文に対しても当てはまるのかは調査されていない。そこで、本研究では前報 [1] と同じ話し者セットが発話した異なる文に対して同様の実験を行い、話し者間類似度を前報と比較することを目的とする。

## 2 方法

## 2.1 刺激音

ATR 音声データベースセット C のうち、前報と同じ男性話し者 20 名で、前報とは異なる文「疲労回復には十分な睡眠が必要だ」の音声を用いた。この文は、短すぎず、長すぎないものとして選択した。文が短すぎる場合には韻律に個人差が現れにくくなる可能性があり、長すぎる場合には個人性の対比較において先に提示された刺激音の個人性を記憶しておく負担が大きくなるためである。音声データの標本化周波数は 20kHz、量子化ビット数は 16bit である。話し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聴覚に異常のない大学生の男女 (男性 36 名, 女性 12 名) 計 48 名が参加した。

## 2.3 聴取実験の方法

2 話し者の刺激音を 1 対として、話し者 20 名すべての組み合わせで提示した (同一話し者が対になることも含まれる)。刺激音間の無音区間の長さは 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音も提示した。刺激対の数は 400 である。この刺激音を各 1 回ランダムに提示し

た。100 試行を 1 セットとし、各実験参加者に計 4 セット実施した。

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。実験に先立ち、実験参加者は 5 つの刺激対を用いてリハーサルを行い、実験の手続きを確認すると共に、受聴レベルを調整した。実験参加者は、各刺激音に対し 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) で類似性を評定した。聞き直しは 1 度だけ許した。以上の手続きは、甲南大学における研究倫理審査にて承認されている。

## 2.4 分析方法

前報と同様に、以下の 2 条件を満たす実験参加者のデータは信頼性が低いと判断し分析対象から除外した。

1. 「同一人物」と評定した回数が 400 試行中 60 回を超える者 (実際に同一話し者の刺激対が提示されるのは 40 回)
2. 「似ていない」と評定した回数が 400 試行中 250 回を超える者

実験参加者の中には、異なる話し者の音声の類似性を段階的に判定することを困難を感じ、ほとんどの刺激対に対して「似ていない」と判定する者がいた。上記 2 はこのようなケースである。結果的に実験では、8 名のデータを除外し 40 名を分析対象とした。

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, 3: やや似ている, 4: 似ている, 5: 同一人物), 実験参加者で平均した類似度を求めた。そして、得られた話し者類似度と前報との相関係数及び差の平均値を求めた。前報で求めた話し者  $i$  と話し者  $j$  の話し者間類似度を  $s_1(i, j)$ , 本研究で求めた話し者  $i$  と話し者  $j$  の話し者間類似度を  $s_2(i, j)$  とすると ( $i =$

\* Measurement of perceptual speaker similarity for sentence speech :A supplementary examination.  
by NISHIKAWA, Fumiya, KITAMURA, Tatsuya (Konan University)

1, ..., 20,  $j = 1, \dots, 20$ ), 二つの話者間類似度との差の平均値  $\bar{d}$  は以下の式で求めた.

$$\bar{d} = \frac{1}{210} \sum_{i=1}^{20} \sum_{j=i}^{20} |s_1(i, j) - s_2(i, j)|$$

さらに, 得られた話者間類似度を非計量 MDS にて分析した. この分析には統計ソフトウェア R の isoMDS を用いた.

### 3 結果

本研究と前報にて得られた話者間類似度間の相関は 0.9 であった (同一人物の話者間類似度を除いた相関は 0.6). この結果から前報と同じ男性 20 名が発話した異なる文音声には相関があることが分かる. また,  $\bar{d}$  の値は 0.37 であった.

話者間類似度から非計量 MDS により求めた第 1-第 2 次元平面上的男性話者の布置を Fig.1 に, 前報の結果を Fig.2 に示す. 図中の数字は話者 20 名の番号に対応している. これらの図の作成においては, ストレス値が 5% を下回る 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 [2]. 話者 11 と話者 17 が対角に布置される, 話者 4, 5 が似た座標に布置されるなど共通点が見られたが, 異なる点も見られた.

### 4 考察

前報 [1] では「冷房では冷えすぎが問題になる」という文に対して話者間類似度を求めた. その際, 異なる実験参加者群を対象に聴取実験を 2 度行った. そして, 2 群間の相関係数は 0.97,  $\bar{d}$  は 0.21 という結果が得られ, 非計量 MDS による話者の布置の類似度も高かった.

一方, 本研究では, 同じ話者セットの異なる文を用いて同様の実験を行い, 相関係数は上記の値より小さく,  $\bar{d}$  は大きかった. また, 非計量 MDS による話者の布置の類似度も高いとはいえなかった. これらの結果は, 話者の知覚的話者間類似度は文により変化しうることを示している. ただし, この結果は本研究及び前報で対象とした話者に依存したものであり, 異なる話者を対象に同様の調査を行えば異なる結果が得られる可能性がある.

### 5 おわりに

本研究では, 前報と同一の話者セットによる異なる文に対して聴取実験により話者間類似度を

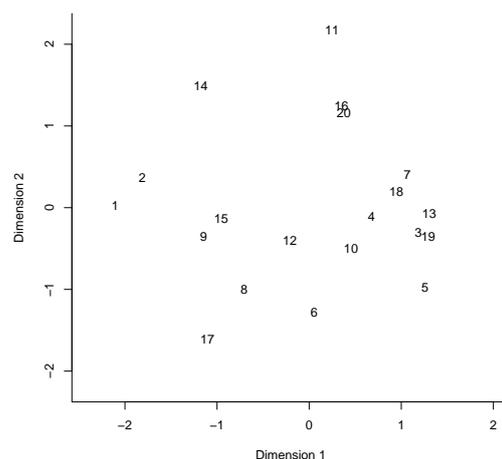


Fig. 1 Similarity based speaker map.

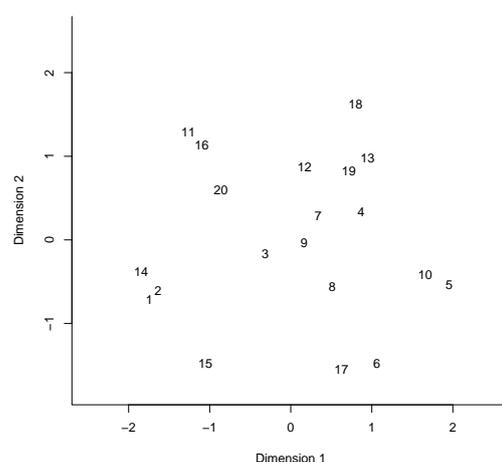


Fig. 2 Similarity based speaker map measured by Kitamura et al. [1]

求めた. その結果, 文が異なると話者間類似度は必ずしも一致しないことを示した.

謝辞 本研究の一部は 2018 年度科研費 (No. 17H02339) の支援を受けた.

### 参考文献

- [1] 北村, 中間, 大村, 川元, 音響誌, 71(10), 516–525, 2015.
- [2] 中村, 多次元データ解析法 (共立出版), 2009.