The impact of duration of speech sample on listeners’ judgements of voice similarity

Kirsty McDougall, Alice Paver and Francis Nolan

Theoretical and Applied Linguistics Section, University of Cambridge

1. Background

- Listeners perceive some speakers as sounding more similar than others, yet the phonetic underpinnings of perceived voice similarity (PVS) are not well understood.
- Improved understanding of PVS will aid foil selection for voice identification paradigms.
- Which phonetic features contribute to voice similarity judgements and does this vary across accent?

2. Previous Study: McDougall (in press) [1]

- Listeners (N = 120) judged the similarity of 120 pairs of voices on a 9-point Likert scale.
- 6 groups of 15 speakers (male, 18-30 years):
  - 1 SSBE
  - 1 York English
  - 1 York vs SSBE[1]
  - 1 Bradford English
  - 1 Wakefield English
  - 1 Bradford English

3. Present Study: Follow-up Experiment

- Experiment repeated on SSBE group 2 with 10s stimuli, to compare against 3s results.
- Longer task so 120 stimuli pairs divided into 4 blocks (80 listeners, 20 per block).
- Listeners judgements normalized and averaged to form combined sets.
- MDS applied to new 3s & 10s matrices.

4. Acoustic Analysis

- Mean f0 (semitones)
- LTF F1-F4 in ERB, Praat tracker.
- F2-F1
- F2’ [5]
- Articulation Rate (AR): number of phonetic syllables per second [6]

5. Results

6. Discussion

- f0: key role for PVS in 3s & 10s.
- LTF relatively consistent; cor.
- Patterns in higher dimensions.
- AR: no sig. cor. for 3s or 10s.
- PVS quite variable; new analysis.
- Does not highlight importance of different phonetic features when longer (10s) stimuli used.
- Short samples (3s) seem to be sufficient for listeners to get hold of speakers’ characteristics, at least within the same accent.
- Further investigation needed re different accents.
- Proceed with 3s samples for MDS ‘fairness’ check in voice parade construction.

7. References


