Perceptual similarity of identical twins across different L1 listeners: the importance of voice quality in Forensic Phonetics

Eugenia San Segundo¹, Paul Foulkes², Peter French¹, Vincent Hughes¹ and Olaf Köster³
¹Dept. Language & Linguistic Science, University of York, UK ²JP French Associates, York, UK ³BKA (Federal Criminal Policy Office, Germany)

BACKGROUND & OBJECTIVE
- Round robin test recently conducted by the Bundeskriminalamt (BKA) to evaluate the performance of experts in speaker identification tasks:
  - Auditory evaluation since the technical characteristics of the recordings prevented most experts from carrying out specific acoustic analyses.
  - Comparisons for a pair of female German twins.
  - Widely assumed that twins’ voices are similar, and thus recognition of voices is especially difficult (e.g. [1,2]).
- Results:
  - Lack of native knowledge of the language spoken by the twins was not a disadvantage for telling the twins apart.
  - Informal feedback from participants suggested that voice quality (VQ) –approached holistically rather than analytically-- was the main cue used by non-native listeners to distinguish the twins.
- Limitations:
  - Limited and idiosyncratic data set (the twins were of advanced age and had lived in different dialectal regions)

WHAT IDENTICAL TWINS
- Monzygotic (MZ) twins occur when a single ovum is fertilized by a sperm cell to form one zygote, which then divides in 2.
- MZ twin pairs share all their genes in common. The fact that they usually share environmental (educational & social) influences makes them examples of extreme similarity, also in voice. Both organic (vocal tract anatomy) and learned (phonetic choices) variation –which usually explain between-speaker variation— are minimized in MZ twin pairs.
- Investigations with twins (identical vocal tract) may then prove useful to assess VQ closeness in very similar-sounding speakers.

MATERIALS & METHOD
Subjects: 10 speakers selected from the corpus collected in [6], i.e. three sets of twin pairs (natives of Montreal, native English speakers). Original corpus contains 54 speakers (aged 18-54), so for the selection of the 10 speakers of this experiment some criteria were established:
- Similar age (mean: 21; sd: 3.7)
- Similar mean f0 (mean: 113 Hz; sd: 13 Hz)
- Similar Euclidean distance (ED) between each speaker and his twin (as calculated in [7]), in order to select only the most similar-sounding twin pairs. EDs are based on the perceptual assessment of their VQ using a simplified version of the Vocal Profile Analysis (VPA) scheme [3]. The mean ED between twin pairs was 0.62, measured in Similarity Matching Coefficients (SMC), a typical distance measure for categorical data where the number of matches for each variable is divided by the number of variables.

RESULTS
MDS analyses were carried out using similarity scores to construct a perceptual map of the speakers. The scatter plot (right) shows the relative magnitude of the sorted Eigenvalues. 7 dimensions necessary to accurately reproduce between-speaker distances in the perceptual space, but MDS results typically visualized using only the first 2 or 3 dimensions.

DCT_JCT AGF_SGF ARI_RE Asm_RSM AMG_EMG
ENGLISH 0.219 0.241 0.349 0.435 0.405
SPANISH 0.343 0.381 0.365 0.349 0.667

All speakers are closer in the perceptual space. Does this imply that knowledge of the linguistic content makes the task more difficult? Distinct effect of the message?
Better detection of very similar speakers (i.e. twins). Smaller distances between these in comparison with English listeners. Note also the different magnitude of the plots.

Most similar twin pair: AGF_SGF and ARI_RE. Most different twin pair: AMG_EMG.

DISCUSSION
- Eigen-decomposition: 7 main dimensions explain similarity decisions by listeners (both English and Spanish).
  - Voice is highly multidimensional; reducing the perceptual space to 2D or 3D may be misleading.
  - Similarity of the relative magnitude of the sorted Eigenvalues suggests that similar perceptual strategies operate for both listener groups.
  - Almost the same ranking of twin similarity for both listener groups can indicate the same cue prominence, i.e. regardless of familiarity with the language spoken or understanding of the linguistic content, both groups seem to rate the same twin pairs as most similar and the same twin pairs are most dissimilar.
  - Exception: AGF-SGF most similar for Spaniards (VQ analysis: tense VT & advanced tongue tip) while DCT-JCT most similar for English (VQ analysis: harshness & raised larynx)

DIFFERENT PERCEPTUAL SCALE OF VQ SETTINGS?
- Equivalent reaction times point to similar listening strategies learned to them.
- Other future work: 1) Divide listeners in musical and non-musical training; (2) Test with listeners of other languages (e.g. Germans - would these obtain similar results as English?)

CONCLUSIONS

REFERENCES

ACKNOWLEDGEMENTS
Thanks to Andrew MacFarlane and Duncan Robertson for references about measuring reaction time. A special thanks to Juana Gil for letting ESS run the test with Spanish listeners at the Phonetics Lab (CSIC, Madrid).