Prosody can help distinguish identical twins: implications for forensic speaker comparison

Eugenia San Segundo[1], Lei He[2] and Volker Dellwo[2]

[1] Department of Language & Linguistic Science, University of York, UK

[2] Institute of Computational Linguistics, University of Zurich, Switzerland

BACKGROUND & OBJECTIVE

• **Background:** Voice similarity of identical twins attracts the attention of researchers (also in forensic studies):

✓ Why? widely assumed that twins' voices are very similar → especially difficult recognition (e.g. [1])

✓ However: hardy comparable results across studies

» because of different number of speakers, speaking style and forensic comparison methods

» so, how to assess the relative importance of different systems or the value of a set of acoustic features over others?

– Some exceptions: Twin Corpus [2]

» so far three main studies using this corpus & task 5

» same twin pairs and speaking style (see *Materials & Method*)

» but still different comparison methods/system output

Twin Pair	Methodological approach (output) MFCCs (scores) [a]	Glottal features (LLRs) [b]	VPA (Euclidean distances) [c]
01	2.59	-0.1	0.8
02	2.65	-1.0	0.7
03	3.45	5.8	0.8
04	3.79	0	0.4
05	3.53	0.2	0.5
06	3.20	0.6	0.7
07	2.31	12.1	0.6
08	3.54	9.9	0.8
09	2.66	12.6	0.5
10	0.64	2.9	0.6
11	4.93	-1.5	1
12	1.34	-14.6	0.3

Table 1. Previous investigations using Task 5 of the Twin Corpus [2]:
[a] Batvox 4.1, Agnitio Voice Biometrics [3], [b] BioMet®Soft [4]
[c] Simplified Vocal Profile Analysis [5]

Different misidentifications produced by each system!
(gray-shaded cells in Table 1)
→ complementary info provided by each system?
→ need for more hybrid approaches in FSC?

• **Objective:** new approach based on prosodic parameters:

a) rhythmic metrics related to the variability and proportion of duration between consonant and vocalic segments

b) syllabic measures related to intensity differences between consecutive syllables

Why?

– Previous studies show that these parameters play an important role in between-speaker differences [6, 7].

– They cover suprasegmental aspects: independent of acoustic features related to vocal tract/source.

Will these features be useful to tell twins apart when other systems failed?

MATERIALS & METHOD

• **Subjects:** 24 speakers from the Twin Corpus collected by ESS [2]
– 12 monozygotic (MZ) twin pairs
– male; aged 20-36
– native speakers of Standard Peninsular Spanish

• **Task:**
– participant - researcher spontaneous conversations
– over the phone (~10min)

• **Speech material:**
– ~ 2 mins net speech * 24 speakers
– Inter-Pause (IP) stretches per speaker: 31 (mean); 6 (SD)

• **Corpus annotation:**
– Manual transcription
– Semi-automatic alignment and segmentation at the phonetic and syllable level using *EasyAlign*

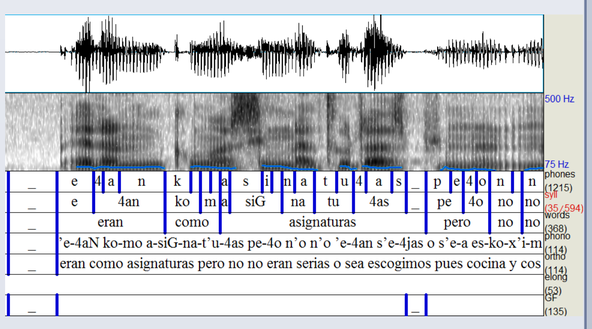

Figure 1. Example of phonetic alignment
CV tier created from the segment tier (phones) at a later stage

Acoustic analysis:

• **Rhythmic measures**:
- From CV intervals tier: 17 measures
e.g. *ΔC(ln), n-PVI-V, n-PVI-C, %V, varcoV*
- From syllable tier: 9 measures
e.g. *ΔSyl, varcoSyl, rPVISyl, nPVISyl*

• **Intensity measures**:
- From syllable tier:
varcoM, varcoP, stdevP, stdevM

Statistical analysis:

1) Principal Component Analysis (PCA)
- In order to reduce the number of variables
- Rotation method: Varimax with Kaiser normalization

2) Dissimilarity measures and significance tests
- Following method described in [8] for twin speaker comparisons
- Both analyses based on only 8 measures after PCA:

a) Euclidean distances
- based on the 8 prosodic measures together
- z-score normalized & rescaled to 0-1 range
→ lower values indicate 'more similar'
→ higher values indicate 'more different'

b) Independent t-tests
- based on the separate prosodic measures
- two-tailed tests with Bonferroni correction

RESULTS

• **PCA analysis:**
- 8 components extracted
- 1st variable selected per component (highest loadings):

$$\Delta V \cdot varcoC \cdot nC \cdot meanCLn \cdot \Delta SylLn \cdot varcoP \cdot nPVI\text{-}C \cdot nPVI\text{-}V$$

• **Dissimilarity measures (ED) and significance tests (*t* test)**

Twin Pair	ED	*t* tests ΔV	varcoC	nC	meanC Ln	ΔSylLn	varcoP	nPVI_ V	nPVI_C
01	0.544	t(55)= 0.64	**t(55)= -2.17** [a]	t(55)= 0.69	t(55)= 0.86	t(55)= -1.10	t(55)= 1.39	t(55)= 1.19	t(55)= -0.91
02	0.488	**t(71)= 2.39** [a]	t(71)= 0.35	t(71)= -1.31	t(71)= 1.94	**t(71)= 2.10** [a]	t(71)= 1.00	t(71)= 0.39	t(71)= 1.76
03	0.468	t(34)= 0.98	t(34)= -0.55	t(34)= 1.11	t(34)= 0.80	t(34)= -0.14	t(34)= 0.81	**t(34)= 2.50** [a]	t(34)= -1.24
04	0.482	**t(57)= 2.27** [a]	**t(57)= 2.16** [a]	t(57)= -0.30	t(57)= 1.79	**t(57)= 2.75** [a]	***t(57)= 3.40*** [b]	t(57)= 1.77	t(57)= 0.94
05	0.190	t(67)= -1.09	t(67)= 0.45	t(67)= -0.60	**t(67)= -2.40** [a]	**t(67)= -2.71** [a]	t(67)= 0.55	t(67)= -0.39	t(67)= 0.31
06	0.007	t(63)= 1.27	t(63)= -0.13	t(63)= -0.26	t(63)= -0.72	t(63)= 0.38	**t(63)= -2.35** [a]	t(63)= -8.42	t(63)= 0.16
07	0.191	t(61)= 0.16	t(61)= -1.23	t(61)= -0.60	t(61)= -1.52	t(61)= -0.004	t(61)= 0.09	t(61)= -1.40	t(61)= 0.11
08	0.308	t(70)= -1.22	t(70)= -0.78	t(70)= -0.54	t(70)= 1.12	t(70)= -1.80	t(70)= -1.13	t(70)= -0.78	t(70)= 0.62
09	0.642	**t(55)= 2.79** [a]	t(55)= -1.59	t(55)= 1.03	t(55)= -0.93	t(55)= 1.55	t(55)= 0.89	***t(55)= 3.16*** [b]	t(55)= 0.37
10	0.345	t(56)= 1.70	t(56)= -0.25	t(56)= -0.94	t(56)= 0.58	**t(56)= 2.19** [a]	t(56)= -0.64	t(56)= 1.95	t(56)= -0.20
11	0.129	t(68)= 1.10	t(68)= 0.84	t(68)= 0.95	**t(68)= -2.09** [a]	t(68)= 0.05	t(68)= 0.87	t(68)= 0.17	t(68)= 0.30
12	0.084	t(65)= 0.77	t(65)= 0.16	t(65)= 0.39	t(65)= -0.57	t(65)= 1.33	t(65)= -1.46	t(65)= 0.76	t(65)= 0.24

[a] Significant at P < 0.05 (without Bonferroni correction)
[b] Significant at P < 0.05 (with Bonferroni correction)

DISCUSSION

• **Euclidean distances (ED)**

- Based on the same number of prosodic measures, some twin pairs are more similar than others.

• **t tests**

- Durational measure ***nVPI_V*** *(normalized pairwise variability index for V intervals)* reveals significant differences between speakers in twin pair 09

- Intensity measure ***varcoP*** *(peak intensity variability across syllables)* reveals significant differences in twin pair 04

. . . . . .

• Overall, we observed variation in the temporal patterns exhibited by twin pairs. As highlighted by the **ED**, very few twin pairs are really similar (twin pairs 06 and 12) when considering the 8 prosodic characteristics.

• Upon further examination, **t-tests** revealed which features contribute the most to distinguish between twins.

✓ Interestingly, both intensity and durational measures allow twin differentiation - depending on the pair.

✓ The finding that **varcoP** can distinguish twin pair 04 is particulary relevant, as these speakers were misidentified by the MFCC-based ASR system. The system based on glottal source features gave LLR = 0 (no decision). [See Table 1]

CONCLUSIONS

• Rhythmic variability exists even between extremely similar speakers (i.e. identical twins).

✓ Prosody offers idiosyncratic information, possibly complementary to that provided by forensic systems based on vocal tract and glottal characteristics.

✓ Some of the investigated measures proved useful to tell certain twins apart where other systems had failed to distinguish them (see Table 1).

✓ Future hybrid approaches should consider adding prosodic measures for a better characterization of speakers and hence for more reliable forensic comparison systems.

• In terms of methods, PCA seems a good method for dimension reduction, especially with highly correlated measures.

• **Limitations:**
– The method used to investigate how similar/different twin pairs are follows previous studies on twins [8] but differ from common forensic approaches / output (e.g. EER, LLRs).

• **Future work:**
– Calculate weighted Euclidean distances.
– Explore different ways to combine the output provided by several forensic comparison systems.
– Take into account typicality apart from similarity measures.

REFERENCES

[1] Decoster, W., Van Gysel, A., Vercammen, J.; Debruyne, F. (2000). Voice similarity in identical twins. *Acta Oto-Rhino-Laryngologica Belgica, 55(1)*, 49-55.

[2] San Segundo, E. (2014). *Forensic speaker comparison of Spanish twins and non-twin siblings: A phonetic-acoustic analysis of formant trajectories in vocalic sequences, glottal source parameters and cepstral characteristics*, Doctoral dissertation, Menéndez Pelayo International University & CSIC

[3] San Segundo, E.; Künzel, H. J. (2015). Automatic speaker recognition of Spanish siblings: (monozygotic and dizygotic) twins and non-twin brothers. *Loquens*, 2(2):e021.

[4] San Segundo, E.; Gómez-Vilda, P. (2014). Evaluating the forensic importance of glottal source features through the voice analysis of twins and non-twin siblings. *Language and Law/Linguagem e Direito,* 1:22-41.

[5] San Segundo, E.; Mompeán, J.A. (2017). A Simplified Vocal Profile Analysis Protocol for the Assessment of Voice Quality and Speaker Similarity. *Journal of Voice*. DOI: http://dx.doi.org/10.1016/j.jvoice.2017.01.005

[6] Dellwo, V.; Leemann, A.; Kolly, M-J. (2015). Rhythmic variability between speakers: Articulatory, prosodic, and linguistic factors. *J. Acoust Soc Am.*, 137(3):1513-1528.

[7] He, L.; Dellwo, V. (2016). The role of syllable intensity in between-speaker rhythmic variability. *International Journal of Speech, Language and the Law,* 23: 245-275.

[8] Whiteside, S. P.; Rixon, E. (2013). Speech tempo and fundamental frequency patterns: a case study of male monozygotic twins and an age-and sex-matched sibling. *Logopedics Phoniatrics Vocology*, *38*(4), 173-181.

ACKNOWLEDGEMENTS

Thanks to Sandra Schwab for help with the corpus alignment tool. This investigation has been possible thanks to an International Short Visit of the Swiss National Science Foundation (SNSF).