

Speaker Profiling: An automatic method?

Georgina Brown, and Jessica Wormald

Department of Language and Linguistic Science, University of York, York, UK

{gab514|jessica.wormald}@york.ac.uk

Increasing attention is being given to the application of automatic speaker systems in forensic casework. The current paper considers an automatic speaker profiling system developed using the ACCDIST metric (Huckvale 2004) to group speakers into accent groups (Huckvale, 2007; Ferragne and Pellegrino, 2007; Hanani *et al*, 2013). The current system allows for the clustering of phones, meaning phoneme categories can be compared, not just individual, context dependent, segments. The system relies on segmental input in the form of mid-point MFCC vectors. Potential features of interest can both be specified, and also potentially identified through the system. This paper investigates the effects of various segmental combinations, exploring the system's strengths and weaknesses in a classification task. It highlights the importance of making considered phonemic choices when training the system before generating an automatic result. The first author has previously used the modified system with speakers of Scottish/English border varieties and observed a 61.2% recognition rate on an eight way recognition task with speakers from four locations and two age groups (Brown & Watt 2014).

The current paper demonstrates the system's ability to classify groups of Panjabi-English (PE) speakers across and within the two English cities of Bradford and Leicester after training with reading passage data. PE speakers are British-born native-English speakers with Panjabi language heritage. Within each location, two age groups are represented.

Results considering 20 PE speakers from Bradford and 26 from Leicester highlight the ability of the system to recognise speakers from different geographical locations and of different ages. Table 1 includes the system's results when including all vowel phonemes. Variation from these results is observed depending upon the combination of features selected, highlighting the importance of grounded sociophonetic choices when training the system. Sociophonetic differences between the respective groups can be exploited to improve the system's results

Table 1. Results from the Phoneme-based ACCDIST system. Results for all vowels and best reduced combination included.

<i>PE Speaker Groups</i>	<i>Features</i>	N correct	% correct
Bfd old vs. Lei old	All vowels	15/18	83.3
	FACE GOAT PRICE MOUTH	18/18	100
Bfd young vs. Lei young	All vowels	25/26	91.2
	FACE GOAT PRICE MOUTH	23/26	88.5
Bfd old vs. Bfd young	All vowels	5/20	25.0
	FACE GOAT PRICE MOUTH + /r/	15/20	75.0
	CHOICE NEAR FLEECE + /r/	13/20	65.0
Lei old vs. Lei young	All vowels	13/24	54.2
	FLEECE KIT GOOSE FOOT	17/24	70.8

References

Brown, G. and D. Watt (2014). *Performance of a novel automatic accent classifier system using geographically-proximate accents*. Poster presented at BAAP, University of Oxford. 7th-9th April

- Ferragne, E. and F Pellegrino. (2007). Automatic dialect identification: A study of British English. In Christian Muller (Ed.), *Speaker Classification, Volume 2 of Lecture Notes in Computer Science*. Berlin Heidelberg: Springer-Verlag. pp 243-257.
- Hanani, A., M. Russell and M. Carey. (2013). Human and computer recognition of regional accents and ethnic groups for British English speech. *Computer Speech and Language*, **27**, 59-74.
- Huckvale, M. (2004). ACCDIST: A metric for comparing speakers' accents. In *Proceedings of the International Conference on Spoken Language Processing*. pp. 29-32. Korea.
- Huckvale, M. (2007). ACCDIST: An accent similarity metric for accent recognition and diagnosis. In Christian Müller (Ed.), *Speaker Classification, Volume 2 of Lecture Notes in Computer Science*. Berlin Heidelberg: Springer-Verlag. pp. 258-274.