**Towards a knowledge-based system for accent classification in the British Isles**

This paper is a preliminary step towards designing a knowledge-based system for automatic accent classification based on acoustic criteria. Our goal is to come up with a parsimonious and phonetically interpretable system for quick accent diagnosis by an expert linguist. Such a system could also be useful for dialectologists and sociolinguists as it may support or contradict classic accent typologies (e.g. Trudgill, 1990; Hughes et al., 2005).

An important feature in such systems is that they should be able to imitate human reasoning, or at least, provide information in such a way that a human expert can use it most effectively. This point raises the following issue: distances between vowels in the Mel-Frequency Cepstral Coefficient (MFCC) space and correlations between individuals' between-vowel distances have been shown to be very effective for an automatic system (Ferragne & Pellegrino, 2007) – especially because within-speaker acoustic distances provide some kind of speaker normalization (Barry et al., 1989; Huckvale, 2004). However human experts do not explicitly reason in terms of distances. For instance, it is well known that *look* and *luck* are traditionally homophones in the north of England. However, a human expert will not normally have to assess the distance between the two vowels in order to produce a reliable judgement. Instead, the expert will be able to tell a northerner – with a typical accent – from a southerner just by hearing the vowel in *luck* or in similar words. As a consequence, we assumed that our knowledge-based system should rely on absolute vowel coordinates in some auditory phonetic space.

A hundred speakers from the Accents of the British Isles corpus were grouped under 6 accent supra-regions (following Trudgill, 1990). The speech material was a list of 11 /hVd/ words (*heed, hid, head, had, hard, hod, hoard, Hudd, heard, hood, who'd*). F1 and F2 values were measured at temporal mid-point, Bark-transformed, averaged for each vowel type, and a speaker's set of averaged F1 and F2 values was subsequently z-scored. The resulting parameters derived from F1 and F2 were fed as predictive attributes to the C4.5 classification-tree algorithm (Kohavi & Quilan, 2002) with accent class as the target.

The decision scenario is shown in Figure 1. The interpretation of Figure 1 is straightforward: each node is represented by a predictive variable – e.g. 'hidF2', means that the decision is based on the F2 value for the vowel in the test word *hid* – whose cut-off value appears above each corresponding leaf. For example, all the speakers from Wales are correctly classified after three rules involving F2 in *hid*, F1 in *hoard*, and F2 in *Hudd*.

The preliminary set of 5 rules shown in Figure 1 allows the classification of 100 speakers into 6 regions with a 16% error rate. In our task, the C4.5 algorithm achieved lower performance than linear discriminant analysis (4% error rate), but the former is probably more useful for the linguist because it provides an explicit scenario with a restricted number of predictive variables. Further improvements include increasing the number of phonetic segments and analyzing consonants, computing more acoustic parameters per phonetic segment, enhancing the legibility of the predictive variables by discretizing them (and perhaps using nominative values such as front/back and open/close), and testing the perceptual validity of the rules with humans performing the decision process.

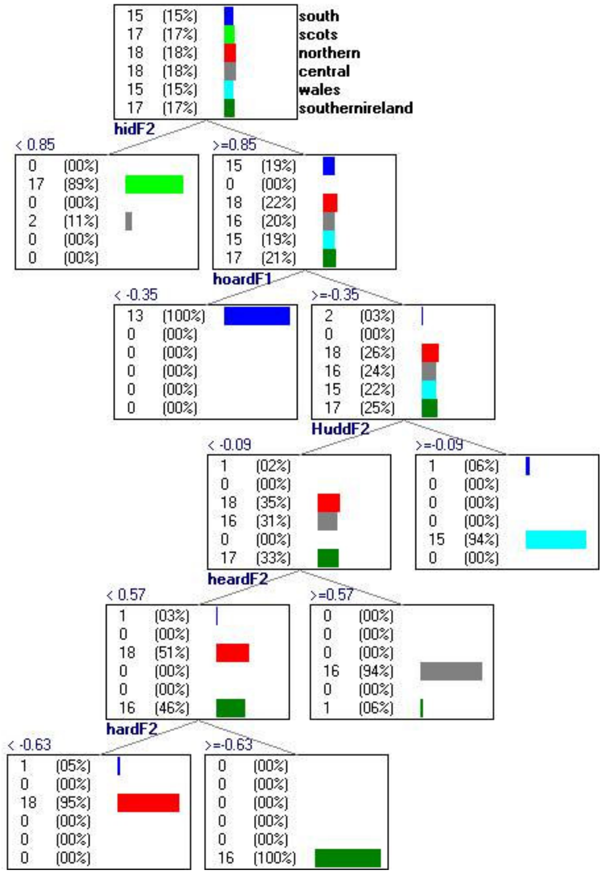


*Figure 1: C4.5 decision tree. The horizontal bars in each node represent the percentage of speakers from each region that have been assigned to this node.*

## References

Barry, W.J., Hoequist, C.E. and Nolan, F.J., 1989, An approach to the problem of regional accent in automatic speech recognition, *Computer Speech and Language*, 3:355–366.

Ferragne, E. and Pellegrino, F., 2007, Automatic dialect identification: A study of British English, in C. Müller [Ed], *Speaker Classification II*, 243–257, Springer, Berlin.

Huckvale, M., 2004, ACCDIST: A metric for comparing speakers' accents, *Interspeech* 29–32, Jeju, Korea.

Kohavi, R. and Quinlan, R., 2002, Decision Tree Discovery, in Klösgen, W. and Zytkow, J.M., [Eds], *Handbook of Data Mining and Knowledge Discovery*, 267–276, Oxford University Press, Oxford, 2002.

Hughes, A., Trudgill, P. and Watt, D., 2005, *English Accents and Dialects. An Introduction to Social and Regional Varieties of English in the British Isles*, Hodder Arnold, London.

Trudgill, P., 1990, *The Dialects of England*, Blackwell, Oxford.