The Role of Frequency and Magnitude Scales for Identifying Plosives' Place of Articulation

Daniel McCarthy and Jalal Al-Tamimi

Newcastle University

d.mccarthy1@newcastle.ac.uk; jalal.al-tamimi@newcastle.ac.uk

How to identify plosive consonants' place of articulation accurately from acoustic information has attracted much attention yet their identification remains less than fool-proof (e.g. Suchato's (2004) methods falsely classified the place of English plosives at a rate of 8%). One possible reason for this is that many studies (e.g. Stevens and Blumstein, 1978; Stevens et al., 1999; Suchato, 2004) derive their acoustic attributes from a linear Hz-dB spectrum rather than exploring more perceptually motivated representations, such as the Bark-Phon spectrum. In addition to this, although there are two main cues for identifying a plosive's place – its release burst and its formant transitions - the formant transitions have tended to be neglected by automatic speech recognition (see e.g. Ali et al., 2001: 838).

To fill these lacunae, the present study compares the performance of some of the existing attributes for identifying place of articulation (Stevens et al., 1999; Suchato, 2004) with some new attributes devised by the first author, including a comparison of attributes based on a Bark-Phon spectrum with those based on the more usual Hz-dB spectrum. The data consist of a manually annotated corpus of sentences read by 19 speakers of British English (9 male, 10 female), and contains a total of 6,779 plosive tokens, which occur in a wide range of phonetic contexts (e.g. in stressed and unstressed syllables, before front, back and central vowels, and as part of consonant clusters).

Using leave-one-out cross-validation, results show that attributes derived from the Bark-Phon spectrum rather than the Hz-dB spectrum classified place of articulation more accurately in three out of the four cases, on average by 3.9 percentage points. For voiced plosives, the two best formant-based attributes classified place almost as well as the two best burst-based ones (77.4% vs 78.9%), but for the voiceless plosives the formant-based attributes were clearly far less informative than the burst-based ones (45.7% vs 79.2%). Possible explanations for these findings as well as recommendations for future research are discussed.

References

Ali, A.M.A., Spiegel, J.V., and Mueller, P. (2001) 'Robust Classification of Consonants using auditory-based speech processing'. *International Conference on Acoustic, Speech, and Signal Processing*, Salt Lake City, pp. 833-840.

Stevens, K.N., Manuel, S.Y., and Matthies, M.L. (1999) 'Revisiting Place of Articulation Measures for Stop Consonants: Implications for Models of Consonant Production'. *Proceedings of the 14th International Congress of Phonetic Sciences*, Volume 2, pp. 1117-1120.

Stevens, K.N., and Blumstein, S.E. (1978) 'Invariant Cues for Place of Articulation in Stop Consonants'. *Journal of the Acoustical Society of America* 64(5), pp. 1358-1368.

Suchato, A. (2004) '*Classification of Stop Consonant Place of Articulation*'. PhD Thesis, Massachusetts Institute of Technology.