

Intra- and inter-speaker variability in acoustic properties of English /s/

Colleen Kavanagh

Department of Language & Linguistic Science, University of York, York, UK
cd519@york.ac.uk

In the field of forensic speech science, it is vital to explore all acoustic parameters that are potential speaker discriminants. Much of the previous research has been disproportionately focused on vowels. Although auditory analysis of consonants is common in forensic casework, the acoustic properties of consonant segments have been under-researched from this perspective. This study investigates the speaker-specificity of some acoustic characteristics of the English fricative /s/ and looks to contribute background population statistics for use in speaker comparison work.

The intra- and inter-speaker variability in duration and spectral properties of /s/ was investigated in data from 30 speakers of two British English dialects. Recordings of young adult male speakers of Cambridge English from the DyViS (Nolan, McDougall, de Jong, & Hudson, 2009) and IViE (Grabe, Post & Nolan, 2001) corpora were used, in addition to young and middle-aged male speakers of Leeds English from the IViE and Morley (Richards, 2008) corpora. Read speech was used in the present study to allow direct comparison across speakers.

Segment duration was normalized for speaking rate using Average Syllable Duration as a measure of local articulation rate. To permit comparison across the full dataset, the spectra for all 30 speakers were filtered at an upper limit of 8 kHz, as the lowest sampling rate used in the recordings was 16 kHz. An additional filter was applied at 4 kHz in order to explore speaker discrimination performance of the parameters at settings mimicking the bandpass filter effect of telephone transmission. Because 18 of the 30 speakers' recordings were sampled at a higher rate (44.1 kHz), further filters were applied to spectra for these 18 speakers only to investigate discrimination with data from wider frequency ranges. Four spectral measures – centre of gravity, standard deviation, skewness, and kurtosis – were calculated in a 40-ms window centered on the midpoint of each token.

Speakers' means and range of values for the five acoustic parameters (filtered at 500-8000 Hz) are presented in Figure 1. The triangles indicate mean values and the vertical bars represent the range of variation (maximum and minimum values) employed by each speaker. Although means display relatively little inter-speaker variation in all of the features, the individuals at the extreme high and low ends of the distributions may be the best discriminated, particularly those at the extremes on more than one parameter.

Discriminant analyses were conducted to investigate the most speaker-specific predictors or combinations of predictors; relative performance was compared across the different filter settings used. The discriminatory ability of these parameters was also evaluated and will be presented using a likelihood ratio framework.

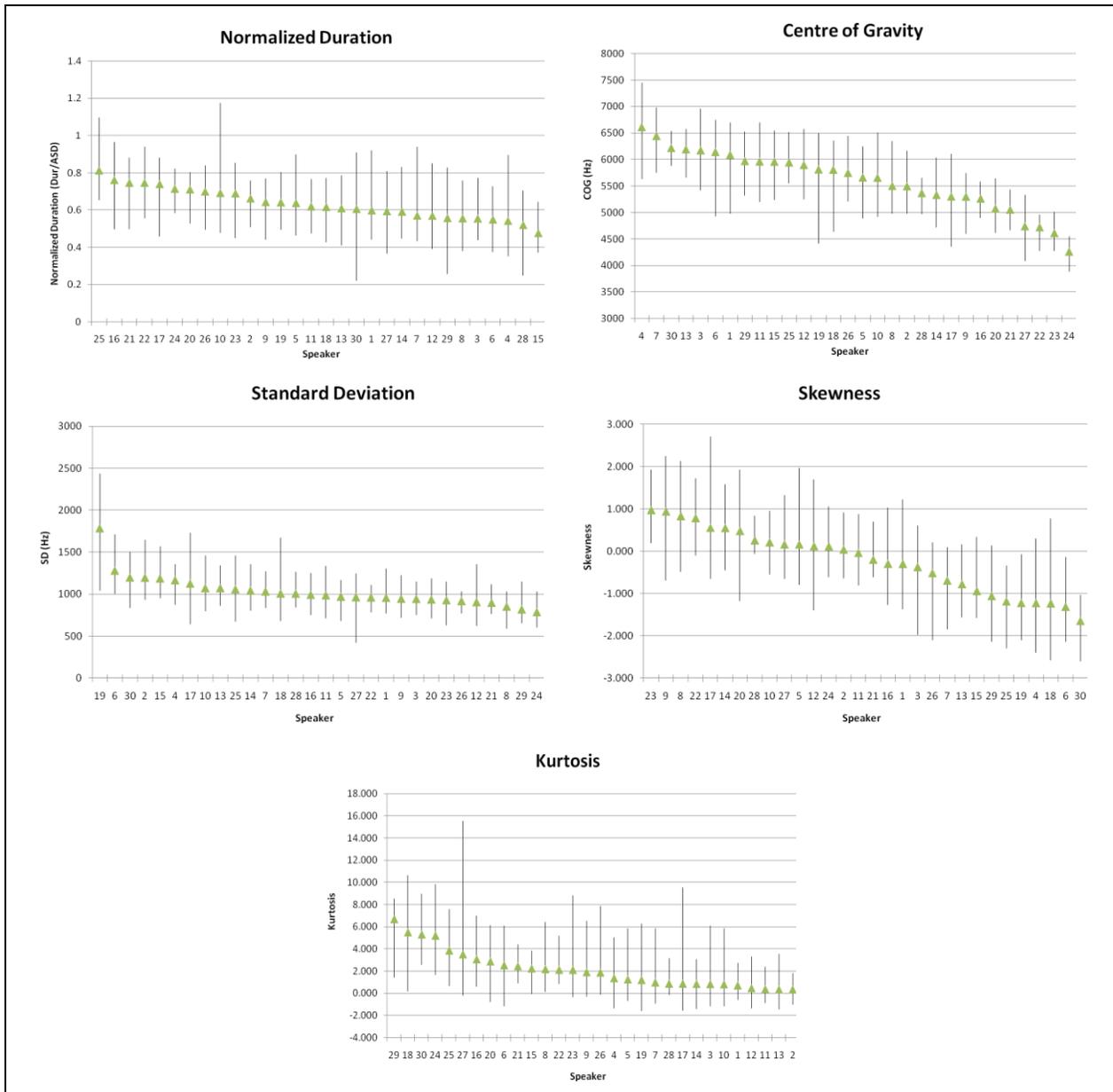


Figure 1. Speaker means and ranges for the five acoustic parameters, duration, centre of gravity, standard deviation, skewness, and kurtosis, for all 30 speakers, filtered at 500-8000 Hz. Speakers are arranged left to right in descending order of mean value for each parameter.

References

- Grabe, E., Post, B., & Nolan, F. (2001). The IViE Corpus. Department of Linguistics, University of Cambridge. <http://www.phon.ox.ac.uk/IViE>.
- Nolan, F., McDougall, K., de Jong, G., & Hudson, T. (2009). The DyViS database: style-controlled recordings of 100 homogeneous speakers for forensic phonetic research. *International Journal of Speech, Language and the Law*, **16**(1), 31–57.
- Richards, H. (2008). Mechanisms, motivations and outcomes of change in Morley (Leeds) English. Unpublished PhD thesis. University of York.