

## Potential pitfalls when choosing to normalise

NICHOLAS FLYNN

Technology is advancing at an ever-increasing rate and, as a consequence, tools designed to automate the extracting, measuring, and normalisation of vowel and formant data are readily available to researchers (e.g. Boersma 2001; Thomas & Kendall 2007). Such tools have substantially reduced the time taken to conduct a thorough analysis of vocalic variables and allow a far greater number of tokens to be measured and included in analyses. While taking advantage of the advanced technology and labour-saving tools available, it is important for researchers to make informed decisions about which methods and tools to use to ensure accuracy of results, avoid skewing and bias and minimise errors of measurement or normalisation (see, for example, Di Paolo et al. 2010; Watt et al. 2010; Thomas 2011). In this presentation, I focus on the normalisation of vowel formant data. There are a large number of available algorithms that can be used when normalising, many available as part of on-line normalisation tools. The choice of normalisation algorithm rests with the researcher and it is crucial that one is chosen that:

- performs well for the criteria of the study
- is suitable for the dataset being used and change can proceed.
- provides robust and replicable results

Using a combination of real and synthetic data, I demonstrate the ease with which errors can be made when using online normalisation tools, and reinforce the importance of checking and cross-checking instrumental results with our own auditory judgements. I will conclude that while normalised formant measurements can offer good cross-speaker comparisons, and are especially useful for visual purposes, it remains essential to still use our own judgements as ear-trained linguists when analysing, as technology does not (yet?!) have the power to determine when mistakes are made.