Automation and sociophonetics

JOSEF FRUEHWALD
University of Edinburgh

This talk will focus on the benefits of automation in vowel analysis, and what insights we can arrive at when we have 10x or 100x the number of vowel measurements than conventional studies. I will begin with a description of the FAVE suite (Rosenfelder et al 2011) which automates the process of formant estimation, and is becoming increasingly more widely used. I will describe how its method of Bayesian Formant. Tracking operates, and address frequent concerns that arise (e.g. it is a black box system, it will only return results that look like general American English, etc). I'll then lay out many of the positive benefits of the FAVE system. First and foremost is its explicitly defined and strict adherance to heuristics for formant estimation, providing the foundation for scientific replacability. Second is the large volume of data it produces, which allows for finer grained investigation of contextual effects than was previously possible. In conventional studies, there may have been 5 or 10 vowel measurements in rarer contexts, but with FAVE, we may now have 50 or 100. While one might expect that a "Big Data" approach might reveal a fuzziness to vowel systems, my research has found the opposite to be true, especially when viewing the data through a diachronic lens. The illustrating example for this will be the case of /ay/ raising before underlying voiced and voiceless flaps in Philadelphia. Pre-flap /ay/ is a relatively rare context in sociolinguistic interviews, and the results show a sharp diachronic division of these contexts according to their underlying phonological status.