Innovative Methods in Sociophonetics, II
Invited & Accepted Abstracts

University of Edinburgh
Edinburgh, Scotland (U.K.)
19th April 2017

http://www.lel.ed.ac.uk/wsc/ims/

CLOSELY ALIGNING OUR QUANTITATIVE METHODS WITH OUR SOCIOLINGUISTIC THEORIES
JOSEF FRUEHWALD (THE UNIVERSITY OF EDINBURGH)

Rereading the founding documents of variationist sociolinguistics (e.g. Labov (1969) and Cedergren & Sankoff (1974)) there are baroque discussions of quantitative detail (e.g. the "geometric ordering" of factors, and comparisons of "additive" and "multiplicative" models). The attention given to such issues makes most sense when we realize that these were not just isolated discussions about quantitative methods, but claims being staked out about linguistic and sociolinguistic theory. More recently, our quantitative methods have become more sophisticated with the introduction of off-the-shelf mixed effects models (Johnson, 2009) but at the same time they have become somewhat bifurcated from our theory building. It is not uncommon, for example, to read a sociolinguistic paper that presents the results of a regression model in one section, then, in a separate section, deploys a mixture of different quantitative and qualitative technique to actually develop the theory.

In this talk, I'll be discussing how we can build bespoke Bayesian statistical models with structures that align with the theories we wish to evaluate. The obvious benefit of such an approach is that it realigns our quantitative methods with our attempts at theory building. I'll specifically fit models to evaluate the degree of speech community and individual level cohesion in a number sociolinguistic variables.

A TIN EAR FOR ACCENTS? HUMAN VS. MACHINE IN ACCENT CLASSIFICATION AND AUTHENTICATION TASKS

GEORGINA BROWN, MEGAN JENKINS, JESSICA WORMALD AND DOMINIC WATT
(THE UNIVERSITY OF YORK & JP FRENCH ASSOCIATES)

Characterising subvarieties of a language according to their phonetic and phonological properties is often – perhaps necessarily – a selective exercise. As sociophoneticians we tend to focus on the features that do the most work in making social and regional accents different from one another, and pay less regard to those features that the accents have in common. Doing so presupposes that the features which embody inter-accent differences are more worthy of our attention than the other sort, and may thereby also betray our tacit assumptions about what features count as more or less perceptually salient to non-linguists (see further Llamas *et al.* 2016). There are therefore benefits to be gained by making use of methods designed to capture properties of accents in a more holistic way. We can do this by expanding the set of phonetic/phonological variables under investigation, and/or by deploying analytical techniques which are fundamentally different in nature from those in the sociophonetician's customary toolkit.

Automated systems such as FAVE (Rosenfelder *et al.* 2011) give us the opportunity to generate and process data on an unprecedented scale, and through their use we can address questions about sociophonetic variation that it may otherwise be impractical to try to answer. On the other hand, it could be argued that FAVE's combination of forced alignment and automatic formant extraction is doing effectively the same job as would a human analyst interested in the acoustic correlates of vowel variation, albeit reproducibly, more transparently, and very much more quickly. It is not qualitatively different in this sense.

If, however, we apply automated methods which work by representing properties of the speech signal at a more abstract level, for instance the MFCC- and i-vector-based algorithms that form the basis of the current generation of speech and speaker recognition systems, we may compare human and machine performance in a variety of interesting ways. In the first part of the current paper we discuss the capacity of a novel automatic accent classification system, Y-ACCDIST (Brown 2016), to correctly sort audio samples representing a variety of British English accents. These are drawn from the *Accent and Identity on the Scottish/English Border* (AISEB) corpus (Watt et al. 2013), the *Panjabi-English Corpus* (Wormald 2016), and the *Northern Englishes Corpus* (Haddican et al. 2013). Y-ACCDIST is shown to perform at a very high level of accuracy in the sorting task, even where its inter-sample comparisons are based just on a parsimonious subset of the vocalic portions of the recordings.

In the second part of the paper we focus on the extent to which YACCDIST is able to distinguish between recordings of a group of speakers composed of laypeople, actors and phoneticians, talking in genuine and feigned

Edinburgh accents (Jenkins 2016). Voice disguise effected by using an accent other than one's own is reasonably common in forensic phonetic casework, and it can be very difficult for human listeners to detect if done competently. We therefore set out to test whether an automatic system might be harder to deceive. In the present case, we set Y-ACCDIST and a panel of approximately 100 human listeners the task of trying to sort the authentic from the feigned accent samples. The results show that both perceptual systems – machine and human – distinguish between the two sets of samples at above chance levels. Most importantly from the forensic standpoint, the false positive rate – i.e., the rate at which fakers are classified as genuine – is substantially lower for Y-ACCDIST (5%) than it is for the human listeners (28%). We conclude by considering the implications of this finding from both sociophonetic and forensic perspectives.

References

Brown, G. (2016). Automatic accent recognition systems and the effects of data on performance. *Proceedings of Odyssey 2016*, Bilbao, June 2016, pp. 94-100.

Haddican, B., Foulkes, P., Hughes, V. & Richards, H. (2013). Interaction of social and linguistic constraints on two vowel changes in northern England. *Language Variation and Change* 25(3): 371-403.

Jenkins, M. (2016). Identifying an imitated accent: Humans vs. computers. Unpublished MSc dissertation, University of York.

Llamas, C., Watt, D. & MacFarlane, A.E. (2016). Estimating the relative sociolinguistic salience of segmental variables in a dialect boundary zone. *Frontiers in Psychology* 7:1163. DOI: 10.3389/fpsyg.2016.01163.

Rosenfelder, I., Fruehwald, J., Evanini, K. & Yuan, J. (2011). *FAVE (Forced Alignment and Vowel Extraction) Program Suite*. URL: http://fave.ling.upenn.edu.

Watt, D., Llamas, C. & Johnson, D.E. (2013). Sociolinguistic variation on the Scottish-English border. In Lawson, R. (ed.). *Sociolinguistics in Scotland*. Basingstoke: Palgrave Macmillan, pp. 79-102.

Wormald, J. (2016). *Regional Variation in Panjabi-English*. PhD thesis, University of York. URL: http://etheses.whiterose.ac.uk/13188/.

3:00-3:30

DISCOURSE ANALYSIS IN ETHNOGRAPHIC, QUALITATIVELY-SKEWED MIXED METHODS IN SOCIOPHONETICS

FERGUS O'DWYER

This presentation will outline how ethnographically informed, qualitatively-skewed mixed methods can investigate the indexical meaning of particular phonetic variants. The benefits of an ethnographic approach include that it can facilitate careful discourse analysis alongside detailed acoustic analysis, to reveal much about the ways in which phonetic variability is harnessed for the communication of interactional information (Hay & Drager 2007). I suggest that one way to advance sociophonetics in this area is to thoroughly attend to the discourse environment in which tokens occur. I exemplify this by outlining my analytical manoeuvres in a sports club, which led me to argue that positions of knowledge and authority are indexed by a low and back PRICE variant.

Statistical analysis examined the correlation between variation in the PRICE vowel in interview data, and three types of predictors (linguistic, social attribute, and interactional variables). These statistical results highlight the patterns and trends of variation that deserve more attention and interpretations: stancetaking in this case. I then highlight key repetitions, which allows me to infer what the data implies, and make interpretations about the social meaning of sound variation.

I then implemented in-depth analysis of speech excerpts where backed and lowered PRICE tokens are found, using a discourse analysis lens (Schiffrin 1994; Schegloff 2007; Gee 2011; Jaffe 2013). A turn by turn take interprets how identities are managed in interaction, and highlights the alignments and other contextualization cues which allow listeners to assess what the speaker intends. The general result of this analysis is an explanation of how speakers place themselves and other in positions ranging from less to more knowledgable, and forms interpretations of the discourse functions that seem to cluster around occurrences of the PRICE variants.

As Drummond and Dray (2015) suggest, the use of certain variants can be better understood by looking at the practices which speakers value, participate in and identify with. My experience suggests that this is best done by using discourse analysis to highlight the salient patterns, specifically the stance commonalities of discourse functions of the variants in question. This process allows researchers to outline how social meanings and interactional identities are indexed in salient sociophonetic variants.

References

Drummond, Rob, and Susan Dray. 2015. "'Gimme Dat Ting'. Word Initial TH-Stopping among Urban British Youth." presented at the *NWAV 44*, Toronto, October 22.
Gee, James Paul. 2011. *How to Do Discourse Analysis*. London: Routledge.
Hay, Jennifer, and Katie Drager. 2007. "Sociophonetics." *Annual Review of Anthropology* 36 (1): 89–103.
Jaffe, Alexandra. 2013. "Anthropological Analysis." In *Research Methods in Sociolinguistics: A Practical Guide*, edited by Janet Holmes and Kirk Hazen, 213– 229. Oxford: Wiley-Blackwell.
Schegloff, Emanuel A. 2007. *Sequence Organization in Interaction: Volume 1: A Primer in Conversation Analysis*. Cambridge: Cambridge University Press.
Schiffrin, Deborah. 1994. *Approaches to Discourse*. Cambridge: Blackwell.

3:30-4:00

ASSESSING TYPICALITY IN FORENSIC VOICE COMPARISON: HOW CAN SOCIOPHONETICS HELP? (AND HOW CAN SOCIOPHONETICS BENEFIT?)

VINCENT HUGHES AND JESSICA WORMALD

In this paper we will talk about how sociophonetics can directly improve the quality of forensic voice evidence in the UK. We will also discuss ways in which sociophonetics can benefit through greater collaboration across disciplines.

What's the problem?
Forensic voice comparison accounts for the majority of work conducted by forensic speech scientists (French & Foulkes 2012). Such cases involve the comparison of an unknown voice (e.g. covert drug deal) with a known voice (e.g. police interview). It is the role of an expert to evaluate the evidence in terms of support for the prosecution and defence. To do this the expert needs to assess not only the similarity between the voices, but also, crucially, the typicality of features in the wider population. This is because strength of evidence is dependent on whether features of the voice are common or rare. Typicality is often assessed based on the experience of the analyst. However, as shown in Ross et al. (2016), there is considerable inter-analyst variability in estimates of the typicality of linguistic features, which raises issues about the reliability of experience-based evidence presented to the courts.

Why is this a problem?
There have been increasing calls for typicality to be assessed in a more robust, replicable, and reliable way (see Morrison 2014). To do this, it is necessary to have up-to-date reference material, such as linguistic descriptions, or corpora of representative speakers which can be used to calculate typicality empirically. A significant hurdle, however, is the current lack of coverage in terms of the range of varieties for which reference data/ information exist. There is also a lack of data which capture the range of potential within-speaker variability across forensically realistic tasks. Furthermore, existing corpora are relatively limited. Corpora collected for sociolinguistic work are often very highly controlled for relevant regional and social factors, but they are usually relatively small and there is currently no UK repository for accessing the data. Corpora collected for automatic speaker recognition have the benefit of containing a considerable number of speakers. However, there are usually limited controls over speaker demographics.

What are the solutions?
We have a number of potential solutions to these issues and we would like to encourage feedback on our proposals. Specifically, we are considering: (i) establishing a set of 'forensic protocols' for researchers to follow

when collecting a corpus, (ii) starting a journal for descriptive works on regional and social varieties of British English, and (iii) creating an open-access repository for corpora. We hope that these resources will be useful not just to forensics and sociophonetics, but also to linguistics and phonetics more generally, speech and language therapy, second language acquisition, speech technology, and education amongst other disciplines.

Why bother?
We will discuss a number of mutual benefits of such collaboration for forensics and sociophonetics. Specifically, we will consider the theoretical and practical implications of analysing more forensically realistic data in sociophonetics.

References

Foulkes, P. and French, J. P. (2012) Forensic speaker comparison: a linguistic-acoustic perspective. In P. Tiersma and L. Solan (eds.) *Oxford Handbook of Language and Law*. Oxford: Oxford University Press. pp. 557-572.

Morrison, G. S. (2014) Distinguishing between science and pseudoscience: testing of validity and reliability, and approaches to forensic voice comparison. *Science and Justice* 54: 245-256.

Ross, S., French, J. P. and Foulkes, P. (2016) UK practitioners' estimates of the distribution of speech variants. Paper presented at the *International Association of Forensic Phonetics and Acoustics (IAFPA)* Conference, York, UK, 24-26 July 2016.

4:00-4:50

Through the looking glass: Changing perspectives on gender and the /s ʃ/ contrast in Glasgow

Jane Stuart-Smith (The University of Glasgow)

A key issue for sociophonetic research is how to capture and characterise the relevant aspects of speech which relate to identified social constructs. Acoustic analysis has until recently been relatively 'static'. For example, the complex dynamic patterning of acoustic vowel quality over the timecourse of a vowel's production is typically represented from a single timepoint, or an average across a few timepoints. Recent advances in acoustic phonetics and statistical analyses now allow for both dynamic representation of the phonetic entities for sociophonetic analysis, and statistical consideration of their social correlates which account in different ways for dynamic acoustic representations (see e.g. Docherty et al 2015; Stuart-Smith et al 2015; Speech dynamics, social meaning, and phonological categories). Socially-conditioned variation in the production of /s/ is now well-established across several varieties of English (e.g. Stuart-Smith 2007; Levon and Holmes-Elliott 2014) and other languages (e.g. Pharao et al 2014). Reidy (2016) showed that dynamic acoustic analysis helps differentiate /s/ and /ʃ/ in English, and /s/ in English and Japanese. The question for this paper is: does a dynamic acoustic representation help resolve the impact of social factors of gender and time on /s/ and /ʃ/ in spontaneous Glaswegian vernacular speech?

Glaswegian vernacular has long been noted to show an auditorily-retracted /s/, especially in male speakers (Macafee 1983). Stuart-Smith (2007)'s acoustic analysis of wordlist data collected in the 1990s confirmed a lower peak frequency for male speakers, but also for working-class girls, who are distinct not from working-class boys but from their middle-class counterparts. Glasgow has shown substantial changes over the twentieth century: are shifts in social gender linked to linguistic change? A preliminary analysis of 5095 tokens of /s/ and /ʃ/ from 32 middle-aged and adolescent working-class speakers recorded in the 1970s and 1990s, showed consistently higher peaks in female speakers for /s/ than males. But over time a significant reversal is also found such that girls born most recently show lower peaks than women born in the 1920s; Figure 1. This paper presents the results of a dynamic acoustic analysis applied to the same /s/ and /ʃ/ data for the same speakers. All instances of stressed initial /s/ and /ʃ/ were extracted and peak, slope, mean and spread measures were taken from a series of spectra (10ms, Hamming window) over the central 70% of the fricative. The resulting time series were then subjected to a Discrete Cosine Transformation (DCT) analysis (Harrington 2010), and the first four coefficients were statistically analysed using linear mixed effects modelling for adjacent phonetic context, gender and decade of birth.
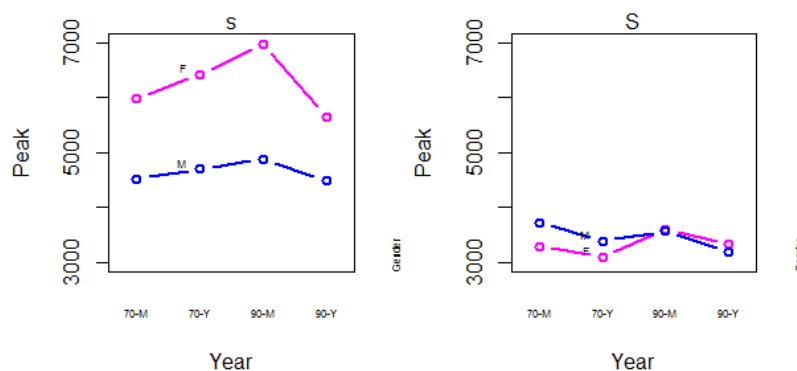
Figure 1: Estimates of peak frequency (Hz) taken from Long Term Average Spectra, from spontaneous speech from 32 speakers, for the significant interaction of sound, gender, and decade of recording (70M = middle-aged speakers recorded in 70s, etc).

References

Docherty, G., Gonzalez, S., and Mitchell, N. (2015) Static vs dynamic perspectives on the realization of vowel nuclei in West Australian English. In: 18th International Congress of Phonetic Sciences, Glasgow, Glasgow, UK, 10-14 Aug 2015

Harrington, J. (2010). *Phonetic analysis of speech corpora*. John Wiley & Sons.

Levon, E., & Holmes-Elliott, S. (2013). East end boys and west end girls:/s/-fronting in Southeast England. *University of Pennsylvania Working Papers in Linguistics*, *19*, 111-120.

Macafee, C. (1983). *Glasgow*. John Benjamins Publishing.

Pharao, N., Maegaard, M., Møller, J. S., & Kristiansen, T. (2014). Indexical meanings of [s+] among Copenhagen youth: Social perception of a phonetic variant in different prosodic contexts. *Language in Society*, 43, 1-31.

Reidy, P. F. (2016). Spectral dynamics of sibilant fricatives are contrastive and language specific. *The Journal of the Acoustical Society of America*, *140*(4), 2518-2529.

Speech dynamics, social meaning, and phonological categories Satellite workshop to LabPhon15, July 13, Cornell University

Stuart-Smith, J., Lennon, R., Macdonald, R., Robertson, D., Soskuthy, M., Jose, B., and Evers, L. (2015) A Dynamic Acoustic View of Real-Time Change in Word-Final Liquids in Spontaneous Glaswegian. In: 18th International Congress of Phonetic Sciences, Glasgow, Glasgow, UK, 10-14 Aug 2015.

Stuart-Smith, J., Timmins, C., and Tweedie, F. (2007) 'Talkin' Jockney'?: variation and change in Glaswegian accent. *Journal of Sociolinguistics*, 11, 221-260.

Stuart-Smith, J. (2007) Empirical evidence for gendered speech production: /s/ in Glaswegian. In: Cole, J. and Hualde, J.I. (eds.) *Laboratory Phonology 9.* Series: Phonology and phonetics. Mouton de Gruyter: New York, USA, 65-86.