Of vowels and 'systems': New methods for the study of vocalic variation

Panel Organisers: Anne Fabricius (Roskilde University) and Dominic Watt (University of York, UK)

Methods XV, Groningen, August 2014

Abstract of the session

The study of variations in vowel production and within vocalic configurations (often conceptualized as 'systems' and 'subsystems' within sociolinguistic theory, e.g. in Labov 1994) forms part of the paradigmatic foundation of sociolinguistics. Recently, new developments in the methods employed in the quantitative study of vowels have emerged, both in terms of methodologies of measurement and of graphical representation of vowel configurations or 'systems'. While formant extraction has been the key form of instrumental analysis in acoustic phonetics since Joos 1948 and Peterson and Barney 1952, and in sociolinguistics since Labov, Yaeger and Steiner 1972, techniques deriving from speech recognition work, for example, have recently made inroads into the field, and speech perception has taken a leading role within sociophonetics in particular. Forced alignment and programming in R for new quantitative analysis methods are becoming ever more widespread tools. In the light of new ways of looking at vowels, we want to ask for example whether there is a need for sociolinguistics/dialectology to challenge the fundamental idea of the vowel system as a 'system' in the first place. To what extent can the system be thought of as a theoretical convenience, or is it 'real' in some sense, and what arguments and evidence dominate each position? Why do we think of the vowels of language X as operating as a coherent and homeostatic series of contrasts, and what are we claiming when we present evidence of the system changing in some way? This panel will address questions such as this while showcasing a set of cutting-edge papers that illustrate some of the new methodologies and conceptualizations of the vowel space, which makes Methods XV an ideal venue for this collection of presentations. The discussion will aim to open up the possibilities these new techniques provide, and to ask how these methodologies and analytical approaches might have consequences for the ways in which we conceptualize vowel variation theoretically. We look at concrete measurement issues, conceptualizations of the vowel space and procedures such as normalization.

References:

Joos, M. (1948). Acoustic Phonetics. Language, 24(2), 5–136. doi:10.2307/522229

Labov, W. (1994). *Principles of Linguistic Change,* volume 1: *Internal Factors*. Oxford: Basil Blackwell.

Labov,W.,Yaeger, M, & Steiner, R. (1972). A quantitative study of sound change in progress. Philadelphia: U.S. Regional Survey.

Peterson, G. E., & Barney, H. L. (1952). Control Methods Used in a Study of the Vowels. *The Journal of the Acoustical Society of America*, 24(2), 175–184. doi:10.1121/1.1906875

TIMETABLE

10.30am Brief Introduction by the organisers

10.30am to 11am CONTRIBUTION 1: *Automation and Sociophonetics*, Josef Fruehwald, (University of Edinburgh)

11am to 11.30am CONTRIBUTION 2: *On the impact of noise on vowel formant measures*, Tamara Rathcke (University of Kent) & Jane Stuart-Smith (University of Glasgow)

11.30am to 12 noon CONTRIBUTION 3: *Measurement error and stability in vowel formant extraction: A simulation experiment,* Tyler Kendall (University of Oregon) and Charlotte Vaughn (Northwestern University)

12 noon to 12.30pm CONTRIBUTION 4: Reconceptualizing the vowel space, Robert A. Fox and Ewa Jacewicz (Ohio State University)

12.30 to 2.30pm LUNCH BREAK

2.30pm to 3pm CONTRIBUTION 5: *Naïve Categorization of American English Vowels*, Douglas S. Bigham (San Diego State University)

3.pm to 3.30pm CONTRIBUTION 7: *Potential Pitfalls when Choosing to Normalise*, Nicholas Flynn (Independent scholar)

3.30pm to 5pm, DISCUSSION, Roeland van Hout (Radboud University, Nijmegen), discussant.

CONTRIBUTION 1

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.

Rosenfelder I., Fruehwald, J., Evanini, K. and Jiahong Yuan. 2011. FAVE (Forced Alignment and Vowel Extraction) Program Suite. http://fave.ling.upenn.edu CONTRIBUTION 5

CONTRIBUTION 2

On the impact of noise on vowel formant measures

Tamara Rathcke (University of Kent) & Jane Stuart-Smith (University of Glasgow)

Acoustic analyses of speech recordings, and subsequent vowel formant measures taken from them, may be compromised by various factors. Recorded in different settings, using various equipment and digitisation procedures, vowel formant values may be affected by technical artefacts (De Decker and Nycz 2010; Hansen and Pharao 2006/submitted; Plichta 2004) or lead to unreliable measures which then require a time-consuming manual correction or even an exclusion from the dataset, thus introducing random variability into the sample or reducing the sample size.

In this paper, we concentrate on the implications of noise components on the F1/F2-measurements of /i a u/ vowels in Scottish English. Our analyses drew upon spontaneous speech corpora of Glaswegian vernacular recorded in the 1970s and 2000s. Most of these recordings were made using lavalier microphones. However, their placement was not controlled for, nor were sources of background noise or

acoustic properties of the respective places where recordings took place. As well as the inherent variability in the recording settings, all these factors may lead to a ground hum, buzz, hiss or extraneous acoustic events on the recording (Jason 2010). With an increase of noise components, formants can appear very faint or have larger bandwidths and therefore be less clearly defined (cf. Plichta 2004). Two acoustic manifestations of noise are examined here: (1) the signal-to-noise ratio and (2) the spectral composition of the noise components. We confirm and extend previous findings by showing that (1) classical LPC-algorithms used for formant measurements are very sensitive to even slight changes in both noise parameters; (2) especially F1 is affected; and (3) the amount of deviation in F1 differs depending on the vowel category (cf. Hansen and Pharao 2006/submitted). We further make suggestions for the best practice when working with recordings of variable sound quality.

References

Corey, J. (2010). Audio Production and Critical Listening: Technical Ear Training. Elsevier.

De Decker, P. and Nycz, J. (2010). For the record: which digital media are good enough for sociophonetics?, Poster presented at NWAV 39, University of Texas-San Antonio, November 4-6, 2010.

Hansen, G. and Pharao, N. (2006). Microphones and Measurements, In Ambrazaitis et al (eds) Proceedings of FONETIK 2006, Department of Linguistics Working Papers 52, University of Lund: Sweden, 49–52

Hansen, G. and Pharao, N. (submitted). Differences in formant values caused by different microphone setups. Journal of Acoustical Society of America.

Plichta, B., 2004. Data acquisition problems. In B. Plichta Signal acquisition and acoustic analysis of speech available at: http://bartus.org/akustyk/signal_aquisition.pdf

CONTRIBUTION 3

Measurement error and stability in vowel formant extraction: A simulation experiment

Tyler Kendall (University of Oregon) and Charlotte Vaughn (Northwestern University)

Recent years have seen increased focus on methodologies for acoustic vowel study. Much work has focused explicitly on techniques for vowel analysis, such as through the development and evaluation of vowel normalization procedures (e.g., Clopper 2009, Fabricius et al. 2009, Flynn 2011, Thomas and Kendall 2007), resulting in more rigorous methods. However, sources of error exist in other facets of acoustic vowel research and these other potential problems have been addressed less frequently. Specifically, it is clear that there are limitations in the accuracy and precision of vowel measurements (e.g., Harrison 2004, 2007), as a function of linear predictive coding (LPC) methods, as well as, of course, noise in the acoustical signal being studied. For instance, different measurement points and different LPC settings (such as for the formant analysis procedure in Praat) are known to yield different results (Boersma & Weenink 2013, Duckworth et al. 2007). Researchers are generally well aware of the need to consider inter-analyst

differences in their acoustic work. Yet, less research has explicitly or quantitatively studied the extent to which these differences matter for the outcome of an investigation (Duckworth et al. 2007, Harrison 2004, 2007). In this presentation, we consider the sources of error in common formant extraction techniques, investigating the extent to which the delimitation of vowel boundaries and software (Praat) settings influence the formant values obtained. To do this, we report on the results of a vowel measurement simulation where, rather than extracting a single measurement for each vowel, thousands of measurements are taken for each vowel with varied settings in jittered measurement locations (seeded by measurements from a human analyst) and vowel tokens are treated as distributions of probable formant frequencies instead of simple points or vectors in scatter plots. Such a consideration, we argue, sheds important insight into the bounds of measurement error in vowel work.

References

Boersma, Paul & David Weenink. 2013. Praat: Doing phonetics by computer. [software & documentation]

Clopper, Cynthia. 2009. Computational methods for normalization acoustic vowel data for talker differences. Language and Linguistics Compass 3(6): 1430-42.

Duckworth, Martin, Kirsty McDougall, Gea de Jong, and Linda Shockey. 2007. The reliability of formant measurements in high quality audio data: the effect of agreeing measurement procedures. Conference presentation: 2007 IAFPA Annual Conference. Plymouth, UK.

Fabricius, Anne H., Dominic Watt, and Daniel Ezra Johnson. 2009. A comparison of three speaker-intrinsic vowel formant frequency normalization algorithms for sociophonetics. Language Variation and Change 21:413-35.

Flynn, Nicholas. 2011. Comparing vowel formant normalisation procedures. York Working Papers in Linguistics (Series 2) 11: 1-28.

Harrison, Philip. 2004. Variability of formant measurements. MA Dissertation. York, UK: University of York.

Harrison, Philip. 2007. Formant measurement errors: Preliminary results from synthetic speech. Conference presentation: 2007 IAFPA Annual Conference. Plymouth, UK.

Thomas, Erik R. and Kendall, Tyler. 2007. NORM: The vowel normalization and plotting suite.

CONTRIBUTION 4

Reconceptualizing the Vowel Space

Robert A. Fox and Ewa Jacewicz

Characterization of the vowel system of a language in terms of the midpoint formant values of a relatively small set of "point" or "corner" vowels has been relatively common in phonetic research, even quite recently. See, for example, Chiung et al., 2012, who compared the vowel systems of three age groups and five languages using the F1 and F2 midpoint values for /i, u, a/. In an extensive project, Vorperian & Kent (2007) reanalyzed a set of published acoustic vowel data from speakers aged 4 through adulthood. Their study focused on vowel space development (including the F1-F2 area) based on the midpoints of the four vowels (/i, u, a, æ/) which comprise the "vowel quadrilateral." However, a limitation of this approach is the assumption that such point vowels are a valid measure of the boundaries of the speaker's vowel space although vowels are often produced outside this delimited area. Concentration on vowel formant values only at the midpoint also ignores vowel inherent spectral change which can differ as a function of both the age and dialect of the speaker (Jacewicz & Fox, 2013) and can be affected by sound change. Utilizing the formant values at a number of different locations for a wider range of individual vowels has significant implications for the size and shape of the resulting vowel space. Our recent approach to analyzing and conceptualizing the vowel space of a speaker (looking at variation in this space as a function of dialect and age) has been to define the boundaries of the vowel space on the basis of produced exemplars of all vowel categories in American English whose formant trajectories are sampled at multiple time points. We have explored different ways of defining the boundaries of the vowel space on the basis of these formant values. In addition, we considered variations in vowel "density" across this vowel space recognizing that vowels may be more concentrated and overlap more in some areas than in others (some of which may be very sparsely inhabited). We have found that the distribution of these dense areas also varies from dialect-to-dialect and generation-to-generation. We will describe the techniques used in this reconceptualization of the vowel space.

References:

Chiung, H., Kong, E.J., Edwards, J., Weismer, G. & Fourakis, M. (2012). Cross-linguistic studies of children's and adults' vowel spaces. Journal of the Acoustical Society of America, 131, 442-454.

Jacewicz, E. & Fox, R. A. (2013). Cross-dialectal differences in dynamic formant patterns in American English vowels. In: Vowel Inherent Spectral Change, ed. by G. S. Morrison and P. Assmann, pp. 177-198. New York: Springer.

Vorperian, H.K. & Kent, R.D. (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. Journal of Speech, Language, and Hearing Research, 50, 1510–1545.

CONTRIBUTION 5

Naïve Categorization of American English Vowels

Douglas S. Bigham (San Diego State University)

Many studies have shown evidence for the categorical perception of speech sounds (see, e.g., Liberman et al., 1957; Beddor & Strange, 1982; Lotto et al., 1998; Diehl et al., 2004). Typically, perception studies investigate the categorical boundaries between percepts with closely related acoustic structures, such as the F1 cross-over point between KIT and DRESS percepts or the VOT boundary between voiced and voiceless stop percepts. Although listener perception generally coincides with the pre-established categories of expert phoneticians, no major studies currently exist that begin without these kinds of a priori category assumptions. Without knowing the cognitive reality of these categories as naïve listeners experience them, categorical perception phenomena cannot be fully interpreted.

This study works toward filling this gap in our understanding of the categorical perception of vowel sounds by presenting results based on the naïve-categorization views of listeners. Instead of providing category or token labels and having listeners perform a difference-judgement task, listeners were asked to group vowel tokens into 2, 3, 4, and 5 categories as they perceived them. All listeners were previously familiar with the speaker who provided the token data. No speakers had previous experience with phonetics or the phonetic categories of vowel sounds. Results are interpreted via a correlation matrix of vowel x category groupings.

The evidence for these patterns may lend objective support to the posthoc construction of the feature [+/-peripheral] in Labov (1994) and related work or help explain the connectedness of vowels for historical work. The ways in which these naïve groupings deviate from the categories of expert phoneticians can not only shed light on language change phenomena but also provide a principled benchmark from which future work on vowel perception, categorization, and change can proceed.

References

Beddor, P. S. and Strange, W. 1982. Cross-language study of perception of the oral-nasal distinction. Journal of the Acoustical Society of America 71, 1551-1561.

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech Perception. Annual Review of Psychology, 55(1), 149–179. doi:10.1146/annurev.psych.55.090902.142028

Labov, W. (1994). Principles of Linguistic Change, volume 1: Internal Factors. Oxford: Basil Blackwell.

Liberman. A.M. , K.S. Harris, H.S. Hoffman, B.C. Griffith. 1957. The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology.54:358-368.

Lotto AJ, Kluender KR, Holt LL. 1998. The perceptual magnet effect depolarized. J. Acoust. Soc. Am. 103:3648–55

CONTRIBUTION 6

Potential Pitfalls when Choosing to Normalise

Nicholas Flynn (Independent scholar)

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 online 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
- 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.

References

Boersma, Paul (2001) “Praat, a system for doing phonetics by computer”. Glot International 5: 341-345.

Di Paolo, Marianna, Malcah Yaeger-Dror & Alicia Beckford Wassink (2010) ‘Analysing vowels’. In: Di Paolo, Marianna & Malcah Yaeger-Dror (eds.) Sociophonetics: A Student’s Guide. London: Routledge. pp. 87-106.

Thomas, Erik R. (2011) Sociophonetics: An Introduction. Basingstoke: Palgrave Macmillan.

Thomas, Erik R. & Tyler Kendall (2007) NORM: The Vowel Normalisation and Plotting Suite. Online Resource. URL: <http://ncslaap.lib.ncsu.edu/tools/norm>

Watt, Dominic J.L., Anne H. Fabricius & Tyler Kendall (2010) ‘More on vowels: Plotting and normalisation’. In: Di Paolo, Marianna & Malcah Yaeger-Dror (eds.) Sociophonetics: A Student’s Guide. London: Routledge. pp. 107-118.