DOI: 10.17469/O2103AISV000002

SUSANNE FUCHS

Changes and challenges in explaining speech variation: A brief review

In this brief overview I review how our understanding of speech variation has changed over the last decades. I argue that depending on the motivation of the scientific investigations and the theoretical points of view, variation has been discussed with respect to biological, social and communicative factors as well as with respect to the nonlinearities between acoustics, articulation, and perception. Since all of these factors can potentially interact with each other, the core challenges now are to explain the underlying mechanisms, to use appropriate methodologies for the analyses, and to understand how far the underlying mechanisms can be generalized to other speakers or communicative events. I conclude that speech variation is not an obstacle, but rather a rich source that allows us to examine the many facets of language.

Key words: variation, sociophonetics, speaker-specific behaviour, acoustic-articulatory relations.

1. *Changes in explaining speech variation*

The aim of this paper is to provide a general review of how our understanding of speech variation has changed over the last decades, and how speech variation has been explained, depending on the perspective of the researcher. The emphasis lies specifically on acoustic and articulatory variations that have their origin in dialectal, social, communicative and biological factors. Moreover, the nonlinear relations between acoustic, articulation and perception are taken into account. These and other constraints are particularly challenging when discussing speech variation. This review is not intended to provide a summary of studies in this area. Instead, some important papers have been selected which, in my view, mirror how variation has been approached at a certain time.

1.1 Changes in understanding dialectal, social and communicative influences

Finding variation in speech is not novel in itself, it has persisted since the early days of empirical speech research performed using technological equipment. Peterson, Barney (1952) are among the pioneers in the area of acoustics, analysing American English vowel formants of 76 speakers (men, women and children) with different dialectal background and the perception of these vowels by 70 listeners (men, women) with the same dialectal background as the speakers. On the basis of their results they were able to show that speaker productions and listener judgements were influenced by their dialectal background. Furthermore, they provided evidence that variation in the acoustic signal has vowel-specific consistencies and cannot be treated as random noise.

Another famous study focused on acoustic variations that can be attributed to social factors. Labov (1963) investigated the frequencies and distribution of diphthongs on Martha's Vineyard, an island with specific geographical patterns (e.g. up-island: a rural area; down-island: an area with larger towns and tourism) and social structures (four ethnic groups: English descents, Portuguese descents, Indian descents and a mixture of other groups; lowest average income; smallest amount of rich people; highest rate of unemployment in the state with a huge dependency on seasonal work in tourism; highest percentage in the state of married working woman with children). The speech analyses were based on 69 interviews with the islanders. Vowel centralization of the diphthongs was found particularly in inhabitants of the up-island while more open variants occurred in the speech of inhabitants living down-island with a larger influence by the mainland. He also noted that seasonal tourists did not directly affect the islanders' speech production, and differences were grounded in long-term processes of social disparities.

More recently, long-term effects of sound change have also been described in regards to a particularly famous social figure, Queen Elizabeth II of the United Kingdom (Harrington, Palethorpe & Watson, 2000; Harrington, 2006, 2007). These authors used the Queen's Christmas broadcast recordings to analyse changes in her speech between the 1950s and the 1980s. Such long-term analyses are exceptional, because longitudinal data from the same speaker are rarely available. Harrington and colleagues were able to show that from the 1950s to the 1980s, the Queen's vowel productions moved in the direction of English speakers who are younger and lower in the social hierarchy.

Within the last decades, there have also been studies focussing increasingly on speech variation that is due to short-term adaptation of a speaker towards a specific listener and a communicative situation. Such an approach does not exclude the possibility of long-term adaptations. Short-term adaptations have been covered under several terms with different definitions, such as "interactive alignment" (e.g. Pickering, Garrod, 2004), "convergence" (e.g. Manson, Bryant, Gervais & Kline, 2013), "inter-personal coordination" (e.g. Tolston, Shockley, Riley & Richardson, 2014), "speech imitation" (e.g. Garnier, Lamalle & Sato, 2013), "accommodation" (Giles, Coupland & Coupland, 1991), and "entrainment" (e.g. Levitan, Gravano & Hirschberg, 2011). For conversations, it has been shown that interlocutors adapt to each other at various levels. They can adapt in the use of words and their meaning (convergence in lexical and semantic representations, see e.g. Garrod, Anderson, 1987), in the use of syntactic structures (convergence in syntactic representations, see e.g. Branigan, Pickering & Cleland, 2000) and in speech rate or fundamental frequency (phonetic convergence, see e.g. Babel, 2009; Babel, Bulatov, 2012). However, it is still unclear where the origin of this convergence lies and how it emerges between humans, despite their not being physically connected.

One view proposes that convergence in dialogue is the result of shared linguistic representations (Pickering, Garrod, 2004). Pickering and Garrod assume that convergence is grounded in an automatic priming process. Priming refers to a process of increased sensitivity to a certain stimulus due to prior experience, e.g. the prior presentation or production of a word enhances later perceptual identification of this word (Jakoby, 1983). This process is automatic and can even occur in cases in which the task is "not to converge" (Issartel, Marin & Cadopi, 2007).

Such a view is somewhat different from the Communication Accommodation Theory (CAT) by Giles *et al.* (1991). In their framework, the degree of convergence is not automatic, as convergence is seen from a sociolinguistic perspective where interlocutors adjust in the direction of increased mutual similarity to facilitate communication and establish appropriate social distance in the respective communicative situation.

Eckert (2008: 455) proposes, "speaking in the social world involves a continual analysis and interpretation of categories, groups, types, and personae and of the differences in the ways they talk". Social meaning and variation of language is then studied with respect to style, which can flexibly change according to the many factors and functions involved in communicative situations and their interpretation (for further review, see Eckert 2012).

Thus, speech variation has been approached from at least two different perspectives: long-term changes due to, e.g., dialectal background, social status, age, and short-term adaptations to the listener and situation. Long-term adaptations and the respective changes are often discussed in light of sound change phenomena in larger communities and short-term adaptations are seen with respect to the flexibility and individual interpretation of a social agent under the situational circumstances.

1.2 Biological factors explaining speech variation

In his pioneering empirical work, Fant (1966) analysed formants of sustained vowel productions in seven male and seven female speakers of Swedish. On the basis of this speech material, he was able to show that on average, females have formant frequencies 18 percent higher than those of males. This scaling factor is inversely proportional to the vocal tract length of the speaker, i.e. higher formant values go hand in hand with smaller vocal tracts in females in comparison to males. However, Fant also reported vowel-specific effects. Specifically, low back vowels are produced with much higher first formants in females than in males. These production differences among male and female speakers are larger than in any other vowel examined. Fant attributed these findings to the relatively long pharyngeal cavity of male speakers in comparison to female speakers. He also mentioned another potential influence on vowel-specific effect – the smaller laryngeal cavity in females.

A substantial body of work followed Fant's seminal study using the progress in technology and computational power. Fitch, Giedd (1999) used a large sample of 129 people (53 females) with a normal Body-Mass-Index in an age range from 2 to 25 years as part of a larger study on brain development. Body features (weight and height) were taken into account as well as selected measurements based on scans of Magnetic Resonance Images of the vocal tract. Vocal tract length strongly correlated with body features: taller and heavier people had a longer vocal tract. When the effect of body size was removed from the statistical analysis, effects of age and sex remained with specific proportions between the oral and pharyngeal cavity for males and females. Additional analysis revealed that sex-differences in pharyngeal length occur after puberty.

So far, Vorperian, Wang, Schimek, Durtschi, Kent, Gentry & Chung (2011) have analysed the largest sample of which I am currently aware (Magnetic Resonance Imaging scans & computer tomography images of 605 people from birth to 19 years of age), adding more details to the nonlinear development of vocal tract proportions and discussing the implications for speech acoustics. Surprisingly few studies, though, have investigated changes in the upper vocal tract in the elderly, probably because the most significant changes occur from birth to adulthood. However, these changes are crucial particularly in the discussion of sound change when comparing older and younger adults (Reubold, Harrington & Kleber, 2010). Xue, Hao (2003) are an exception, analysing data from 38 younger speakers (20-30 years old) and 38 older speakers (65-87 years old, all healthy). They found a larger oral cavity volume in the elderly, a slightly longer mouth cavity, and no difference in pharyngeal length. Overall vocal tract length did not differ significantly. The larger mouth cavity in the elderly coincided with lowered first formant values. Further empirical work is needed to describe and evaluate vocal tract changes on a continuous scale over the entire lifespan.

From the description of the overall vocal tract shape, more recent studies have focused on specific parts of the vocal tract, e.g. the size and shape of the palate (e.g., Brunner, Fuchs & Perrier, 2009; Fuchs, Toda, 2010; Yunusova, Rosenthal, Rudy, Baljko & Daskalogiannakis, 2012; Lammert, Proctor & Narayanan, 2013; Weirich, Fuchs 2013; Weirich, Fuchs, Simpson, Winkler & Perrier, 2016). For instance, Brunner *et al.* (2009) investigated differences in the coronal plane of the palate and discussed these with respect to speakers' articulatory precision in the production of high vowels and /j/. All these sounds are realized with a considerable amount of tongue-palate contacts. The authors assumed that speakers with a flat palate shape have to limit their articulatory variability in order to keep acoustic variability within a tolerable range, while speakers with a dome-shaped palate are less constrained. These assumptions were generally confirmed by means of electropalatographic data from 32 speakers of different languages. Speakers with a flat palate consistently showed a lower level of articulatory variability than speakers with dome shaped palates (who were either variable or not), while acoustic variability did not differ among the speakers.

The palatal inclination angle has been discussed with respect to the production of the phonemic /s/-/ʃ/ contrast in German (Weirich, Fuchs, 2013). The authors tested their hypotheses on the basis of speech samples of mono- and dizygotic twin pairs as well as a group of unrelated speakers. They were able to provide evidence for speaker-specific articulatory strategies depending on the alveolo-palatal inclination angle. Speakers with a flat angle only retracted the tongue for the postalveolar fricative while speakers with a steep angle additionally elevated the tongue.

The particular shape of the palate has even been discussed as a bias in the development of clicks in Khoisan languages, even if it is clear that speakers can in principle learn every language independently of their vocal tract properties (Moisik, Dediu, 2015). The authors follow up on earlier work that reports a lack of a clear alveolar ridge in speakers of these click languages. Moisik, Dediu suggest that less articulatory effort is needed for the production of clicks when speakers have a smooth palatal profile. They provide evidence for their suggestion by means of simulations with a flat and a steeper alveolar ridge using a 3D biomechanical tongue model.

Different modelling approaches were required for a recent investigation (Weirich *et al.*, 2016) of the degree of jaw opening in males' and females' speech, because it is unclear whether speakers only adapt to their particular vocal tract anatomy or compensate for it. Weirich and colleagues considered the possibility that the longer pharynx in male speakers coincides with a greater distance between the condyle and the gnathion, which in turn also results in a larger jaw displacement. Simulations with prototypical male and female models showed a complete linguo-pharyngeal closure for the male model producing a prominent low vowel, while such a closure was not found in the female model. The authors therefore suggest that in the production of low prominent vowels, males do not open their jaw very widely, because this may cause a linguo-pharyngeal closure if the tongue does not compensate for it. The study is particularly interesting in light of the reported sociophonetic finding that male mumbling comes across as macho (Heffernan, 2010).

Consistent speaker-variation has not only been described at the level of vocal tract morphology, but also in regards to parts which had been considered rather invariant among human subjects for a long time. Golestani and colleagues (Golestani, Price & Scott, 2011) carried out an anthropomorphic analysis of brain structures in speech areas for two different participant groups: expert phoneticians who had received years of training in sound production and perception and a control group with no particular training. Golestani *et al.* were able to find differences in the surface area of the pars opercularis, a structure involved in phonological processing. The size of the area correlated positively within the group of expert phoneticians with years of phonetic training. Thus, training can affect brain plasticity in speech-related areas. The authors also found a greater gyrification in the auditory cortex of the expert phonetician brain, but this did not show any correlation with training; it most likely already develops in utero.

Finally, one might come to the conclusion that even if there can be consistent inter-speaker variation at the level of vocal tract or brain morphology, it might not occur at the genetic level in the non-clinical population. But even at this level, our understanding has changed over the last decades. The popular assumption that genes defining our physical properties are fixed and assigned at birth without any further changes is no longer viable (Dediu, 2015). Furthermore, even if humans are genetically very close to their ancestors, the idea that the mutation of a single gene brought language to the human species (as is often proposed by researchers in generativism) has become nothing more than a fantasy (Dediu, Christiansen, 2016). According to the authors, a certain gene can produce different proteins at different times in different tissues. "The genetic foundations of language and speech are extremely complex and there is no gene 'for' language (Fisher, 2006). Instead, there are many genes interacting in complex regulatory networks tuned to many contextual cues and influencing many aspects of the phenotype. Genes are not monolithic units with simple and clear functions but instead there is pervasive gene regulation at multiple levels and constant interaction with the environment" (Dediu, Christiansen, 2016: 367).

To summarize, consistent biological between-speaker variation can be found at different macroscopic and microscopic levels. They have been discussed in terms of speech motor control, individual differences (Fuchs, Pape, Petrone & Perrier, 2015a), and even language evolution (Dediu, Jannsen & Moisik, 2016). The first robust empirical data focussed on visible and audible differences in vocal tract morphology. More recently, these have been complemented by work looking behind the surface structures and examining, for example, the level of the brain, the level of gene regulation, and the level of biomechanics, adding increasing detail to the complex picture and sometimes challenging theoretical models of speech and language production.

1.3 Relation between different levels

A complicating factor in the description of speech variation comes into play when considering different levels, i.e. acoustics, articulation and perception. In many instances, these levels have a nonlinear relationship and a certain amount of variation at one level does not necessarily coincide with variation at another level.

Different relationships have yielded very influential theoretical concepts in phonetics. In his famous paper on the quantal nature of speech, Stevens (1989) described the non-linear relationship between acoustics and articulation as well as between acoustics and perception in three different regions of the articulatory domain. Regions I and III characterize a relation of relative acoustic stability (with only small changes visible as plateau-like shapes), but substantial articulatory variation. Both regions differ qualitatively with respect to their acoustic values. Region II is a threshold area between regions I and III, characterized by stability in articulation and huge variability in the acoustics. The articulatory-acoustic relations, according to Stevens (see page 5), are quantal in nature and can also be applied to the relation between perception and acoustics. Stevens interprets these relationships in connection with distinctive phonological features. Stable acoustic regions (regions I and III) would be favoured in phonological inventories. Thus Stevens supposes a primacy of invariant or relatively stable acoustic regions over variable articulatory motions for speech production (Blumstein, Stevens, 1979; Stevens 1989).

Further support for this primacy comes from studies on motor equivalence in speech production, following the general motor control principle that several possibilities exist to reach a defined goal. For speech production, trading relations have been shown, e.g., between tongue body raising and lip protrusion during the production of /u/ (among others, Perkell, Matthies, Svirsky & Jordan, 1993; Savariaux, Perrier & Orliaguet, 1995). This leads the authors to infer that goals are defined in the acoustic/auditory domain, because this domain is relatively stable across conditions, while articulatory motions vary for this particular sound.

Another very influential approach discussing the relation between speech production and perception is the theory of hypo- and hyper articulation (H&H by Lindblom, 1990). It rests on empirical evidence that there is no invariance in the speech signal. Lindblom suggests that speech production varies along a continuum between hyper- and hypo articulation. Hyper articulation may be a requirement under certain situational conditions (e.g. the interlocutor is in another room; there is background noise, etc.). The effort manifests on the side of the speaker who must produce a message and increase discriminability for the listener. Hypo-articulation follows the principles of minimal motor effort for the speaker, but goes hand in hand with less perceptual discriminability and increased effort by the listener to understand the message. The H&H theory suggests that flexible adaptations of speakers to reception (social and communicative constraints) and production (physiological and cognitive constraints, see Lindblom 1990: 418) are responsible for the variation between hyper- and hypo-speech.

More recently, Perkell and colleagues (Perkell, Guenther, Lane, Matthies, Stockmann & Tiede, 2004a; Perkell, Matthies, Tiede, Lane, Zandipour, Marrone 2004b) found speaker-specific effects in the relation between speech production and perception. They hypothesized that inter-speaker variation may be attributed to individual perceptual capabilities. Those speakers who are able to accurately discriminate phonemic contrasts are also the ones who could distinctively produce a phonemic contrast. 19 speakers were recorded with electromagnetic articulography and acoustics producing the words *cod*, *cud*, *who'd*, and *hood*. The same speakers underwent a perception experiment with an annotation and a discrimination task of the speech material. Speakers with high perceptual discriminability scores were the ones who produced the vowel contrasts with a larger distance in the acoustic and articulatory space in comparison to the speakers with less perceptual discriminability, providing a reason for speaker variation based on perceptual capacities.

A study by Cangemi, Krüger & Grice (2015) showed that the speaker-specific behaviour in perceptual discriminability and articulatory precision is not a general property of a "universally intelligible speaker" or a "universally proficient listener", but rather that how a speaker behaves with respect to a particular listener depends on the particular dyad, at least for intonational contrasts of focus. Cangemi and colleagues recorded five native speakers of German using electromagnetic articulography with target words under different focus conditions. Lip motions and acoustic pitch accents were analysed. The production experiment was followed by a perception experiment with 20 different listeners judging the productions of the five speakers. The most original finding was that one particular subject can be more intelligible than most others for a particular listener and less intelligible than most other speakers for another listener. Thus, the same subject can be judged on very different ends, depending on the respective dyad.

The complex interplay and the nonlinearities between different levels involved in speech production and perception make an investigation of speech variation challenging, since this variation may be specific to one particular level and may depend on the speech material, the speaker task, and communicative constraints.

In summary, in the last decades researchers dedicated their attempts to the description of nonlinearities between acoustics, articulation, and perception with the primary interest being to find regularities and explain phonemic inventories. More recent studies focus on the flexibility of these nonlinear relations due to speaker-specific or communicative constraints.

2. *Challenges in explaining speech variation*

Researchers trying to explain the variation in speech production and perception face various challenges. I will provide four of them related to the following topics: generalization, intra- and inter-speaker variability, single time point and time series analysis, and multidimensional factors. I will try to offer explanations on the basis of selected examples.

2.1 How far can we generalize?

Experiments are often designed in such a way that a few factors with a few levels are varied and the dependent variables studied. Increasing the number of factors to higher than three has the disadvantage that it is hard to interpret potential interactions among them. If a factor has several levels, additional post-hoc analyses are required to understand the relation between the different levels. In this respect, experimental designs with just a few independent variables and levels are favoured. However, less comprehensive designs may restrict the researcher with regards to the generalizations that can be drawn. In a recent investigation, Koenig, Fuchs (2015) studied the effect of loudness (normal versus loud speech) on German vowel production in different speech tasks. The investigation was motivated by previous efforts which a) focussed to a large extent on low vowels rather than sampling broadly across the vowel space, and b) provided average values for the whole vowel space without being explicit about vowel-, speaker-, or task effects. Moreover, loud speech is often taken as an intervention method for speech therapy (e.g. within the Lee Silverman voice treatment). Understanding how much loud speech affects formant values across several vowels, speakers, and tasks is fundamental to evidence-based treatment.

Figure 1 - *F1 and F2 space (in Hz) consisting of tense German /a, i, u/ vowels, spoken by 11 female speakers in normal (grey) and loud (black) speech conditions; from Koenig & Fuchs (2015)*

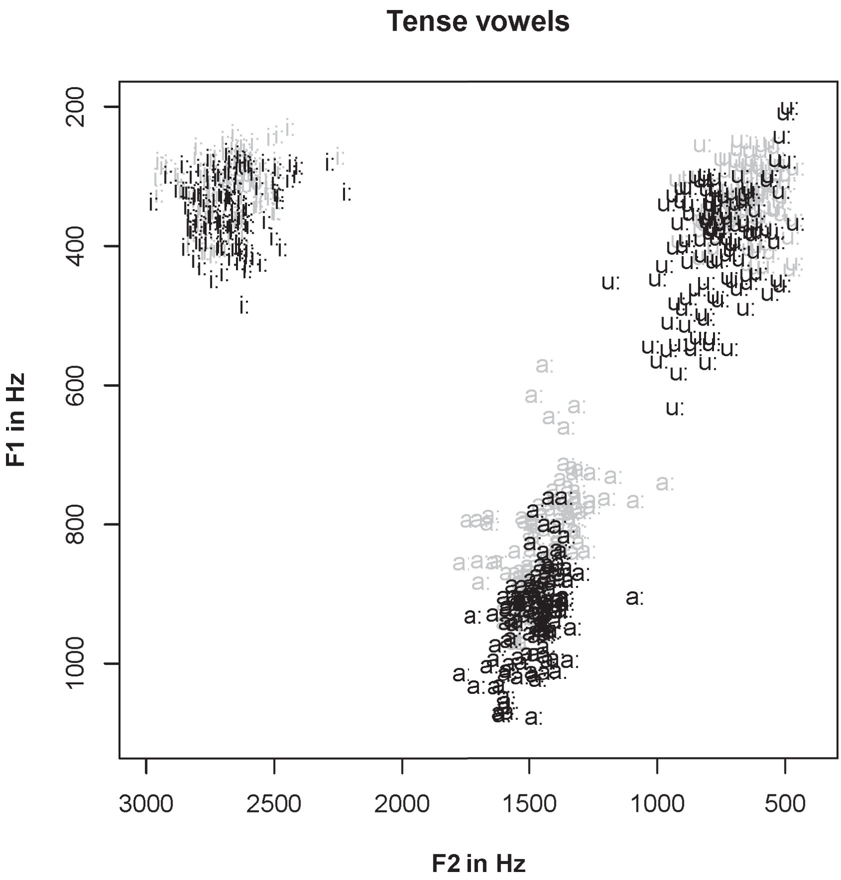
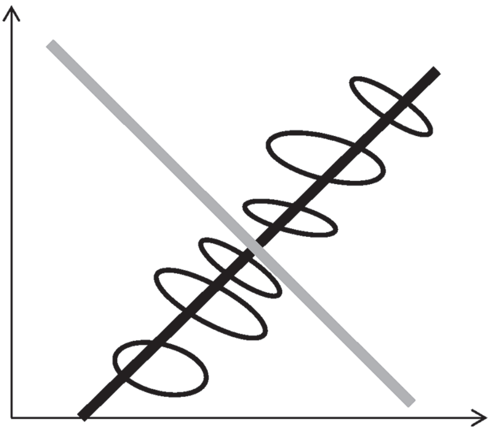


Figure 1 displays the results for tense vowels in a question-answer task. Normal vowel productions are visualized in grey and loud speech in black. Significant differences due to loudness are produced in the first formant of low vowels, while in high vowels these differences are weak or absent. The study is a nice example that investigating only low vowels does not allow generalizations to all other vowels. The often described increase in vowel space (based on the area between the corner vowels /a, i, u/ in loud speech may actually be a result of the low F1 value in /a/ only and not a result of more peripheral vowels in general. Hence, it is advisable to draw conclusions that are based on the specific speech material and task.

2.2 How much are global effects mirrored in intra-speaker variation?

The second example is related to intra- and inter-speaker variation. In most studies, researchers are interested in global behaviour across speakers and a given task, while single speakers' behaviour is treated as random noise. If several data points are recorded from single subjects, the distribution of these must be considered. Figure 2 shows a schematic example for an extreme case. The six black dispersion ellipses correspond to distributions of six different subjects. The main direction of variance in these ellipses clearly shows a negative within-speaker correlation (depicted as the schematic grey regression line). If statistical analysis were to be based solely on average values of all speakers, a positive correlation would be found (depicted as the black regression line). Even though this example may be an extreme case, it makes sense to look at intra- as well as inter-speaker variation, to record several data points for each subject, to graphically explore a potential bias between intra- and inter-speaker effects, and to include speaker-specific slopes in statistical models.

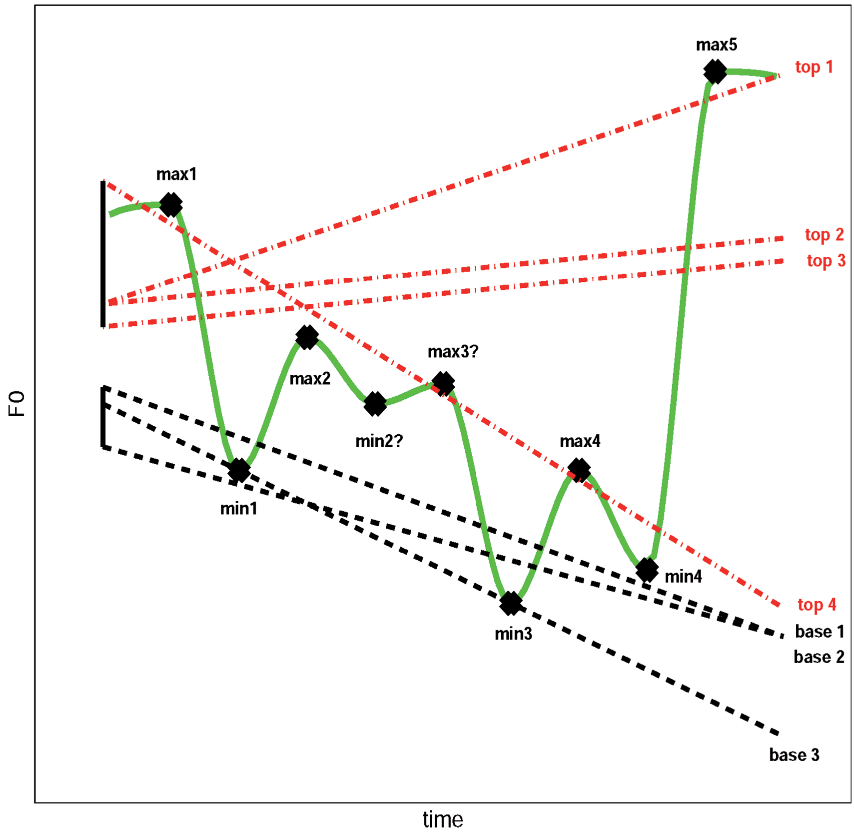
Figure 2 - *Schematic view of intra- and inter-speaker variation of two measured parameters. The dispersion ellipses correspond to variation in six single speakers. The grey regression line schematically displays the variation within speakers while the black regression line to variation between speakers*



2.3 Single time point analyses versus time series analyses

Another important challenge we are currently facing is the choice of either analysing selected temporal landmarks or the entire time series in the temporal window of interest. More and more studies oppose "magic moment measures" (Vatikiotis-Bateson, Barbosa & Best, 2014), because the choice of the selected time points may be primarily driven by certain theoretical concepts. Single time point analysis can be valuable in one case, but unreliable in another, since speech production and perception are complex dynamic processes involving the flow of coordinated articulatory motion and transitory acoustic states. The recent advances in tools able to take the whole time series into account are particularly helpful in this respect. The third example will illustrate this point. The example deals with a phenomenon called f0 declination, reflecting the gradual decrease of the fundamental frequency (see Fuchs, Petrone, Rochet-Capellan, Reichel & Koenig, 2015b, and references therein). F0 declination has been measured in various ways: As the regression line that is fit through all f0 values within a given time interval, as the topline, a regression line going through all high pitch accents, a baseline moving through the low tones, and the midline, a regression line between the top- and the baseline. One of the difficulties for the calculation of top-, base- and midlines is the selection of the local high and low tones, which very much depends on the phonological theory. In Figure 3, it is shown that the slope of the topline and baseline can vary considerably when including or excluding some of the pitch accents. For instance, topline 4 has a negative slope when the boundary tone (max 5) is not included, while all other toplines show a positive slope when the boundary tone is included.

Figure 3 - *Stylistic f0 contour with potential pitch maxima and minima (black dots). Based on the inclusion of certain maxima and minima, different f0 top- and baselines (dashed and dashed-dotted lines) were calculated. This figure was adopted from Reichel, Mády, 2013. Thanks to Uwe Reichel for making it available*



When all data points are taken into account for the calculation of the regression line, the results will be less affected by local peaks and valleys, but may also be subject to noise, due for instance to micro-prosodic perturbations.

2.4 Teasing apart different influences

The largest challenge of all might be to tease apart the different factors (see § 1.1-§ 1.3). Doing so presupposes a researcher or a research team that has a broad background expertise and is aware of all the factors that could potentially have an impact on the observed dependent variables. For example, researchers in sociophonetics should be aware of the biological factors that might also contribute to their findings and know the literature in this area. Reversely, this is also true for researchers who are primarily interested in the biological origin of language. Thus, researchers must critically question at which level they expect variation, how much this expectation is driven by their own theoretical and conceptual thinking, and to what extent they may be "blind" to other areas. Luckily, the existence of vast publication databases allows researchers to carry out comprehensive inter-disciplinary literature reviews reaching beyond their main research area.

Additionally, speaker-specific physically realistic models (e.g. Winkler, Fuchs, Perrier & Tiede, 2011; Stavness, Nazari, Perrier, Demolin & Payan, 2013) are available that help us to better separate articulatory behaviour that is an adaptation to specific vocal tract properties from cases in which speakers compensate for particular properties. It also allows investigation of the complex relations between speaker-specific anatomy, muscle recruitment, articulation and acoustics with the drawback of being time-consuming and computationally expensive.

Finally, open access articles and respectful data sharing among researchers may be a rich source for explaining speech variation from different angles and at the same time cutting costs and reducing effort.

3. *Conclusion*

This review has shown that variability in speech has been found at various levels since the earliest empirical studies on the topic were first conducted. This should not imply that there were no approaches supporting the idea of invariance. However, within the last decades, the body of empirical evidence indicating that variation is everywhere has continued to grow. This is the case not only at the macroscopic level (such as body size and vocal tract length), but also at the microscopic level (such as in brain morphology or genetics). Speakers differ in terms of their body features, language and culture, and social behaviour, and they can flexibly adapt to communicative constraints.

Depending on the focus of their studies, scientists have provided explanations for variation that range from social to communicative to biological in nature. To overcome the boundaries between research disciplines, we should be aware of the limits of our own conceptual thinking when interpreting variation. A comprehensive knowledge of different research perspectives, theoretical plurality, critical thinking, and/or working in interdisciplinary teams are among the factors which could help to allow future work to disentangle potentially co-occurring processes (Fuchs, Lancia, 2016). I believe we now have access to many sources (e.g. multidimensional and multisensory recording techniques, statistics, computational power for processing huge datasets, open access libraries, servers for scientific data sharing), which allow us to proceed in such a direction.

What is the main takeaway from this brief review?

I wish to conclude with something Georg Meyer said at our first summer school "Cognitive and physical models of speech production, perception and perception-production interaction" in Lubmin 2004: "variability is not the enemy, variability is our friend". I would even go a step further and say that variability is a rich source that allows us to examine the many facets of language in detail, something which would be impossible if variation did not exist.

Acknowledgments

This work was supported by the German Bundesministerium für Bildung und Forschung (BMBF) (Grant Nr. 01UG1411) and the Leibniz Society. I would like to thank my colleagues at the Laboratory Phonology Group at ZAS, the editors, Barbara Gili Fivela and my external collaboration partners for the fruitful exchanges, comments and discussions related to these topics. Thanks to Olivia Maky for proofreading.

Bibliography

Babel, M.E. (2009). Phonetic and social selectivity in speech accommodation. PhD Dissertation, University of California, Berkeley.

Babel, M.E., Bulatov, D. (2012). The role of fundamental frequency in phonetic accommodation. In *Language and Speech*, 55(2), 231-248.

Blumstein, S.E., Stevens, K.N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. In *The Journal of the Acoustical Society of America*, 66(4), 1001-1017.

Branigan, H.P., Pickering, M.J. & Cleland, A.A. (2000). Syntactic coordination in dialogue. In *Cognition*, 75, B13-25.

Brunner, J., Fuchs, S. & Perrier, P. (2009). On the relationship between palate shape and articulatory behavior. In *The Journal of the Acoustical Society of America*, 125(6), 3936-3949.

Cangemi, F., Krüger, M. & Grice, M. (2015). Listener-specific perception of speaker-specific productions in intonation. In Fuchs, S., Pape, D., Petrone, C. & Perrier, P. (Eds.), *Individual differences in speech production and perception.* Frankfurt am Main-Berlin-Bern-Bruxelles-New York-Oxford-Wien: Peter Lang Publisher, 123-145.

Dediu, D. (2015). *An introduction to genetics for language scientists*. Cambridge: Cambridge University Press.

Dediu, D., Christiansen, M.H. (2016). Language evolution: constraints and opportunities from modern genetics. In *Topics in Cognitive Science*, 8(2), 361-370.

Dediu, D., Jannsen, R. & Moisik, S.R. (2017). Language is not isolated from its wider environment: Vocal tract influences on the evolution of speech and language. In *Language and Communication*, 54, 9-20.

Eckert, P. (2008). Variation and the indexical field. In *Journal of Sociolinguistics*, 12(4), 453-476.

Eckert, P. (2012). Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. In *Annual Review of Anthropology*, 41, 87-100.

Fant, G. (1966). A note on vocal tract size factors and non-uniform F-pattern scalings. In *STL-QPSR*, 7(4), 22-30.

Fisher, S.E. (2006). Tangled webs: Tracing the connections between genes and cognition. In *Cognition*, 101(2), 270-297.

Fitch, T., Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. In *The Journal of the Acoustical Society of America*, 106(3), 1511-1522.

Fuchs, S., Lancia, L. (2016). Seeing speech production through the window of complex interactions: Introduction to supplement containing selected papers from the Xth International Seminar on Speech Production (ISSP) in Cologne. In *Journal of Speech, Language, and Hearing Research*, 59, S1555-S1557.

Fuchs, S., Pape, D., Petrone, C. & Perrier, P. (Eds.) (2015). *Individual differences in speech production and perception.* Frankfurt am Main-Berlin-Bern-Bruxelles-New York-Oxford-Wien: Peter Lang Publisher.

Fuchs, S., Petrone, C., Rochet-Capellan, A., Reichel, U.D. & Koenig, L.L. (2015). Assessing respiratory contributions to f0 declination in German across varying speech tasks and respiratory demands. In *Journal of Phonetics*, 52, 35-45.

Fuchs, S., Toda, M. (2010). Can biological or sociophonetic factors explain the difference in male versus female /s/. In Fuchs, S., Zygis, M. & Toda, M. (Eds.)., *Turbulent sounds. An interdisciplinary guide*. Berlin: Mouton de Gruyter, 281-302.

Garnier, M., Lamalle, L. & Sato, M. (2013). Neural correlates of phonetic convergence and speech imitation. In *Frontiers in Psychology*, 4, 600.

Garrod, S., Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. In *Cognition*, 27, 181-218.

Giles, H., Coupland, N. & Coupland, J. (1991). Accomodation Theory: Communication, context, and consequence. In Giles, H., Coupland, J. & Coupland, N. (Eds.), *Contexts of Accomodation: Developments in Applied Sociolinguistics*. New York: Cambridge University Press, 1-68.

Golestani, N., Price, C.J. & Scott, S.K. (2011). Born with an ear for dialects? Structural plasticity in the expert phonetician brain. In *Journal of Neuroscience* 31(11), 4213-4220.

Harrington, J. (2007). Evidence for a relationship between synchronic variability and diachronic change in the Queen's annual Christmas broadcasts. In *Laboratory phonology*, 9, 125-143.

Harrington, J. (2006). An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. In *Journal of Phonetics*, 34(4), 439-457.

Harrington, J., Palethorpe, S. & Watson, C.I. (2000). Does the Queen speak the Queen's English? In *Nature*, 408, 927-928.

Heffernan, K. (2010). Mumbling is macho: Phonetic distinctiveness in the speech of American radio DJs. In *American Speech*, 85, 67-90.

Issartel, J., Marin, L. & Cadopi, M. (2007). Unintended interpersonal co-ordination: Can we march to the beat of our own drum? In *Neuroscience Letters*, 411, 174-179.

Jakoby, L.L. (1983). Perceptual enhancement: Persistent effects of an experience. In *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(1), 21-38.

Koenig, L.L., Fuchs, S. (2015). Acoustic effects of loud speech and interrelationships among measures. In The Scottish Consortium for ICPhS (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, UK: the University of Glasgow. Paper number 424.

Labov, W. (1963). The social motivation of a sound change. In *Word*, 19(3), 273-309.

Lammert, A., Proctor, M. & Narayanan, S. (2013). Morphological variation in the adult hard palate and posterior pharyngeal wall. In *Journal of Speech, Language, and Hearing Research*, 56(2), 521-530.

Levitan, R., Gravano, A. & Hirschberg, J. (2011). Entrainment in speech preceding backchannels. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers-Volume 2. Association for Computa tional Linguistics*, 113-117.

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In Hardcastle, W.J. & Marchal, A. (Eds.), *Speech production & speech modeling*. Netherlands: Kluwer Academic Publishers, 403-439.

Manson, J.H., Bryant, G.A., Gervais, M.M. & Kline, M.A. (2013). Convergence of speech rate in conversation predicts cooperation. In *Evolution and Human Behavior*, 34(6), 419-426.

Moisik, S.R., Dediu, D. (2015). Anatomical biasing and clicks: Preliminary biomechanical modelling. In Little, H. (Ed.), *Proceedings of the Satellite event: The evolution of phonetic capabilities: causes, constraints, consequences*, Glasgow, UK, 12 August 2015, 8-13.

Perkell, J.S., Guenther, F.H., Lane, H., Matthies, M.L., Stockmann, E. & Tiede, M. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. In *Journal of the Acoustical Society of America*, 116, 2338-2344.

Perkell, J.S., Matthies, M.L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., Stockmann, E. & Guenther, F.H. (2004). The distinctness of speakers' /s/-/ʃ/ contrast is related to their auditory discrimination & use of an articulatory saturation effect. In *Journal of Speech, Language, and Hearing Research*, 47, 1259-1269.

Perkell, J.S., Matthies, M.L., Svirsky, M.A. & Jordan, M.I. (1993). Trading relations between tongue-body raising & lip rounding in production of the vowel /u/: a pilot motor equivalence study. In *Journal of the Acoustical Society of America*, 93, 2948-2961.

Peterson, G.E, Barney, H.L. (1952). Control methods used in a study of the vowels. In *Journal of the Acoustical Society of America*, 24(2), 175-184.

Pickering, M.J., Garrod, S. (2004). The interactive-alignment model: Developments and refinements. In *Behavioral and Brain Sciences*, 27(2), 212-225.

Reichel, U.D., Mády, K. (2013). Parameterization of F0 register and discontinuity to predict prosodic boundary strength in Hungarian spontaneous speech. In Wagner, P. (Ed.), *Elektronische Sprachverarbeitung. Studientexte zur Sprachkommunikation*. Dresden: TUDpress, 223-230.

Reubold, U., Harrington, J. & Kleber, F. (2010). Vocal aging effects on F0 and the first formant: a longitudinal analysis in adult speakers. In *Speech Communication*, 52(7), 638-651.

Savariaux, C., Perrier, P. & Orliaguet, J.P. (1995). Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: a study of the control space in speech production. In *Journal of the Acoustical Society of America*, 98, 2428-2842.

Stavness, I., Nazari, M.A., Perrier, P., Demolin, D. & Payan, Y. (2013). A biomechanical modeling study of the effects of the orbicularis oris muscle and jaw posture on lip shape. In *Journal of Speech, Language, and Hearing Research*, 56(3), 878-890.

Stevens, K.N. (1989). On the quantal nature of speech. In *Journal of Phonetics*, 17, 3-45.

Tolston, M.T., Shockley, K., Riley, M.A. & Richardson, M.J. (2014). Movement constraints on interpersonal coordination and communication. In *Journal of Experimental Psychology: Human Perception and Performance*, 40(5), 1891-1902.

Vatikiotis-Bateson, E., Barbosa, A.V. & Best, C.T. (2014). Articulatory coordination of two vocal tracts. In *Journal of Phonetics*, 44, 167-181.

Vorperian, H.K., Wang, S., Schimek, E.M., Durtschi, R.B., Kent, R.D., Gentry, L.R. & Chung, M.K. (2011). Developmental sexual dimorphism of the oral and pharyngeal portions of the vocal tract: An imaging study. In *Journal of Speech, Language, and Hearing Research*, 54, 995-1010.

Weirich, M., Fuchs, S. (2013). Palatal morphology can influence speaker-specific realizations of phonemic contrasts. In *Journal of Speech, Language, and Hearing Research*, 56(6), S1894-S1908.

Weirich, M., Fuchs, S., Simpson, A., Winkler, R. & Perrier, P. (2016). Mumbling: Macho or morphology? In *Journal of Speech, Language, and Hearing Research*, 59(6), S1587-S1595.

Winkler, R., Fuchs, S., Perrier, P. & Tiede, M. (2011). Speaker-specific biomechanical models: From acoustic variability via articulatory variability to the variability of motor commands in selected tongue muscles. In Laprie, Y. (Ed.), *Proceedings of the ISSP Montreal*, 219-226.

Xue, S.A., Hao, G.J. (2003). Changes in the human vocal tract due to aging and the acoustic correlates of speech production: a pilot study. In *Journal of Speech, Language, and Hearing Research*, 46(3), 689-701.

Yunusova, Y., Rosenthal, J.S., Rudy, K., Baljko, M. & Daskalogiannakis, J. (2012). Positional targets for lingual consonants defined using electromagnetic articulography. In *The Journal of the Acoustical Society of America*, 132(2), 1027-1038.