The handle https://hdl.handle.net/1887/3505668 holds various files of this Leiden University dissertation.

Author: Puggaard-Rode, R.
Title: Stop! Hey, what's that sound? the representation and realization of Danish stops
Issue Date: 2023-01-11

Stop! Hey, what's that sound? the representation and realization of Danish stops

Puggaard-Rode, R.

CHAPTER 7

Conclusion

7.1 Towards a synthesis

This dissertation has had two main goals: to provide a complete overview of the existing literature on Danish stops, and to fill some specific gaps in our knowledge. In this chapter, I aim to synthesize the resulting research, and pinpoint some areas for future research. In what follows, I assess the impact of the dissertation from five different perspectives: history, phonetics, phonology, variation, and data analysis. The first four of these link back to the topical subdivision made in the literature review in Chapter 2; data analysis is a more general perspective, but it has been equally impactful for the work presented in the dissertation. In Section 7.2, I give a concise chapter-by-chapter summary of the dissertation, and in the sections that follow, I discuss each of these five perspectives in turn.

7.2 Chapter-by-chapter summary

In Chapter 1, I present and motivate the subject area of the dissertation. I also provide a general overview of the Danish language, and

introduce the transcription conventions used throughout. The rest of the dissertation is divided into two parts: Part I gives a 'big picture' perspective on Danish stops, and Part II consists of corpus studies on more narrow topics.

Chapter 2 provides a comprehensive overview of the existing literature on Danish stops. As mentioned above, this overview is divided into four perspectives: history, phonetics, phonology, and variation. This chapter mostly sticks to summarizing the literature, but also provides some critical discussion of abstract underlying consonant clusters proposed in the literature, as well as the predictions that follow from some approaches to the representation of laryngeal contrast.

Chapter 3 deals with Danish consonant gradation, the phenomenon which is a.o. responsible for alternations between unaspirated stops [p t k] and semivowels [u̯ ɣ ɹ]. The traditional phonological analysis of Danish consonants holds that these are positional allophones of the phonemes /b d g/. The chapter isolates a number of problems with this analysis: it results in rampant neutralization, it is phonetically 'unnatural', and the morphophonological evidence in favor of the analysis is weak, since it relies exclusively on irregular and unproductive morphology. An alternative analysis is proposed, where [u̯ ɹ] are always considered allophones of /v j/. The traditional analysis does account for some structural generalizations that are not accounted for in the alternative analysis; it is argued that these are best understood as the result of well-known phonetic pressures operating during earlier stages of the language's history.

Chapter 4 presents a corpus study of intervocalic closure voicing in Danish stops – the first empirical study of this topic. Some previous research assumes that intervocalic voicing is essentially categorical, while other research assumes that intervocalic voicing is essentially non-existent; both of these positions have come up in discussions of the phonological representation of laryngeal contrast in Danish and beyond. The study finds that intervocalic voicing is very negligible in Danish /p t k/, and quite rare in /b d g/, presumably because /b d g/ are produced with an active glottal spreading gesture which serves to block voicing. In statistically modeling the data, I tested how a range of different variables affect the rate of closure voicing, and concluded that it is first and foremost a lenition phenomenon. The results are predicted

by a gesture-based approach to phonological representation (such as Articulatory Phonology); other approaches which have made explicit predictions about Danish stops either incorrectly predict, or are unable to predict, aspects of the results.

Chapter 5 presents an exploratory corpus study of the spectral characteristics of Danish aspirated stop releases, focusing on how they change through time, and how they are affected by phonetic context. It is often brought up in the literature that /t/ is highly affricated, but previous descriptions have all been impressionistic or introspective. /p k/ releases are rarely mentioned. The results indeed show that /t/ is affricated across all phonetic contexts, but also that affrication is gradually replaced by aspiration towards the end of releases; the proportion of affrication and aspiration is modulated by phonetic context. /k/ releases are highly affected by phonetic context, presumably since the precise place of articulation in velar consonants is very variable. /p/ releases are also affected by phonetic context, but they are more prone to interspeaker variation than contextually deter-mined variation. The main contribution of Chapter 5 is methodological. The data are fitted to function-on-scalar regression models using the entire spectrum as the response variable. This is a method which arguably overcomes a more general problem faced by researchers in analyzing the complex multidimensional information in noisy spectra, and which allows for intuitive visualization of how the spectrum changes through time under different conditions.

Chapter 6 combines two corpus studies of stop realization in the traditional regional varieties of Danish spoken on the Jutland peninsula. One study investigates variation in voice onset time, and the other investigates variation in the spectral characteristics of aspirated stop releases. The chapter also briefly discusses closure voicing in northern Jutlandic varieties, where voicing is much more widespread than in Modern Standard Danish. Voice onset time is generally shorter in traditional Jutlandic varieties than in Modern Standard Danish, and shows complex regional patterning, which is accounted for with a combination of social and historical factors. /t/ releases in particular also vary in their release characteristics; the vast majority of the peninsula has not adapted the prominent /t/-affrication of Modern Standard Danish discussed in Chapter 5. This

chapter models some complex dynamic variables: the regional variable is approached with generalized additive mixed models, and the spectral variable is approached with functional principal component analysis, which is adapted to solve some of the same problems as function-on-scalar regression in Chapter 5.

7.3   History

Inspired by Evolutionary Phonology, the dissertation has suggested phonetic accounts of the series of sound changes that eventually resulted in stop gradation, i.e. the process whereby voiceless unaspirated stops [p t k] in some contexts came to alternate with semivowels [u̯ x̣ ɪ̯] in Modern Standard Danish. Evolutionary Phonology builds on the neogrammarian insight that previous stages of languages were subject to the same phonetic pressures as synchronic languages; in the past century, our understanding of those phonetic pressures has drastically improved. While the account in Section 3.5 concerns language history, much of the explanation comes from modern experimental phonetics.

I argue in Section 3.5 that Danish had voiced stops in a previous stage of the language. This is not a common assumption in Germanic historical linguistics, but it makes it much easier to account for subsequent sound changes. I proposed that the sounds which currently alternate between [p t k]–[u̯ x̣ ɪ̯] were previously voiced stops [b d g] in all positions; the voiced stops gradually weakened in post-vocalic position, and later lost their voicing in pre-vocalic position. In other words, there was a decrease in sonority in pre-vocalic position, and a drastic increase in sonority in post-vocalic position. Both changes can be considered reactions to the well-known articulatory pressure against obstruent voicing.

Closure voicing at an earlier stage was arguably a prerequisite for the post-vocalic lenition patterns. This idea finds support in data from regional varieties of Danish. The varieties traditionally spoken in northern Jutland have short voice onset time across the board and much more widespread closure voicing than Modern Standard Danish; they also display extensive lenition in post-vocalic position.

Meanwhile, varieties traditionally spoken in southern Jutland have long voice onset time across the board and generally avoid closure voicing; here, stops historically weakened to voiceless fricatives in post-vocalic position (if they weakened at all). In other words, consonant gradation affected all varieties, but varieties with more widespread voicing were subject to lenition in a more 'sonorous' direction. Since consonant gradation in Modern Standard Danish resulted in drastically higher sonority, this variety likely also had more widespread closure voicing during the early stages of gradation.

7.4   Phonetics

This dissertation has shown that intervocalic voicing is relatively rare in Modern Standard Danish spontaneous speech, and has provided multidimensional analyses of how the sound spectrum changes over time in different phonetic contexts during aspirated stop releases. The studies of intervocalic voicing and spectral changes both focus on the acoustic signal, but explanations of the data rely on the articulation–acoustics link, and draw extensively on previous studies of e.g. glottal activity during stop production.

Although intervocalic voicing in Danish has never been studied empirically before, it has often been discussed in the literature, and has been used in support of (sometimes opposing) proposals about the representation of laryngeal contrast in Danish and beyond. The study reported in Chapter 4 shows that intervocalic voicing is relatively rare even in /b d ɡ/. I tested how a range of factors affect the likelihood of voicing, and found that voicing patterns as a lenition phenomenon. Earlier articulatory research has shown that /b d ɡ/ are produced with an active glottal spreading gesture, and this gesture has been argued to enforce voicelessness. Building on these previous studies, Chapter 4 argues that voicelessness is the default setting in Danish stops across all contexts; when they are sometimes voiced intervocalically, this is due to reduction of the glottal spreading gesture in lenition-prone environments.

The study of intervocalic voicing shows that aerodynamic variables (such as vocalic environment) have little effect on whether stop

releases are voiced throughout. This is somewhat unexpected, and may be due to the binary voicing variable being too coarse-grained. I assume that aerodynamic variables do affect the relative proportion of closure voicing, and that this could in principle be tested in a study with a continuous measure of voicing. The statistical model in such a study would be similar to the one reported in Chapter 4 (although the regression would be linear and not logistic), but the data processing would be significantly more time-demanding.

In the terminology of Kingston and Diehl (1994), who propose a dichotomy between 'automatic' and 'controlled' phonetics, the patterns of intervocalic voicing are surely controlled. Generally speaking, voicing is phonetically natural in intervocalic position. Voicelessness in Modern Standard Danish is managed with a small glottal spreading gesture which is not found in comparable languages, and voicing is found in higher rates in other regional varieties of Danish. These observations strongly suggest that when voicing is blocked intervocalically in Modern Standard Danish /b d ɡ/, this has to reflect cognitively controlled, learned behavior.

The dissertation also includes the first acoustic study of affrication in Danish aspirated stops in Chapter 5; previous discussions of this topic have been impressionistic. The results show that the distribution and prominence of place cues throughout releases vary across different places of articulation. Furthermore, /p t k/ respond quite differently to phonetic context. Little attention has been paid to /p k/ previously. /p/ mostly shows evidence of vowel–consonant coarticulation during the beginning of the release, and the energy distribution in the spectrum is highly speaker-specific. /k/ shows prominent signs of vowel–consonant coarticulation throughout the release. /t/ is, as often assumed, invariably affricated; however, unlike the impression given by most previous descriptions, affrication almost always gives way to aspiration well before the end of the release.

Affrication is likely controlled separately from aspiration to some extent, but the two are not entirely independent. Compared to aspirated stops in other languages, Danish /p t k/ are produced with short closure duration and with a glottal spreading gesture which peaks late. These factors may contribute to the prominent /t/-affrication. Longer closure duration and early glottal spreading

serves to ensure high intraoral air pressure, which in turn ensures a prominent burst; if these mechanisms are not available in Danish, an affricated release may indeed be the result. As such, /t/-affrication can partially be explained as a physical reaction to differences in glottal behavior. Like most sound patterns, however, /t/-affrication likely follows from a combination of physical pressures (automatic phonetics) and learned behavior (controlled phonetics); targeted articulatory studies would be necessary to determine the balance between the two.

In Chapter 5, I also provided measurements of voice onset time from aspirated stops in stressed and unstressed syllables. These measurements were not the focus of the chapter and were not discussed much, but they are worth briefly returning to. Previous studies of Danish voice onset time have focused on stressed syllables only. The measurements in Chapter 5 unsurprisingly show that voice onset time is much shorter in unstressed syllables. This further supports the claim that the magnitude of glottal gestures is reduced in unstressed syllables. This mechanism also causes a higher rate of intervocalic voicing in /b d ɡ/ when they are unstressed. The magnitude of glottal spreading in /b d ɡ/ is modest, so the gesture may be elided entirely in unstressed position, paving the way for intervocalic voicing; the magnitude is greater in /p t k/, so reducing this gesture will usually only result in a shorter aspiration phase.

7.5   Phonology

It has been discussed at several points in the dissertation how different approaches to phonology make different predictions about the phonetics–phonology interface. In some frameworks, phonetic patterns are not assumed to influence phonology at all; in others, phonology and phonetics are intimately connected. In what follows, I will proceed from the assumption that they are closely connected, and that phonetic evidence is useful in discussions of phonological representation.

Chapter 3 argued against the traditional analysis of Danish consonants, which assumes that [ʊ̯ ɪ̯] are both possible positional allophones

of /g/, and that /b/ in some contexts displays stylistic alternations between [ʊ̯ ~ p]. I proposed an alternative analyses where [ʊ̯] is always an allophone of /v/, and [ɹ] is always an allophone of /j/. The traditional analysis requires the /g/ phoneme to be underspecified with regards to both place, manner, and voicing, and also poses synchronically active alternations which are difficult to derive. Perhaps most gravely, it relies exclusively on alternations found in irregular and unproductive morphology; even if we accept that language learners have no problem establishing the unnatural phonological processes, the evidence they encounter in favor of such an analysis is scarce.

The discussion in Chapter 3 is largely theoretical, but it begs several questions which are in principle testable and empirical, particularly with regards to whether speakers categorize [k ʊ̯ ɹ] as the same phoneme. There are psycholinguistic experiments which are well-suited for gauging this type of categorization, including concept formation experiments (e.g. Ohala 1983b), acceptability judgment experiments (e.g. Ohala and Ohala 1986), and perceptual similarity experiments (e.g. Flege et al. 1994). I discussed in Section 3.3.1 how computer simulations in the framework of Bidirectional Phonetics and Phonology could be used to test the feasibility of establishing the traditional analysis on the basis of the data presented to language learners, and building on ongoing work in BiPhon on the phonology–orthography interface may help to determine whether speakers' knowledge of orthography affect how they analyze the alternations (Hamann and Colombo 2017).[1]

In the alternative analysis of stop gradation, I remained agnostic as to whether [χ] can be considered an allophone of /d/, as proposed in the traditional analysis. The evidence in favor of such an analysis is limited, but unlike the proposed /g/ phoneme, it does not result in problematic neutralizations. [χ] has to be established by learners as either 1) an allophone of /d/, or 2) a separate phoneme /χ̞/. Due to the lack of morphophonological evidence in favor of the first analysis, it can only be maintained if it is phonetically grounded;

---

[1]The proposals for future research given here were included in a funding application with Camilla Søballe Horslund as principal investigator, and should be attributed to her.

whether or not this is the case remains an open question, which may be resolved with future articulatory studies. Several sources have mentioned a secondary dorsal or pharyngeal gesture in the articulation of [x̠], and Siem (2019) showed that this gesture is sometimes more prominent than the coronal gesture. If [x̠] is indeed primarily dorsal with a secondary coronal gesture, this would be evidence in favor of a separate /x̠/ phoneme. The nature of the relevant gestures can be examined with e.g. ultrasound tongue imaging, which is a relatively inexpensive method often used to investigate unusual tongue shapes (e.g. Lawson et al. 2013; Mielke 2015), or with real-time magnetic resonance imaging, which is much more costly but also captures more precise spatial information (Carignan et al. 2020).

In Chapter 4, I discussed how different approaches to the underlying representation of laryngeal contrast make different predictions about Danish intervocalic stop voicing. Representational approaches with more phonetic integration generally fare best. Approaches which assume an abstract [voice] feature make no useful predictions about intervocalic voicing; approaches which assume a [spread glottis] feature can explain why intervocalic voicing is very rare in /p t k/, and perhaps also why it is fairly rare in /b d g/; approaches which use articulatory gestures as phonological primitives can straightforwardly account for the results of the corpus study, including why voicing in /b d g/ patterns as lenition.

7.6 Variation

Variation is most explicitly discussed in Chapter 6, which covers regional variation, but it has been relevant in all chapters. For example, speakers of Modern Standard Danish vary in whether they have active [p ~ w̥] alternations, and if so, which lexical items allow for the alternation; this was relevant for the discussion in Chapter 3. Chapter 4 finds that speaker age is a significant predictor of intervocalic stop voicing, with voicing being less common among older speakers. Chapter 5 finds that speaker sex is often a strong predictor of the spectral composition of stop releases, with female speakers having more affricated releases. Additionally, all statistical models

have included random variables filtering out speaker-specific information, due to the baseline expectation that results will show some degree of systematic interspeaker variation.

The patterns of regional variation are too complex to recap here, but it was shown that there is systematic variation in both voice onset time and release characteristics, and seemingly also closure voicing. Some of this comes down to differences in laryngeal activity during stop closures, but there are likely also other systematic articulatory differences at play. Danish stops show regional variation not just in phonetics, but in history and phonology as well: regional varieties have been subject to different historical developments, which sometimes led to differences in phonology.

The phonetic results in Chapter 6 are useful in determining whether the phonetic results in earlier chapters reflect automatic or controlled behavior. Voice onset time in /p t k/ shows complex regional variation, which indicates that the duration of aspiration is cognitively controlled in a highly granular way. The differences in constraints on voicing in some varieties cement that intervocalic voicing patterns are to some extent controlled. The spectral characteristics of stop releases also show variation, particularly in /t/. As discussed in Section 7.4 above, this may be partially due to differences in glottal behavior, but the control of glottal behavior and affrication is presumably partially independent.

An open question concerns the current status of variation in the Jutland peninsula. While some studies have argued that non-standard Danish dialects are essentially extinct, more recent studies show that they are still spoken to varying extents in some of the areas that were explicitly discussed in Chapter 6, also by younger speakers. My intuition is that the unaffricated 'dry t' is still used in northern Jutland, but it would require targeted fieldwork to uncover whether this is the case, whether its use is socially stratified, and whether it is treated as an overt regional feature.

7.7   Data analysis

Many scholars are currently grappling with how to solve deep-seated structural problems in quantitative research. In this section, I briefly outline some of these currents, and reflect on how I have tried to take them into account in my approach to data analysis.

Quantitative research is facing a replication crisis. This was famously pointed out in a large-scale replication study of 100 major psychological findings, where the authors were able to replicate less than half (Open Science Collaboration 2015). The problem is not isolated to psychological research, and has also been discussed for several other scientific disciplines, including linguistics (Roettger and Baer-Henney 2019). It is a gnarly issue, perhaps especially prevalent in confirmatory experimental research, where a significant p-value at $\alpha = 0.05$ has often been a *de facto* threshold for publication. This potentially rather lenient 'filter' is not a guarantee that a study will replicate (Vasishth et al. 2018a), especially since the coveted $p < .05$ can almost always be achieved with some flexibility in data collection and analysis (Simmons et al. 2011; see Stefan and Schönbrodt 2022 for a recent overview of so-called 'p-hacking' strategies). Current academic incentive structures are favorable towards publishing high volumes of novel significant findings (Smaldino and McElreath 2016); in addition to promoting questionable research practices in general, other adverse consequences of this include a bias in the publication record against findings that are not considered 'impactful', such as null results and replication studies (Roettger 2021), and a tendency to reframe exploratory studies as confirmatory during the publication process (Roettger et al. 2019).[2]

Researchers are increasingly taking steps to mitigate these problems, both at the individual and institutional levels. In linguistics, preregistering research (i.e. specifying procedures for data selection and analysis prior to collection) is gaining popularity (Roettger 2021); linguists are increasingly adapting Bayesian approaches to data analysis, which counteracts the dependence on p-values to quantify significance (Nicenboim and Vasishth 2016; Vasishth et al. 2018b);

---

[2]Strictly speaking, studies are only confirmatory if a single statistical model, which was motivated prior to data collection, is fitted to the data (Baayen et al. 2017).

meta-analytic methods are adopted to quantify the robustness of established findings (e.g. Nicenboim et al. 2018; Cristia et al. 2020), and large-scale international collaborations also seek to test how robust our methods and results are, including the ManyBabies Consortium (e.g. 2020) working on first language acquisition in infancy, and the currently ongoing Many Speech Analyses project. At a more institutional level, journals are increasingly adapting policies mandating open sharing of data and code for quantitative studies, which can drastically improve analytic reproducibility (e.g. Hardwicke et al. 2018).

Not all of these solutions are approachable in early career research; preregistration, large-scale collaborations, and focus on less 'impactful' research are obviously more readily available to established researchers in permanent positions, in large part due to the aforementioned incentive structures. Luckily, Roettger (2019) also proposes some best practices that are in principle available to anyone. These can largely be summarized under the umbrella of *transparency*.

Throughout the dissertation, I have sought to be as transparent as possible with regards to analytical decision-making. In describing the corpus studies in Chapters 4–6, I go into detail about each step of the analytical process, including token selection, acoustic analysis, technical details about statistical models, and interpretation of results. I do not go into much detail about how analyses are implemented in practice, i.e. how they are coded, since these details are liable to become obsolete with software changes. This information, however, is included in online appendices in the *DataverseNL* repository (Puggaard-Rode et al. 2022b; Puggaard-Rode 2022a). These include all materials needed for analytical reproducibility.[3] All code is written in open-source software, viz. Praat (Boersma 2001; Boersma and Weenink 2021) and R (R Core Team 2021; RStudio Team 2022). All R code is further written in Markdown format, meaning raw code and output are interspersed with text, where I aim to illustrate and motivate each analytical decision (providing sources if necessary), and describe the purpose of each

---

[3]The audio recordings themselves are not included. I do, however, provide enough information for the individual recordings to be findable for anyone with access to the corpora.

command used in the code. In principle, this allows anyone to evaluate the viability of the analyses. Additionally, it may serve as inspiration for others who need to carry out similar analyses.

The quantitative studies in this dissertation are largely exploratory; I have not aimed at answering simple yes/no questions. While I do report p-values in most chapters, I have tried not to rely on them as a proxy for practical significance; instead, in most cases, I have discussed coefficients and the certainty attributed to them. Ideally, confirmatory studies should be carried out to test whether the observations made in the dissertation replicate. In the case of Chapters 4–5, this would involve collecting new semi-spontaneous speech data with a more balanced group of participants with regards to age and gender, and setting well-motivated criteria for statistical modelling and sample size prior to the analysis. This would be relatively straightforward for the intervocalic voicing results, but less so for the affrication results; the observations made in Chapter 5 may not be suitable for null hypothesis significance testing. In the case of regional variation, it would be possible to use different portions of the same recordings as in Chapter 6, since only a small subset of the corpus is used in the analysis; gathering new data of the traditional variants is probably not possible.

Logistic mixed-effects regression models (see Chapter 4) are well-known in linguistic research, and generalized additive mixed models (see Chapter 6) have become very widely used in the past half decade, although their use in spatial variation has been relatively limited. These methods require no further discussion. On the other hand, functional principal component analysis (see Section 6.7) has only been sparsely used in linguistics, and to my knowledge has not previously been used in the analysis of spectral variance. Function-on-scalar regression (Chapter 5) is a completely new method in linguistics. The use of functional data analysis in analyzing speech spectra is a major contribution of this dissertation. I believe these methods can be used to overcome many problems researchers currently face with selecting variables to reflect the complex multidimensional information of the spectrum. Measures derived from the spectrum are used for many kinds of research question, so the potential of these methods go well beyond analyzing stop releases. An important avenue for further research is to develop best practices for selecting, fine-tuning, and

interpreting these models, and to determine how they are best used for hypothesis testing.