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

Puggaard-Rode, R.

CHAPTER 6

Regional variation in stops

6.1 Introduction

An overt feature of varieties of Jutlandic Danish is the use of a variant of /t/ known colloquially as *tørt t*, 'dry t'. As we saw in the previous chapter, the Standard Danish variant of /t/ invariably has an affricated release; the dry t does not. There are different opinions in the literature as to which varieties use the dry t, ranging from "all the country's dialects" (Brink and Lund 1975: 353), western Jutlandic (Petersen et al. 2021: 156ff.), conservative Standard Danish (Grønnum 2005: 51), and northern Jutlandic (Petersen 2009b); see Section 2.5.3. I have previously shown in a pilot study that variation in the realization of /t/ goes beyond just affrication, and is difficult to delimit geographically: the dry t also has shorter voice onset time (VOT) than affricated variants, and a short variant of /t/ with little affrication is found in both the

Parts of this chapter report collaborative work with Yonatan Goldshtein. Section 6.2.2 is partially based on Goldshtein and Puggaard (2019), and Section 6.5 is a revised version of Puggaard (2021). The account presented here has been developed throughout a number of presentations (Puggaard 2019b, 2019c, 2020a, 2020b; Puggaard and Goldshtein 2019, 2020). The audio data is freely available (DS 1971–1976), as are replication data and code (Puggaard-Rode 2022a).

north and the center of the Jutland peninsula (Puggaard 2018a). In this chapter, I expand on that study with much more data, a more suitable statistical methodology, and a broader focus on geographical variation in stop realization. I focus specifically on VOT and spectral characteristics, comparing VOT measurements and spectra from speakers throughout Jutland. The data come from a large corpus of legacy recordings which preserves an older stage of regional variation in Danish (Andersen 1981b; Pedersen 1983). Parts of the corpus have been used as a source for the Dictionary of Insular Dialects (ØMO 1992–; cf. Gudiksen and Hovmark 2008), but the recordings from Jutland have never before been used systematically for research.

A section of this chapter is dedicated to introducing this corpus, *Dialektsamlingen* 'the dialect collection' (DS 1971–1976). This means that I will briefly divert from the dissertation's main focus in the first sections of this chapter. I do this for two reasons: First, an understanding of the dialect situation in Denmark and the historical context for the recordings is important both in order to understand the results of this chapter, and in order to understand why a study like this is needed in the first place. Second, this is the first introduction to the corpus written in English.

There are many descriptions of Danish dialects, including partial dictionaries, grammars, (morpho-)phonological descriptions, and topical descriptions of individual dialects (for an overview, see Ejskjær 1993). There are also holistic descriptions of the Danish dialect landscape which define geographical boundaries between dialects, primarily on the basis of isogloss bundles (Bennike and Kristensen 1898–1912; Brøndum-Nielsen 1927; Skautrup 1944–1970: IV). With few exceptions, however, the descriptive work has lain dormant since the 1970s, leaving much of the existing work theoretically dated (see Section 6.2.1).[1] As a consequence, methodological progress in acoustic phonetics has barely improved our knowledge of regional phonetic variation (although see Ejstrup and Hansen 2003; Ejstrup 2010). Our knowledge of phonological variation is rich if spotty, while our

[1]The lexicographic work, however, is still very much ongoing, centered around *Jysk Ordbog* 'Jutlandic dictionary' (Skautrup et al. 1970–; Hansen 2020), and *Ømålsordbogen* 'Dictionary of insular dialects' (ØMO 1992–.)

knowledge of subphonemic systems is much poorer and mostly limited to impressionistic description. Similarly, the recent great strides in statistics for corpus phonology have not been applied to regional variation in Danish.

The initial hypothesis of the research reported here, following Puggaard (2018a), is that the dry t is found in large parts of Jutland, and that cues to place contrasts and laryngeal contrasts show geographical variation. A number of theoretically motivated hypothesis follow.

Chodroff and colleagues (Chodroff and Wilson 2017; Chodroff et al. 2019) have recently shown that variation in VOT tends to co-vary across laryngeal settings and places of articulation. With this in mind, I hypothesize that variation is not limited to /t/, but rather that all stops follow similar patterns of variation. Early findings in VOT research showed that voiced, voiceless, and aspirated stops are cross-linguistically consistent categories (e.g. Lisker and Abramson 1964). Later research has not been able to support this neat, constrained three-way division. As VOT is measured in more languages, typological variation in VOT increasingly looks continuous rather than categorical (Cho and Ladefoged 1999; Ladd 2011), suggesting that the only principal limits on variation in VOT comes from limits on perceptual acuity. The Jutlandic data were gathered in a relatively small geographical area shared by one language community with the same set of distinctive stops, so it is an interesting test case for limits of variation in VOT. I use generalized additive mixed modeling to investigate which aspects of the observed variation are attributable to geography, with the assumption that this relationship is non-linear (Wieling et al. 2011, 2014); using this method also makes it possible to test a number of other hypotheses about the influence of contextual variables on VOT.

Chodroff and Wilson (2018) do not find strong signs of covariation in stop release characteristics across places of articulation, but the study presented in Chapter 5 makes it interesting to test whether there are differences in affrication patterns of /p t k/ across varieties, and whether the Jutlandic varieties pattern similarly to Modern Standard Danish. Comparing spectral characteristics across varieties is much less straightforward than comparing VOT. Using function-on-scalar regression as in Chapter 5 is technically possible, but the added complexity of non-linear geographical variation would make the results prohibitively difficult to interpret. I solve this by focusing on the entire spectrum, but only at the midpoint of stop releases, i.e. removing variation in time from the equation. I use a combination of functional principal components analysis (FPCA) and generalized additive mixed modeling to investigate the main sources of variance in release spectra, and how these relate to geography. To the extent of my knowledge, this is the first study to use FPCA to explore spectral variance. As with the model of VOT, this also allows for testing hypotheses about the influence of contextual variables on spectral shape, and whether the influence of contextual variables is similar in Modern Standard Danish and in the traditional Jutlandic varieties.

Throughout this dissertation, a lot of weight has been placed on the relationship between (known) data about articulatory mechanisms, (new) corpus data revealing acoustic patterns, and how the articulation–acoustics link may come to affect phonological systems. I return to all the previously covered topics in this chapter. I briefly discuss some data on intervocalic voicing in Jutlandic varieties, and compare it to the results in Chapter 4; I discuss variation in the spectral characteristics in stop releases, and compare it to the results in Chapter 5; and I relate these results to the well-known outcomes of stop gradation in these varieties, comparing it to the Modern Standard Danish patterns discussed in Chapter 3. I argue that the default in Jutlandic varieties is to differ from Modern Standard Danish on various parameters: VOT is usually shorter in /p t k/, voicing is more widespread in /b d g/, and (particularly) /t/-affrication is less widespread. When varieties deviate from the default pattern, it requires an explanation; I suggest such explanations for the relatively long VOT in southern Jutlandic, and the /t/-affrication in the area around Djursland.

In Section 6.2, I give a brief survey of the field of Danish dialectology, and a longer introduction to the corpus used throughout the chapter. In Section 6.3, I survey the major dialect areas of Jutland, and in Section 6.4, I discuss the relationship between linguistic variation and geography in traditional dialectology and modern dialectometry. Section 6.5 describes a corpus study of regional variation in VOT in the corpus recordings. In Section 6.6, I present a few non-systematic observations on closure voicing in the corpus. Section 6.7 describes a corpus study of variation in spectral characteristics at the midpoint of aspirated stop releases, based on a subset of the material used for the VOT study. Finally, in Section 6.8, I discuss how the results of the chapter compare with our knowledge of Modern Standard Danish and how they relate to categorical phonology, and speculate on the sources of the observed variation.

6.2 Documentation of Danish dialects

In this section, I give a brief historical account of dialectology in Denmark (Section 6.2.1), and a detailed introduction to the collection of recordings at the dialect research centres in Aarhus and Copenhagen, including but not limited to the DS (1971–1976) corpus, which is the source of the data for the subsequent corpus studies (Section 6.2.2).

6.2.1 Dialect descriptions

There is a long tradition of dialectological research in Denmark. Description of regional varieties has been practised unsystematically for several centuries, and dialectology as a serious scientific discipline has been around since the mid-19th century. For political reasons, coverage of the southern Jutlandic varieties of Danish spoken near the unstable border to Germany were of particular interest; an important early example is Lyngby's (1858) highly detailed grammar. The tendency for meticulous annotation and long word lists increased as phonetics gained popularity as a discipline in the late 19th century; Jensen's (1897–1902) description of the morphology and phonology of Northern Jutlandic is a noteworthy example of this trend.

Towards the end of the 19th century, dialectology throughout Europe took a turn towards more comparative and geographically oriented research (see Section 6.4). In Denmark, this a.o. resulted in a detailed collection of dialect maps by Bennike and Kristensen (1898–1912). The authors were teachers at a *højskole* – a type of boarding school for (mostly young) adults from all over the country – and used their students as informants for the project. Another important publication in this vein is Brøndum-Nielsen's (1927) comparative survey of Danish dialects.

Structuralism (particularly glossematics; see Hjelmslev 1943) became dominant in Danish dialectology in the 20th century. This led to many detailed descriptions of individual dialects, which can be quite difficult to follow for readers unfamiliar with the framework. Key publications include Bjerrum (1944), Jensen (1944), and Andersen (1959). Glossematics was openly uninterested in phonetic substance (Hjelmslev 1943: 46), which is also reflected in these descriptions. A notable exception to the glossematic dominance of this time is found in the work of Skautrup (e.g. 1930, 1944–1970).

In recent years, most work in Danish dialectology has been lexically oriented. There are still two dialect research centres in operation – the Peter Skautrup Centre for Jutlandic Dialect Research in Aarhus, and the Department of Dialect Research in Copenhagen – and both are occupied with long-ongoing dictionary projects (Skautrup et al. 1970–, ØMO 1992–). Otherwise, traditional dialectology has been mostly replaced by variationist sociolinguistics, focusing mainly on urban varieties (e.g. Gregersen and Pedersen 1991) or the waning status of traditional dialects (e.g. Kristensen 1977; Kristiansen 1998; Pedersen 2003; Maegaard and Monka 2019).

For more extensive overviews on Danish dialectology, see Ejskjær (1993) and Hovdhaugen et al. (2000: chs. 4.5.5.1 and 5.6.4.1).

6.2.2 Audio recordings

For as long as it has been possible, the Danish dialect research centers have been recording speakers of traditional dialects for research and posterity (Andersen 1981b; Pedersen 1983). Their combined collections include recordings from as far back as 1934, with speakers born as early as 1849. The collections contain more than 2,100 hours of audio, and recordings from 1,080 different parishes spread throughout the country. Figure 6.1 shows the number of recordings made per year, and Figure 6.2 shows the geographical distribution of recording locations. The corpus is eclectic; it contains sociolinguistic interviews, radio interviews, poetry readings, read word lists, free interaction, musical performances, etc., at very inconsistent levels of audio quality. The

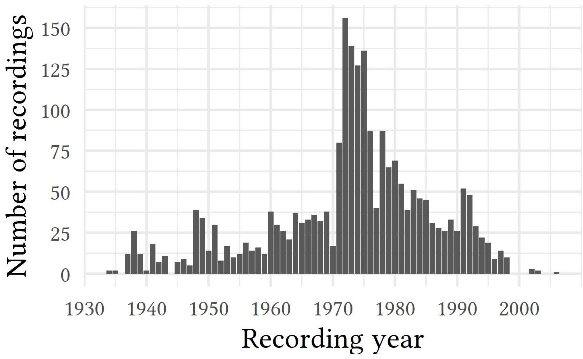

Figure 6.1: *Number of recordings made per year in the combined corpora of dialect recordings.*

corpus maintained by the Peter Skautrup Center for Jutlandic Dialect Research is digitized and freely available on site. These recordings were originally transferred from tape recordings to CD in the 1990s, and later digitized from the CD recordings as MPEG-1 files with a bit rate of 320 kb/s. This format is practical for storage, but the resulting loss of detail is non-trivial (e.g. Bounds et al. 2011); for reference, uncompressed CD quality corresponds to a bit rate of 1411 kb/s. The corpus maintained by the Department of Dialect Research at the University of Copenhagen has not been digitized, but is available on CD on site.

In 1971, the research centers received a grant for a five-year project which aimed at (and almost achieved) gathering tape recordings from every fourth parish in the country (Andersen 1981b; see Figure 6.3). At this time, they already had a broad network of informants, and portable tape recorders had become more readily available. A total of 525 recording sessions took place between 1971–1976, and these recordings are much more uniform than the collection at large. They mostly consist of sociolinguistic interviews with a single informant at their place of residence. Most speakers conform to the NORM (non-mobile older rural males) often found in dialectological studies (Chambers and Trudgill 1998), although the proportion of female informants is higher than commonly found in traditional dialectology. The ages and birth years of the informants in the 1971–1976 recordings are

Figure 6.2: *The geographical coverage of the combined corpora of dialect recordings.*

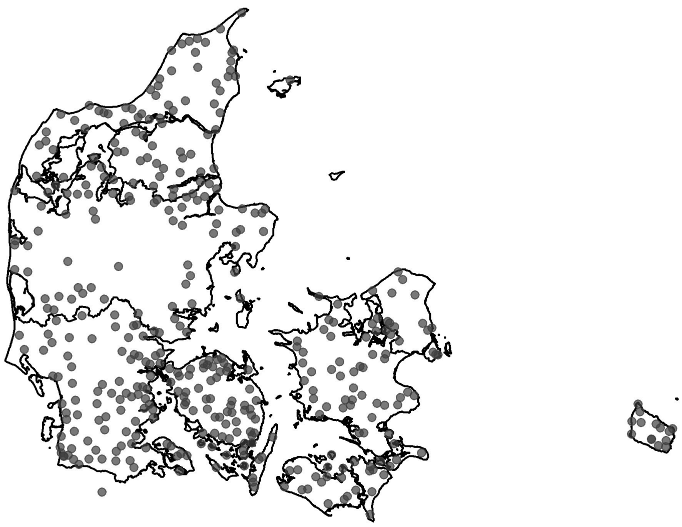

Figure 6.3: *The geographical coverage of dialect recordings made in 1971–1976 and digitized by the Royal Danish Library.*

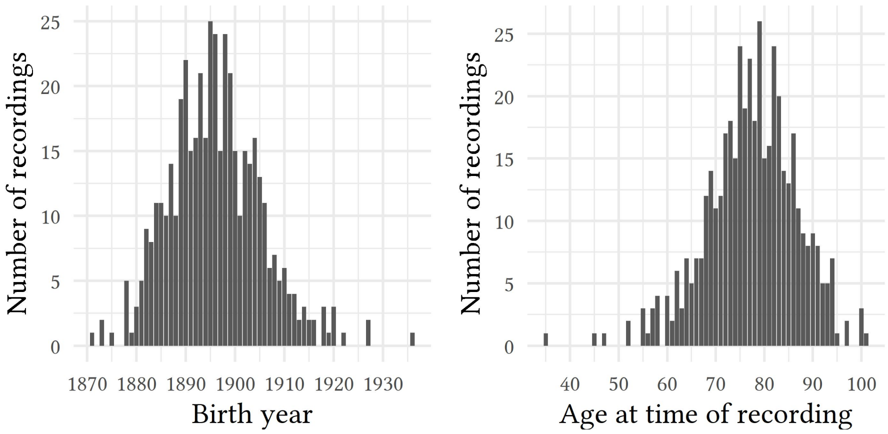

Figure 6.4: *Summary of the ages and birth years of informants in the 1971–1976 recordings.*

shown in Figure 6.4. Approximately two thirds of the recordings have one male informant, one fourth has one female informant, and the rest are group interviews with two or more informants.

Early descriptions of the recordings (Andersen 1981b; Pedersen 1983) clearly state that their primary purpose was to serve as material for dialect dictionaries (Skautrup et al. 1970–; ØMO 1992–).[2] A positive effect of this is that topics generally revolve around old cultural customs and methods of agriculture, food preparation, etc. As Hay and Foulkes (2016) show, speech about older events also tends to elicit conservative pronunciation. A negative effect is that phonetics research clearly did not factor into the considerations at the time; little effort is made to avoid background noise and, the speech of informants often overlaps with discourse particles and questions from the interviewers.

A considerable advantage of these recordings is that the original tapes have all been digitized by the Royal Danish Library and are freely

[2]In the end, the recordings have been used only sparingly for the dictionary of Insular Danish (Gudiksen and Hovmark 2008), and the recordings have not been used systematically for research on Jutlandic varieties until very recently (Puggaard 2018a; Goldshtein and Ahlgren 2021).

available online in high quality.[3] These audio files are of much higher quality than those directly available from the dialect research centers.

This section has given an overview of one particular corpus of recordings of regional varieties, but recordings have also been made for several other projects. For example, Ejstrup (2009) gathered materials in connection with his dissertation on regional variation in vowels, and material was more recently gathered in connection with the *Dialekt i periferien* 'Dialect in the Periphery' project (Maegaard and Monka 2019).

6.3 The dialects of Jutland

Modern Standard Danish is now the primary means of communication throughout Denmark (Kristiansen 1998; Pedersen 2003), but Kristiansen (2003a) judges that the majority of the speech community spoke other regional varieties until the 1960s. In the late 1960s, Skautrup (1944–1970: IV:96ff.) wrote that the dialects were in poor condition, and that the most likely features to survive were phonetic ones which would not affect mutual intelligibility. Skautrup noted at the time that this development was quite advanced on the island of Zealand, where the standard variety had mostly spread from Copenhagen to the rest of the island. The development was less advanced in the Jutland peninsula and the smaller islands, but still well underway.

In the 19th century, an obligatory education system was introduced and agrarian reforms led to increased mobility both in cities and rural areas (Skautrup 1944–1970; Kristiansen 2003b). This led to disruption in the traditional regional varieties and the rise of the current standard language, based on the traditional High Copenhagen variety (Kristiansen 2003a). In the mid-20th century, dialect leveling was accelerated through the spread of national broadcasting in Standard Danish, and through government policies enforcing the use of Standard Danish in the education system (Kristiansen 1990); a more detailed overview of these developments is given in Section 1.3. A recent research project shows that dialects are alive and well in parts of southern Denmark (Monka and Hovmark 2016; Monka 2019),

[3] See the bibliography entry for DS (1971–1976).

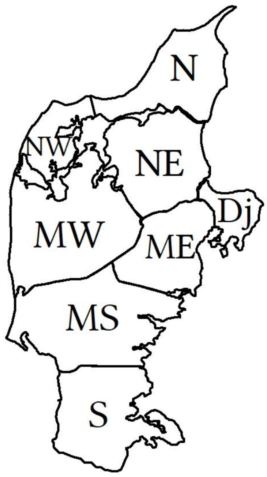

Figure 6.5: *Traditional dialect areas of Jutland as defined by Skautrup et al. (1970–: K.03). N = Northern; NW = North-Western; NE = North-Eastern; Dj = Djursland; MW = Mid-Western; ME = Mid-Eastern; MS = Mid-Southern; S = Southern.*

but that same project also finds complete leveling in other regions that have traditionally been strongly associated with dialect use (Stæhr and Larsen 2019). In yet other regions, regionalized versions of the standard language have replaced traditional dialects (Mortensen 2019). Dialect features may coexist with standard features but take on different social functions that are less geographically delimited than they were in the past (Scheuer et al. 2019).

Figure 6.5 shows the major traditional dialect areas of Jutland as defined by Skautrup et al. (1970–: map K.03).[4] Skautrup (1937, 1944–1970: IV:97) bemoans the fact that there has been relatively little discussion of the basis of these divisions. He maintains that there are no sharp borders between Jutlandic dialect areas. Lines between areas are essentially drawn in transition areas between dialect 'cores' (see

[4]Dialects from this map are written with initial capital letters throughout the chapter. Note that Skautrup et al. (1970–) refer to the two southernmost dialects as *sønderjysk* and *sydjysk* respectively; as there are no fitting English translations for these terms, Southern and Mid-Southern are used here.

Aakjær 1925). Skautrup judges that Danish dialects are mostly defined on the basis of isophones in the form of common phonological developments. Differences in morphology and lexicon also play a role, but less so.

Skautrup (1944–1970: IV:97ff.) provides the clearest diagnostics for the dialect areas. The most important diagnostic is *artikelgrænsen* 'the article boundary', which is responsible for the relatively strict division between eastern and western dialects. In eastern dialects, as in Modern Standard Danish, definiteness in nouns is marked with a suffix [-ən], while in western dialects, it is marked with a phrase-initial article [ə]. This is an exception to the generalization that the defining isoglosses in Danish dialectology are phonological, but note that the article boundary also has major implications for both sentence prosody and segmental phonology (Skautrup 1952). In Section 2.5.3, I discussed a number of further phonological boundaries which involve stops in particular.

6.4 Language variation and geography

Dialectology in Europe became very preoccupied with geography in the late 19th century. Rather than focusing on individual dialects, scholars started drawing detailed maps of distributions of features or lexical items; dialect atlases were drawn of several countries, including Germany (Wenker and Wrede 1895), France (Gilliéron and Edmont 1902–1910), and Denmark (Bennike and Kristensen 1898–1912). In the wake of this work, a debate ensued about whether individual dialects exist at all, or whether geographical variation is of a purely continuous nature (Paris 1888; Gauchat 1903). The conclusion seems to be that although the geographical distribution of features can be chaotic, there are usually areas with adjacent bundles of important isoglosses, and other areas not crossed by significant isoglosses (i.e. dialects do exist). The field of dialect geography has yielded much rich descriptive work, but a common feature of studies from this era is that geography (in a pre-theoretical sense) is typically the only predictor of language variation (Chambers 2000; Britain 2010). Perhaps as a counterreaction, early variationist sociolinguistics (e.g. Labov 1963) was relatively uninterested in geography, with a major exception being the work of Trudgill (e.g. 1974).

Research into the relationship between geography and language variation remains active in the data-driven field of dialectometry (Séguy 1973; see Wieling and Nerbonne 2015 for a recent overview). Dialectometry has made large strides towards estimating the geographical basis of language variation using aggregated features and modern statistical and computational methods. An explicit goal is to estimate how much variation can be explained by geographical distance. By aggregating pronunciations of a large number of words in a single analysis of variation in northern Dutch rather than focusing on well-known loci of variation, Nerbonne and Heeringa (2007) find that geographical distance accounts for more than half of the variation found in their data, making it the most influential predictor. While dialectometry often works with simple Euclidian space ('distance as the crow flies'), the framework also allows for more socially meaningful measures of space, as in e.g. Gooskens' (2005) study of variation in Norwegian using traveling time rather than geographical distance as the main predictor.

In this chapter, I simply model geography in terms of longitude and latitude. As such, I implicitly make the assumption that there are no obvious differences between natural geography and human geography (for more on this notion, see Britain 2010). This is not a good assumption, but it is a highly practical one; quantifying human geography is in itself a difficult task, and more so quantifying human geography as it looked a century ago. This is also part of the motivation for limiting the area of study in this chapter to the Jutland peninsula. This area is relatively densely populated, flat, and has few major obstructions to human movement. Denmark is an island nation, so using Euclidian space as a predictor of variation in the whole country would be much more problematic.

6.5 Variation in voice onset time

In this section, I present a quantitative study of variation in VOT in the traditional Jutlandic varieties. For general background on VOT and information on previous studies of VOT in Danish, see Section 2.3.1. In Section 6.5.1, I present the strategy for selecting the recordings and tokens used in the study, and cover the acoustic and statistical methodology. Finally, in Section 6.5.2, I provide some descriptive statistics, and present the results of a generalized additive mixed model of the data.

6.5.1 Methods

6.5.1.1 Selection of recordings

All recordings used in the current study come from the DS (1971–1976) corpus presented in section 6.2.2 above. DS contains recordings from 230 parishes in Jutland. 17 of these parishes were excluded from the current study. There were three reasons for exclusion: 1) The audio quality was too poor, 2) the recording was too short to include a sufficient number of stops, and 3) group interviews were excluded unless they contained long stretches of speech from a single informant. This only affects areas with a reasonably high density of recordings. If multiple recordings were available from a single parish, one recording was chosen on the basis of either the dialect authenticity judgments made by the original interviewers[5] or audio quality. The audio quality is similar across recordings, and good enough that VOT could be delimited without any major difficulties.

The geographical coverage is shown in Figure 6.6; the peninsula is generally densely covered, although coverage is a little sparse in parts of the Mid-Eastern and Mid-Western dialect areas. The informants' median year of birth is 1896 (range: 1871–1927), and their mean age at the time of recording was 77.4 years (ranging from 45–101 years); the age distribution in these recordings is very similar to that seen in Figure 6.4 above. For thirteen of the informants, no year of birth is reported; I presume that these fall within the reported range. Most recording sessions consist of multiple tapes; whenever possible, the second tape was used for the analysis, so that the informant would have had time to accommodate to the presence of a recording device. All metadata and coordinates, including information about where to

[5] Most recordings in the DS corpus have an accompanying note by the original interviewers evaluating the dialect 'purity' of the informant. This notion is problematized by Goldshtein and Ahlgren (2021).

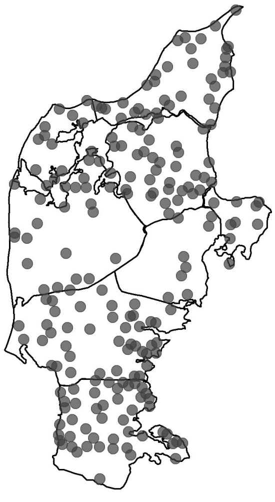

Figure 6.6: *Geographical coverage in the current study.*

Table 6.1: *Distribution of informants across dialect areas by speaker sex.*

Dialect area	**Informants**	**Male**	**Female**
Southern	48	58.3% (*n*=28)	41.7% (*n*=20)
Mid-Southern	40	70% (*n*=28)	30% (*n*=12)
Mid-Eastern	9	66.7% (*n*=6)	33.3% (*n*=3)
Mid-Western	25	92% (*n*=23)	8% (*n*=2)
Djursland	9	66.7% (*n*=6)	33.3% (*n*=3)
North-Eastern	35	85.7% (*n*=30)	14.3% (*n*=5)
North-Western	12	91.7% (*n*=11)	8.3% (*n*=1)
Northern	35	94.1% (*n*=32)	8.6% (*n*=3)
Total	213	77% (*n*=164)	23% (*n*=49)

find to the original recordings, are available in an online appendix (Puggaard-Rode 2022a).

The distribution of informants across dialect areas by speaker sex is seen in Table 6.1. SEX has been shown to affect VOT (e.g. Swartz 1992), particularly among elderly speakers (Torre and Barlow 2009), and by including sex in a statistical model of the data, I will test if the current study can lend credence to those findings. Since speakers were explicitly chosen from a relatively uniform background (non-mobile, rural, previously employed in agriculture) there is little point in attempting to quantify social factors like class.

6.5.1.2 Selection of tokens

As in previous chapters, I distinguish between two laryngeal categories, /b d g/ and /p t k/. This distinction is contrastive in all dialects, although the implementation of the contrast differs.

For each speaker, all stop releases were segmented until the 50th member of /p t k/ had been located. As a result, more tokens of /p t k/ are segmented than of /b d g/. The motivation for this is both practical and theoretical; I have more concrete hypotheses about variation in /p t k/, and only /p t k/ are relevant in the study of affrication patterns in Section 6.7 below. Furthermore, since these recordings are not transcribed, segmentation is very time-demanding. For this reason, /b d g/ are prioritized less.

Only stops in simple onset position were segmented, although palatalized tokens and tokens in stop + /j/ clusters were both treated as simple onsets. Palatalized tokens were included because several dialects show allophonic palatalization of /k/ and /g/ (Bennike and Kristensen 1898–1912: 84ff.); tokens in stop + /j/ clusters were included, because the implementation of these appeared phonetically identical to phonologically palatalized tokens. There were different criteria for the inclusion of the two laryngeal categories. /b d g/ were very often weakened to fricatives or fully voiced when appearing in function words (prepositions, pronouns, and high-frequency adverbs). This would often make segmentation difficult or impossible. For this reason, function words with /b d g/ were excluded unless they were either stressed or initial in an intonational phrase, since gestures are enhanced in these prosodic environments (Steriade 1994). All instances of the pronoun [te] *det* 'it, that' were excluded due to its extremely high frequency (Puggaard 2019a). Function words with initial /p t k/ were not excluded, since they display less weakening than /b d g/ and there are fewer high-frequency function words beginning with /p t k/ than /b d g/. This likely means that this study somewhat underestimates the actual difference between the laryngeal series. This discrepancy is

Table 6.2: *Distribution of stops used in the study by phoneme.*

Consonant	Number
/b/	2,212
/d/	2,369
/g/	2,273
/p/	1,386
/t/	5,169
/k/	4,095
Total	17,504

important to keep in mind, but it is not too concerning, since I am generally more concerned with /p t k/.

The distribution of stops used in the study is shown in Table 6.2. While /b d g/ are reasonably evenly distributed across places of articulation, /p t k/ are more skewed, with relatively few instances of /p/. This was also the case in the study of VOT in Standard Danish by Mortensen and Tøndering (2013), and to a lesser extent in the DanPASS-based studies in Chapters 4–5. This is presumably due to /p/-initial words being rare in Proto-Germanic (see Section 2.2.1).

6.5.1.3 Acoustic analysis

Stop releases were segmented manually in Praat (Boersma 2001; Boersma and Weenink 2021).[6] The beginning of a stop release was demarcated at the burst, which was identified from the waveform. Whenever multiple bursts were visible on the waveform, the final one was chosen (following Cho and Ladefoged 1999: 215).[7] The end of a

[6] There are semi-automatic methods of measuring VOT, such as the AutoVOT software (Sonderegger and Keshet 2012); however, this method relies on training data, and due to the highly variable nature of stop implementation in the Jutlandic data, it was not feasible to provide suitable training data.

[7] This phenomenon is not well-understood. It often comes up in studies of individuals with speech disorders, but it is also relatively common in the general population, particularly in velar stops (see Parveen and Goberman 2012 and references therein). The choice of burst has a non-trivial influence on resulting VOT measurements (Gráczi and Kohári 2014), but this influence is presumably similar across varieties, since there are no indications that the presence of multiple bursts is a regional feature

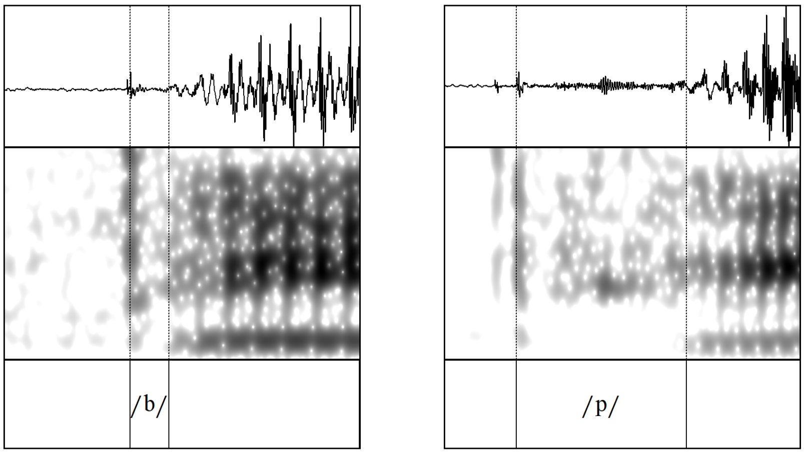

Figure 6.7: *Demarcation of VOT exemplified for unaspirated* /b/ *and aspirated* /p/.

stop was demarcated at voicing onset, identified from the onset of periodicity in the waveform, in accordance with Francis et al.'s (2003) findings that this landmark is most similar to physiological measurements of the onset of vocal fold vibration. This choice of landmark leads to relatively low VOT values, and inflates the differences between dialects as reported in this chapter and the studies of Standard Danish summarized in Section 2.3.1, where the onset of higher formants is typically used to demarcate the end of a stop. It is, however, the same method as were used to delimit stop releases in Modern Standard Danish in Chapter 5, which makes the VOT measurements reported in that chapter comparable to those reported here. The demarcation is exemplified in Figure 6.7. The VOT values for all tokens were extracted using a Praat script, which is freely available in an online appendix along with all TextGrids and measurements (Puggaard-Rode 2022a).

This study only looks at voicing lag, i.e. positive VOT. Measuring pre-voicing in running speech comes with a number of difficulties, particularly in rapid speech. In intervocalic stops, the first portion of the closure is essentially always voiced, due to passive voicing from the preceding vowel; in rapid speech, passive voicing may continue throughout most or all of the closure (Davidson 2016; see Chapter 4).

There is no logical place to start measuring pre-voicing when voicing is continuous; its duration would essentially be a measure of closure duration. When a stop is pre-voiced, there is often still a brief pause in voicing around the release resulting in a short voicing lag. Other recent large-scale studies of VOT have also relied exclusively on voicing lag (e.g. Stuart-Smith et al. 2015; Chodroff and Wilson 2017; Chodroff et al. 2019).

Each token was coded for a number of phonetic environmental predictors, which previous studies indicate influence VOT, and which serve as linear predictors in the statistical model. This allows us to separate the influence of geography from the influence of environmental predictors, and potentially allows us to lend further credence to previous findings. The predictors are:

HEIGHT of the following vowel, which has been shown to influence VOT by e.g. Fischer-Jørgensen (1980), Higgins et al. (1998), and Berry and Moyle (2011); cf. Mortensen and Tøndering (2013), who only found an influence on /b d g/ in Modern Standard Danish. In spite of the great complexity in Modern Standard Danish vowel height distinctions (see Section 1.4), I limited this variable to three levels based on auditory impression. This decision follows Mortensen and Tøndering's (2013) study, which found roughly the same results using a simplified three-way classification of vowel height and Grønnum's (2005: 105) physiological four-way classification of vowel aperture.[8] Danish dialects show a great degree of variability in vowel implementation (Ejstrup and Hansen 2003; Ejstrup 2010), so coding more levels of vowel height would be either too impressionistic or much too time-demanding. Previous studies show that higher vowels increase VOT.

ROUNDNESS of the following vowel, which has been shown to influence VOT in interaction with place of articulation. Fischer-Jørgensen (1972c) found that bilabials have longer VOT before rounded vowels in Standard Danish, while other stops have longer VOT before unrounded vowels.

[8]Note that there are major differences between Grønnum's (2005) vowel height levels, which are based on acoustics, and her vowel aperture levels, which are based on articulation.

BACKNESS of the following vowel, which has also been shown to influence VOT in interaction with place of articulation. Gósy (2001) found that bilabials have longer VOT before back vowels in Hungarian, while other stops have longer VOT before front vowels. Vowels are coded as back or non-back.

PALATALIZATION, which is hypothesized to increase VOT, since palatalized stops are more complex. It is coded as a binary distinction on the basis of auditory impression; no distinction is made between allophonic palatalization and biphonemic sequences of stop + /j/.

STRESS, which has been shown to increase VOT (e.g. Lisker and Abramson 1967; see Section 5.5.1). Stress is coded as a binary distinction.

PLACE OF ARTICULATION and LARYNGEAL CATEGORY. The laryngeal distinction, however it is phonetically implemented, is trivially expected to account for most of the variation in the data. The literature further suggests that the place distinction will influence VOT. A decent rule of thumb is that an occlusion further back in the oral cavity increases VOT, i.e. bilabial < alveolar < velar, although Lisker and Abramson (1964) and Cho and Ladefoged (1999) also find a number of languages not following this pattern. Studies of VOT in Modern Standard Danish have generally found longer VOT for /t/ than /k/ (Fischer-Jørgensen 1954, 1980; Mortensen and Tøndering 2013; see Section 2.3.1). This is also the case with my measurements from DanPASS, as reported in Table 5.1 above. This may be due to the salient affrication of /t/ ensuring a longer release.

Speech rate has also been shown to influence VOT (Andersen 1981a); measuring speech rate of these recordings is far from straightforward, due to the lack of systematic transcriptions of the data, the presence of both informant(s) and interviewer(s), and the general problem with delimiting Danish phonetic syllables (Schachtenhaufen 2010a), particularly in some varieties of Jutlandic (Hansen 1978, 1981). Allen et al. (2003) report that speech rate only partially accounts for idiolectal differences in VOT, which suggests that modeling the individual informant with random effects should largely account for global speech rate effects.

6.5.1.4 Statistical modeling

In order to model the relationship between VOT and geography, the data was fitted to a generalized additive mixed model (GAMM), which can model a potentially non-linear effect of geographical area. Furthermore, descriptive statistics are provided based on the dialect areas. Statistics were calculated in R (R Core Team 2021; RStudio Team 2022) using a number of add-on packages.[9] All R code is freely available (Puggaard-Rode 2022a).

GAMMs were already discussed as one of the smoothing-based approaches to dynamic data analysis in Section 5.4. They are a method of non-linear statistical analysis that is well-suited for data that varies dynamically across time or space (see Wood 2017a for a general introduction, and Sóskuthy 2017 and Wieling 2018 for linguistics-themed introductions). While a linear analysis of e.g. vowel formants across time will have to either measure formants at a chosen landmark or normalize across time steps, a GAMM-based analysis can take into account a full formant trajectory, as demonstrated by Sóskuthy (2017). Similarly, rather than normalizing across dialect areas, a GAMM-based analysis can take into account the full scope of geographical variation in a given area (see also e.g. Wieling et al. 2011, 2014).

The GAMM has VOT as its dependent variable. Regional variation is included in the model through thin plate regression spline smooths (Wood 2003) for the interaction between longitude and latitude; one smooth variable models the main effect of geography, and individual smooth variables model how the two laryngeal categories vary as a function of geography. The contextual phonetic variables discussed above are all included in the model as fixed effects: HEIGHT, ROUNDNESS, BACKNESS, PALATALIZATION, STRESS, SEX, PLACE OF ARTICULATION, and LARYNGEAL CATEGORY.

The fixed effects are contrast coded (Schad et al. 2020; see Sections 4.5.3 and 5.5.2). Binary variables are coded such that the direction with highest expected VOT is positive, and lowest expected VOT is

[9] The following packages were used: `dplyr` (Wickham et al. 2022) for data management; `mgcv` (Wood 2017a, 2021) for fitting GAMMs; `itsadug` (van Rij et al. 2020b) for likelihood ratio tests; `mgcViz` (Fasiolo et al. 2020, 2021) for three-dimensional map-based visualization of GAMMs, and `ggplot2` (Wickham 2016; Wickham et al. 2021) for other visualizations.

Table 6.3: *Overview of contrast coding for categorical variables.*

Variable	Contrast
ROUNDNESS	-½ round, +½ non-round
BACKNESS	-½ back, +½ non-back
PALATALIZATION	-½ non-palatal, +½ palatal
STRESS	-½ unstressed, +½ stressed
SEX	-½ female, +½ male
LARYNGEAL CATEGORY	-½ /b d g/, +½ /p t k/
HEIGHT	-⅓ low, -⅓ mid, +⅔ high
	-½ low, +½ mid
PLACE OF ARTICULATION	-⅔ bilabial, +⅓ alveolar, +⅓ velar
	-½ alveolar, +½ velar

negative. For ternary variables, two theoretically motivated contrasts were coded. For HEIGHT, I test the distinction between 1) high and non-high vowels, and 2) mid and low vowels. For PLACE OF ARTICULATION, I test the distinction between 1) labials and non-labials, and 2) alveolars and velars. The contrast codes are summarized in Table 6.3.

As mentioned in Section 6.5.1.3, we have some specific hypotheses about the interactions between PLACE OF ARTICULATION : BACKNESS and PLACE OF ARTICULATION : ROUNDNESS, respectively. These only pertain to /b p/, so they are included in the GAMM for the labial–non-labial only. The model is further fitted with by-speaker random slopes corresponding to each of the fixed effects, with the exception of SEX.

The model is run with fast restricted maximum likelihood estimation (fREML) with discretized values for covariates to decrease computing load (Wood et al. 2017), assuming the scaled-*t* error distribution to account for heavy-tailed residuals. In order to estimate the influence of the geographical variables, a nested model without the effect of area was fitted, and minimized smoothing parameter scores (fREML scores; see Section 5.5.2) of the two models are compared using a likelihood ratio test (van Rij 2016).

The model described here differs from the one included in a previously published paper (Puggaard 2021) in a number of ways. In the previous paper, categorical variables were not contrast coded. Instead of having separate variables for PLACE OF ARTICULATION

Table 6.4: *Mean VOT in ms for each phoneme by dialect area (see Figure 6.5 for abbreviations; MSD = Modern Standard Danish). First and third quantile in parentheses.*

Area	**/b/**	**/p/**	**/d/**	**/t/**	**/g/**	**/k/**
S	9.3 (5–12)	42.7 (28–55)	14 (8–18)	53.7 (40–66)	16.7 (11–22)	54.5 (41–66)
MS	8.1 (4–10)	38.5 (26–48)	12.4 (6–16)	52.3 (38–64)	13.1 (5–19)	45.9 (34–57)
ME	5.7 (3–7)	41.5 (23–53)	11.4 (5–15)	51.7 (31–70)	12.7 (6–17)	49.9 (37–62)
Dj	9.2 (4–13)	46.2 (30–60)	14.5 (9–19)	53.8 (39–67)	14.4 (7–20)	48.3 (35–60)
MW	5.8 (3–7)	32 (19–44)	10.2 (4–13)	40.1 (28–50)	10.8 (3–15)	39.8 (27–51)
NE	6.8 (3–10)	30.6 (17–40)	12.7 (5–17)	42.5 (29–53)	10.2 (4–14)	41.5 (27–52)
NW	8.2 (4–11)	31.8 (20–42)	12.9 (6–16)	36.3 (24–46)	13.3 (6–20)	42.9 (31–53)
N	6.9 (3–10)	30.6 (18–40)	11.5 (5–16)	42.2 (27–53)	11.8 (4–16)	41.8 (29–52)
MSD	–	44.3 (29–56)	–	69.8 (54–81)	–	50.1 (38–59)

and LARYNGEAL CATEGORY, the stops were coded with a categorical PHONEME variable with the levels /b d g p t k/. The model reported here furthermore has a much more elaborate random effects structure, and includes two interaction variables. These changes should make the results more interpretable and reliable. There are no substantial concomitant changes to the interpretation of the results.

6.5.2 Results

6.5.2.1 Descriptive statistics

Table 6.4 gives VOT values for the different phonemes as grouped by dialect area and for Modern Standard Danish /p t k/ (see Section 5.5.1 above). Mortensen and Tøndering (2013) measured the VOT of

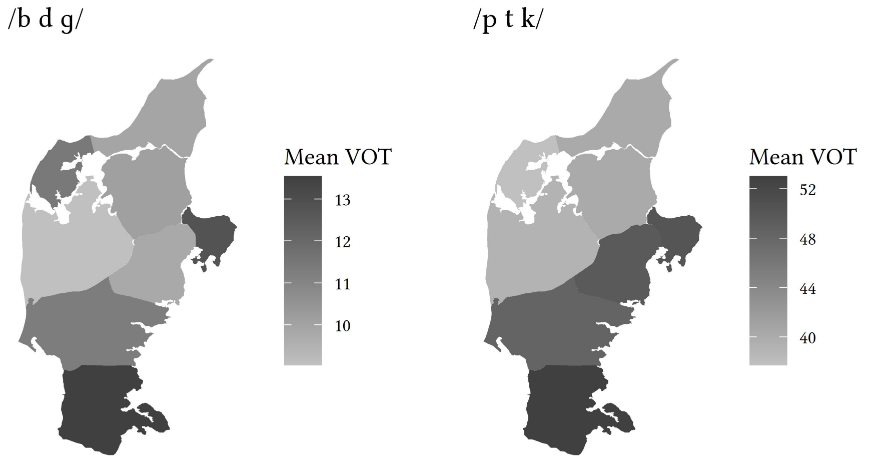

Figure 6.8: *Mean VOT values for the two laryngeal categories by dialect area projected onto maps.*

/b d g/ in the DanPASS corpus, but given the differences in methodology between that study and the present one, I will not be making comparisons with their results here. The results for the two laryngeal categories are projected onto maps in Figure 6.8, and the results for individual phonemes are projected onto maps in Figure 6.9. With only a few counterexamples, VOT is shorter in the traditional regional varieties than in Modern Standard Danish. Modern Standard Danish /t/ is much longer on average than /t/ in all Jutlandic varieties.

For /p t k/, dialect area clearly influences VOT. The pattern is roughly similar for all stops. It is most pronounced for /t/ and least for /k/. The dialect areas seem to form clusters: essentially, south(-eastern) varieties have longer VOT, and north(-western) varieties have shorter VOT. Interestingly, in most cases, the minimum gap in VOT between a member of either dialect cluster seems to be approximately 10 ms, which was found by Blumstein et al. (2005) to be the lower limit of what the human neural system can perceive. This suggests that the limits on regional variation in VOT is only constrained by limits on perception. The only thing approaching a pattern in /b d g/ is the relatively high values across the board in the Southern and Djursland varieties.

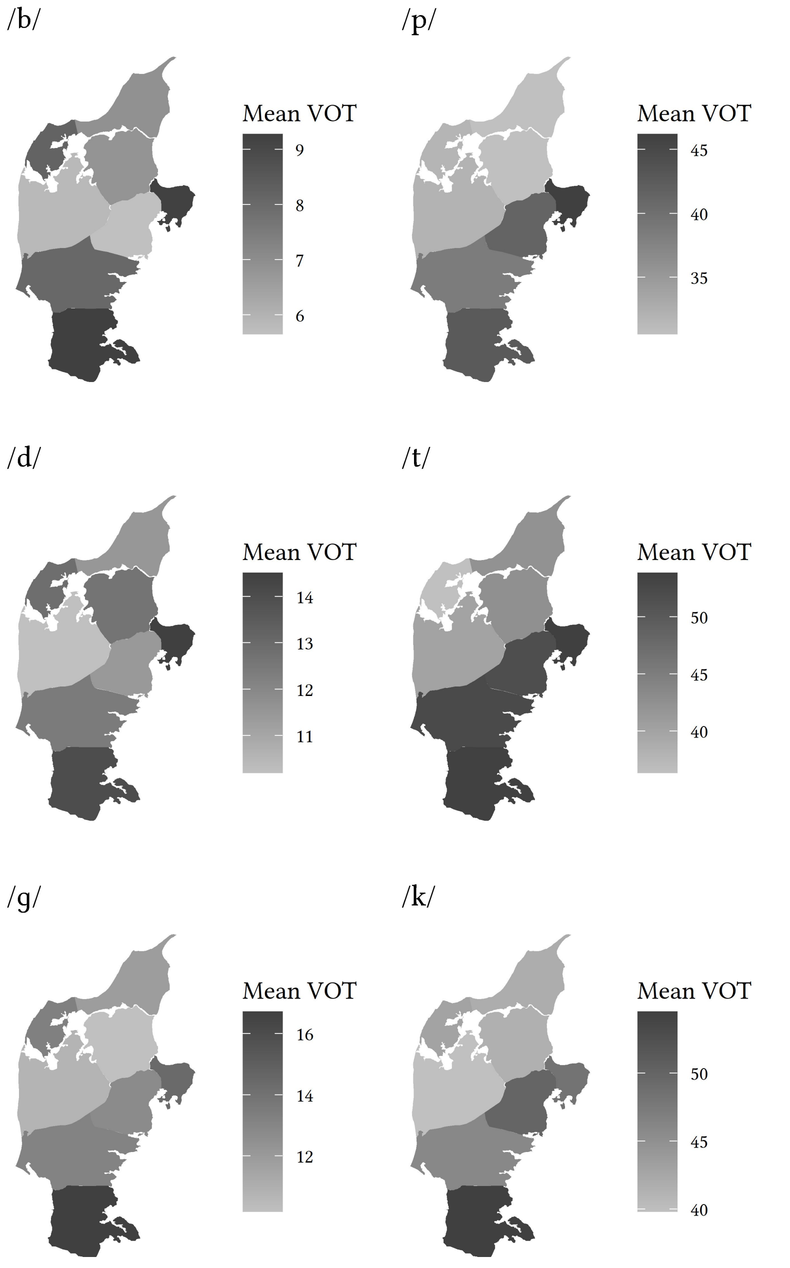

Figure 6.9: *Mean VOT values for individual phonemes by dialect area projected onto maps.*

Table 6.5: *Parametric coefficients of generalized additive mixed model.*

	estimate	SE	*t*	*p*	
intercept	31.03	0.6	51.36	<.001	***
LARYNGEAL CATEGORY	31.18	0.63	49.54	<.001	***
HEIGHT: high *vs.* non-high	2.7	0.28	9.74	<.001	***
HEIGHT: low *vs.* mid	-0.07	0.24	-0.31	0.76	
ROUNDNESS	2.86	0.33	8.7	<.001	***
BACKNESS	2.49	0.38	6.64	<.001	***
PALATALIZATION	9.39	0.94	10	<.001	***
STRESS	2.6	0.22	11.96	<.001	***
PLACE: labial *vs.* non-labial	5.56	0.29	18.9	<.001	***
PLACE: alveolar *vs.* velar	0.82	0.38	2.15	0.03	*
SEX	0.03	0.45	0.07	0.94	
ROUNDNESS : labial contrast	1.6	0.66	2.42	0.02	*
BACKNESS : labial contrast	3.04	0.74	4.11	<.001	***

6.5.2.2 Generalized additive mixed model

This section presents the results of the GAMM. A likelihood ratio test found that a model including the geographical variable performs significantly better than a nested model without the geographical variable, with χ^2 (9) = 72.8, p < .001. With a high effect size of R^2 = .66, the model explains the data quite well. The parametric coefficients and estimated significance of smooth terms are given in Tables 6.5 and 6.6. Table 6.5 summarizes the influence and significance of the linear predictors, and Table 6.6 summarizes the non-linear influence of geography. I unpack this information below, starting with the linear predictors. The intercept (β_0 = 31.01) indicates that the mean VOT in the data set is 31 ms when all other variables are kept at zero. As this averages over both laryngeal categories, the intercept is not particularly meaningful.

LARYNGEAL CATEGORY. There is a (trivially) strong effect of laryngeal category on VOT, such that /p t k/ are longer than /b d g/. This is by far the strongest categorical predictor.

HEIGHT. The data support the hypothesis that VOT is increased when a high vowel follows, but does not support a more complex effect of vowel height.

Table 6.6: *Approximate significance of smooth terms modeling geographical variation. (edf = estimated degrees of freedom, ref.df = referential degrees of freedom).*

	edf	**ref.df**	***F***	***p***	
lon,lat	10.98	12.96	3	<0.001	***
lon,lat : /b d g/	2	2	1.51	0.22	
lon,lat : /p t k/	2	2	11.6	<.001	***

ROUNDNESS. The hypothesis was that following vowel roundness influences VOT in interaction with place of articulation, such that VOT is longer before unrounded vowels, unless the stop is bilabial. This is supported by the analysis, which shows that VOT is generally longer before unrounded vowels, and more so if the stop is not bilabial.

BACKNESS. As with ROUNDNESS, the hypothesis was that following vowel backness influences VOT in interaction with place of articulation, such that VOT is longer before non-back vowels, unless the stop is bilabial. This is supported by the analysis, which shows that VOT is generally longer before non-back vowels, particularly if the stop is non-labial.

PALATALIZATION. The data strongly support the hypothesis that palatalization increases VOT.

STRESS. The data support the hypothesis that stress increases VOT. This effect is stable, although the magnitude is surprisingly small.

PLACE OF ARTICULATION. The data support the hypothesis that VOT is modulated by place of articulation. There is strong support for the generalization that labials are shortest, and more moderate support for alveolar < velar. This is rather different from Modern Standard Danish, where the results reported in Section 5.5.1 and earlier studies of VOT (Section 2.3.1) all find much higher VOT in /t/ than /k/. If this pattern in Modern Standard Danish is due to the salient affrication in /t/, this suggests that /t/-affrication will be much less common in the Jutlandic varieties. The descriptive statistics suggest that some areas show a stable pattern of alveolar < velar.

SEX. The model shows no signs of a stable correspondence between sex and VOT.

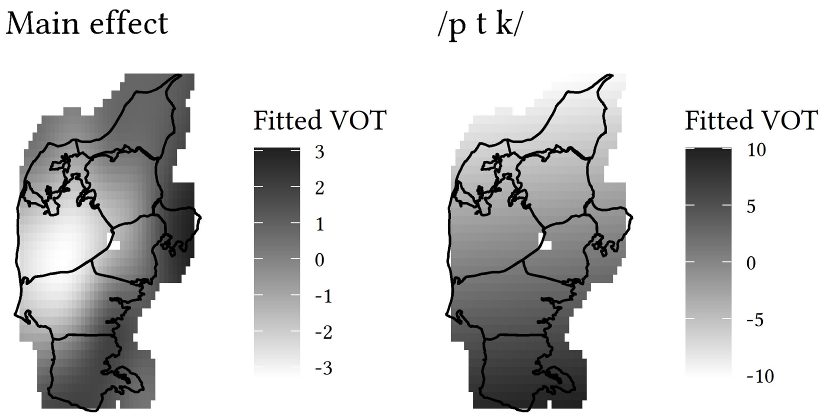

Figure 6.10: *Fitted VOT values attributable to area. Darker shading indicates higher fitted values. Black lines indicate traditional dialect boundaries.*

GEOGRAPHY. The main purpose of this study is to examine the non-linear influence of geography on VOT. There is a strong main effect, suggesting that a primary geographical pattern is shared by all stops. Additionally, /p t k/ show further patterns of geographical variation. Table 6.6 only tells us that significant differences exist; in order to inspect these differences, we need to visualize the fitted values. In Figure 6.10, the fitted main effect of geography on VOT is projected onto a map (left), as is the fitted effect of VOT in /p t k/ (right). There is relatively high VOT in a small area around Djursland, and in the Southern variety, as well as the eastern part of the Mid-Southern variety. Fairly high VOT is also found in a small area covering parts of the Northern and North-Eastern varieties. A large area in the mid-western part of the peninsula has very short VOT. The main effect of geography is highly non-linear, and with a few exceptions does not follow the major traditional dialect areas.[10] The effect attributed to /p t k/ shows a strong, if much simpler, continuous effect of increasing VOT going north–south.

[10]The completely white squares in the middle of both maps are due to the scarcity of recordings in this area, and have nothing to do with the results.

The plot legends in Figure 6.10 refer to the variance in VOT that the model attributes to coordinates relative to the intercept β_0. The magnitude of this variance is of course bigger in /p t k/, as VOT is overall higher in these sounds; the *p*-value for this effect is lower, which is likely because the model's confidence in the geographical fit is lower. However, both results are quite strong. The results are discussed further in Section 6.8 below.

6.5.3 Interim conclusions

This study of VOT in Jutlandic varieties of Danish follows up on Puggaard (2018a), which showed that a variant of /t/ with short VOT is not limited to any specific part of the peninsula, but is found to different extents throughout. I have provided stronger support for that finding, and shown that variation is not limited to /t/, but reflects more general patterns in stop phonetics. Shorter variants of stops than in Modern Standard Danish are nearly consistent throughout Jutland. The longest VOT is found in southern Jutland, parts of mid-eastern Jutland, and Djursland. Very short VOT is found in the centre of the peninsula, and (specifically for /p t k/) in the far north. There is no good explanation for why VOT values are so short in the center of the peninsula; we must assume that it was simply a feature of the traditional local variety. Variation follows consistent but complicated geographical patterns, and the GAMM results suggest that there are multiple continua of variation. The study further finds support for most claims in the literature about how environmental phonetic variables influence VOT.

6.6 Some observations on closure voicing

In Chapter 4, I discussed the extent of intervocalic stop voicing in Modern Standard Danish in the light of previous studies of glottal activity in stop production. Unsurprisingly, there are no corresponding studies of glottal activity in the traditional regional varieties. We only know that glottal spreading is needed to maintain aspiration, and on the basis of the results presented in the previous section, we know that aspirated releases are generally shorter in traditional

Jutlandic varieties than in Modern Standard Danish. This implies that the glottal opening gesture in /p t k/ either peaks earlier or has a smaller magnitude. The results do not allow us to compare /b d g/ in Jutlandic varieties and Modern Standard Danish, since there are no studies of Modern Standard Danish /b d g/ with comparable methodology. It seems a reasonable assumption, however, that the glottal spreading gesture in /b d g/ would also be smaller (or non-existent), in order to achieve a more dispersed contrast. I will not present an empirical study of closure voicing in regional varieties here, but simply discuss some of my observations from working with the corpus, and how they relate to the situation in Modern Standard Danish.

In Modern Standard Danish, intervocalic voicing is common in inflectional suffixes, but otherwise uncommon. It is especially rare in stressed syllables, and after syllables with stød. I argued in Chapter 4 that this relative rarity is caused by a glottal spreading gesture during the closure of /b d g/, as reported by e.g. Hutters (1985). If this glottal gesture is smaller or non-existent in other varieties of Danish, the predicted voicing patterns would be different, as discussed in Section 4.2: intervocalic voicing should overall be more likely, and should be found in more contexts. A smaller or non-existent glottal opening gesture will not result in closure voicing in absolute initial position, as initial closure voicing does not follow automatically from a narrow glottis, but requires separate articulatory effort.

While working with the Jutlandic recordings, I observed that continuous voicing in intervocalic (or intersonorant) position was much more common than in Modern Standard Danish. This was especially the case in the northern part of the peninsula. For most Northern Jutlandic Danish speakers in the corpus, intersonorant voicing is indeed categorical or near-categorical (as has sometimes been incorrectly claimed for Modern Standard Danish). This is not only the case in contexts where voicing is relatively common in Modern Standard Danish, but also e.g. in initial stops in stressed syllables and after syllables with stød.

Consider the phrase in (1), which is spoken by an informant from Bindslev.

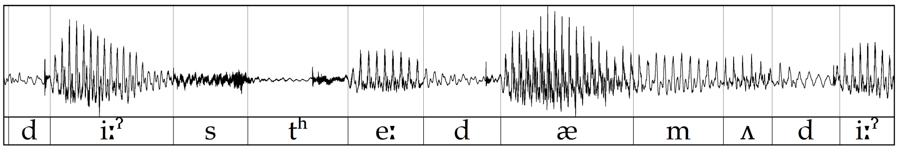

Figure 6.11: *Waveform showing three instances of fully voiced intersonorant* [d] *initially in stressed syllables; spoken by informant from Bindslev; see* (1).

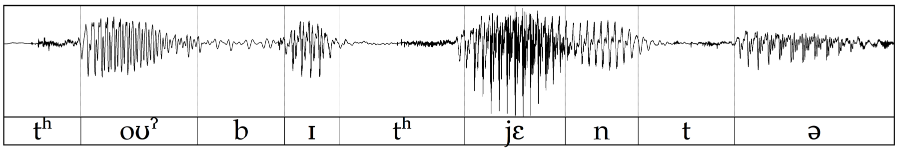

Figure 6.12: *Waveform showing* [b] *with creaky voice after a vowel with stød; spoken by informant from Asdal; see* (2).

(1) ʌ ˈçɛlˀp ʌ ˈpʰluh noʊn ˈdiːˀs tʰeː ˈdæm ʌ ˈdiːˀs tʰe ə ˈseːˀl
and help to pluck some partly for them and partly for to sell
'and help pluck some, in part for them, and in part to sell'

All tokens of [d] in (1) are fully continuously voiced, in spite of being initial in stressed syllables. This is evident from Figure 6.11, which shows the waveform of the relevant parts of (1).

The phrase in (2), which is spoken by an informant from Asdal, shows creaky voice continuing throughout the closure of /b/, which follows a stressed vowel with stød, as evident from Figure 6.12.

(2) ˈtʰoʊ̯ˀ bɪˈtʰjɛntə
'two officers'

This pattern is common in Northern Jutlandic varieties, despite being exceedingly rare in Modern Standard Danish.

As argued in Section 4.2, closure voicing is articulatorily natural in intervocalic position, and as such, does not require a phonological

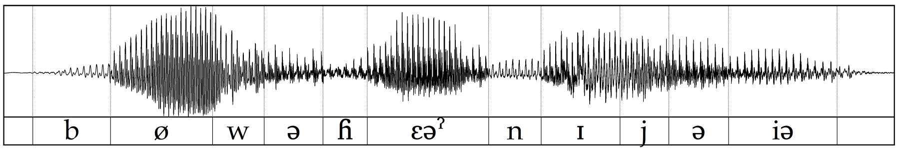

Figure 6.13: *Waveform showing post-pausal, word-initial, and stressed* [b] *with closure voicing; spoken by informant from Elling; see* (3).

account.[11] It is noteworthy, however, that voicing is usually maintained even after stød, which should significantly impact the 'naturalness' of voicing. This clearly suggests that these speakers of Northern Jutlandic varieties lack the glottal spreading gesture found in Modern Standard Danish intervocalic /b d g/: while voicing is not usually found initially, there is no effort to block voicing in other positions.

Closure voicing in initial position, however, requires articulatory effort. I have found no examples in the Northern Jutlandic data of voiced /d g/ in absolute initial position, but some speakers realize /b/ with closure voicing near-consistently, including in absolute initial position. This is the case for (at least) the informants from Elling, Volstrup, Voer, and Bording.

The phrase in (3), which is spoken by an informant from Elling, shows closure voicing in /b/, which is spoken in a word-initial stressed syllable after a pause; the notation (.) indicates a pause. This is evident from Figure 6.13.

(3) ˈlilə (.) ˈhɑ̃õ (.) ˈbøwə ɦɛəˀˈnɪjə iə (.) ˈstʁɑjby
little (.) harbor (.) built here.down in (.) Strandby
'little harbor built down here in Strandby'

When post-pausal voicing is only found in /b/, this is likely because the supraglottal cavity behind the closure is particularly large in /b/ (see Section 4.2). This corresponds to a well-known typological observation: if languages have just one voiced stop, it is bilabial, as demonstrated by a number of Mayan languages, including Tzeltal and Huatesco

[11] The arguments given in that section all extend to intersonorant position.

(Gamkrelidze 1975). This leaves two possible explanations for the voicing pattern in those Northern Jutlandic speakers who display /b/ voicing: 1) Some articulatory effort is put into voicing of /b d g/, and it is sufficient for achieving voicing in /b/ but insufficient in /d g/. 2) No articulatory effort is put into voicing of /d g/. The actual explanation is likely a combination of the two.

The lone pre-voicing in /b/ may either signal that closure voicing was gaining ground in Northern Jutlandic, or it may be the last vestige of a system with more closure voicing. The latter option seems more plausible: as discussed in Section 3.5.3, there is reason to believe that voicing generally used to be more widespread in earlier stages of Danish. The fact that Northern Jutlandic had more widespread closure voicing relatively recently also helps to explain the short VOT in /p t k/ in this variety.

6.7 Variation in spectral characteristics

In Chapter 5, the spectral characteristics of stop releases are discussed in some detail. As pent-up air behind a supraglottal closure is released, frication noise is generated which reflects properties of the primary constriction in the oral cavity. As the constriction opens, low-frequency aspiration noise generated near the glottis gradually becomes dominant, and the turbulent airstream excites the natural frequencies of the oral cavity. In Chapter 5, I explored how the distribution of noise in Modern Standard Danish stops gradually changes throughout the release. As discussed above, there is reason to believe that spectral characteristics during stop releases in some Jutlandic varieties are either different from Modern Standard Danish, or change at a different rate throughout the release. This is particularly to be expected in /t/, which is prominently affricated in Modern Standard Danish.

I used function-on-scalar regression models to explore Modern Standard Danish stop releases in Chapter 5. Ideally, similar models would be used to test how time-varying spectral characteristics vary throughout the Jutland peninsula. This is mathematically possible; the software used to fit these models does not 'care' how complex and multidimensional the underlying data structure is, although computing time may drastically increase with more complex models. However, a statistical model including interrelationships between amplitude, frequency, time, longitude, and latitude would be very tricky to interpret and visualize. For this reason, and in spite of the counterarguments I gave in Section 5.3, this exploration of variation in spectral characteristics makes use of 'magic moments' (Mücke et al. 2014). The full spectrum is analyzed, but in order to model spectral characteristics in an interpretable manner, the complexity of the underlying data must first be reduced.

In the following section, I discuss methods of dimension reduction in statistics, focusing specifically on functional principal component analysis. In Section 6.7.2 below, I present the token selection procedure and statistical methods used in the current study, and in Section 6.7.3, I discuss variation in the spectral characteristics of stop releases in Jutlandic varieties.

6.7.1 Clustering and dimension reduction

Phonological contrasts are regulated with many phonetic cues of varying importance. For various reasons, phoneticians have often had to focus on one cue at a time. Some reasons are theoretical: similar to how some phonologists have attempted to boil phonological grammars down to as few features as possible, some phoneticians have tried to boil phonetic contrasts down to as few invariant acoustic cues as possible (e.g. Stevens and Blumstein 1981). Other reasons are practical; for example, many popular statistical methods in linguistics require a single continuous or categorical dependent variable (see Sections 5.3–5.4).

There are several statistical methods available for reducing a large number of variables to a smaller, more manageable number. Some linguists have used these methods for classification of languages and varieties. For example, Dunn et al. (2005) use hierarchical cluster analysis to arrive at phylogenetic relationships between languages on the basis of large number of data points describing different aspects of language structure; Nerbonne (2009) uses multidimensional scaling to classify dialects on the basis of pronunciation differences in cognate lexical items (see also Wieling 2012; Prokić and Nerbonne 2013).

A common approach to reducing the complexity of multidimensional data is principal component analysis (PCA). In an early example of PCA applied to acoustic data, amplitudes for Dutch vowels were measured in 18 frequency bins, and PCA was used to reduce the original 18 dimensions to 10 principal components (PCs) with no loss in explanatory value (Plomp et al. 1967; Pols et al. 1969; Klein et al. 1970). Instead of being classified on the basis of 18 binned amplitude measurements, each observation can now be classified on the basis of coordinates in a ten-dimensional space. These coordinates are called PC scores, abbreviated s_{n}. The 10 PCs are ranked by how much variance in the original dimensions they account for, and the studies show that the 3–4 PCs with the highest explanatory value can account for the vast majority of information in the original measurements. Classification of the vowels based on these 3–4 PCs is strongly correlated with formant-based classification of the vowels.

PC scores can serve as variables in regression models. For example, if there are several continuous predictor variables in a study, PCA can reduce them to a smaller number, ideally making the results more easy to interpret, and overcoming the problem of fitting models with potentially correlated variables. If a study has a high number of continuous response variables, requiring many separate regression models to be fitted, PCA can be used to reduce the number of necessary models.

Similar to the functional extension of regression we saw in Chapter 5, there is also a functional extension of PCA (FPCA; Castro et al. 1986; Ramsay and Silverman 2005). PCA is used to reduce a large number of continuous variables, and FPCA is used to reduce the complex information found in functions (such as curves) to a small number of continuous variables. This has an obvious appeal in phonetic research, where we often work with variables that change as a function of time, such as pitch (Zellers et al. 2010). FPCA can even be applied to multiple curves simultaneously; Gubian et al. (2015) use FPCA to jointly analyze the patterns of $F1$ and $F2$ in the realization of diphthongs and hiatuses in Spanish, and Gubian et al. (2019) use a similar technique to track an ongoing vowel merger in New Zealand English. In the first study, the authors find that the majority of variation in the formant contour pairs can be expressed with just one PC. A linear regression with s_1 (scores for the first PC) as its dependent variable can predict the diphthong–hiatus contrast in Spanish quite well.

Unlike most previous studies applying FPCA to phonetic data, I do not analyze how some variable changes as a function of time, but rather amplitude as a function of frequency (i.e. spectra). Each PC is associated with an aspect of variance in the input spectra, and the influence of each PC can be visualized by showing how it deviates from the mean spectrum. As I show in Section 6.7.3.1, it is also possible to largely reconstruct any of the original spectra *amp(f)* by weighting the mean spectrum in the data *μ(f)* by the PC scores of *amp(f)*.

6.7.2 Methods

6.7.2.1 Token selection and acoustic analysis

The tokens used in the analysis of spectral characteristics are the same as those used in the analysis of VOT in Section 6.5, although some tokens were excluded here. Similar to the study in Chapter 5, this study is limited to /p t k/.

As mentioned above, this study relies on 'magic moments' in the sense of Mücke et al. (2014). A Praat script was used to extract a 5 ms sound file from the midpoint of each annotated stop release. A bandpass filter was applied to remove frequencies below 500 Hz and above 12.5 kHz. The results in Chapter 5 suggest that the noise distribution at release midpoint will reflect different sources in affricated and non-affricated stops. 5 ms is presumably long enough that the derived spectra are not overly affected by random noise, and short enough that the derived spectra do not span across multiple phases of the release. Stops with VOT < 10 ms were excluded from the study, since 5 ms taken from the midpoint of such short releases likely do span across multiple phases. The distribution of stops by phoneme, as well as the number of excluded stops by phoneme, are given in Table 6.7.

Multitaper spectra were generated in R for each sound file, using the same procedure as described in Section 5.5.1, fn. 13. Within each spectrum, the intensity values (in W/m^2) were log-transformed and standardized (following Gelman and Hill 2006) to keep all measurements on the same scale. Since the values are standardized, simple

Table 6.7: *Distribution of stops used in the study by phoneme.*

Consonant	Number	Excluded
/p/	1,295	91
/t/	5,103	66
/k/	4,042	53
Total	10,440	210

log transformation is equivalent to converting W/m^2 values to dB.[12] Measurements at frequencies > 8 kHz were removed from the analysis, as they did not contribute much to the models other than noise.

The variables used in the analysis of VOT are all included in this study as well; the potential influence of these variables on spectral characteristics of stop releases are discussed in Chapter 5.

6.7.2.2 Statistical modeling

Since a model including both geographical coordinates and dynamic spectral characteristics would be prohibitively difficult to interpret, I first applied FPCA to the spectra within each stop phoneme. Scores for the most influential PCs subsequently serve as dependent variables in (potentially) spatial GAMMs. Statistics were calculated in the R environment (R Core Team 2021; RStudio Team 2022) using a number of add-on packages.[13]

Unlike the study of affrication in Chapter 5, the sound recordings used to compute multitaper spectra for this study were all of equal duration, which means that all spectra have the same number of frequency bins. In other words, the data are regularly sampled and densely distributed. Dense data are often presumed to be relatively noiseless (Gajardo et al. 2021b), but stop release spectra are inher-

[12]Recall that I did not log-transform the intensity values for the study in Chapter 5. My impression from exploratory analysis is that using a logarithmic scale is too fine-grained for fitting FOSR models, and using the W/m^2 scale is too coarse-grained for FPCA of the spectrum, although differences across scales are in both cases relatively minor. It remains a problem for future research to establish best practices.

[13]In addition to the packages used in the study of VOT (see fn. 9), `fdapace` (Gajardo et al. 2021a) was used to implement FPCA.

ently noisy, in spite of multitaper spectra being much less noise-prone than spectra computed from Fourier transformation. For this reason, the spectra are smoothed using a local linear smoother (see Zhang and Wang 2016), with smoothing parameters set automatically using generalized cross-validation (GCV). This procedure is used to select a combination of parameters which yields a good compromise between overfitting and underfitting.[14] For each stop, I computed as many PCs as necessary to account for 95% of the variance in the data. However, for each stop, I only analyzed the three most influential PCs, as the variance accounted for by lower PCs was negligible and difficult to interpret.

For each stop, s_{1-3} served as dependent variables in separate GAMMs. The model selection procedure used here was rather different from previously discussed models. I have had specific theoretical reasons for including all independent variables in previously discussed models, so variables were included regardless of whether their inclusion improved the model fit. In Chapter 5, I motivated how a number of variables may influence the shape of release spectra, and these were all included in the FOSR models. Those reasons also apply to the spectra studied here, but I cannot evaluate how (or if) they will affect specific PC scores computed from spectra. For this reason, I used a step-up model selection procedure, starting with a minimal model including only the geographical variable and by-speaker random intercept. For each stop, I added variables and associated by-speaker random slopes (except in the case of SEX) in the order of their effect size in Chapter 5, i.e. their associated reduction in fREML score. For example, for the /k/ PCs, the order was first BACKNESS, HEIGHT, ROUNDNESS, SEX, STRESS (and finally PALATALIZATION and VOICE ONSET TIME, which were not included in the models in Chapter 5). Variables were kept only if the cost of fitting a more complex model was sufficiently outweighed by an improved model fit; in practice, I ran likelihood ratio tests to test if variables caused a significantly reduced fREML score. Finally, I compared the final model with a nested model

[14] This approach differs from Gubian et al. (2015), who recommend using B-spline smoothing with parameters guided by a combination of GCV and domain-specific knowledge. They implement FPCA using a different R package, `fda` (Ramsay et al. 2009), which requires input functions to be pre-smoothed by the user.

which did not include the geographical variable. When I report *p*-values for the geographical variable below, they are computed from these likelihood ratio tests (see van Rij 2016).

As with the model presented in Section 6.5.1.4, all models are run with fREML with discretized values for covariates. Unlike the spatial GAMM discussed in Section 6.5.1.4, these GAMMs are run with the presumption of a Gaussian error distribution, and the resulting residuals are reasonably normally distributed (for more details, see Puggaard-Rode 2022a). The categorical variables inherit the same contrast coding from the VOT study; see Table 6.3. VOT is standardized within each phoneme, following the procedure of Gelman and Hill (2006).

6.7.3 Results

In this section, I will present the results of the analyses described above in turn, starting with /t/. The results for /t/ spectra will be described in more detail than /k/ and /p/, as I will use /t/ to exemplify some applications of FPCA.

6.7.3.1 /t/

Five PCs are needed to account for 95% of the variance in /t/ spectra. PC1 accounts for 58.4% of the variance, PC2 accounts for 18.2%, and PC3 accounts for 9.3%. The influence of the first three PCs is shown in Figure 6.14. The top left plot shows the variance captured by the respective PCs. The other plots all show the mean spectrum as well as the shape of the mean spectrum when weighted by PC1–3, respectively. This gives an indication of what spectra with a relatively high score s_n and a relatively low s_n look like. The ‘low’ scores correspond to the mean spectrum weighted by the first quantile of all scores, and the ‘high’ scores correspond to the mean weighted by the third quantile of all scores.

The mean spectrum of /t/ shows little energy at lower frequencies, relatively high amplitude at frequencies between 1.5–4.5 kHz, a peak just below 2.5 kHz, and gradual loss of energy at higher frequencies. The main source of variance in the data (PC1) is the location and magnitude of the primary peak. High s_1 corresponds to a lower, more

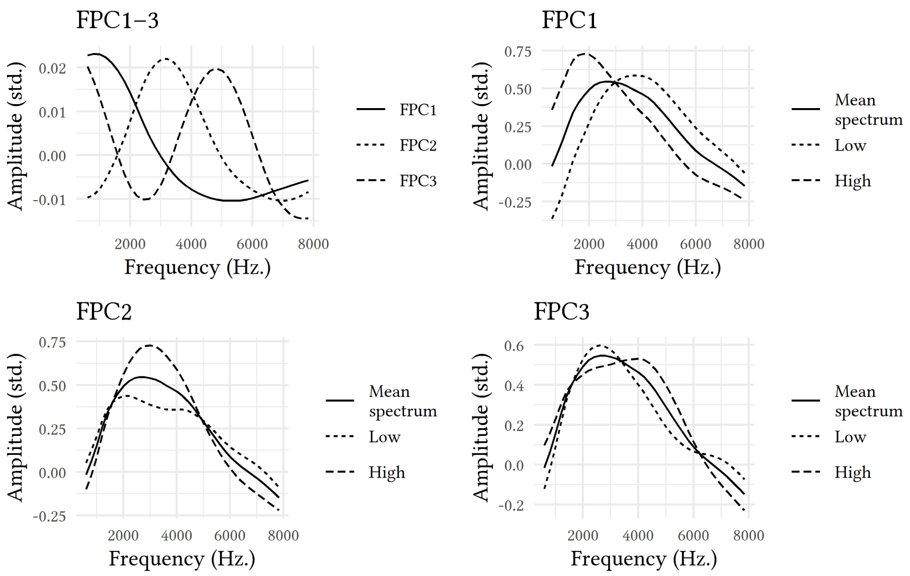

Figure 6.14: *Plots showing the variance in* /t/ *spectra captured by PC1–3 (top left) and the effect of weighting the mean by high and low scores of PC1–3.*

prominent peak and more energy at lower frequencies, while low s_1 corresponds to a higher peak and less energy at lower frequencies. Another important source of variance in the data (PC2) is in the magnitude of the peak. High s_2 corresponds to a more prominent peak, and correspondingly somewhat less energy at higher frequencies, while low s_2 corresponds to a less prominent peak, and somewhat more energy at higher frequencies. PC3 clearly accounts for less variance, and it mostly corresponds to the peakedness of the energy distribution. High s_3 corresponds to a more restricted and somewhat more prominent peak in the same location as in the mean spectrum, and low s_3 corresponds to a more broad distribution of energy.

It may be conceptually helpful to compare these results to spectral moments (see Section 5.3). The mean spectrum shows significant negative skew. PC1 interacts with both the mean of the distribution (i.e. COG) and skewness; higher s_1 corresponds to lower COG and more negative skew. PC2 mostly interacts with kurtosis, with high s_2 corre-

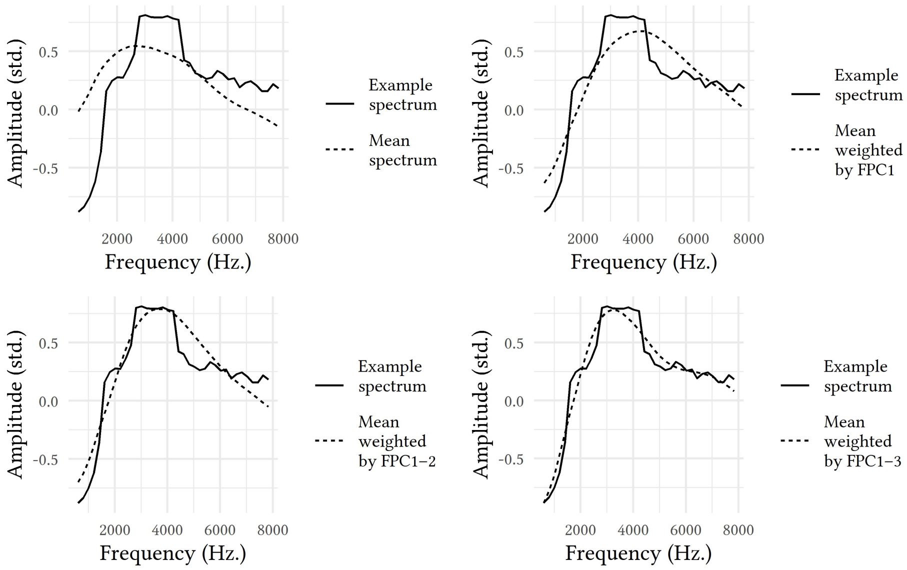

Figure 6.15: *Plots showing a random function* amp(f) *and the mean function* μ(f)*, and the effect of weighing* μ(f) *by* amp(f)*'s scores for FPC1–3.*

sponding to a more peaked distribution. PC3 interacts with all of the four first spectral moments, with high s_3 simultaneously corresponding to a higher COG, higher SD, less skew, and lower kurtosis.

For each input spectrum and each PC, a score is computed. These scores can be used to roughly reconstruct any input curve from the mean, using the equation in (4) (adapted from Gubian et al. 2015: 21).

(4) $amp(f) \approx \mu(f) + s_1 \cdot PC1(f) + s_2 \cdot PC2(f) + \ldots + s_n \cdot PC\mathrm{n}(f)$

The equation in (4) states that any input spectrum *amp(f)* can be approximated by weighting the mean spectrum *μ(f)* by *amp(f)*'s scores $s_{1\ldots n}$ multiplied by the PC functions *PC*1…n*(f)*. Figure 6.15 shows how this is implemented for a random spectrum *amp(f)* in the data, which has relatively low scores for PC1 and PC3, and a relatively high score for PC2. Note that PC1–3 together account for just 85.9% of variance in the data, so weighting the mean by PC1–3 for any curve will only yield a rough approximation of *amp(f)*; note also that *amp(f)* as seen in Figure 6.15 has not yet been smoothed. In the top left figure, we see *amp(f)* and $\mu(f)$. They have quite different shapes. In the top right figure, we see $\mu(f)$ weighted by $s_1 \cdot PC1(f)$ (in this case, s_1 = -27). This serves to move the primary peak in the weighted mean to a higher frequency, increase the energy somewhat at higher frequencies, and drastically decrease the energy at lower frequencies. In the bottom left figure, $\mu(f)$ is further weighted by $s_2 \cdot PC2(f)$ (in this case, s_2 = 7). This has the effect of increasing the energy at the peak. In the bottom right figure, $\mu(f)$ is further weighted by $s_3 \cdot PC3(f)$ (in this case, s_3 = -9.2). This has the effect of narrowing the main peak and skewing it towards somewhat lower frequencies, and aligning both the lowest and highest frequencies more with *amp(f)*. Considering that *amp(f)* in Figure 6.15 is not pre-smoothed, the weighted mean approximates the original function quite well; weighting the mean by the scores of lower-ranked PCs would, of course, yield an even closer approximation.

Figure 6.16 shows the mean PC scores by dialect area projected onto maps. It is immediately noteworthy that Djursland behaves differently from other dialect areas, in having either very low or very high PC scores. s_2 and s_3 do not differ much by dialect area, but s_1 shows quite different patterns in Mid-Eastern and Djursland Danish relative to the rest of the peninsula. The upshot is that mid-point spectra from /t/ releases in Mid-Eastern and Djursland Danish show amplitude peaks at higher frequencies, which suggests that these are areas with /t/-affrication. The upshot of the exceptional s_2–s_3 levels in Djursland is that this variety shows less of a narrow peak around 2.5 kHz in /t/ spectra, and generally a broader distribution of energy in the spectrum.

A GAMM was fitted with s_1 as the dependent variables and all candidate independent variables except STRESS. This model has a medium strong effect size of R^2 = .39. A likelihood ratio test found that a model which included the geographical variable performed significantly better than a nested model without this variable, with χ^2 (3) = 4.1, p = 0.043. The influence of the geographical variable is mapped in Figure 6.17; these patterns are similar to the corresponding map in Figure 6.16. They also show particularly low fitted s_1 in Djursland and the surrounding area. All other independent variables also significantly influence s_1 (see Table 6.8). s_1 is lower – i.e. there is

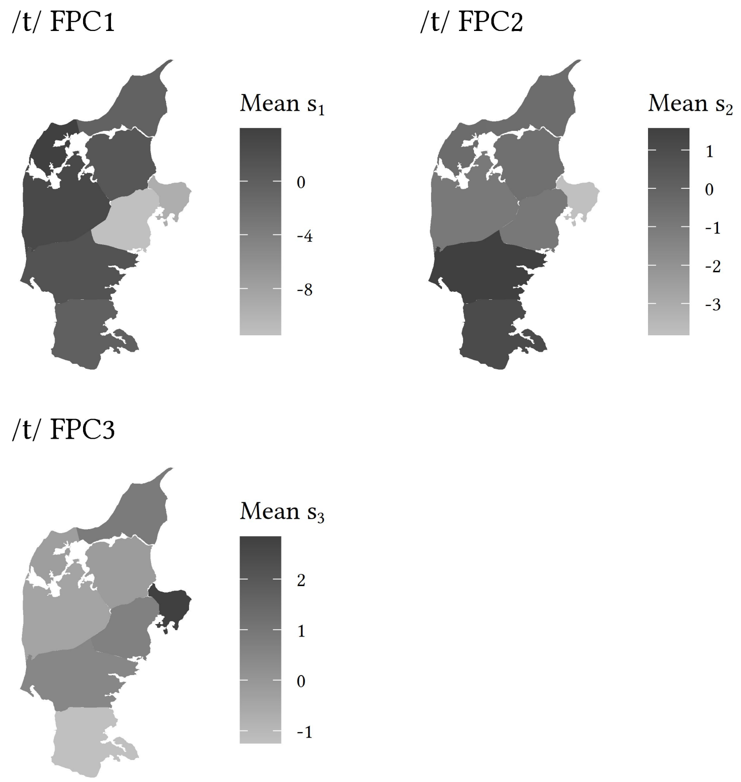

Figure 6.16: *Mean PC scores for /t/ spectra by dialect area projected onto maps.*

less energy at higher frequencies – before non-high, and particularly low, vowels; before rounded vowels; before back vowels; and in non-palatalized /t/ tokens. Men have much higher s_1 than women. Finally, there is an inverse relationship between VOT and s_1, such that lower VOT corresponds to higher s_1. These are all expected patterns, and similar to the results in Chapter 5.

Fewer variables contribute to the GAMM modeling s_2: HEIGHT, ROUNDNESS, STRESS, and BACKNESS. The geographical variable does not significantly improve the fit of this model. The final model has a medium strong effect size, with R^2 = .29. Only two variables significantly influence s_2, namely BACKNESS, with $\hat{\beta}$ = 1.97, SE = 0.7, t = 2.82,

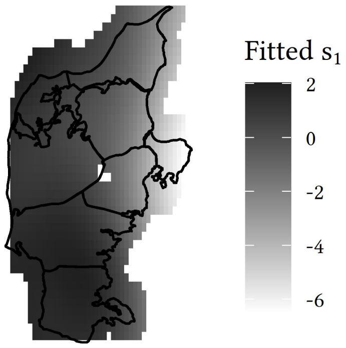

Figure 6.17: *Fitted* s_1 *values for /t/ attributable to area. Darker shading indicates higher fitted values. Black lines indicate traditional dialect boundaries.*

Table 6.8: *Parametric coefficients of generalized additive mixed model with* s_1 *of /t/ spectra as the independent variable.*

	estimate	**SE**	***t***	***p***	
intercept	-1.42	1.38	1.03	0.3	
HEIGHT: high *vs.* non-high	-3.65	0.74	-4.96	<.001	***
HEIGHT: low *vs.* mid	-4.38	0.91	-4.84	<.001	***
ROUNDNESS	3.47	0.97	3.58	<.001	***
BACKNESS	-3.91	1.21	-3.22	<.01	**
PALATALIZATION	-6.79	2.03	-3.35	<.001	***
SEX	6.26	1.87	3.35	<.001	***
VOICE ONSET TIME	-2.03	0.69	-2.95	<.01	**

$p < .01$, and ROUNDNESS, with $\hat{\beta} = 2.21$, $SE = 0.61$, $t = 3.6$, $p < .001$. In other words, high s_2, associated with an especially prominent energy peak around 2.5 kHz, is especially found before non-back and round vowels.

All variables contribute to the GAMM modeling s_3; the geographical variable does not significantly improve the fit of this model. The resulting model has a medium small effect size of $R^2 = .25$. The only variable that significantly influences s_3 is ROUNDNESS, with $\hat{\beta} = 2.62$, $SE = 0.45$, $t = -5.84$, $p < .001$. In the case of both s_2 and s_3, the few significant variables can certainly not fully

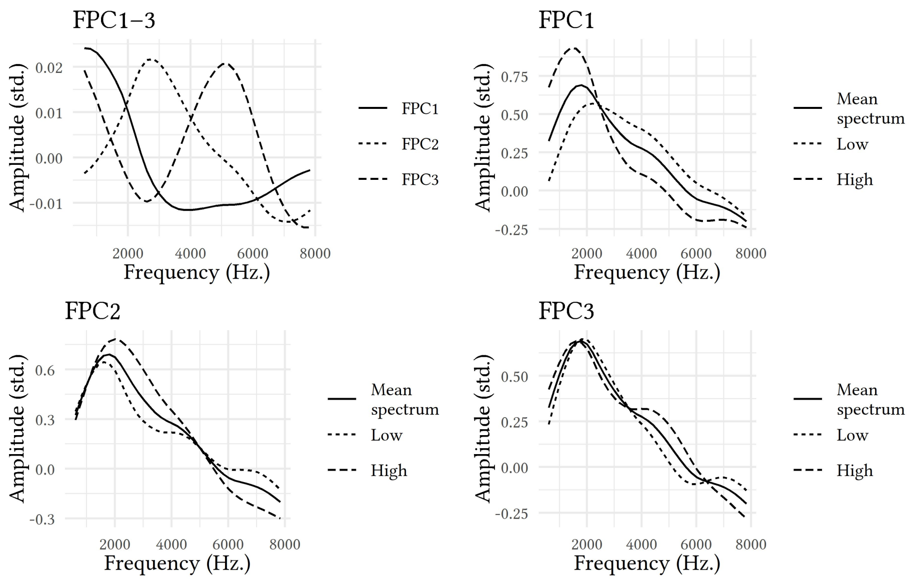

Figure 6.18: *Plots showing the variance in* /k/ *spectra captured by PC1–3 (top left) and the effect of weighting the mean by high and low scores of PC1–3.*

account for all variance in the data; lower-ranked PCs are often more difficult to interpret, and in turn, more difficult to account for.

6.7.3.2 /k/

Five PCs are needed to account for 95% of variance in the /k/ spectra. PC1 accounts for 54.4% of the variance, PC2 accounts for 19.9%, and PC3 accounts for 9.6%. The influence of the first three PCs are shown in Figure 6.18, which is structured like Figure 6.14 above.

PC1–3 look strikingly similar for /t/ and /k/, but it should be kept in mind that the mean /k/ spectrum looks quite different from the mean /t/ spectrum. The main peak of the mean /k/ spectrum falls just below 2 kHz; energy gradually decreases at higher frequencies, and there is very little energy above 6 kHz. PC1 corresponds to more energy at lower frequencies, and less energy at higher frequencies. This means that a token with high s_1 has more energy at very low frequencies relative to the mean, a more prominent peak at slightly

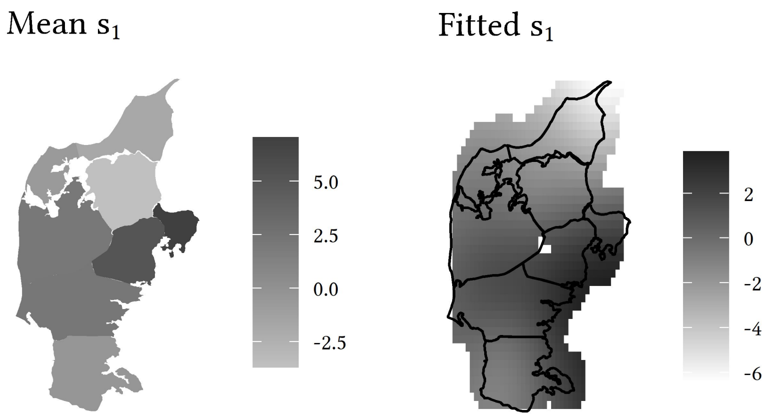

Figure 6.19: *Mean and fitted* s_1 *values for* /k/ *projected onto maps. Darker shading indicates higher* s_1 *values. Black lines indicate traditional dialect boundaries.*

lower frequencies, and much less energy at frequencies above 2.5 kHz. Low s_1 corresponds to a less prominent peak located somewhat higher than the mean, and more equally distributed energy at higher frequencies. PC1 seemingly interacts with all four spectral moments, and low s_1 shows characteristics of velar affrication. PC2 corresponds to more energy at higher frequencies, particularly around 2.5–3 kHz. Tokens with high s_2 have a more prominent peak at slightly higher frequencies than the mean, and generally more energy at frequencies between 1.5–4.5 kHz. PC3 corresponds to more energy both at the lowest frequencies and at frequencies well above the peak. The most prominent difference in a token with high s_3 is a somewhat more even distribution of energy above the peak and below 6 kHz.

A GAMM modeling s_1 was fitted with all candidate independent variables except SEX. This model has a medium strong effect size of $R^2 = .33$. A likelihood ratio test found that the geographical variable significantly improves the fit of such a model, with $\chi^2\ (3) = 4.61$, $p = 0.026$. Figure 6.19 shows two maps: on the left, mean s_1 by dialect area is projected onto a map, and on the right, fitted s_1 values from the GAMM projected onto a map This map shows particularly high s_1 in the area around Djursland and Mid-Eastern Jutland, and particularly

Table 6.9: *Parametric coefficients of generalized additive mixed model with* s_1 *of /k/ spectra as the independent variable.*

	estimate	**SE**	***t***	***p***	
intercept	-3.12	0.96	-3.25	<.01	**
HEIGHT: high *vs.* non-high	-4.21	0.86	-4.88	<.001	***
HEIGHT: low *vs.* mid	-5.1	0.97	-5.24	<.001	***
ROUNDNESS	5.23	1.35	3.89	<.001	***
BACKNESS	-5.51	1.26	-4.38	<.001	***
PALATALIZATION	-8.26	1.76	-4.7	<.001	***
STRESS	-1.4	0.64	-2.2	0.028	*
VOICE ONSET TIME	-1.67	0.69	-2.41	0.016	*

low s_1 in the Southern and South-Eastern areas. This is also essentially what the GAMM finds.

All other variables also significantly influence s_1, as seen in the summary in Table 6.9. These results show that phonetic context has a similar influence on s_1 in /k/ spectra and /t/ spectra, and the resulting patterns are similar to what was found for Modern Standard Danish in Chapter 5. s_1 is higher, i.e. energy is more concentrated at lower frequencies, before non-high vowels, particularly low vowels; before rounded vowels; before back vowels; in non-palatalized tokens; and in unstressed syllables. There is an inverse relationship between VOT and s_1, such that lower VOT correlates with higher s_1. These are all contexts where we would not expect affrication.

A separate GAMM was fitted modeling s_2, where only HEIGHT, BACKNESS, PALATALIZATION, and SEX contribute. The geographical variable was excluded from this model, as it did not significantly improve the model fit. The model has a medium small effect size of $R^2 = .244$. Other than the intercept β_0, only two variables significantly influence s_2: BACKNESS, with $\hat{\beta} = 4.66$, $SE = 0.63$, $t = 7.43$, $p < .001$, and PALATALIZATION, with $\hat{\beta} = 4.17$, $SE = 1.05$, $t = 3.96$, $p < .001$. Unsurprisingly, high s_2, corresponding to more energy at higher frequencies, is found before non-back vowels and in palatalized /k/ tokens.

Finally, a GAMM modeling s_3 was fitted. Only HEIGHT, BACKNESS, PALATALIZATION and VOICE ONSET TIME contribute to this model. A likelihood ratio test found that the geographical variable does not

Table 6.10: *Parametric coefficients of generalized additive mixed model with* s_3 *of* /k/ *spectra as the independent variable.*

	estimate	SE	t	p	
intercept	-0.88	0.44	-2	0.045	*
HEIGHT: high *vs.* non-high	-0.93	0.4	-2.33	0.02	*
HEIGHT: low *vs.* mid	-0.63	0.37	-1.72	0.086	.
BACKNESS	-1.14	0.44	-2.6	<.01	**
PALATALIZATION	-1.89	0.71	-2.65	<.01	**
VOICE ONSET TIME	-0.68	0.31	-2.15	0.031	*

significantly improve the model fit. The final model has a medium small effect size of R^2 = .237. Several of the variables show a significant influence on s_3; the results are summarized in Table 6.10. The upshot is that high s_3, which corresponds to a broader distribution of energy at higher frequencies, is found before non-high vowels, before back vowels, in non-palatalized tokens, and in tokens with shorter VOT. This is tricky to interpret. A possible explanation is that PC3, more so than the higher PCs, signifies a bimodal energy distribution in the spectrum; in this case, high s_3 may be indicative of a (weak) formant structure midway through the release, which we may well expect in these particular contexts. (PC3 in /t/ spectra may have a similar explanation, but this would not particularly help explain the results of the statistical model of s_3 in /t/.)

6.7.3.3 /p/

Five PCs are needed to account for 95% of variance in the /p/ spectra. PC1 accounts for 54.3% of the variance, PC2 accounts for 20.6%, and PC3 accounts for 9.5%. The influence of the first three PCs is shown in Figure 6.20, which is structured in the same way as Figures 6.14 and 6.18.

The mean /p/ spectrum has a fair amount of energy at lower frequencies, peaks at around 1.5 kHz, and shows linear reduction in energy at higher frequencies. PC1 is associated with much more energy at lower frequencies, and somewhat less energy at frequencies above 2.5 kHz. A token with high s_1 has a low and prominent peak relative to the mean. PC2 is associated with more energy between 2–4 kHz.

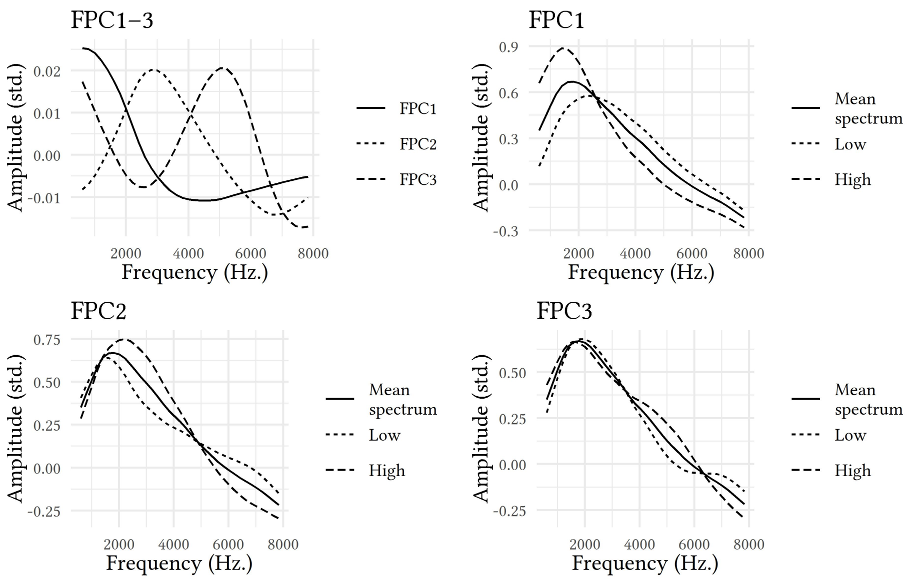

Figure 6.20: *Plots showing the variance in* /p/ *spectra captured by PC1–3 (top left) and the effect of weighting the mean by high and low scores of PC1–3.*

A token with high s_2 has a higher and more prominent peak than the mean, and generally more energy at mid-high frequencies. PC3 is associated with more energy both at the lowest frequencies and around 5 kHz. High s_3 seems to add a second, less prominent peak at higher frequencies around 4–5 kHz; I suggested that a similar pattern in /k/ spectra could be due to weak formant activity during the stop release.

Geography is not a significant predictor of any of the /p/ PCs.

In the GAMM modeling s_1, all variables except BACKNESS and VOICE ONSET TIME contribute. The model has a medium low effect size with $R^2 = .281$. Only two variables significantly influence s_1, namely HEIGHT (high *vs.* non-high) with $\hat{\beta} = -4.24$, $SE = 1.26$, $t = -3.36$, $p < .001$, and SEX with $\hat{\beta} = 4.45$, $SE = 1.76$, $t = 2.53$, $p = 0.012$. The upshot is that s_1 is higher, i.e. energy in the spectrum peaks at lower frequencies and has a narrower distribution, before non-high vowels and in tokens from male speakers.

Table 6.11: *Parametric coefficients of generalized additive mixed model with* s_2 *of* /p/ *spectra as the independent variable.*

	estimate	**SE**	***t***	***p***	
intercept	0.45	0.58	0.77	0.44	
HEIGHT: high *vs.* non-high	3.25	0.9	3.62	<.001	***
HEIGHT: low *vs.* mid	1.96	0.97	2.02	0.043	*
BACKNESS	3.92	0.84	4.67	<.001	***
STRESS	0.68	0.65	1.06	0.292	
SEX	-2.57	1.1	-2.35	0.019	*

In the GAMM modeling s_2, HEIGHT, BACKNESS, STRESS, and SEX were the only variables to contribute. This model has a medium high effect size of R^2 = .329. Several variables significantly influence s_2; the model is summarized in Table 6.11. The results show that s_2 – associated with an energy peak at a higher frequency than the mean, and more evenly distributed energy at higher frequencies – is found before mid and particularly high vowels, before non-back vowels, and in female speakers. This is unsurprising, and in line with the results from Chapter 5.

Finally, the GAMM modeling s_3 is not very informative; only the HEIGHT variable contributes, and the model has a rather low effect size of R^2 = .178. The 'high *vs.* non-high' contrast is significant, with $\hat{\beta}$ = -1.73, *SE* = 0.53, t = -3.27, p = 0.001. This indicates that the somewhat more bimodal spectral energy distribution associated with s_3 is found particularly before non-high vowels; this is sensible, but given the low effect size, this model should be taken with a grain of salt.

6.7.4 Interim conclusions

In this section, I have applied FPCA to multitaper spectra computed from the midpoint of stop releases. This method was used to summarize main sources of variance in the spectra, which were then analyzed with regression models. It was shown that midpoint spectra in /p t k/ vary in largely analogous ways, the two most immediately interpretable ones being 1) the location of the main energy peak, and 2) the prominence of that peak. These sources of variance are to a high extent determined by phonetic context in predictable ways, similar to what we saw for Modern Standard Danish in Chapter 5. For example, the main peak is usually higher before high, non-back, and non-round vowels, in palatalized tokens, in stressed syllables, and in female speakers. It is an important finding in itself that the results are largely in line with expectations, since this is the first study to use FPCA to analyze variance in speech spectra.

In /t k/, the primary source of variance s_1 displayed regional variation. In /t/, s_1 was particularly low in Djursland and the area immediately to the north-west of Djursland, indicating that /t/ in this area is associated with a main energy peak at relatively high frequencies at the release midpoint; in other words, /t/-affrication is relatively common in this area, but otherwise not common in the traditional varieties of Jutland. The pattern is different in /k/; here, s_1 is high in Djursland and surrounding areas, but very low in the Northern and parts of the North-Eastern dialect areas. Very low s_1 in /k/ corresponds to a peak at high frequencies relative to the mean, but also a broader distribution of energy in high frequencies, which is consistent with velar friction. As such, regional patterns of affrication at the release midpoints are different in /t/ and /k/; there are no signs of regional patterns of affrication in /p/. These results, as well as the results for VOT and the observations on closure voicing, are discussed in more detail below.

6.8 Discussion

In this section, I will discuss the results presented above from three different perspectives. In Section 6.8.1, I discuss how the different patterns of VOT, spectral energy distribution, and closure voicing in Jutlandic varieties compare to Modern Standard Danish. In Section 6.8.2, I briefly discuss possible parallels between the phonetic findings in this chapter and regional variation in stop gradation (see Chapter 3). Finally, in Section 6.8.3, I hypothesize about possible sources of the observed variation; I argue that the variation can shed a light on both lesser understood aspects of the early stages of stop gradation, and on sociolinguistic aspects of the spread of the standard language.

6.8.1 Comparison with Modern Standard Danish

In this chapter, I have reported the results of corpus studies on VOT and spectral characteristics in the stops of Jutlandic varieties of Danish. I have also provided some preliminary data on closure voicing in these varieties. These topics have all been discussed for Modern Standard Danish earlier in the dissertation, so it is worth comparing Jutlandic varieties and Modern Standard Danish.

In Section 2.3.1, I mentioned several previous studies of VOT in Modern Standard Danish. Those studies are not directly comparable to the one presented here due to differences in methodology. However, even without direct comparison, it is obvious that the difference in VOT between /b d g/ and /p t k/ is more extensive in Modern Standard Danish than in the Jutlandic varieties. This signals that the laryngeal contrast is cued differently across varieties. The northern varieties as well as Mid-Western show particularly small differences in VOT between /b d g/ and /p t k/; in northern varieties, we further see closure voicing in contexts where it would be rare or simply non-existent in Modern Standard Danish.

In Chapter 4, I discussed how a reported glottal spreading gesture in Modern Standard Danish /b d g/ (Frøkjær-Jensen et al. 1971; Jeel 1975; Hutters 1985) may account for the rarity of intervocalic voicing in these sounds. Some speakers of (particularly northern varieties of) Jutlandic have continuous voicing in stressed syllables and following stød, which strongly implies that these speakers do not have such a phonologized glottal spreading gesture. Likewise, comparison with the DanPASS corpus (see Table 6.4) shows that VOT in /p t k/ is generally shorter in Jutlandic varieties than in Modern Standard Danish. This is also particularly the case for northern varieties. This implies that the glottal spreading gesture in /p t k/ in these varieties is timed differently and/or is of a smaller magnitude than in Modern Standard Danish.

With regards to spectral characteristics of stop releases, the results in Section 6.7 can not be directly compared to those in Chapter 5, but it can be inferred that phonetic and phonological environment affects stop releases in similar ways in Jutlandic varieties and Modern Standard Danish. The mean spectral energy distribution at /t/ release midpoints is, however, quite different from Modern Standard Danish in most of the traditional Jutlandic varieties. In Modern Standard Danish, the /t/ release midpoint shows spectral properties that are consistent with alveolar frication, regardless of phonetic and phonological context. The mean energy peak in the Jutlandic varieties is much lower, suggesting that affrication is *not* an invariant feature of /t/ in these varieties, except perhaps in Djursland and immediately surrounding areas. As in Modern Standard Danish, /k/ spectra in Jutlandic varieties show strong coarticulation effects. In addition, they also show regional variation, with energy peaks at relatively high frequencies (consistent with affrication) being common in the far north-west of the peninsula.

The underlying articulatory mechanisms of these seemingly unrelated findings are likely related. If we take Northern Jutlandic as an example, this variety shows low VOT across the board, widespread intervocalic voicing, and no /t/-affrication. These findings can all be explained if we assume a glottal spreading gesture that peaks early in /p t k/, and no glottal spreading gesture in /b d g/. The early-peaking gesture in /p t k/ ensures short VOT, but it also ensures higher intraoral air pressure at the time of the release. This in turn ensures a prominent burst, and removes the 'need' for /t/-affrication. The lack of a glottal spreading gesture in /b d g/ is functionally motivated by the need to maximally disperse the laryngeal contrast; this in turn ensures short VOT, and a greater propensity for intervocalic voicing. These hypotheses are in principle empirical questions, but they may not be practically testable given the current status of these varieties (see Section 6.3).

6.8.2 Parallels between phonetics and phonology

VOT was found to be low in the center of the peninsula, and VOT in /p t k/ in particular was found to be quite low in the far north; VOT was rather high around Djursland, and VOT in /p t k/ in particular was high in the far south. These results show parallels to regional variation in stop gradation in Jutland (following Bennike and Kristensen 1898–1912; see Sections 2.4.4 and Chapter 3 for more discussion of stop gradation, and Section 2.4.4 for maps showing the various outcomes of some of these processes).

In the center of the peninsula, stop gradation generally resulted in voiced fricatives across the board, although Old Danish /t/ evolved further into a (post-)alveolar approximant [j ~ ɹ] in some areas. In the far north of the peninsula, stop gradation often progressed further, with Old Danish /t/ being lost completely, and Old Danish /k/ developing into [j] after front vowels.[15] In other words, areas with low VOT also showed significant reduction of Old Danish stops, resulting in voiced fricatives or sonorants. In southern parts of the peninsula, however, Old Danish /p k/ developed into voiceless fricatives; and in the far south (e.g. in Viöl, present-day Germany), /p t/ were both retained as stops in most contexts (Bjerrum 1944). In areas with high VOT, reduction was more constrained, and generally resulted in voiceless obstruents.

I suggested above that regional differences in VOT are the result of differences in the timing and magnitude of phonologized glottal spreading gestures. The differences in stop gradation follow directly from this observation: /b d g/ are more likely to alternate with voiceless fricatives in areas with a glottal spreading gesture, and more likely to alternate with voiced fricatives in an area without such a gesture. Further reduction to voiced approximants is unsurprisingly more common in areas where voiced fricatives developed in the first place.

6.8.3 The spread of aspiration and affrication

In Section 3.5.3, I argued that /b d g/ in Danish were likely voiced by the time the first steps of stop gradation took place; this is rather difficult to verify. It is easier to verify that /t/-affrication is a relatively recent development in Standard Danish (Brink and Lund 1975).

If we assume that /b d g/ were voiced in earlier stages of the language, and /t/ was not affricated, then the widespread Jutlandic pattern of short VOT and non-affricated /t/ requires no explanation. However, two findings from this chapter do need to be explained: 1) Why do southern Jutlandic varieties have relatively high VOT across the board, and 2) why does Djursland and surrounding areas to the south have relatively high VOT and /t/-affrication? In other

[15] The situation is somewhat more complex in the area around Djursland, where Old Danish /p/ was retained as a stop, but /t/ developed into an approximant.

words, we do not need to explain why many areas are different from Modern Standard Danish, but rather why a few areas share similarities with Modern Standard Danish. I argue in this section that southern Jutlandic traditionally did *not* have a voicing-based contrast, and that the area around Djursland was relatively early in adapting prominent aspiration and /t/-affrication from Modern Standard Danish.

Step 2 of stop gradation is repeated in (5):

(5) /b d g/ → [β ð ɣ] / weak position

As mentioned in the previous section, a more general version of (5), viz. loss of closure in weak position, is found in almost all varieties of Danish. In most varieties, the result was indeed [β ð ɣ], and these often weakened further into approximants or other sonorants. In southern Jutland, however, /b g/ developed into [f χ] in weak position. This suggests that Step 2 of stop gradation resulted in spirantization of /b d g/ without changing their voicing status. Just as the result of stop gradation in Modern Standard Danish is difficult to explain if /b d g/ were voiceless at the time of Step 2, it is difficult to explain why the outcome in southern Jutlandic varieties would include [f χ] *unless* /b g/ were voiceless at the time. This is supported by the relatively high VOT found in southern Jutlandic varieties in the 20th century. A possible reason why the historical southern Jutlandic varieties did not have voiced stops is areal influence from German.

The situation is different in the area surrounding Djursland, where Step 2 generally resulted in voiced fricatives or glides, as in Modern Standard Danish.[16] This suggests that in this area, unlike in southern Jutland, /b d g/ were voiced at the time of Step 2. In this area, lengthy VOT is a more recent innovation. Since /t/-affrication is also found in this area, both features were likely borrowed from Modern Standard Danish relatively recently. This calls for a sociolinguistic explanation.

[16]In fact, Bennike and Kristensen (1898–1912: maps 49–52) report a lot of variation in this area; Old Danish /p/ is mostly realized as [v], but a small area north-west of Djursland retains a stop which the authors transcribe as [b], but which may well be voiceless; Old Danish [t] has a lot of reflexes in this general area, including [ð ɹ j ɣ]; Old Danish [k] is consistently realized as [ɣ]. See Section 2.5.3 for further details.

The findings are in line with the *cascade model* of interdialectal influence (Labov 2003), which is a more general version of Trudgill's (1974, 1983) *gravity model.* Both models predict that change does not spread purely geographically, but rather spreads between population centers in a manner that is predictable from a combination of population size and geographical distance. From that perspective, high VOT and /t/-affrication in and around Djursland, particularly south of Djursland along the east coast, may be due to Modern Standard Danish influence first reaching Aarhus (the largest city of Jutland) and spreading from there to other major cities. Such an effect would of course be somewhat obscured in this study, as the DS recordings all feature speakers from rural areas, but the cascade model predicts that rural areas sufficiently close to cities will be affected by changes originating in those cities relatively early. There is no indication of a similar effect in the northern part of the peninsula, where the two largest urban areas at the turn of the 20th century (Aalborg and Randers) show no signs of high VOT or /t/-affrication.[17] Locations of major cities in Jutland at the turn of the 20th century are shown in Figure 6.21.

It has been argued that a simpler and more effective way to account for aggregated linguistic variation is to view variation simply as a sublinear function of geographical distance (Séguy 1971; Nerbonne and Heeringa 2007; Nerbonne 2009).[18] Geographical distance in itself, however, can be a poor predictor in a country such as Denmark, where a significant portion of the population live on islands.[19] The cascade model seems to work quite well for our purposes, since Aarhus – a coastal city quite far removed from Copenhagen – *does* appear to be ground zero of the innovations of long VOT and /t/-affrication in Jutland.

---

[17] Likewise, the city of Esbjerg is in an area with rather low VOT and no signs of /t/-affrication. Esbjerg is a special case, as it was a very young city at the turn of the 20th century, settled only a few decades earlier (Matthiessen 1985).

[18] Others have criticized the cascade model for taking into account only population size and density, and not other social factors such as prestige and age (e.g. Bailey et al. 1993; Boberg 2000; Horvath and Horvath 2001).

[19] A possible solution is calculating traveling time instead of raw geographical distance, following Gooskens (2005).

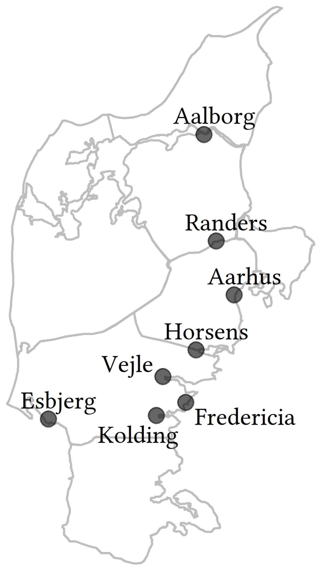

Figure 6.21: *Cities with population sizes above 10,000 in the year 1901. Data from Matthiessen (1985).*

To recap, I have argued that long overall VOT in southern Jutland, especially in /p t k/, is an old feature of these varieties, and that this is reflected in how stop gradation progressed in this area. Conversely, I have argued that long overall VOT and /t/-affrication in the area surrounding Djursland is a recent innovation which spread directly from Modern Standard Danish to the highly urbanized central–southern east coast of Jutland, most clearly targeting the area around the largest city, viz. Aarhus.

6.9 Conclusions

In this chapter, I have provided the first English-language introduction to the DS (1971–1976) legacy corpus of Danish dialect recordings from the 1970s. I further used the recordings from Jutland for two large-scale quantitative studies of phonetic variation. It is the first time that this incredible resource has been used systematically for research.

In the first study, I compare the VOT of /b d g p t k/ from 213 recordings across the peninsula. The data are analyzed using a spatial generalized additive mixed model which also accounts for a number of contextual variables that are known to affect VOT. In the second study, I compare spectral characteristics of stop release midpoints of /p t k/ in the same recordings. Here, I use functional principal component analysis to summarize the main sources of variance in release spectra; principal component scores then serve as dependent variables in (potentially) spatial generalized additive mixed models. Finally, I provided some preliminary data on closure voicing in the northern varieties in particular.

The results of the first study show that VOT is generally rather low in the Jutlandic varieties relative to Modern Standard Danish, and that there are complex patterns of regional variation. The second study shows that the main source of variation in /p t k/ release midpoint spectra is the frequency and prominence of the primary energy peak; in /t k/, this source of variance is partially dependent on geography, although in different ways. In Djursland, for example, /t/ spectra show a prominent peak at relatively high frequencies, while /k/ spectra show a prominent peak at relatively low frequencies; this indicates that /t/ is affricated, whereas /k/ is not.

I have suggested in earlier chapters that Danish stops used to have a voicing-based laryngeal contrast, and we know from previous research that /t/-affrication in Standard Danish is a relatively recent development. On the basis of this, I suggested that only the relatively long VOT in some areas and the /t/-affrication in some areas require an explanation; short VOT and unaffricated /t/ are the default in Jutlandic Danish. On the basis of earlier phonological developments, I suggest that the relatively long VOT in the Southern variety has long been a feature of that dialect. I further suggest that the relatively long VOT and /t/-affrication in Djursland and surrounding areas are borrowed features from the Copenhagen area, which spread directly to the major city of Aarhus and from there to nearby cities and the surrounding rural areas.