THE /t/ RELEASE IN JUTLAND DANISH: DECOMPOSING THE SPECTRUM WITH FUNCTIONAL PCA

Rasmus Puggaard-Rode

Institute for Phonetics and Speech Processing (IPS), LMU Munich
r.puggaard@phonetik.uni-muenchen.de

ABSTRACT

This study concerns variation in the realization of /t/ in Danish. /t/ is prominently affricated in Modern Standard Danish, but affrication is known to have been missing in some of the regional varieties that used to be spoken in Jutland. It remains unclear exactly which varieties lacked affrication. Due to relatively recent major changes in language policy and mobility, the traditional Danish dialect landscape has undergone rapid levelling, so we explore this problem using a large corpus of recordings of elderly speakers from the 1970s. We propose using the functional extension of principal component analysis to decompose the sources of spectral variance in stop release midpoints, and use the results as a proxy for determining affrication status. We find that /t/ affrication was limited to parts of eastern Jutland, perhaps due to early adaptation of the feature in some of the area's largest cities.

Keywords: Danish, functional PCA, regional variation, spectral decomposition, stop affrication

1. INTRODUCTION

The aims of this study are twofold: 1) to explore the geographical distribution of (missing) /t/ affrication in traditional varieties of Jutland Danish (see Fig. 1), and 2) to explore the use of functional principal component analysis (FPCA) [1] for compressing the information in speech spectra into a small number of easily interpretable discrete variables.

In Modern Standard Danish (MSD), /p t k/ are voiceless aspirated, and /t/ in particular is prominently affricated [2, 3, 4]. Until relatively recently, the majority of the Danish speech community spoke traditional dialects [5], but in the past century, rampant standardization has largely led to the levelling of these varieties [6]. An overt feature of some Jutlandic varieties was the use of a variant of /t/ without salient affrication, which is known colloquially as *tørt t* 'dry t' [7]. There is no consensus about which Jutlandic varieties lacked affrication. Sources have variously claimed that it was missing from all

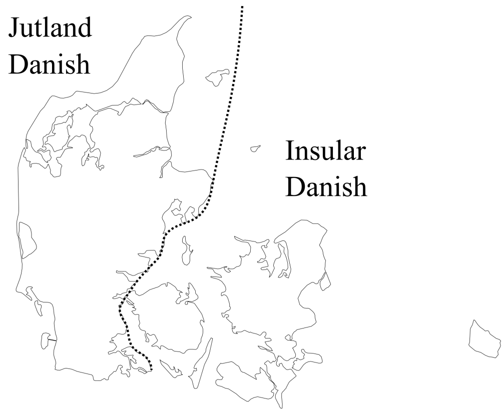

Figure 1: Map showing the areas associated with the two primary Danish dialect groups.

Jutlandic varieties [8], Northern Jutlandic [9], Western Jutlandic [10], or all varieties except Eastern Jutlandic [11]. We explore this mystery using a large legacy corpus of sociolinguistic interviews which was explicitly recorded to document the traditional language and culture of rural Denmark [12, 13]. These recordings (particularly the ones from Jutland) largely preserve a stage of language variation unaffected by the recent wave of standardization.

The acoustics of alveolar frication and aspiration are well-understood. In alveolar frication, a jet of air impinges on a hard surface (the upper front teeth) immediately in front of the coronal constriction, resulting in noise in a broad range of high frequencies (mainly above 4 kHz) [14]. In aspiration, lower frequency turbulence noise in a narrower frequency range (mainly below 1 kHz) is generated at and near the glottis [15]. This noise is filtered through the supraglottal cavity, where the resonance frequencies of the following vowel significantly affect the resulting noise. Analyzing how these spectral characteristics vary regionally is not at all straightforward.

Spectral moments are commonly used to decompose spectra generated from aperiodic portions of speech. The spectral mean, or *center of gravity* (COG) [16], is a particularly popular measure. Since spectra rarely resemble Gaussian distributions, the mean frequency alone is not very informative with respect to overall shape; furthermore, studies have found conflicting evidence about which spectral moments (if any) can be used to consistently determine noise sources [17, 18]. Several other measures have been proposed for decomposing aperiodic spectra [19, 20, 21], but none have been as influential as COG.

For the reasons outlined above, we do not use spectral moments here. Instead, we adopt FPCA as an alternative method. FPCA is used to determine the main sources of variance in curves without imposing predetermined shapes. This makes it particularly suitable for exploratory work. Each resulting principal component (PC_i) corresponds to a source of variance relative to the average spectral shape, and each individual spectrum gets a score s_i for each PC indicating how well it corresponds to this source of variance. These scores can then be fitted to conventional statistical models; here, we fit PC scores to spatial generalized additive mixed models (GAMMs) [22] in order to test whether any of the variance in spectral shape is geographically determined. These models also include a range of phonetic contextual predictors. The influence of phonetic contextual predictors largely pattern as predicted, which we take as an indication that FPCA is indeed a suitable tool for decomposing spectra. FPCA has previously been used to analyze various time-varying phonetic measures, such as F_0 and formants [1, 23]; to our knowledge, it has not previously been used to analyze speech spectra.

2. METHODS AND MATERIALS

2.1. Recordings

Our materials consist of sociolinguistic interviews with elderly speakers of traditional regional varieties recorded between 1971–1976 [13, 24]. Interviews were conducted with informants in 230 parishes in Jutland, of which 213 were used in this study. 23% of the informants are women. Informants' median year of birth is 1896 (range 1871–1927). The geographical distribution of informants can be seen in Fig. 2 (which shows the location of Djursland for reference.) In spite of the large scope of these recordings, they have not been used for systematic research until very recently [7, 25].

2.2. Acoustic processing

5,169 /t/ releases were manually segmented. The release burst and the onset of periodicity were used as delimitation landmarks, as identified from the waveform. Using Praat [26], 5 ms snippets were extracted from the midpoint of each /t/ release, filtered to in-

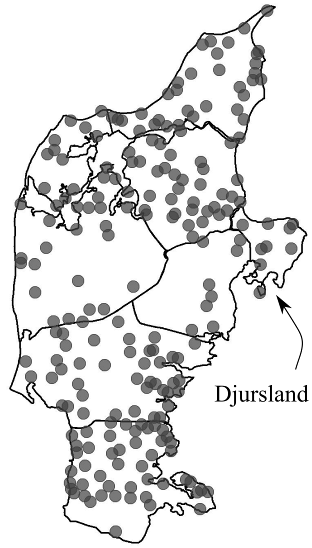

Figure 2: The geographical distribution of recordings used in this study.

clude only frequencies between 0.5–12.5 kHz. For each of these, multitaper spectra were generated in R [27] following the method described in [28]. Intensity values were log-transformed and standardized in order to keep all spectra on the same scale. Measurements at frequencies > 8 kHz were not included in the analysis, as they did not contribute much to the analysis other than noise.

2.3. Statistical analysis

All statistics were calculated in R [27]. The package `fdapace` [29] was used for FPCA. Spectra were smoothed using a local linear smoother with parameters set automatically using generalized cross-validation. Five PCs are needed to account for 95% of spectral variance. We focus on the first three $PC_{1–3}$, which account for 85.9% of the variance.

PC scores $s_{1–3}$ served as dependent variables in separate GAMMs with the following maximal structure: s_i ~ LON,LAT + HEIGHT + STRESS + SEX + BACK + ROUND + PALATALIZATION + VOT. (HEIGHT, ROUND, BACK, and STRESS refer to features of the following vowel; SEX refers to the sex of the speaker.) Most of these features have been shown to affect spectral shape in MSD stop releases [4]. There were various phonological palatalization rules in different Jutland Danish varieties [30]; we make no distinction between different sources of palatalization. All categorical variables are contrast coded [31]; the HEIGHT variable tests two Helmert contrasts, viz. 'high *vs.* non-high' and 'low *vs.* mid'. All other categorical variables are binary and coded with sum contrasts. The VOT variable is standardized. By-speaker random slopes were included for each variable except SEX. The geographical variable LON,LAT is modelled with two-dimensional thin plate regression spline smooths [32]. GAMMs were fitted using the package `mgcv` [33]. The packages `itsadug` [34] and `mgcViz` [35] were used for health checks and visualizations of the models.

The source material is spontaneous speech, so the data are rather imbalanced: 41.9% of /t/ tokens were followed by non-high vowels, of which 15.4% were low. 33.7% of tokens were found in stressed syllables. 11.5% of tokens were followed by back vowels, and 22.7% were followed by rounded vowels. Only 2.5% of the tokens were palatalized.

We used a step-up model selection procedure, adding variables in the order given above (based on these variables' influence on MSD /t/ release spectra [4]). Variables were only kept if the cost of fitting a more complex model was sufficiently outweighed by an improved model fit, as determined with likelihood ratio tests [36]. Data and annotated analysis code are available online [37]. More details of the study are available in my dissertation [38].

3. RESULTS

PC_1 accounts for 58.4% of spectral variance, PC_2 accounts for 18.2%, and PC_3 accounts for 9.3%. The variance explained by $PC_{1–3}$ is visualized in Fig. 3. The plots all show the mean spectrum and the shape of the mean spectrum when weighted by the first and third quantile of all PC scores. This gives an indication of what spectra with a relatively high score s_i and a relatively low s_i look like.

The average spectrum has little energy below 1.5 kHz, relatively high amplitude at frequencies between 1.5–4.5 kHz, gradual energy loss above that, and peaks just below 2.5 kHz. The main source of variance, captured by PC_1, is the location and magnitude of the primary peak. Positive s_1 corresponds to a lower frequency peak, and generally more energy at lower frequencies. COG would likely have captured some of the same information as PC_1, but they are not equivalent: in addition to the location of the peak, PC_1 also captures information about variance in spectral slope, peak amplitude, and skew. Given what we know about aspiration and alveolar frication, strong negative s_1 seems to represent an alveolar noise source, and strong positive s_1 seems to represent a glottal noise source. PC_2 captures variance in the magnitude of the main peak. Positive s_2 corresponds to a more prominent peak, and less energy at higher frequencies, while negative s_2 corresponds to a less prominent peak, and more energy at higher frequencies. PC_3 mostly captures information about the peakedness of the energy distribution. Positive s_3 corresponds to a more restricted and more prominent peak in the same location as in the mean spectrum, and negative s_3 corresponds to a broader distribution of energy.

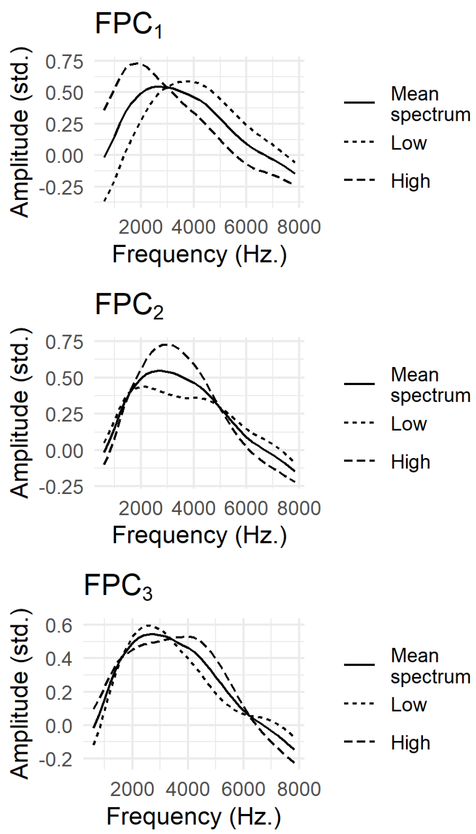

Figure 3: Variance in release spectra captured by $PC_{1–3}$ and the effect of weighting the mean by high and low scores $s_{1–3}$.

The GAMM modelling s_1 was fitted with 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 including the geographical variable LON,LAT performed significantly better than a nested model without this variable, with $\chi^2(3) = 4.1$, $p = 0.043$. The influence of the geographical variable is plotted on a map in Fig. 4; Fig. 4 shows that negative s_1 is common in Djursland and the surrounding area in mid-eastern Jutland (see Fig. 2); we return to this below. All other variables also significantly influence s_1 (see Table 1). The upshot is that there is less energy at high frequencies before non-high, particularly low, vowels; before rounded vowels; before back vowels; in non-palatalized releases; and in tokens from male speakers. Finally, there is an inverse relationship between VOT and s_1, such that shorter VOT corresponds to higher s_1. These are all expected patterns, similar to what is found in MSD [4].

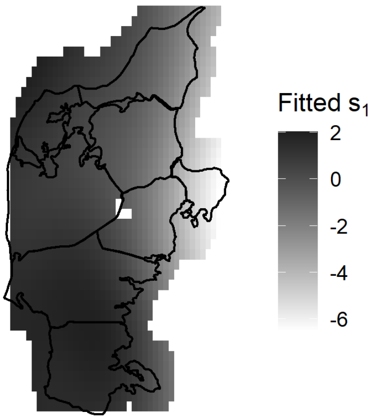

Figure 4: Fitted s_1 values attributable to area.

	est.	*SE*	*t*	*p*
intercept	-1.42	1.38	1.03	0.3
height				
- non-high, + high	-3.65	0.74	-4.96	<.001
- low, + mid	-4.38	0.91	-4.84	<.001
round				
- rd, + non-rd	3.47	0.97	3.58	<.001
back				
- back, + non-back	-3.91	1.21	-3.22	<.01
palatalization				
- non-pal., + pal.	-6.79	2.03	-3.35	<.001
sex				
- female, + male	6.26	1.87	3.35	<.001
VOT	-2.03	0.69	-2.95	<.01

Table 1: Parametric coefficients of GAMM modeling s_1.

Fewer variables contribute to the GAMM modeling s_2: HEIGHT, ROUND, STRESS, and BACKNESS. The geographical variable LON,LAT does not significantly improve the fit of this model. The final model has a medium effect size, with R^2 = .29. Only two variables significantly influence s_2, namely BACK with $\hat{\beta}$ = 1.97, *SE* = 0.7, t = 2.82, $p < .01$, and ROUND with $\hat{\beta}$ = 2.21, *SE* = 0.61, t = 3.6, $p < .001$. In other words, positive s_2, associated with an especially prominent energy peak around 2.5 kHz, is especially found before non-back and non-round vowels. This is also in line with previous findings for MSD [4].

All parametric candidate 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 found to significantly influence s_3 is ROUND, with $\hat{\beta}$ = 2.62, SE = 0.45, t = -5.84, $p < .001$.

4. DISCUSSION AND CONCLUSIONS

Spectral shapes that suggest a coronal noise source are found in /t/ release midpoints in the varieties of eastern Jutland, particularly in Djursland, but not in other traditional varieties of Jutland Danish. This suggests that /t/ affrication was not historically a feature of most varieties of Jutland Danish.

Indeed, it is possible that /t/ affrication was not a feature of any traditional Jutlandic varieties. Southern Djursland coincides with the main water route between Jutland and the island of Zealand, which is where the capital (Copenhagen) and consequently the locus of Standard Danish is located. This is also where the largest city of Jutland (Aarhus) is located, and several of the major historical cities of Jutland are located immediately south of Aarhus [7]. Affrication may be a traditional feature of this area, but the feature may also have spread early from Standard Danish to Aarhus and from there to adjacent rural areas. This would be in line with the *cascade model* of interdialectal influence [39], which predicts that change spreads between population centers in a manner that is predictable from a combination of population size and geographical distance, thus also affecting rural areas close to major cities relatively early.

PC_1 and PC_2 capture the two most readily interpretable sources of variance in the release midpoint spectra: the location and relative magnitude of the energy peak. PC_1 seems to largely capture the difference between aspiration and affrication noise, which is modulated by both geography and phonetic context; lower PCs seem to capture effects of coarticulation, and these do not differ substantially across regional varieties. The influence of phonetic context on $PC_{1–2}$ is largely predictable, and is similar to what has been shown for MSD [4]; for example, the main peak is usually found at higher frequencies before high, non-back, and non-round vowels, in palatalized tokens, in stressed syllables, and in female speakers. The fact that these contextual phonetic effects all behave as expected are an indication that FPCA can indeed be fruitfully used to decompose variance in spectral characteristics. A possible venue to explore in future research would be using FPCA to decompose variance in the trajectories of spectral measures over the time course of stop bursts and releases.

5. ACKNOWLEDGMENTS

Many thanks to Janet Grijzenhout, Bert Botma, Johanna Cronenberg, Jiayin Gao, James Kirby, Michele Gubian, Kathleen Jepson, Vanessa Reichel, Josiane Riverin-Coutlée, and two anonymous reviewers for helpful comments on the presentation of this paper.

6. REFERENCES

[1] Gubian, M., Torreira, F., Boves, L. 2015. Using functional data analysis for investigating multidimensional dynamic phonetic contrasts. *J. Phon.* 49, 16–40.

[2] Jespersen, O. 1897–1899. *Fonetik. En systematisk fremstilling af læren om sproglyd.* Schuboteske.

[3] Fischer-Jørgensen, E. 1954. Acoustic analysis of stop consonants. *Le Maître Phonetique* 32(69), 42–59.

[4] Puggaard-Rode, R. 2022. Analyzing time-varying spectral characteristics of speech with function-on-scalar regression. *J. Phon.* 95, art. 101191.

[5] Kristiansen, T. 2003. Danish. In: Deumart, A., Vandenbussche, W. (eds), *Germanic standardizations. Past to present*. John Benjamins, 69–91.

[6] Pedersen, I. L. 2003. Traditional dialects of Danish and the de-dialectalization 1900–2000. *Int. J. Soc. Lang.* 159, 9–28.

[7] Puggaard, R. 2021. Modeling regional variation in voice onset time of Jutlandic varieties of Danish. In: Van de Velde, H., Hilton, N. H., Knooihuizen, R. (eds), *Language variation – European perspectives VIII*. John Benjamins, 79–110.

[8] Brink, L., Lund, J. 1975. *Dansk rigsmål. Lydudviklingen siden 1840 med særligt henblik på sociolekterne i København*. Gyldendal.

[9] Petersen, N. R. 2009. Aspirater. In: *Den store danske*. Gyldendal.

[10] Petersen, J. H., Juul, H., Pharao, N., Maegaard, M. 2021. *Udtalt. En introduktionsbog til dansk fonetik*. Samfundslitteratur.

[11] Heger, S. 1981. *Sprog og lyd. Elementær dansk fonetik*. Akademisk Forlag.

[12] Andersen, T. A. 1981. Dialektbånd og databehandling. *Ord & Sag* 1, 11–18.

[13] Goldshtein, Y., Puggaard, R. 2019. Overblik over danske dialektoptagelser. *Ord & Sag* 39, 18–28.

[14] Stevens, K. N. 1993. Models for the production and acoustics of stop consonants. *Speech Comm.* 13(3), 367–375.

[15] Stevens, K. N. 1998. *Acoustic phonetics*. MIT Press.

[16] Forrest, K., Weismer, G., Milenkovic, P., Dougall, R. N. 1988. Statistical analysis of word-initial voiceless obstruents. Preliminary data. *J. Acoust. Soc. Am.* 84(1), 115–123.

[17] Stoel-Gammon, C., Williams, K. A., Buder, E. 1994. Cross-language differences in phonological acquisition. Swedish and American /t/. *Phonetica* 51(1), 146–158.

[18] Shadle, C. H., Mair, S. J. 1996. Quantifying spectral characteristics of fricatives. *Int. Conf. Spoken Lang. Proc.* 4, 1521–1524.

[19] Jongman, A., Wayland, R., Wong, S. 2000. Acoustic characteristics of English fricatives. *J. Acoust. Soc. Am.* 108(3), 1252–1263.

[20] Bunnell, H. T., Polikoff, J., McNicholas, J. 2004. Spectral moment vs. bark cepstral analysis of children's word-initial voiceless stops. *Int. Conf. Spoken Lang. Proc.* 8.

[21] Koenig, L. L., Shadle, C. H., Preston, J. L., Mooshammer, C. R. 2013. Toward improved spectral measures of /s/. Results from adolescents. *JSLHR* 56(4), 1175–1189.

[22] Wieling, M., Nerbonne, J., Baayen, R. H. 2011. Quantitative social dialectology. Explaining linguistic variation geographically and socially. *Plos One* 6(9).

[23] Gubian, M., Boves, L., Cangemi, F. 2011. Joint analysis of F_0 and speech rate with functional data analysis. *IEEE ICASSP*, 4972–4975.

[24] *Dialektsamlingen*, 1971–1976. URL: https://dansklyd.statsbiblioteket.dk/samling/dialektsamlingen/.

[25] Goldshtein, Y., Ahlgren, L. M. 2021. Ideologies of language and place. Conflicting expectations to dialectal speech between informants and dialectologists. *J Postcolonial Ling* 5, 178–203.

[26] Boersma, P., Weenink, D. 2019. Praat. Doing phonetics by computer, v6.0.55.

[27] R Core Team, 2021. R. A language and environment for statistical computing, v4.1.2.

[28] Reidy, P. F. 2013. An introduction to random processes for the spectral analysis of speech data. *OSUWPL* 60, 67–116.

[29] Gajardo, A., Satarupa, B., Carroll, C., Chen, Y., Dai, X., Fan, J., Hadjipantelis, P. Z., Han, K., Ji, H., Zhu, C., Müller, H.-G., Wang, J.-L. 2021. `fdapace`, R package v0.5.8.

[30] Bennike, V., Kristensen, M. 1912. *Kort over de danske folkemål med forklaringer*. Gyldendal.

[31] Schad, D. J., Vasishth, S., Hohenstein, S., Kliegl, R. 2020. How to capitalize on a priori contexts in linear (mixed) models. A tutorial. *J. Mem. Lang.* 110, art. 104038.

[32] Wood, S. N. 2003. Thin plate regression splines. *J. Roy. Stat. Soc. B* 65(1), 95–114.

[33] Wood, S. N. 2021. `mgcv`. R package v1.8-38.

[34] van Rij, J., Wieling, M., Baayen, R. H., van Rijn, H. 2020. `itsadug`. R package v2.4.

[35] Fasiolo, M., Nedellec, R., Goude, Y., Wood, S. N. 2021. `mgcViz`. R package v0.1.9.

[36] van Rij, J. 2016. Testing for significance. `itsadug` R package vignette.

[37] DataverseNL, DOI: 10.34894/FJKFD1.

[38] Puggaard-Rode, R. 2023. *Stop! Hey, what's that sound? The realization and representation of Danish stops*. PhD dissertation, Leiden University.

[39] Labov, W. 2003. Pursuing the cascade model. In: Britain, D., Cheshire, J. (eds), *Social dialectology*. John Benjamins, 9–22.