LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

ECR KOLLOQUIUM
IPS, LMU
8 DECEMBER 2022

Analyzing the spectrum with functional data analysis

Rasmus Puggaard-Rode

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

We have the tools to statistically model spectral shape directly, and doing so forgoes some of the problems with other methods of describing the spectrum

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Roadmap

- The case: Danish /t/ releases
- Analyzing aperiodic spectra
  - FDA as an alternative
- Two corpus studies
  - Affrication in Modern Standard Danish
    - Function-on-scalar regression
  - Affrication in traditional Jutland Danish varieties
    - Functional principal component analysis
- General discussion

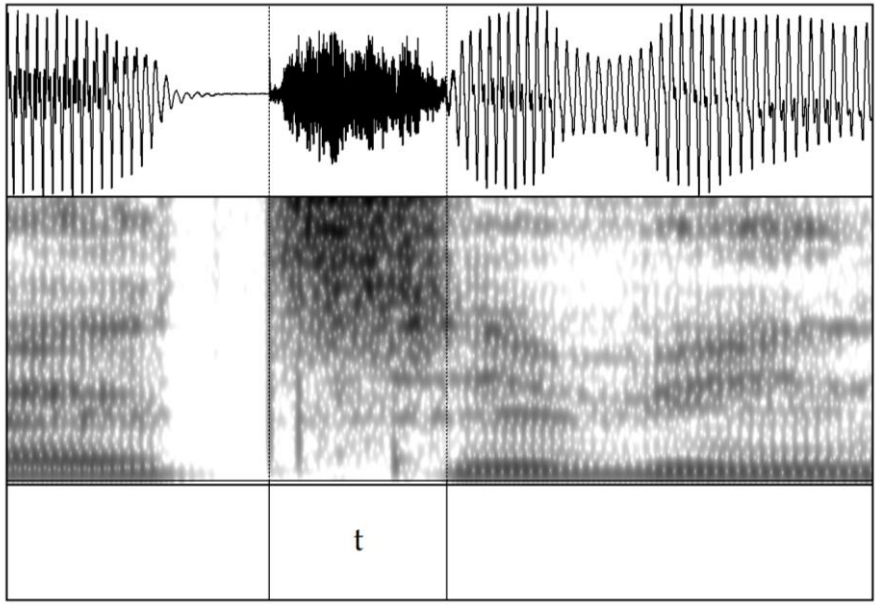
LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

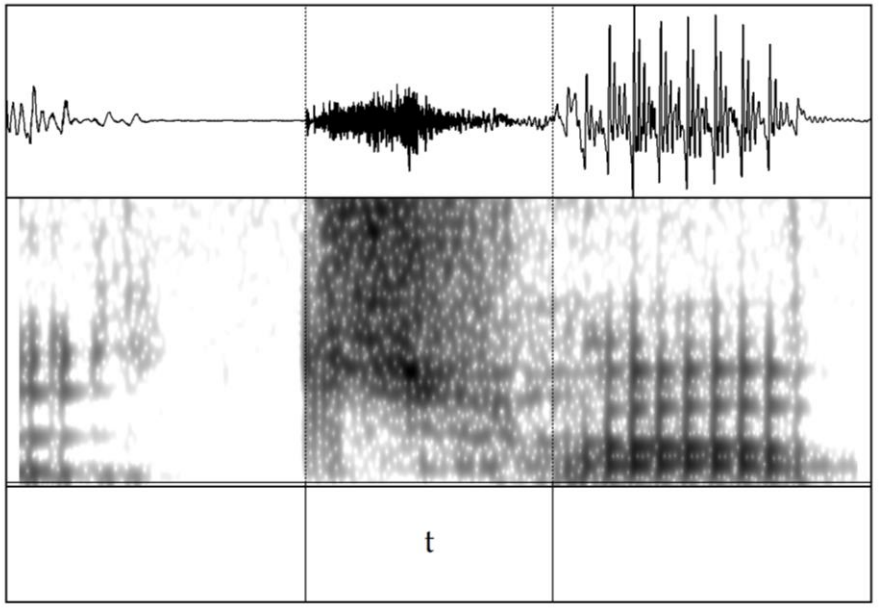
Danish /t/ releases

- Modern Standard Danish has an aspiration-based laryngeal contrast in stops
  - Unaspirated /b d g/
  - Aspirated /p t k/
- /t/ is saliently affricated
  - Aspirated stop? Affricated stop? Affricate?
- Transcription strategies
  - [d̥ˢ] (Basbøll 1969, 2005; Grønnum 1998)
  - [d̥ˢʰ] (Petersen 1983)
  - [d̥ˢh] (Brink & Lund 1975)
  - [tsʰ] (Basbøll & Wagner 1985)
  - [tˢ] (Grønnum 1998, 2005)
  - [ts] (Schachtenhaufen 2022)

Danish /t/ releases

- Modern Standard Danish has an aspiration-based laryngeal contrast in stops
  - Unaspirated /b d g/
  - Aspirated /p t k/
- /t/ is saliently affricated. What is it?
  - Aspirated stop? Affricated stop? Affricate?
- Transcription strategies
  - [d̥ˢ] (Basbøll 1969, 2005; Grønnum 1998)
  - [d̥ˢʰ] (Petersen 1983)
  - [d̥ˢh] (Brink & Lund 1975)
  - [tsʰ] (Basbøll & Wagner 1985)
  - **[tˢ] (Grønnum 1998, 2005)**
  - **[ts] (Schachtenhaufen 2022)**





Grønnum (2009)

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Danish /t/ releases

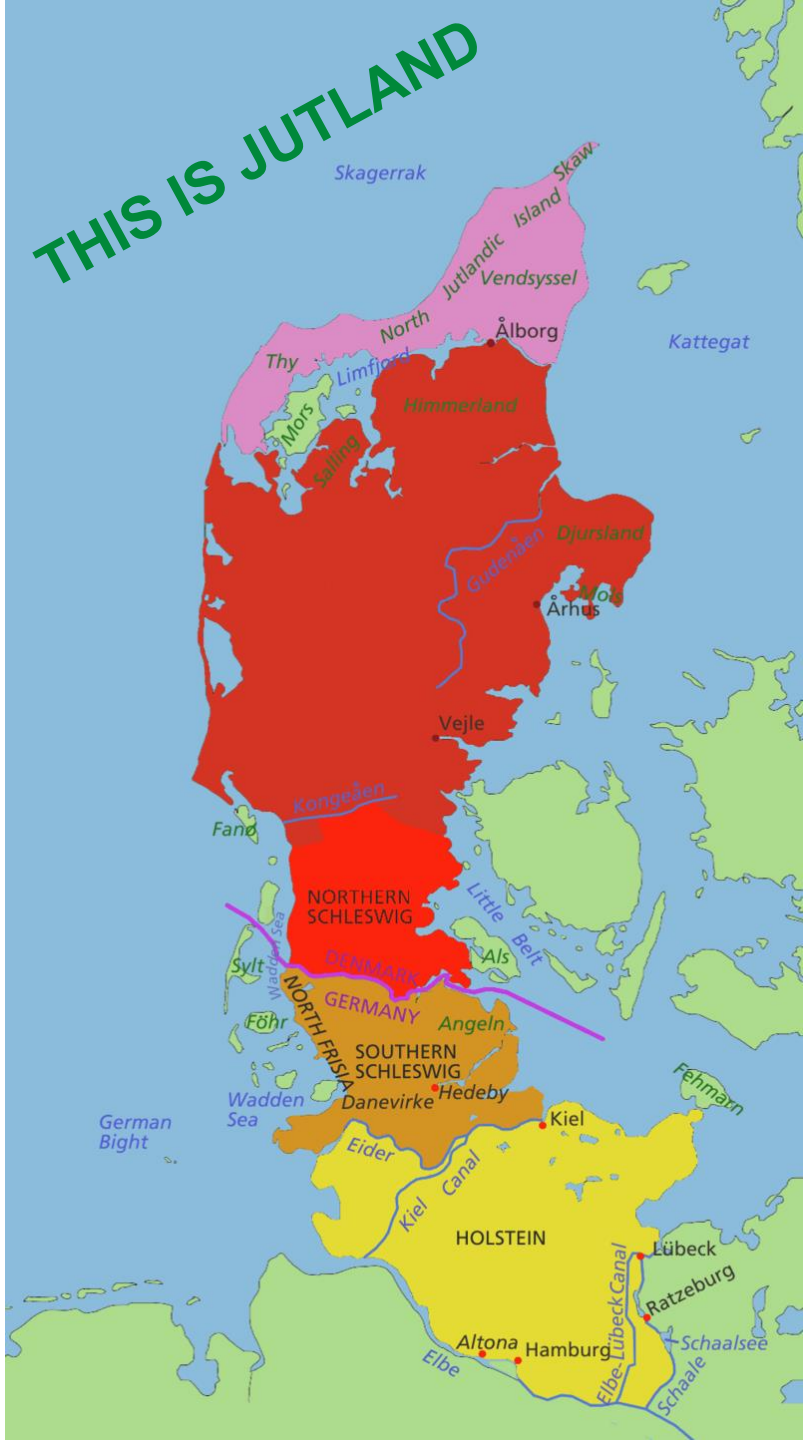
- Fischer-Jørgensen (1954, 1972)
  - /t/ releases contain both affrication and aspiration proper
  - The affrication spectrum is similar to the /s/ spectrum
  - Superimposing /t/ aspiration on /p k/ will lead listeners to perceive the sound as /t/

- Brink & Lund (1975)
  - /t/ affrication was widespread in mid-19C in Copenhagen
  - In mid-20C it was exceptionless in Copenhagen

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Variation

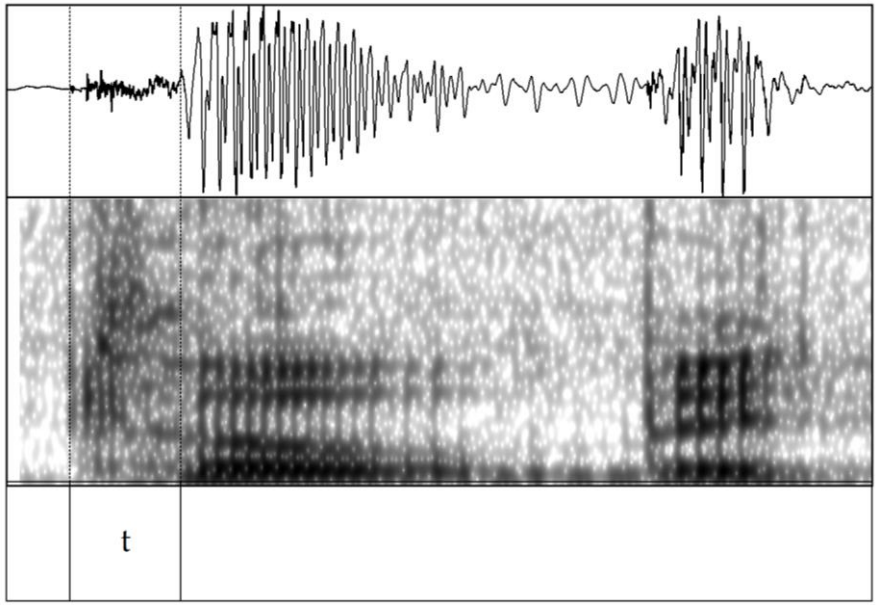
- Not all varieties have saliently affricated /t/
- **But** it's a complete mystery which varieties lack it
  - Missing from
    - 'All dialects' (Brink & Lund 1975)
    - High and formal styles (Grønnum 2005)
    - Northern Jutlandic (Petersen 2009)
    - Western Jutlandic (Petersen et al. 2021)
    - All Jutland varieties except Eastern Jutlandic (Heger 1981)
- The sociolinguistic situation of Danish dialects is complex, but these traditional varieties are presumably (nearly) extinct (Pedersen 2003; but see also Maegaard & Monka 2019)

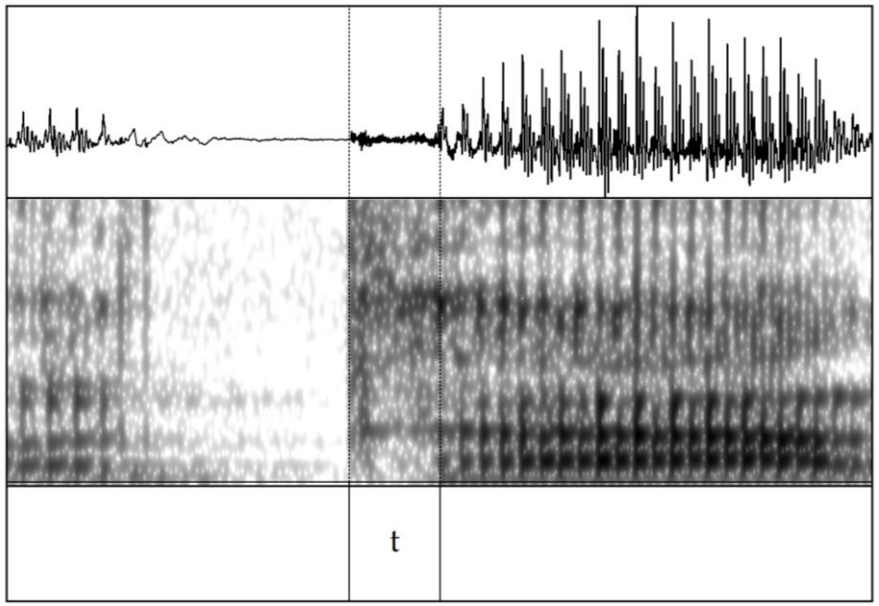


LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Variation

- Not all varieties have saliently affricated /t/
- **But** it's a complete mystery which varieties lack it
  - Missing from
    - 'All dialects' (Brink & Lund 1975)
    - High and formal styles (Grønnum 2005)
    - Northern Jutlandic (Petersen 2009)
    - Western Jutlandic (Petersen et al. 2021)
    - All Jutland varieties except Eastern Jutlandic (Heger 1981)
- The sociolinguistic situation of Danish dialects is complex, but these traditional varieties are presumably (nearly) extinct (Pedersen 2003; but see also Maegaard & Monka 2019)





DS (1971–1976)

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Who cares?

- These are unresolved issues in a very well-described language
  - But too often they're treated as solved

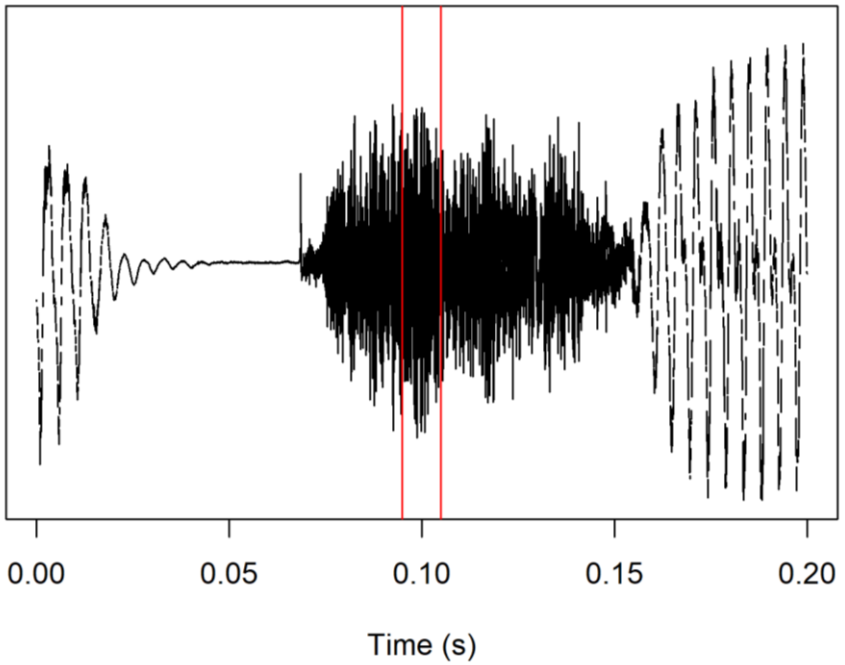
- How much does fine phonetic detail vary?
- What governs variation in fine phonetic detail?

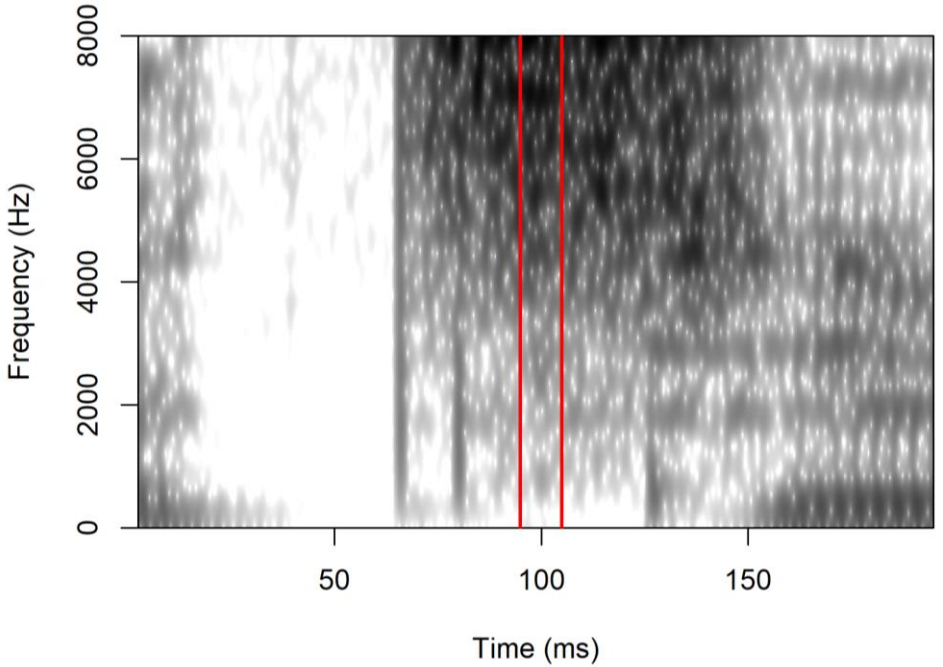
LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

RQ1: How do spectral characteristics vary over time in Modern Standard Danish /t/ releases, and how are they affected by phonetic context?

RQ2: How do spectral characteristics of /t/ releases vary geographically in traditional dialects of Jutland?

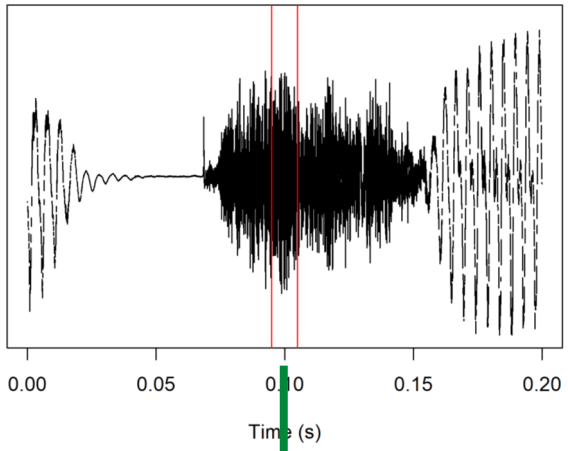
Aperiodic spectra

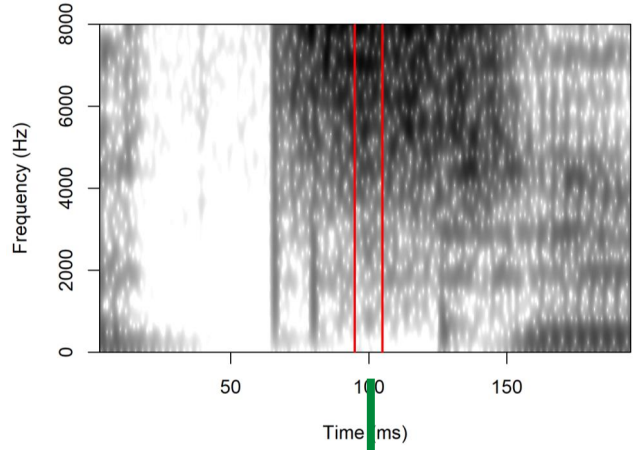


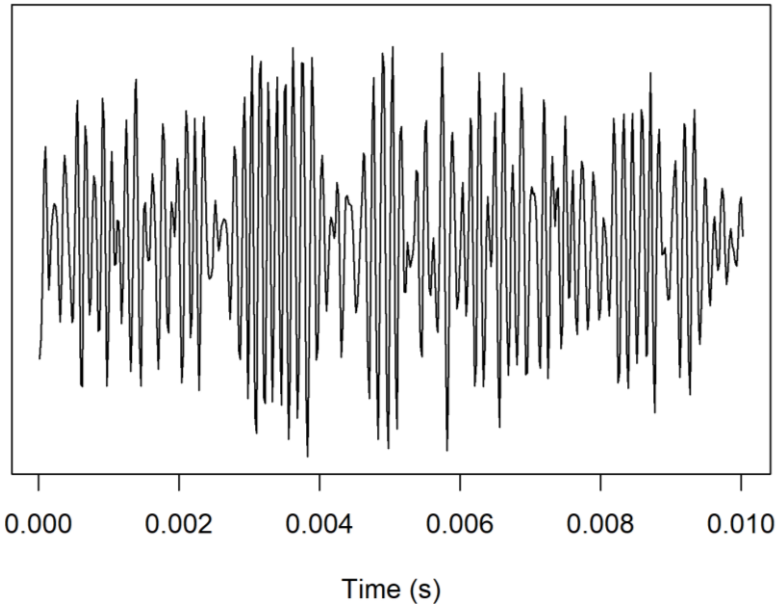


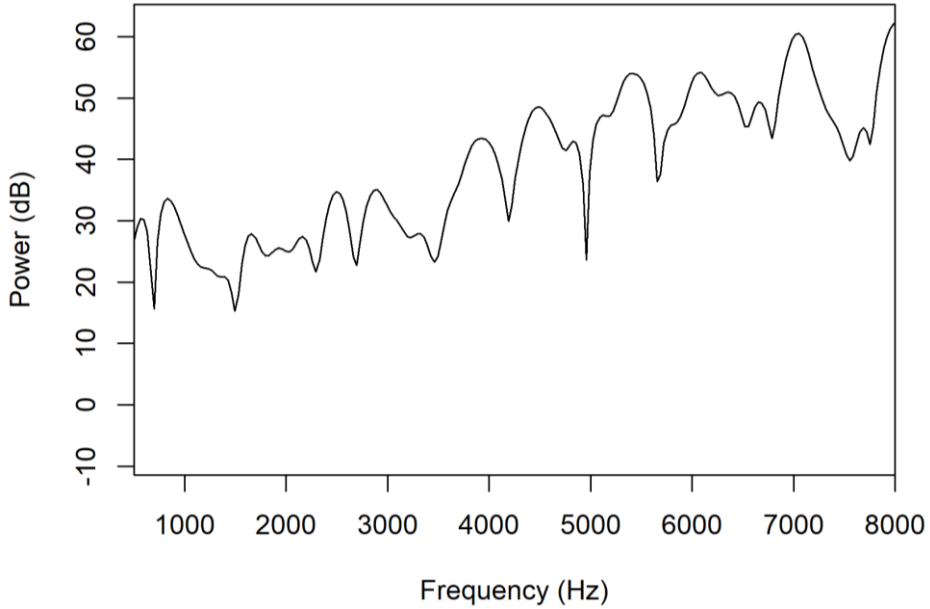
LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Aperiodic spectra





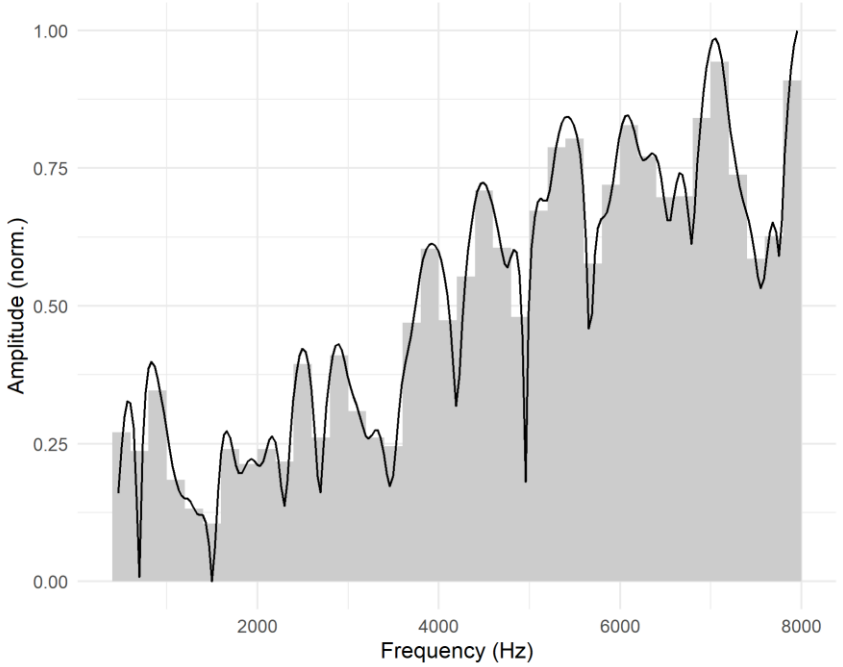




LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Spectral moments

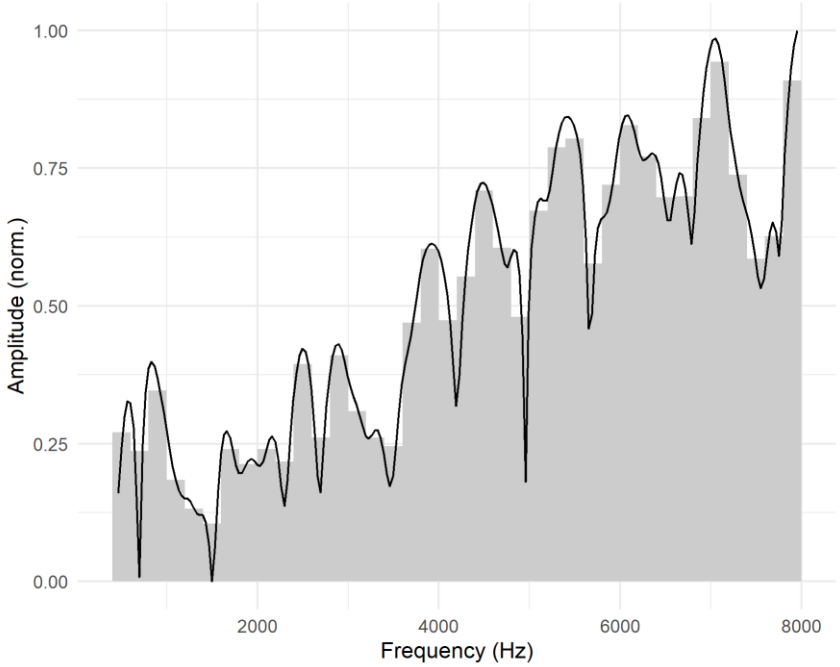
- Treating the spectrum as a probability distribution, and deriving moments from it (Forrest et al. 1988)

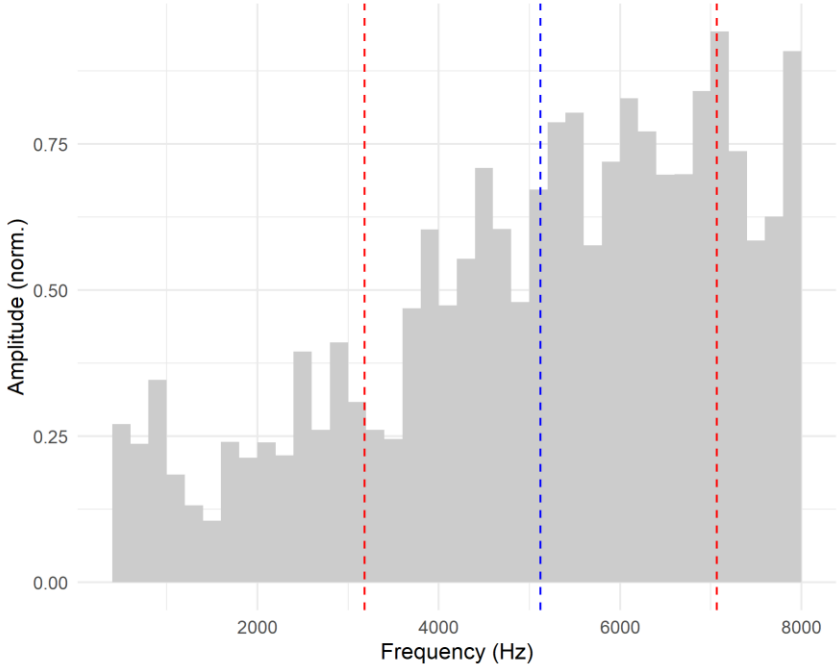


LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Spectral moments

- Treating the spectrum as a probability distribution, and deriving moments from it (Forrest et al. 1988)
  - Centroid frequency, or **center of gravity**
    - 5,124 Hz
  - Standard deviation
    - 1,941 Hz
  - Skewness
    - -1.2
  - Kurtosis
    - 1.36

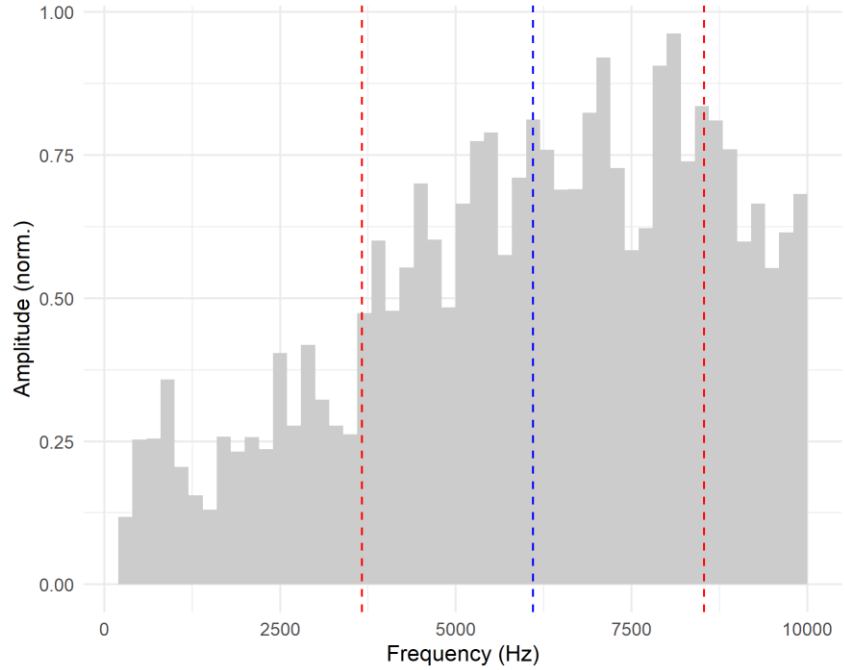




LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

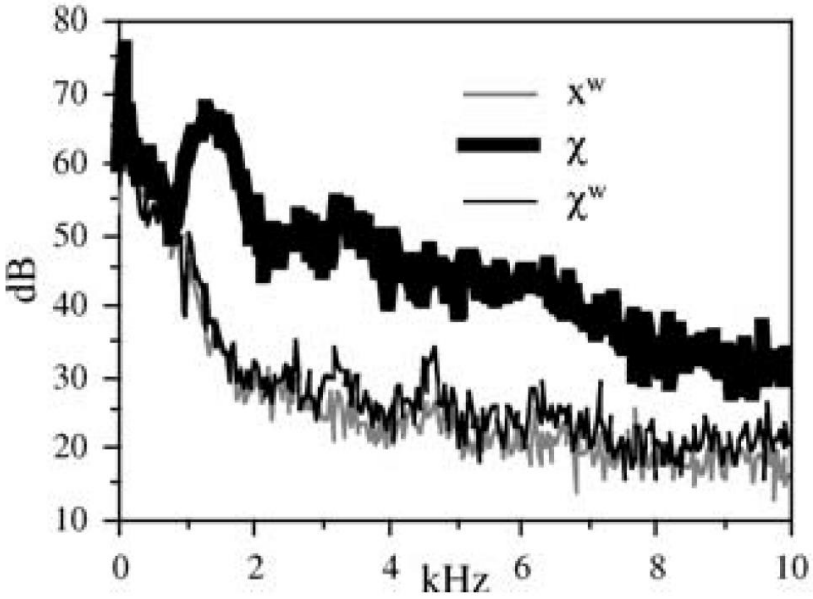
Spectral moments

- Subsequent studies have not given consistent results about which moments are useful
  (Stoel-Gammon et al. 1994; Shadle & Mair 1996)
- Spectral moments are highly dependent on how the spectrum is filtered
  - Our spectrum was filtered at 500–8,000 Hz
  - If we'd chosen 300–10,000 Hz instead
    - COG would be ~1,000 Hz higher
    - SD would be ~500 Hz higher



Spectral moments

- Subsequent studies have not given consistent results about which moments are useful
  (Stoel-Gammon et al. 1994; Shadle & Mair 1996)
- Spectral moments are highly dependent on how the spectrum is filtered
  - Our spectrum was filtered at 500–8,000 Hz
  - If we'd chosen 300–10,000 Hz instead
    - COG would be ~1,000 Hz higher
    - SD would be ~500 Hz higher
- Higher moments are often ignored, and **COG alone says little about spectral shape**
  - [χ χ^w] have roughly identical COG in Montana Salish

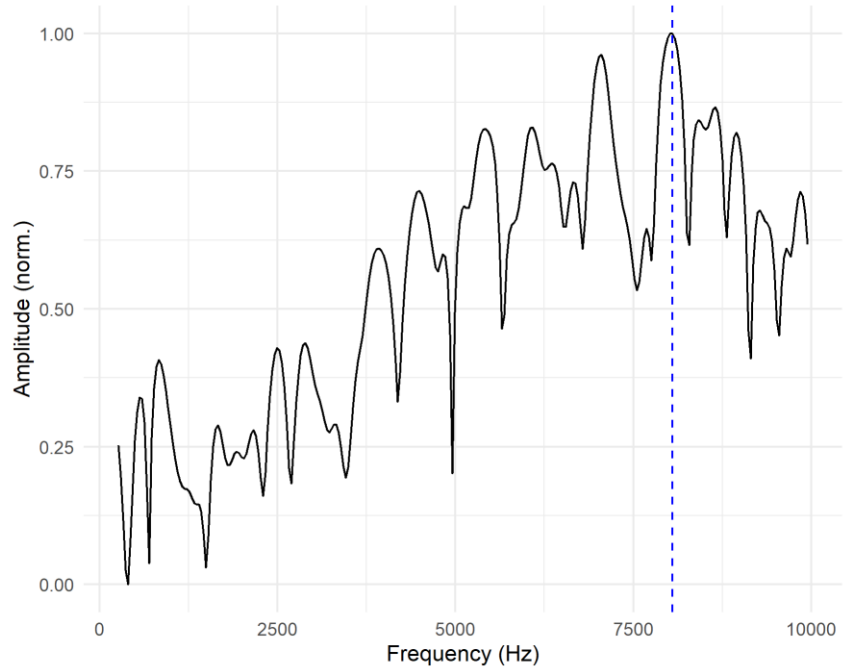


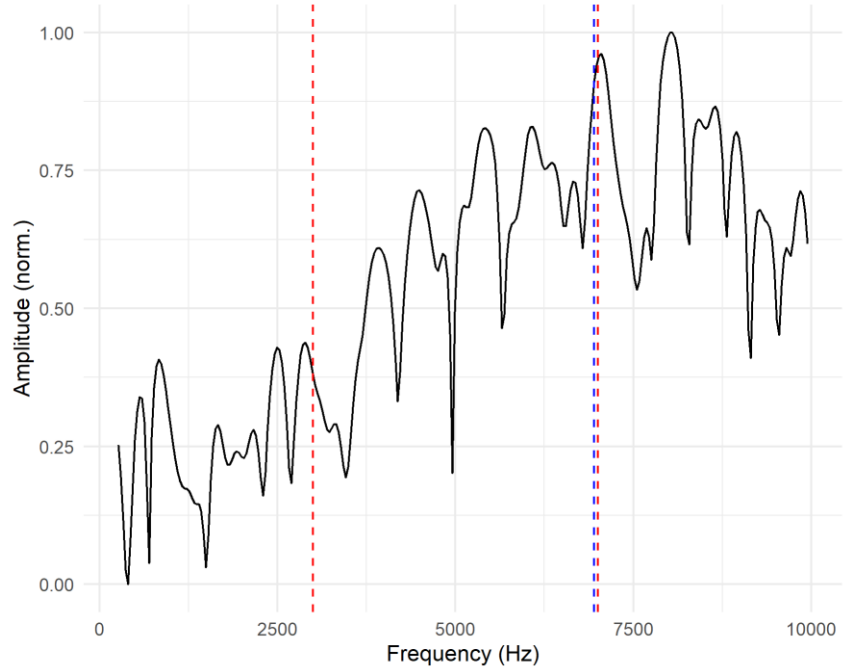
Gordon et al. (2002)

Peak frequency

- Peak frequency (Jongman et al. 2000)
  - 8,050 Hz

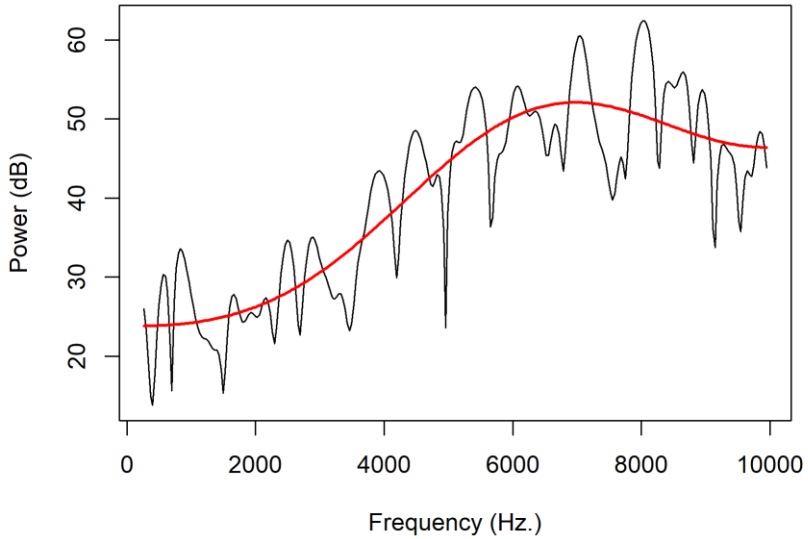
- Mid-frequency spectral peak
  (Koenig et al. 2013; Chodroff & Wilson 2022)
  - 6,952 Hz





DCT coefficients

- Smoothing aperiodic spectra with four DCTs usually retains gross spectral shape (Bunnell et al. 2004; Harrington 2010; Spinu & Lilley 2016; Jannedy and Weirich 2017; Kokkelmans 2021)
  - DCT coefficients then tell us something about
    - Mean (k0)
    - Slope (k1)
    - Curvature (k2)
    - Amplitude at higher frequencies (k3)



LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

We need to operationalize RQ1 and RQ2, but we are now lost in the garden of forking paths

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Functional data analysis

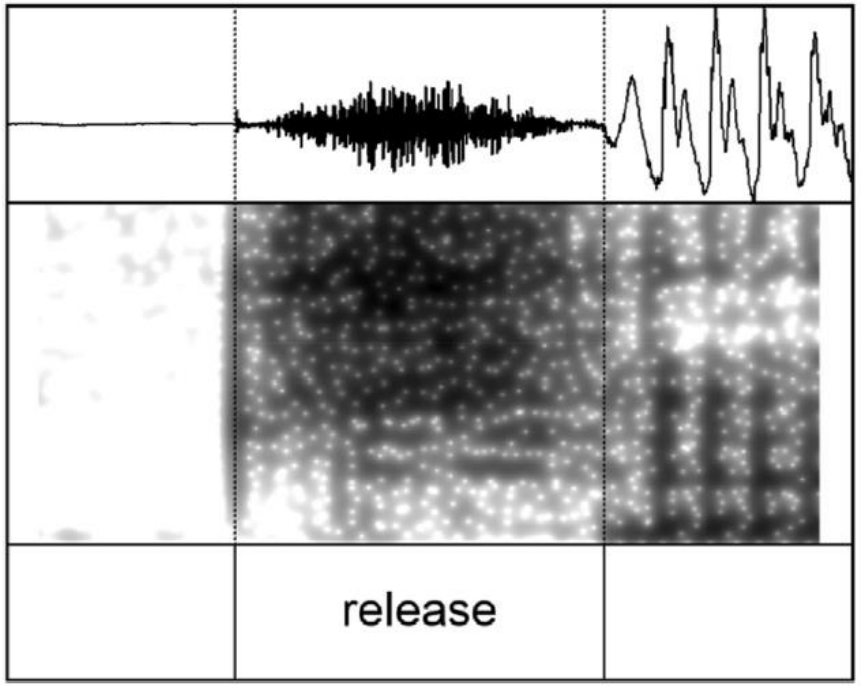
- Extensions of statistical methods for functional data (Ramsay & Silverman 2005; Kokoszka & Reimherr 2017)
  - Functional PCA (Gubian et al. 2015)
  - Functional regression models (Greven & Scheipl 2017; Pouplier et al. 2017)

- Often applied to time series data *Y(t)*
  - But should work just as well for spectral data *A(f)*

RQ1: Materials

- Monologs from DanPASS corpus (Grønnum 2009)
  - ~3 hours of speech, 18 speakers (5F/13M)
  - 850 /t/ tokens
    - Mean VOT: 79 ms (stressed), 68 ms (unstressed)

- Each stop split into 20 normalized time steps
  - Multitaper spectra generated in R for each of these
  - Frequency range 500–10,000 Hz



Puggaard-Rode (2022)

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Function-on-scalar regression

- Modeled with **function-on-scalar** models fitted with `refund::pffr` (Bauer et al. 2018; Goldsmith et al. 2021)
  - Wrapper for `mgcv` functions, so won't be totally unfamiliar for GAMM users (Wood 2017)
  - Allows problems to be formalized as functional regression

- GAMM: amplitude ~ frequency × time ☹
- FOSR: spectrum ~ time ☺

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

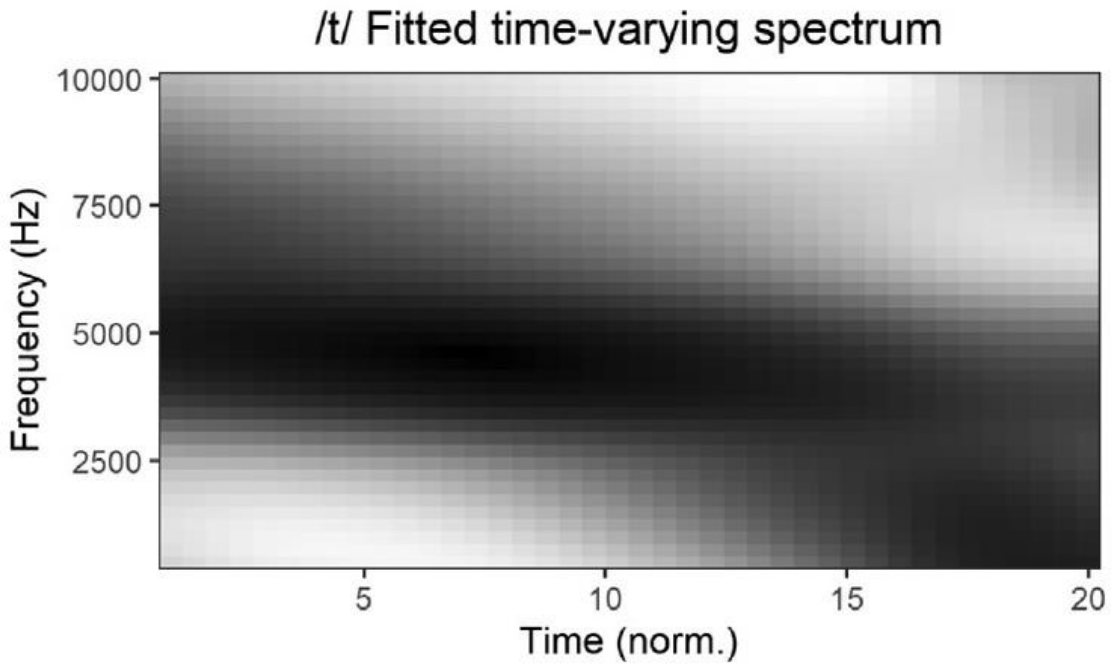
Function-on-scalar regression

- Model structure
  - The spectrum has a smooth functional intercept
  - Spectral shape can vary smoothly over time
  - For each of the following fixed effects, spectra can vary smoothly over time
    - Speaker sex
    - Following vowel height, roundness, backness
    - Stress
  - By-speaker functional random slopes for each fixed effect
  - Corrected for autocorrelated errors

```
pffr(Y ~ s(timestep, k=16) +
    s(timestep, k=16, by=stressn) +
    s(timestep, k=16, by=sexn) +
    s(timestep, k=16, by=high_vn) +
    s(timestep, k=16, by=back_vn) +
    s(timestep, k=16, by=round_vn) +
    s(speaker, timestep, bs="re") +
   s(speaker, timestep, by=stressn, bs="re") +
   s(speaker, timestep, by=high_vn, bs="re") +
   s(speaker, timestep, by=back_vn, bs="re") +
   s(speaker, timestep, by=round_vn, bs="re"),
data=t_df, ydata=t_y,
bs.yindex = list(bs="ps", k=6, m=c(2,1)),
bs.int = list(bs="ps", k=round(nrow(t_y) / nrow(t_df), 0), m=c(2,1)),
rho = ar1+0.1,
```

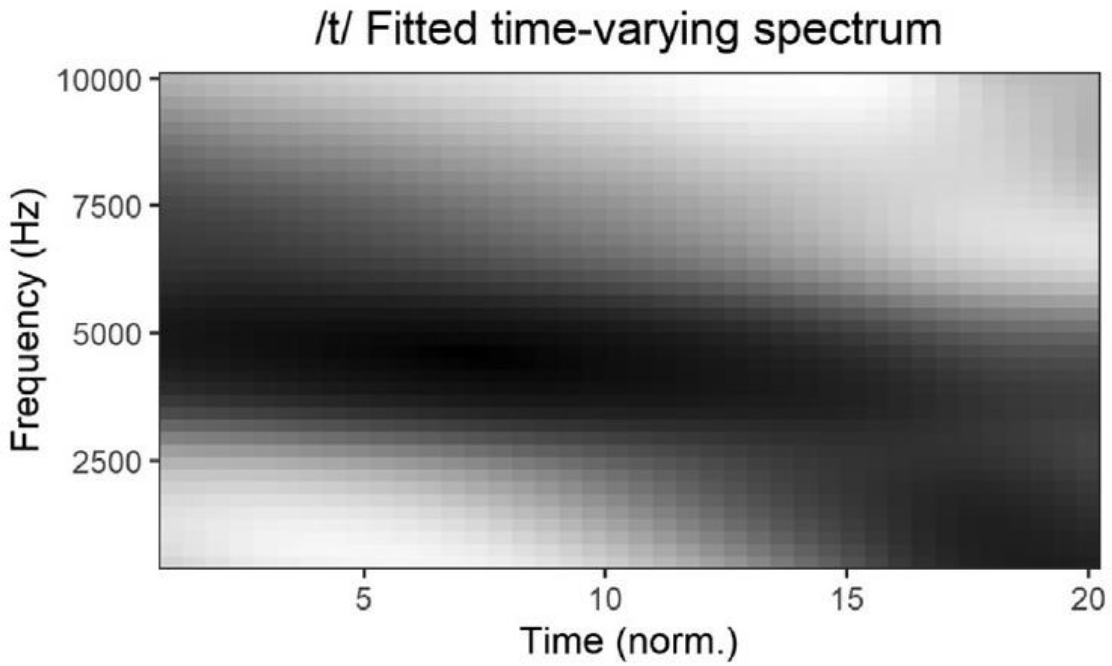
LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Results

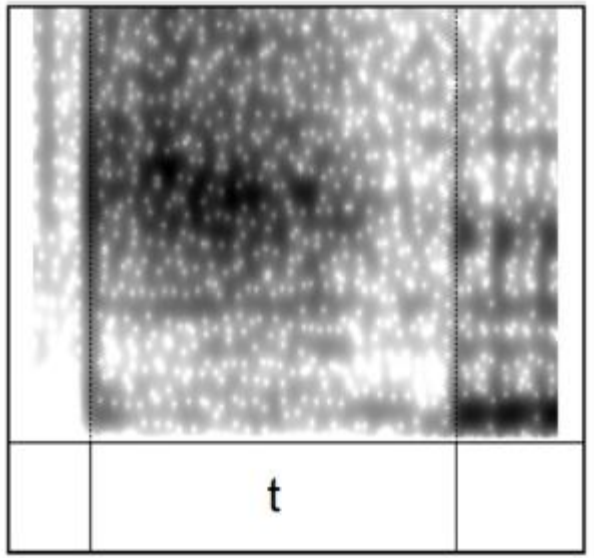


LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Results

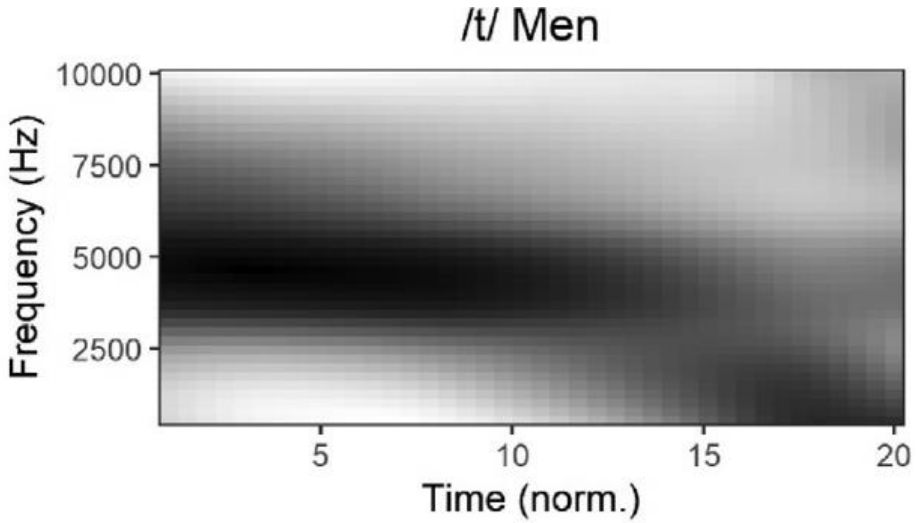


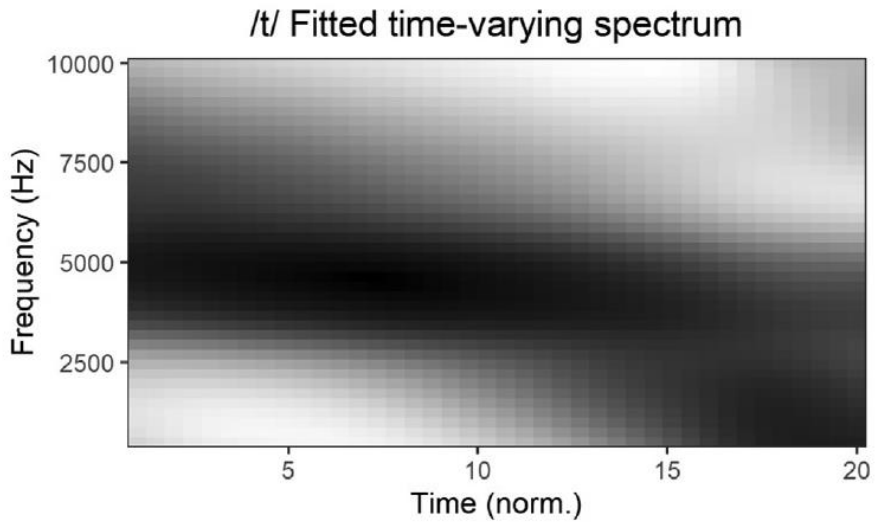
cp. a real spectrogram



LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

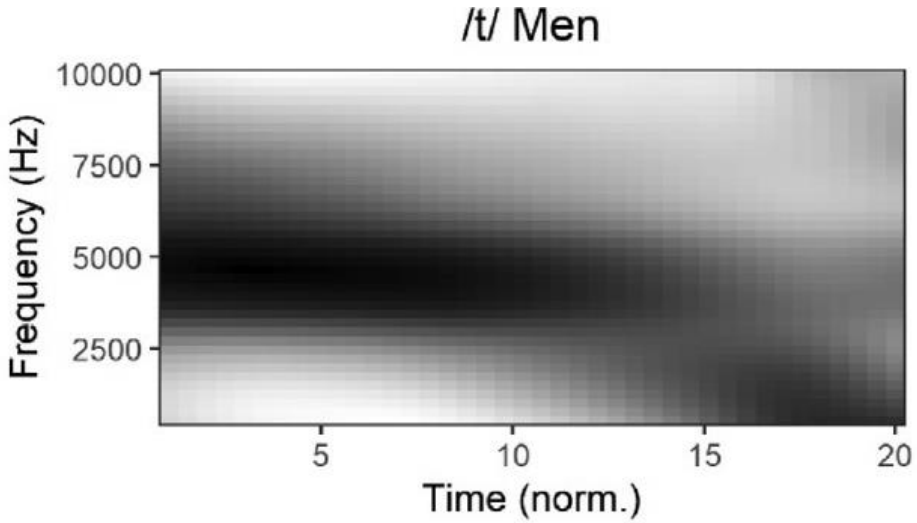
Results

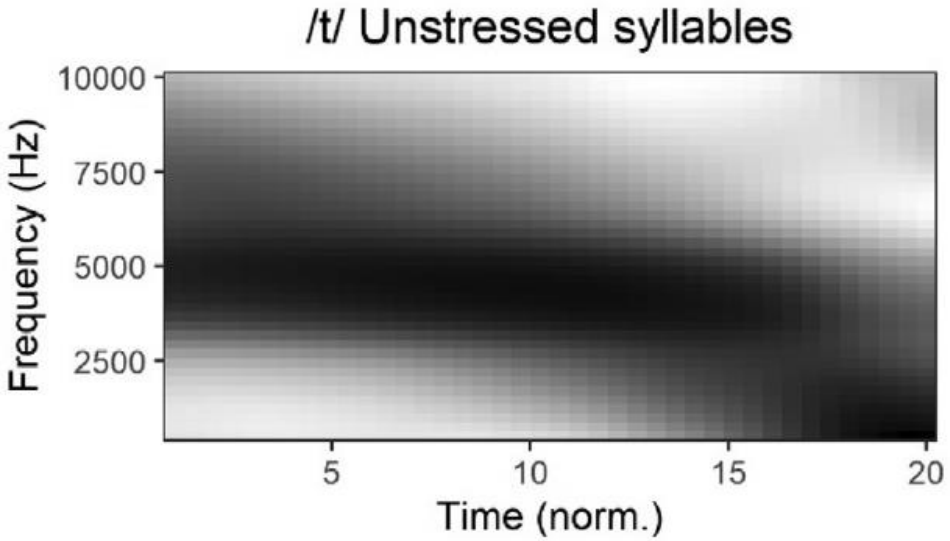


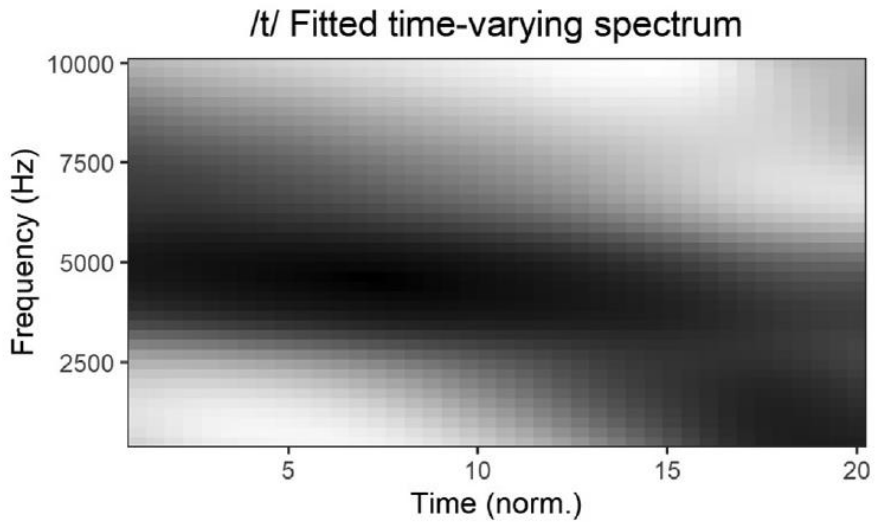


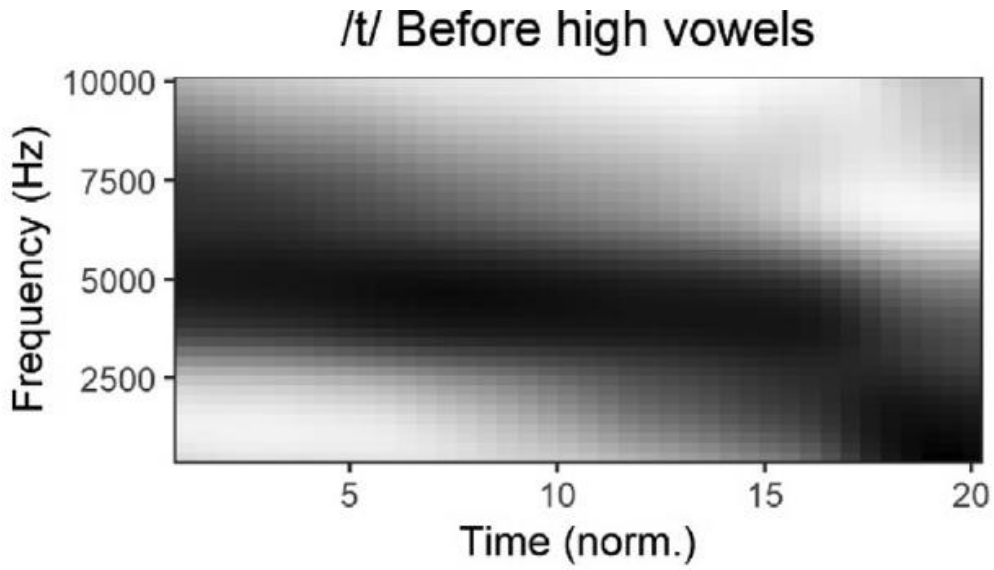
LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Results



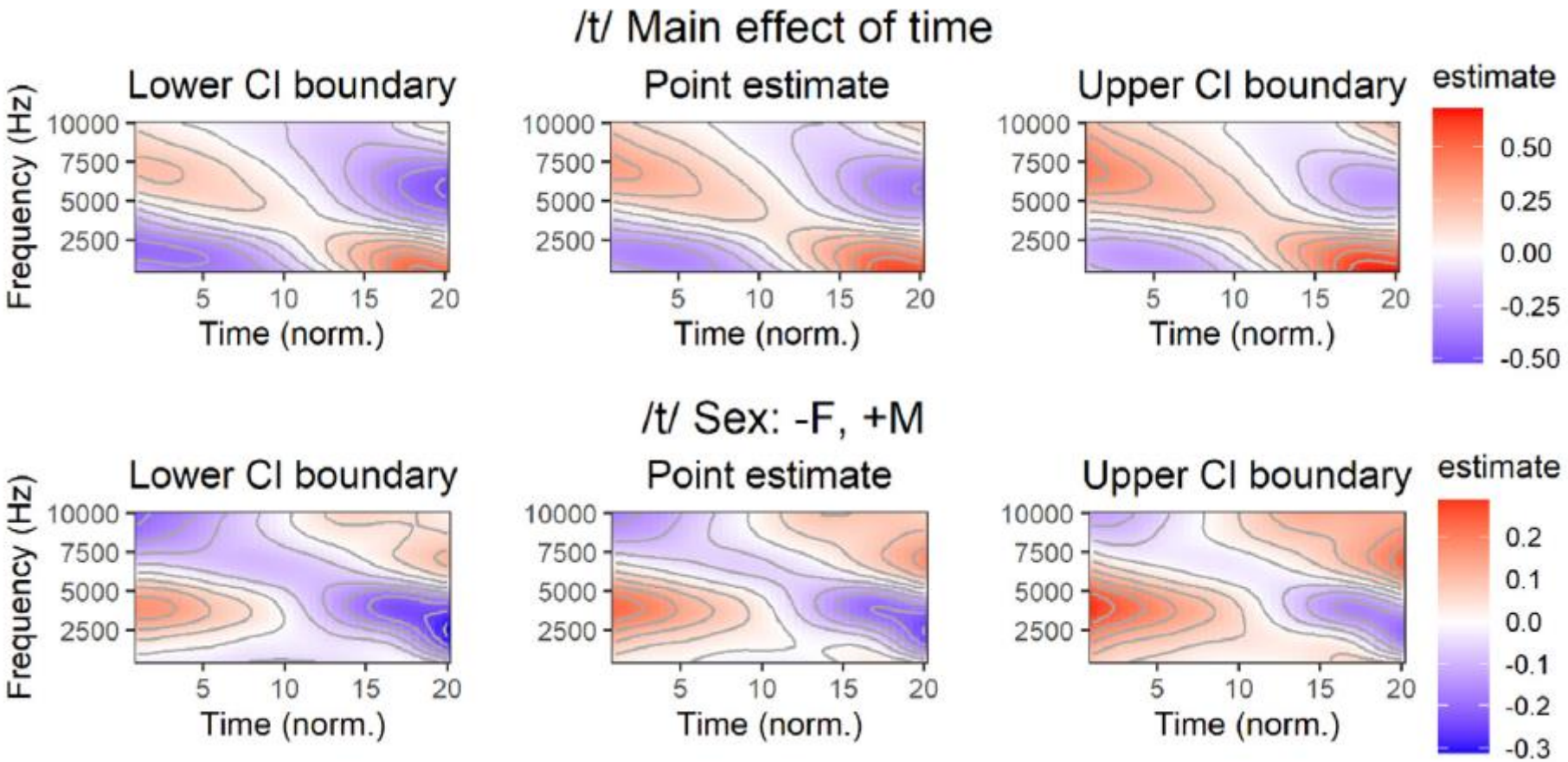






LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Results



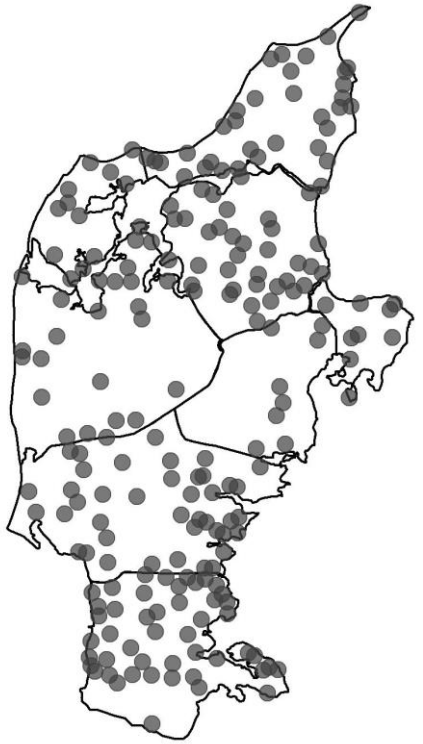
LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Interim discussion

- MSD /t/ is invariably affricated
- A coronal noise source generally dominates most of the release
  - But **not** the final portion of the release
- The transition from dominant alveolar noise source to dominant glottal noise source is smooth
- The proportion of alveolar <> glottal is modulated by phonetic context

RQ2: Materials

- Legacy corpus of sociolinguistic interviews with elderly speakers recorded between 1971–1976
  (DS 1971–1976; Goldshtein & Puggaard 2019)
- Recordings from 213 different parishes
- 5,103 /t/ tokens
- 5 ms extracted from each release midpoint
  - Multitaper spectrum generated in R

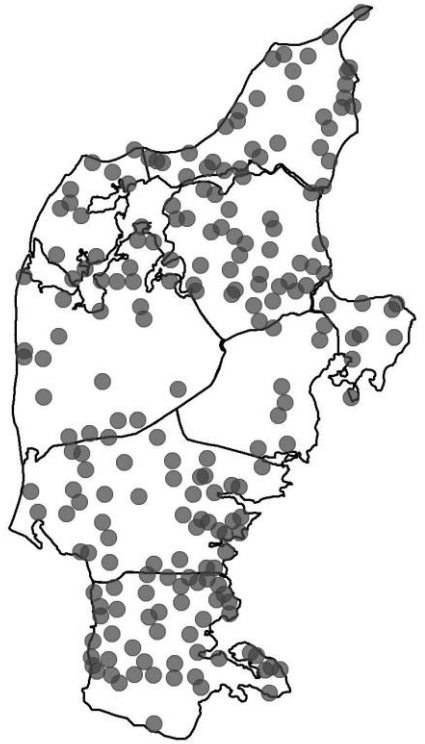


LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

RQ2: Materials

- Legacy corpus of sociolinguistic interviews with elderly speakers recorded between 1971–1976
  (DS 1971–1976; Goldshtein & Puggaard 2019)
- Recordings from 213 different parishes
- 5,103 /t/ tokens
- 5 ms extracted from each release midpoint
  - Multitaper spectrum generated in R

- *Hold up, can't we just model the full /t/ release?*
  - Computationally, sure
  - But how to visualize the results?
    Regional variation in how spectral shape varies over time?



LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

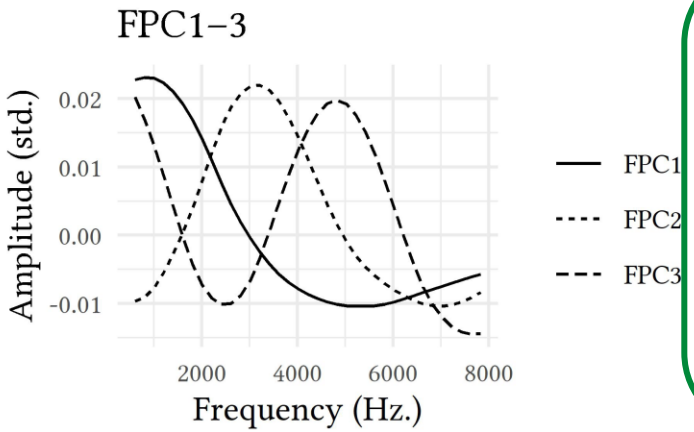
Functional PCA

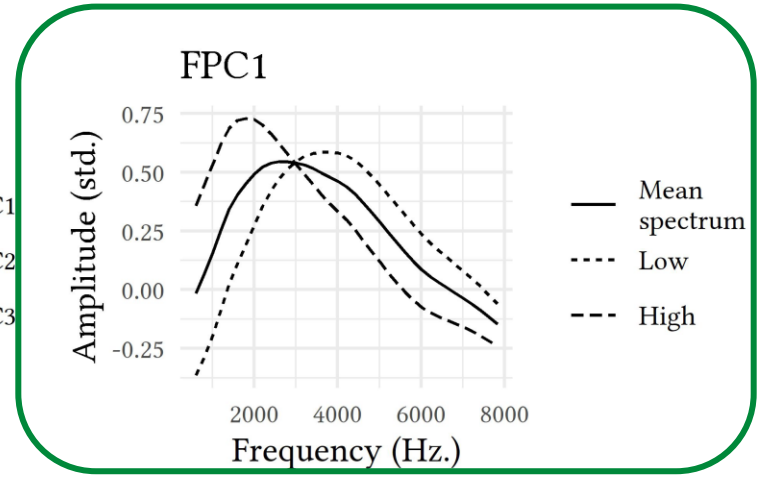
- Spectral variance analyzed with functional PCA
  - Identifies main sources of variance
  - Bottom-up approach is particularly suitable here
    - This is exploratory research, and the existing literature doesn't allow for strong predictions
  - Principal components are independent by definition

- FPCA done in R with `fdapace` (Gajardo et al. 2021)
  - PC scores used as dependent variables in spatial GAMMs
    - Predicting PC scores from geography (longitude × latitude), phonetic context, and speaker sex

Results

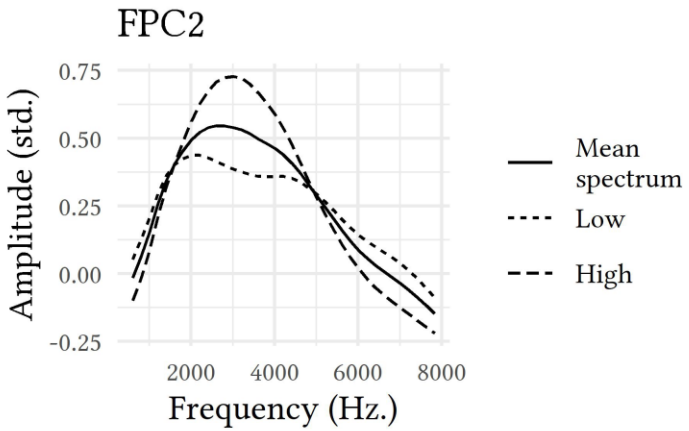
5 PCs account for 95% of variance

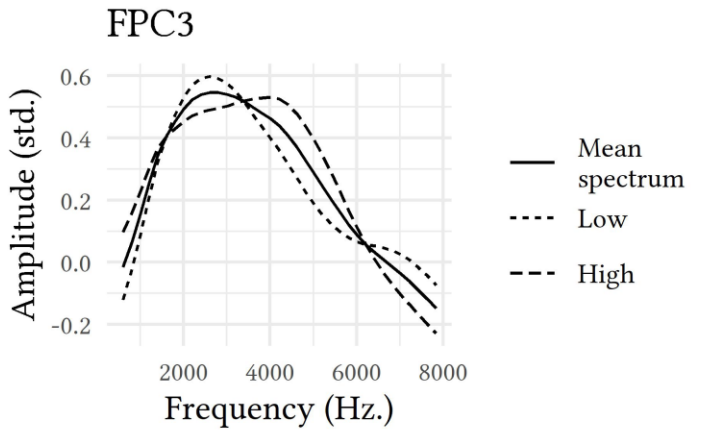




58.4%

18.2%





9.3%

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

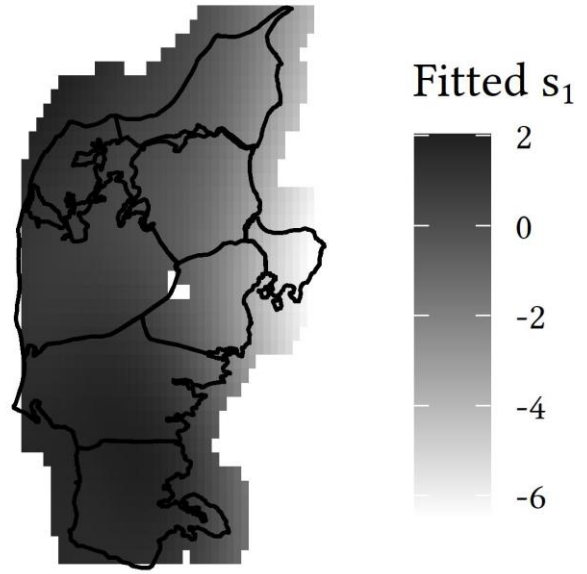
Results

- Negative s_1 (=spectrum reflecting alveolar noise source)
  - Unlikely before low vowels
  - Unlikely before rounded vowels
  - Unlikely before back vowels
  - More likely from female speakers

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Results

- Negative s_1 (=spectrum reflecting alveolar noise source)
  - Unlikely before low vowels
  - Unlikely before rounded vowels
  - Unlikely before back vowels
  - More likely from female speakers
  - ➢More likely in eastern Jutland (particularly Djursland)
    - Could be a traditional feature of this area
    - Could be because this coincides with the biggest city of the peninsula (Aarhus)



LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

General discussion

- Time-varying spectral characteristics of MSD /t/ releases suggest the presence of both salient affrication and aspiration
  - Smooth transition from one to the other
  - Relative proportion modulated by speaker and phonetic context
- In most traditional Jutland Danish varieties, the /t/ release midpoint does not reflect a coronal noise source
  - Affrication more likely in certain contexts
  - Affrication more likely in eastern Jutland

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

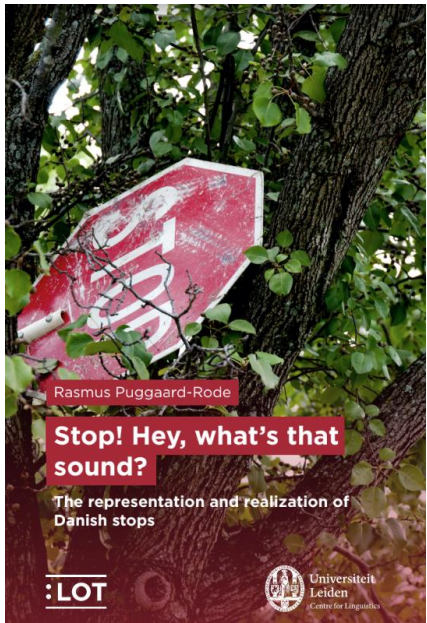
General discussion

- FDA allow us to model spectral shape directly
  - Analyzing main sources of variance with FPCA
  - Using spectral shape as a dependent variable with FOSR
- Largely forgoes the problems of other approaches to analyzing the spectrum
  - Bottom-up approach
  - Fairly easy to interpret
  - Familiar scale

References

- Basbøll, H. 1969. The phoneme system of Advanced Standard Copenhagen. *ARIPUC* 3, 33–54.
- Basbøll, H. 2005. *The phonology of Danish.* OUP.
- Basbøll, H & J Wagner. 1985. *Kontrastive Phonologie des Deutschen und Dänischen. Segmentale Wortphhonologie und -phonetik.* Max Niemeyer. doi:10.1515/9783111358345.
- Bauer, A, F Scheipl, H Küchenhoff & A-A Gabriel. 2018. An introduction to semiparametric function-on-scalar regression. *Statistical Modelling* 18(3/4), 346–364. doi:10.1177/1471082X17748034.
- Blacklock, OS. 2004. *Characteristics of variation in production of normal and disordered fricatives, using reduced-variance spectral methods*. PhD dissertation, Uni Southampton.
- Brink, L & J Lund. 1975. *Dansk rigsmål.* Gyldendal.
- Bunnell, HT, J Polikoff & J McNicholas. 2004. Spectral moment vs. bark cepstral analysis of children's word-initial voiceless stops. *Proc Interspeech* 8. doi:10.21437/Interspeech.2004-81.
- Chodroff, E & C Wilson. 2022. Uniformity in phonetic realization. Evidence from sibilant place of articulation in American English. *Language* 98(2), 250–289. doi:10.1353/lan.2022.0007.
- DS = *Dialektsamlingen*. 1971–1976. URL:dansklyd.statsbiblioteket.dk
- Fischer-Jørgensen, E. 1954. Acoustic analysis of stop consonants. *Le Maître Phonetique* 32(69), 42–59.
- Fischer-Jørgensen, E. 1972. Tape cutting experiments with Danish stop consonants in initial position. *ARIPUC* 6, 104–168.
- Forrest, K, G Weismer, P Milenkovic & RN Dougall. 1988. Statistical analysis of word-initial voiceless obstruents. Preliminary data. *JASA* 84(1), 115–123. doi:10.1121/1.396977.
- Gajardo, A, S Bhattacharjee, C Carroll, Y Chen, X Dai, J Fan, PZ Hadjipantelis, K Han, H Ji, C Zhu, H-G Müller & J-L Wang. `fdapace`. R pkg v0.5.8.
- Goldshtein, Y & R Puggaard. 2019. Overblik over danske dialektoptagelser. *Ord & Sag* 39, 18–28.
- Goldsmith, J, F Scheipl, L Huang, J Wrobel, C Di, J Gellar, J Harezlak, MW McLean, B Swihart, L Xiao, C Craicineau & PT Reiss. 2021. `refund`. R pkg v0.1-24.
- Gordon, M, P Barthmaier & K Sands. 2002. A cross-linguistic acoustic study of voiceless fricatives. *JIPA* 32(2), 141–174. doi:10.1017/S0025100302001020.
- Greven, S & F Scheipl. 2017. A general framework for functional regression modelling. *Statistical Modelling* 17(1/2), 1–35. doi:10.1177/1471082X16681317.
- Grønnum, N. 1998. Illustrations of the IPA. Danish. *JIPA* 28(1/2), 99–105. doi:10.1017/S0025100300006290.
- Grønnum, N. 2005. *Fonetik og fonologi. Almen og dansk*. 3rd ed. Akademisk Forlag.
- Grønnum, N. 2009. A Danish phonetically annotated spontaneous speech corpus (DanPASS). *Speech Comm* 51(7), 594–603. doi:10.1016/j.specom.2008.11.002.
- Gubian, M, F Torreira & L Boves. 2015. Using Functional Data Analysis for investigating multidimensional dynamic phonetic contrasts. *JPhon* 49, 16–40. doi:10.1016/j.wocn.2014.10.001.
- Harrington, J. 2000. *Phonetic analysis of speech corpora*. Wiley-Blackwell.
- Heger, S. 1981. *Sprog og lyd. Elementær dansk fonetik*. Akademisk Forlag.
- Jannedy, S & M Weirich. 2017. Spectral moments vs discrete cosine transformation coefficients. Evaluation of acoustic measures distinguishing two merging German fricatives. *JASA* 142(1), 395–405. doi:10.1121/1.4991347
- Jongman, A, R Wayland & S Wong. 2000. Acoustic characteristics of English fricatives. *JASA* 108(3), 1252–1263. doi:10.1121/1.1288413.
- Koenig, LL, CH Shadle, JL Preston & CR Mooshammer. 2013. Toward improved spectral measures of /s/. Results from adolescents. *JSLHR* 56(4), 1175–1189. doi:10.1044/1092-4388(2012/12-0038).
- Kokkelmans, J. 2021. *The phonetics and phonology of sibilants. A synchronic and diachronic OT typology of sibilant inventories*. PhD dissertation, Uni Verona.
- Kokoszka, P & M Reimherr. 2017. *Introduction to Functional Data Analysis.* CRC Press. doi:10.1201/9781315117416.
- Maegaard, M & M Monka. 2019. Patterns of dialect use. Language standardization at different rates. In M Maegaard, M Monka, KK Mortensen & AC Stæhr (eds.), *Standardization as sociolinguistic change. A transversal study of three traditional dialect areas*, 27–46. Routledge. doi:10.4324/9780429467486-2.
- Pedersen, IL. 2003. Traditional dialects of Danish and the de-dialectalization 1900-2000. *Int J Soc Lang* 159, 9–28. doi:10.1515/ijsl.2003.012.
- Petersen, JH, H Juul, N Pharao & M Maegaard. 2021. *Udtalt. En introduktionsbog til dansk fonetik*. Samfundslitteratur.
- Petersen, NR. 1983. The effect of consonant type on fundamental frequency and larynx height in Danish. *ARIPUC* 17, 55–86.
- Petersen, NR. 2009. Affrikater. In *Den store danske*. Gyldendal.
- Pouplier, M, J Cederbaum, P Hoole, S Marin & S Greven. 2017. Mixed modeling for irregularly sampled and correlated functional data. Speech science applications. *JASA* 142(2), 935–946. doi:10.1121/1.4998555.
- Puggaard-Rode, R. 2022. Analyzing time-varying spectral characteristics of speech with function-on-scalar regression. *J Phon* 95. doi:10.1016/j.wocn.2022.101191.
- R Core Team. 2021. R v4.1.2.
- Ramsay, JO & BW Silverman. 2005. *Functional data analysis.* 2nd ed. Springer. doi:10.1007/b98888.
- Reidy, PF. 2013. An introduction to random processes for the spectral analysis of speech data. *OSUWPL* 60, 67–116.
- Reidy, PF. 2015. A comparison of spectral estimation methods for the analysis of sibilant fricatives. *JASA Express Letters* 137(4), 248–254. doi:10.1121/1.4915064.
- Schachtenhaufen, R. 2022. *Ny dansk fonetik*. Books-on-demand.
- Shadle, CH & SJ Mair. 1996. Quantifying spectral characteristics of fricatives. *Proc Interspeech* 4, 1521–1524.
- Spinu, L & J Lilley. 2016. A comparison of cepstral coefficients and spectral moments in the classification of Romanian fricatives. *JPhon* 57, 40–58. doi:10.1016/j.wocn.2016.05.002.
- Stevens, KN. 1993. Modelling affricate consonants. *Speech Comm* 13(1/2), 33–43. doi:10.1016/0167-6393(93)90057-R.
- Stoel-Gammon, C, KA Williams & E Buder. 1994. Cross-language differences in phonological acquisition. Swedish and American /t/. *Phonetica* 51(1–3), 146–158. doi:10.1159/000261966.
- Wood, SN. 2017. *Generalized additive models. An introduction with R*. 2nd ed. CRC Press. doi:10.1201/9781315370279.
- Wood, SN. 2021. `mgcv`. R pkg v1.8-38.

For more info, find this book in a store near you soon



LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Tak for opmærksomheden!

r.puggaard@phonetik.uni-muenchen.de