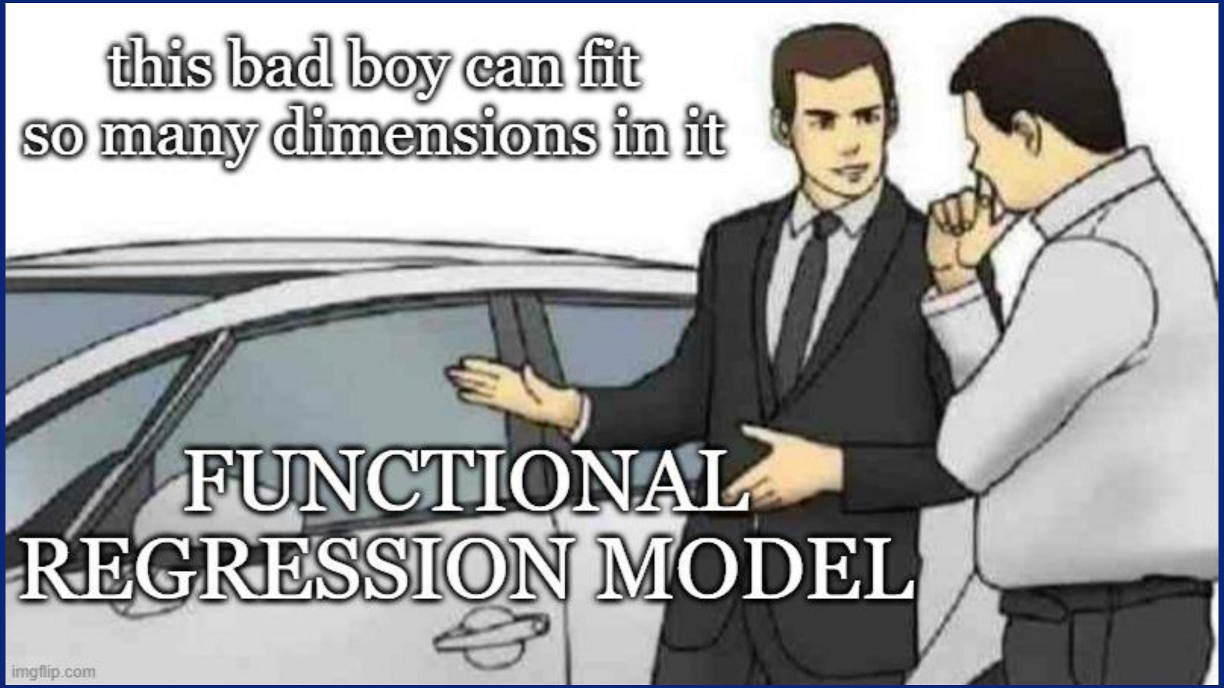
Holistic analysis of Danish stop releases



Rasmus Puggaard-Rode | r.p.hansen@hum.leidenuniv.nl

22 September 2021

Sounds of Language and Speech, Aarhus University

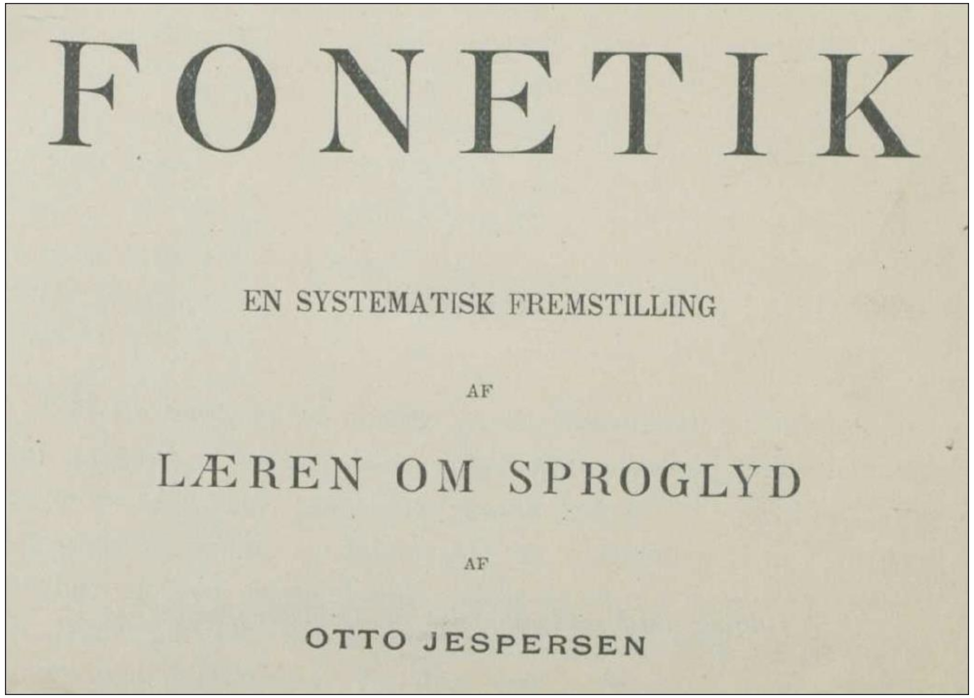
Universiteit
Leiden

Roadmap

- What's going on with Danish (aspirated) stop releases?
- What to measure?
- Analysis of dynamic data
- The study
  - Methods
  - Results
  - Discussion and remaining issues

Danish stops – or affricates?

- Danish /t/ is affricated quite dramatically
  - This was widespread already around 1850 (Brink & Lund 1975)
- Otto Jespersen assumed that Danish was undergoing the sound change /p t k/ → /pf ts kx/
- This either
  - Hasn't happened yet, but might
  - Won't happen
  - Is underway, and hasn't been noticed



FONETIK

EN SYSTEMATISK FREMSTILLING

AF

LÆREN OM SPROGLYD

AF

OTTO JESPERSEN

Jespersen 1899

/t/ transcription strategies

- [d̥^{s}] (Basbøll 1969, 2005; Grønnum 1998 narrow)
- [d̥^{sh}] (Petersen 1983)
- [d^{s}h] (Brink & Lund 1975)
- [ts^{h}] (Basbøll & Wagner 1985)
- [t^{s}] (Grønnum 1998 broad)
- [ts] (Schachtenhaufen in prep)

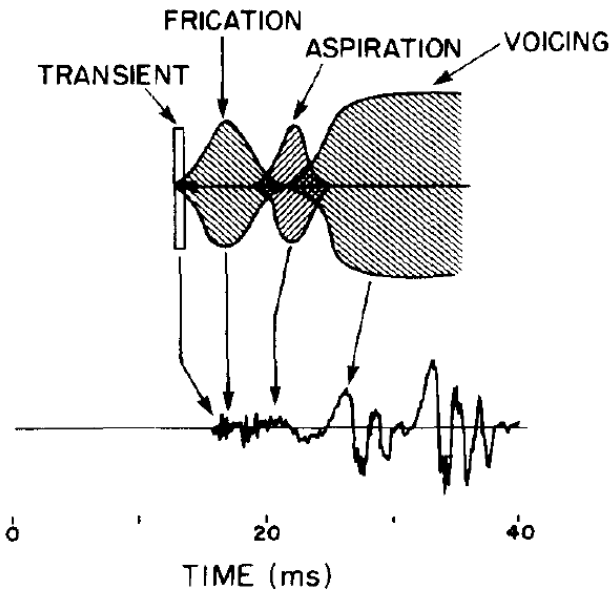
So are /p k/ stops, and /t/ an affricate?

- Well, that depends…
- Phonologically /p t k/ behave similarly
  - Similar phonotactic restrictions
  - Deaspiration in unstressed position and after /s/
  - Unreleased syllable-finally, but optionally released phrase-finally
- Phonetically, there are no clear heuristics for making the call
  - All aspirated stops have some amount of affrication – it's a design feature!
  - What amount of affrication is "enough"?

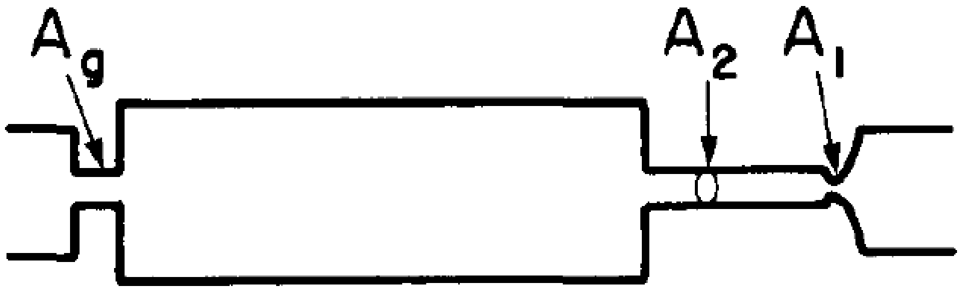
Vestergaard 1967; Grønnum 2005

Stops and affricates

- Frication in stops is transitional
  - This transitional noise is a crucial cue to place of articulation!



- Affricates have a secondary constriction



Stevens 1993a, 1993b

Stops, affricates, and activities of the glottis

- In aspirated stops in related languages (Eng, Swe, Ice, Ger),
  - Closures are relatively long   Lisker 1957; Löfqvist 1976; Stathopoulous & Weismer 1983; Braunschweiler 1997
  - The glottis is fully spread before the release   Sawashima 1970; Pétursson 1976; Löfqvist 1980; Hoole et al 1984
  - = High intraoral air pressure at the time of release
    → strong place cues in burst!
- In Danish aspirated stops,
  - Closures are relatively short   Fischer-Jørgensen 1968, 1972
  - Glottis is only fully spread just after the release   Frøkjær-Jensen et al 1971
  - Strong place cues are not ensured in initial burst
    → constriction retained to cue place!

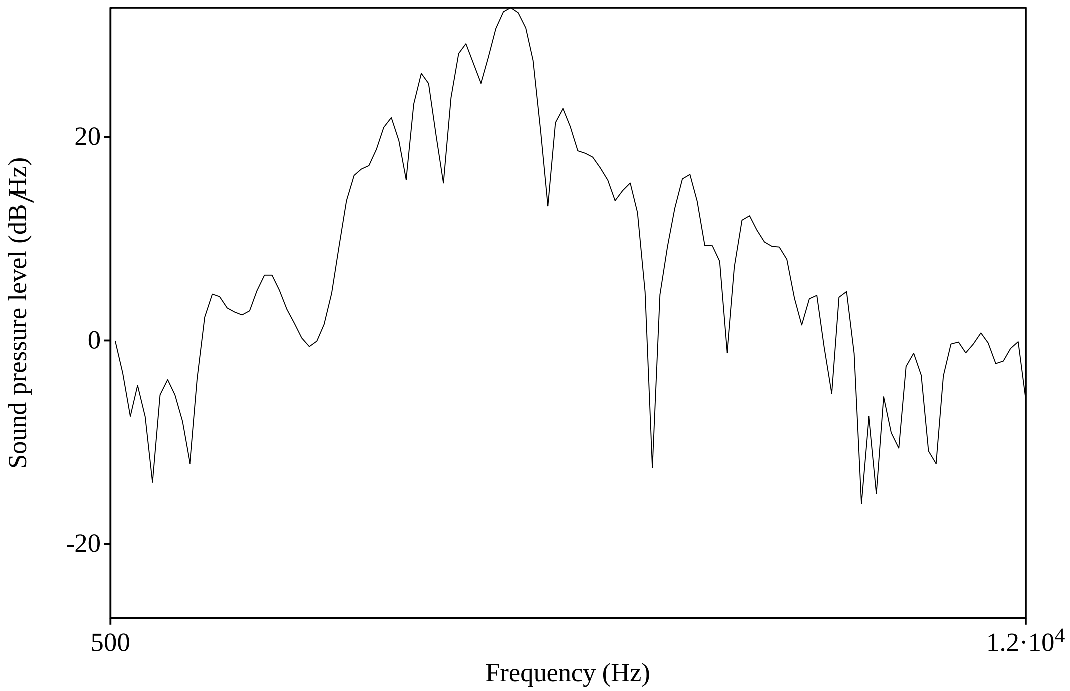
Research questions

- How do the spectral characteristics of Danish stop releases vary across time?
- How are spectral characteristics affected by different phonetic contexts?
- Are these effects robust across speakers?

- How to measure and model this??!!

Measuring frication

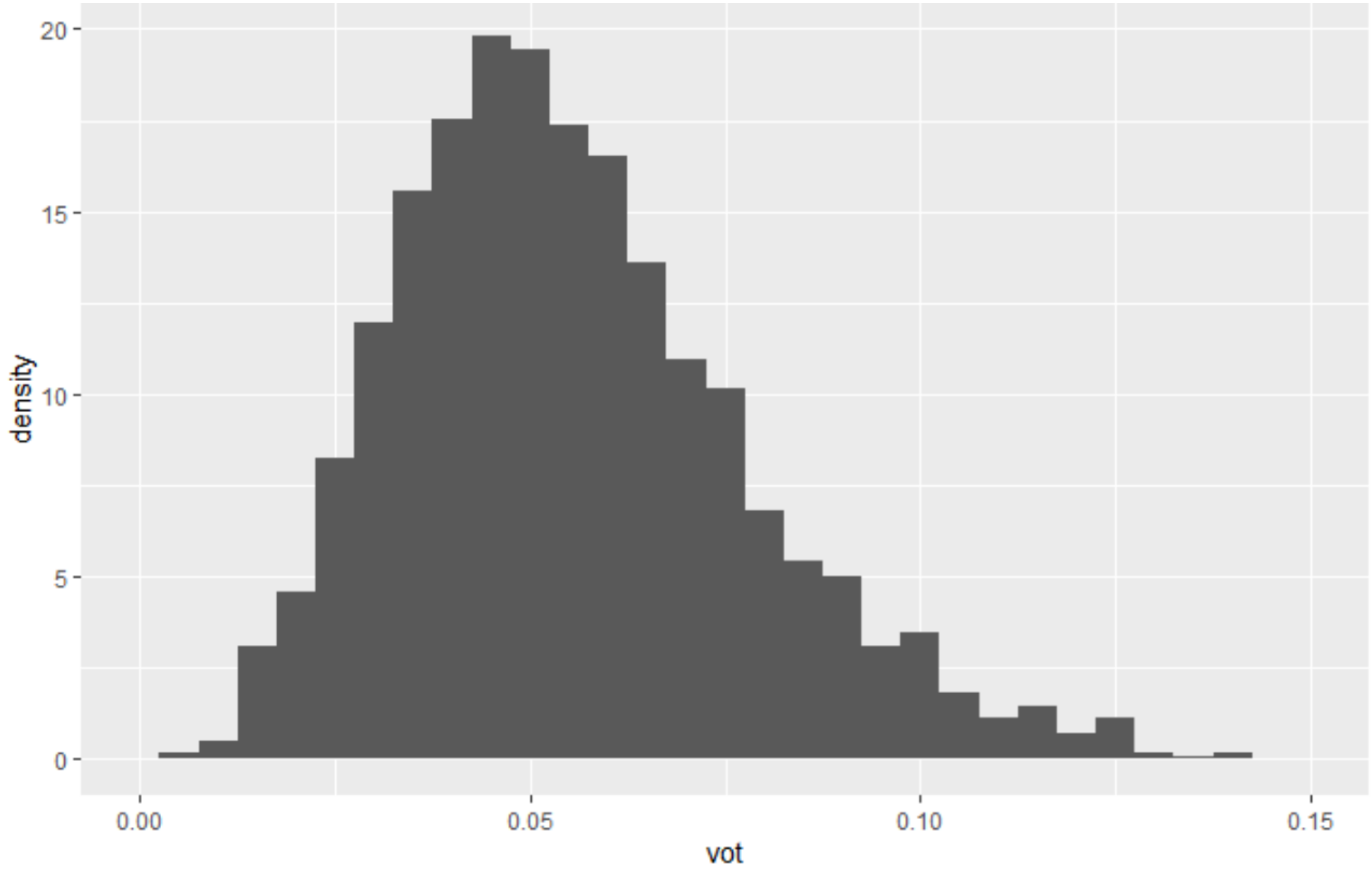
- The problem:
  - Sound is reflected in a complex spectrum
  - The spectrum contains highly multi-dimensional information
  - Most statistical models require discrete dependent variables
  - Discrete values representing entire spectra are often **insufficient** or **difficult to interpret**



The spectrum as probability mass function

- 4 highest moments

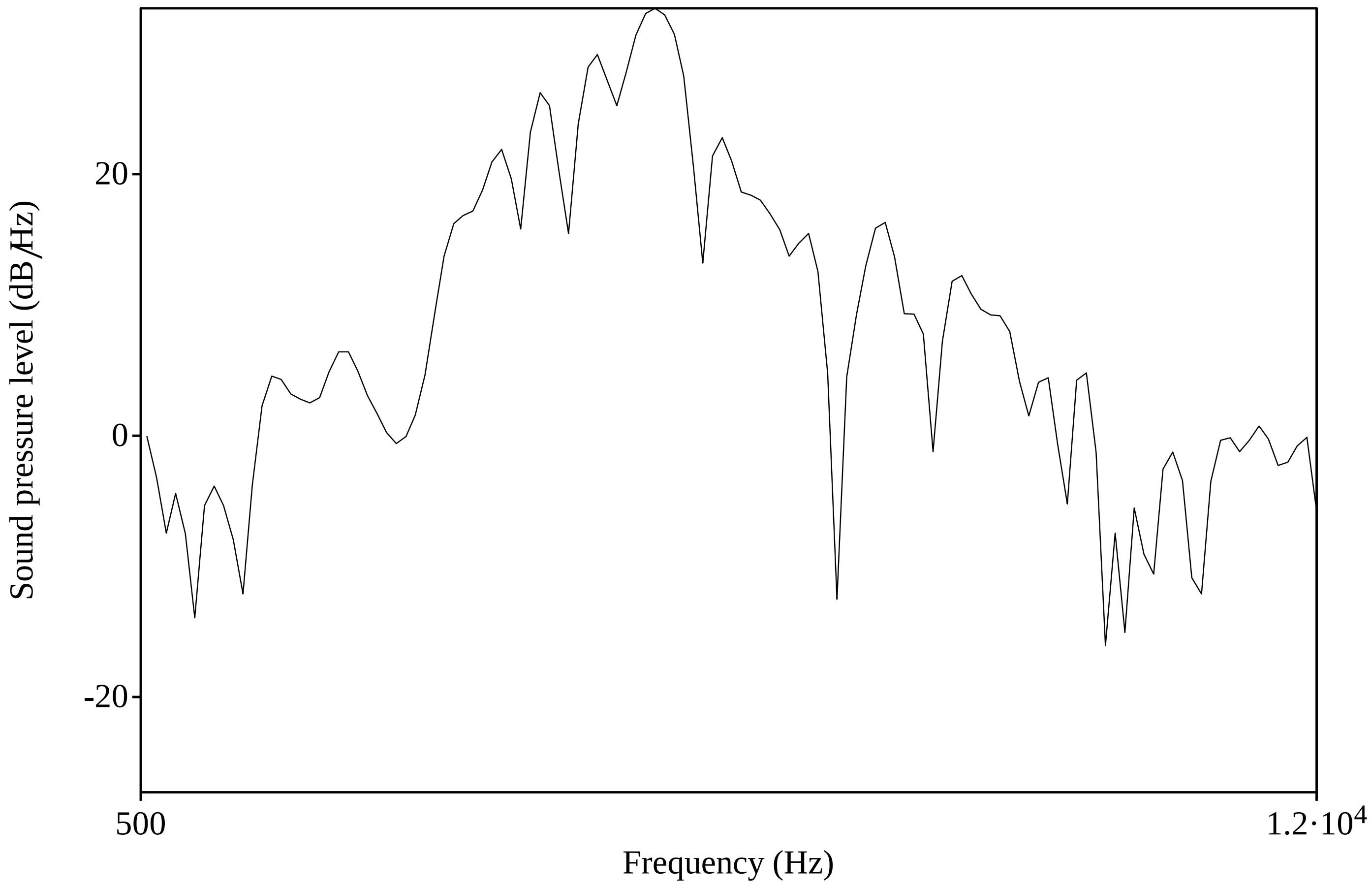
- Mean
  55.7 ms
- Standard deviation
  24.1 ms
- Skewness
  1.94
- Kurtosis
  20.12



The spectrum as probability mass function

- 4 highest moments

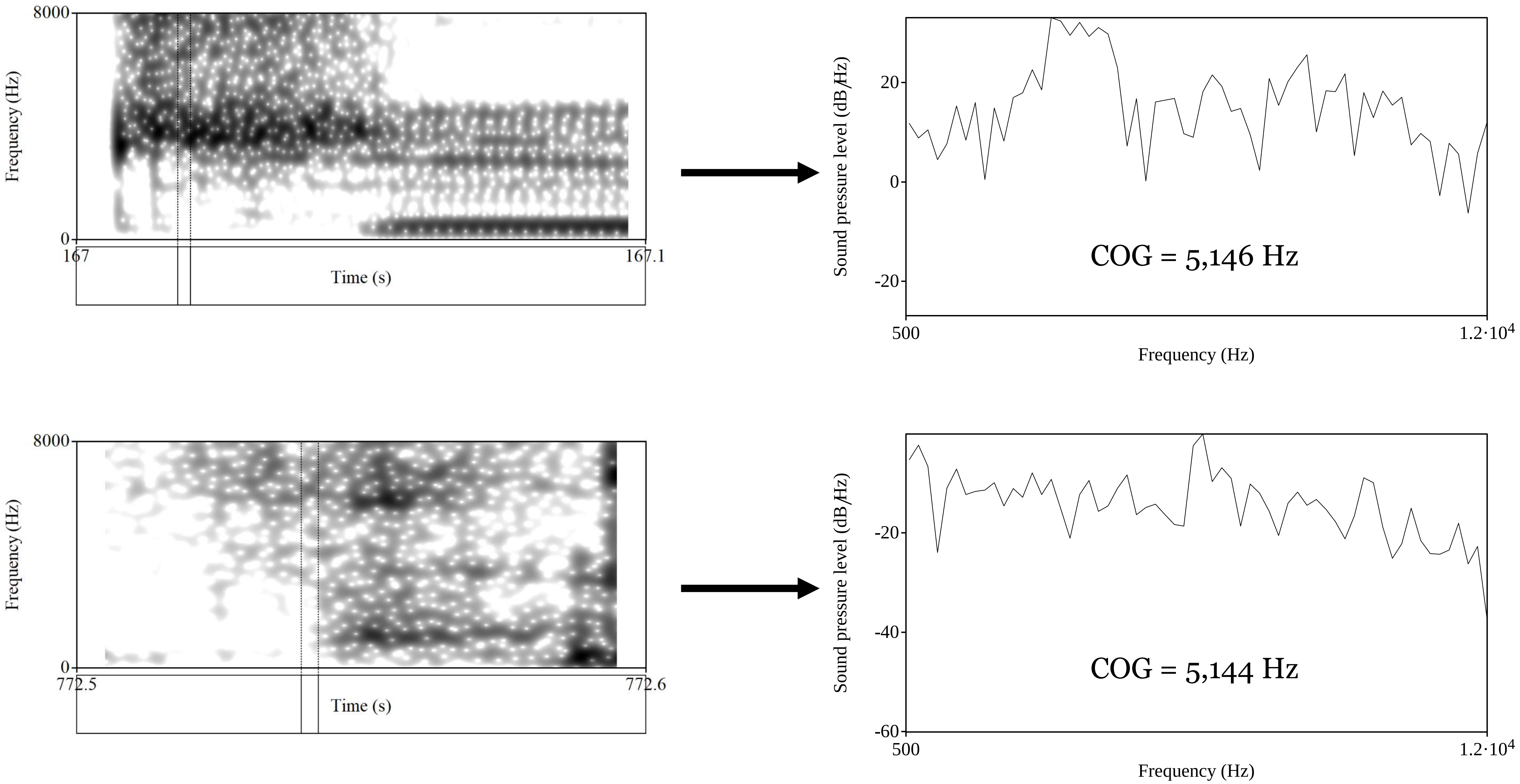
- Mean (center of gravity)
  5369 Hz
- Standard deviation
  709 Hz
- Skewness
  1.05
- Kurtosis
  9.55



Forrest et al 1988

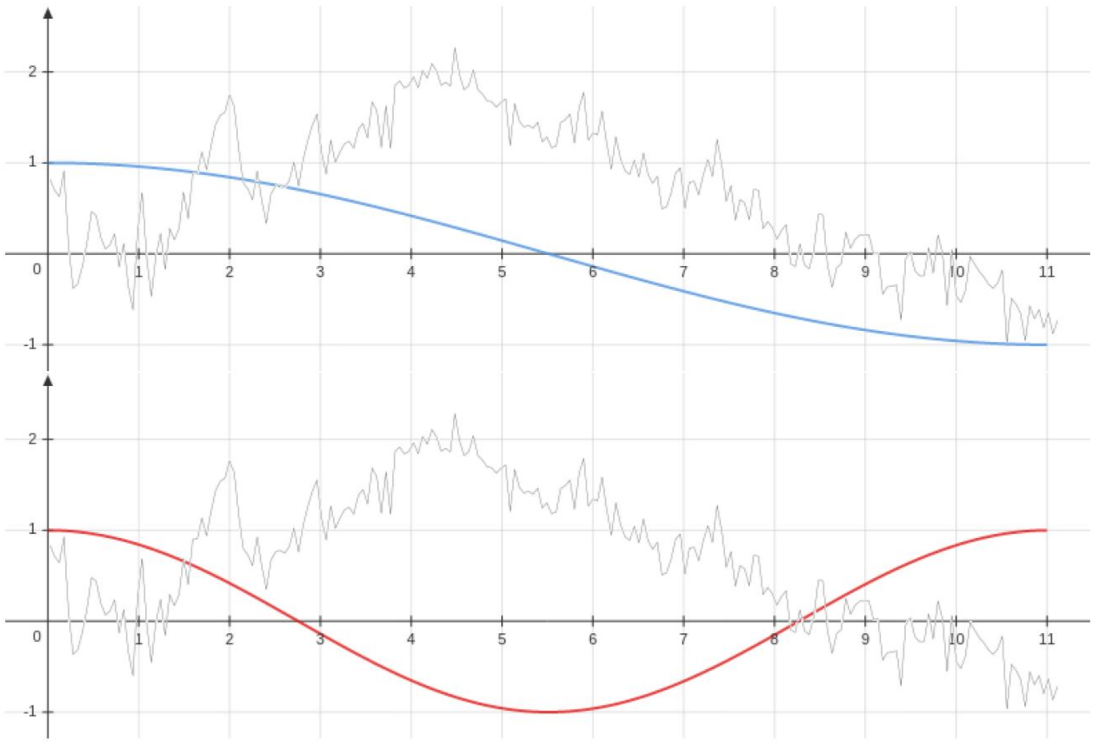
The spectrum as probability mass function

- COG and skewness are sometimes effective at predicting place of articulation
  - Especially for stop bursts
- Results are not all that stable though!
- Many studies use only COG
  - I’ve been very guilty of this
  - It’s intuitively easy to interpret
  - Has all the problems expected from reporting only the mean of a non-normally distributed function


8000
Frequency (Hz)
0
167
167.1
Time (s)
20
0
-20
Sound pressure level (dB/Hz)
COG = 5,146 Hz
500
1.2·10^4
Frequency (Hz)
8000
Frequency (Hz)
0
772.5
772.6
Time (s)
-20
-40
-60
Sound pressure level (dB/Hz)
COG = 5,144 Hz
500
1.2·10^4
Frequency (Hz)

Discrete cosine transform coefficients

- Discrete cosine transform decomposes a signal into (usually four) cosine waves
- DCT coefficients reflect how similar spectra are to the cosine waves
  - DCT0 = mean amplitude
  - DCT1 = slope
  - DCT2 = curvature
  - etc…



Kokkelmans 2021; Jannedy & Weirich 2017

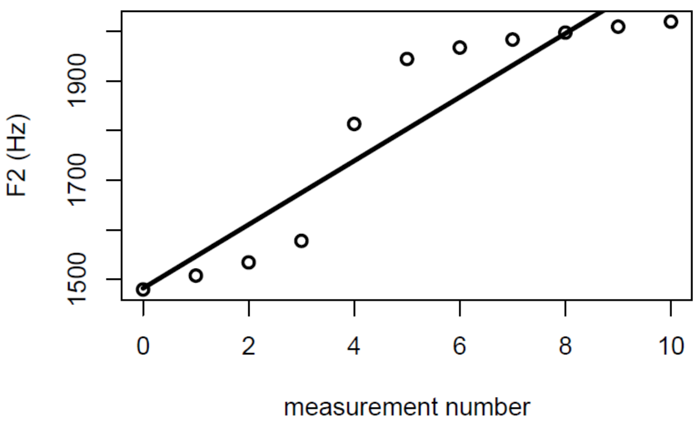
Moments vs DCT

- Both analyses usually yield 4 discrete values
  - Ideally we want just 1 ☹
  - Multiple statistical models = results more difficult to interpret
- Moments
  - Sorta problematic assumptions
  - Fairly intuitive
- DCT
  - More reflective of actual perception
  - More information about the spectrum
  - Not very intuitive at all

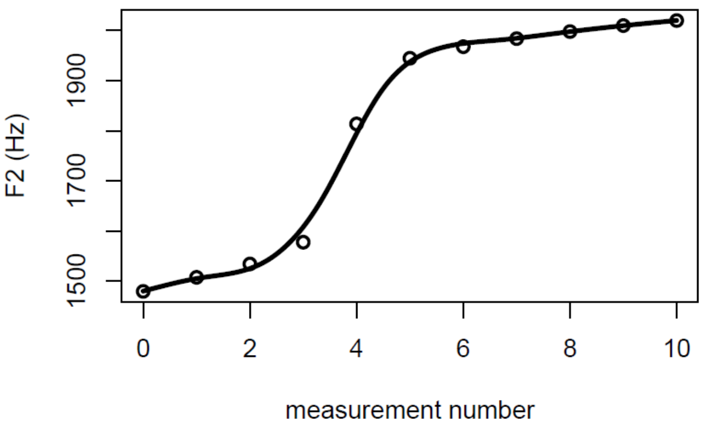
Analysis of dynamic data: GAMMs

- In the past ~5 years, the **generalized additive mixed model** has become a very popular tool in linguistics
- Particularly popular for variables that vary dynamically as a function of time

Linear model



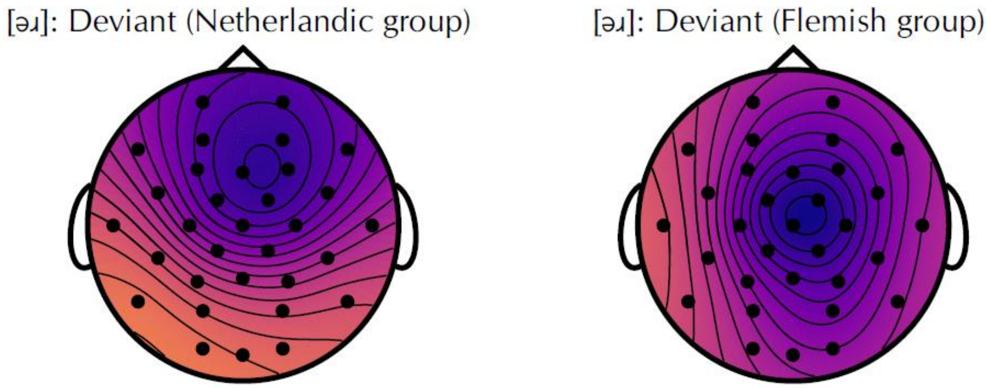
GAM



Sóskuthy 2017

Analysis of dynamic data: GAMMs

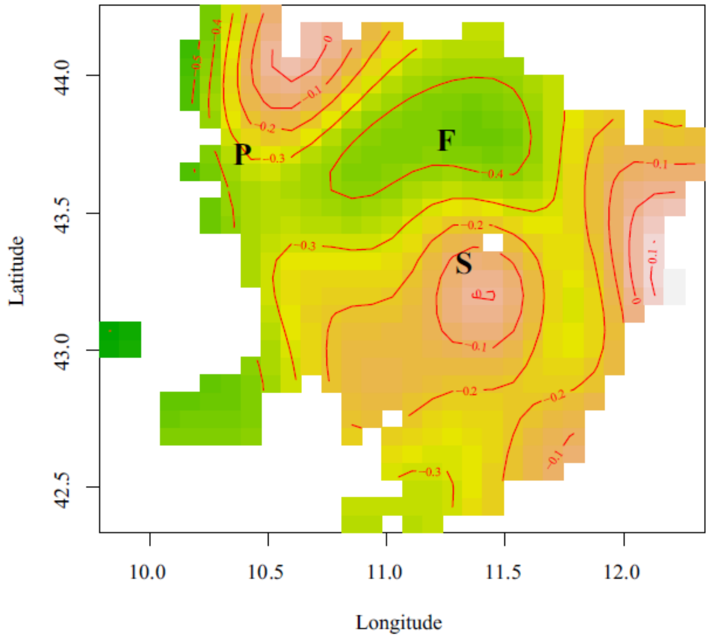
- GAMMs have also been used to analyze other kinds of dynamic linguistic data
  - EEG data

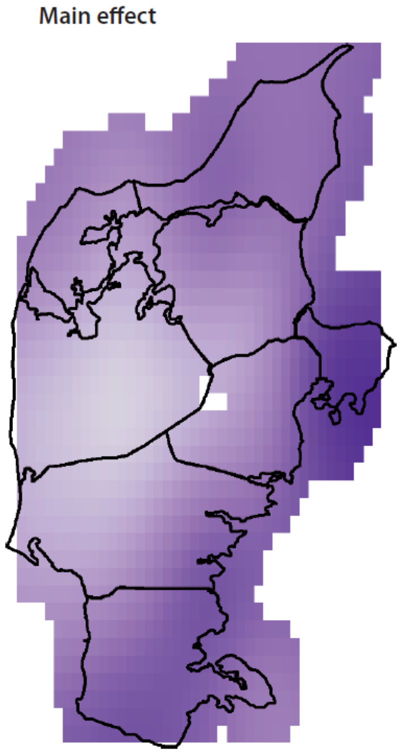


Voeten 2020

Analysis of dynamic data: GAMMs

- GAMMs have also been used to analyze other kinds of dynamic linguistic data
  - Geolinguistic variation

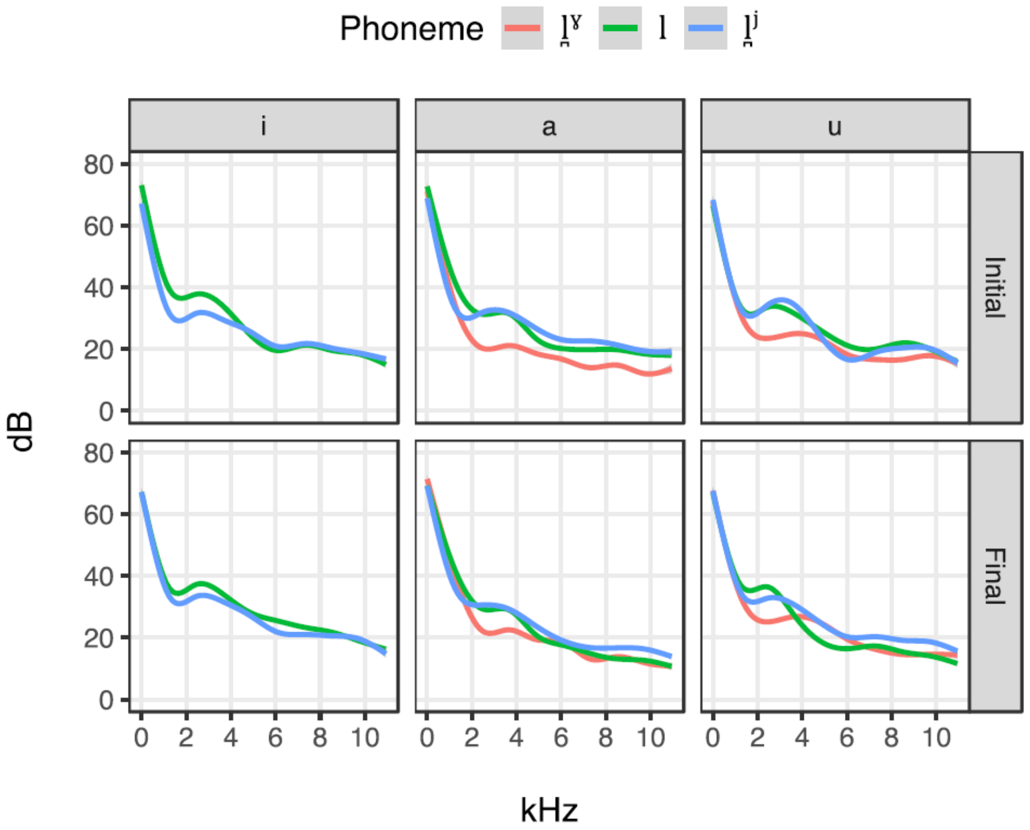




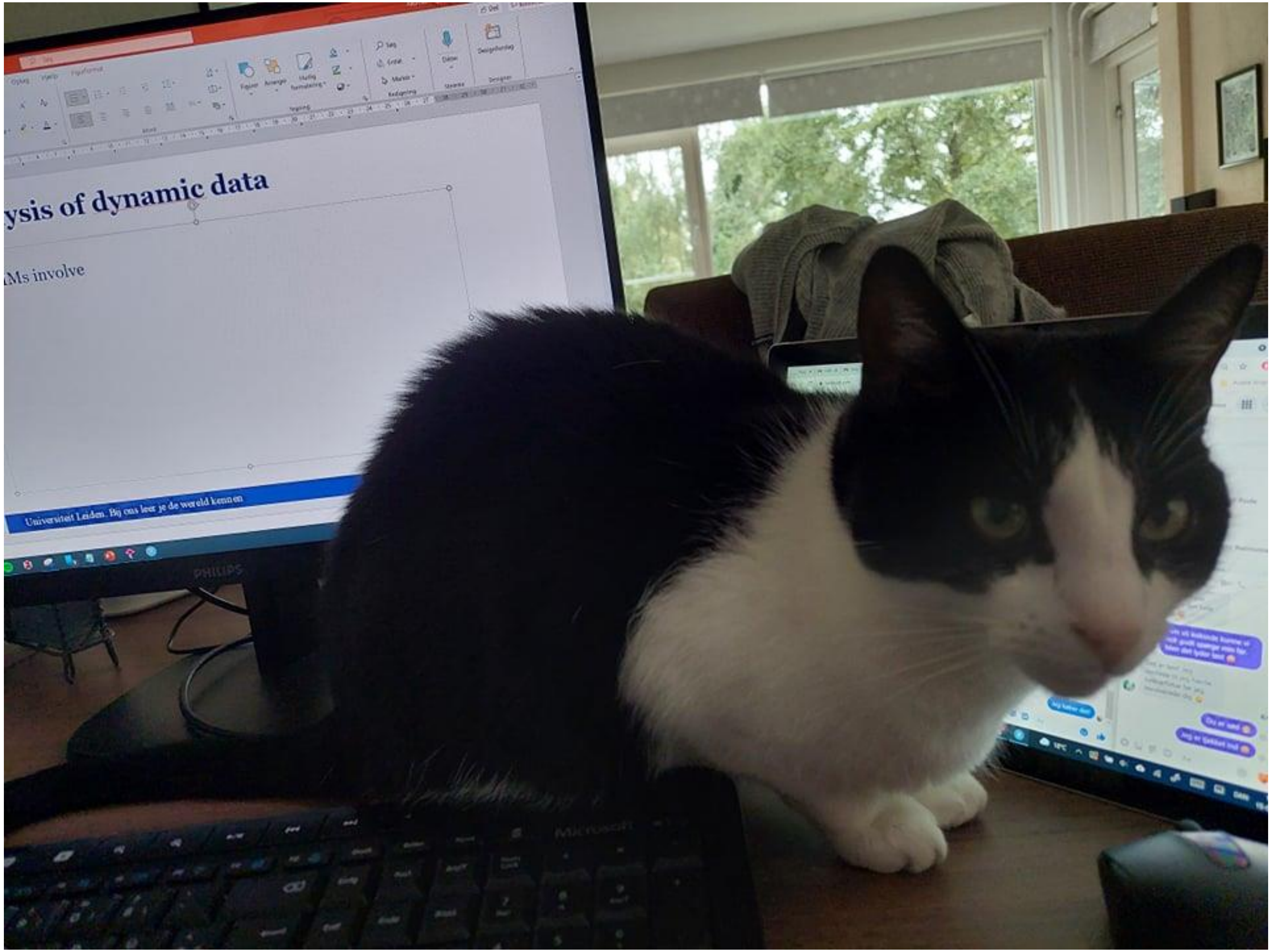
Wieling et al 2014; Puggaard 2021

Analysis of dynamic data: GAMMs

- GAMMs have also been used to analyze other kinds of dynamic linguistic data
  - Sound spectra

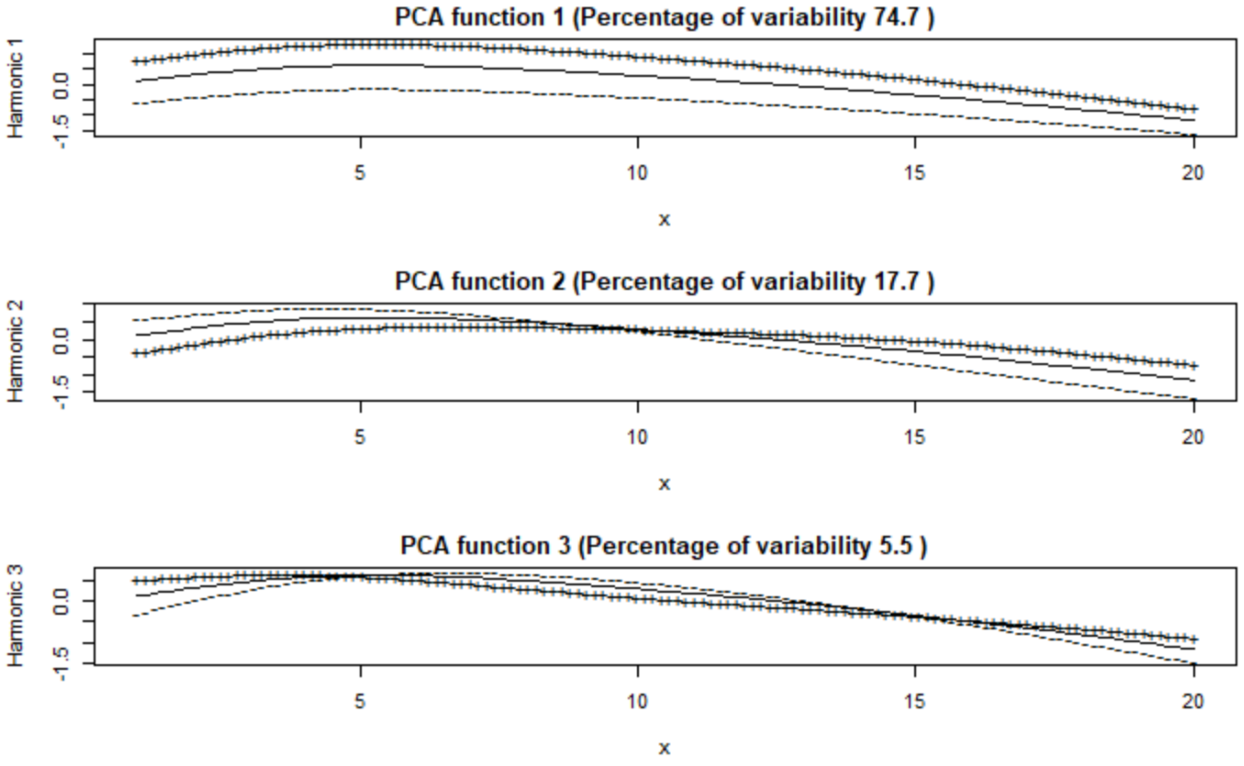


Nance & Kirkham 2020


ysis of dynamic data
Ms involve
Universiteit Leiden. Bij ons leer je de wereld kennen
PHILIPS

Analysis of dynamic data: FDA

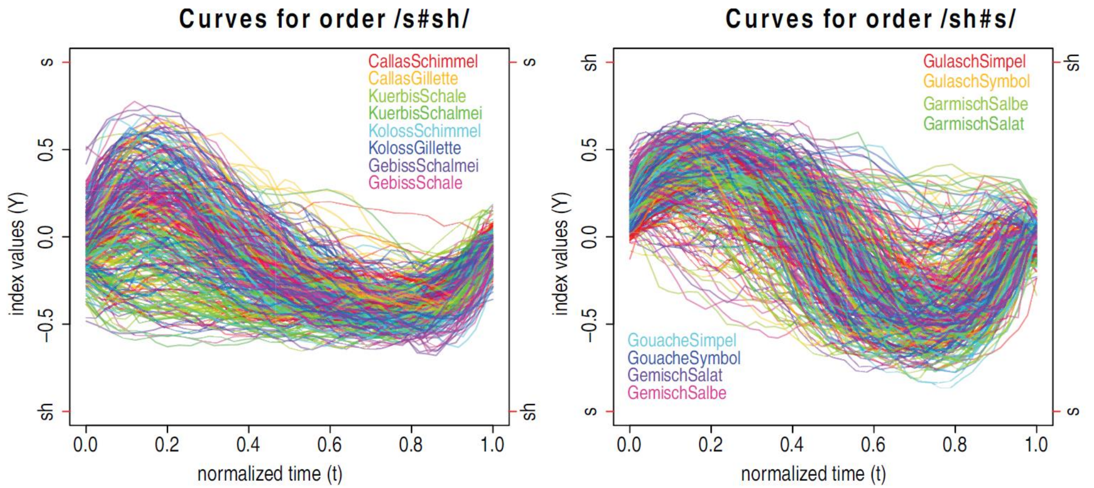
- Functional data analysis has had less influence on linguistics
- Family of methods that uses functions as input variables
  - E.g. curves
  - Functional principal components analysis
  - Scalar-on-function regression
  - **Function-on-scalar regression**
  - Function-on-function regression



Gubian et al 2015; Greven & Scheipl 2017

Function-on-scalar regression

- Similar to GAMMs
- Dependent variable is a function instead of something discrete
- Has been used for **spectra** before
  - As in… infrared spectra, and protein mass spectra
- Has been used in linguistics before
  - Well… sort of
- Never used for sound spectra before



Cederbaum et al 2016; Wood 2017; Bauer et al 2018; Goldsmith et al 2021

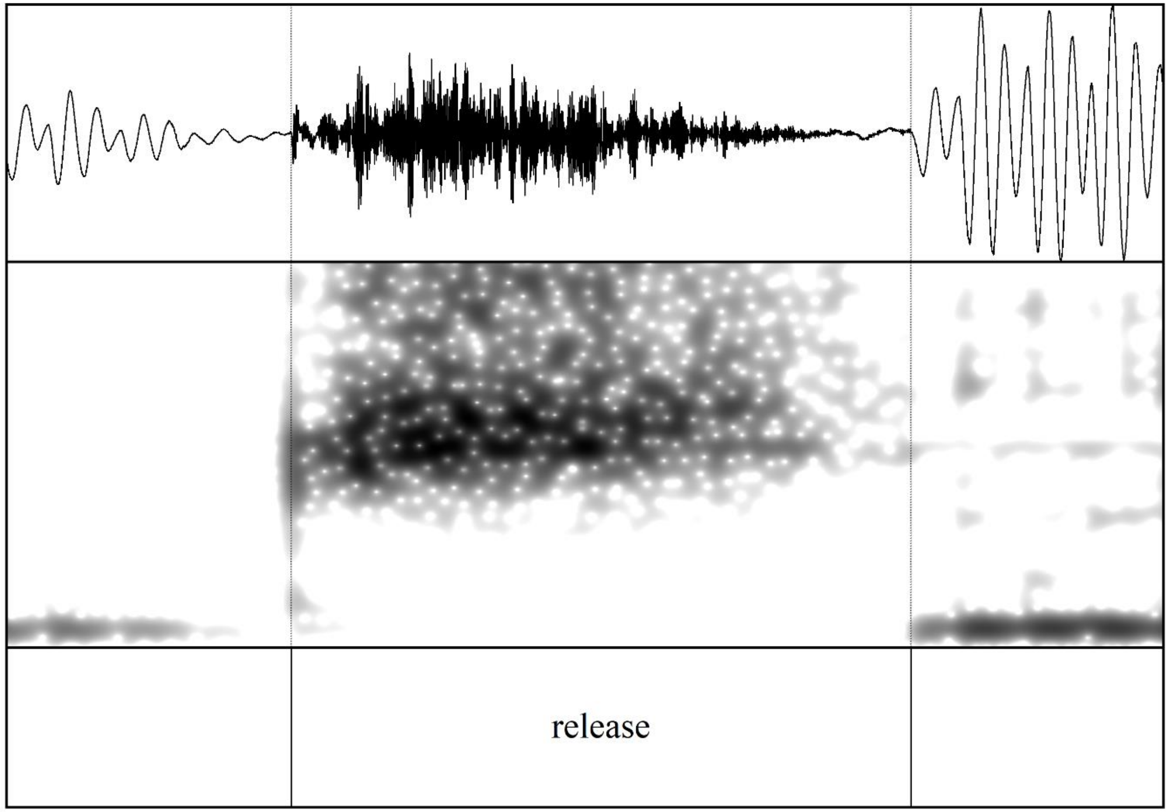
The corpus

- Danish Phonetically Annotated Speech (DanPASS)
- Specifically the 1996 monologues
  - 18 speakers
  - 171 minutes of speech
  - 13 men, 5 women
  - Mean age = 29 years
  - Quite thorough transcriptions

Grønnum 2009

Acoustic analysis

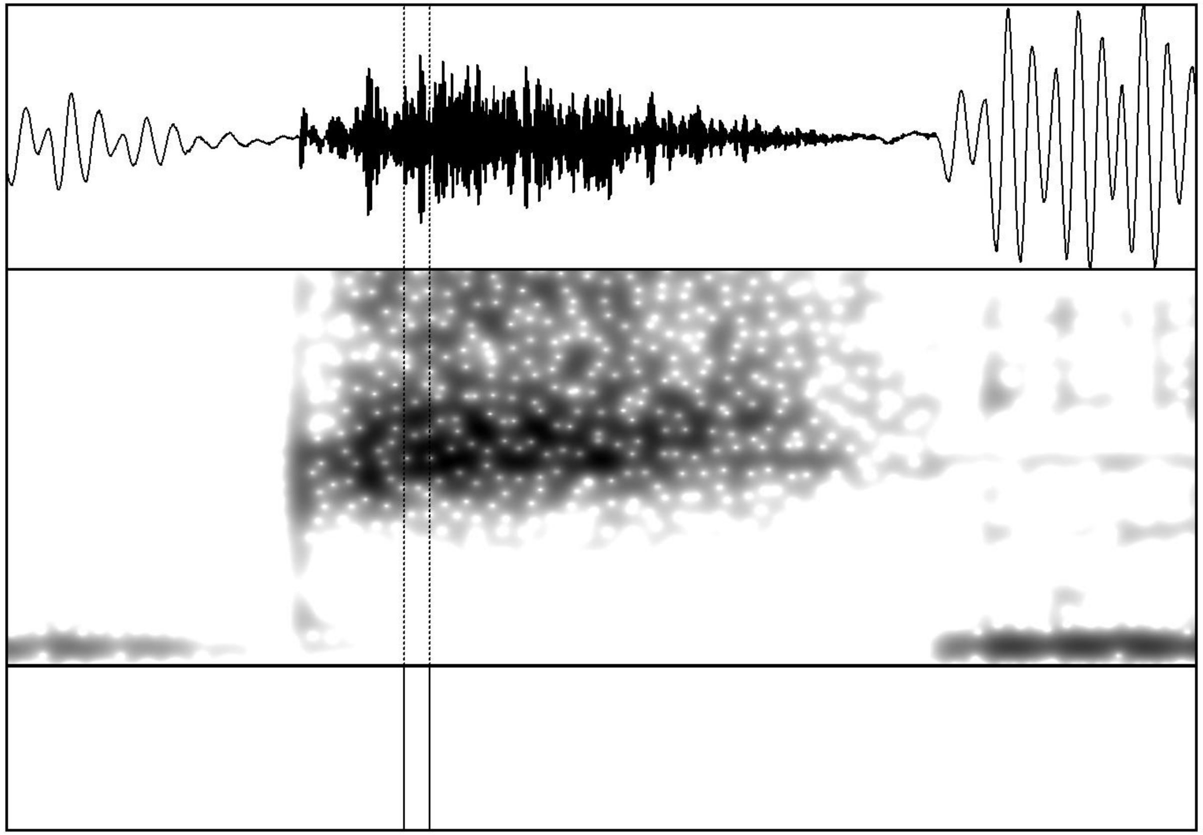
- Praat script used to isolate /p t k/ in simple onset
- Partially automatic segmentation of stop releases
- Info about phonetic context extracted from transcriptions
- 2,539 stops
  - 205 stops excluded if no distinct closure phase could be located



Boersma & Weenink 2019; many thanks for Dirk Jan Vet for help with Praat scripting!

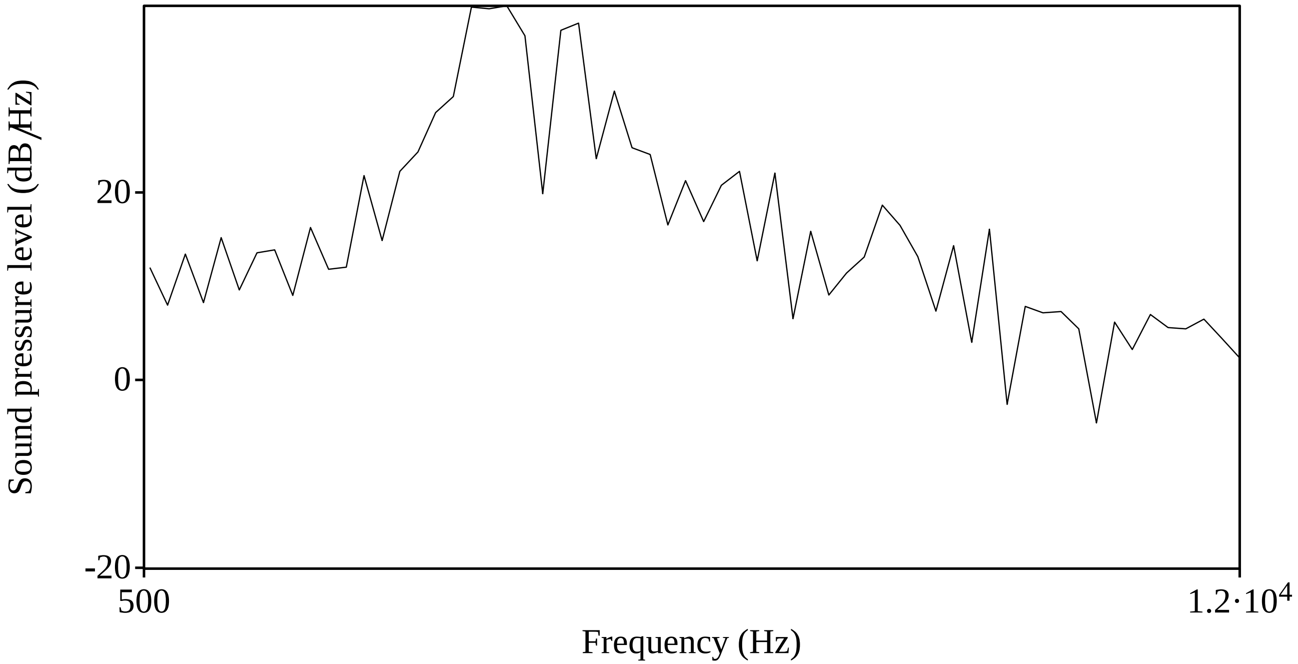
Acoustic analysis

- Each token split into 20 equally long time steps
- Frequencies >500 Hz and <12,000 Hz filtered away
  - Only frequencies below <10,000 Hz used for modeling /t/, <8,000 Hz for /p k/
- Multitaper spectra generated for each time step in R
  - Preferable to FFT spectra if the signal is dynamic

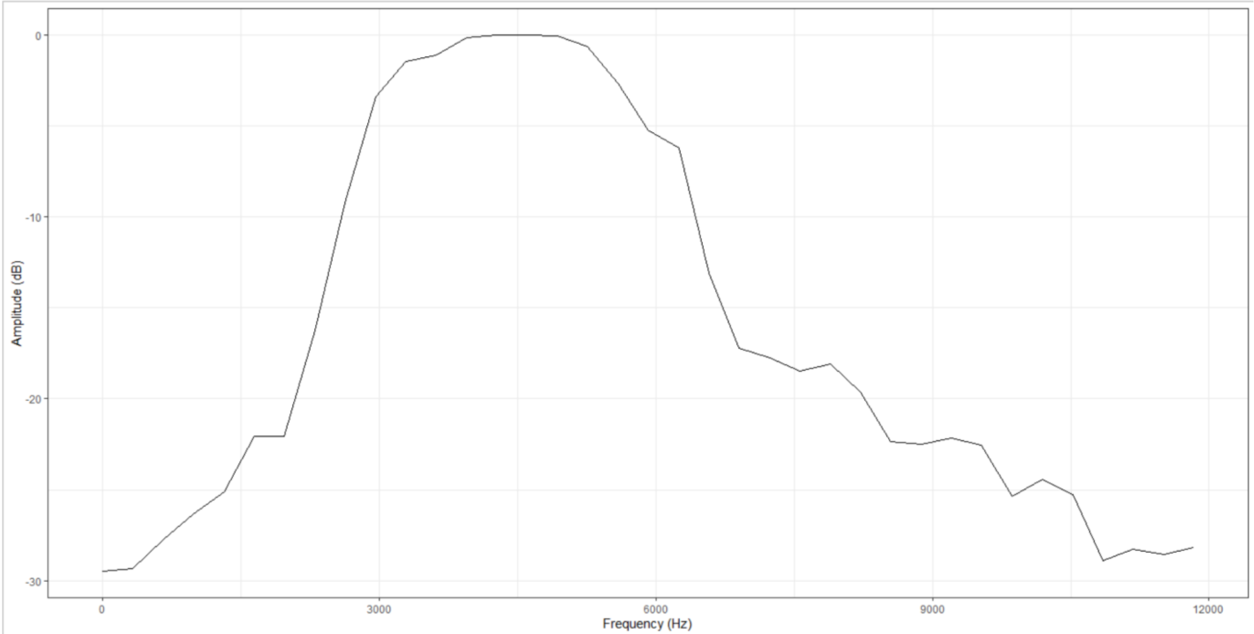


Blacklock 2004; Reidy 2016; R Core Team 2020

Acoustic analysis



- FFT spectrum



- Multitaper spectrum
- Each of these correspond to a vector of amplitude values for frequencies

Statistical analysis

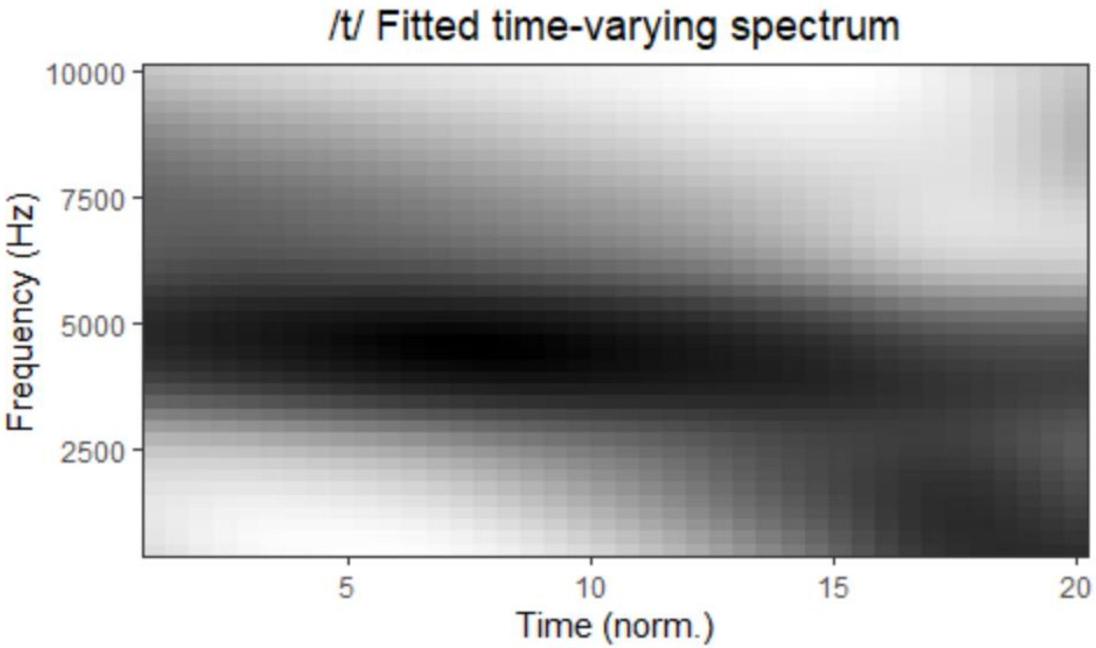
- Amplitude values were standardized within each curve
- Categorical variables were contrast-coded
- For each stop, a model was fitted of the form

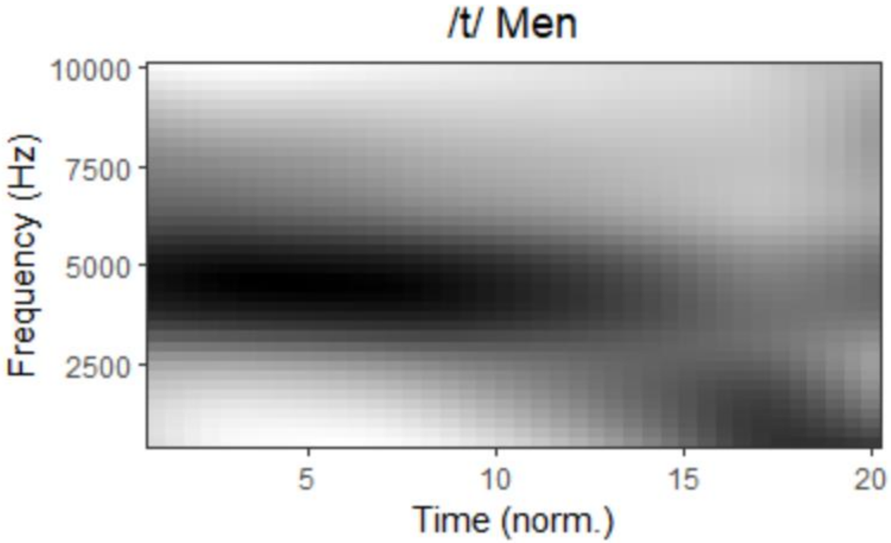
  - Spectrum ~ time + stress + sex + high vowel + round vowel + back vowel

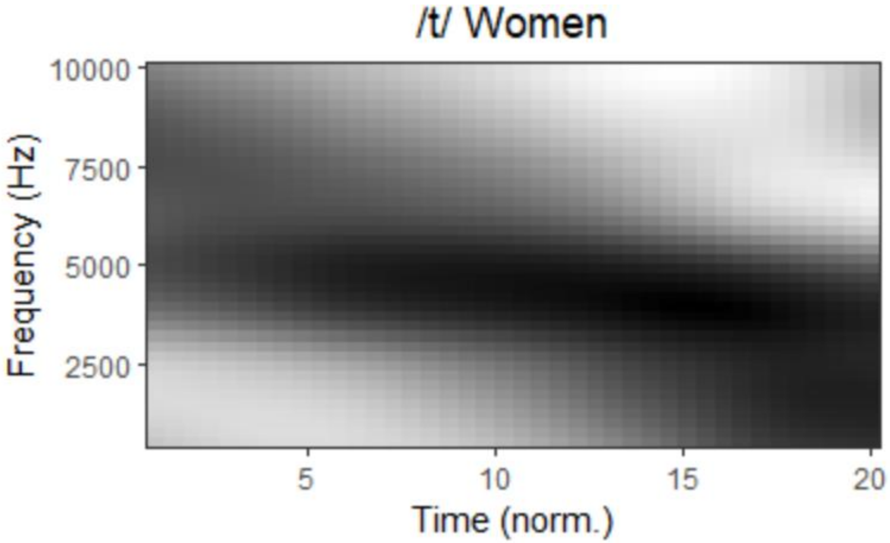
- All categorical variables were modeled as time-varying
- By-speaker random slopes were fitted for all variables (except sex)

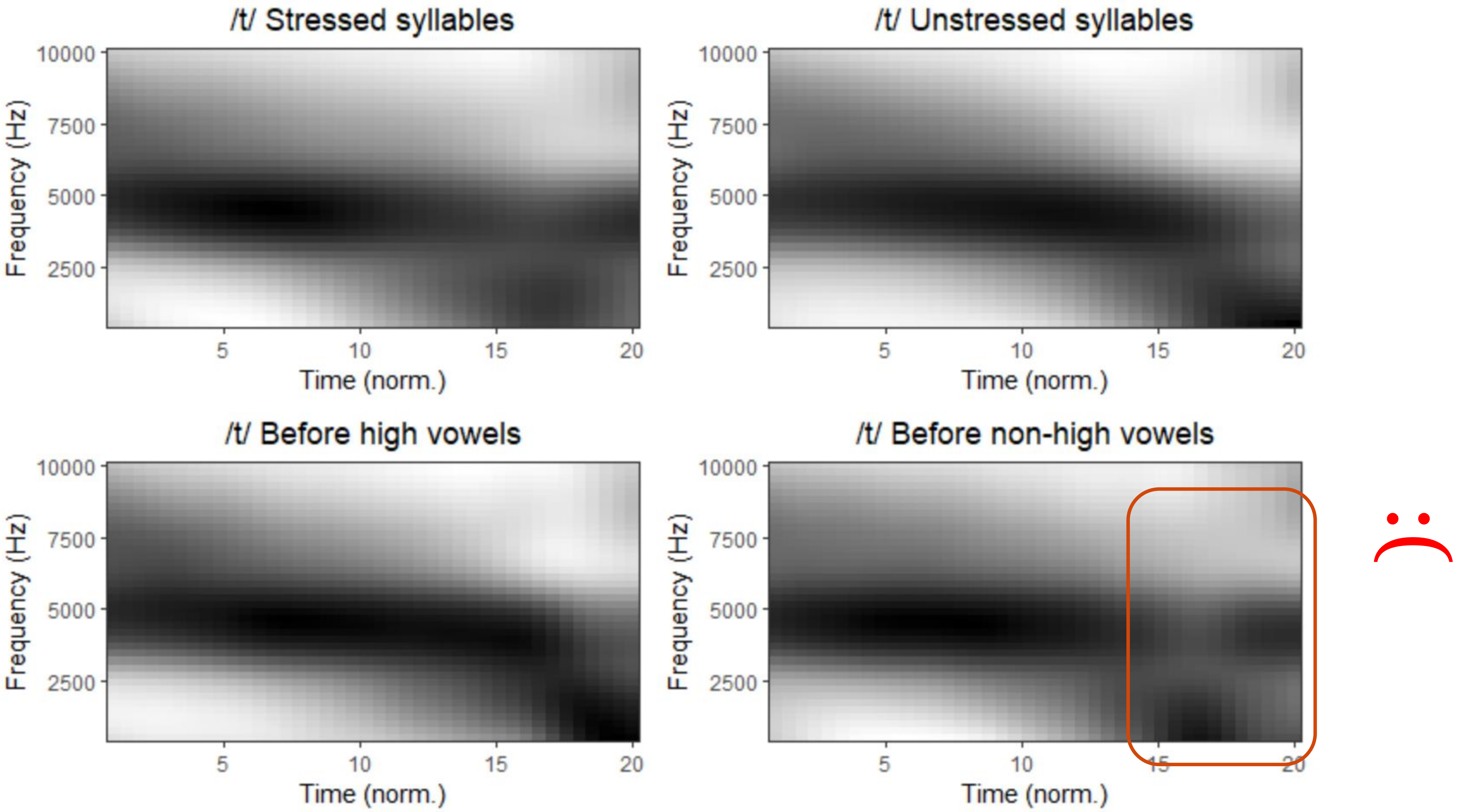
Statistical analysis

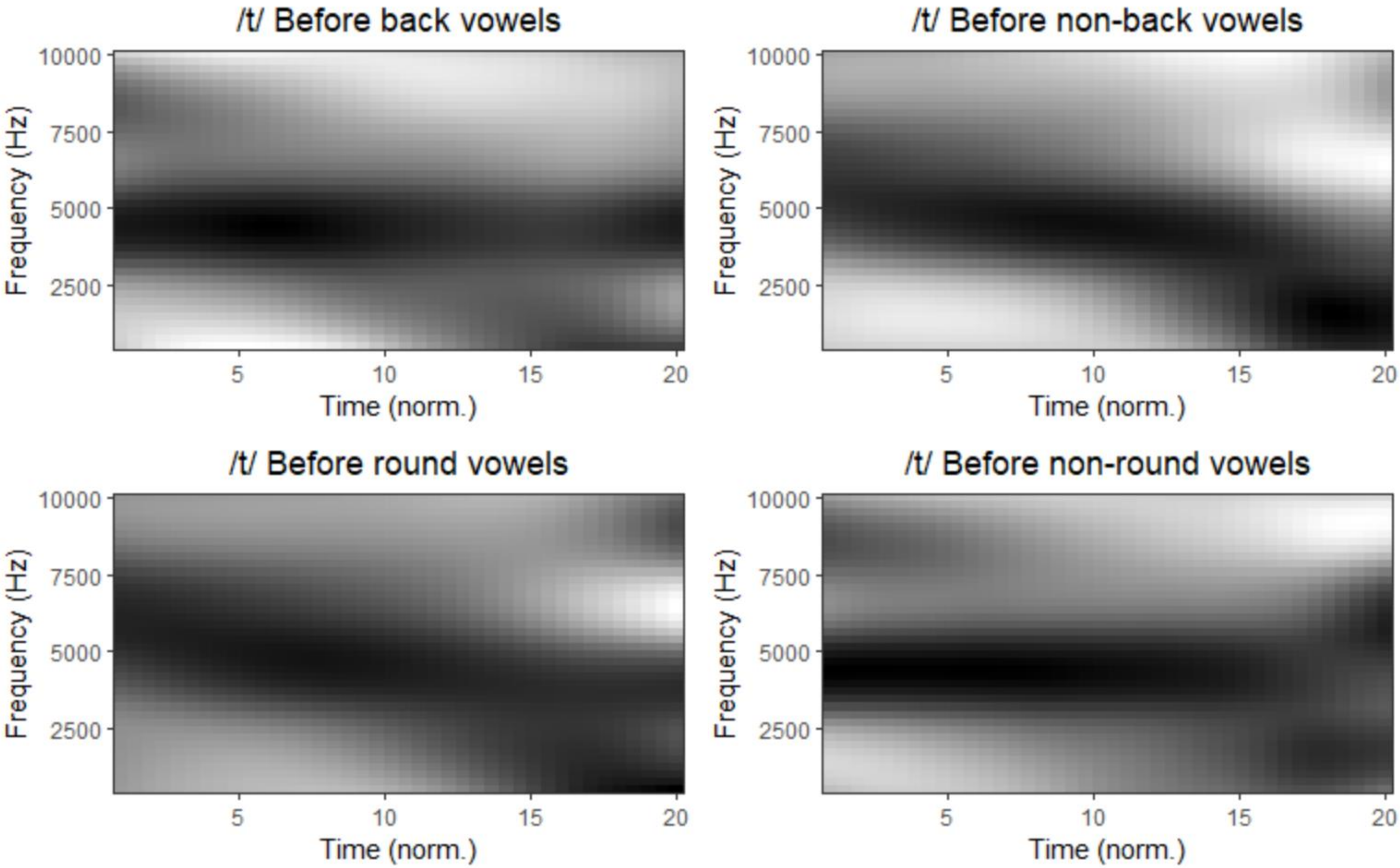
```
t_mod_ar = pffr(Y ~ s(timestep, k=16) +
                s(timestep, k=16, by=stressn) +
                s(timestep, k=16, by=sexn) +
                s(timestep, k=16, by=high_vn) +
                s(timestep, k=16, by=back_vn) +
                s(timestep, k=16, by=round_vn) +
                s(speaker, 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,
            control=list(trace=TRUE))
```


/t/ Fitted time-varying spectrum
10000
7500
5000
2500
Frequency (Hz)
5
10
15
20
Time (norm.)


/t/ Men
10000
7500
5000
2500
Frequency (Hz)
5
10
15
20
Time (norm.)

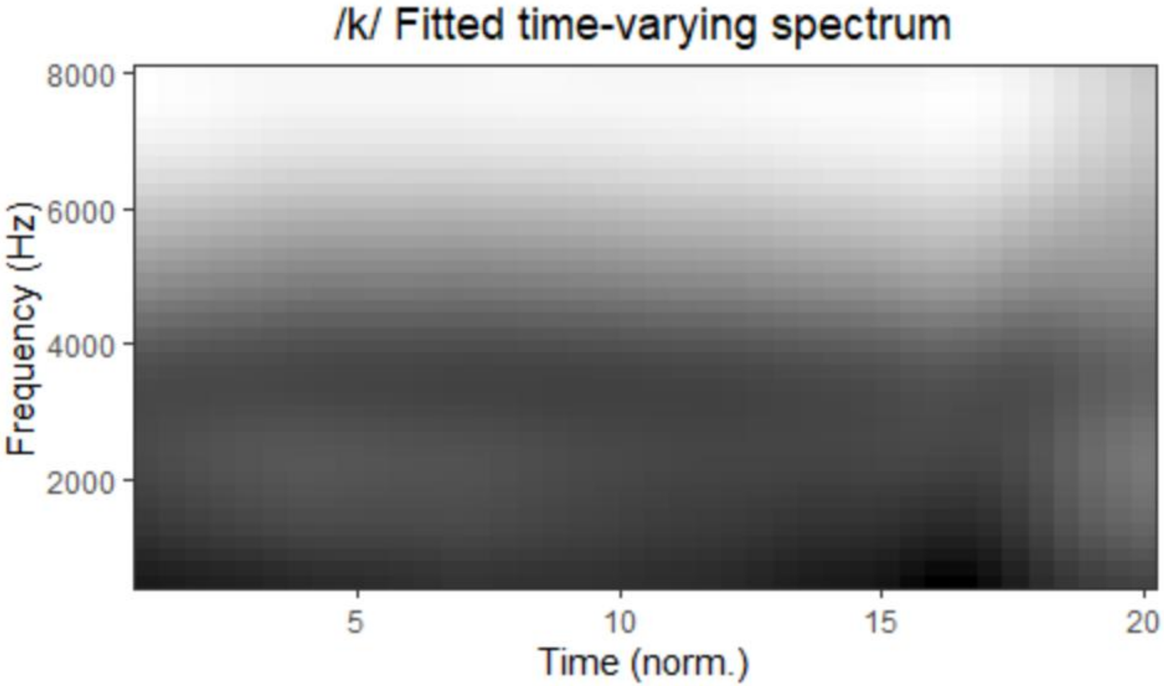

/t/ Women
10000
7500
5000
2500
Frequency (Hz)
5
10
15
20
Time (norm.)

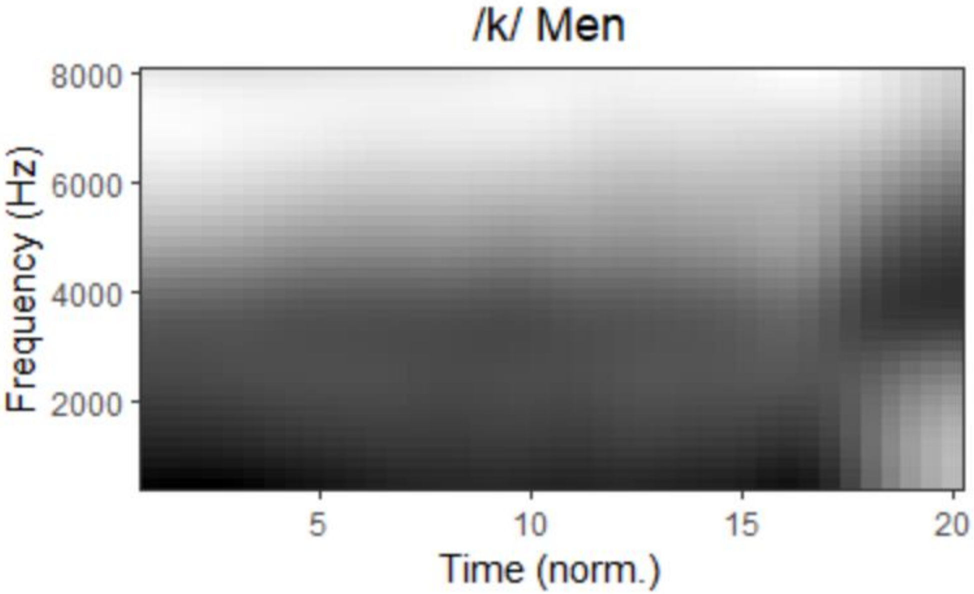

/t/ Stressed syllables
/t/ Unstressed syllables
/t/ Before high vowels
/t/ Before non-high vowels
Frequency (Hz)
10000
7500
5000
2500
5
10
15
20
Time (norm.)

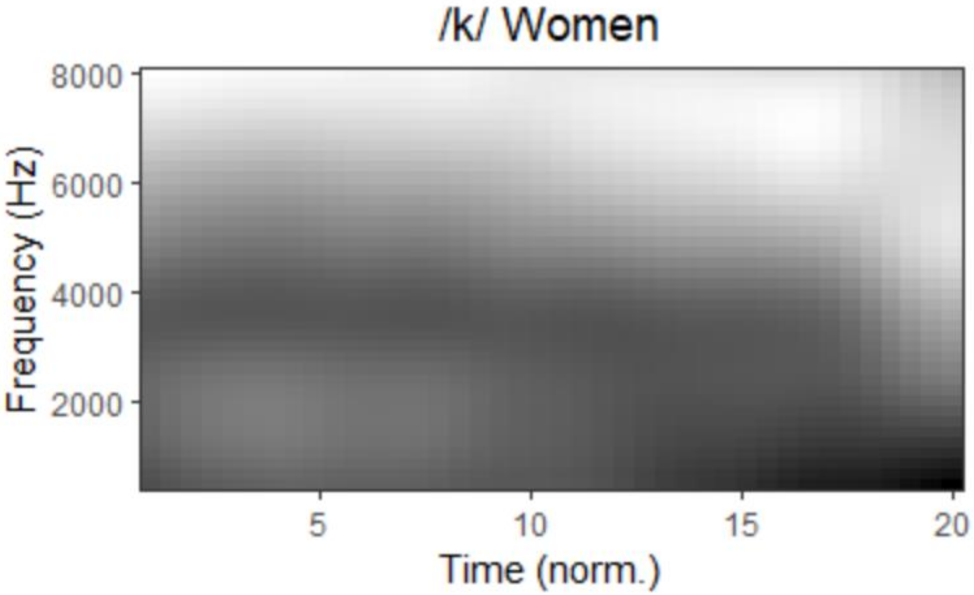

/t/ Before back vowels
/t/ Before non-back vowels
/t/ Before round vowels
/t/ Before non-round vowels
Frequency (Hz)
10000
7500
5000
2500
Time (norm.)
5
10
15
20

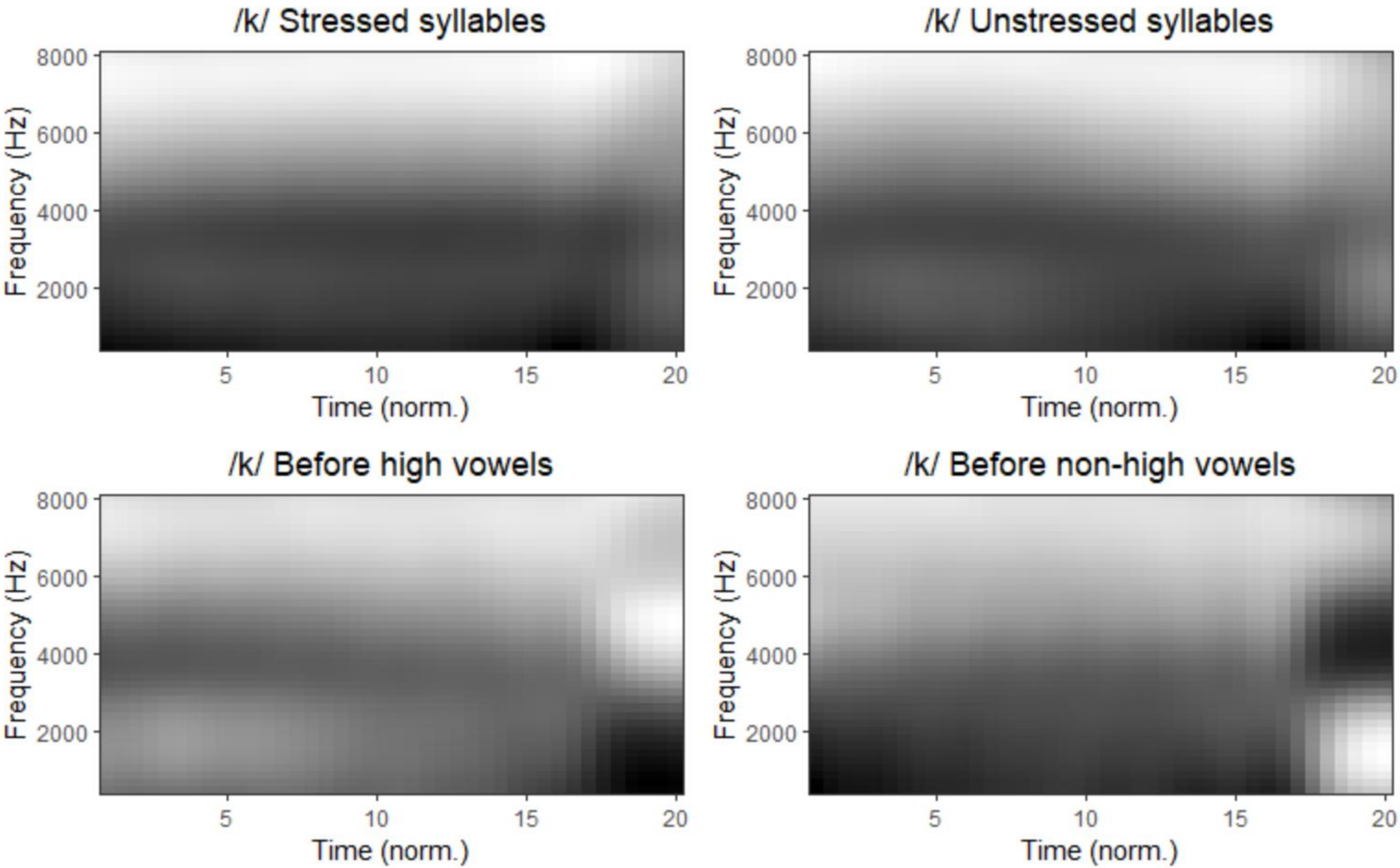
Discussion

- /t/
  - Energy distribution consistent with alveolar noise source throughout most of the release
  - Particularly for women, in unstressed syllables, and when following vowel is high/non-back

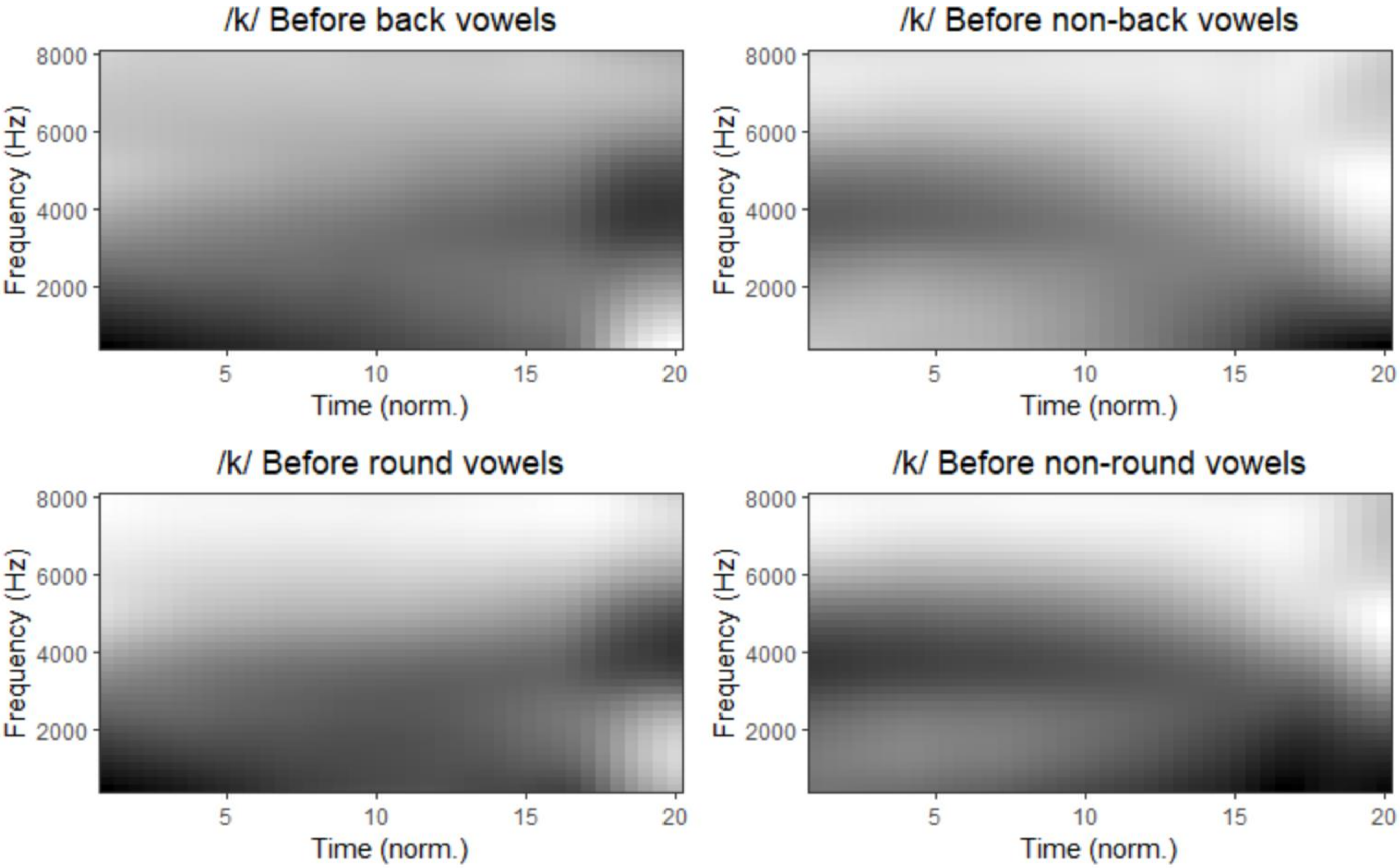

/k/ Fitted time-varying spectrum
Frequency (Hz)
8000
6000
4000
2000
5
10
15
20
Time (norm.)


/k/ Men
Frequency (Hz)
8000
6000
4000
2000
5
10
15
20
Time (norm.)


/k/ Women
Frequency (Hz)
8000
6000
4000
2000
5
10
15
20
Time (norm.)

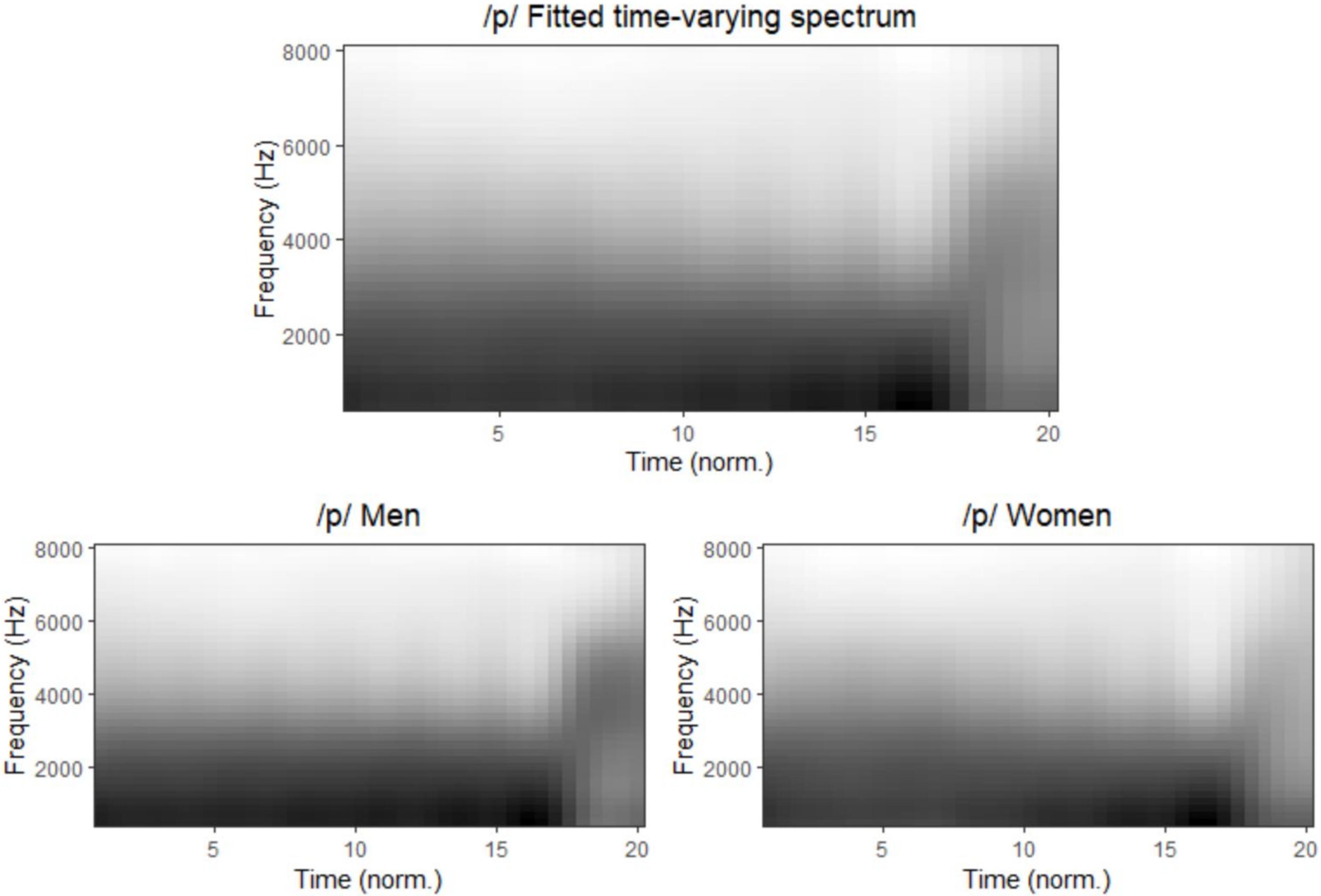


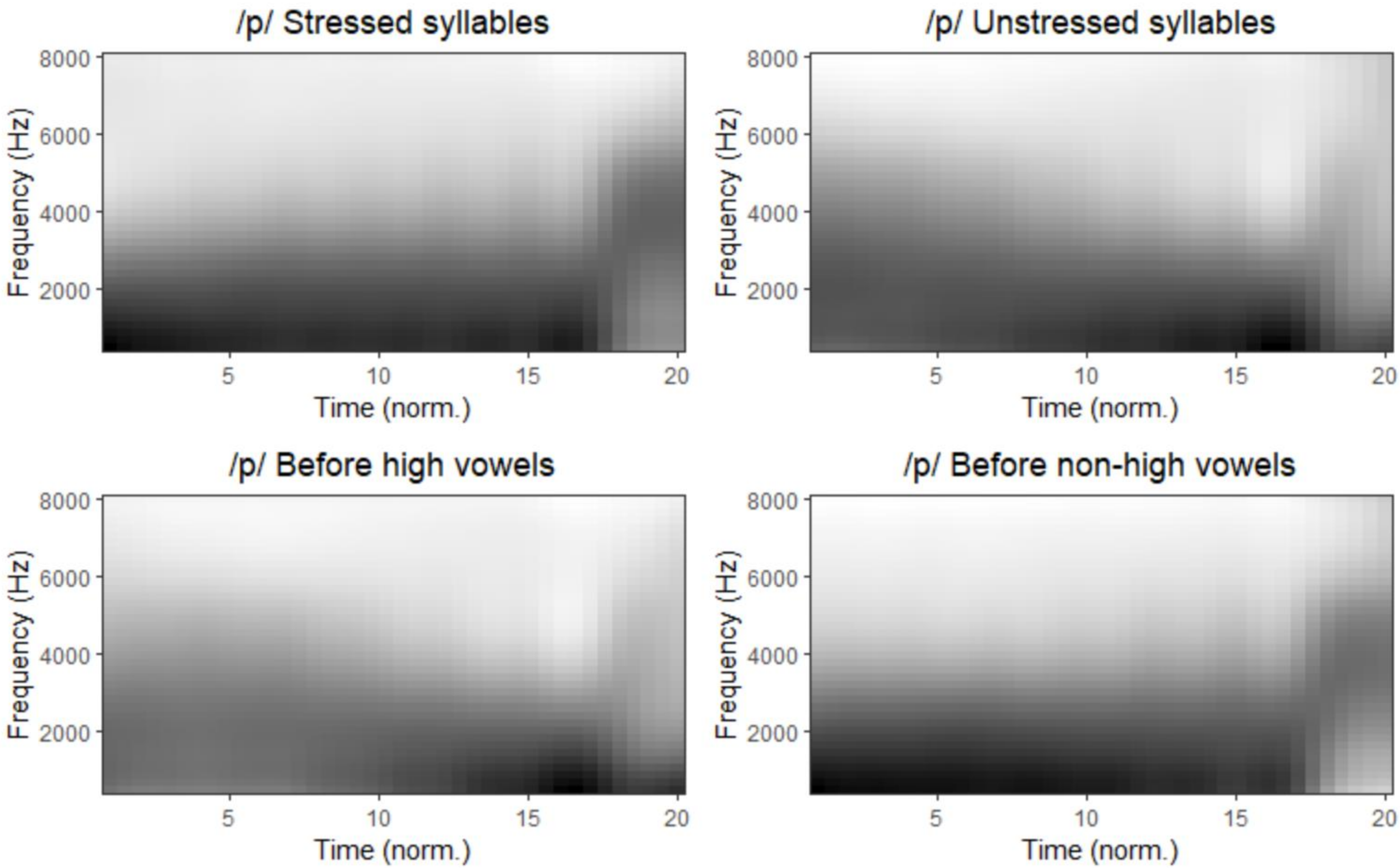
/k/ Stressed syllables
/k/ Unstressed syllables
/k/ Before high vowels
/k/ Before non-high vowels
Frequency (Hz)
8000
6000
4000
2000
Time (norm.)
5
10
15
20

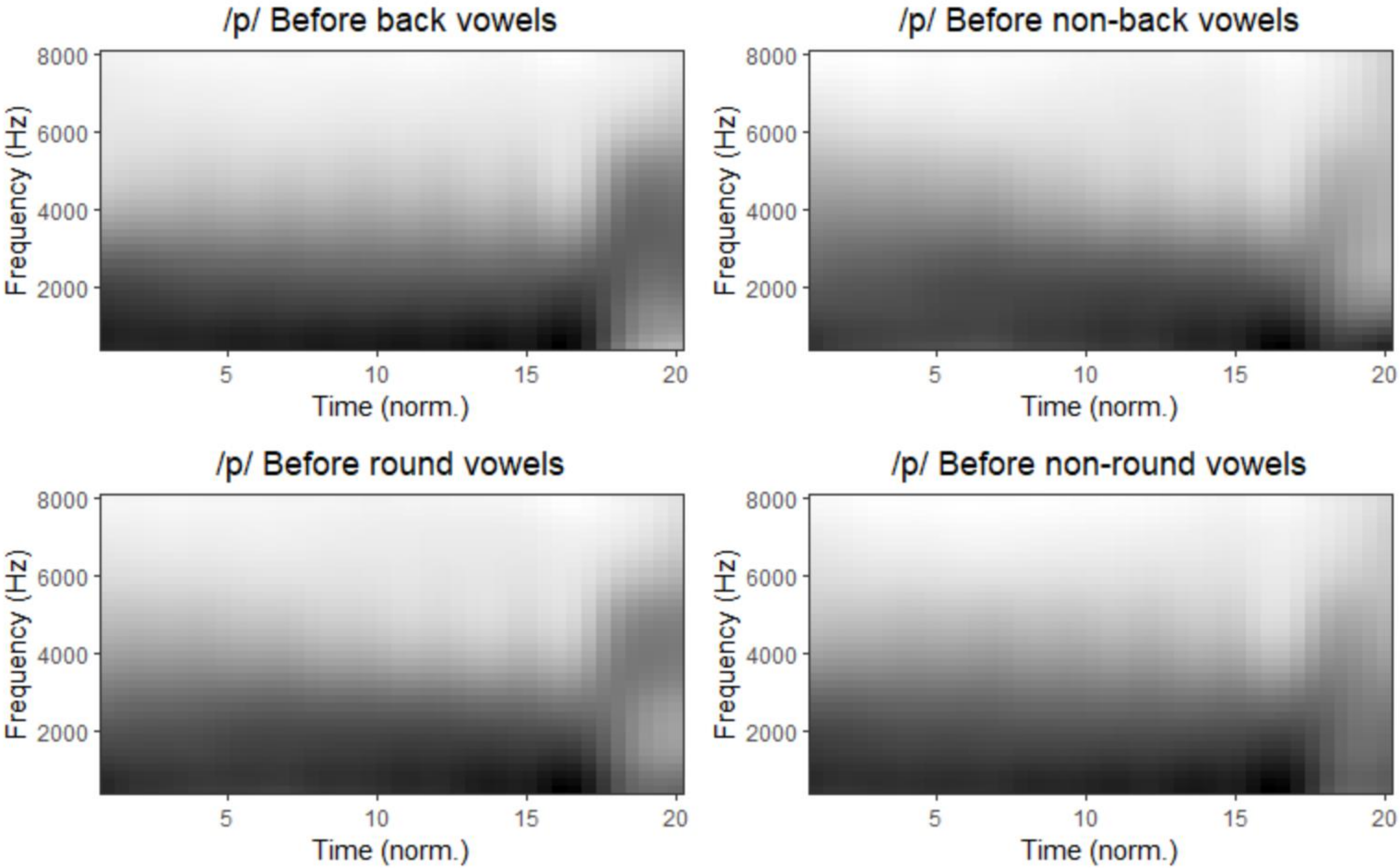

/k/ Before back vowels
/k/ Before non-back vowels
/k/ Before round vowels
/k/ Before non-round vowels
Frequency (Hz)
8000
6000
4000
2000
Time (norm.)
5
10
15
20

Discussion

- /k/
  - Energy distribution highly dependent on phonetic context
  - Consistent with velar noise source throughout half or more of the release before front, non-back, and non-round vowels
  - Mostly consistent with glottal noise source otherwise


/p/ Fitted time-varying spectrum
8000
6000
4000
2000
Frequency (Hz)
5
10
15
20
Time (norm.)
/p/ Men
8000
6000
4000
2000
Frequency (Hz)
5
10
15
20
Time (norm.)
/p/ Women
8000
6000
4000
2000
Frequency (Hz)
5
10
15
20
Time (norm.)


/p/ Stressed syllables
/p/ Unstressed syllables
/p/ Before high vowels
/p/ Before non-high vowels
Frequency (Hz)
8000
6000
4000
2000
Time (norm.)
5
10
15
20


/p/ Before back vowels
/p/ Before non-back vowels
/p/ Before round vowels
/p/ Before non-round vowels
Frequency (Hz)
8000
6000
4000
2000
Time (norm.)
5
10
15
20

Discussion

- /p/
  - Energy distribution during the first half of release dependent on phonetic context
  - Diffuse energy distribution (consistent with labial noise source) throughout the spectrum in unstressed syllables, before high vowels, before back vowels
  - Otherwise consistent with glottal noise source

Discussion

- I'm overall very happy with the results
  - We avoid using measurements that are either problematic or very difficult to interpret
  - The results are easy to grasp if you're used to looking at spectrograms (or at least I think so…)

- **Why** these strange patterns at the end of each release?
- Would ERB be better than Hz?
- Are there other disadvantages to this that I've overlooked?

References

- Basbøll, H. 1969. The phoneme system of Advanced Standard Copenhagen. *ARIPUC* 3, 33–54.
- 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 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, University of Southampton.
- Boersma, P & D Weenink. 2019. *Praat. Doing phonetics by computer*. Version 6.0.55.
- Braunschweiler, N. 1997. Integrated cues of voicing and vowel length in German. A production study. *Language & Speech* 40(4), 353–376. doi:10.1177/002383099704000403.
- Brink, L & J Lund. 1975. *Dansk rigsmål 1. Lydudviklingen siden 1840 med særligt henblik på sociolekterne i København*. Gyldendal.
- Cederbaum, J, M Pouplier, P Hoole & S Greven. 2016. Functional linear mixed models for irregularly or sparsely sampled data. *Statistical Modelling* 16(1), 67–88. doi:10.1177/1471082X15617594.
- Fischer-Jørgensen, E. 1968. Voicing, tenseness and aspiration in stop consonants, with special reference to French and Danish. *ARIPUC* 3, 63–114.
- –. 1972. Kinesthetic judgement of effort in the production of stop consonants. *ARIPUC* 6, 59–73.
- 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.
- Frøkjær-Jensen, B, C Ludvigsen & J Rischel. 1971. A glottographic study of some Danish consonants. In LL Hammerich, R Jakobson & E Zwirner (eds.), *Form and substance*, 123–140. Akademisk Forlag.
- Goldsmith, J, F Scheipl, L Huang, J Wrobel, C Di, J Gellar, J Harezlak, et al. 2021. *refund. Regression with functional data*. R package version 0.1-24.
- 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.
- –. 2005. *Fonetik og fonologi. Almen og dansk*. 3rd ed. Akademisk Forlag.
- –. 2009. A Danish phonetically annotated spontaneous speech corpus (DanPASS). *Speech Communication* 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.
- Hoole, P, B Pompino-Marschall & M. Dames. 1984. Glottal timing in German voiceless occlusives. In MPR Van den Broecke & A Cohen (eds.), *ICPhS 10*, 399–403. Foris. doi:10.1515/9783110884685-059.
- 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.
- Jespersen, O. 1899. *Fonetik. En systematisk fremstilling af læren om sproglyd*. Det Schubotheske Forlag.
- Kokkelmans, J. 2021. *The phonetics and phonology of sibilants. A synchronic and diachronic OT typology of sibilant inventories*. PhD dissertation, University of Verona.
- Lisker, L. 1957. Closure duration and the intervocalic voiced-voiceless distinction in English. *Language* 33(1), 42–49. doi:10.2307/410949.
- Löfqvist, A. 1976. Closure duration and aspiration for Swedish stops. *Lund Phonetics WP* 13, 1–40.
- –. 1980. Interarticulator programming in stop production. *JPhon* 8(4), 475–490. doi:10.1016/S0095-4470(19)31502-5.
- Nance, C & S Kirkham. 2020. The acoustics of three-way lateral and nasal palatalisation contrasts in Scottish Gaelic. *JASA* 147(4), 2858–2872. doi:10.1121/10.0000998.
- Petersen, NR. 1983. The effect of consonant type on fundamental frequency and larynx height in Danish. *ARIPUC* 17, 55–86.
- Pétursson, M. 1976. Aspiration et activité glottale. Examen expérimental à partir de consonnes islandaises. *Phonetica* 33(3), 169–198. doi:10.1159/000259721.
- Puggaard, R. 2021. Modeling regional variation in voice onset time of Jutlandic varieties of Danish. In H Van de Velde, NH Hilton & R Knooihuizen (eds.), *Language Variation - European Perspectives VIII*, 79–110. John Benjamins. doi:10.1075/silv.25.04pug.
- R Core Team. 2020. *R. A language and environment for statistical computing*. Version 4.0.3.
- Reidy, PF. 2016. *spectRum. Compute periodogram and multitaper spectral estimates of waveform data.* Convenience functions for R. https://github.com/patrickreidy/spectRum.
- Sawashima, M. 1970. Glottal adjustments for English obstruents. *Status Report on Speech Research* 21/22, 187–200.
- Schachtenhaufen, R. In prep. Illustrations of the IPA. Danish.
- Sóskuthy, M. 2017. Generalised additive mixed models for dynamic analysis in linguistics. A practical introduction. arXiv:1703.05339v1.
- Stathopoulos, ET & G Weismer. 1983. Closure duration of stop consonants. *JPhon* 11(4), 395–400. doi:10.1016/S0095-4470(19)30838-1.
- Stevens, KN. 1993a. Modelling affricate consonants. *Speech Comm* 13(1/2), 33–43. doi:10.1016/0167-6393(93)90057-R.
- –. 1993b. Models for the production and acoustics of stop consonants. *Speech Comm* 13(3/4), 367–375. doi:10.1016/0167-6393(93)90035-J.
- Vestergaard, T. 1967. Initial and final consonant combinations in Danish monosyllables. *Studia Linguistica* 21(1), 37–66. doi:10.1111/j.1467-9582.1967.tb00547.x.
- Voeten, CC. 2020. *The adoption of sound change. Synchronic and diachronic processing of regional variation in Dutch*. PhD dissertation, Leiden University.
- Wieling, M, S Montemagni, J Nerbonne & RH Baayen. 2014. Lexical differences between Tuscan dialects and Standard Italian. Accounting for geographic and sociodemographic variation using generalized additive mixed modeling. *Language* 90(3), 669–692. doi:10.1353/lan.2014.0064.
- Wood, SN. 2017. *Generalized additive models. An introduction with R*. 2nd ed. CRC Press. doi:10.1201/9781315370279.

[redacted]ak for jeres opmærksomhed!

Universiteit Leiden

AARHUS UNIVERSITY