On the feasibility of the Danish model of intonational transcription: Phonetic evidence from Jutlandic Danish

Anna Bothe Jespersen[1], Pavel Šturm[2], Míša Hejná[1]

[1]Aarhus University, Denmark
[2] Charles University, Czech Republic

anna.jespersen@cc.au.dk

Abstract

Most of our knowledge of Danish f0 variation and intonation is based on the work of Grønnum and colleagues, who developed an a-phonological model in which a series of repeated "default" contours are superpositioned onto an overarching f0 slope. The current paper tests a range of predictions stemming from this model, most importantly the adequacy of analysing f0 modulations as a string of repeated contours differing in range but not in shape. To facilitate comparison with earlier work in the area, our material is based on read speech, 45 speakers of Jutland Danish participated in the experiment. Analyses of f0 in sentences of differing complexity supplied little evidence in favour of the existence of default contours. Instead, our acoustic data revealed an array of f0 shapes associated with various prosodic anchor points, which are influenced in both range and shape by positional context and the presence or absence of focus.

Index Terms: Danish, fundamental frequency, intonation.

1. Introduction

While Autosegmental-Metrical systems for prosodic transcription (e.g. ToBI, [1]) have been introduced in several language contexts, such a model is not yet available for Danish. Instead, intonologists working on this language have almost exclusively made use of a model developed specifically for Danish. The aim of the present paper is to test a number of hypotheses generated by this model, which we refer to as the *Superpositional Model (SM)*. Several aspects of the model's predictions run counter to findings from other languages, and previous evaluation of the SM has primarily been carried out by its creators. It has thus been applied mostly to Copenhagen Danish, and this validation has predominantly relied on read speech, on introspection and on speech produced by phoneticians. In this paper, we aim to examine the SM and test several of its predictions from the perspective of a different demographic area, using non-phonetically trained speakers and a much larger speaker sample.

The work on Danish intonation through which the Superpositional Model has been developed and tested (cf. [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12]; see [13, pp. 530–534] for a review) has assumed a hierarchical ordering of melodic components. Underlying the entire utterance is a global trend for f0 to be falling, which is in part influenced by the physiological constraints of the lungs [14]. The sentence type (yes/no question, declarative, and others) adds another modification to the global f0 trend, with declaratives further lowering the global f0, and syntactically unmarked questions driving f0 towards a more horizontal level. At the other end of

the spectrum, Grønnum's model also incorporates f0 modifications caused by stød, and microprosodic segmental influences, such as the effect of voiced vs. voiceless segments on the f0 of adjacent vowels (see also [15]). In between the global and local levels, the model distinguishes a third main level, which embodies the main melodic variation of the utterance, namely a sequential structure of pitch contours. These are associated with individual stress groups, running from a stressed syllable through any unstressed material up until the next stressed syllable. The stress groups correlate with f0 movements termed "default contours", which are 'recurring and fairly constant f0 patterns', repeated throughout the utterance, and are modified by both global and local pitch trends [5, pp. 131–132]. Within stress groups, f0 *shape* is thus held constant, while f0 *range* is mediated by positional context (Figure 1).

Figure 1. *An illustration of Grønnum's Superpositional Model, from [6, pp. 532]. Line 1 illustrates f0 trajectories with syntactically unmarked questions, lines 2 show various types of questions, and line 3 represents declarative statements.*

These characteristics of the model mean that declaratives, such as those investigated in the current paper, are predicted to consist of several falling contours of decreasing steepness as the global f0 trajectory reaches lower and lower levels (see [11]). In other words, the model predicts a series of similarly shaped contours of gradually decreasing f0 (more precisely, of rise-falls that start to fall after a post-tonic rise). Importantly, it does not predict final rising contours with any sentence types for most dialects of Danish, and especially not with declaratives (though indications that this is possible are offered by [8], [9]). Such a prediction goes against the crosslinguistic tendencies whereby questions are commonly produced with a rising final contour (e.g. [16], [17], [18]), and more and more languages related to Danish are reported to use rising final contours with declaratives [19, pp. 163–169].

The f0 stretch associated with the default contour is furthermore subject to variation caused by prominence.

Grønnum [5, pp. 141–144] notes that emphatic focus enlarges the f0 span of the focused default contour while reducing the f0 and temporal spans of the preceding contours. In other words, Grønnum's model generally predicts that the *shape* of contours will not be influenced by focal prominence, which only affects the *range* of contours.

The default contour is thus supposed to be structurally identical throughout all utterances within a single variety. However, dialectal variation is expected: each variety of Danish is seen to have one default pitch contour, which is unique to the variety. Effectively, then, the model accounts for regional variation as follows: '[A]lthough other regional variants of Danish have the exact same rhythmic patterning within an utterance as Copenhagen Danish, the melodic contour is different for each one' [6, p. 340, translation by the first author]. Figure 2 (reproduced from [6]) provides an illustration of such default contours from four dialects of Danish, including the standard.

Figure 2. Default pitch contours suggested for 4 of the regional varieties of Danish, including the Copenhagen standard. Blue lines represent the accented vowel. Adapted from [6].

In this paper, we will not consider dialectal differences. However, such variation was examined while undertaking the present research, and no major differences have been located. A brief summary of these efforts will be presented in the results section.

While the researchers working within the framework of the SM have modelled Danish intonational variation based to a large extent on the default contour, researchers have not included in their considerations an analysis based on phonological structures within an intonational inventory (cf. Autosegmental-Metrical phonology, [13], or the British school of transcription, [20]). This feature makes comparisons with intonational work on related languages – which have been described using other frameworks – burdensome. Another drawback of this line of work is that acoustic studies of f0 variation in Danish (cf. [2]–[5], [8]–[11]) have almost exclusively focused on Copenhagen Danish, the (Zealand) standard variety. Most non-acoustic work has relied on speech produced by Grønnum and her colleagues, as have some of the acoustic studies. One exception to these points is [4], which relies on auditory analysis of native speakers performed by Grønnum and includes data from a total of 7 speakers of 3 Jutlandic "standard dialects" [4, p. 183]. Thus, to the best of our knowledge there are currently no acoustic studies on f0 variation in Jutlandic Danish, nor any available phonological model of Danish intonation. Furthermore, early work has often relied on read speech and sentences in isolation recorded in laboratory conditions (again excepting [4]).

The present paper speaks to the lack of acoustic descriptions of Jutlandic Danish f0 with an eye to providing an acoustically informed initial exploration of Danish melodic variation. We address this aim by investigating the claim made in earlier work that the Danish speech melody is formed through a series of repeated default contours differing not in shape but in height and span only. In doing so, we examine stretches of phonetic material corresponding to Grønnum's "stress groups", and the f0 contours associated with them. Our overarching research question is the following:

Are there consistent differences in f0 shape between contours associated with particular intonational landmarks?

Specifically, we investigate three comparisons:

a. *Phrase-initial versus phrase-medial pre-nuclear contours.* This distinction lets us investigate Grønnum's claim that contours retain their f0 shape throughout the utterance.

b. *Nuclear contours associated with intermediate versus full intonational phrases (ip/IP).* Here, we tap into the claim that declarative f0 does not rise by examining two contexts in which other languages produce rising contours [21].

c. *Contours associated with unfocused versus focused IP nuclei.* This lets us investigate the claim that in non-truncating contexts, focus affects only the f0 span of contours, not their shape [22], [23].

2. Methods

2.1. Materials

The analysis is based on the Danish translation of the Rainbow Passage [24] provided by the first author and checked by another native speaker of Danish. This passage was recorded by a sample of 60 native Danish speakers for purposes that are not linked with intonation (see [25]). Within the Rainbow Passage, we selected the following 5 sentences, which enabled us to consider various pre-nuclear and nuclear phenomena.

a. *En regnbue er en opbrydning af hvidt lys i mange smukke farver.* 'The rainbow is a division of white light into many beautiful colours.'

b. *Folk kigger efter den, men finder den aldrig.* 'People look, but no one ever finds it.'

c. *Når folk søger noget udenfor egen formåen siger man, at de kigger efter en krukke guld for enden af regnbuen.* 'When a man looks for something beyond his reach, his friends say he is looking for the pot of gold at the end of the rainbow.'

d. *Andre har forsøgt at forklare fænomenet i fysiske termer.* 'Others have tried to explain the phenomenon physically.'

e. *Man har haft mange komplicerede ideer om regnbuen.* 'Many complicated ideas about the rainbow have been formed.'

2.2. Speakers

The speakers represent variation within 5 adjacent Jutlandic dialects areas: South (3 males, 1 female), Mid South (7 males, 2 females), Mid East (9 females, 8 males), Mid West (5 males, 6 females), and North East (2 males, 2 females). As the data was collected for another study, there is some imbalance regarding speaker's self-reported physiological sex for all 5 dialect groups. All the participants are between 18 and 41 years of age (Females: mean = 27, SD = 4; Males: mean = 27,

SD = 6). We do not investigate variation in these factors here, but instead control for such variation by including dialect and physiological sex as fixed effects in our statistical models (see below).

2.3. Data processing

The speakers were recorded in a sound-attenuated booth at the Department of English at Aarhus University, Denmark. The H5 Zoom Handy recorder was combined with a condenser microphone positioned approximately 30 cm in front of the speakers. The recording was stored as a mono wav file with a sampling rate of 44.1 kHz.

F0 measures were extracted from intervals corresponding to f0 contours anchored to pre-nuclear prominent syllables and to nuclei associated with intermediate and full intonational phrases. The intervals thus encompass the accented (tonic) syllable and the following unstressed syllables. Nuclear tunes were defined as tunes starting with the final prominent syllable of an intonational phrase (ip or IP). IP edges were differentiated from ip edges based on a combination of the syntax of the read sentences and the following phonetic criteria. A phrase edge was marked as an IP where (i) pre-final lengthening was present, (ii) phrase-final pauses were present, (iii) two or more syllables were realized with noticeable creak. (Narrowly) focused IP nuclei were distinguished from unfocused nuclei in that the nucleus was not associated with a default reading of the last metrically prominent syllable, but rather with any other syllable in the phrase. Following the focused nucleus, all other material undergoes reductions of e.g. pitch range and duration (for details on narrow focus, see [26, p. 254–256], [27], [28]).

F0 contours associated with pre-nuclear and nuclear anchor points were labelled in Praat [29] and ProsodyPro [30]. A Praat script was used to extract f0 measures at 10 equidistant points throughout the labelled interval from the beginning of the prominent vowel to the end of the f0 offglide as defined by the end of voiced material, or the f0 elbow associated with a new f0 movement. The 10-point measures allow for a time-normalised presentation of f0 curves, which to some extent addresses differences in speech rate across the corpus and differences in stress group length. Raw f0 measures (in Hz) were converted to semitones with separate benchmarks for male and female speakers, as recommended by Laver [31]. In this paper, the benchmarks chosen were 120 Hz for females and 75 Hz for males, in line with Laver's recommendations.

2.4. Analysis and statistics

The statistical analysis was conducted through the fitting of Generalised Additive Mixed Models (GAMMs) via the *mgvc* [32] and *itsadug* [33] packages in *R* [34], using *tidymv* [35] for visualization. GAMMs allow for dynamic analyses of 'measurable quantities of speech that vary in space/time' [36, p. 1], such as f0 changes, through the application of smooth terms, which account for the *shape* of dynamic changes (as contrasted with *height* differences, modelled by parametric terms). Furthermore, GAMMs allow separate random smooths to be fitted to account for the variation in f0 shape introduced by random variables. For this paper, models were fitted using *bam()* and hierarchical model comparisons were carried out using *compareML()* from the *itsadug* package. Significance testing of individual levels relies on the ANOVA-generated p-values of difference smooths, as the analyses are primarily concerned with shape differences, and on visualisations of difference smooths and their confidence intervals in *itsadug*, as described in [37] and recommended in [36, p.p. 19–20].

3. Results

In this section, we focus on the four contrasts of particular interest laid out in the Introduction. To dismiss the null hypothesis, the statistical model needs to find evidence of differences between the shapes of f0 trajectories associated with different tonal categories. Figure 3 shows the time-normalised f0 trajectories associated with each of the 5 tonal categories investigated.

Figure 3. *Time-normalised f0 in stress-groups as a function of tonal category.*

All five tonal categories differ in height as predicted by the SM, with initial pre-nuclear tunes being higher than medial pre-nuclear tunes, and ip nuclei being higher than IP nuclei. However, it is clear that there are also differences in the shapes of each tune's f0 trajectory. Only two tonal categories look to be relatively similar in shape: initial pre-nuclear tunes and ip nuclei.
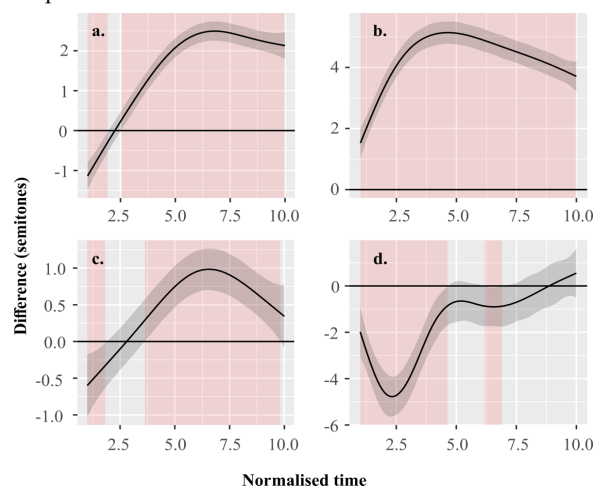

Figure 4. *Plots of difference smooths between a. initial and medial pre-nuclear tunes, b. initial pre-nuclear tunes and ip nuclei, c. ip and IP nuclei, d. focused and unfocused IP nuclei. Red shading = areas of significant difference between comparisons. Grey shading = 95% CIs.*

Figure 4 displays the difference smooths arising from statistical comparison of 4 relevant pairs of tonal categories

(see Table 1). In this graph, red shadings indicate temporal areas of statistical significance, i.e. intervals where the two contours compared in each pane have significantly different shapes. Figure 4a illustrates the difference smooth between initial and medial pre-nuclear tunes. These two tonal types differ at nearly all points along their trajectories ($F=68.3$, ***). Similarly, initial pre-nuclear tunes and ip nuclei (illustrated in Fig. 4b) differ significantly in shape along their entire trajectory ($F=10.1$, ***). Their perceived similarity in Figure 3 thus arises solely from their similar heights. Figure 4c is a comparison between nuclei associated with ips and IPs. These two tunes differ primarily at the ends of their trajectories, and the difference is significant ($F=22.7$, ***). Finally, Figure 4d illustrates the difference smooth between tunes associated with unfocused and focused IP nuclei. In Figure 3, there was some overlap of the two tune types towards the end of their f0 trajectories. The difference (Fig. 4d) between the two tunes is nevertheless significant due to the initial part ($F=11.2$, ***).

Table 1: *GAMM table.*

Smooth	Edf	Ref.df	F	Sig.
Reference (Initial PN)	2.69	3.69	0.08	.99
Shape:Medial Pre-Nuclear	4.56	5.26	69.97	<.001
Shape:Nucleus(ip)	3.11	3.86	10.61	<.001
Shape:Nucleus(IP)	5.54	6.72	52.40	<.001
Shape:Nucleus(Focus)	5.88	7.07	27.70	<.001
Random = Speaker	108.02	385.0	20.96	<.001
Random = Sentence	260.6	276.5	20.97	<.001

R-sq.(adj) = .614; Deviance explained = 62.1%.

Finally, we very briefly illustrate in Figure 5 the relative paucity of dialectal differences in our corpus.

Figure 5. *Plot of time-normalised f0 trajectories by tonal category for five Jutlandic dialect groups.*

None of the dialect groups differed significantly from the rest with respect to pre-nuclear tunes or unfocused IP nuclei. The main differences lie in the ip nuclei and focused IP nuclei. There were significant differences in focused contexts. Here,

three dialect areas yielded steep falling contours (North East, Mid East, South), whereas the two other dialects (Mid West and Mid South) showed a rise-fall with a prominent early peak. This difference in shape was statistically significant (**). Furthermore, the North Eastern speakers displayed a rising tune with ip nuclei (*). We have to keep such differences in mind when discussing nuclear contours, but we argue that the limited variation in shape across the dialects does not impact our findings on the predictions of the SM.

4. Discussion and conclusions

This paper has provided little evidence to support the main prediction of the Superpositional Model (SM), namely that "default" f0 movements associated with so-called stress groups are identical in form throughout the utterance, differing only in height and span [2–12]. Instead, the paper has unveiled that speakers of Jutlandic Danish produce a variety of average f0 shapes associated with stress groups at various metrical and prosodic anchor points. Indeed, while the SM would predict a series of pre-nuclear contours that differ only in f0 height and span, this paper found that pre-nuclear contours at different points in the rhythmic context were systematically different both in range *and* in shape, with differences along almost the entire f0 trajectory. As further evidence for the need for a more varied intonational inventory, the paper has also found systematic differences between contours associated with ip and IP nuclei, and between focused and unfocused IP nuclei. In this way, it is clear that the default contour is not a practicable way of describing pre-nuclear or nuclear f0 stretches. Finally, our data shows that declarative phrases do not always display falls at the global level, contrary to predictions of the SM, and may indeed rise locally (e.g. as shown by rising declarative ip nuclei in the North East data).

Because of its aim to test the predictions of an a-phonological model of intonation, the analyses presented here have dealt with surface f0 variation examined through the lens of movements associated with prosodic anchor points. We have thus not attempted to provide phonologically-based descriptions. Future work can fruitfully tease out and examine systematic phonological variation associated with prosodic anchor points. Furthermore, while our reliance on read speech should impact the amount and type of f0 variation in our data, it does allow us to make a more direct comparison to previous work (e.g. [3], [7]), which is also based on read material. Moreover, Grønnum [3] posits that default pitch patterns are highly engrained in speakers of regional standards and can be found even in laboratory settings. Nevertheless, an extension of our findings to conversational speech would provide a valuable way of exploring stylistic variation in f0 realisation. It would thus allow for stronger conclusions to be made about the phonological realities of Jutlandic intonation, and about adequate ways of modelling Danish and Jutlandic f0 movements.

5. Acknowledgements

6. References

[1] M. E. Beckman, J. Hirschberg and S. Shattuck-Hufnagel, "The original ToBI system and the evolution of the ToBI framework," in *Prosodic Typology - The Phonology of Intonation and Phrasing*, S.-A. Jun, Ed., Oxford: Oxford University Press, 2005, pp. 9–54.

[2] N. Thorsen, "An Acoustical Investigation of Danish Intonation," *J. Phon.*, vol. 6, pp. 151–75, 1978.

[3] N. Grønnum, "Prosodic Parameters in a Variety of Regional Danish Standard Languages, with a View towards Swedish and German," *Phonetica*, vol. 47, pp. 182–214, 1990.

[4] N. Grønnum, "Superposition and subordination in intonation: A non-linear approach," in *Proc. XIIIth Int. Congr. Phon. Sci.*, K. Elenius and P. Branderud, Eds., Stockholm, 1995, vol. 2, pp. 124-131. Available: https://danpass.hum.ku.dk/ng/papers/proc-phon13_1995_124-131.pdf

[5] N. Grønnum, "Intonation in Danish," in *Intonation Systems: A Survey of Twenty Languages*, D. Hirst and A. D. Cristo, Eds., Cambridge, UK: Cambridge University Press, 1998, pp. 131–151.

[6] N. Grønnum, *Fonetik & Fonologi*. Copenhagen: Akademisk Forlag, 2005.

[7] N. Reinholdt Petersen, "Modelling Danish sentence and phrase intonation," in *Proc. XIVth Int. Congr. Phon. Sci.*, J. Ohala, Ed., San Francisco, 1999, pp. 925–928. Available: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS1999/papers/p14_0925.pdf

[8] J. Tøndering, "Intonation contours in Danish spontaneous speech," in *Proc. XVth Int. Congr. Phon. Sci.*, M. J. Solé, D. Recasens and J. Romero, Eds., Barcelona, 2003, pp. 1241–1244.

[9] J. Tøndering, "Intonation og informationsstatus i spontane monologer-belyst ved prominensmålinger," Ph.D. dissertation, Univ. Copenhagen, 2004. Available: https://danpass.hum.ku.dk/links_2014/student%20projects/johnt_speciale.pdf

[10] J. Tøndering, "Preplanning of intonation in spontaneous versus read aloud speech: Evidence from Danish," in *Proc. XVIIth Int. Congr. Phon. Sci.*, Hong Kong, 2011, pp. 2010–2013.

[11] N. Grønnum and J. Tøndering, "Question intonation in non-scripted Danish dialogues," in *Proc. XVIth Int. Congr. Phon. Sci.*, J. Trouvain and W. Barry, Eds., Saarbrücken, Germany, 2007, pp. 1229–1232.

[12] H. Basbøll, *The Phonology of Danish*. Oxford: Oxford University Press, 2005.

[13] J. Pierrehumbert, J, "The Phonetics and Phonology of English Intonation", Ph.D. dissertation, MIT, 1980.

[14] S. Fuchs, C. Petrone, A. Rochet-Capellan, U. D. Reichel and L. L. Koenig, "Assessing respiratory contributions to f0 declination in German across varying speech tasks and respiratory demands," *Journal of Phonetics*, vol. 52, pp. 35–45, 2015.

[15] J. P. Kirby and D. R. Ladd, "Effects of obstruent voicing on vowel F0: evidence from 'true voicing' languages," J. Acoust. Soc. Am., vol. 140, pp. 2400–2411, 2016.

[16] E. L. Asu, "Downtrends in different types of question in Estonian," in *Proc. Speech Prosody*, Aix-en-Provence, 2002, pp. 143–146.

[17] N. Hedberg and J. M. Sosa, "The phonetics of final pitch accents in American English polar questions," in *Proc. XVIIth Int. Congr. Phon. Sci.*, Hong Kong, 2011, pp. 847–850.

[18] J. Michalsky, "Phonetic effects of speaking style on final rises in German questions and statements," in *Proc. XVIIIth Int. Congr. Phon. Sci.*, Glasgow, United Kingdom, 2015, paper 719.

[19] P. Warren, *Uptalk: The Phenomenon of Rising Intonation*. Cambridge: Cambridge University Press, 2016.

[20] F. Nolan, "Intonation", in *Handbook of English Linguistics*, B. Aarts and A. McMahon, Eds., Oxford: Blackwell, 2006, pp. 433–456.

[21] A. Jespersen, "Acoustic, pragmatic and sociolinguistic cues to the distinction between uptalk and other declarative rises in Australian English," *J. Int. Phon. Ass.*, submitted.

[22] M. Baltazani and S.-A. Jun, "Focus and topic intonation in Greek," in *Proc. XIVth Int. Congr. Phon. Sci.*, J. Ohala, Ed., San Francisco, 1999, pp. 1305−1308.

[23] Y. Xu and C. X. Xu, "Phonetic realization of focus in English declarative intonation," *Journal of Phonetics*, vol. 33, pp. 159−197, 2005.

[24] G. Fairbanks, *Voice and Articulation Drill Book*. NY: Harper & Row, 1960.

[25] M., Hejná, P. Šturm, L. Tylečková and T. Bořil, "Normophonic breathiness in Czech and Danish: Are females breathier than males?" *Journal of Voice*, vol. 35, pp. 498.e1−498.e22, 2021.

[26] D. R. Ladd, *Intonational Phonology*. Cambridge: Cambridge University Press, 1996.

[27] A. Jespersen and M. Hejná, "Rødgrød med fløde? Hvad kæmper andetsprogstalere af dansk med, og hvordan imødegås problemerne af sprogskolerne?" *17. Møde om Udforskningen af Dansk Sprog Rapporter*, Aarhus, Denmark, 2019, pp. 311−327.

[28] E. Norcliffe and T. F. Jaeger, "Accent-free prosodic phrases? Accents and phrasing in the post-nuclear domain," in *Proc. Interspeech 2005*, Lisbon, Portugal, 2005.

[29] *Praat: doing phonetics by computer.* (Version 6.0.47). P. Boersma and D. Weenink. Accessed: Feb. 8, 2019. Available: http://www.praat.org/

[30] Y. Xu, "ProsodyPro: A Tool for Large-scale Systematic Prosody Analysis" in *Proc. Tools Resources Analysis of Speech Prosody (TRASP 2013)*, Aix-en-Provence, France, 2013, pp. 7−10.

[31] J. Laver, *Principles of Phonetics*. Cambridge: Cambridge University Press, 1994.

[32] *Package 'mgcv'.* (2018). S. Wood. Available: https://cran.r-project.org/web/packages/mgcv/mgcv.pdf

[33] *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs.* (2019). J. van Rij, M. Wieling, H. Baayen and H. van Rijn. Available: https://cran.r-project.org/web/packages/itsadug/index.html

[34] *R: A language and environment for statistical computing.* (2018). R Core Team, R Foundation for Statistical Computing, Vienna. Available: https://www.R-project.org/

[35] *tidymv: Plotting for generalized additive models.* (2017). S. Coretta. Available: https://github.com/stefanocoretta/tidymv

[36] M. Sóskuthy, "Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction," *arXiv,* 1703.05339, 2017.

[37] S. Coretta, "An Introduction to GA(M)Ms," 2019. Available: https://github.com/stefanocoretta/gamm-workshop