Speech Task and Prosodic Context Influence Glottal Stop Variation in Tahitian

Janet Fletcher, Adele Gregory

School of Languages and Linguistics, University of Melbourne
janetf@unimelb.edu.au;adele.gregory@unimelb.edu.au

Abstract

It is well accepted that phonemic /ʔ/ can have multiple variants ranging from a canonical voiceless stop with full glottal occlusion to more vowel-like non-modal voiced variants. This study examines acoustic correlates of glottal stop variation in Tahitian. While it has a number of glottal variants like closely related Hawaiian, it differs significantly in that the major allophone in Tahitian is a fully or partially occluded glottal stop. While like Hawaiian more vowel-like variants are realised in less-constrained discourse, more voiceless variants and lower levels of Harmonics-to-Noise ratio are observed in word initial- versus word medial contexts.

Index Terms: glottal stop, Tahitian, duration, Harmonics-to-Noise ratio

1. Introduction

Tahitian (Reo Tahiti) is the dominant indigenous language of French Polynesia. It is part of the Eastern Polynesian sub-group of Polynesian languages [1]. It is estimated that between 60-70% of the population of French Polynesia speak, read, or write Tahitian or one of the other six indigenous languages of the archipelago. Tahitian is the most widespread in terms of daily usage although undergoing a process of reduced transmission from older to younger generations [2].

Table 1. *Phoneme inventory of Tahitian*

Consonants

	labial	*dentialveolar*	*Glottal*
Stops	p	t	ʔ
Nasals	m	n	
Fricatives	f v		h
Liquids		r	

Vowels

	front		back
High	i i:		u u: (+Sequences)
Mid	e e:		o o:
Low		a a:	

The sound system of Tahitian is often described as "simple" (e.g. [3], [4], [5]) because of the small number of contrastive consonant and vowel phonemes relative to French or English, for example. Table 1 illustrates the consonant and vowel inventories of Tahitian (after [4]). The inventory consists of nine consonant phonemes with a simple stop series /p/ /t/ /ʔ/, five primary vowels that contrast in quantity and a number of vowel sequences/diphthongs. It is widely claimed that Oceanic languages like Tahitian have undergone a high degree of "phonetic erosion" over time (e.g. [6]). One key example that is often cited is the case of the glottal stop /ʔ/ which is an important contrastive phoneme in Tahitian and the other indigenous languages of French Polynesia. To date, there has been very limited experimental phonetic investigation of any aspect of spoken Tahitian and this paper focuses on variation in the acoustic realisation of the glottal stop /ʔ/.

It is commonly accepted that whilst complete closure of the glottis is the textbook realisation of a glottal stop, there can be variable realisations within and across the world's languages [7, 8]. Glottal stop variants range from a period of complete glottal occlusion to a period of creaky voice or even breathy voice along a continuum. It is also well attested that prosodic conditioning can also influence glottal stop realisation with more fully occluded or gesturally strengthened variants in Intonational Phrase-initial or word-initial positions [9,10]. In a large-scale survey of 131 Illustrations of the International Phonetic Alphabet published in the Journal of the International Phonetic Association, [8] showed that acoustic measures of glottal state including percentage of voicing and voicing intensity using strength of excitation measures (SoE) were effective in quantifying differences among glottal stop variants with more vowel-like variants showing higher % voicing and stronger levels of voicing compared to glottal stops. Findings were less clear for phrase or word position however. Similarly, in Arapaho, [11] found that harmonics-to-noise ratio (HNR) was a useful indicator of different glottal stop variants with more modal phonation showing a higher ratio compared to more creaky, laryngealised phonation. Fully occluded glottal stops show similar levels of HNR to voiceless /t/ in Arapaho and are always acoustically longer than other variants.

In Hawaiian, another Eastern Polynesian language, an investigation of glottal stop variation in naturalistic speech data showed that relatively few /ʔ/ are produced with full glottal closure (7%) with most variants produced with creaky voice variants[12]. These include a variant that consists of a period of creaky voice flanked by two modal vowels: 'mid-creak', and a variant 'whole creak' where an entire VʔV sequence is produced with creaky voice. Other more vowel-like realisations of /ʔ/ include a so-called 'intensity dip' variant that is a period of modal phonation of reduced intensity compared to surrounding modal vowels. Two further 'vocalic' variants noted by [12] are modal voice and breathy voiced variants. By contrast word-initial /ʔ/ is more likely to be produced with at least partial if not complete glottal closure suggesting a degree of prosodic strengthening in this location. Similar to Arapaho, however, full glottal closure variants tend to be somewhat longer than other more vowel-like variants in Hawaiian.

As yet, there has been no focused experimental study of glottal stop variation in Tahitian. One study [3] suggests that glottal stops are always voiceless but also mentions a variant that may well be similar to the so-called intensity dip noted by [12] for Hawaiian. [3] suggests that more detailed phonetic analyses are required. [13] also discusses laryngealised variants but concludes that variation is somewhat unpredictable whereas [4] claims that modal and more creaky voice realisations are likely when flanking vowels are identical in a VʔV sequence. In view of this, and also findings for the closely related language Hawaiian [12] and other languages that have a phonemic glottal stop [8][11], it is highly likely that Tahitian exhibits phonetic variation in /ʔ/ realisation. It remains to be seen whether prosodic context is a relevant factor in determining whether there are more voiceless, fully occluded variants in prosodic constituent-initial position, for example. If prosodically conditioned variation is present, we predict that more controlled laboratory phonology-type tasks where narrow focus is invariably realised on specific experimental tokens of interest may also result in more fully-occluded variants compared to connected speech [12]. In fact we would predict that the proportion of full closure glottal stops to more "vocalic" creaky voice variants in connected speech tasks will be similar to other studies that have focused on more naturalistic discourse.

2. Methodology

2.1. Speakers and materials

Speech data were obtained from five female speakers aged between 18 and 50 at the time of the recordings. All participants were born in French Polynesia and are L1 speakers of Tahitian along with Tahitian French. The recording materials consisted of two elicited reading tasks. The first set of materials (Experiment 1, henceforth EXP1) were designed to illustrate all contrastive stops and vowels, and consisted of two disyllabic words written in Tahitian orthography inserted in a carrier phrase. Ninety words were included in the study but only tokens containing the glottal stop or oral stops will be examined in this analysis. An example of the EXP1 carrier phrase is shown in (1). Tahitian orthography uses ‘ to indicate a glottal stop or *‘eta*.

(1) I roto i te reo tahiti, e parauhia ***pata*** e'ere ***pa'a***.
"In Tahitian we say **scorpion** and not **bark**."

The second experimental task consisted of a reading and retelling of the Aesop fable La Bise et Le Soleil (the north wind and the sun), henceforth BES, following the materials used in the Atlas sonore des langues régionales de France [14]. All materials were presented by PowerPoint.

The speakers were recorded in Papeete in a quiet room at the Université de la Polynésie Française in November 2018 using a Zoom H6 recorder through a Countryman ISOMAX head-mounted microphone. Recordings had a sampling rate of 44.1kHz and 16 bit quantisation. Speakers produced four repetitions of each token in EXP1 and two or three repetitions of the fable in the BES corpus. Table 2 shows the distribution of glottal and oral stops across the two experimental tasks investigated in this study.

Table 2. *Number and distribution of /ʔ/ and combined oral stops /p/ /t/ according to experimental task and word position*

	EXP1		BES	
	Initial	Medial	Initial	Medial
/ʔ/	377	701	302	306
Oral stops	1502	1057	484	209

2.2. Data processing and annotation procedures

The audio files were forced aligned using a version of the general WebMaus protocol [15] that has been adapted for French Polynesian language inventories, and vowel and consonant boundaries were subsequently checked and hand corrected using the waveform and spectrogram [16]. Additional annotation of both datasets was undertaken to identify the different types of glottal stop variants adapting the fine-grained annotation procedures in [12] for Hawaiian. Glottal stop variants were assigned to different categories ranging from the most consonantal to least consonantal (i.e. more vocalic) as follows:

Full glottal closure (FGC)
Mid creak (creaky phonation during most of the C interval)
Whole creak (creaky phonation across a VCV sequence)
Intensity dip (reduced intensity relative to surrounding Vs)
Breathy phonation
Modal phonation

Where there was a period of voicelessness with some pulses at either the beginning or end of the closure interval, these were identified as Full Glottal Closure (FGC). In many cases a portion of the following vowel was also included if there was a period of creaky voice at the beginning of the vowel. Two examples of /ʔ/ with complete glottal closure (i.e. canonical glottal stops) are shown in Figure 1. Cases of /ʔ/ realised as non-modal phonation (mid creak, whole creak, breathy voice) were also visually identified from waveforms and spectrograms. Cases of creaky phonation showed irregular and damped glottal pulses and breathy voice showed high frequency noise components (after [7]).

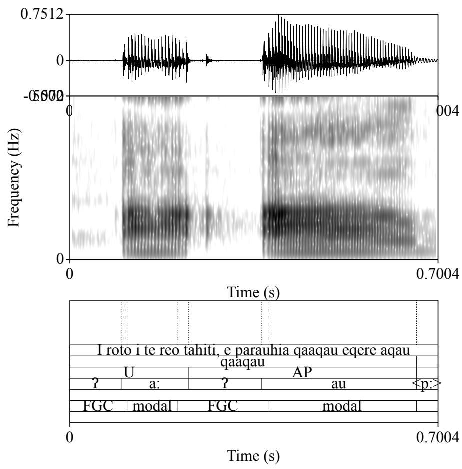

Figure 1: *Waveform and spectrogram of the token /ʔaa'ʔau/ 'conscience/soul' showing two examples of /ʔ/ produced with full glottal closure (FGC) in the EXP1 task.*

Mid creak variants were noted if /ʔ/ was produced with a period of intermittent creaky voice with no real evidence of full glottal closure. Whole creak realisations showed creaky voice across the entire VʔV sequence with no evidence of devoicing. Instances of whole creak were more likely to occur when the flanking vowels were identical. Intensity dip realisations were labelled if /ʔ/ was realised as a period of modal phonation with reduced amplitude. Initial annotations were conducted by the second author and checked by the first author.

2.3. Acoustic analysis procedures

Duration values were extracted for consonant phonemes (separately for glottal and oral stops) and for each annotated glottal variant using the emu-sdms system in R [16]. The oral stop durational values were calculated to enable comparison with glottal stop duration. Similar to [11], we included closure interval plus post-release phases in the overall C measures for voiceless stops. For this reason, only consonants that were not preceded by a pause were included in this analysis. Durational values for each variant type were also calculated after [11,12].

Following [10], VoiceSauce [17] implemented in Matlab was used to extract a number of voice quality parameters including harmonics-to-noise ratio < 1500 Hz (HNR15) which is a measure of inharmonic noise. Creaky phonation is generally associated with lower ratio values compared to modal phonation. Values were extracted for all labelled /ʔ/ and /p, t/ across the two experimental tasks (EXP1 and BES). Other parameters were also extracted including spectral tilt measures (H1*-H2*), and cepstral peak prominence (CPP) but these will be investigated in a forthcoming study.

2.4. Statistical Analysis

The duration data were analysed with linear mixed effects models (LMM) using lmerTest [18] in R [19] with fixed effects consonant category (oral, glottal) and experimental task (EXP1, BES) plus interactions. Preliminary statistical analysis showed that prosodic prominence was not significant so this was not included in the final model. Random effects were included for speaker and word. The voicing measure (HNR15) was submitted to a generalised additive mixed model (GAMM) using mgcv [20] and itsadug [21] in R. The model included parametric difference terms for experimental task and position, smooths over normalised glottal stop duration for position and experimental task, and random smooths by participant and experimental task and word and experimental task. Basis terms were set at k=11 and an AR1 error term was also included in the model to take into account autocorrelation, after [22].

3. Results

3.1. Distribution of glottal variants

Figures 2 and 3 show the distribution of major glottal stop variants plotted separately for experimental task. In this section descriptive rather than inferential statistics are presented to allow comparison with similar studies [8,11,12]. In both experimental tasks, the dominant variant of /ʔ/ has full glottal closure with 77% of all variants in EXP1 (Figure 2) and 59% in the BES corpus (Figure 3) so experimental task is an important factor determining the occurrence of canonical glottal stops. A higher proportion of creaky voiced (i.e. mid creak, whole creak) and other variants (breathy and modal) are realised in the BES task compared to the more controlled EXP1 task. In EXP1, 96% of word-initial variants are canonical glottal stops compared to 65% in word-medial position. It should also be pointed out that there are more than twice as many word-medial glottal variants in EXP1 compared to initial /ʔ/ overall as shown in Table 2. In the BES task, around 66% of word initial glottal stops are produced with full glottal closure with 50% of medial /ʔ/ realised as canonical glottal stops. While creaky voiced variants (i.e. mid creak, whole creak) are produced in medial contexts in both tasks, more whole creak variants (i.e. with creaky voice realised across the entire VCV sequence) and modal voiced realisations are observed in medial position in the BES corpus. Only a handful are produced by the same speakers in EXP1. There are more identical vowel VʔV sequences in the BES corpus which is a prime location for either whole creak or modal realisations according to [4]. Moreover most medial modal realisations in EXP1 are found in the minimal pair /roʔa/ *ro'a* "heartwood" and /roːʔaː/ *rō'ā* "shrub used for fishing lines" which is produced by two speakers without any glottalisation. There are only a handful of intensity dip and modal or breathy variants of /ʔ/ produced by speakers compared to full glottal closure variants.

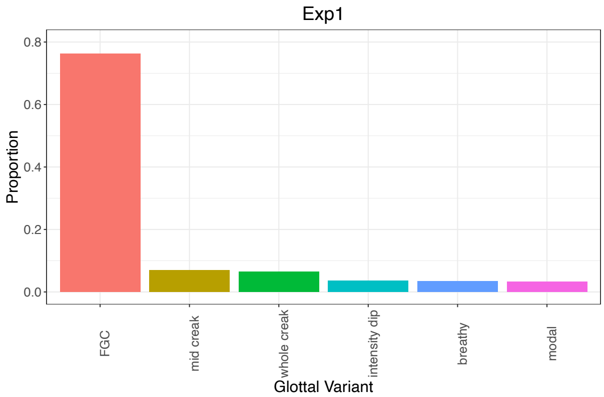

Figure 2: *Distribution of glottal variants of /ʔ/ in Experiment 1*

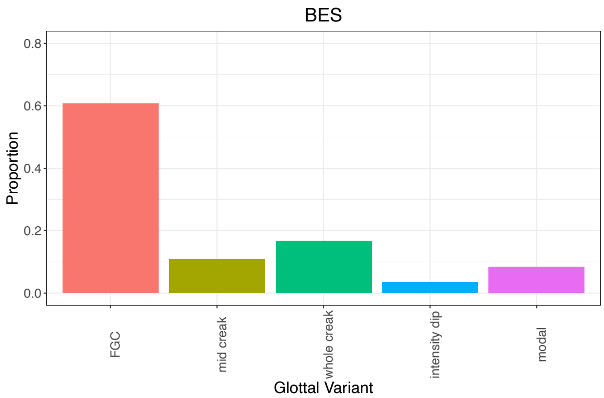

Figure 3: *Distribution of glottal variants of /ʔ/ in the BES corpus*

3.2. Duration

Figure 4 shows a boxplot summarising the simplified distribution of acoustic duration values for all phonemic glottal stops /ʔ/ compared to the combined oral stops /p,t/. There were significant effects of stop type (oral versus glottal) and experimental task on consonant duration, with no interaction between the two factors (consonant type: F=133.28 $p<0.0001$; task: F= 4.39, $p<0.05$). Oral stops are significantly longer than glottal stops in general by around 60 ms. Stops are also generally longer in the controlled experimental task compared to the read task.

Figure 5 plots the durational distribution of the main variants of /ʔ/ in both experimental tasks collapsed across word position for ease of visualisation. Recalling the skewed distribution of full glottal closure (FGC) variants across the two tasks from 3.1, only descriptive statistics are presented here. For the most part, variants are longer in EXP1 compared to the BES corpus. The dominant FGC category is the longest variant with a mean duration value of 122 ms in EXP1 and 61 ms in BES. Mid creak variants are also much shorter in the BES corpus compared to EXP1 (58 ms versus 98 ms). In sum, creaky voice variants are shorter than more "consonantal" variants overall. As noted above, the plot also reflects the important differences between the two tasks noted previously with EXP1 showing some variants with modal phonation in particular lexical items, whereas BES contains more variants with whole creak realisations given the higher instance of VʔV sequences where the flanking vowels are identical.

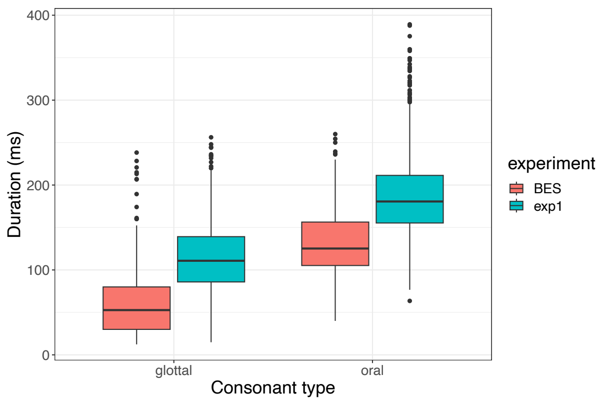

Figure 4: *Glottal stop /ʔ/ and combined oral stop /p,t/ duration plotted by experimental task and word position*

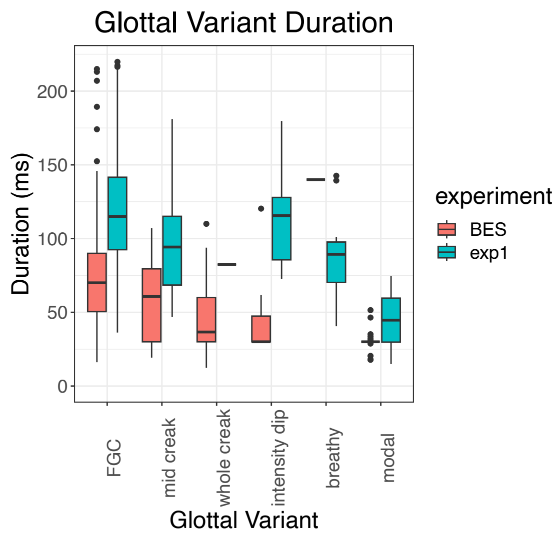

Figure 5: *Acoustic duration of /ʔ/ variants plotted according to experimental task.*

3.3. Voice quality measure: HNR

The time course of harmonics-to-noise ratio is shown in Figure 6 for FGC realisations given this was the main /ʔ/ variant across both tasks. Experimental task was a strong predictor of HNR level with lower HNR values observed in the more constrained experimental task reflecting higher levels of full or partial glottal closure and different trajectory shape (parametric: t=-5.88, p<0.0001, non-linear:F=5.8, p<0.0001). Position was also significant (parametric:t=5.45 (non-linear: F=9.2, p<0.0001) with medial /ʔ/ having higher overall HNR values and a different HNR trajectory shape than initial /ʔ/ as shown in Figure 6.

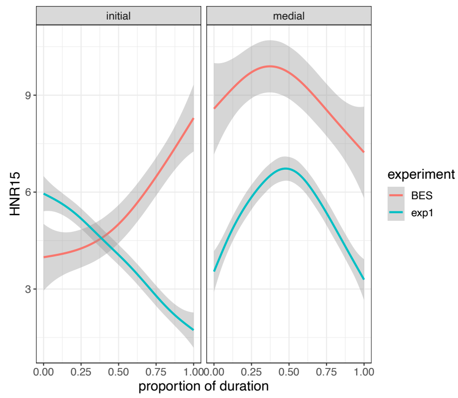

Figure 6: *Loess-smoothed plot of the time course of harmonics-to-noise ratio (HNR) for all /ʔ/ realised with full glottal closure*

4. Discussion

As predicted, /ʔ/ has a number of different glottal variants in Tahitian echoing findings for a range of other languages of the world which also have phonemic glottal stops [8]. Moreover there are different distributions of variants according to experimental task. More canonical glottal stops are produced in the controlled token experiments as a proportion of all glottal variants. Many /ʔ/ variants in the BES corpus, whilst showing a period of voicelessness (i.e. FGC variants) often also show partial closure reinforcing the highly gradient nature of glottal stop production [8,9,11,12]. Compared to Hawaiian, the proportion of canonical glottal stops in both tasks is somewhat higher, with 59% in our BES corpus and 77% in the more controlled experiment compared to 7% reported in [12].

In general, glottal stops in Tahitian tend to be more "consonantal" rather than "vocalic". Even mid-creak realisations tend to be more "consonantal" compared to Hawaiian mid-creak variants [2] particularly in the more controlled task. In terms of acoustic duration, oral stops are longer than glottal stops and both glottal and oral stops are longer in the more controlled experimental task which is not particularly surprising given well documented task and contextual effects on acoustic duration of segments[23]. FGC variants tend to be longer than other variants, particularly in the controlled task, suggesting gestural strengthening, and there are relatively low harmonic-to-noise ratio values similar to the Arapaho canonical /ʔ/. High proportions of true glottal stops are also observed in initial and medial position in the controlled task with the less constrained task showing more non-canonical realisations particularly in medial position, as observed by [8,12]. More vowel-like whole creak and modal variants are observed when /ʔ/ is flanked by identical vowels similar to Hawaiian [12] and as observed by [13].

In sum, there is a degree of expected gradience in the realisation of /ʔ/ in Tahitian together with task-specific inter- and intra-speaker variation (e.g. see 3.1) but at somewhat different levels compared to studies of other languages. Future analyses will examine this further and will include other measures of voice quality including spectral tilt, glottal strength of excitation, and cepstral peak prominence.

5. Acknowledgements

Aspects of this research were supported by the Australian Research Council Centre of Excellence for the Dynamics of Language CE1401000 and a mission de recherche award from the Université de la Polynésie Française to the first author. Particular thanks are expressed to Jacques Vernaudon and Mirose Paia for assistance with subject recruitment in Papeete. Ben Volchok and Eleanor Lewis also contributed to material preparation and data processing of earlier versions of the corpus.

6. References

[1] Walworth, M. “Eastern Polynesia”, in E. Adamou and Y. Matras. The Routledge Handbook of Language Contact, pp 1-17. 2020.
[2] Paia, M. and J. Vernaudon.. “Le Tahitien : plus de prestige, moins de locuteurs” *Hermès, La Revue*, 2002/1, no 32-33, p. 395-402. 2002.
[3] Bodin, V. “Tahiti:la langue et la société”. Éditions ‘Ura. 2006.
[4] Lemaître, Y. “Lexique du Tahitien contemporain”. Paris:Éditions Orstom 1995.
[5] Perini,A-D. “La phonologie du tahitien et l'intégration des emprunts”. *La Linguistique*, 23 (2), 131-142, 1987.
[6] Blust, R. “*t to k: An Austronesian Sound Change Revisited”. *Oceanic Linguistics* 43 (2), 365-410. 2004.
[7] Ladefoged, P. and Maddieson, I. “Sounds of the world’s languages”. Oxford: Wiley-Blackwell. 1987
[8] Garellek , M, Chai, Y., Huang, Y., and van Doren, M. “Voicing of glottal consonants and non-modal vowels. *Journal of the International Phonetic Association* 53, 305-332. 2021.
[9] Kohler, K. “Glottal stops and glottalization in German”. *Phonetica*, 51, 38-51.1994
[10] Garellek, M. “Voice quality strengthening and glottalization” *Journal of Phonetics 45*(1): 106–113.2014.
[11] Whalen, D., DiCanio, C. and Chen, W-R. “Variable realization of the Arapaho glottal stop despite its being distinctive and frequent”. Proceedings of ICPHS2023, 3276-3280. 2023.
[12] Davidson, L. “Effects of word position and flanking vowel on the implementation of glottal stop: Evidence from Hawaiian”. *Journal of Phonetics* 88, 1-14. 2021.
[13] Lemaître, Yves. “La Phonologie du Tahitien”. Unpublished PhD Manuscript .1971
[14] Boula de Mareüil, P. Vernier F., Adda G., Rilliard, A, Vernaudon, J. “A speaking atlas of indigenous languages of France and its Overseas Territories”, International Conference Language Technologies for All (LT4All), Paris (pp. 155–159). 2019.
[15] Kisler, T., Reichel, U., Schiel, F. “Multilingual processing of speech via web services”. *Computer, Speech, and Language*, 45, 326-347. 2017.
[16] Winkelmann, R., Harrington, J., Jänsch, K. “EMU-SDMS: Advanced speech database management and analysis in R. *Computer, Speech and Language* 45, 392-410. 2017.
[17] Shue, Y.-L., P. Keating , C. Vicenik, K. Yu. “VoiceSauce: A program for voice analysis”, Proceedings of the ICPhS XVII, 1846-1849. 2011.
[18] Kuznetsova A, Brockhoff P.B., Christensen R.H.B.“lmerTest Package: Tests in Linear Mixed Effects Models.” Journal of Statistical Software, 82(13), 1-26. 2017
[19] R Core Team. “R: A language and environment for statistical computing”. R. Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org. 2021.
[20] Wood, S. N. “ Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models.” *J Royal Statistical Society*. 73, 3-36. 2011.
[21] van Rij, J., Wieling, M., Baayen, R.H., van Rinj, H. “Itsadug: interpreting time series and autocorrelated data using GAMMS”. https:/CRAN-R-project.org/package=itsadug. 2022.
[22] Sóskuthy, M. “Evaluating generalised additive mixed modelling strategies for dynamic speech analysis”. *Journal of Phonetics*, 84 https://doi.org/10.1016/j.wocn.2020.101017.2021
[23] Fletcher, J. “The prosody of speech: timing and rhythm”. Handbook of the Phonetic Sciences, 2nd edition, Cambridge, Mass: Wiley-Blackwell, 523-602. 2010