The many shapes of H*

Stella Gryllia, Amalia Arvaniti, Cong Zhang, Katherine Marcoux

Radboud University, Netherlands

stella.gryllia@ru.nl, amalia.arvaniti@ru.nl, cong.zhang@ru.nl, katherine.marcoux@ru.nl

Abstract

We examined individual and task-related variability in the realization of Greek nuclear H* followed by L-L% edge tones. The accents (N = 748) were elicited from native speakers of Greek, producing scripted and unscripted speech, and examined using functional Principal Components Analysis. The accented vowel onset was used for landmark registration to capture accent shape and the alignment of the fall. The resulting PCs were analysed using LMEMs (fixed factors: speaker; task type (scripted, unscripted); accented syllable distance from the analysis window offset, to examine the effects of tonal crowding). Tonal scaling and the steepness of the fall (reflected in PC1 and PC2 respectively) changed by task in ways that differed across speakers. PC3, which captured accent shape, also varied by speaker, reflecting shape differences between a rise-fall and (the expected) plateau-plus-fall realization. Tonal crowding did not have consistent effects. In short, the overall accent shape and the alignment of the accentual fall varied by speaker and task. These results hint at substantial variability in tonal realization. At the same time, they indicate that tonal alignment is not as consistent as is sometimes portrayed and thus it should not be the sole criterion for tone categorization.

Index Terms: pitch accent variability, Greek, functional principal component analysis, tonal alignment, tonal crowding

1. Introduction

In intonation research, especially within the autosegmental-metrical theory of intonational phonology (henceforth AM [1], [2]), the scaling and alignment of tones is considered critical for the categorization of intonational events. Specifically, the tonal targets of a tune are said to have stable scaling, i.e., a relatively invariable value in the frequency scale of choice, and stable alignment, i.e., a stable timing relationship with the segmental material. This timing relationship between targets and segments is often referred to as *segmental anchoring* [3]. This idea was originally supported by the findings of [4] who reported that rising accents in Greek, analysed as L*+H in their work, showed a F0 dip at the onset of the accented syllable while the peak of the subsequent rise did not occur until the onset of the first postaccentual vowel. This relationship between the F0 minimum and maximum associated with the L*+H accent was constant in the data of [4]. Similarly stable tonal alignment has been reported for a number of languages (see [5] for a review).

However, as noted in [6], this consistency across languages may be due to the fact that materials in intonation research tend to be relatively uniform and often scripted: as the Romani data of [6] indicate, stable alignment is not as reliably observed in spontaneous speech. Indeed, there are several studies which show that there is more variability in tonal alignment than anticipated based on the tonal anchoring hypothesis (e.g., [7] on Serbian and Croatian, [8] on Manchego Spanish, [9] on German, [10] on Italian and German). These findings have led researchers to explore alternative methods of capturing phonetic differences between accents, such Tonal Center of Gravity [11] and curve analysis (e.g., [12] on Greek, [13] on Samoan). Despite the increasing popularity of these new methods, tonal target scaling and alignment are still widely used (e.g., [14] on Spanish, [15] on English). Thus, documenting variability in tonal realization and the repercussions it has for traditional measurements is worth considering further. To this end, we examined the effects of speaking style and individual variation on H* in Greek scripted and unscripted speech.

H* is part of the Greek inventory of accents [12], [16], [17] and appears in nuclear position in declaratives with broad focus, i.e., on utterance-final words. In utterances with more than one content word, H* is preceded by L*+H prenuclear accents, such that the last prenuclear accent and the final H* form a plateau that is slightly declining [12], [16]. Although this is the typical realization of Greek H*, it is not clear from the existing literature whether H* is variable and affected by speaking style, since most available data are scripted and uniform. However, [12] report that even in such data, the realization of H* can be influenced by tonal crowding – specifically the location of the accent on the final or penultimate syllable of the utterance – which lowers scaling and brings the fall forward in time. This suggests that variability is present and worth investigating further, particularly because H* is in contrast in Greek with two other accents, H*+L and L+H* [12], [16]. Thus, variability in its realization can be problematic both for speakers, who need to differentiate the accents in conversation, and researchers, who need to do so during analysis.

2. Methods

2.1. Participants

Eight Greek speakers (4F, 4M), 22–27 years old (mean = 24; S.D. = 1.7) took part. Three speakers were from Athens, and the other five from Patras and Aigio in Northern Peloponnese; the accent differences between Athens and Peloponnese are minimal. Participant AP01 was bilingual in Greek (her dominant language) and (heritage) Albanian; she was brought up in Greece and was included as her data did not deviate from the average more than those of the other participants. As is the norm in Greece, all participants spoke English and other European languages learned via formal instruction (Italian, N = 3; French, N = 3; Spanish, N = 2; German, N = 1). None reported any speech or hearing disorders.

2.2. Materials

The materials on which the present study is based form part of a larger recording session that included tasks not reported here.

The scripted data were of two types: (a) short dialogues which the participants read together with the experimenter (see (1) for an example); (b) a fable and a news item. The fable (*You can't please everyone*) was 271 words long; it was translated into Greek by the second author from an English version of the story. The news item was 253 words long and published on 29/02/2020 on https://www.lifo.gr, a Greek online news outlet; it was edited to simplify sentence structure and length and thus facilitate reading.

(1) – Πώς *τη λένε για να την ψάξω στο ίντερνετ;*
[ˈpos ti ˈlene ˈʝa na ti ˈbzakso sto ˈinternet]
"What's her name so I can look her up online?"
– *Ραλλού Ζέρβα.*
[raˈlu ˈzerva]
"Ralou Zerva" [FN LN]

The participants also produced two types of unscripted monologues: (a) they retold the fable and news item to the experimenter from memory; (b) they used a story-telling app ("Story Dice" available on Android and iPhone), which consisted in throwing virtual dice with icons on them (six dice in the reported task) and improvising a story that connected the items depicted on the dice.

2.3. Procedure

The study took place during COVID-19 restrictions. The recordings were made using the experimenter's computer and the participants' phones in conjunction with the AVR application which is available for both Android devices and iPhones. Previous studies (e.g., [18]) have shown that this recording method does not pose problems for F0 analysis, provided the audio is saved in a non-lossy format. The files used here were saved as mono wav files with a sampling rate of 44.1 kHz and 256 kbps. Due to experimenter error, some files were saved only as lossy m4a files and were discarded.

The recordings took place in quite conditions in either the experimenter's or participant's home. The recording phone was placed in front of the participant, with other devices used in the task placed further away. The two recommended setups given to experimenters are illustrated in Fig. 1.

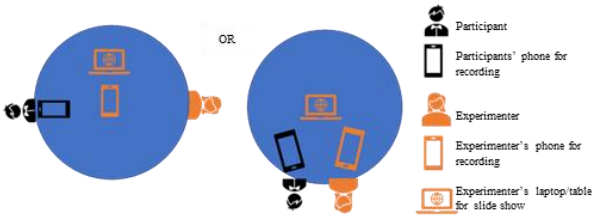

Figure 1: *Set ups for the at-home recordings.*

2.4. Annotation and analysis

The collected data were segmented and annotated in Praat [19] The first two authors checked and corrected the annotations to ensure consistency and accuracy.

First, the audio files were transcribed in Greek, and the transcripts were imported into a textgrid tier. Next, annotators determined whether intonational phrases ended in a H* accent followed by L-L%. If so, they selected and marked in a Praat textgrid tier the *analysis window*. This window included the accented word and any clitics accompanying it. For example, in an utterance ending in *της Αφρικής* [tis afriˈcis] "the Africa, GEN." or *και νερό* [ce neˈro] "and water" these sequences would form one analysis window if the content word was accented. This was done to examine the accents in context, as some Greek accents have consistent effects on the F0 of unaccented syllables, especially those preceding the accent [12].

Once the analysis window was determined, forced alignment was performed using the "Align Interval" function in Praat with the language set to Greek. Forced alignment followed the selection of the analysis window because the forced aligner for Greek worked well for short stretches of speech (up to 500 ms approximately) but produced quite significant errors in longer stretches. The annotators corrected any erroneous segment boundaries and also marked, on a separate tier, the onset and offset of the accented syllable. The type of nuclear accent was annotated in a separate tier; three categories were used, H*, L+H*, and H*+L, though only the realization of H* is reported here.

The resulting corpus consisted of 748 accents annotated as H*. They were relatively evenly divided between scripted and unscripted speech and among speakers; see Table 1.

Table 1: *Distribution of accents by task and speaker.*

	Speaker								
	01	**02**	**03**	**04**	**05**	**06**	**07**	**08**	**Total**
scripted	54	48	57	49	57	53	52	53	423
unscripted	41	39	42	32	56	27	22	66	325
Total	95	87	99	81	113	80	74	119	748

The F0 contours of the analysis windows were extracted using Praat with an octave cost of 0.1 and the following pitch range and time steps: AP01, AP05, AP07, AP08: 120 – 500 Hz, 5 ms time step; AP02, AP06: 70 – 280 Hz, 10 ms time step; AP03, AP04: 70 – 300 Hz, 10 ms time step. These values were empirically tested and chosen to minimize tracking errors due to creaky and breathy voice (which led to pitch halving and doubling respectively). The contours were spot checked but were not manually corrected: as the aim of fPCA is to uncover principal modes of variation in the data, any remaining mistrackings are unlikely to have a significant impact on the analysis. Undefined F0 values were interpolated using stine interpolation (imputeTS package, version 3.2 [20]). These F0 tracks were then scaled by speaker.

The onset of the accented vowel was used for landmark registration, to allow for a connection between traditional AM modelling that relies on tonal alignment and the understanding from fPCA. Landmark registration included time warping. The F0 tracks were smoothed with settings of knots = 8 and lambda = 10^4. These settings were determined empirically to best capture the contours without oversmoothing but also without retaining excessive microprosodic variation.

The smoothed and landmark registered F0 tracks were subjected to functional principal components analysis (fPCA [21], [22]) which captures the principal modes of variation in curves, so that each curve in the input can be treated as the outcome of the principal components. The relationship is expressed as an equation in which the contribution of each component to a given curve is its coefficient (or *score*). The

coefficients were used for statistical analysis. The overall procedure and analysis followed those of [11], [21], and [22].

The scores were statistically analysed using LMEMs (R: 4.0.3 [23]; lme4: 1.1.26 [24]) with speaker, task (scripted, unscripted), *crowding*, and the interactions of speaker × task and speaker × crowding as fixed factors. Speaker was treated as a fixed factor as an aim of the study was to uncover speaker variation. *Crowding*, an operationalization of tonal crowding, was defined as the temporal distance (in ms) of the accented syllable offset from the end of the analysis window.

We started with the full model in (2), and ran a series of competing models. Non-significant factors as determined by the function *anova*() (stats package in R: 4.0.3 [23]) were removed from the model sequentially, one by one, starting with the interactions and then with factors not involved in interactions, until all the remaining factors were statistically significant. If a factor was not significant but it was involved in a significant interaction it was retained in the model.

(2) lmer(PC# ~ speaker × crowding + speaker × task + (1|word), data = data, REML = FALSE)

3. Results

3.1. Average curves and fPCA

Fig. 2 shows the smoothed average curve per speaker in the analysis window; the dotted vertical line represents the onset of the accented vowel. As can be seen from these average curves, the speakers did not all produce accents of the same shape.

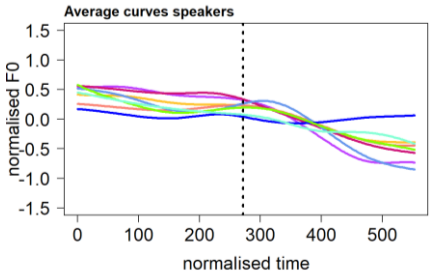

Figure 2: *Smoothed average F0 curves by speaker; the dotted line represents the onset of the accented vowel.*

The fPCA captures the most important modes of this variation. As can be seen in Fig. 3, 60% of the variability in the data relates to differences in accent scaling (PC1). PC2 captures 23.4% of the variability relating to shape; e.g., an accent with low PC1 and PC2 would be realized as a low plateau-fall. Finally, PC3 captures 9.2% of the variation and contributes to distinguishing a fall and a rise-fall shape. [We note that occasional final rises may be due to residual tracking errors.]

3.2. Speaker, task, and tonal crowding effects

The question that arises is whether the variability captured by fPCA reflects differences that relate to speaker-specific patterns, the task, or their combination. To address this question, we ran LMEMs, as noted in section 2.4.

For PC1, the final model was PC1~speaker × task+ (1|word); see Table 2. Tonal crowding was not significant and did not interact with speaker. The model showed a significant speaker × task interaction. As shown in Fig. 4, for some speakers, task did not have any effect on scaling (AP01, AP05, AP06), for others scaling was higher when reading (AP02, AP04, AP07), while a third group showed the reverse pattern (AP03, AP08). See also Fig. 5 for individual average curves by task, illustrating these differences.

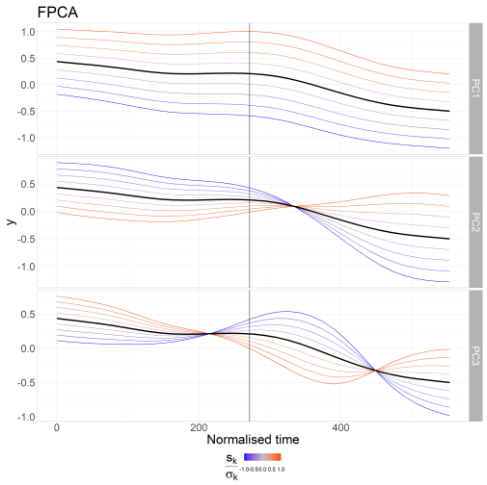

Figure 3: *PC curves for data pooled across speakers; the black line represents the mean curve; red and blue lines present the changes in this curve when a PC is within 1 and -1 standard deviation from the mean respectively; the vertical line indicates the onset of the accented vowel.*

For PC2, the final model was PC2~crowding + speaker × task; see Table 2. Crowding was significant. Additionally, both speaker and its interaction with task were significant, though relevelling indicated that the task differences were significant only for AP02 and AP03. AP02 produced rise-falls in unscripted speech but falls in read speech; AP03 produced rise-falls in both styles but timed the rise differently, so that the peak was before the accented syllable in unscripted speech and after it in read speech (see Fig. 4 and 5).

For PC3, the final model was PC3~speaker×crowding+ (1|word); see Table 2. There was no effect of task, but an effect of speaker (see Fig. 4 and 5), and an interaction with crowding. The interaction did not reveal any consistent trends and is not presented here due to space limitations. The effect of speaker indicated that there were different degrees of variability among the speakers with respect to PC3 (see Fig. 4).

Table 2: *Results of statistical analysis on PC1, PC2 and PC3 scores;*** < 0.001 ; ** < 0.01; * < 0.05; for details, see text.*

PC1	SumSq	MeanSq	DF	DenDF	F	Pr (> F)
speaker	622.3	88.9	7	734.0	0.39	0.91
task	99.6	99.62	1	220.88	0.44	0.51
sp × task	21076.8	3010.98	7	733.20	13.17	***

PC2	SumSq	MeanSq	DF		F	Pr (> F)
crowding	786	786.48	1		7.25	**
speaker	5922	846.03	7		7.80	***
task	42	42.18	1		0.39	0.53
sp × task	2176	310.91	7		2.87	**
Res	79278	108.45	731			

PC3	SumSq	MeanSq	DF	DenDF	F	Pr (>F)
speaker	488.34	69.76	7	685.75	2.31	*
crowding	23.50	23.50	1	437.65	0.78	0.38
sp × crowd	563.94	80.56	7	713.60	2.67	**

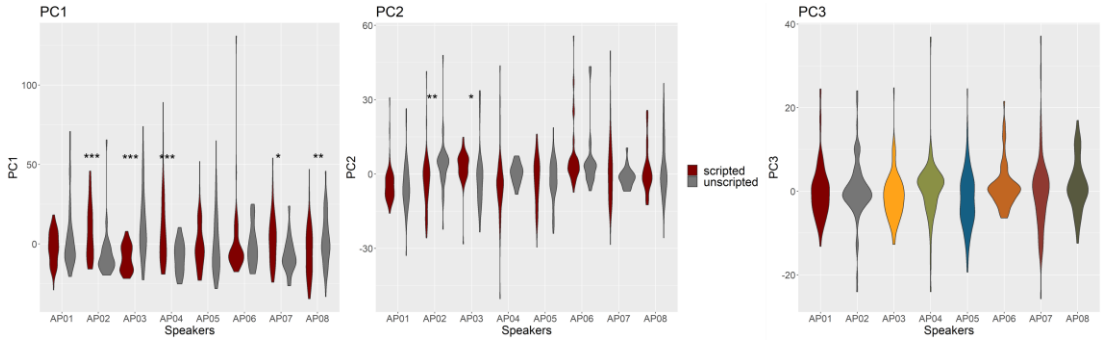

Figure 4: *Box plots of PC scores, by speaker and task for PC1 (left) and PC2 (middle), and by speaker only for PC3 (right).*

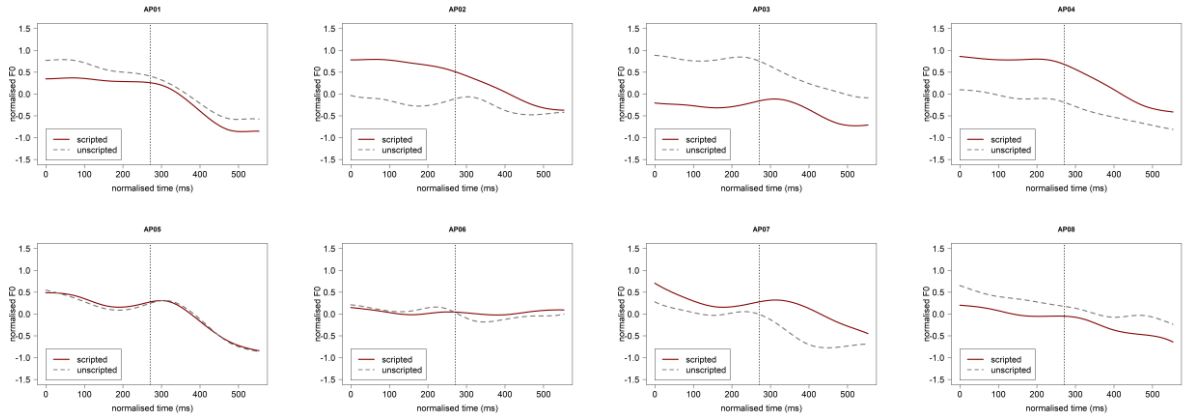

Figure 5: *Smoothed average F0 curves by speaker, separately for scripted and unscripted speech (solid and broken line respectively); the dotted line represents the onset of the accented vowel.*

4. Discussion and Conclusions

These results confirm that, on average, the shape of the Greek H* L-L% tune is a slightly declining plateau followed by a fall to the bottom of the speaker's range. However, this average belies substantial variation, illustrated in Figs 2 and 5, which present average curves by speaker (Fig. 2) and average curves by speaker and task (Fig. 5).

These figures, supported by the statistical analysis of the fPCA output, show considerable differences among speakers. First, task significantly affected scaling: some speakers were more animated in unscripted tasks, while others did the same when reading. Consequently, F0 level was higher in one task vs. the other but not in a consistent way (cf. AP2 and AP3, Figs. 4 and 5). In addition, curves varied more in one task than the other; consider, e.g., the variability in PC1 for unscripted speech in the data of AP03 and AP05, or the variability in PC2 scores for scripted speech in the data of AP04 and AP08 (Fig. 4). Finally, the exact shape of H* differed both across tasks and speakers. This was the case even though H* must be distinguished from L+H* and H*+L in Greek, as mentioned above, [12], [25], and its realization has been shown to be very stable in scripted speech [12]. The present data replicate the overlap between H*, L+H*, and H*+L reported in [12] and provide evidence of additional sources of variability for H*.

The differences observed here have important methodological repercussions. It is standard practice in AM, as noted in section 1, to use scaling and alignment as diagnostics of the phonological status of tonal targets, such as F0 minima (e.g., [26], [27], among many). It is also accepted practice to generalize about the phonetics of tonal events based on data collected from a single task. This stems from the assumption that targets reflecting phonological categories have stable scaling and alignment and thus will be minimally affected by task and similar factors. As the present data show, however, this does not always apply: scaling may vary substantially, as does accent shape and with it the alignment of F0 minima and maxima. When such variability is observed, it is tempting to try to discretize it in the form of "allotones" that appear in specific contexts, or to posit distinct phonological categories ([28], among many). However, neither approach is always possible or advisable (see [6] for a discussion). If such an approach had been taken in the present study, one would have had to conclude that different allotones were used in the same context by different speakers, or that the speakers used different accents altogether. There is no sound motivation for such conclusions, however, unless one focuses exclusively on phonetic detail.

Instead of adopting the above practices, here we followed Lohfink *et al.* [12] in using fPCA, and showed that fPCA can help researchers deal with documented variability while better understanding its sources.

5. Acknowledgements

The financial support of the European Research Council through grant ERC-ADG-835263 (SPRINT) to Amalia Arvaniti is hereby gratefully acknowledged. We thank our speakers who agreed to participate during the COVID-19 pandemic, Eleftheria Politi and Danae Tsinivits for data collection, Dafni Bagioka and Vicky Ioannidou for data annotation, and Kathleen Jepson for processing, data annotation and annotation coordination in the early stages of this work.

6. References

[1] J. B. Pierrehumbert, *The Phonetics and Phonology of English Intonation*. MIT PhD dissertation, 1980.

[2] D. R. Ladd, *Intonational Phonology*, Cambridge: Cambridge University Press, 2008.

[3] S. Frota, A. Arvaniti, and M. D' Imperio, "Prosodic Representations: Prosodic Structure, Constituents, and Their Implementation Segment-To-Tone Association Tonal Alignment", in A. C. Cohn, C. Fougeron, and M. K. Huffman (eds.), *The Oxford Handbook of Laboratory Phonology*, Oxford: Oxford University Press. pp. 265-274, 2011.

[4] A. Arvaniti, D. R., Ladd, and I. Mennen, "Stability of tonal alignment: The case of Greek prenuclear accents", *Journal of Phonetics*, vol. 26, pp. 3-25, 1998.

[5] A. Arvaniti, and J. Fletcher, "The Autosegmental-Metrical theory of intonational phonology", in C. Gussenhoven, and A. Chen (eds.), *The Oxford Handbook of Language Prosody*, Oxford: Oxford University Press, pp. 78-95, 2020.

[6] A. Arvaniti, "Analytical decisions in intonation research and the role of representations: Lessons from Romani", *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, vol. 7, no. 1:6, pp. 1-43, 2016.

[7] R. Smiljanic, "Early vs. late focus: pitch-peak alignment in two dialects of Serbian and Croatian", in L. Goldstein, D. H. Whalen, C. T. Best (eds.), *Laboratory Phonology* 8, Berlin: Mouton de Gruyter, pp. 495-518, 2009.

[8] N. Henriksen, "Style, prosodic variation, and the social meaning of intonation", *Journal of the International Phonetic Association*, vol. 43, no. 2, pp. 153-193, 2013. DOI:10.1017/S0025100313000054.

[9] M. Grice, S. Ritter, H. Niemann, and T. Roettger, "Integrating the discreteness and continuity of intonational categories", *Journal of Phonetics*, vol. 64, pp. 90-107, 2017.

[10] O. Niebuhr, M. D'Imperio, B. Gili Fivela, and F. Cangemi. "Are there "shapers" and "aligners"? Individual differences in signalling pitch accent category", in *Proceedings of ICPhS XVII*, Hong Kong, 17-21 August 2011, pp. 120-123.

[11] J. Barnes, A. Brugos, N. Veilleux, and S. Shattuck-Hufnagel. "On (and off) ramps in intonational phonology: Rises, falls, and the Tonal Center of Gravity", *Journal of Phonetics*, vol. 85, (2021). 101020.

[12] G. Lohfink, A. Katsika, and A. Arvaniti. "Variability and category overlap in the realization of intonation," in *Proceedings of ICPhS 2019*, Melbourne, Australia, Aug. 2019, pp. 701-705. https://assta.org/proceedings/ICPhS2019.

[13] K. M., Yu. "Tonal marking of absolutive case in Samoan", *Natural Language & Linguistic Theory*, *39*(1), pp. 291-365, 2021. doi:http://dx.doi.org/10.1007/s11049-020-09470-2.

[14] C. Huttenlauch, I. Feldhausen, and B. Braun. "The purpose shapes the vocative: Prosodic realisation of Colombian Spanish vocatives", *Journal of the International Phonetic Association, vol. 48*(1), pp. 33-56, 2018. doi: http://dx.doi.org/10.1017/S0025100317000597.

[15] E. Levon. "Same difference: The phonetic shape of high rising terminals in London", *English Language and Linguistics*, vol. 24(1), pp. 49-73, 2020. doi: http://dx.doi.org/10.1017/S1360674318000205.

[16] A. Arvaniti, and M. Baltazani. "Intonational analysis and prosodic annotation of Greek spoken corpora," in Sun-Ah Jun (ed), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford University Press, pp. 84–117, 2005.

[17] T. Georgakopoulos, and S. Skopeteas, "Projective vs. interpretational properties of nuclear accents and the phonology of contrastive focus in Greek," *The Linguistic Review*, vol. 27, no. 3, pp. 319-346, 2010. https://doi.org/10.1515/tlir.2010.012

[18] C. Zhang, K. Jepson, G. Lohfink, and A. Arvaniti. "Comparing acoustic analyses of speech data collected remotely," *The Journal of the Acoustical Society of America*, vol. 149, pp. 3910-3916, 2021. https://doi.org/10.1121/10.0005132.

[19] P. Boersma, and D. Weenink, "Praat: doing phonetics by computer." 2020, [Online]. Available: http://www.praat.org/.

[20] S. Moritz, and T. Bartz-Beielstein, "imputeTS: Time Series Missing Value Imputation in R," *The R Journal*, vol. 9(1), pp. 207-218, 2017. https://doi.org/ 10.32614/RJ-2017-009.

[21] M. Gubian, F. Torreira, and L. Boves, "Using Functional Data Analysis for investigating multidimensional dynamic phonetic contrasts," *Journal of Phonetics*, vol. 49, pp. 16–40, 2015. doi: 10.1016/J.WOCN.2014.10.001.

[22] Y. Asano, and M. Gubian.""Excuse meeee!!": (Mis)coordination of lexical and paralinguistic prosody in L2 hyperarticulation, " *Speech Communication*, vol. 99, pp. 183-200, 2018.

[23] R Core Team, "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, 2020, [Online]. Available: http://www.r-project.org/.

[24] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models Using lme4," *Journal of Statistical Software*, vol. 1, no. 1, pp. 1–48, Oct. 2015, [Online]. Available: https://www.jstatsoft.org/v067/i01.

[25] A. Arvaniti, "Crosslinguistic variation, phonetic variability, and the formation of categories in intonation," in *Proceedings of ICPhS 2019*, Melbourne, Australia, Aug. 2019, https://assta.org/proceedings/ICPhS2019/

[26] J. Borràs-Comes, M. del M. Vanrell, and P. Prieto, "The role of pitch range in establishing intonational contrasts", *Journal of the International Phonetic Association*, vol. 44, no. 1, pp. 1-20, 2014.

[27] V. Makarova, "The Effect of Pitch Peak Alignment on Sentence Type Identification in Russian", *Language and Speech*, vol. 50, pp. 385-422, 2007.

[28] J. I. Hualde, and P. Prieto, "Towards an International Prosodic Alphabet (IPrA)", *Laboratory Phonology*, vol. 7(1), 2016. DOI: http://doi.org/10.5334/labphon.11.