VARIABILITY AND CATEGORY OVERLAP IN THE REALIZATION OF INTONATION

Georg Lohfink[1], Argyro Katsika[2], Amalia Arvaniti[1]

[1]University of Kent, [2]University of California, Santa Barbara
G.Lohfink@kent.ac.uk akatsika@linguistics.ucsb.edu A.Arvaniti@kent.ac.uk

ABSTRACT

Functional Principal Component Analysis (FPCA) was used to model within-category variability and cross-category overlap of F0 features associated with three Greek pitch accents (H*, L+H*, H*+L), and thus disentangle categorical differences from gradient variability. FPCA is an analysis of curves, returning the dominant modes of curve variation as functions, called Functional Principal Components (PCs); each input curve receives a coefficient for identified PCs, representing the contribution of each PC to that curve's shape. The three accents, which have distinct pragmatic meanings, were utterance-final in declaratives and elicited from thirteen Greek speakers. PC1 and PC2 captured 87.7% of the data variance. Statistical modelling of the coefficients revealed presence of multiple cues, including overall shape, curve height, and position of curve peak. Though PCs showed cross-category overlap, together they distinguished the three accents, providing evidence that tonal events are realized by combining multiple cues, and in a variable manner.

Keywords: intonation, variability, Greek, Functional Principal Components Analysis

1. INTRODUCTION

A major issue in intonation research is modelling the variability of F0 contours while capturing significant generalizations that guide phonological abstraction, [3]. Some intonation models ignore variability altogether by dealing with idealized contours, e.g. [11, 21]; others, e.g. [9, 22], focus instead on capturing variability. However, this often happens at the expense of generalization; see [1, 3].

The autosegmental-metrical model of intonational phonology (henceforth AM, [17]) captures phonological generalizations but has difficulty dealing with variability, as phonetic invariance is used as a primary diagnostic criterion. In AM, pitch contours are said to consist of a string of tonal targets, that are treated as the reflexes of phonological tones if they show invariant scaling and alignment relative to some structural position (e.g. the onset of the accented syllable). This criterion is based on the findings of [4] and is known as *segmental anchoring*. It implies that tonal targets are invariant, remaining distinct for different tonal events, such as different pitch accents, and showing little cross-category overlap. Segmental anchoring has proved a useful heuristic for intonational analysis [13, 14]), but has its limitations: because it relies on invariance to determine phonological structure, it is at odds with the extent of variability documented in intonation in natural speech, e.g. [1]. Thus the idea of segmental anchoring requires radical rethinking.

Here, Functional Principal Components Analysis (henceforth FPCA) is used as the basis for a different approach to the study of intonation [10]. FPCA is an analysis of curves; it returns the dominant modes of variation in functional form, called *Functional Principal Components* (PCs). Every input curve receives a coefficient for identified PCs, representing the contribution of each PC to that curve's realised shape. Thus, the PCs can be seen as components that together determine the shape of each curve. The coefficients of the PCs, henceforth *scores*, can be statistically analysed together with other dimensions of the signal (such as duration) to quantify their joint contribution to the realization of tonal categories. FPCA allows us to consider both specific features of F0 curves, and to uncover commonalities between curves that may differ superficially. By choosing the window of analysis, it is also possible to consider distal effects, instead of focusing on F0 differences local to specific syllables.

Here, FPCA involved three pitch accents, H*, L+H*, and H*+L, all found in utterance-final position in Greek declaratives [2]. According to [2], these accents are used in different pragmatic situations. H* indicates that the accented item is new in discourse. L+H* indicates that the accented item is the one from a small set of alternatives that should be placed in the common ground; L+H* is often used for contrastive focus [5]. H*+L indicates that the accented item is new in discourse, but suggests that for the speaker it should have *already* been in the common ground (it is a way to implicate that one is stating the obvious).

The three accents are realized similarly, as they are all found in utterance-final position in declaratives followed by L-L% edge tones. Both H*+L and H* are falls; L+H* is a rise-fall. Further, Greek restricts the location of stress to the last three syllables of a word. This leads to tonal crowding, since the accent and

following edge tones (H* L-L%, L+H* L-L%, H*+L L-L%) must be realized on at most three syllables, and may have to be realized on one syllable only. Given that tonal crowding alters tonal realization [3, 6, 15], and the three accents are similar, they provide an ideal ground for testing and modelling effects of tonal crowding and variability.

2. METHODS

2.1. Speakers

The data were elicited from 13 native speakers of standard Greek (10 F, 3 M), aged from early 20s to early 40s (mean = 34, S.D. = 8).

2.2. Materials

The present analysis is based on the three test words shown in (1). As can be seen in (1a), (1b), and (1c), their stress is on the antepenult, the penult and the ultima respectively.

(1) a. [**laðo**ˈ**le**mono]# *oil-and-lemon-sauce*
b. [**lemo**ˈ**na**ða]# *lemonade*
c. [**ɣala**ˈ**na**]# *light blue*

The test words were always final in utterances that were either one or two content words long. These utterances were answers to questions, with the QA pairs presented as mini-dialogues to the participants. Representative dialogues, one with a short (one word) and one with a long (two words) test utterance, are given in (2) together with the expected tune and its association with the segmental string. In this instance, the expected accent was H* as the test utterances show new information. In longer utterances like (2b), the first content word, in (2b) [sinaˈɣriða] "sea-bream", carries a prenuclear L*+H accent [2].

2.3. Recording procedure

The participants were recorded in quiet locations in Athens, Greece, using a DAT recorder at a sampling rate of 44.1 kHz. Each dialogue was typed on a card in Greek orthography. The participants read them aloud as part of a larger dataset with different types of utterances which acted as distractors.

Each participant read the dialogues four times, with the cards being shuffled between repetitions. In total, 936 tokens were elicited [13 speakers × 3 accents × 3 test words × 2 utterance lengths × 4 repetitions], of which 844 were used for analysis. The other 92 tokens were discarded because speakers either did not use the intended tune, or had extensive stretches of creaky voice, which led to unreliable F0 tracking rendering the data unsuitable for FPCA. The discarded data are a small fraction of the corpus (9.8%), and given how FPCA functions, the omission is unlikely to have affected the FPCA outcome.

(2) a. Short utterance:
[ˈti naˈfto]
What's this?
[[laðoˈlemono]$_{ip}$]$_{IP}$
| | |
H* **L- L%**
Oil-and-lemon-sauce.

b. Long utterance:
[eˈsis ˈti θa ˈparete]
What will you have?
[[sinaˈɣriða laðoˈlemo]$_{ip}$]$_{IP}$
| | | |
L*+H **H*** **L-L%**
Sea-bream in oil-lemon-sauce.

2.4. Annotation and analysis

The data were annotated in Praat [7]. The window of analysis was the three-syllable interval ending with the stressed syllable of the test-word, shown in bold in (1). The onset of the accented syllable was also annotated and used for *landmark registration* (see below). The F0 of the three-syllable intervals was extracted using STRAIGHT [16], normalized by speaker, and submitted to FPCA following [10] and using the following smoothing parameters: k=8 and lambda=10^6. These values were selected from a number of possible combinations, after comparison between smoothed and original curves. Smoothing aimed at minimizing measurement errors produced by STRAIGHT. Analysis was based on F0 curves in Hz, since this is the most frequent measurement of F0 used in annotation. Time was scaled to range from 0 to mean duration. Landmark registration was conducted to align the onsets of the accented syllables across all curves, as shown in Fig. 1.

3. RESULTS

Fig. 2 illustrates the mean F0 over the three-syllable window for each accent pooled over stress positions, utterance lengths, and speakers. These data illustrate the basic shape of each accent, but also show the difficulties with locating tonal targets, particularly for H* which is realized as a gently falling plateau.

The first two components of FPCA, PC1 and PC2 captured 87.7% of the data variance (64.8% for PC1, and 22.9% for PC2). Fig. 3 illustrates the analysis: the black line is the average curve for the entire corpus (and thus the same for PC1 and PC2). Curves with + signs show the effect of +1 standard deviation of the coefficients to the shape of the curve; curves with – signs show the effect of -1 standard deviation. F0 curves in the corpus are a composite of the PC1 and

PC2 contribution. As shown in Fig. 3, PC1 reflects primarily (but not exclusively) differences in scaling: higher PC1 scores result in curves with higher scaling, but also in an earlier fall; lower PC scores result in lower scaling and a later fall. PC2 reflects a combination of contour shape and peak alignment: higher scores result in a "scooped" curve with a high late peak, while low scores result in a low scaled plateau and early fall.

Figure 1: Time-scaled F0 curves before landmark registration (left) and after (right).

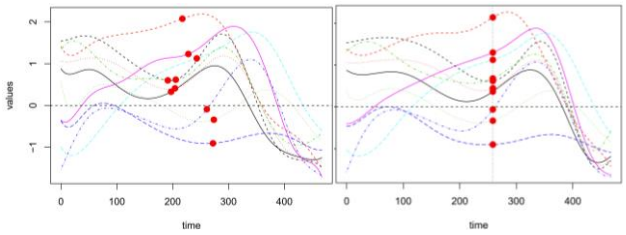


Figure 2: Normalized F0 per accent type across stress positions, utterance lengths, and speakers.

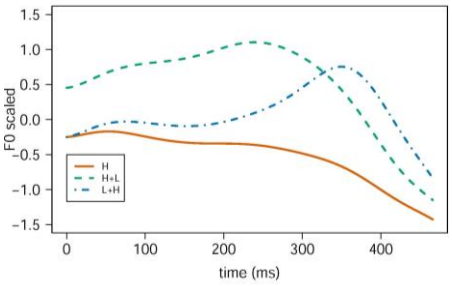
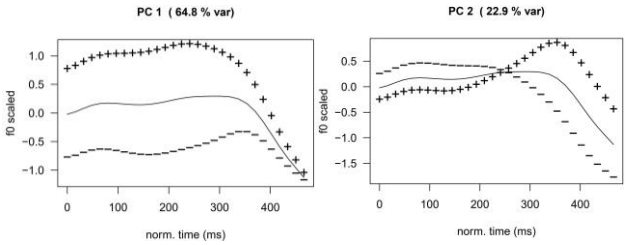
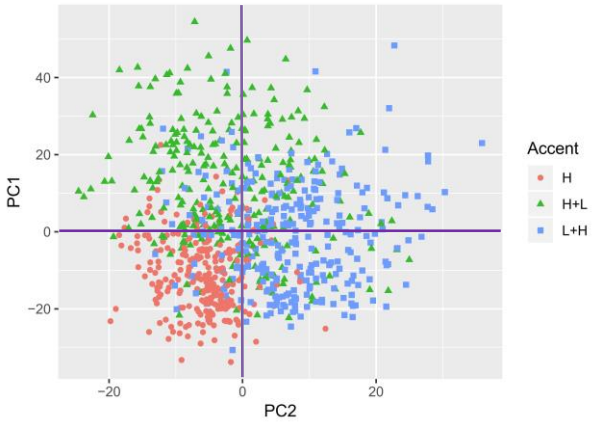


Figure 3: PC1 and PC2 curves modelling the data [solid black line = mean curve; + = higher PC scores; - = lower PC scores]; see text for details.



These trends apply to the pooled curves, but what is of interest here is how the PCs contribute to the F0 curves associated with each accent; this is illustrated in Fig. 4. For H*, both PC1 and PC2 have mostly negative scores, indicating that the location of the peak (i.e. the fall onset) is variable (since negative PC1 corresponds to a late fall, and negative PC2 to an early fall); overall, however, H* is scaled low. H*+L has consistently positive PC1, indicating consistently high scaling, while its PC2 is more variable suggesting variability in the location of the pitch fall. Finally, L+H* has variable PC1 scores, but predominantly positive PC2 scores, indicating that for L+H*, the most important feature is the scooped shape captured by positive values of PC2.

Figure 4: Distribution of accents on the PC1 × PC2 plane. Lines separate positive from negative scores.



The PC scores were modelled using linear mixed effects models in R [8, 20], with accent type (H*, L+H*, H*+L), stress position (antepenultimate, penultimate, final), and utterance length (short, long) as fixed factors, and speaker as random slopes (for PC1) or random intercepts (for PC2, due to model convergence issues); H*, antepenultimate, and long were the levels of comparison.

Table 1: Results of model for PC1; $p < .05$ = *; $p < .01$ = **; $p < .0001$ = ***; $p < .0001$ = ****

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-10.14	1.94	-5.23	****
H*+L	**21.81**	**2.99**	**7.29**	****
L+H*	**14.60**	**3.42**	**4.27**	***
stress_penult	**3.95**	**1.85**	**2.14**	*
stress_final	-2.26	1.64	-1.38	
U_lengthshort	**-3.02**	**1.30**	**-2.32**	*
H*+L:stresspenult	**-4.37**	**2.21**	**-1.97**	*
L+H*:stresspenult	-3.51	2.18	-1.61	
H*+L:stressfinal	-2.00	2.20	-0.91	
L+H*:stressfinal	3.37	2.17	1.55	
H*+L:U_lengthshort	**6.91**	**1.80**	**3.84**	***
L+H*:U_lengthshort	**-8.78**	**1.77**	**-4.97**	****

The statistical analysis indicates that PC1 and PC2 are affected by stress and utterance length, with both factors interacting with accent (see Tables 1 and 2). Stress affects mostly PC2, which is significantly lower when the accent is on the ultima. The shape of the PC curves suggests that stress effects differ by accent: for H* and H*+L final stress results in an earlier fall, while for L+H* it results in undershoot of the accent's "scooped" shape (see Fig. 5, PC2). Utterance length affects both PC1 and PC2. For PC1, short utterances lead to lower scores (i.e. lower scaling overall) for H* and L+H*; for PC2 the effect is found only for L+H*, with PC2 being significantly higher in short utterances (see Fig. 6). As a result of these interactions, while the PCs are statistically distinct by accent, they still show overlap across accents, as indicated by the density plots in Fig. 7.

Table 2: Results of model for PC2; $p < .05$ = *; $p < .01$ = **; $p < .0001$ = ***; $p < .0001$ = ****

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-3.54	0.97	53.55	***
H*+L	**7.06**	**1.03**	**820.1**	****
L+H*	**13.6**	**1.03**	**820.59**	****
stress_penult	**-3.63**	**0.9**	**820.54**	****
stress_final	**-3.95**	**0.89**	**820.38**	****
U_lengthshort	-0.23	0.73	819.85	
H*+L:stresspenult	-1.82	1.26	820.67	
L+H*:stresspenult	-1.19	1.24	820.38	
H*+L:stressfinal	**-8.45**	**1.26**	**820.3**	****
L+H*:stressfinal	**-8.61**	**1.24**	**820.73**	****
H*+L:U_lengthshort	0.54	1.03	819.76	
L+H*:U_lengthshort	**6.11**	**1.01**	**820.84**	****

Figure 5: Box plots of PC1 (left) and PC2 (right) as a function of accent and stress position.

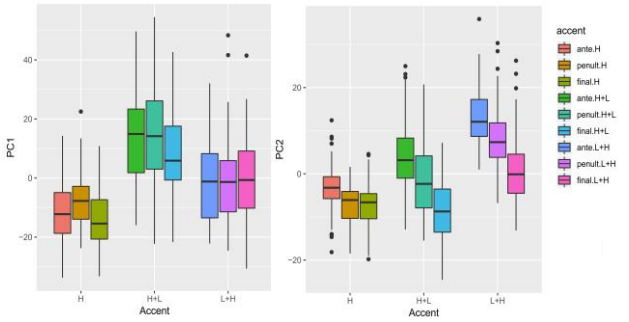
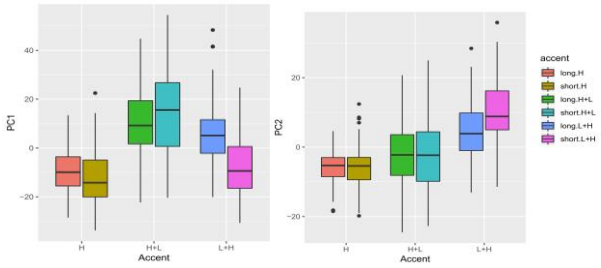


Figure 6: Box plots of PC1 (left) and PC2 (right) as a function of accent and utterance length.

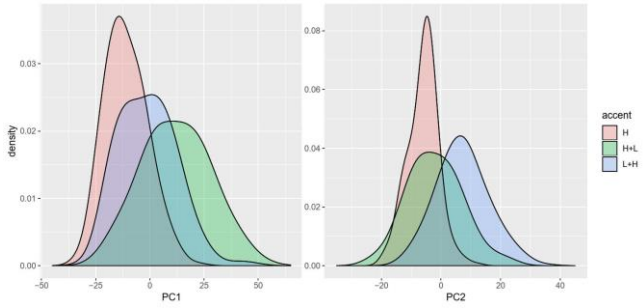


4. DISCUSSION

With respect to AM, the findings indicate that participants produced distinct accents in response to distinct pragmatic contexts, consistent with the descriptions and representations in [2].

In addition, FPCA offered new insights into the realization of the accents. First, FPCA showed that neither scaling nor alignment is invariant. As illustrated in Fig. 7, the PC scores show considerable overlap between accents, even though the differences between them are statistically significant. This should not be surprising, either based on statistics or on what we know about the realization of phonetic categories, such as VOT, and the extent to which they overlap, even when contrastive [18, 19]. However, because invariance has been an important criterion for determining intonation structure and the primitives of each system, this kind of overlap is seen as problematic and even in need of fine-grained representation [12]. The overlap uncovered here indicates that such granularity in representation is unrealistic.

Figure 7: Density plots of PC1 scores (left) and PC2 scores (right), separately for each accent.



Additionally, FPCA captured co-dependencies between scaling and alignment: PC scores led to either an earlier and higher peak (PC1), or an earlier and lower peak (PC2). Further, the positive PC2 values for L+H* indicate that *shape* may be more important than alignment for the realization of some tonal categories. Overall, these results question AM assumptions about the realization of tonal targets: scaling and alignment are not independent of each other, while invariant alignment is an idealization, and may not be the most important component of a tonal category's realization. Additional results, not shown here, further indicate that dimensions like duration are used in a trade-off relationship with F0 features to encode tonal categories like the accents investigated here.

Finally, the data reveal not only well-known effects of tonal crowding [3, 6] but also distal effects of context, here due to the presence of a preceding accent. The results show significant scaling changes *throughout the three-syllable window* of analysis, with the effects differing by accent: H* and L+H* have higher PC1 when a L*+H accent precedes (Fig. 6): H* creates a plateau with it [2], while for L+H*, the preceding L*+H leads to undershoot of its L tone (as also indicated by its lower PC2; see Fig. 6). For H*+L, on the other hand, the preceding L*+H does not create a plateau with the H*+L but, rather, leads to a lowering of the H*+L accent's scaling (see Fig. 6, PC1).

Overall, the results suggest that the established research focus on localized F0 targets and invariance as criteria for the phonological status of tonal events risks positing categories that are too fine-grained and capture phonetic variability rather than essential contrasts. Instead the results argue in favour of treating tonal events similarly to segments, i.e. as being expressed by a number of phonetic parameters that show variability and are in trading relationships with each other.

5. REFERENCES

[1] Arvaniti, A. 2016. Analytical decisions in intonation research and the role of representations: Lessons from Romani. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7(1), 6. DOI: http://doi.org/10.5334/labphon.14

[2] Arvaniti. A., Baltazani, M. 2005. 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. 84–117.

[3] Arvaniti, A., Ladd, D. R. 2009. Greek wh-questions and the phonology of intonation. *Phonology* 26, 43–74.

[4] Arvaniti, A., Ladd, D. R., Mennen, I. 1998. Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics* 26, 3–25.

[5] Arvaniti, A., Ladd, D. R., Mennen, I. 2006a. Tonal association and tonal alignment: evidence from Greek polar questions and contrastive statements. *Language and Speech* 49, 421–450.

[6] Arvaniti, A., Ladd, D. R., Mennen, I. 2006b. Phonetic effects of focus and "tonal crowding" in intonation: Evidence from Greek polar questions. *Speech Communication* 48, 667–696.

[7] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer [Computer program].

[8] Douglas B., Maechler, M., Bolker, B., Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1), 1–48.

[9] Fujisaki, H. 1983. Dynamic characteristics of voice fundamental frequency in speech and singing. In: MacNeilage, P. F. (ed) *The production of speech*. Heidelberg: Springer-Verlag, 39–55.

[10] Gubian, M., Torreira, F., Boves, L. 2015. Using functional data analysis for investigating multidimensional dynamic phonetic contrasts. *Journal of Phonetics* 49, 16–40.

[11] Hirst, D., Di Cristo, A., Espesser, R. 2000. Levels of representation and levels of analysis for the description of intonation systems. In: Horne, M. (ed), *Prosody: Theory and Experiment, Studies presented to Gösta Bruce*. Dordrecht: Kluwer Academic Publishers, 51–87.

[12] Hualde, J. I., Prieto, P. 2016. Towards an International Prosodic Alphabet (IPrA). *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7(1), 5. DOI: http://doi.org/10.5334/labphon.11

[13] Jun, S.-A. (ed). 2005. *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.

[14] Jun, S.-A. (ed). 2014. *Prosodic Typology II: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.

[15] Katsika, A. 2016. The role of prominence in determining the scope of boundary-related lengthening in Greek. *Journal of Phonetics* 55, 149–181.

[16] Kawahara, H., de Cheveigné, A., Banno, H., Takahashi, T., Irino, T. 2005. Nearly defect-free F0 trajectory extraction for expressive speech modifications based on STRAIGHT. *Proceedings of Interspeech 2005*.

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

[18] Nakai, S., Scobbie, J. M. 2016. The VOT category boundary in word-initial stops: Counter-evidence against rate normalization in English spontaneous speech. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7(1), 13. DOI: http://doi.org/10.5334/labphon.49

[19] Piccinini, A. & A. Arvaniti. 2018. Dominance, mode, and individual variation in bilingual speech production and perception. *Linguistic Approaches to Bilingualism*. https://doi.org/10.1075/lab.17027.pic

[20] R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/.

[21] 't Hart, J., Collier, R., Cohen, A. 1990. *A Perceptual Study of Intonation*. Cambridge: Cambridge University Press.

[22] Xu, Y. 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46(3-4), 220–251.