DOI: 10.17469/O2112AISV000008

RICCARDO ORRICO, STELLA GRYLLIA, NA HU, AMALIA ARVANITI

The role of metrical structure in prominence perception

Despite the increased interest on prominence, there is still lack of consensus of what prominence is. A recent review (Ladd, Arvaniti, 2023) has pointed out that the work conducted so far has painted a contradictory picture: on the one hand, some studies report that prominence assessment changes as a function of acoustic properties of words, pointing to the interpretation of prominence as a universal phenomenon; other evidence, however, suggests that the role of prominence is determined by phonology, therefore language-specific. We report a prominence study in Greek with the aim of testing the hypothesis that phonological factors linked to metrical strength trump acoustic salience in the assessment of prominence. The results met our predictions and revealed that accented words are more likely to be selected as prominence than unaccented words, regardless of their acoustic properties and, crucially, even when accented words are less salient than unaccented words.

Keywords: Prominence, Metrical structure, Greek, Rapid Prosody Transcription.

1. *Introduction*

The last decades have witnessed an increased interest in the study of prosodic prominence. Such interest may have been encouraged by the development of experimental paradigms that have facilitated the investigation of prominence and other prosodic phenomena. A popular paradigm used in prominence studies is Rapid Prosody Transcription (henceforth RPT, Cole, Mo & Baek, 2010a; Cole, Mo & Hasegawa-Johnson, 2010b; see Cole, Shattuck-Hufnagel, 2016, for an overview). In RPT, linguistically untrained participants are asked to listen to utterances and mark prominent words and the location of boundaries on a transcript of those utterances. The researchers then examine the factors contributing to word prominence based on the characteristics of the words most commonly chosen by participants. Specifically, the results are analyzed by considering inter-rater agreement (typically, Fleiss' Kappa) and calculating p-scores, the percentage of participants who marked a word as prominent (a similar measure is also used to analyze the properties boundary placement, b-scores, see Cole et al., 2010a). The linguistic correlates of prominence are statistically validated by correlating participants' responses with the properties of individual words. The ultimate aim of many RPT studies is to test how likely a word is to be selected as a function of the properties of that word. Some studies used p-scores as the outcome variable of interest for their statistical testing, such as Cole et al. (2010b) who performed correlation analysis between p-scores and properties of the signal in American English, or Baumann and Winter (2018) who used p-scores to perform a random forest analysis to investigate the relative importance of phonetic, phonological, and lexical cues on p-score variation in German. Other studies have used the raw binary response prominent/not prominent for individual words as the dependent variable of Generalized Linear Mixed-Effect Models (among others, Baumann, Winter, 2018; Cole, Hualde, Smith, Eager, Mahrt & De Souza, 2019; Bishop, Kuo & Kim, 2020; Arvaniti, Gryllia, Zhang & Marcoux, 2022; Im, Cole & Baumann 2023; Orrico, Gryllia, Kim & Arvaniti, 2023; 2025; Orrico, Gryllia, Hu, Kim & Arvaniti, 2024).

These studies have provided useful insight into the perception of prominence and have shown that prominence perception is affected by several parameters, including but not limited to acoustic and phonological factors (see also Wagner, Origlia, Avesani, Christodoulides, Cutugno, D'Imperio, Escudero, Gili Fivela, Lacheret, Ludusan, Moniz, Chasaide, Niebuhr, Rousier-Vercruyssen, Simon, Šimkz, Tesser & Vainio, 2015 on this point). Amplitude and duration of words have been strongly correlated to the perception of a word as prominent; this has been shown for American English (Cole et al., 2010b; 2019; Bishop et al., 2020; Im et al., 2023), German (Baumann, Winter, 2018; Riesberg, Kalbertodt, Baumann & Himmelmann, 2020), French and Spanish (Cole et al., 2019), and Italian (D'Imperio, 1997). The role of F0 in prominence perception has also been shown for different languages (for German, Baumann, Winter, 2018; for English, Bishop et al., 2020; Im et al., 2023; for English, Spanish, and French, Cole et al., 2019), though some studies have reported that its contribution is weaker than other acoustic variables (Cole et al., 2010b and Kochanski, Grabe, Coleman & Rosner, 2005 for American and British English respectively). Other prosodic factors include the presence of a pitch accent, which has been considered a strong prominence predictor in English (Bishop et al., 2020; Im et al., 2023) and German (Baumann, Winter, 2018) and the type of pitch accent, with low accents being considered less prominent than high ones, and high accents being considered less prominent than rising ones (see Baumann, Winter, 2018 for evidence in German, and Arvaniti et al., 2022 and Orrico et al., 2023, 2025 for evidence in British English). Finally, prominence perception is also affected by non-prosodic cues. Recent studies show the effect of pragmatics: Orrico et al. (2025) reported the effect of contrast on word selection in British English; Im et al., 2023 reported the effect of givenness and correction in American English. Other factors include word frequency (Cole et al., 2010b; Baumann, Winter, 2018) and syntactic properties (Calhoun, Wollum & Kruse Va'ai, 2019). Some studies show that the relative use of prosodic and non-prosodic cues for prominence selection during RPT depends greatly on individual variability. Orrico et al. (2025) reported that, in British English, the selection of a word as prominent varies as a function of both accents' shape and function, though individual variability has a role in defining which one of the two criteria is more important. Similar findings have also been reported for German: while both prosodic and non-prosodic factors have a role in predicting prominence, individuals may have different preferences towards the criteria used, with some preferring prosodic and others non-prosodic cues (Baumann, Winter, 2018).

1.1 Prominence beyond acoustic salience

The studies reported in the previous section have unveiled crucial aspects of the processes behind prominence perception; however, a clearer understanding of what constitutes prominence is still lacking (cf. Ladd, Arvaniti, 2023). Prominence is often defined very vaguely, as any property of a word that makes it stand out compared to others (see the definitions in Cole et al., 2010b and Bauman, Winter, 2018). The lack of a restrictive definition is partly the reason why researchers have followed exploratory approaches (correlation between responses and properties of words) rather than testing specific hypotheses. Ladd and Arvaniti's (2023) pointed out that recent results provide a contradictory picture of prominence rather than clarifying its properties. As reported above, some findings correlate prominence to gradient dimensions of the signal, like F0, duration, and amplitude, and suggest a straightforward relationship between the two. In other words, prominent words are expected to have specific features: they are longer, louder, or have higher pitch than non-prominent words. Additionally, when considering the role of pitch accent, factors linked to F0 shape (e.g. the phonetic difference between a L* and a L+H*) have been argued to play a crucial role in determining relative prominence, thereby indicating the importance of acoustic salience over phonological structure. This points towards a universal concept of prominence.

In contrast, other studies suggest that the role of prominence is language-specific. We can indeed find evidence that such a view of prominence – as pertaining to words that are acoustically more salient than others – does not apply across languages. Riesberg et al. (2020) report an RPT study in which German and Papuan Malay listeners marked prominent words in both German and Papuan Malay. Inter-rater agreement on prominence selection revealed that German listeners showed good levels of agreement in both languages, while Papuan Malay listeners showed i) higher agreement when rating prominence in German than in their native language and ii) lower agreement than the German participants when rating Papuan Malay. In addition, the cues used by Papuan Malay speakers were consistent with those used by the German ones when rating prominence in German. These results indicate that while prominence rating in German might be based on universally recognizable cues that make words acoustically salient (both German and Papuan Malayan listeners had similar intuitions about German), the same set of cues does not define prominence in Papua Malay. In fact, while German listeners recognized those cues as prominence lending also in Papuan Malay, Papuan Malayan listeners could not agree on what prominence is in their own language, showing that they do not have the same notion of prominence as German speakers. Related to this point, Cole et al. (2019) report an RPT investigation on English, Spanish, and French under different instruction conditions, with one group being given a definition of prominence based on acoustic factors and another being given a definition based on pragmatic criteria. The results show that while all three languages were affected by the instructions, the effect was larger in French when compared to English. This outcome is most likely related to typological differences between French and

English: French is a language that lacks stress, and in which pitch accents are not prominent-lending and are not used to signal focus (Welby, 2006; Cole et al., 2019) which may have led the French speakers to adhere more closely to the instructions.

Additionally, the investigation of prominence using other paradigms has suggested that the link between prominence and acoustic cues is mediated by phonological structure. Rump and Collier (1996) report two experiments in Dutch on the relationship between relative pitch height and focus. In the first experiment, participants were asked to adjust the F0 height of the pitch accents in the utterance *Amanda gaat naar Malta* 'Amanda goes to Malta', bearing one pitch accent on *Amanda* and another on *Malta*; the aim was to have participants adjust the F0 contour so that it corresponded to different focus conditions (broad focus, corrective focus on *Amanda*, *Malta*, or both). In the second experiment, listeners were asked to map pitch contours with the focus condition. Overall, the results showed that while the presence of a corrective focus on *Amanda* required a high peak on that word and the absence of f0 movement on *Malta*, a corrective focus on Malta did not require that *Amanda* had flat f0; in fact, the adjustments created for that condition still had a visible movement on *Amanda*. Such a relationship between focus and the relative F0 height on the two words cannot be explained within a framework in which prominence is directly linked to F0. On the contrary, this shows the role of metrical structure on the interpretation of prominence: the first utterance (focus on *Amanda*) is interpreted as having a nuclear accent (*Amanda*), which is followed by a deaccented *Malta*; the second utterance (focus on *Malta*) has a prenuclear accent on *Amanda*, which is followed by the nuclear accent on *Malta*. Thus, the interpretation of prominence does not depend on which of the two words has the highest pitch but rather on which word bears the metrically strongest element, i.e. the nuclear pitch accent (cf. Ladd, Arvaniti, 2023).

These findings, together with methodological investigations of the RPT paradigm, cast doubts on the purported connection between acoustic salience and prominence perception. Indeed, while it is true that RPT has been proven to be a robust paradigm (see, for example, Orrico et al. 2025 which showed that RPT results can be replicated, both in relation to the role of phonetics and pragmatics in prominence assessment and in relation to the individual variability in the responses), some studies have reported the relationship between RPT results and the type of task being used. As mentioned above, the instructions can have an effect on the way listeners approach the task (Cole, Mahrt & Hualde, 2014; Cole et al.; 2019). Additionally, Orrico et al. (2024) showed that RPT responses are also affected by the number of words that listeners are allowed to select. They reported two RPT studies in Greek, one in which they asked participants to mark as many prominent words as they saw fit (following the traditional paradigm, Cole et al., 2010a; 2010b), and one in which they asked participants to mark only the most prominent word (a task that had been tested by Calhoun et al., 2019 in English and Samoan). While the general picture of prominence did not change across tasks, the effect of acoustic cues on prominence selection was much smaller when listeners were allowed to select only the most prominent word; e.g., F0 height did not affect responses. Such task-dependent results require that findings on prominence should be interpreted with caution.

In sum, recent studies have painted a picture of prominence that is connected to the acoustic properties of the signal. While this is a result that is reported in several studies employing RPT, therefore robust, the investigation of non-Germanic languages (e.g. Cole et al., 2019; Riesberg et al., 2020), as well as evidence collected using other paradigms (e.g. Rump, Collier, 1996), challenges such a view of prominence. One possible factor that can explain the results that do not fit the standard understanding of prominence is offered by Ladd and Arvaniti (2023). These authors highlight the need of taking into account language-specific metrical structure in the definition of prominence: prominent words are those that are parsed in metrically strong positions and not those that are the most salient from an acoustic perspective.

1.2 Overview of Greek intonation

In this section we provide a brief overview of Greek intonation couched within the Autosegmental-Metrical (AM) framework (Pierrehumbert, 1980; Ladd, 2008). We describe the tunes frequently used for questions and statements, showcasing that metrically strong words are not necessarily acoustically more salient than weaker words (e.g. unaccented words), as in polar questions with non-final focus. Most of these tunes were used in the current study.

In statements with broad focus, the default tune is (L*+H) H* L-L %, with the nuclear pitch accent H* being realized as a high and declining plateau on the last content word (Baltazani, 2003; Arvaniti, Baltazani 2005; Lohfink, Katsika & Arvaniti, 2019). All content words before the nucleus carry a L*+H prenuclear accent (Arvaniti, Ladd & Mennen, 1998). In statements with narrow focus, the nuclear pitch accent is L+H* followed by L-L % edge tones (Arvaniti, Baltazani, 2005; Arvaniti, Ladd & Mennen, 2006; Georgakopoulos, Skopeteas, 2010; Lohfink et al., 2019). As in broad focus statements, all content words before the nucleus have a L*+H prenuclear accent, while words following the nucleus are typically deaccented.

Greek polar questions are string identical to declaratives and are distinguished from them only by intonation. The default tune for polar questions is (L*+H) L* H-L %; all content words preceding the nuclear accent bear a L*+H accent, and words following the nucleus are typically deaccented. The nuclear L* accent is realized as a low F0 stretch on the stressed syllable of the word in focus. The edge tone configuration is realized as a rise-fall movement, whose alignment depends on the location of the focus. When the word in focus is the last word of the utterance, the L* accent is realized on the stressed syllable of that word, and the H- phrase accent is realized on its last syllable followed by the L % boundary tone. When the word in focus is not the final word of the question, the H- phrase accent is realized on the *stressed syllable* of the final word (Grice, Ladd & Arvaniti, 2000;

Arvaniti 2002). Note that polar question tunes with non-final focus represent a mismatch between metrical strength and acoustic salience, as opposed to narrow focus statements in which the two dimensions cooccur on the same word. The focused element in questions is indeed low in F0, while the final word, despite being unaccented, shows a rise-fall movement (H-L %) on the stressed syllable (where typically pitch accents are realized) that is acoustically salient. Despite its resemblance to a pitch accent, native speakers distinguish between the H- phrase accent and phonetically comparable accentual rises, like the L+H* of narrow focus statements (Arvaniti et al., 2006).

Finally, wh-questions are realized with the nuclear pitch accent associated with the stressed syllable of the wh-word, which is fronted, followed by a dip or low F0 stretch (depending on the length of the question), followed by the boundary tone, realized on the last vowel of the question (among others Baltazani, 2002; Baltazani, 2003; Arvaniti, Baltazani, 2005; Arvaniti, Ladd, 2009; Grice et al., 2000; Baltazani, Gryllia & Arvaniti, 2020). Different tonal configurations have been attested, depending on the pragmatic function of the question. The most frequent tonal patterns for wh-questions include either a L*+H pitch accent in nuclear position, typically followed by a L-H % combination of edge tones, or a L+H* pitch accent typically combined with flat L-L % edge tones. Fig. 2 in § 2.3 provides examples of some of the contours that have been discussed above and were used as stimuli.

1.3 Research question

The present study aims to investigate the role of metrical structure on prominence assessment. We focus on prominence at the sentence level (prominent words in an utterance) and do not consider lexical prominence (prominent syllable in a word); though see Ladd (2008: Ch. 2 and Ch. 7) and Ladd and Arvaniti (2023) for a discussion about the relationship between lexical and sentence prominence. Our study investigates the relationship between sentence prominence assessment in RPT and phonological cues linked to the presence and type of accents. The study does not follow the traditional use of RPT (i.e. investigate all possible properties of a word that make it prominent) but uses the paradigm to test the hypothesis that prominence perception reflects the metrical organization of the utterance regardless of the differences in the phonetic realization of prosodic events. Our study is couched within the AM theory of intonation, in which the metrical structure (presence or absence of an accent, whether that accent is nuclear or prenuclear) and the phonetic realization of tonal events are orthogonal dimensions. This allowed the design of stimuli in which metrical structure and acoustic cues do not always match (see differences in questions and statements in § 1.2) and tested the prediction that the selection of a word as prominent would be determined by whether that word is metrically strong (accented or not), regardless of what accent it bears (e.g. L* or L+H*) or its acoustic salience.

2. *Methods*

2.1 Participants

Twenty-three Greek speakers participated in the study, and responses of four participants were excluded because three of them self-reported a history of language- or hearing-related disorders and one selected more than 85 % of the words in more than 10 % of the stimuli, suggesting carelessness during the task.

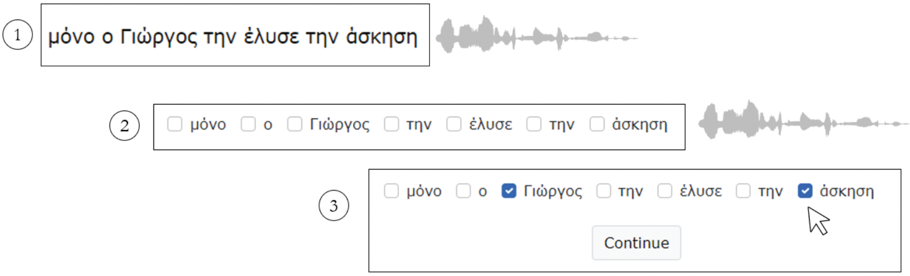
Therefore, the analysis involved the responses from 19 participants (12 females, 5 males, 1 non-binary, and 1 preferred not to say; ages ranged from 20 to 61 years, Median = 24.0, Mean = 31.8). They were all native speakers of Greek, brought up in Greece. They were recruited via personal networks and participated voluntarily.

2.2 Procedure

The study was conducted online using ROLEG (Radboud Online Linguistic Experiment Generator, https://www.roleg.nl/), a web platform developed at the Centre for Language Studies (CLS) at Radboud University. The participants were provided with a web link that they used to access the study page; they were asked to carry out the task in a quiet environment, listen carefully to the stimuli via headphones, and regulate the volume to a comfortable level; they were also encouraged to contact us in case of questions about the study or of technical problems. They were asked to select the words that in their opinion "stand out from the rest, that is, sound more important or stressed or as if the speaker has emphasized them" by checking a box next to that word. The wording of the instructions was chosen to avoid biasing participants towards focusing on either acoustic or semantic aspects of the words (cf. Cole et al., 2014; 2019). Participants were instructed to select as many words as they saw fit.

The study started with three practice trials followed by 74 experimental trials and included four prompts for self-paced breaks. Trial randomization was done by creating three lists containing all the stimuli in different orders; the lists were randomly assigned to participants. In each trial, participants listened to an utterance twice while seeing its transcript on screen; the transcript lacked punctuation (other than apostrophes in contractions) and capitalization (except for proper names). During the second repetition, a checkbox appeared next to each orthographic word, as shown in the schematic representation of a sample trial in Fig. 1.

Figure 1 - *Schematic representation of experimental trials. Participants first heard the utterance while seeing a transcription (1), then they heard a repetition of the same utterance and a selection box appeared next to each word to select prominent words (2)-(3)*

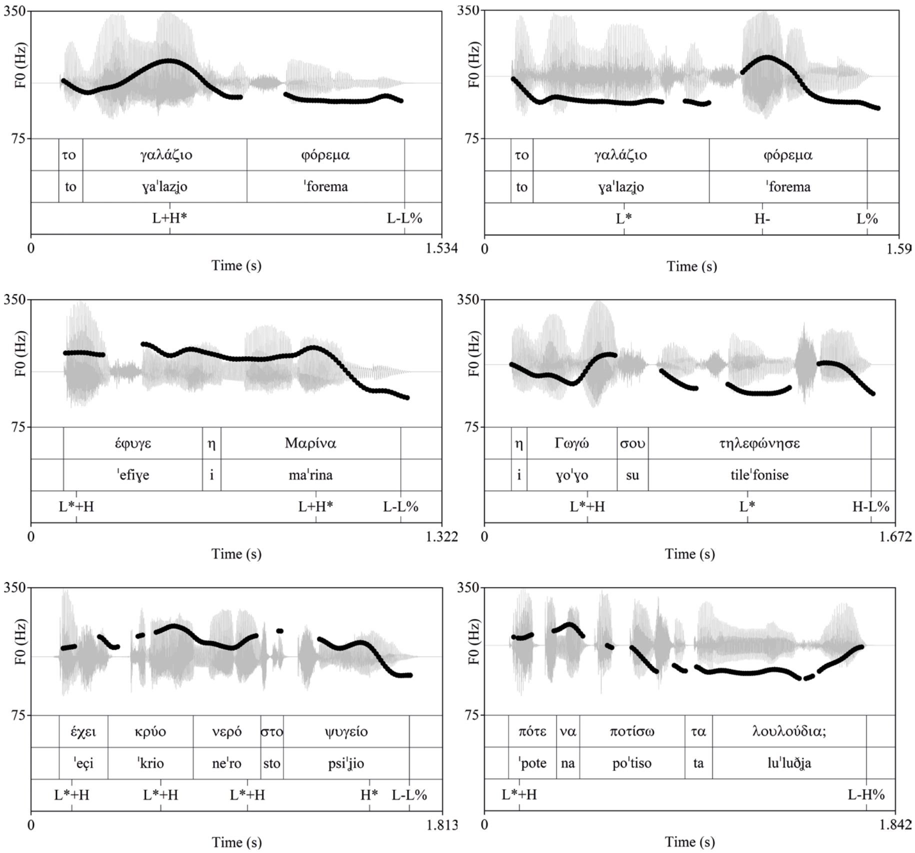


2.3 Stimuli

The stimuli were 74 utterances recorded from one female native speaker of Greek. They varied in length (2-8 orthographic words, 277 in total), sentence type, and focus position, all consisting of one intonational phrase with only one intermediate phrase; their tunes (pitch accents and edge tones) reflected those that are most typically used for the sentence types according to previous descriptions of Greek intonation (see § 1.2). After recording the stimuli, two authors listened to and inspected the utterances to ensure they were produced with the intended tune, focus location, and phrasing. The stimuli included the following utterance types, which were chosen to have stimuli with different types of nuclear accents. See Fig. 2 for examples.

- Narrow focus statements with final (10 items) and non-final focus (16 items); they were realized with a L+H* nuclear accent on the focused word, a prenuclear L*+H on all preceding content words, and L-L % edge tones. In items with final focus, the nuclear pitch accent was realized on the last content word in the utterance; in items with non-final focus, the nuclear pitch accent was realized on a non-final word, with the following content words being deaccented.
- Polar questions with final (10 items) and non-final focus (16 items), realized with a L* nuclear accent on the focused word, a L*+H on all preceding content words, and H-L % edge tones. In items with final focus, the nuclear accent was realized on the stressed syllable of the last content word, while the H-L % edge tones were realized on the last syllables of that word; in non-final focus items, the nuclear accent was realized on non-final words, while the H- rise was aligned with the stressed syllable of the last word.
- Broad focus statements (16 items), realized with a H* nuclear accent on the final word, L*+H prenuclear accents on all preceding content words, and L-L % edge tones.
- Information-seeking wh-questions (6 items), realized with a nuclear L*+H accent on the sentence-initial wh-word and L-H % edge tones.

Figure 2 - *Sample stimuli for the following sentences (from top-left to bottom-right): i) non-final narrow focus statement [to ɣaˈlazʝo ˈforema]* 'The blue dress'*; ii) polar question with non-final focus [to ɣaˈlazʝo ˈforema]* 'The blue dress'*; iii) narrow-focus statement with final focus [ˈefiʝe i maˈrina]* 'Marina left'*; iv) polar question with final focus [i ɣoˈɣo su tileˈfonise]* 'Did Gogo call you?'*; v) broad focus statement [ˈeçi ˈkrio neˈro sto psiˈʝio]* 'There's cold water in the fridge'*; and vi) wh-question [ˈpote na poˈtiso ta luˈluðʝa]* 'When should I water the flowers'



Here we report an analysis of these utterances that shows how the acoustic properties of the words relate to their phonological status; the analysis will confirm that metrical strength and acoustic salience do not always correlate (see § 1.3). We used Praat (Boersma, Weenink, 2023) to extract $F0_{max}$, duration, and Root-Mean-Square (RMS) amplitude from the stressed syllables of all words in all utterances and fitted three models in R, each having the acoustic value as the dependent variable, Accent (unaccented, prenuclear L*+H – henceforth pL*+H, – H*, L+H*, L*, L*+H) as the fixed effect and Item (the utterance) as random intercept. Results from multiple comparisons are shown in Tables 1 to 3 and plotted in Fig. 3. The results for duration showed that unaccented words were shorter than accented ones, though not shorter than L*+Hs; among accent types, differences in duration were found between H*s and prenuclear accents and L*+Hs, with H*s being longer than both; additionally, L+H*s were longer than L*+Hs. In the case of $F0_{max}$, as expected, L* accents were the ones with the lowest values and they were significantly lower than all other categories, including unaccented words. Unaccented words also had low F0 max values, though not lower than H*s. The model for RMS showed that L+H* accents had the highest amplitude values; unaccented words did not have significantly lower amplitude than all accent types, but only L+H*s and pL*+H.

Overall, the results confirmed the arguments reported above for the analysis of target stimuli, showing that accented words are not necessarily more acoustically salient than unaccented words; additionally, it shows that different accents are not equally salient from an acoustic perspective.

Table 1 - *Pairwise comparison for the effect of Accent on duration*

contrast	*estimate*	*SE*	*t*	*p*
unaccented – pL+H*	-0.54	0.14	-3.82	<.01
*unaccented – H**	-1.30	0.24	-5.50	<.0001
*unaccented – L+H**	-1.14	0.19	-5.96	<.0001
*unaccented – L**	-0.93	0.19	-4.87	<.0001
unaccented – L+H*	0.05	0.37	0.13	n.s.
pL+H – H**	-0.76	0.25	-3.00	<.05
pL+H – L+H**	-0.60	0.21	-2.82	n.s.
pL+H – L**	-0.39	0.21	-1.83	n.s.
pL+H – L*+H*	0.59	0.38	1.55	n.s.
H – L+H**	0.16	0.28	0.57	n.s.
H – L**	0.37	0.28	1.30	n.s.
H – L*+H*	1.35	0.43	3.16	<.05
L+H – L**	0.21	0.25	0.84	n.s.
L+H – L*+H*	1.19	0.40	2.94	<.05
L – L*+H*	0.98	0.40	2.43	n.s.

Table 2 - *Pairwise comparison for the effect of Accent on F0 max*

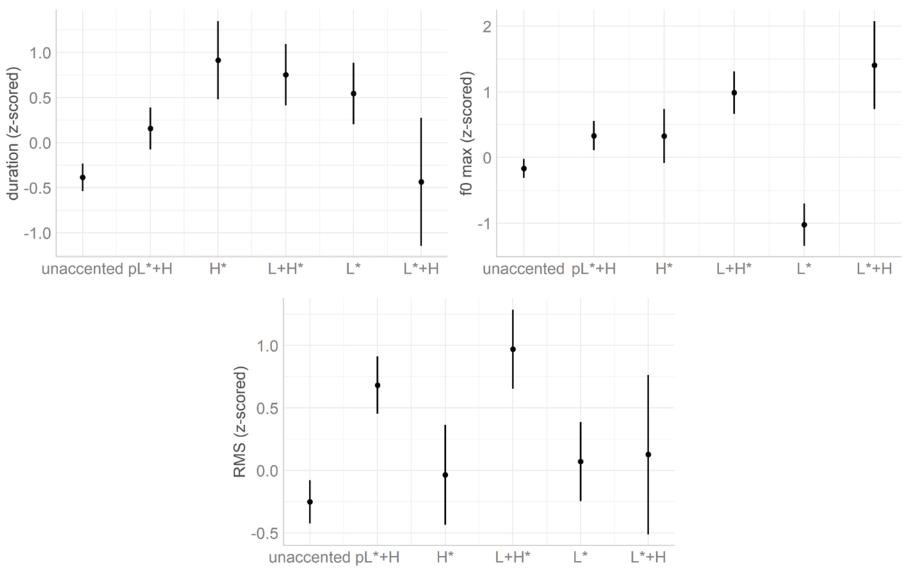
contrast	*estimate*	*SE*	*t*	*p*
unaccented – pL+H*	-0.50	0.14	-3.70	<.01
*unaccented – H**	-0.49	0.22	-2.22	n.s.
*unaccented – L+H**	-1.15	0.18	-6.37	<.0001
*unaccented – L**	0.85	0.18	4.72	<.001
unaccented – L+H*	-1.57	0.35	-4.47	<.001
pL+H – H**	0.00	0.24	0.02	n.s.
pL+H – L+H**	-0.65	0.20	-3.27	<.05
pL+H – L**	1.35	0.20	6.78	<.0001
pL+H – L*+H*	-1.07	0.36	-2.96	<.05
H – L+H**	-0.66	0.27	-2.46	n.s.
H – L**	1.35	0.27	5.04	<.0001

contrast	*estimate*	*SE*	*t*	*p*
H – L*+H*	-1.08	0.40	-2.67	n.s.
L+H – L**	2.01	0.23	8.59	<.0001
L+H – L*+H*	-0.42	0.38	-1.10	n.s.
L – L*+H*	-2.43	0.38	-6.35	<.0001

Table 3 - *Pairwise comparison for the effect of Accent on RMS amplitude*

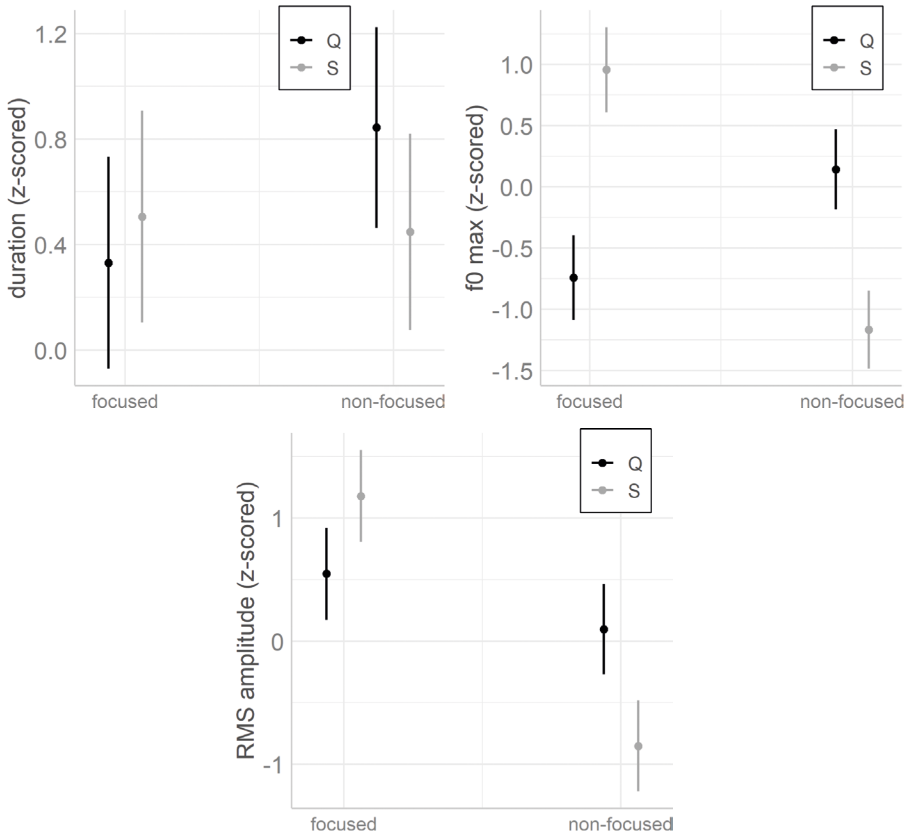
contrast	*estimate*	*SE*	*t*	*p*
unaccented – pL+H*	-0.94	0.13	-7.46	<.0001
*unaccented – H**	-0.22	0.21	-1.03	n.s.
*unaccented – L+H**	-1.22	0.17	-7.31	<.0001
*unaccented – L**	-0.32	0.17	-1.92	n.s.
unaccented – L+H*	-0.38	0.33	-1.17	n.s.
pL+H – H**	0.72	0.22	3.35	<.05
pL+H – L+H**	-0.29	0.19	-1.55	n.s.
pL+H – L**	0.61	0.19	3.31	<.05
pL+H – L*+H*	0.56	0.34	1.63	n.s.
H – L+H**	-1.01	0.26	-3.93	<.01
H – L**	-0.11	0.26	-0.41	n.s.
H – L*+H*	-0.16	0.38	-0.43	n.s.
L+H – L**	0.90	0.22	4.03	<.01
L+H – L*+H*	0.84	0.36	2.35	n.s.
L – L*+H*	-0.06	0.36	-0.16	n.s.

Figure 3 - *Duration, $F0_{max}$, and RMS amplitude values as a function of presence and type of accent, See Tables 1, 2, and 3 for information about statistical significance*



An additional set of analyses was performed considering only narrow focus statements and polar questions with non-final focus. As mentioned in § 1.3, these two utterance types represent a clear case of match (statements) and mismatch (questions) between metrical strength and acoustic salience, therefore they can provide crucial insights for the understanding of prominence. Their acoustic analysis was done by extracting duration, $F0_{max}$, and Root Mean Square (RMS) amplitude from the stressed syllable of the word in focus (the word bearing the nuclear accent) and the stressed syllable of the word in final position. The values of these measures were z-scored and used as the dependent variables in individual Linear Mixed-Effect Models (LMMs) in R (R Core Team, 2021); all three models had Focus (Focused, Non-focused), Modality (Question, Statement), and their interaction as fixed effects and item as random intercept; we report the model output for the interaction effect, which is the most informative in relation to our design. The models showed that duration was not affected by Focus or Modality (Focus × Modality: Est. = .57 (.39), $t = 1.46$, $p > .1$; see Fig. 4). For $F0_{max}$, the interaction between Focus and Modality was significant (Est. = 3.01 (.34), $t = 8.81$, $p < .0001$); specifically, a pairwise comparison showed that, in statements, the focused word had higher F0 max than the non-focused word (Est. = 2.12 (.24), $t = 8.76$, $p < .0001$), while in questions the focused words had lower $F0_{max}$ than their non-focused counterparts (Est. = .88 (.24), $t =3.62$, $p < .001$), see the middle panel in Fig. 3. Finally, there was a significant interaction effect for amplitude (Est. = 1.58 (.31), $t = 5.15$, $p < .0001$): amplitude was higher in focused words, but the effect was larger in statements than questions (Est. = 2.03 (.22), $t = 9.27$, $p < .0001$ and Est. = .45 (.22), $t = 2.08$, $p < .05$, respectively); see Fig. 4. In sum, these analyses confirm the prediction that in polar questions with non-final focus, acoustic salience and metrical strength do not match.

Figure 4 - *Duration,* $F0_{max}$*, and RMS amplitude as a function of focus × modality. All comparisons were significant except for duration*



2.4 Statistical analysis of the RPT responses

The RPT responses were analyzed using Generalized Linear Mixed-Effect Models (GLMMs) in R with the packages *lme4* (for the model fitting, Bates, Maechler, Bolker & Walker, 2015) and *emmeans* (for pairwise comparisons, Lenth, 2023). The analysis was conducted by running two models. The first model tested the binary RPT responses (whether a word was selected or not as prominent) as a function of Phonological status of words: unaccented, prenuclear accented, nuclear accented. The second model tested the binary responses as a function of Nuclear accent type: H* (broad focus statements), L+H* (narrow focus statements), L* (polar questions), L*+H (wh-questions). The random structure in both models included random intercepts for Speaker and Item.

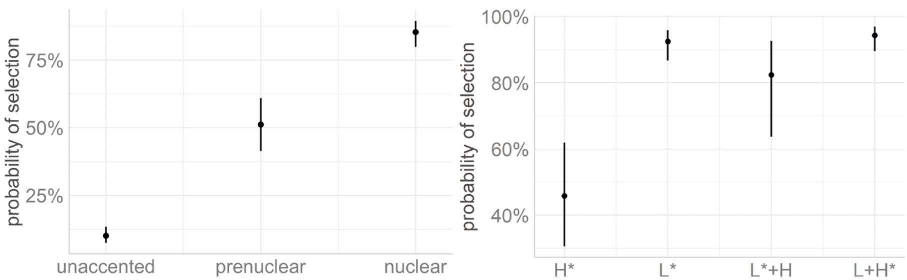
We additionally ran a third model including only statements and polar questions with non-final focus, as they represented clear cases of matches and mismatches between metrical strength and acoustic salience. For this analysis we only included words bearing the nuclear pitch accent and those bearing the edge tones. Specifically, we analyzed responses for nuclear accented (L+H* and L*) and final words (L-L % and H-L %) in these two utterance types. The model was built with the binary response as dependent variable, and Modality (Statement, Question), Focus status (Focused, Non-focused), and their interaction as fixed effects; the random structure included random intercepts for Speaker and Item.

3. *Results*

The results of two GLMMs fitted for the aggregate RPT responses largely met our predictions. The first model tested the selection of a word as a function of the phonological status of the word (i.e. unaccented, prenuclear accented, nuclear accented) and showed that unaccented words are less likely to be marked as prominent than both prenuclear accented words (Est. = 2.18 (.20), $z = 11.16$, $p < .0001$) and nuclear-accented words (Est. = 3.83 (.19), $z = 20.01$, $p < .0001$). Additionally, nuclear-accented words were more likely to be selected than words bearing prenuclear accents (Est. = 1.64 (.22), $z = 7.58$, $p < .0001$).

The second model tested the effect of accent type in nuclear position. The only accent that was rated significantly different than the others was H*, the nuclear accent in broad focus statements. Specifically, H*-bearing words were less likely to be selected as prominent than words accented with L* (Est. = 2.66 (.37), $z = 7.20$, $p < .0001$), L*+H (Est. = 1.71 (.54), $z = 3.20$, $p < .01$), and L+H* (Est. = 2.98 (.39), $z = 7.69$, $p < .0001$). There were no differences among the other accents L* – L*+H: Est. = .96 (.52), $z = 1.84$, $p > .1$; L* – L+H*: Est. = .30 (.35), $z = .85$, $p > .1$; L*+H – L+H*: Est. = 1.26 (.53), $z = 2.37$, $p > .05$). The outputs of the two models are plotted in Fig. 5.

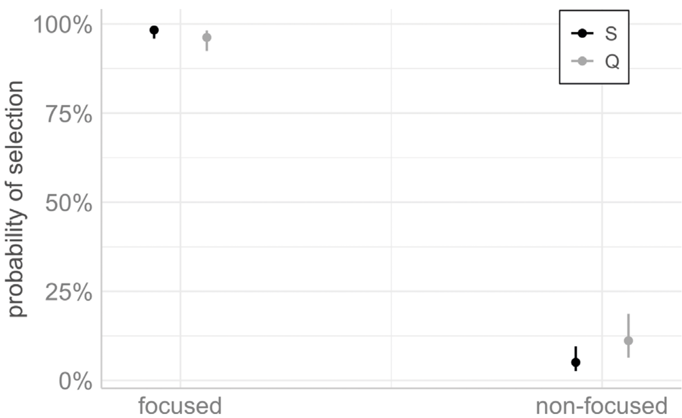
Figure 5 - *The left panel shows the probability of selection as predicted as a function of phonological status; all comparisons were significant; the right panel shows the same probability as a function of nuclear accent type; only H* in comparison with all other accents was significant, the other comparisons were not*



3.1 Statements and polar questions with non-final focus

The GLMM fitted for the statements and questions with non-final focus showed that, in line with our predictions, focused words were more likely to be selected than non-focused words, regardless of modality (for Statements: Est. = 6.98 (.49), $z = 14.18$, $p < .0001$; for Questions: Est. = 5.23 (.37), $z = 14.21$, $p < .0001$). The interaction between focus status and modality was significant (Est. = 1.75 (.54), $z = 3.27$, $p < .01$), and the pairwise comparison showed that the responses only differed for unaccented words in questions (bearing H-L %) vs statements (bearing L-L %), with higher probability of selection in questions than statements (Est. = .94 (.28), $z = 3.31$, $p < .001$). Crucially, no difference was found between L* and L+H* (Est. = .81 (.45), $z = 1.80$, $p > .05$). These effects are shown in Fig. 6.

Figure 6 - *Probability of selection as predicted by the GLMM for focused and non-focused words in questions (black) and statements (grey). All comparisons were statistically significant, except for the difference between Q and S in the focused condition*



4. *Discussion*

We reported an RPT study conducted on Greek with the aim of investigating the relationship between prominence assessment and phonological information linked to metrical structure. We hypothesized that the selection of a word as prominent would be determined by its metrical status and not by the phonetic shape of the accent it bore or how salient it sounded. To test this hypothesis, we designed a set of stimuli comprising both matching and mismatching cases between metrical strength and acoustic salience.

In general, our predictions were largely met. RPT responses varied as a function of the presence of pitch accent rather than specific acoustic cues. The general analysis showed that the likelihood of selecting a word aligned with the metrical organization of the utterances: unaccented words were the least likely to be selected, while nuclear-accented words were the most likely, with prenuclear-accented words falling in between. This outcome, accompanied by the fact that unaccented words were not necessarily shorter, less loud, or lower in pitch than accented ones (see § 2.3), supports our predictions on prominence selection.

When looking at statements and questions with non-final focus, this result becomes even clearer. We showed that in polar questions with non-final focus, the word in focus is acoustically less salient than the final word, as opposed to statements, in which salience and metrical strength overlap. Specifically, the L+H*s in statements were always followed by a deaccented and less salient (much lower in pitch and amplitude) word, while the L* accents in questions were always followed by an unaccented word that was of comparable duration and amplitude but all had much higher pitch. The RPT results for questions, however, showed that the nuclear-accented words were the most likely to be selected in the utterance, despite not being the most salient; in contrast, the final word bearing the H-L % edge was much less likely to be selected than the accented words, despite being acoustically more salient. By and large, the size of the effect of focus condition on the probability of word selection was the same for the two modalities (see § 3.1 and Fig. 6), indicating that neither the presence nor the magnitude of the effect changed as a function of differences in acoustic characteristics between accented and unaccented words.

This shows that when Greek listeners assessed the prominences of these utterances, they did not select the words that stood out by virtue of being acoustically more salient, but interpreted the role of the words in the metrical organization of the utterance. As reported in § 1.1, similar results are reported in Rump and Collier (1996), in which Dutch listeners interpreted the focus of an utterance by considering the metrical strength of the words rather than their relative F0 height. This is in line with our view of prominence as determined by metrical strength rather than acoustic salience. This is also compatible with previous studies showing that the interpretation of prominence is mediated by the phonological representation of that utterance (cf. Ladd, Arvaniti, 2023).

Another important result is that L*s had the same probability of selection as L+H*s and that, more generally, the phonetic shape of a pitch accent was not an important drive for prominence selection (contrary to results reported in, among others, Baumann, Winter, 2018, and Cole et al., 2019). Indeed, despite differences in amplitude, F0 height, and duration, the effect of pitch accent type was very weak: the pairwise comparison showed that all nuclear accents were highly likely to be selected, with the only exception being H*. As a matter of fact, the likelihood for H* to be selected as prominent was the lowest among all accents, including the prenuclear L*+H. Though unexpected, this result is not entirely incompatible with our view of prominence. One possible explanation for this outcome could be related to the function that these accents have in the discourse. As reported in § 1.2, H* is the nuclear accent in broad focus statements, where the whole utterance represents new information conveyed to the addressee. In some frameworks, such utterances are considered to lack focus altogether (Katz, Selkirk, 2011). This represents a crucial difference between H*s and other accents. While accents like

L* or L+H* indicate narrow focus, narrowing down the focus to the accented word, H* accents have the function of presenting every word in the utterance as equally important. Furthermore, the fact that H* accents were not less salient than other accents also shows that their low likelihood of being selected does not simply depend on information related to the signal and is therefore linked to the way those accents are processed at a higher-order level. Such mechanisms involving the meaning of the accents have also been found in previous studies investigating other languages. Orrico et al. (2025) found that H* accents are less prominent than L+H* accents in British English, with a clear difference between the two observed only when the former had a non-contrastive function and the latter a contrastive function, confirming the importance of cues beyond the acoustic details. Similarly, D'Imperio's (1997) investigation of prominence perception in Neapolitan Italian showed that in narrow focus statements, the word bearing the nuclear pitch accent was consistently interpreted as the most prominent word in the utterance, but the interpretation of prominence in broad focus statements was around chance level, which is consistent with our results. Crucially, the fact that these results about prominence perception in Greek relate to results collected for other languages suggests that these mechanisms are not specific to Greek, but are part of prominence perception across languages.

We also found that acoustic salience did matter to some extent: a slightly higher probability of selection was observed for unaccented words with a H- phrase accent. Both the size of this effect and the fact that it was observed only for unaccented words show that the role of acoustic cues is subordinate to that of metrical strength. Additionally, this effect might also be explained as epiphenomenal, linked to methodological facts, like the type of stimuli used or the type of task we adopted. As mentioned in the introduction, RPT responses can be influenced by task details. Orrico et al. (2024) showed that when participants were asked to select *all* prominent words (as in the current study), the effect of acoustic variables (such as F0 height) was larger than when the participants were asked to select only the most prominent word. The fact that the effect of these acoustic variables is small and susceptible to changes in the task suggests that they are not the main drive for prominence selection.

In conclusion, the investigation reported here, following the lead of Ladd and Arvaniti (2023), provides empirical evidence in support of a view of prominence that considers acoustic elements of utterances as cues to prominence only by virtue of the fact that they are part of well-formed intonation contours, associated to the text according to language-specific metrical rules that ultimately define which words are prominent.

Acknowledgements

This work was supported by the European Research Council through grant ERC-ADG-835263 (SPRINT) to Amalia Arvaniti. We thank our participants for voluntarily participating in the study, Vasiliki Kyritsi and Vasiliki Barouta for their assistance with data collection, and Bob Ladd and Jennifer Cole for many helpful discussions on theoretical and methodological aspects linked to prominence.

References

Arvaniti, A., Baltazani, M. (2005). Intonational analysis and prosodic annotation of Greek spoken corpora In Jun, S-A. (Eds.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford University Press. 84-117.

Arvaniti, A., Gryllia. S, Zhang, C. & Marcoux, K. (2022). Disentangling emphasis from pragmatic contrastivity in the English H* ~ L+H* contrast. *Proc. Speech Prosody* 2022, 837-841, doi: 10.21437/SpeechProsody.2022-170

Arvaniti, A., Ladd, D.R. & Mennen, I. (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics*, 26: 3-25.

Arvaniti, A., Ladd, D.R. & Mennen, I. (2006). Tonal association and tonal alignment: evidence from Greek polar questions and contrastive statements. *Language and Speech*, 49: 421-450.

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

Baltazani, M. (2002). Quantifier scope and the role of intonation in Greek. Unpublished Ph.D. dissertation, University of California, Los Angeles.

Baltazani, M. (2003) Broad Focus across sentence types in Greek. *Proceedings of Eurospeech-2003*, Geneva, Switzerland.

Baltazani, M., Gryllia, S. & Arvaniti, A. (2020). The Intonation and Pragmatics of Greek wh-Questions. *Language and Speech*, 63 (1), 56-94. doi: 10.1177/0023830918823236

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01

Baumann, S., Winter, B. (2018). What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *Journal of Phonetics*, 70, 20-38.

Bishop, J., Kuo, G. & Kim, B. (2020). Phonology, phonetics, and signal-extrinsic factors in the perception of prosodic prominence: Evidence from Rapid Prosody Transcription. *Journal of Phonetics*, 82, 100977.

Boersma, P., Weenink, D. (2023). Praat: doing phonetics by computer [Computer program]. Version 6.3.13, retrieved 31 July 2023 from http://www.praat.org/

Cole, J., Hualde, J.I., Smith, C.L., Eager, C., Mahrt, T. & de Souza, R.N. (2019). Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish. *Journal of Phonetics*, 75, 113-147. DOI: https://doi.org/10.1016/j.wocn.2019.05.002

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Cole, J., Mahrt, T. & Hualde, J.I. (2014). Listening for sound, listening for meaning: Task effects on prosodic transcription. In *Proceedings of Speech Prosody*, Dublin, Ireland 20-23 May 2014, 859-863.

Cole, J., Mo, Y. & Baek, S. (2010a). The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language and Cognitive Processes* 25 (7):1141. DOI: http://dx.doi.org/10.1080/01690960903525507

Cole, J., Mo, Y. & Hasegawa-Johnson, M. (2010b). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, 1(2), 425-452.

Cole, J., Shattuck-Hufnagel, S. (2016). New methods for prosodic transcription: Capturing variability as a source of information. *Laboratory Phonology*, 7(1).

D'Imperio, M. (1997). Breadth of focus, modality, and prominence perception in Neapolitan Italian. In Ainsworth-Darnell, K., D'Imperio, M. (Eds.) *Working papers in linguistics:* Volume 50, 19-39.

D'Imperio, M. (2000). Acoustic-perceptual Correlates of Sentence Prominence in Italian. In Muller, J.S., Huang, T. & Roberts, C. (Eds.) Working papers in linguistics: Volume 54, 59-77.

Georgakopoulos T., Skopeteas S. (2010). Projective versus interpretational properties of nuclear accents and the phonology of contrastive focus in *Greek. Linguistic Review*, 27(3), 319–346.

Grice, M., Ladd, D.R. & Arvaniti, A. (2000). On the place of phrase accents in intonational phonology. *Phonology*, 17(2), 143-185.

Im, S., Cole, J. & Baumann, S. (2023). Standing out in context: Prominence in the production and perception of public speech, *Laboratory Phonology*, 14(1). doi: https://doi.org/10.16995/labphon.6417

Katz, J., Selkirk, E. (2011). Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language*, 771-816.

Ladd, D.R. (2008). *Intonational phonology*. Cambridge University Press.

Ladd, D.R., Arvaniti, A. (2023). Prosodic prominence across languages. *Annual Review of Linguistics*, 9, 171-193.

Lenth, R. (2023). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.8.8, <https://CRAN.R-project.org/package=emmeans>

Lohfink, G., Katsika, A. & Arvaniti, A. (2019). Variability and category overlap in the realization of intonation. In Calhoun, S., Escudero, P., Tabain, M. & Warren, P. (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia 2019, 2946-2950

Orrico, R., Gryllia, S., Hu, N., Kim, J. & Arvaniti, A. (2024). Prosodic prominence in Greek: methodological and theoretical considerations. *Proc. Speech Prosody 2024*.

Orrico, R., Gryllia, S., Kim, J. & Arvaniti, A. (2023). The influence of empathy and autistic-like traits in prominence perception. In Skarnitzl, R., Volín J. (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 1280-1284). GUARANT International.

Orrico, R., Gryllia, S., Kim, J. & Arvaniti, A. (2025). Individual variability and the H* ~ L+H* contrast in English. *Language and Cognition*, 17, e9. doi: 10.1017/langcog.2024.62

Pierrehumbert, J.B. (1980). The phonology and phonetics of English intonation (Doctoral dissertation, Massachusetts Institute of Technology).

Riesberg, S., Kalbertodt, J., Baumann, S. & Himmelmann, N. (2020). Using Rapid Prosody Transcription to probe little-known prosodic systems: The case of Papuan Malay, *Laboratory Phonology*, 11(1): 8. doi: https://doi.org/10.5334/labphon.192

Wagner, P., Origlia, A., Avesani, C., Christodoulides, G., Cutugno, F., D'Imperio, M., Escudero, D., Gili Fivela, B., Lacheret, A., Ludusan, B., Moniz, H., Chasaide, A., Niebuhr, O., Rousier-Vercruyssen, L., Simon, A., Šimkz, J., Tesser, F. & Vainio, M. (2015). Different Parts of the Same Elephant: A Roadmap to Disentangle and Connect Different Perspectives on Prosodic Prominence. In The Scottish Consortium for ICPhS 2015 (Ed.), Proceedings of the 18th International Congress of Phonetic Sciences. Glasgow, UK: the University of Glasgow (10th-14th of August 2015). ISBN 978-0-85261-941-4. Paper number 0202.1-5

Welby, P. (2006). French intonational structure: Evidence from tonal alignment. *Journal of Phonetics*, 34(3), 343-371.