Corresponding Author: Mary Baltazani, University of Oxford, 41 Wellington Square, Oxford, OX1 2JF, UK. Email: mary.baltazani@ling-phil.ox.ac.uk

The intonation and pragmatics of Greek wh-questions

Mary Baltazani[1] Stella Gryllia[2] Amalia Arvaniti[3]

[1] University of Oxford, Oxford, UK

[2] University of Leiden, Leiden, The Netherlands

[3] University of Kent, Canterbury, UK

Abstract

We experimentally tested three hypotheses regarding the pragmatics of two tunes (one high-ending, one flat-ending) used with Greek wh-questions: (a) the high-ending tune is associated with information-seeking questions, while the flat-ending tune is *also* appropriate when wh-questions are not information-seeking in which case their function can instead be akin to that of a statement; (b) the high-ending tune is more polite, and (c) more appropriate for contexts leading to information-seeking questions. The wh-questions used as experimental stimuli were elicited from four speakers in contexts likely to lead to either information-seeking or non-information-seeking uses. The speakers produced distinct tunes in response to the contexts; acoustic analysis indicates these are best analysed as L*+H L-!H% (rising), and L+H* L-L% (flat). In a perception experiment where participants heard the questions out of context, they chose answers providing information significantly more frequently after high-ending than flat-ending questions, confirming hypothesis (a). In a second experiment testing hypotheses (b) and (c), participants evaluated wh-questions for appropriateness and politeness in information- and non-information-seeking contexts. High-ending questions were rated more appropriate in information-seeking contexts, and more polite independently of context relative to their flat-ending counterparts. Finally, two follow-up experiments showed that the interpretation of the two tunes was not affected by voice characteristics of individual speakers, and confirmed a participant preference for the high-ending tune. Overall, the results support our hypotheses and lead to a compositional analysis of the meaning of the two tunes, while also showing that intonational meaning is determined by both tune and pragmatic context.

Keywords

Intonation; pragmatics; Greek; wh-questions; politeness; gender

Introduction

Research on intonation, especially within the autosegmental-metrical framework of intonational phonology (AM), is by now quite extensive and encompassing an increasingly large number of languages (Jun, 2005, 2014; for reviews see e.g., Gussenhoven, 2004; Ladd, 2008; Arvaniti, in press). Intonational meaning, however, has not received as much attention as the phonetics and phonology of intonation (but see Gunlogson, 2003; Vanrell, Mascaró, Torres-Tamarit, & Prieto, 2013; Armstrong & Prieto, 2015; Brown & Prieto, 2017). Here we provide evidence that contributes to

this new understanding of the importance of studying the pragmatics of intonation alongside its phonology and phonetics. Specifically, we report the results of two perception experiments and the phonetic analysis of the stimuli; together, the three studies and two additional follow-up experiments address the phonetics, phonology, and pragmatics of two tunes that are used with wh-questions in Greek, allowing speakers to employ such questions for different pragmatic purposes.

Possibly the most important reason why intonational meaning has not received sufficient attention so far has to do with the lack of consensus on what intonational meaning involves. For those studying intonation from the perspective of semantics and syntax, the emphasis is on accentuation and its role in encoding information structure (among many, Selkirk, 1984; Féry & Kügler, 2008; Büring, 2012); for example pitch accents can be analysed as marking the words that carry them as new information (for Greek see Baltazani, 2002, 2006; Gryllia, 2009a, b). Studies of intonation from the perspective of psycholinguistics and sentence processing, concentrate on semantic properties like scope relations (e.g., Martí, 2001; Baltazani, 2002, 2006; Huang & Snedeker, 2009) or the relation between prosodic structure and sentence or discourse processing (e.g., Schafer, Speer, Warren, & White, 2000; Carlson, Clifton & Frazier, 2001; Hwang & Schafer, 2009).

In AM, on the other hand, the focus has been on the role of intonation in conveying pragmatic meaning. For example, Pierrehumbert and Hirschberg (1990), following the tenets of AM, consider tunes to be composed of distinct elements, pitch accents, phrase accents, and boundary tones, all of which contribute to the meaning of an utterance: each primitive in a tune is a *morpheme* that encodes pragmatic meaning. For example, in Pierrehumbert and Hirschberg (1990: 291), a L* accent "marks items that S [the speaker] intends to be salient but not to form part of what S is predicating in the utterance". Edge tones (in English) have been analysed as conveying speaker or hearer commitment to the truth of the proposition: a falling declarative conveys the speaker's commitment while a rising declarative shifts commitment to the hearer (e.g., Gunlogson, 2003). In combination, the elements of a tune allow speakers to specify a particular relationship between the propositional content of their utterance and the *mutual beliefs* of the discourse participants (beliefs participants arrive at as a result of the conversational interaction). Thus, intonational meaning goes beyond the encoding of information structure and allows interlocutors to contribute to and alter mutual beliefs (see also Steedman, 2007, 2014; Büring, 2012, 2016; Portes et al., 2014).

Crucially, in this view meaning is the product of the tune's morphemes in combination with the utterance's lexical and propositional content and, importantly, the pragmatic context in which the utterance is realized. As Pierrehumbert and Hirschberg put it "S's beliefs are *not* specified by choice of tune—the 'declarative' contour H* L-L%, for example, will *not* be translated as *S believes x*. But S's belief in *x* **may be inferred** from the combined meanings of pitch accents, phrase accents, and boundary tone, **as they are used in particular contexts** [emphasis in bold added]" (Pierrehumbert & Hirschberg, 1990: 286). This is a crucial difference between AM approaches to intonational meaning and other frameworks in which tunes are often treated as holistically conveying *communicative functions*, such as question or statement (e.g., Xu, 2005) or *attitudes*, such as interest, excitement, or lack of certainty (e.g., Cruttenden, 1997, ch. 4). Treating tunes as gestalts and linking them to

particular functions or attitudes can account neither for the fact that tunes may convey radically different meanings when used with different utterances in different contexts, nor for the commonalities that such uses evince (see Arvaniti, in press, for examples and further discussion). At the same time, these holistic approaches emphasize the role of the speaker and ignore the role of the addressee who must infer the speaker's intended meaning and may not always be successful at doing so (cf. Elder, 2017).

The results we report here support these AM tenets regarding intonational meaning and the importance of context in interpreting tunes. They further show that additional inferences can be drawn from intonation, for example, including inferences related to politeness (cf. Brown & Prieto, 2017; Astruc, Vanrell, & Prieto, 2016; Vanrell, Mascaró, Torres-Tamarit, & Prieto, 2013). Finally, our data show that the role of the addressee is paramount in understanding intonational meaning because these inferences are not totally deterministic: it is possible for the addressee to disregard the contribution of the tune and base their interpretation on pragmatic factors and the propositional content of the question.

1.1 Greek wh-questions: Background

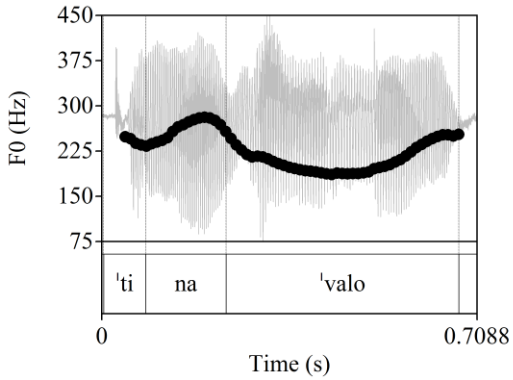
Greek wh-questions are marked both morphologically and syntactically as such: Greek is a wh-fronting language that is, wh-words are utterance-initial, as shown in examples (1) and (2) below (Agouraki, 1990; Tsimpli, 1995; Anagnostopoulou, 1999; Kotzoglou, 2005). Greek does have some marked constructions where the wh-word can remain in situ (Sinopoulou, 2008; Vlachos, 2010; Alexopoulou & Baltazani, 2012); these constructions, however, are rare.

(1) [ˈpços aˈɣorase to viˈvlio]
who.NOM bought.3RD.SG the book.ACC
"Who bought the book?"

(2) [ˈti aˈɣorase o ˈnikos]
what.ACC bought.3RD.SG the nikos.NOM
"What did Nikos buy?"

Wh-questions are typically uttered with the tune illustrated in Figure 1 with two utterances of different length: the contour consists of a rise associated with the stressed syllable of the wh-word, followed by a dip or low F0 stretch (depending on the length of the question), and ends with a small rise. In AM terms, this tune has been analysed as L*+H L-!H%, i.e., as consisting of a L*+H pitch accent associated with the wh-word, a L- phrase accent of variable realization depending on utterance-length, and a downstepped !H% boundary tone realized on the last vowel of the question and rising roughly to the middle of the speaker's range (among others, Baltazani, 2002, 2003; Arvaniti & Baltazani, 2005; Arvaniti & Ladd, 2009; Grice, Ladd, & Arvaniti, 2000). As Figure 1 indicates, the tune has only one pitch accent independently of utterance-length; this (nuclear) pitch accent always associates with the wh-word (Arvaniti & Ladd, 2009).

(a)



(b)

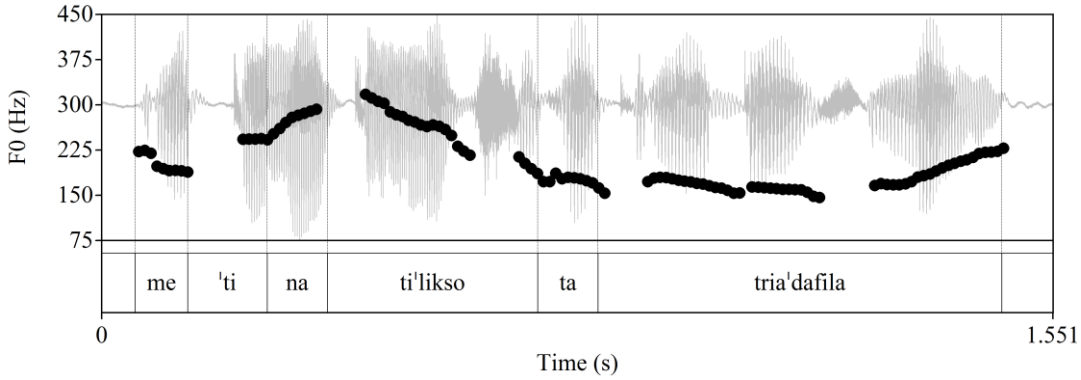
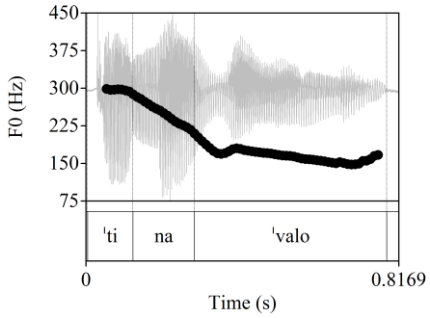


Figure 1: Waveforms and F0 contours of two wh-questions, [ˈti na ˈvalo] “what should I wear?” in panel (a), and [meˈti na tiˈlikso ta triaˈdafila] “what should I wrap the roses with?” in panel (b), both uttered with the high-ending tune.

Previous studies of wh-questions in Greek note that the final pitch rise is optional and that wh-questions can instead end with flat low F0 (Arvaniti & Baltazani, 2005; Arvaniti & Ladd, 2009). Though full discussions of this flat-ending tune are not available, Arvaniti & Baltazani (2005) and Arvaniti & Ladd (2009) assume that it differs from the high-ending tune only in terms of its boundary tone; thus, they represent it as L*+H L-L%, with the wh-word carrying the only pitch accent, L*+H, as in the high-ending tune. This flat-ending tune is illustrated in Figure 2 where it can be observed that, contrary to the assumption of previous studies, the onset of the tune is not the same as in Figure 1: in panel (a) where the question starts with the stressed syllable of the wh-word [ˈti] “what”, the contour in Figure 1(a) shows a short rise to a late peak, while the contour in Figure 2(a) starts high; in panel (b) where the question starts with an unstressed syllable [meˈti] “with what”, the contour in Figure 1(a) starts low and rises to a late peak, while that in Figure 2(b) shows a curtailed rise from roughly the middle of the speaker’s range, with the pitch peak co-occurring with the wh-word’s stressed vowel. These differences in the realization of the rise and the alignment of the peak indicate that the pitch accent of the flat-ending tune may not be L*+H as assumed in earlier work. We return to this point in section 2, and more extensively in section 6.1.

(a)



(b)

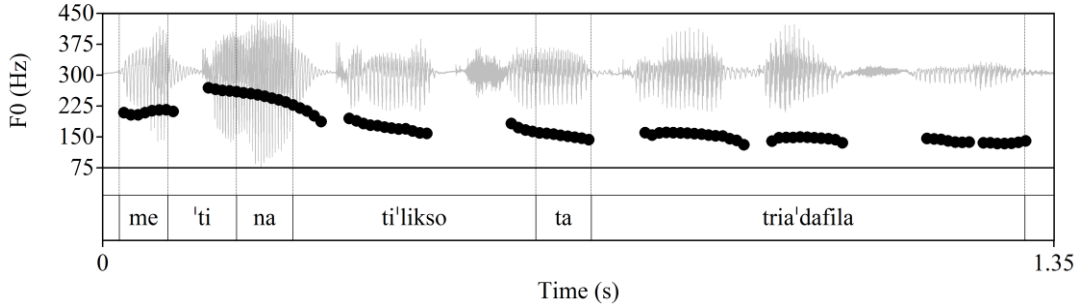


Figure 2: Waveforms and F0 contour of two wh-questions, [ˈti na ˈvalo] "what should I wear?" in panel (a), and [meˈti na tiˈlikso ta triaˈdafila] "what should I wrap the roses with?" in panel (b), both uttered with the flat-ending tune.

Existing descriptions of wh-question intonation in Greek provide limited comments on the meaning and use of the two tunes. Arvaniti and Ladd (2009) only note that the flat-ending tune was occasionally used by male participants in their study (cf. Clopper & Smiljanic, 2011, on gendered tune frequency). Arvaniti and Baltazani (2005: 95) say that both tunes are used for wh-questions and describe the high-ending tune as "involved".

Our own intuitions as native speakers of Standard Greek together with the above observations in previous work suggest the following regarding the differences in interpretation between the two tunes. The high-ending tune (represented above as L*+H L-!H%, following Arvaniti & Ladd, 2009) is suitable for a variety of contexts in which a wh-question is used to elicit information from the addressee, what we will henceforth call *information-seeking contexts* and *information-seeking questions*. The flat-ending tune is also appropriate in this context, since the question is overtly marked as such and questions with this tune are used in Greek to seek information. However, it may be seen as a less polite tune for asking a question. For instance, a typical question such as [ˈpos se ˈlene] "what's your name?" uttered with the flat-ending tune would be appropriate if the speaker is a policeman interrogating a suspect, but not if she is a kindergarten teacher asking her new students to introduce themselves. If correct in our assumptions about this difference, the two tunes should be rated differently with respect to politeness (for details see below).

In addition to this politeness-related difference between the two tunes when used with information-seeking wh-questions, the flat-ending tune can also be used in Greek in situations where a wh-question need not function as a question per se: specifically,

when a wh-question is uttered using the flat-ending tune, it can be used simply as a means of eliciting information from the addressee (i.e. as an information-seeking question), or it may lead to additional inferences that make its function akin to that of a statement. The latter use means that a wh-question uttered with the flat-ending tune can be *non-information-seeking*; instead, it can serve or be interpreted as being implicitly a statement, typically one with a negative flavour. How a flat-ending wh-question is interpreted in a given situation – as information-seeking or non-information-seeking – depends on the context and inferences the addressee can draw from it. We provide an example below to illustrate this point, the investigation of which is the focus of the research reported here. When a teenager arrives home after curfew, an irate parent may express her disapproval by using a wh-question: [ˈti ˈora ˈine afˈti] "what time is this?". This question can only be uttered with the flat-ending tune if it is to convey the parent's disapproval of the teenager's late return. In English, the same force could be conveyed by a question like *what time do you call this?*. In turn, the teenager who knows he's late will have no doubt that he is not being asked to tell the time, even though he hears an utterance that is grammatically a question. Now, our teenager can and may choose to take the flat-ending question as information-seeking and respond by telling the time, since the utterance is grammatically a question; doing so, however, would be a breach of convention and would only be interpreted as cheeky, not helpful.

This non-informational use of wh-questions with the flat-ending tune is similar but not identical to rhetorical questions; following Caponigro and Sprouse (2007), we assume that rhetorical questions have an obvious answer known to both speaker and addressee. The Greek wh-questions discussed here do not have such an obvious answer; for example, if a speaker utters [ˈpu ˈine ta buˈfan] "where are the coats?" with a flat-ending tune, there is no answer known to them. Rather, there are two possible interpretations: (a) the question may be information-seeking, in which case the speaker simply conveys that they do not know where the coats are and wants to find out; (b) the question is used as an indirect way of conveying the speaker's annoyance at not finding the coats where she expected them to be (and carries a whiff of an accusation that the addressee many be responsible for this state of affairs). As in the previous example, in English this meaning would be expressed by a marked question like *What have you done with the coats?*. Note that in this instance reaching interpretation (b) does not stop the addressee from responding with information, i.e. interpreting the question as both information-seeking and as implying that something is wrong with the location of the coats. We return to this point in section 6.1.

As mentioned, the above description of the tunes' pragmatics is based on observation, intuition and the limited previous data on the matter; it lacks, however, solid empirical evidence provided by a large body of native speakers without training in intonation. The studies reported here address this need: we examined the pragmatic interpretation and evaluation of the two tunes by means of two perception experiments; further, in order to ensure our stimuli fell into distinct melodic categories based on pragmatic context, we acoustically analysed the small corpus of questions from which the stimuli were drawn and briefly present the results.

Based on the above description of the two tunes, we predicted the following. In terms of realization, we anticipated that the two tunes would differ both in terms of the boundary tone (!H% vs. L%) but also in terms of the pitch accent on the wh-word:

previous analyses show that the pitch accent in the high-ending tune is L*+H (Arvaniti & Ladd, 2009); utterances like those in Figure 2 indicate that the pitch accent of the flat-ending tune lacks the extended rise from a low F0 point and has an early peak, characteristics consistent with a L+H* pitch accent instead (Arvaniti, Ladd, & Mennen, 2006). In terms of pragmatics, we anticipated that listeners would be more likely to interpret questions with high-ending tunes as information-seeking, while they would be more likely to make additional inferences about questions with flat-ending tunes and thus interpret those as non-information-seeking. Because of this difference in interpretation we further expected that in information-seeking contexts listeners would find high-ending tunes more appropriate than flat-ending tunes. In contrast, we anticipated flat-ending tunes to be rated more appropriate in non-information-seeking contexts, and to be generally seen as less polite: in information-seeking contexts this would be because the flat-ending tune shows less involvement with the conversation; in non-information-seeking contexts, this would be because the tune leads to negative inferences, as shown above. In the remainder of the paper, we briefly present the essential differences in the realization of the two tunes (section 2) followed by the two perception experiments (sections 3 and 4); these are followed by two follow-up experiments that further probe the role of speaker in the interpretation of the tunes (section 5), and by a pragmatic analysis of the melodies (section 6).

2. Acoustic analysis and selection of the perception stimuli

We recorded a set of wh-questions in both information-seeking and non-information-seeking contexts, with the aim of selecting from among them the stimuli of the perception experiments. Before proceeding with stimulus selection we acoustically analysed this set of data to ensure that two distinct contours were produced in response to the two types of pragmatic contexts. We briefly report the results here for completeness focusing on elements that, based on previous research and our own observations, are critical for differentiating the two tunes; they confirm that our stimuli had distinct contours and thus were appropriate for the perception experiments (for a full-scale production study, see Gryllia, Baltazani, & Arvaniti, 2018).

2.1 Method

2.1.1 Speakers

The stimuli for the perception experiments were elicited from four speakers of Greek, two male and two female, between 30 and 48 years of age (average age = 44.75). Three of them (F1, F1, and M2) were speakers of the standard Athenian dialect; M1 had been raised in the North of Greece but had moved to Athens in his late teens and retained only some traits of his native accent. The male speakers were naïve as to the exact purposes of the experiment; the two female speakers are among the authors of this study. As our statistical analysis below indicates, there were no significant differences in the realization of the tunes based either on phonetic training or dialectal differences (see section 2.2.).

2.1.2 Materials and procedures

The recorded materials consisted of 16 wh-questions (see Appendix 1). The questions started with a number of different wh-words (e.g., [ˈti] “what”, [jaˈti] “why”), and

also varied in length and lexical makeup, as the primary aim of these recordings was to create pragmatically plausible stimuli for the perception experiments, not a balanced corpus typical of controlled production studies.

Since our aim was to elicit the same set of 16 wh-questions produced with both the high-ending and the flat-ending tune, each question was presented to the speakers in two contexts; one of the contexts (context A) was information-seeking, so designed to elicit the question with a high-ending tune, while the other (context B) could be interpreted as non-information-seeking, i.e., it was designed to elicit the flat-ending tune.˙ An example of the two contexts used with one of the questions is provided in (3) and (4); see also Appendix 7. Context A in (3) describes a situation in which a question would most plausibly be used to request information, namely a speaker asking for directions. Context B in (4) describes a situation in which a question could be non-information-seeking along the lines discussed in section 1.1: in this instance the speaker could use the question not to ask for directions, but instead to express annoyance and imply that it is impossible to access Syntagma because of the protest march (though of course the possibility that this context could be interpreted as information-seeking cannot be completely excluded: interactional pragmatics shows that no context will elicit only one possible response; cf. Bateson, 1972; Goodwin & Duranti, 1992). The design of the dialogues was based on previous research (e.g., Arvaniti & Baltazani, 2005) and our own native speaker intuitions (see section 1.1.).

(3) Context A: *Lena, who is visiting Athens for the first time, stops a passer-by for directions:*
Question: ['pos θa 'pao sto 'sidaɣma]
"How will I get to Syntagma?"

(4) Context B: *A protest march in Syntagma is scheduled for the time Kostas has an interview there; as they listen to the news, Kostas says to his wife:*
Question: ['pos θa 'pao sto 'sidaɣma]
"How will I get to Syntagma?"

Each context-question pair was written on a card. The cards were grouped into two sets, one with contexts designed to elicit the high-ending tune, and one with contexts designed to elicit the flat-ending version. The speakers were asked to consider the context and then read aloud the question in as natural a way as possible. For the male speakers, the differences between the two melodies were briefly explained; they did not have difficulties producing them naturally for the recordings. All speakers were recorded in a quiet room using a laptop and the facilities of Praat (Boersma & Weenink, 2016) at default settings (sampling rate 44100 Hz, 16-bit quantization). In total 128 stimuli were recorded (16 stimuli × 2 contexts × 4 speakers).

2.1.3 Measurements

The questions were acoustically analysed by measuring the scaling and, where appropriate, alignment of a number of tonal targets manually annotated in Praat. We followed a methodology similar to that of Arvaniti and Ladd (2009), marking the

tonal targets shown below.[1] The measurements are illustrated in Figure 3.

- *Accentual Low* (AL): the lowest F0 point at the onset of the wh-word's stressed vowel after which the pitch rise unambiguously began (such a low point was not always present; for details see section 2.2). AL was expected to be lower in the high-ending than the flat-ending tune.
- *Accentual High* (AH): the F0 maximum in the vicinity of the wh-word's stressed syllable; in cases of a high plateau the first point in the plateau was chosen. We measured both the scaling and temporal alignment of the AH. Alignment was defined as the distance of the H from the onset of the stressed vowel in the wh-word (SV). We expected AH to be aligned later in the high-ending than the flat-ending tune, but we were agnostic with respect to scaling.
- *Boundary High* (BH): the highest non-spurious value at the end of the utterance in high-ending contours.
- *Boundary Low* (BL): the lowest non-spurious F0 value at the end of the utterance in flat-ending contours. We expected BL to be lower than BH.

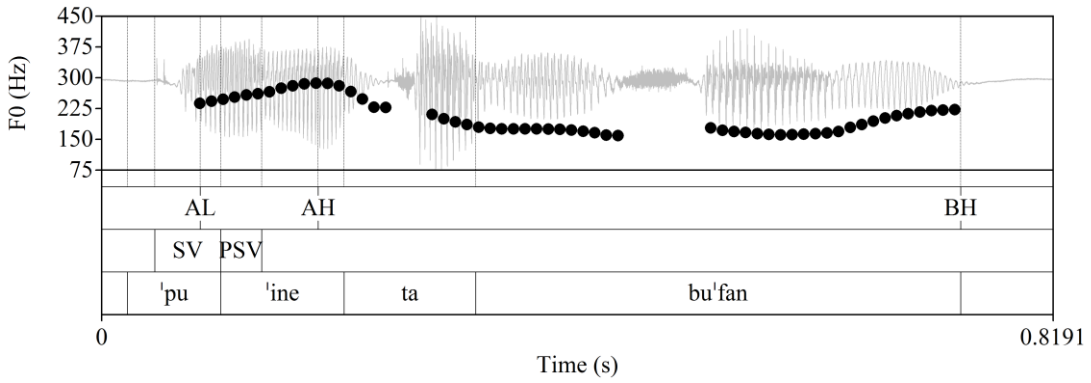


Figure 3: Waveform and F0 contour of the questions, [ˈpu ˈine ta buˈfan] "where are the coats?" together with a textgrid illustrating the annotations.

2.1.4 Statistical analysis

We ran a series of linear mixed-effects models using the *lmer* function of the *lme4* package (Douglas, Maechler, Bolker, & Walker, 2015) in R (R Core Team, 2017).

Specifically, all reported F0 results are based on models with the relevant F0 measurement as the dependent variable, with context, speaker and their interaction as fixed factors, and with item as a random factor. These models were found to perform best for all F0 measurements, based on the likelihood ratio test for accentual high [AH], $\chi^2 = 23.915$, df = 6, $p < 0.001$; for boundary High [BH] and boundary Low [BL], $\chi^2 = 222{,}462$, df = 6, $p < 0.001$.; Pinheiro & Bates, 2000; Bolker et al., 2009). [For AH alignment, on the other hand, the best fit model included only context as a fixed factor and item as a random factor [$\chi^2 = 42.109$, df = 6, $p < 0.001$]. The interested reader is referred to Appendices 2-4, which present the outputs of the best-fit models.

[1] In addition to the tonal targets reported in the text, we annotated the beginning of the low plateau and examined its scaling and alignment relative to the first stressed syllable following the wh-word. In the interest of space, these results are not presented as no differences were found between the two tunes.

2.2 Results

Context A triggered high-ending tunes and context B flat-ending tunes [η = 0.708]. All four speakers consistently produced distinct tunes in response to the two types of context. Since production across speakers was largely consistent, below we report only on statistically significant differences that pertain to context, unless additional detail is needed.

First, our results revealed an asymmetry in the behaviour of the Accentual Low (AL) as a function of context, with AL missing much more frequently after context B than context A: 67% of tokens with the flat-ending tune lacked a discernible AL compared to 11% for the high-ending tune. As this difference did not leave a sufficient number of tokens for comparison, we did not analyse AL any further (but see 2.3 for a discussion).

BH was scaled significantly higher than BL [est. = −88.804, S.E. = 6.645, $t(126)$ = −13.365]; see Appendix 2 for details. On average, BH was 212 Hz [SD = 53, N = 64], while BL was 122 Hz [SD = 35, N = 64]. Paired *t*-tests showed that this difference holds for all four speakers [F1, $t(15) = 13.624$, $p < 0.001$; F2, $t(15) = 16.502$, $p < 0.001$; M1, $t(15) = 10.273$, $p < 0.001$), M2, $t(15) = 13.504$, $p < 0.001$].

Overall, the scaling of AH was not affected by context [est. = -27.445, S.E. = 13.817, $t(126)$ =-1.986; for context A, $\bar{x}$ = 284 Hz, SD = 41, N = 64; for context B, $\bar{x}$ = 289 Hz, SD = 46, N = 64]. See Appendix 3 for details. Speakers varied with respect to the scaling of AH, however. In particular, paired *t*-tests showed that for F1, AH was significantly higher in high ending tunes ($\bar{x}$ = 310.8 Hz, SD = 15.6) than in flat-ending tunes ($\bar{x}$ = 283.3 Hz, SD = 29.3), [$t(15) = 4.999$, $p < 0.001$]. F2 showed a similar pattern (high-ending tunes: $\bar{x}$ = 302.8 Hz, SD = 31.2, flat-ending tunes: $\bar{x}$ = 292.5 Hz, SD = 8.8), but the difference did not reach statistical significance [$t(15) = 1.624$, $p > 0.05$]. M2 had the opposite pattern, with AH in flat-ending tunes ($\bar{x}$ = 292.0 Hz, SD = 8.7) being significantly higher than in high-ending tunes ($\bar{x}$ = 247.4, SD = 23.7), [$t(15) = -4.701$, $p < 0.001$]. M1 showed a similar pattern to M2 (flat-ending tunes: $\bar{x}$ = 287.1 Hz, SD = 75.1, high-ending tunes: $\bar{x}$ = 275.2, SD = 52.2) which did not reach significance [$t(15) = -0.502$, $p > 0.05$].

Context did systematically affect AH alignment [est. = −148.19, S.E. = 14.56, $t(126)$ = −10.178]; see Appendix 4 for details. In the pooled data, in context B, the AH peak appeared on average 18 ms after the beginning of the accentual vowel; in context A, on the other hand, AH appeared significantly later, on average 114 ms after the same segmental landmark. Taking into account the duration of the stressed vowel, which was on average 52 ms in context A and 57 ms in context B, these results indicate that AH appeared around 30% into the stressed vowel in context B, but well after the stressed vowel in context A. Indeed, given that the average duration of the post-accentual consonant was 69 ms, the results indicate that in context A, AH appeared near the beginning of the post-accentual vowel. In short, in string identical questions, there was an early accentual peak co-occurring with the wh-word's stressed vowel, in response to context B, but a late peak, co-occurring with the vowel of the following syllable, in response to context A; this difference is illustrated in Figure 4.

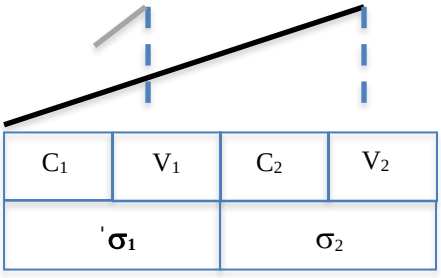


Figure 4: Stylized F0 curve illustrating the alignment of the accentual peak in response to context A (black line) and context B (grey line); σ stands for syllable; ˈσ1 is the accented syllable of the wh-word.

2.3 Interim discussion

The acoustic analysis of the production data indicates that the four speakers produced two distinct tunes, each one used in response to a different context: a high-ending tune in response to context A in which a wh-question is interpreted as information-seeking, and a flat-ending tune in response to context B in which a wh-question can be interpreted either as information- or as non-information-seeking. These differences were consistent across speakers.

As noted, the two tunes showed a systematic difference in the final boundary scaling, in that F0 was consistently lower in flat-ending than high-ending contours, as we had anticipated. In addition to differences regarding the boundary tone, we found differences regarding the pitch accent associated with the wh-word. In our data, the high-ending tune had an accent that started relatively low and rose to a late peak occurring after the end of the stressed vowel of the wh-word (on late peak alignment in Greek, see Arvaniti, Ladd, & Mennen, 1998). The accent in the flat-ending tune, on the other hand, had only an optional rise, while the accentual peak occurred early, well within the stressed vowel of the wh-word.

Both the alignment of the peak and the different behaviour of the preceding rise are of interest in interpreting the results. As mentioned in section 2.1.2, the length of the wh-word varied from monosyllabic, for example, [ˈpos] “how”, to longer constructions such as [apoˈpu] “from where”. It is known that lack of sufficient segmental material can lead to tonal crowding which can in turn affect the realization of tones (cf. Arvaniti & Ladd, 2009, on Greek wh-questions; Arvaniti, Żygis, & Jaskula, 2016, on Polish calling contours). Despite tonal crowding, however, AL – which reflects the rise to the accentual peak – was typically present in high-ending contours. In contrast, AL was not realized in flat-ending contours unless there was enough segmental material (e.g., with wh-expressions like [apoˈpu] “from where” but not with monosyllabic wh-words like [ˈpos] “how”). Figure 5 illustrates this difference using a question with a short wh-word, [ˈti] “what”. These differences in realization between the accents in the two tunes are consistent with the phonetics of the accents represented as L*+H and L+H* respectively in earlier work on Greek intonation (see Arvaniti et al., 1998; Arvaniti & Ladd, 2009, on L*+H; Arvaniti & Baltazani, 2005; Arvaniti et al., 2006, on L+H*). Thus our results indicate that our speakers produced

two distinct tunes: L*+H L-!H% and L+H* L-L%.[2]

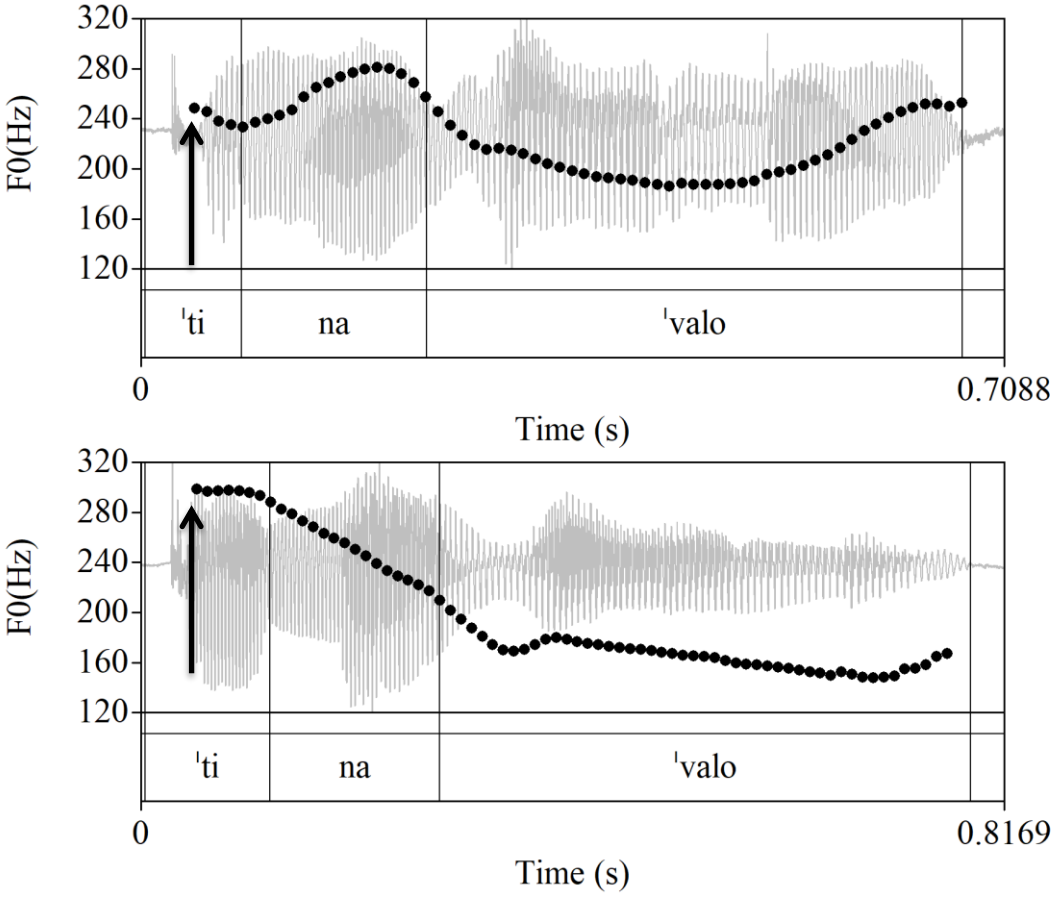


Figure 5: Waveform and two renditions of the same question, [ˈti na ˈvalo] “What should I wear?”, with the high-ending contour (top), and the flat-ending contour (bottom), illustrating the different onset of the two contours; arrows show the relevant points.

3. Experiment 1: Pragmatic interpretation

In Experiment 1 participants heard questions with the two tunes and were asked to choose appropriate responses to them (for details see 3.1.2 below). We anticipated that when participants listened to the questions out of context they would tap into prototypical out-of-the-blue interpretations of the two tunes. Specifically, we hypothesized that participants would be more likely to interpret high-ending questions as information-seeking and thus choose an information-providing response. On the other hand, flat-ending questions, *though they could also be interpreted as*

[2] This should not be taken to imply that no other tunes are used with wh-questions in Greek. Different tunes are possible when wh-questions are used rhetorically, as exclamations, or to indicate surprise; a discussion of these tunes is beyond the scope of the present paper. Further, our brief description of the phonetic differences between the two tunes does not preclude the presence of additional cues, such as durational or amplitude cues; for a discussion of such cues using a large corpus of questions, see Gryllia et al. (2018).

information-seeking, would be more likely than high-ending questions to be interpreted as non-information-seeking.

3.1 Method

3.1.1 Participants

The results reported here are based on 50 participants (38 female and 12 male). They were all monolingual native speakers of standard Greek studying at the University of Ioannina, and ranged in age from 18 to 32 years. The data from an additional fourteen participants were discarded after their questionnaires showed that they were bilingual (N = 4), or speaking a dialect other than the standard (N = 8), or had a history of speech or hearing disorders (N = 2). We also discarded the data from an additional seven participants as their questionnaires showed that they had given the same response in more than 85% of the stimuli; we considered their responses to be due to *carelessness* or *insufficient effort* (Huang et al., 2012; Meade & Craig, 2012).

3.1.2 Stimuli and procedure

The stimuli consisted of six sets, each containing eight wh-questions; each set included four high-ending and four flat-ending versions of the same question as produced by each of the four speakers in the production study (6 wh-questions × 2 tunes × 4 speakers). These 6 questions (48 stimuli) were selected from the set of 16 questions (which yielded a corpus of 128 questions) discussed in section 2. The six questions were chosen based on the following criteria: (a) The realization of the tune: we chose realizations prototypical for each of the two tunes to ensure they were maximally distinct, as we were interested in exploring the different pragmatic interpretation of the two tunes; (b) Pragmatic plausibility of the responses each question could elicit. The questions had to be equally plausible as information- and non-information seeking questions to avoid biasing the participant responses towards a particular interpretation; further, we wanted both replies to be equally plausible and easy to understand (i.e., not convoluted or far-fetched).

The experiment included 48 trials. Each question was followed by two responses, as shown in (5) below: Response A provides information about the questioned constituent; choosing it implied that the participant took the question "Who is Manolopoulos?" as information-seeking. Response B does not answer the question; choosing this response implied that the participant assumed the question was non-information-seeking and meant instead as a comment about the worth of Manolopoulos, an evaluation with which the addressee agreed. We hypothesized that high-ending questions would lead to participants' choosing more A-type responses and that flat-ending questions would lead them instead to choosing more B-type responses (see Appendix 5 for all questions and responses used in the experiment).

(5) Stimulus: [ˈpços ˈine o manoˈlopulos]
"Who is Manolopoulos?"

Response A: *He's a friend of Danae's*
Response B: *Indeed! He is utterly useless!*

Each trial started with a warning tone (441 Hz, 200 ms), followed by 250 ms of silence, and then the question. Participants had 3.5 s to respond before the next warning tone started. Their task was to choose one of the two possible responses to each question, presented to them in hard copy response sheets in counterbalanced order. They were told that each time they would hear a question that was part of a dialogue, and their task was to decide how the dialogue continued. The experiment ran on Praat and took approximately 20 minutes. It was presented aurally via loudspeakers (free field) in a classroom at the University of Ioannina. The free field method introduces some "noise" into the data, in that participants seated in different locations do not hear the stimuli in exactly the same way. It was chosen because we estimated that these differences were unlikely to have a material effect on the responses, since these dealt with pragmatic interpretation, which often takes place in noisy conditions, while free field allowed us to collect data from a large sample of participants (we note that free field has been successfully employed even for psychoacoustic experiments where listening conditions are critical, e.g. Iversen, Patel, & Ohgushi, 2008).

3.1.3 Statistical analysis

We ran a number logistic regression models using the *glm* function in *lme4* package (Douglas et al., 2015) in R (R Core Team, 2017). The results reported here are based on the model that included participants' responses as the dependent variable, and tune, speaker and their interaction as fixed factors. The likelihood ratio test showed that this model performed best [$\chi^2 = 43.111$, df = 6, $p < 0.001$]. For details, the interested reader is referred to Appendix 6.

3.2 Results

The results showed an effect of tune: participants chose information-providing responses significantly more frequently after high-ending than flat-ending questions [est. = 0.9872, S.E. = 0.1701, $p < 0.001$, ref. F1]6. As can be seen in Figure 6, the effect was not identical for the four speakers, however. To address this issue, we ran additional models with speakers F1, F2, and M1 as reference categories. The results showed that despite the differences in effect magnitude across speakers, for all of them participants chose information-providing responses significantly more frequently after high-ending than flat-ending questions. For details, see Appendix 6.

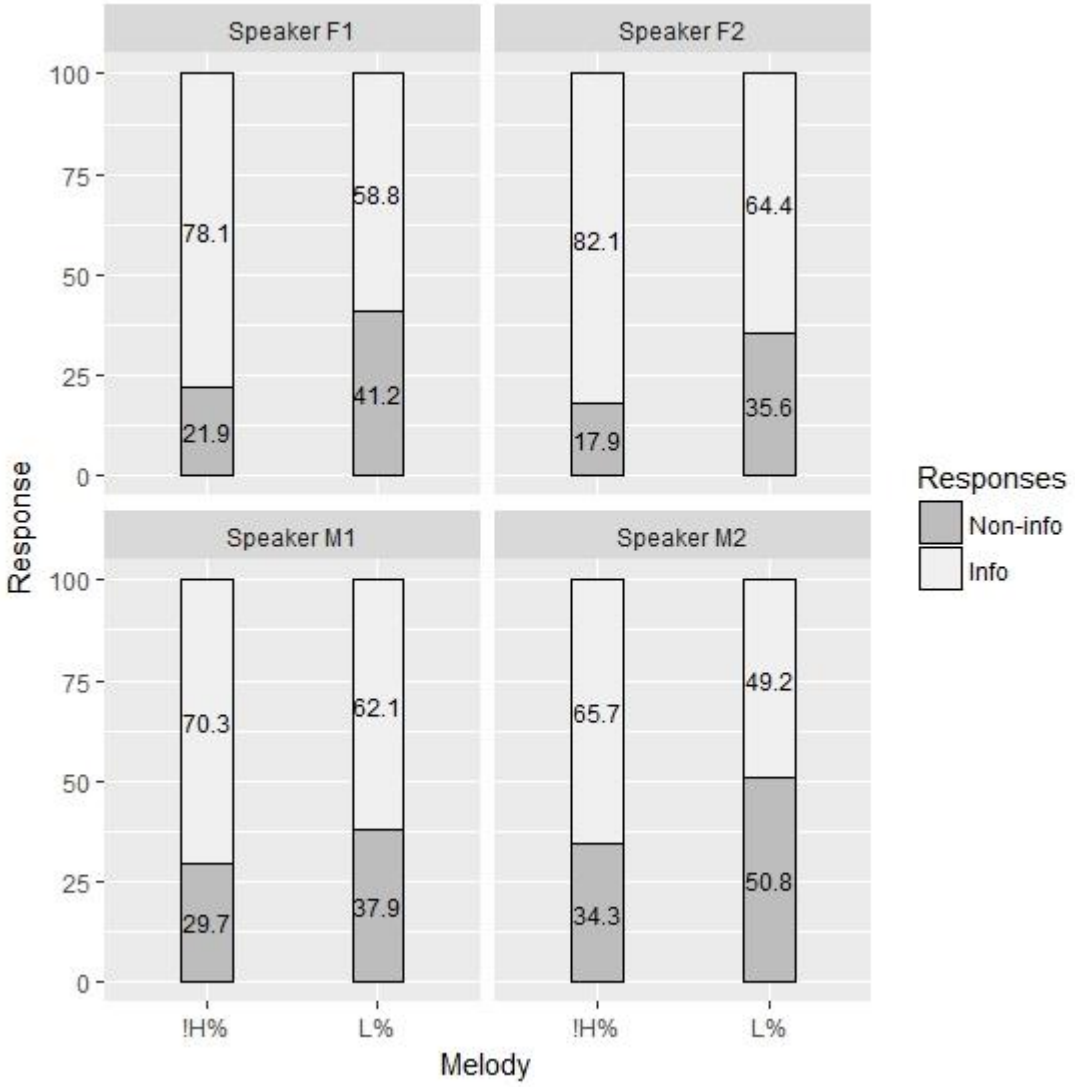


Figure 6: Percentages of participant responses as a function of tune and speaker.

3.3 Interim discussion

The aim of Experiment 1 was to test if the two tunes, L*+H L-!H% and L+H* L-L%, would lead to different pragmatic interpretations of string identical wh-questions. Our results showed that they did and that the differences were consistent with our overall interpretation of the pragmatics of the two tunes. As anticipated, the differences were not categorical, in that participants did not exclusively choose one type of response per tune. Rather, they showed a *preference* for one or the other interpretation, showing a statistically stronger preference for information-seeking responses after L*+H L-!H% tunes relative to questions with L+H* L-L%. This difference indicates that Greek speakers preferentially associate each tune with a pragmatic interpretation, but do not do so in an absolute fashion. Further, as the results in Figure 6 show, both information-seeking and non-information-seeking responses were selected by participants in response to both tunes; this further suggests that wh-questions in Greek can generally be information-seeking or non-information-seeking, and that ultimately a question's pragmatic interpretation rests with the listener. We return to this point in section 6.1.

4. Experiment 2: Appropriateness and politeness

In Experiment 2, we tested how the two tunes are evaluated by Greek listeners, since, as noted in the introduction, there is previous work indicating that the high-ending tune is more involved (Arvaniti & Baltazani, 2005), a description indicating that it is considered more polite. Further, the results of Experiment 1 showed a general preference for the high-ending tune, so a question we wished to address was whether this tune was seen as generally more appropriate for questions. In order to address these research questions, we provided participants with contexts similar to those used for the elicitation of the stimuli and followed by questions; they were asked to rate each question for its appropriateness and politeness in the context in which it was heard. We expected that the high-ending tune would be considered more appropriate in contexts that led to an information-seeking question. In contrast, we expected that the flat-ending tune would be considered more appropriate when the context indicated that the question was non-information-seeking. Thus, we anticipated an interaction between tune and context with respect to the appropriateness rating. Finally, we hypothesized that the flat-ending tune would be considered less polite than the high-ending tune overall (see also section 1.1.)

4.1 Method

4.1.1 Participants

The reported results are based on 70 participants (54 female and 16 male). They were all monolingual native speakers of standard Greek studying at the University of Ioannina, and ranged in age from 18 to 23 years. None had taken part in Experiment 1. The data of an additional 18 participants were discarded because of one of the following reasons: they were bilingual (N = 7), spoke a dialect other than the standard (N = 5), had a history of speech or hearing disorders (N = 4), provided no response (N = 1) or the same response (N = 1) to most trials.

4.1.2 Stimuli and procedure

The stimuli were the same 16 questions as in Experiment 1. These were deemed the most appropriate for the same reasons as before: (a) the realization of each tune was prototypical; (b) both versions of each question were equally plausible in information- and non-information seeking contexts, while the contexts that could be constructed for these questions were relatively brief and easy to follow (see also section 3.1.2).

For each question we constructed two contexts, such that one context would likely lead to an information-seeking question as in (6), or to a non-information-seeking one as in (7) (cf. Portes et al., 2014, for use of a similar paradigm). Specifically, the context in (6) describes a situation in which a speaker may plausibly use a question to request information. In contrast, a question following a context such as (7) could be interpreted either as information-seeking or as an indirect way of telling the addressee that he has no good reason to complain about his tax situation.

(6) Context A: *Alkis and Haris are discussing their finances. Haris says:*
Question: [ˈposa ˈpliroses stin efoˈria]
“How much have you paid in taxes?”

(7) Context B: *Haris complains that he pays high taxes. Alkistis, who knows Haris is on the dole, says to him:*
Question: [ˈposa ˈpliroses stin efoˈria]
"How much have you paid in taxes?"

The contexts were read by a native speaker of Greek who was not among the speakers who recorded the stimuli (for the contexts see Appendix 7). [3] Contexts and questions were crossed for a total of 96 trials (48 melodies × 2 contexts) so that each question was heard after a context that made asking for information the most plausible choice (context A), and after a context that make such a request less plausible (context B). Trial order was pseudo-randomized so that participants would not hear a high-ending (!H%) and a flat-ending (L%) version of the same question in consecutive order.

Each trial started with a warning tone (441 Hz, 200 ms), followed by 250 ms of silence, the context and, after 200 ms of silence, the question; participants then had 3.5 s to respond before the next warning tone was heard. The 96 experimental trials were preceded by 4 training trials. No fillers were used, as the experiment was long and complex, and we did not wish to complicate it further.

The experiment ran on a PC using Praat and lasted approximately 35 minutes. It was presented aurally via loudspeakers in a classroom at the University of Ioannina. Participants filled response sheets in hard copy. They were told they would hear a "short story" followed by another utterance and would have to answer questions about the latter. They were tasked with rating how appropriate and polite each question was *in the context that preceded it*. The two rating questions appeared one after the other on the response sheet in the same order, with appropriateness first; no counterbalancing was attempted as the experiment was quite lengthy and complicated already. Both questions were answered using a 1-7 Likert scale.

4.1.3 Statistical analysis

The responses to the two rating variables were z-transformed (for appropriateness, N = 6702; for politeness, N = 7252). We ran a number of linear mixed-effect models using the *lmer* function of the *lme4* package (Douglas et al., 2015) in R (R Core Team, 2017).
The reported results are based on the best fit models, according to the likelihood ratio test [for appropriateness, $\chi^2 = 29.0923$, df = 6, $p < 0.001$; for politeness, $\chi^2 = 232.5602$, df = 6, $p < 0.001$; Pinheiro & Bates, 2000; Bolker et al., 2009]. These models included tune, context, and speaker as fixed factors, and the tune × context, and tune × speaker interactions; participants and items were included as random factors. The interested reader is referred to Appendix 8 for the full outcomes of the appropriateness model, and to Appendix 9 for the politeness model.

[3] These contexts were almost identical to the contexts used in the production study. To illustrate, in the production study, the contexts used to elicit the question "How much have you paid in taxes" only differed from the perception contexts shown in (6) and (7) in terms of the final verb, which in the production was *asks* instead of *say*s. Further, in order to avoid repetition, in the perception experiment we varied the contexts somewhat, such that no two instances of the same question were preceded by identical contexts. The types of differences can be seen if one compares the contexts in (6) and (7) to those for stimulus 4 in Appendix 7.

4.2 Results

With respect to appropriateness, we found that participants judged questions more appropriate when they were preceded by a context that made information-seeking plausible, that is, context A [est. = 0.342, S.E. = 0.028, *t*(6702) = 12.053]. There was no effect of tune [for tune: est. = 0.005, S.E. = 0.045, *t*(6702) = 0.104], however, the interaction between context and tune was significant [est. = 0.132, S.E. = 0.040, *t*(6702) = 3.294]: participants rated high-ending questions as more appropriate than flat-ending questions after information-seeking contexts, while there was no difference for non-information-seeking contexts after which both tunes had lower ratings (see Figure 7, and Table 1). For more details see Appendix 8.

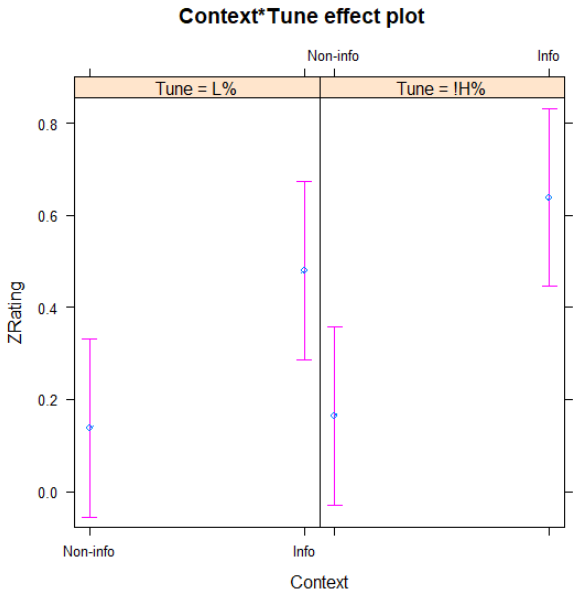


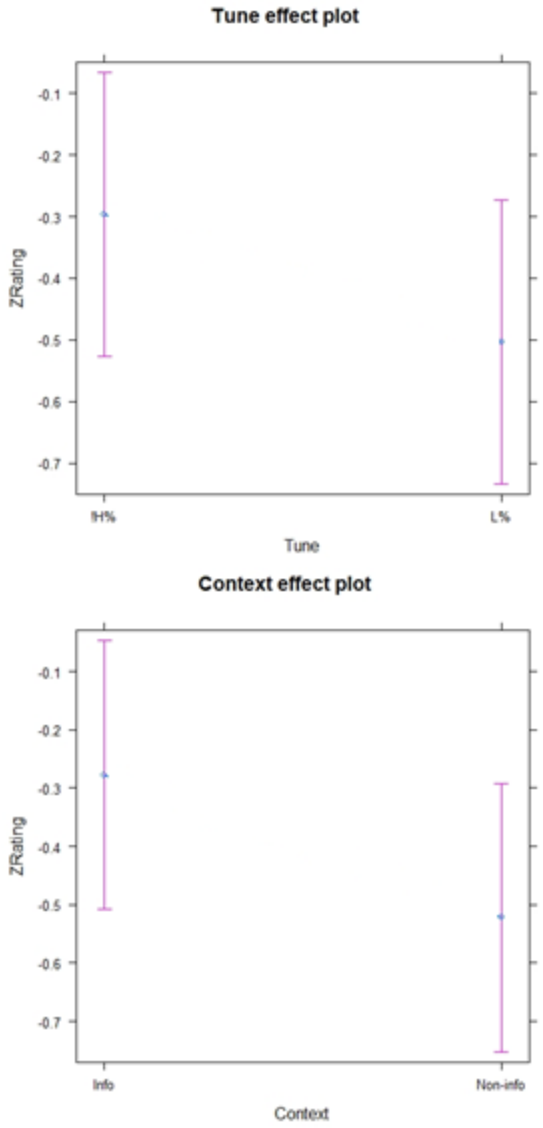
Figure 7: Appropriateness z-ratings as a function of context and tune interaction.

Table 1: Raw ratings on the Likert scale and z-ratings of appropriateness and politeness; standard deviations are given in parentheses.

		Appropriateness		Politeness	
Tune	Context	Rating	Z-Rating	Rating	Z-Rating
L%	Non-info	4.8 (1.9)	0.1 (1.0)	3.5 (1.7)	-0.6 (0.9)
L%	Info	5.5 (1.6)	0.5 (0.9)	3.8 (1.7)	-0.4 (0.9)
!H%	Non-info	4.9 (1.8)	0.2 (1.0)	3.8 (1.8)	-0.4 (0.9)
!H%	Info	5.8 (1.4)	0.6 (0.8)	4.3 (1.7)	-0.2 (0.9)

With respect to politeness, we found that high-ending questions were judged more polite than flat-ending questions [est. = 0.415, S.E. = 0.041, *t*(6679) = 10.151]; see Figure 8(a) (and Table 1). Context had an effect on politeness ratings, such that questions heard after type A contexts (information-seeking) were judged more polite than questions heard after type B contexts [est. = 0.211, S.E. = 0.025, *t*(6679) =

8.178]; see Figure 8(b); for details see Appendix 9. The interaction between tune and context was not significant [est. = 0.064, S.E. = 0.036, *t*(6679) = 1.775]. The results also showed significant interactions between tune and speaker; see Figure 9. For speakers F1, F2 and M2 flat-ending questions were judged less polite than high-ending questions. For speaker M1, there was no effect of tune on the rating of politeness; flat-ending and high-ending questions were judged similarly. We return to this point in section 5.



(a) (b)

Figure 8: Politeness ratings as a function of tune (panel a) and context (panel b).

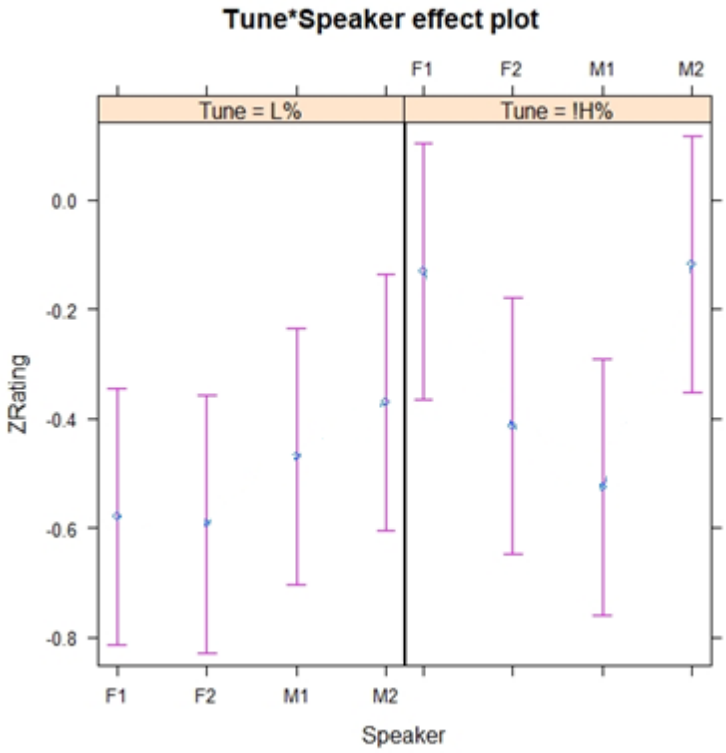


Figure 9: Politeness z-ratings as a function of tune and speaker interaction.

4.3 Interim Discussion

The results confirmed our prediction that high-ending questions are considered more polite than their flat-ending counterparts, in both information-seeking and non-information-seeking contexts.

The results regarding appropriateness support those of Experiment 1, by showing that the high-ending tune is considered significantly more appropriate than the flat-ending tune after contexts meant to trigger information-seeking questions. In addition, however, we found that in non-information-seeking contexts, both tunes received lower ratings (see Figure 7). Based on our analysis and the results of Experiment 1 this was not anticipated: we had expected that flat-ending melodies would be rated as more appropriate in non-information-seeking contexts, while high-ending melodies would seem inappropriate in these contexts and thus rated lower; indeed the use of a high-ending tune in some non-information-seeking contexts could be interpreted as sarcastic or mocking by feigning ignorance. The fact that our prediction was not borne out, however, is in line with the preference for questions to be treated as information-seeking, evinced in Experiment 1. This indicates that computing the additional inferences required for the non-information-seeking meaning of questions is more complicated and thus avoided by the participants, at least in an experimental setting. In turn this means that it is impossible to hinder participants from interpreting any context in a way that allows them to treat the following question as information-seeking (cf. Bateson, 1972; Goodwin & Duranti, 1992). This seems to be the case. Consider, for instance, a trial from the experiment in which the context in (8) is followed by the question in (9) produced with the high-ending tune; it is possible for a

listener to consider that Alexia's aim is not to express her displeasure at the ordeal awaiting her the following day but to genuinely seek the addressee's assistance in finding alternative means of transport (while also expressing her annoyance at the inconvenience that awaits her).

(8) *All public transport will be on strike tomorrow. Alexia, who needs to deliver an order in Syntagma Square, says to her husband:*
(9) [ˈpos θa ˈpao sto ˈsidaɣma]
"How will I get to Syntagma?"

In addition to the general conclusions drawn from this experiment our results revealed some speaker-specific differences, particularly in how polite the questions were rated, in that the high-ending questions of F1, F2, and M2 were considered more polite than their flat-ending counterparts, but the same did not apply to M1. Since these different results were coupled with some speaker-specific (albeit not statistically significant) differences in Experiment 1 (see section 3.2), we wished to further investigate the effect that our speakers may have had on the results. This was of particular importance since, as noted in section 2.1.1, F1 and F2 had phonetic training, and a M1 retained traces of a Northern Greek accent.

5. Experiments 3 and 4: The role of the speaker

We conducted two follow-up experiments using the same stimuli as before elicited from six speakers, the four original and two additional ones, a female speaker without phonetic training, and a male speaker who was phonetically trained; both were native speakers of Standard Greek in their late thirties to early forties. The set-up of the two experiments was the same (see 5.1 below) except that in the first experiment, the participants were asked to comment on the questions, and in the second on the speakers' voices. We refer to these as the *naturalness* and *voice* experiment respectively. The aim of the former was to test whether the questions of the trained speakers were rated differently from those of the untrained ones (since they had patterned similarly in Experiment 1); the aim of the latter was to test whether specific voices elicited different reactions from participants (since we found some speaker-specific differences in politeness ratings in Experiment 2).

5.1 Methods

5.1.1 Participants

For the naturalness experiment, we collected data from 24 monolingual native speakers of Standard Greek (19 female, 5 male, age range: 22-36 years old). For the voice experiment, we collected data from 25 monolingual native speakers of Standard Greek (10 female, 15 male, age range: 22-46 years old). No participants in either experiment reported any speech or hearing disorders. We used social media to advertise the experiments, which were conducted online.

5.1.2. Stimuli and procedures

For both experiments the stimuli consisted of 72 questions, 36 wh-questions uttered with the high-ending tune and 36 questions uttered with the flat-ending tune. As mentioned, 48 of these were the same questions used as stimuli in Experiments 1 and 2. The other 24 stimuli were the same questions elicited from two additional speakers, a phonetically trained male (M3), and a female speaker without phonetic training (F3). Their stimuli were elicited and recorded in a similar manner to that of the original speakers.

For both the naturalness and voice experiment, these 72 stimuli were organized into six blocks each containing 12 questions, evenly divided by tune. Trial order was randomized within each block. There were no fillers or training session. The experiments ended with a series of questions regarding the linguistic history of the participants and the method they used to run the experiment (specifically, whether they used headphones or their computer's loudspeakers; the experiments ran only on computers, not on mobile devices).

The experiments were prepared and run online using *psytoolkit* (psytoolkit.org) set to experiment mode (Stoet, 2010, 2017). Each trial started with a warning tone (440 Hz, 600 ms), followed by 500 ms of silence. Then participants heard a question, at the offset of which they were asked one of two questions, depending on the experiment. In the naturalness experiment, they were told they would hear a series of questions and had to answer how they found each question. Participants were given three choices: (i) fine, (ii) so and so, and (iii) somewhat odd. In the voice experiment, participants were told that we solicited input on the voices they would hear, specifically, that we wanted to know how much they would like to hear these voices on the radio. They were again given three choices, (i) pleasant, (ii) so and so and (iii) not pleasant. Each experiment lasted approximately 10 minutes.

5.1.3. Statistical analysis

Since there were only three ratings, we ran multinomial logistic regressions using the *multinom* function of the *nnet* package (Venables & Ripley, 2002) in R (R Core Team, 2017). Rating was included in the models as the dependent variable, while tune, speaker, and the interaction tune × speaker were included as fixed factors.

5.2 Results

The naturalness experiment yielded a total of 1728 responses. There was an effect of tune [est. = 0.273, SE = 0.133, $p < 0.05$], but no interaction between tune and speaker: flat-ending tunes were more likely to be rated as *somewhat odd* than *fine* relatively to the high-ending tunes. Finally, there were some differences among speakers: the stimuli of M1 were more likely to be rated as *somewhat odd* rather than *fine* compared to those of F1, F2, and F3 [for F1: est. = -0.830, SE = 0.357, $p < 0.05$; for F2: est. = -1.664, SE = 0.428, $p < 0.01$; for F3: est. = -0.920, SE = 0.342, $p < 0.01$]; the stimuli of M2 were more likely to be rated as *somewhat odd* rather than *fine* compared to those of F1, F2, F3, and M3 [for F1: est. = -1.096, SE = 0.344, $p < 0.01$; for F2: est. = -1.930, SE = 0.417, $p < 0.01$; for F3: est. = -1.188, SE = 0.329, $p < 0.01$; for M3: est. = -0.731, SE = 0.304, $p < 0.01$]; see Figure 10 and Appendix 10.

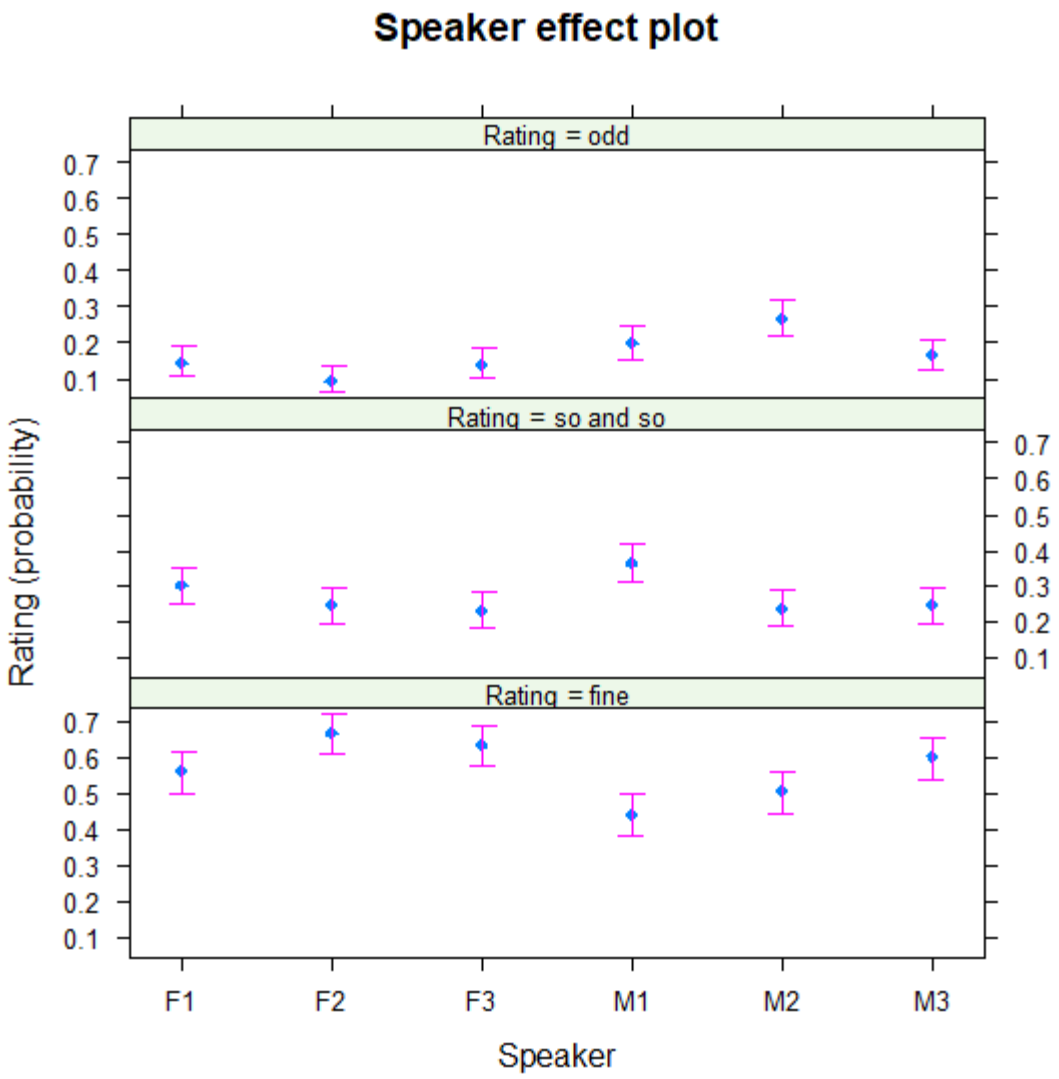


Figure 10: Naturalness ratings as a function of speaker.

The voice experiment yielded a total of 1800 responses. There was no effect of tune on voice rating [est. = -0.405, SE = 0.303, $p > 0.05$]. The comparison across speakers showed that in high-ending questions and with the voice of F1 as reference, F2, F3 and M2 were judged more pleasant, M1 less pleasant, and M3 equally pleasant [for F2, est. = - 1.792, SE = 0.359, $p < 0.01$; for F3, est. = - 1.423, SE = 0.341, $p < 0.01$; for M2, est. = - 0.657, SE = 0.296, $p < 0.01$; for M1, est. = 2.069, SE = 0.379, $p < 0.01$; for M3, est. = -0.272, SE = 0.307, $p > 0.05$]; see Figure 11. In flat-ending questions and with the voice of F1 as reference, F2 was judged more pleasant, M1 less pleasant, and F3, M2 and M3 equally pleasant to F1 [for F2, est. = -1.872, SE = 0.431, $p < 0.01$; for M1, est. = 2.671, SE = 0.418, $p < 0.01$; for F3, est. = -0.194, SE = 0.326, $p > 0.05$; for M2, est. = -0.407, SE = 0.315, $p > 0.05$; for M3, est. = 0.273, SE = 0.323, $p > 0.05$]. For details see Appendix 11.

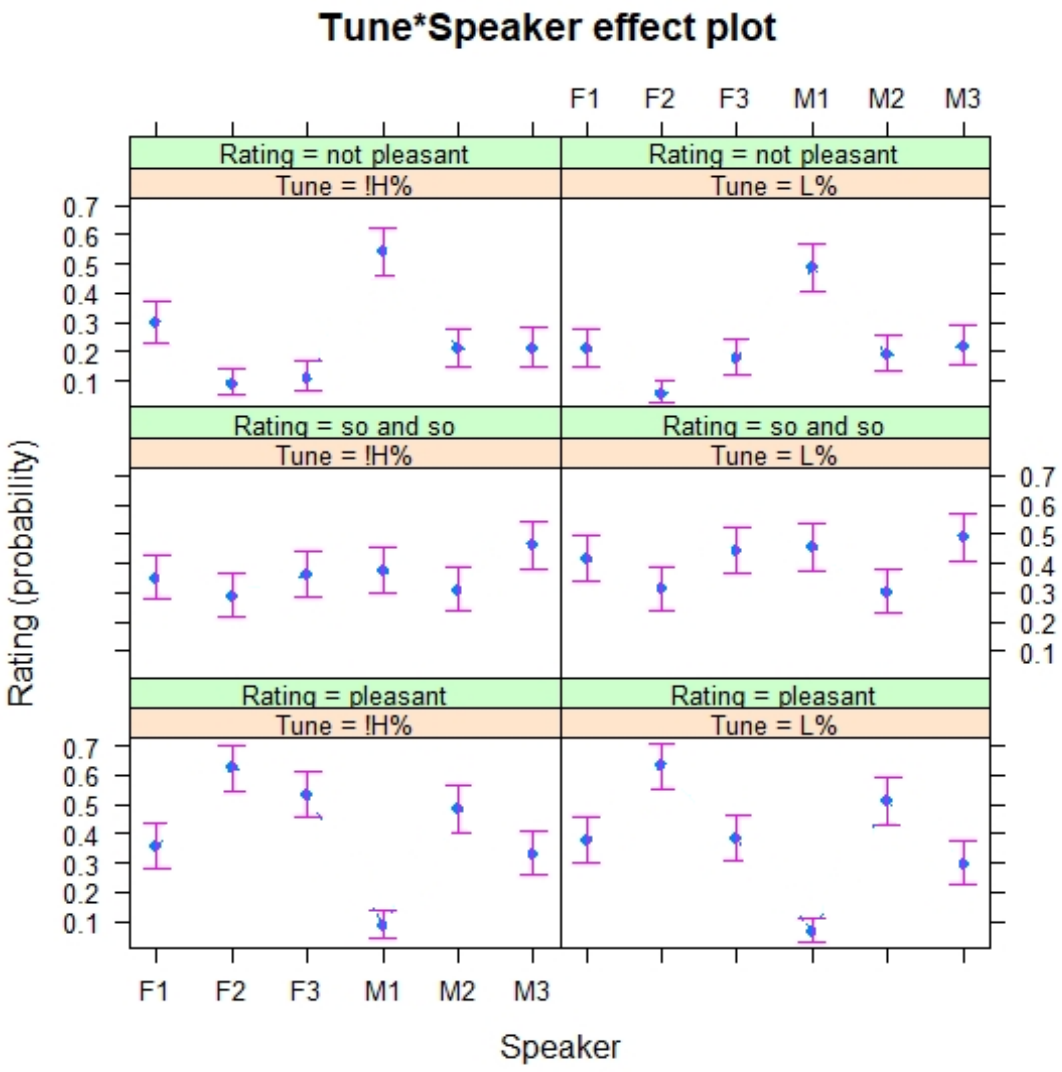


Figure 11: Voice ratings as a function of tune and speaker.

5.3 Interim discussion

The aim of the naturalness and voice experiments was to explore possible effects that the voices of our speakers, the training of F1 and F2, and the accent traces in the speech of M1 could have had on the pragmatic interpretation and evaluation of the two tunes under investigation. These follow-up experiments showed that if such effects were present, they were minimal; rather, the results supported our conclusions from the original experiments. Differences regarding how the voices and questions were rated were present but did not follow a discernible pattern. First, the questions and voices of F1 and F2 were not rated differently from those of the other speakers, independently of training (see Figure 10). The overall higher rating of high-ending questions supports the results on appropriateness (Experiment 2), indicating that the high-ending tune is preferred by Greek speakers. This is not surprising, given that the interpretation of this tune is straightforward with questions, as explained in more detail in section 6.1. Further, no effect in terms of the questions was observed with respect to speaker M1, suggesting that his tunes were deemed comparable to those of the other speakers; this result corroborates the acoustic analysis of the stimuli which also showed no distinction between this speaker and the rest. On the other hand, the

voice of M1 was rated less favourably than other voices (see Figure 11). This could be due to voice quality or to his accent, but it is not possible to disentangle the two effects. This unfavourable rating may be the reason why there was no differentiation between his high-ending and flat-ending questions in terms of politeness. Overall, these follow-up experiments support our main results.

6. Discussion

The acoustic analysis of the stimuli confirmed that our speakers produced questions with two distinct tunes depending on the context in which the questions were elicited. The production data also showed that the tunes differed systematically not only in their boundary tones, as reported in previous research, but also in the pitch accent associated with the wh-word. Within the intonational system of Greek, these two tunes are best represented as L*+H L-!H% and L+H* L-L%.

The two tunes also led to differences in the interpretation of string-identical questions, and were evaluated differently with respect to how appropriate they were and how polite they sounded. Experiment 1, in particular, showed that these tune differences influenced pragmatic interpretation: participants overwhelmingly interpreted high-ending tunes as information-seeking and chose answers that provided the requested information; for flat-ending tunes participants chose significantly more frequently responses consistent with an interpretation of the questions as non-information-seeking. The fact that the switch was not complete and that participants still preferred information-providing answers, while they also chose non-information providing responses to high-ending questions, indicates that there is not a one-to-one relationship between meaning and tune; rather, the tune can preferentially lead to some interpretation over another.

6.1 Pragmatic interpretation

A question that arises is how these different interpretations emerge and how they can be connected to the composition of the two tunes. We address this below by providing a compositional analysis of the two tunes, starting with the interpretation of the high-ending tune, L*+H L-!H%.

One property of information-seeking wh-questions in general and in Greek in particular is that they "…show non-exhaustive quantification, in the sense that they do not require a presupposed set of alternative values to be assigned to the wh-variable" (Roussou, Vlachos & Papazachariou, 2013: 484). This property accords well with the melodic makeup of the Greek information-seeking, high-ending questions: their wh-word carries a L*+H pitch accent, which is typically used in Greek either in the prenuclear domain or as the nucleus in calls and continuation rises (Arvaniti & Ladd, 1995; Arvaniti, Ladd, & Mennen, 1998, 2000; Arvaniti & Baltazani, 2005; Baltazani et al., 2015); all are open-ended contexts in which follow-up is possible. The high boundary tone denotes speaker lack of commitment and acts as an invitation to the addressee to provide the missing information (i.e., select from an open set). For example, when the question [ˈti na su etiˈmaso] 'What [food] should I prepare for you?' is uttered by a parent to a child with the L*+H L-!H% tune, it is interpreted as *I*

don't know what you would like, please provide me with your preference.

This interpretation does not hold for flat-ending questions, that is, when the intonation is L+H* L-L%. First, the L+H* nuclear accent signals narrow or contrastive focus in Greek (Botinis, 1998; Baltazani & Jun, 1999; Georgiafentis & Sfakianaki, 2004; Arvaniti & Baltazani, 2005; Arvaniti et al., 2006; Gryllia, 2009a, b; Georgakopoulos & Skopeteas, 2010; Haidou, 2012; Skopeteas, 2016). Narrow focus gives rise to exhaustive quantification, that is, a closed set of alternatives (Rooth, 1985, 1992). Thus, a substitution of the L*+H with the L+H* pitch accent on the wh-word invites the addressee to choose one among a (small) number of choices in a closed set of (known) alternatives (cf. Dennison & Schafer, 2017; Tomlinson, Gotzner, & Bott, 2017). Second, in addition to the L+H* accent, the flat-ending tune ends with L-L% edge tones. The L+H* L-L% is a tune also used with narrow focus declaratives (e.g. Arvaniti et al., 2006). In declaratives, the L-L% edge tones imply speaker commitment to the truth of the utterance (which is the selection of a specific alternative, the L+H* accented item, out of a closed set). We maintain that the use of typically declarative edge tones with wh-questions also implies speaker commitment in Greek; thus, commitment is not shifted to the addressee, giving an assertive force to the flat-ending questions. By extension, this move adds the implicature that every choice in the closed set of alternatives denoted by L+H* is false; from the point of view of the speaker, this renders any choice made by the addressee an infelicitous answer.

An alternative way to understand the pragmatics of the flat-ending questions is to compute their meaning by backward implication: *because* the speaker utters the question as a statement (i.e., with L-L% edge tones), they imply that there's nothing for the addressee to answer, therefore none of the alternatives in the closed set indicated by the wh-word accented with L+H* is true or appropriate. Note, however, that the addressee can either agree with this discourse move of the speaker and express acquiescence (cf. response B in example (5)), or they can remonstrate and instead add a new element to the common ground. To illustrate, a caregiver uttering the question [ˈti na su etiˈmaso] "What [food] should I prepare for you?" with the L+H* L-L% tune could be conveying something like *We have bread, pastrami, cheese, jam, and peanut butter, but you like none of these, so there is nothing I could fix you*. An inference that the addressee can make here would be that the speaker's utterance amounts to *you're a fussy eater, there's nothing you like*. The addressee can then agree with this inference (*yes, there's nothing I like*), or, treat the question as information-seeking (responding, e.g. with *make me a cheese sandwich*), by either not making or choosing to ignore the negative inference. Alternatively, if the addressee is uncooperative, they can add a new but unavailable item to the set (*I would like a PARMA HAM sandwich!*). If the addressee chooses this response, however, they are likely to use a L+H* accent to indicate contrast and convey to the speaker that they were wrong in implicating that the set of alternatives is null. Note that a L+H* accent would be infelicitous in the response if the question were treated as information-seeking (*make me a CHEESE SANDWICH*), as in this case the addressee is choosing from an open set and thus provides new, not contrastive information.

This possible denotation of wh-questions as non-information-seeking is not unique to Greek and can be added to wh-questions either by using intonation or by other means. Bartels (1997) reports an intonational contrast between English wh-rhetorical

questions, which end in a fall, and wh-ordinary questions which end in a rise; as noted in 1.1. the Greek questions with the flat-ending tune are similar to (though not identical) to rhetorical questions (cf. Sadock, 1974; Han, 2002; Caponigro & Sprouse, 2007). In Russian, on the other hand, the particle *že* is reported to add such an implicature to wh-questions: "Rendering the meaning of utterances containing *že* into English usually involves the use of either some contrastive lexeme, such as *but*, or some prosodic means of indicating contrast. Specifically, the English translation of statements containing *že* can employ the so-called "contradiction contour" (Liberman & Sag, 1974; Pierrehumbert & Hirschberg, 1990)" (McCoy, 2003:325). Especially for the use of *že* in Russian wh-questions, it is observed that "... the question is asked but presupposes that no (reasonable) answer is true, from the speaker's point of view" (McCoy, 2003:326).

It should be clear from the above and from our results that the non-information-seeking interpretation of wh-questions uttered with the flat-ending tune is not always adopted by the addressee. As previous research indicated, Greek speakers produce genuine questions with both the high-ending and the flat-ending tune (Arvaniti & Ladd, 2009; see also Gryllia et al., 2018). Our results confirm this earlier observation and further show that flat-ending questions can be interpreted either as information- or non-information-seeking. This dual interpretation should not be surprising: intonation alone cannot determine the pragmatics of an utterance (Ward & Hirschberg, 1992; Hirschberg & Ward, 1995; Pierrehumbert & Hirschberg, 1990; Gunlogson, 2003; Steedman, 2007, 2014; Armstrong & Prieto, 2015; Brown, Salverda, Gunlogson, & Tanenhaus, 2015). Intonation works in tandem with lexical pragmatics, information structure, the propositional semantics of the utterance, and often specific information about the particular context. Consider for instance a teenager who asks her father for his car keys and hears back [ˈti ta ˈθes] "what do you want them for?" with the flat-ending tune. If our teenager believes her father trusts her, she is likely to interpret the question as a request for information about where she needs to go. On the other hand, if she believes that her father suspects her of wanting the car for some questionable purpose (such as going to a party instead of doing homework), she is more likely to infer that her father is really implying something like *I don't think you have a good reason for borrowing the car*. Note also that this interpretation can arise not from the father's intent, but from the teenager's own knowledge that they want to go out but have not finished their homework. In a different context, a teenager who believes her father does not trust her driving may instead interpret the same question as implying that her father does not want to give her the car keys because he is afraid she will have an accident. Finally, the addressee can entertain both interpretations of the question at once, responding by both providing information and remonstrating against negative implicatures. In short, the interpretation and the inferences the addressee can draw are not determined exclusively by the speaker and their choice of tune, but are partly at the addressee's discretion and can even vary on an individual basis (cf. Degen & Tanenhaus, 2015; Franke & Degen, 2016; Elder, 2017). In the case of wh-questions, the presence of a wh-word is clearly important in determining the pragmatic interpretation of the utterance and its role can outweigh the contribution of the tune; if so, then the straightforward interpretation of the utterance as information-seeking wins, independently of the speaker's intention.

The above analysis rests on the idea that the meaning of the tunes is compositional, but our present data only show that L*+H combines with L-!H%, while L+H* combines with L-L%. Compositionality implies, however, that the other two possible combinations of pitch accents and edge tones, L+H* L-!H% and L*+H L-L%, should also be possible and attested. A large production study indicates that these combinations are possible (Gryllia et al., 2018), though in the absence of a large corpus of prosodically annotated spontaneous Greek it is impossible to know their frequency. We would expect, however, that such combinations are rare, as they are likely to be pragmatically infelicitous in many contexts. Thus, L+H* L-!H% is possible when questions are used as a rhetorical device to introduce an alternative from a closed set; e.g. [ˈpços tileˈfonise ˈsimera] uttered with L+H* L-!H% would be the equivalent of *guess who called today* or *who do you think called today?*, when the person is someone that the speaker and addressee had been recently discussing as someone who never calls. Note that the set of alternatives is closed, so the L+H* accent on the wh-word is appropriate, while the use of L-!H% invites the addressee to select from this closed set (which in this example may include the set of people who rarely call). L*+H L-L%, on the other hand, seem less plausible pragmatically, as the pitch accent and edge tones provide contradictory cues; as a listener of synthetic stimuli of this type commented to us, "it sounds as if the speaker changed their mind half way through". Clearly, however, such gaps do not argue against compositionality, only against pragmatically odd combinatorial possibilities.

Finally, we note that the differences in interpretation may not rest exclusively on the tune, but may also involve additional cues, such as changes in speaking rate and in the duration, amplitude or quality of particular segments. Figure 5 provides an indication of such possible differences, showing that the vowel of [ˈti] when uttered with the L*+H accent (top panel) is shorter than the same vowel with the L+H* accent (bottom panel), while the opposite obtains with respect to the last vowel of [ˈvalo]. We do not dispute that such effects are possible but consider them to be part of the realization of the tunes (cf. Gryllia et al., 2018, on the predictive value of such effects in a large corpus of Greek wh-questions). On the other hand, our two follow-up experiments showed that the interpretation of the two tunes was largely unaffected by characteristics of the speaker. Though the changes in interpretation were larger for some speakers than others in Experiment 1, the differences between tunes were consistent across all of them, as was the evaluation of the tunes in terms of appropriateness, with a preference for the high-ending tune overall. The only speaker effect we found was the somewhat more negative evaluation of M1, the speaker who retained traces of a non-standard accent, and whose voice was rated less positively than that of the other speakers. Although it is not possible to ascertain whether the effect was due to accent or to voice quality, the fact that his voice was negatively evaluated but his questions were not, suggests the latter. If so, this would also suggest that the lack of difference in the evaluation of the politeness in his questions in Experiment 2 may be related to the same overall less positive impression of his voice. Findings like these indicate that speaker-specific features, such as voice quality, should be taken into consideration when testing for socio- or paralinguistic effects of intonation. However, our results at least do not lead us to believe that voice quality can affect the linguistic (i.e. pragmatic) function of intonation.

7. Conclusion

In conclusion, the production and perception results reported here provide empirical evidence for the use of two distinct melodies, L*+H L-!H% and L+H* L-L%, with Greek wh-questions. Contrary to earlier studies our data show that the melodies differ both in pitch accent and final boundary tone. These differences in phonological composition result in different pragmatic interpretations of the two tunes, such that the former is more suitable for questions used to request information, while the latter can also be used when wh-questions are non-information-seeking. Although these two different interpretations arise from the tonal composition of the tunes, the pragmatic contribution of the tunes to the overall interpretation of a wh-question varies, depending on pragmatic context and is at the discretion of the addressee, who can ultimately override the contribution of the tune and base their interpretation on pragmatic factors and the propositional content of the question: the weighting of these factors accounts for the fact that questions with the flat-ending tune can be taken at face value. These results support the contention of autosegmental-metrical intonational phonology that intonational meaning is compositional and modulated through pragmatics, information structure and propositional semantics.

Acknowledgements
We thank our speakers and experiment participants, Nikos Vergis for his assistance with Experiments 3 and 4, Ivano Caponigro for discussions on rhetorical questions, and *Language and Speech* Associate Editor Pilar Prieto and four anonymous reviewers for their helpful comments on an earlier version of this manuscript. We also thank various audiences in Greece and elsewhere for providing us with useful feedback on our results. A preliminary analysis of the main data has been published as Arvaniti, Baltazani, & Gryllia (2014).

Funding
This research was partially funded by BA grant SG160538 to Amalia Arvaniti (PI), and Mary Baltazani & Stella Gryllia (Co-Is).

References

Agouraki, G. (1990). On the projection of maximal categories: the case of CP and FP in Modern Greek. *UCL Working Papers in Linguistics* 2 (pp. 183-200).

Alexopoulou, T., & Baltazani, M. (2012). Focus in Greek wh-questions. In A. Neeleman & I. Kucerova (Eds.), *Information Structure: Contrasts and Positions* (pp. 206– 246). Cambridge University Press.

Anagnostopoulou, E. (1999). On experiencers. In A. Alexiadou, G. Horrocks, & M. Stavrou (Eds.), *Studies in Greek Syntax* (pp. 67-93). Kluwer.

Armstrong, M., & Prieto, P. (2015). The contribution of context and contour to perceived belief in polar questions. *Journal of Pragmatics* 81, 77–92.

Arvaniti, A. (In press). The Autosegmental-Metrical model of intonational phonology. In S. Shattuck-Hufnagel & J. Barnes (Eds.), *Prosodic Theory and Practice*. The MIT Press.

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

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

Arvaniti, A., Baltazani, M., & Gryllia. S. (2014). The pragmatic interpretation of intonation in Greek wh-questions. In *Proceedings of Speech Prosody* 7, 1144–1148. Available at http://fastnet.netsoc.ie/sp7/.

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. (2000). What is a starred tone? Evidence from Greek. In M. Broe & J. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon* (pp. 119–131). Cambridge: Cambridge University Press.

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., Żygis, M., & Jaskula, M. (2016). The phonetics and phonology of the Polish calling melodies. *Phonetica*, 73(3-4), 338–361.

Astruc, L., Vanrell, M., & Prieto. P. (2016). Cost of the action and social distance affect the selection of question intonation in Catalan. In M. E. Armstrong, N. Henriksen, & M. Vanrell (Eds.), *Intonational Grammar in Ibero-Romance: Approaches across linguistic subfields,* (93–114). John Benjamins.

Baltazani, M. (2002). *Quantifier Scope and the Role of Intonation in Greek*. PhD. Thesis, University of California, Los Angeles.

Baltazani, M. (2003). Broad Focus across sentence types in Greek. *Proceedings of INTERSPEECH 2003,* 89–92. Available at http://www.isca-speech.org/archive/eurospeech_2003.

Baltazani, M. (2006). Intonation and pragmatic interpretation of negation in Greek. *Journal of Pragmatics*, 38, 1658–1676.

Baltazani, M., & Jun, S.A. (1999). Focus and topic intonation in Greek. *Proceedings of the XIVth International Congress of Phonetic Sciences*, vol. 2, 1305–1308.

Baltazani, M., Kainada, E., Lengeris, A. and Nicolaidis, K. (2015). The prenuclear field matters: questions and statements in Standard Modern Greek. *Proceedings of the 18th International Congress of Phonetic Sciences.* Available at https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0941.pdf.

Bateson, G. (1972). *Steps to an ecology of mind: Collected essays in anthropology, psychology, evolution and epistemology*. San Francisco, CA: Chandler.

Bartels, C. (1997). *Towards a compositional interpretation of English statement and question intonation*. PhD. Thesis, UMass, Amherst. Available at http://scholarworks.umass.edu/dissertations/AAI9721430.

Boersma, P., & Weenink, D. (2016). *Praat: doing phonetics by computer* [Computer program]. Version 6.0.14, retrieved 11 February 2016 from http://www.praat.org/.

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H., & White, J.S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24(3), 127–35.

Botinis, A. (1998). Intonation in Greek. In D. Hirst & A. Di Cristo (Eds.), *Intonation Systems: A Survey of Twenty Languages* (pp. 288–310). Cambridge: Cambridge University Press.

Brown, L. & Prieto. P. (2017). (Im)politenes: prosody and gesture. In J. Culpeper, M. Haugh, & D. Kádár (Eds.), *The Palgrave Handbook of Linguistic (Im)politeness* (pp. 357–379). London: Palgrave Macmillan.

Brown, M., Salverda, A. P., Gunlogson, C., & Tanenhaus, M. K. (2015). Interpreting prosodic cues in discourse context. *Language, Cognition and Neuroscience*, 30(1-2), 149–166.

Büring, D. (2012). Focus and Intonation. In G. Russell & D. Graff Fara (Eds.), *Routledge Companion to the Philosophy of Language* (pp. 103–115). London: Routledge.

Büring, D. (2016). *Intonation and Meaning*. Oxford: Oxford University Press.

Caponigro, I., & Sprouse, D. (2007). Rhetorical questions as questions. In E. Puig-Waldmüller (Ed.), *Proceedings of Sinn und Bedeutung* 11 (pp.121–133). Barcelona: Universitat Pompeu Fabra.

Carlson, K., Clifton C., & Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language*, 45, 58-81.

Clopper, C. G., & Smiljanic, R. (2011). Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics*, 39(2), 237-245.

Cruttenden, A. (1997). *Intonation* (2nd edition). Cambridge: Cambridge University Press.

Degen, J., & Tanenhaus, M. K. (2015). Availability of alternatives and the processing of scalar implicatures: A visual world eye-tracking study. *Cognitive Science*, 1–30.

Dennison, H., & Schafer, A. (2017). Processing Intonationally Implicated Contrast versus Negation in American English. *Language and Speech,* 60(2), 174–199.

Douglas B., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48.

Elder, C.-H. (2017). When implicatures fail: Miscommunication in post-Gricean pragmatics. Paper presented at the *15th International Pragmatics Conference*. Belfast, 1–21 July 2017.

Féry, C., & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics*, 36, 680–703.

Franke M. & Degen. J. (2016). Reasoning in reference games: individual- vs. population-level probabilistic modeling. *PLoS ONE* 11(5): e0154854. https://doi.org/10.1371/journal.pone.0154854

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.

Georgiafentis, M., & Sfakianaki, A. (2004). Syntax interacts with prosody: VOS order in Greek. *Lingua*, 114, 935–961.

Goodwin, C. & Duranti, A. (1992). *Rethinking context: language as an interactive phenomenon.* Cambridge: Cambridge University Press.

Grice, M., Ladd, D. R., & Arvaniti, A. (2000). On the place of phrase accents in intonational phonology. *Phonology,* 17, 143-186.

Gryllia, S. (2009a). *On the nature of preverbal Focus in Greek. A theoretical and experimental approach*. PhD thesis no. 200. Utrecht: LOT Dissertation Series.

Gryllia, S. (2009b). Greek marks topics instead of foci. *The Linguistic Review*, 26(2-3), 177–205.

Gryllia, S., Baltazani, M., & Arvaniti, A. (2018). The role of pragmatics and politeness in explaining prosodic variability. *Speech Prosody* 2018. Poznań, 12-15 June 2018.

Gunlogson, C. (2003). *True to Form: Rising and Falling Declaratives as Questions in English*. New York: Routledge.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Haidou, K. (2012). *The Syntax-Pragmatics Interface of Focus Prominence in Greek*. PhD thesis, SOAS, London.

Han, C. (2002). Interpreting interrogatives as rhetorical questions. *Lingua,* 112, 201–229.

Hirschberg, J., & Ward, G. (1995). The Interpretation of the High-Rise Question Contour in English. *Journal of Pragmatics,* 24, 407-412.

Hlavac, M. (2018). *stargazer: Well-Formatted Regression and Summary Statistics Tables*. R package version 5.2.1. https://CRAN.R-project.org/package=stargazer

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, 27(1), 99–114.

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: Insight into the semantics– pragmatics interface. *Cognitive Psychology,* 58, 376–415.

Hwang, H., & Schafer, A. (2009). Constituent Length Affects Prosody and Processing for a Dative NP Ambiguity in Korean. *Journal of Psycholinguistic Research,* 38, 151–175.

Iversen, J. R., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *Journal of the Acoustical Society of America,* 124, 2263–2271.

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

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

Kotzoglou, G. (2005). *Wh-Extraction and Locality in Greek*. Ph. D. thesis, University of Reading.

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

Liberman, M. & Sag, I. (1974). Prosodic form and discourse function. In M. La Gally, R.A. Fox, & A. Bruck (Eds.), *Proceedings of the Chicago Linguistic Society* 10, (pp. 416-427). Chicago Linguistics Society.

Lüdecke, D. (2018). *sjPlot: Data Visualization for Statistics in Social Science*. doi: 10.5281/zenodo.1308157, R package version 2.6.0, https://CRAN.R-project.org/package=sjPlot.

Martí, L. (2001). Intonation, Scope, and Restrictions on Quantifiers. In K. Megerdoomian & L. A. Barel (Eds.) *WCCFL 20 Proceedings*, (pp. 372-385). Somerville, MA: Cascadilla Press.

McCoy, S. (2003). Connecting Information Structure and Discourse Structure through "Kontrast": The Case of Colloquial Russian Particles -TO, ŽE, and VED'. *Journal of Logic, Language and Information,* 12, 319–335.

Meade, A. W. & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437–55.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonation in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in Communication*, (pp. 271–311). Cambridge, MA: MIT Press.

Pinheiro, J. & Bates, D. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer.

Portes, C., Beyssade, C., Michelas, A., Marandin, J. M., & Champagne-Lavau, M. (2014). The dialogical dimension of intonational meaning: Evidence from French. *Journal of Pragmatics*, 74, 15-29.

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

Rooth, M. (1985). *Association with Focus*. Ph. D. thesis, University of Massachusetts.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics,* 1, 75-116.

Roussou, A., Vlachos, C. & Papazachariou, D. (2013). In situ, ex situ and (non) echo questions. In N. T. Lavidas, Alexiou, & A.M. Sougari (Eds.), *Major Trends in Theoretical and Applied Linguistics*: *Selected papers from the 20th International Symposium on Theoretical and Applied Linguistics (ISTAL 20)*, vol. 3, (pp. 475–494).

Sadock, J.M. (1974). *Towards a linguistic theory of speech acts*. New York: Academic Press.

Schafer, A.J., Speer, S.R., Warren, P., & White, S.D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research, 29,* 169-182.

Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: The MIT Press.

Sinopoulou, O. (2009). Απλές ερωτήσεις με ερωτηματική λέξη in situ: η περίπτωση των Ελληνικών. [Simple questions with wh-in situ: the case of Greek]. In M. Baltazani, G. K. Giannakis, A. Tsangalidis, & G. J. Xydopoulos (Eds.), *Proceedings of the 8th International Conference on Greek Linguistics*, (pp. 1118–1132).

Skopeteas, S. (2016). Information structure in Modern Greek. In C. Féry & S. Ishihara (Eds.), *Handbook of Information Structure*, (pp. 686–708). Oxford: Oxford University Press.

Steedman, M. (2007). Information-structural semantics for English intonation. In C. Lee, M. K. Gordon, & D. Büring (Eds.), *Topic and Focus: Cross-linguistic Perspectives on Meaning and Intonation*, (pp. 254–264). Dordrecht: Springer.

Steedman, M. (2014). The surface-compositional semantics of English intonation. *Language*, 90, 2–57.

Stoet, G. (2010). PsyToolkit - A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42(4)*, 1096-1104.

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology,* 44(1), 24-31.

Tomlinson Jr., J.M., Gotzner, N. & Bott, L. (2017). Intonation and Pragmatic Enrichment: How Intonation Constrains Ad Hoc Scalar Inferences. *Language and Speech,* 60(2), 200–223.

Tsimpli, I. M. (1995). Focusing in Modern Greek. In K.E. Kiss (Ed.), *Discourse Configurational Languages*, (pp. 176–206). Oxford: Oxford University Press.

Vanrell, M. M., Mascaró, I., Torres-Tamarit, F., & Prieto, P. (2013). Intonation as an encoder of speaker's certainty: Information and confirmation yes-no questions in Catalan. *Language and Speech*, 56(2), 163–190.

Venables, W. N. & Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN 0-387-95457-0.

Vlachos, C. (2010). Wh-in situ: the case of Greek. In V. Torrens, L. Escobar, A. Gavarró, & J. Gutiérrez (Eds.), *Movement and Clitics: Adult and Child Grammar*, (pp. 84–111). Cambridge: Cambridge Scholars Publishing.

Ward, G., & Hirschberg, J. (1992). The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. *Journal of Phonetics,* 20, 241–251.

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

Appendix 1. Questions (in transcription) and glosses of all 16 questions used in the production study; last column shows the number of syllables in the question. Wh-words are underlined; shaded questions were those used as stimuli in the two perception experiments.

	Question	Gloss	Syllables
1	ˈpos θa ˈpao sto ˈsidaɣma	“How will I get to Syntagma?”	8
2	ˈpços ˈpire to paɣoˈto	“Who ordered the ice cream?”	7
3	ˈti θa ˈfame ˈsimera	“What are we having today?”	7
4	ˈposa ˈpliroses stin efoˈria	“How much did you pay for taxes?”	10
5	ˈposi ˈora θa se ˈpari	“How long will it take you?”	8
6	me ˈti na tiˈlikso ta triaˈdafila	“What shall I wrap the roses with?”	12
7	ˈpu ˈine ta buˈfan	“Where are the coats?”	6
8	ˈpços ˈine o manoˈlopulos	“Who is Manolopoulos?”	9
9	ˈpu to ˈafises	“Where did you leave it?”	5
10	ɟaˈti kles	“Why are you crying?”	3
11	apo ˈpu ˈine i ˈtaɲa	“Where’s Tania from?”	8
12	ˈtora apo ˈpu na ˈstripso	“Where should I turn now?”	8
13	ˈti na su etiˈmaso	“What can I cook for you?”	7
14	ˈti na ˈvalo	“What should I wear?”	4
15	ˈti ta θes	“What do you want them for?”	3
16	apo ˈpu to ˈpires to padeˈloni	“Where did you get the trousers from?”	11

The function *sjt.lmer* in package *sjPlot* in R (Lüdecke 2018) was used to generate the tables presented in Appendices 2-4.

Appendix 2. Results of linear mixed effects model for final boundary scaling (BH vs. BL); results in bold reach statistical significance.

	BH vs. BL (Hz)		
	B	*CI*	*p*
Fixed Parts			
(Intercept)	246.86	237.65 – 256.06	<.001
Context (ref. info)	**-88.80**	**-101.83 – -75.78**	**<.001**
Speaker F2 (ref. Speaker F1)	**13.15**	**0.12 – 26.17**	**.048**
Speaker M1	**-109.44**	**-122.46 – -96.42**	**<.001**
Speaker M2	**-43.29**	**-56.31 – -30.27**	**<.001**
Context × Speaker M1	**32.69**	**14.27 – 51.11**	**<.001**
Context × Speaker M2	-13.64	-32.06 – 4.77	.147
Context × Speaker F1	**-22.83**	**-41.24 – -4.41**	**.015**
Random Parts			
σ^2		353.199	
$\tau_{00,\ Item}$		0.000	
N_{Item}		16	
ICC_{Item}		0.000	
Observations		128	
R^2 / Ω_0^2		.912 / .912	

Appendix 3. Results of linear mixed effects model of Accentual High (AH) scaling; results in bold reach statistical significance.

	AH (Hz)		
	B	*CI*	*p*
Fixed Parts			
(Intercept)	310.76	291.30 – 330.21	<.001
Context (ref. info)	-27.44	-54.53 – -0.36	.051
Speaker F2 (ref. Speaker F1)	-7.90	-34.98 – 19.18	.568
Speaker M1	**-35.58**	**-62.66 – -8.50**	**.010**
Speaker M2	**-63.34**	**-90.43 – -36.26**	**<.001**
Context × Speaker M1	**39.34**	**1.04 – 77.64**	**.044**
Context × Speaker M2	**72.00**	**33.71 – 110.30**	**<.001**
Context × Speaker F1	17.09	-21.21 – 55.39	.382
Random Parts			
σ^2		1527.323	
$\tau_{00,\ Item}$		49.080	
N_{Item}		16	
ICC_{Item}		0.031	
Observations		128	
R^2 / Ω_0^2		.219 / .217	

Appendix 4. Results of linear mixed effect model for Accentual High alignment (AH – SVO); results in bold reach statistical significance.

	AH alignment (ms)		
	B	*CI*	*p*
Fixed Parts			
(Intercept)	162.33	140.69 – 183.97	<.001
Context (ref. info)	**-148.19**	**-176.73 – -119.66**	**<.001**
Speaker F2 (ref. Speaker F1)	**-36.60**	**-65.13 – -8.06**	**.012**
Speaker M1	**-101.33**	**-129.87 – -72.80**	**<.001**
Speaker M2	**-56.60**	**-85.14 – -28.07**	**<.001**
Context × M1	**41.01**	**0.65 – 81.37**	**.046**
Context × M2	**100.97**	**60.61 – 141.32**	**<.001**
Context × F2	**66.37**	**26.02 – 106.73**	**.001**
Random Parts			
σ^2		1695.921	
$\tau_{00,\ Item}$		254.189	
N_{Item}		16	
ICC_{Item}		0.130	
Observations		128	
R^2 / Ω_0^2		.681 / .680	

Appendix 5. Stimuli (questions and responses) for Experiment 1.

	Stimulus	Information-providing response	Non-information-providing response
1	[ˈpos θa ˈpao sto ˈsidaɣma]	*You take metro line 3 and get off at Syntagma.*	*You're right, you can't go, there will be chaos.*
4	[ˈposa ˈpliroses stin efo'ria]	*A thousand euros in cash and the remaining 4000 in installments.*	*Well, the truth is, I didn't pay much.*
5	[ˈposi ˈora θa se ˈpari]	*I estimate around 3 hours.*	*You're right, it won't take me long, I'm going.*
8	[ˈpços ˈine o mano'lopulos]	*He's Kostas' brother, but they have a different surname.*	*That's true! He thinks he can get away with anything because he's the owner's son!*
12	[ˈtora a'popu na ˈstripso]	*At the pastry shop turn right.*	*Blimey, I didn't know Ipirou is a one-way street!*
15	[ˈti ta ˈθes]	*I'm gathering toys for the bazaar.*	*OK, I don't need them after all.*

Appendix 6. Tables generated using the function *sjt.glm* in package *sjPlot* in R (Lüdecke 2018). Results in bold reach statistical significance

Appendix 6.1. Experiment 1: Results of mixed effects logistic regression model.

	Response		
	Odds Ratio	*CI*	*P*
(Intercept)	1.42	1.14 – 1.80	<.003
Tune (ref. !H%)	**2.50**	**1.76 – 3.57**	**<.001**
Speaker (ref. Speaker F1)			
Speaker F2	1.27	0.91 – 1.75	>.05
Speaker M1	1.15	0.83 – 1.59	>.05
Speaker M2	**0.68**	**0.49 - 0.93**	**<.02**
Tune × Speaker F2	1.01	0.61 – 1.70	>.05
Tune × Speaker M1	**0.58**	**0.35 – 0.94**	**<.03**
Tune × Speaker M2	0.79	0.49 – 1.28	>.0.5
Observations		2442	

Appendix 6.2. Experiment 1: Results of mixed effects logistic regression model.

	Response

	Odds Ratio	*CI*	*P*
(Intercept)	1.64	1.30 – 2.07	<.001
Tune (ref. !H%)	**1.44**	**1.03 – 2.02**	**<.03**
Speaker (ref. Speaker M1)			
Speaker F1	0.87	0.63 – 1.21	>.05
Speaker F2	1.10	0.79 – 1.53	>.05
Speaker M2	**0.59**	**0.43 – 0.81**	**<.001**
Tune × Speaker F1	**1.73**	**1.06 – 2.83**	**<.03**
Tune × Speaker F2	**1.75**	**1.06 – 2.91**	**<.03**
Tune × Speaker M2	1.37	0.86 – 2.19	>.0.5
Observations		2442	

Appendix 6.3. Experiment 1: Results of mixed effects logistic regression model.

	Response		
	Odds Ratio	*CI*	*P*
(Intercept)	1.80	1.43 – 2.29	<.001
Tune (ref. !H%)	**2.53**	**1.74 – 3.71**	**<.001**
Speaker (ref. Speaker F2)			
Speaker F1	0.79	0.57 – 1.10	>.05
Speaker M1	0.91	0.65 – 1.26	>.05
Speaker M2	**0.54**	**0.39 – 0.74**	**<.001**
Tune × Speaker F1	0.99	0.59 – 1.65	>.05
Tune × Speaker M1	**0.57**	**0.34 – 0.94**	**<.03**
Tune × Speaker M2	0.78	0.47 – 1.28	>.0.5
Observations		2442	

Appendix 7. Stimuli and contexts for Experiment 2.

	Stimulus	Information-seeking context	Non-information-seeking context
1	[ˈpos θa ˈpao sto ˈsidaɣma]	*Lena, who is visiting Athens for the first time, stops*	*A public transport strike is announced for*

		somenone in the street for directions:	*tomorrow. Giannis, who must deliver an order at Syntagma tells his friend:*
4	[ˈposa ˈpliroses stin efoˈria]	*Lukas returns home after negotiations with the taxman. His friend, who's been waiting to hear the news, tells him:*	*Thanos is complaining that he pays too much in taxes. Niovi, who knows that Thanos does not give receipts in his shop, tells him:*
5	[ˈposi ˈora θa se ˈpari]	*Nikos works in an estate agency with Ksenia and is leaving for an appointment. Ksenia will be left alone in the office and she also needs to go a little later so she tells him:*	*Nikitas asks his son Alkis to go buy him some cigarettes, but Alkis protests that he's waiting for a phone call in 15 minutes. Nikitas insists because he believes Alkis will be back very quickly and tells him:*
8	[ˈpços ˈine o manoˈlopulos]	*Vassilis and Filippos are looking at old high-school photographs, when Vassilis says "Ah, there's Manolopoulos!". Filippos does not remember him and says:*	*Everyone in Lina's office is taken with the minister's new advisor, Manolopoulos. In a conversation about him, Lina, who does not like him because she believes he got his position due to nepotism, says:*
12	[ˈtora apoˈpu na ˈstripso]	*Stavros is guiding Lena who's driving him home. He forgets himself because they're talking, so when they reach a crossroads Lena tells him:*	*Dimos is driving, following Lefteri's instructions, until they reach a cul-de-sac. Dimos then tells him:*
15	[ˈti ta ˈθes]	*Fani asks her mother for 100 euros. Her mother, who is curious, tells her:*	*Fenia asks her mother for the car keys. Her mother, who does not trust her as a driver, tells her:*

Appendix 8. Tables generated suing the function sjt.lmer in package sjPlot in R (Lüdecke 2018).

Appendix 8.1. Experiment 2: Results of linear mixed effects model on appropriateness ratings; results in bold reach statistical significance.

	ZRating

	B	*CI*	*p*
Fixed Parts			
(Intercept)	0.20	-0.00 – 0.40	.053
Context (ref. info)	**0.34**	**0.29 – 0.40**	**<.001**
Tune (ref. !H%)	0.00	-0.08 – 0.09	.917
Speaker F2 (ref. Speaker F1)	**-0.13**	**-0.21 – -0.05**	**<.001**
Speaker M1	-0.04	-0.12 – 0.04	.333
Speaker M2	-0.06	-0.14 – 0.02	.126
Context × Tune	**0.13**	**0.05 – 0.21**	**<.001**
Tune × Speaker F2	0.02	-0.09 – 0.13	.725
Tune × Speaker M1	-0.03	-0.15 – 0.08	.545
Tune × Speaker M2	0.10	-0.01 – 0.21	.081
Random Parts			
σ^2		0.674	
$\tau_{00,\ Participant}$		0.108	
$\tau_{00,\ Sentence}$		0.047	
$N_{Participant}$		70	
$N_{Sentence}$		6	
$ICC_{Participant}$		0.131	
$ICC_{Sentence}$		0.056	
Observations		6702	
R^2 / Ω_0^2		.237 / .237	

Appendix 8.2. Experiment 2: Results of linear mixed effects model on appropriateness ratings; results in bold reach statistical significance.

	ZRating		
	B	*CI*	*p*
Fixed Parts			
(Intercept)	0.06	-0.14 – 0.26	.531
Context (ref. info)	**0.34**	**0.29 – 0.40**	**<.001**

Tune (ref. !H%)	0.02	-0.06 – 0.11	.583
Speaker F1 (ref. Speaker F2)	**0.13**	**0.05 – 0.21**	**<.001**
Speaker M1	**0.09**	**0.02 – 0.17**	**.019**
Speaker M2	0.07	-0.01 – 0.15	.073
Context × Tune	**0.13**	**0.05 – 0.21**	**<.001**
Tune × Speaker F1	-0.02	-0.13 – 0.09	.725
Tune × Speaker M1	-0.05	-0.17 – 0.06	.339
Tune × Speaker M2	0.08	-0.03 – 0.19	.164
Random Parts			
σ^2		0.674	
$\tau_{00,\ Participant}$		0.108	
$\tau_{00,\ Sentence}$		0.047	
$N_{Participant}$		70	
$N_{Sentence}$		6	
$ICC_{Participant}$		0.131	
$ICC_{Sentence}$		0.056	
Observations		6702	
R^2 / Ω_0^2		.237 / .237	

Appendix 8.3. Experiment 2: Results of linear mixed effects model on appropriateness ratings; results in bold reach statistical significance.

	ZRating		
	B	*CI*	*p*
Fixed Parts			
(Intercept)	0.16	-0.04 – 0.36	.120
Context (ref. info)	**0.34**	**0.29 – 0.40**	**<.001**
Tune (ref. !H%)	-0.03	-0.12 – 0.06	.508
Speaker F2 (ref. Speaker M1)	**-0.09**	**-0.17 – -0.02**	**.019**

Speaker F1	0.04	-0.04 – 0.12	.333
Speaker M2	-0.02	-0.10 – 0.06	.575
Context × Tune	**0.13**	**0.05 – 0.21**	**<.001**
Tune × Speaker F2	0.05	-0.06 – 0.17	.339
Tune × Speaker F1	0.03	-0.08 – 0.15	.545
Tune × Speaker M2	**0.13**	**0.02 – 0.24**	**.019**
Random Parts			
σ^2		0.674	
$\tau_{00,\ Participant}$		0.108	
$\tau_{00,\ Sentence}$		0.047	
$N_{Participant}$		70	
$N_{Sentence}$		6	
$ICC_{Participant}$		0.131	
$ICC_{Sentence}$		0.056	
Observations		6702	
R^2 / Ω_0^2		.237 / .237	

Appendix 9.

Appendix 9.1. Experiment 2: Results of linear mixed effects model on politeness ratings; results in bold reach statistical significance.

	ZRating		
	B	*CI*	*p*
Fixed Parts			
(Intercept)	0.01	-0.23 – 0.24	.950
Context (ref. info)	**-0.28**	**-0.33 – -0.23**	**<.001**
Tune (ref. !H%)	**-0.48**	**-0.56 – -0.40**	**<.001**
Speaker F2 (ref. Speaker F1)	**-0.28**	**-0.35 – -0.21**	**<.001**
Speaker M1	**-0.39**	**-0.47 – -0.32**	**<.001**

Speaker M2	0.01	-0.06 – 0.09	.710
Context × Tune	0.06	-0.01 – 0.14	.076
Tune × Speaker F2	**0.27**	**0.17 – 0.37**	**<.001**
Tune × Speaker M1	**0.50**	**0.40 – 0.60**	**<.001**
Tune × Speaker M2	**0.20**	**0.09 – 0.30**	**<.001**
Random Parts			
σ^2		0.559	
$\tau_{00,\ Participant}$		0.226	
$\tau_{00,\ Sentence}$		0.063	
$N_{Participant}$		70	
$N_{Sentence}$		6	
$ICC_{Participant}$		0.267	
$ICC_{Sentence}$		0.074	
Observations		6680	
R^2 / Ω_0^2		.379 / .379	

Appendix 9.2. Experiment 2: Results of linear mixed effects model on politeness ratings; results in bold reach statistical significance.

	ZRating		
	B	*CI*	*p*
Fixed Parts			
(Intercept)	-0.28	-0.51 – -0.04	.022
Context (ref. info)	**-0.28**	**-0.33 – -0.23**	**<.001**
Tune (ref. !H%)	**-0.21**	**-0.29 – -0.13**	**<.001**
Speaker F1	**0.28**	**0.21 – 0.35**	**<.001**
Speaker M1	**-0.11**	**-0.18 – -0.04**	**.002**
Speaker M2	**0.30**	**0.23 – 0.37**	**<.001**
Context × Tune	0.06	-0.01 – 0.14	.076
Tune × Speaker F1	**-0.27**	**-0.37 – -0.17**	**<.001**

Tune × Speaker M1	**0.23**	**0.13 – 0.34**	**<.001**
Tune × Speaker M2	-0.07	-0.17 – 0.03	.159
Random Parts			
σ^2		0.559	
$\tau_{00,\ Participant}$		0.226	
$\tau_{00,\ Sentence}$		0.063	
$N_{Participant}$		70	
$N_{Sentence}$		6	
$ICC_{Participant}$		0.267	
$ICC_{Sentence}$		0.074	
Observations		6680	
R^2 / Ω_0^2		.379 / .379	

Appendix 9.3. Experiment 2: Results of linear mixed effects model on politeness ratings; results in bold reach statistical significance.

	ZRating		
	B	*CI*	*p*
Fixed Parts			
(Intercept)	-0.39	-0.62 – -0.15	.001
Context (ref. info)	**-0.28**	**-0.33 – -0.23**	**<.001**
Tune (ref. !H%)	0.02	-0.06 – 0.10	.580
Speaker F1	**0.39**	**0.32 – 0.47**	**<.001**
Speaker F2	**0.11**	**0.04 – 0.18**	**.002**
Speaker M2	**0.41**	**0.34 – 0.48**	**<.001**
Context × Tune	0.06	-0.01 – 0.14	.076
Tune × Speaker F1	**-0.50**	**-0.60 – -0.40**	**<.001**
Tune × Speaker F2	**-0.23**	**-0.34 – -0.13**	**<.001**
Tune × Speaker M2	**-0.31**	**-0.41 – -0.21**	**<.001**

Random Parts	
σ^2	0.559
$\tau_{00,\ Participant}$	0.226
$\tau_{00,\ Sentence}$	0.063
$N_{Participant}$	70
$N_{Sentence}$	6
$ICC_{Participant}$	0.267
$ICC_{Sentence}$	0.074
Observations	6680
R^2 / Ω_0^2	.379 / .379

Appendix 10. Naturalness experiment: Results of multinomial logistic regression model; results in bold reach statistical significance. Tables generated using the function *stargazer* in the package *stargazer* (Hlavac 2018).

	Dependent variable:	
	so and so	odd
	(1)	(2)
Tune (ref. !H%)	0.067	0.703*
	(0.271)	(0.360)
Speaker M3	-0.568**	0.367
	(0.288)	(0.362)
Speaker F3	-0.886***	-0.091
	(0.301)	(0.384)
Speaker M2	-0.123	1.098***
	(0.282)	(0.344)
Speaker M1	0.216	0.830**
	(0.268)	(0.357)
Speaker F2	-0.441	-0.835*
	(0.273)	(0.462)
Tune × Speaker M3	0.563	-0.624
	(0.396)	(0.490)
Tune × Speaker F3	0.974**	-0.141
	(0.404)	(0.506)
Tune × Speaker M2	-0.037	-0.755*
	(0.401)	(0.458)
Tune × Speaker M1	0.444	-0.551
	(0.380)	(0.483)
Tune × Speaker M2	0.092	0.422
	(0.392)	(0.568)
Constant	-0.658***	-1.711***
	(0.188)	(0.281)
Akaike Inf. Crit.	3,306.340	3,306.340
Note:	*p<0.1; **p<0.05; ***p<0.01	

Appendix 11. Voice experiment: Results of multinomial logistic regression model; results in bold reach statistical significance. Tables generated using the function *stargazer* in the package *stargazer* (Hlavac 2018).

	Dependent variable:	
	so and so	not pleasant
	(1)	(2)
Tune (ref. !H%)	0.121	-0.405
	(0.268)	(0.303)
Speaker F2	-0.763***	-1.792***
	(0.268)	(0.359)
Speaker F3	-0.374	-1.423***
	(0.263)	(0.341)
Speaker M1	1.559***	2.096***
	(0.373)	(0.370)
Speaker M2	-0.429	-0.657**
	(0.272)	(0.296)
Speaker M3	0.347	-0.272
	(0.271)	(0.307)
Tune × SpeakerF2	-0.053	-0.080
	(0.372)	(0.561)
Tune × SpeakerF3	0.419	1.230***
	(0.369)	(0.472)
Tune × SpeakerM1	0.346	0.575
	(0.547)	(0.559)
Tune × SpeakerM2	-0.197	0.249
	(0.378)	(0.432)
Tune × SpeakerM3	0.058	0.545
	(0.379)	(0.445)
Constant	-0.019	-0.186
	(0.195)	(0.204)
Akaike Inf. Crit.	3,542.845	3,542.845
Note:	*p<0.1; **p<0.05; ***p<0.01	