

Underspecification in intonation revisited: a reply to Xu, Lee, Prom-On & Liu

Amalia Arvaniti and D. Robert Ladd

Appeared in *Phonology* 32: 537-541

We are naturally pleased that Xu and his colleagues have taken the trouble to address our critique of PENTA, and it is useful to have a concise restatement of PENTA's aims and assumptions. However, in our opinion their reply does not answer the key point of our earlier paper (henceforth A&L09), which was that syllable-by-syllable specification of F0 does not make theoretical sense in a language where F0 functions at the phrase or utterance level, and does not permit adequate quantitative modelling of complex intonation contours in short utterances.

To begin with the theoretical issue, A&L09 focused on a central problem in describing intonation, namely, the fact that contours with similar functions and globally similar shapes can apply to utterances of very different lengths. An abstract representation in terms of phonological landmarks such as local peaks provides a way of expressing the systemic equivalence of such contours irrespective of the length of the utterance to which they are applied. Defining contours in terms of such landmarks entails the existence of what we termed *sparse tonal specification*: there need not be an intonational target for every syllable, and the F0 on any given syllable may reflect nothing more than a transition between an earlier target and a later one. Conversely, in short utterances, a syllable may bear two or more intonational specifications. This idea is not, of course, original with A&L09; it is implicit in Bruce's pioneering analysis of the Swedish accent distinction (1977), and sparse tonal specification as a general principle was explicitly discussed with respect to Japanese by Pierrehumbert and Beckman (1988). The purpose of A&L09 was simply to show how this principle, in addition to making phonological sense, provides insight into various phonetic details of the contours on Greek WH-questions, and to show that the same phonetic details are difficult to account for under PENTA's assumption of *syllable-sized pitch targets*. To avoid misunderstanding, we emphasise that what we mean by this phrase is simply that each syllable has an underlying pitch specification, a *pitch target* in PENTA. The details of the F0 are determined by context in combination with these targets; the issue is whether every syllable needs an underlying pitch specification at all.

In their reply, Xu and his colleagues do not address this fundamental challenge. They simply restate the assumption (p. xxx):

PENTA's imperative for a pitch target for each syllable comes from its core assumption about speech articulation, as represented by the TA model shown in Figure 2. That is, the F0 contour of every syllable comes from a single mechanism: articulatory approximation of an underlying pitch target in synchrony with the syllable. Thus there is no other way of generating an F0 contour for a syllable besides assigning it an underlying pitch target.

They justify their unwillingness to abandon this core assumption in two principal ways. First, they believe that they have a superior conception of intonational function; second, they claim that the qTA component of PENTA is successful at modelling and predicting the phonetic detail of a wide variety of contours based on this function-centred view. We briefly address these two points in turn.

With regard to function, Xu et al. state that the autosegmental-metrical (AM) approach to intonation is concerned purely with form. This statement betrays a fundamental misunderstanding. AM phonology, like any phonological analysis, examines form together with meaning, attempting to determine which phonetic differences signal meaning distinctions. Unlike PENTA, that is, it does not assume that certain very specific communicative functions like 'focus' are easily definable and identifiable across languages. Rather, the AM literature includes several accounts of intonational meaning (e.g. Gussenhoven 1984, Pierrehumbert & Hirschberg 1990, Steedman 2014) based on the assumption that intonation can be used to encode a variety of often very broad or general pragmatic meanings, and that specific intonational nuances are determined by intonational form and context operating in tandem (see Ladd 2008, ch. 1, for further discussion). These researchers do not agree on one single analysis of intonational meaning, because, instead of defining a limited set of communicative functions a priori, AM theory considers intonational meaning to be subject to empirical investigation with ordinary assumptions about the relation between meaning and form.

As for the argument based on modelling, it has two clear weaknesses. First, any argument based on quantitative modelling needs to acknowledge that models and quantitative predictions can be reasonably successful even in the absence of sound theoretical understanding. To take an extreme example, the ancient Babylonians were able to predict eclipses with remarkable accuracy based solely on empirically observed

periodicities and without any clear idea of the earth's position relative to the sun and the moon (Steele 1997); closer to the topic at hand, Lindblom (e.g. 2004) has often cautioned against confusing phonetic 'curve-fitting' with genuine understanding. There is no doubt that Xu's early work on tonal coarticulation in Mandarin, based as it is on serious attempts to understand the physical basis of speech F0 control (e.g. Xu 1999, Xu and Wang 2001, Xu and Sun 2002), makes an important contribution to our knowledge, but the fact that it yielded a fairly accurate model of spoken F0 contours in Chinese is no guarantee that its theoretical insights into speech production are either correct or more widely applicable.

Second and more important, Xu and his colleagues have not answered our specific points about the ways in which PENTA is in principle unable to describe certain features of the Greek WH-question contours discussed in A&L09. In their section 4 they present qTA simulations of two medium-length illustrative contours, focusing primarily on the problem of stress clash. They avoid the more general problem of comparing very short and long contours, which was our central point, and they simply ignore some of our relevant findings. Space does not permit a detailed discussion, but we would note at least the following:

- They account for our finding that the nuclear high peak is aligned earlier in stress clash contexts by invoking the 'target strength' of the immediately following stressed syllable. They note that 'because there is no anticipatory mechanism in qTA', more distant stressed syllables would not be expected to have any such effect, which is consistent with A&L09. However, they do not mention our finding (A&L09: 58) that the effect of stress clash is significantly greater in short sentences than in long ones, which does seem to require look-ahead.
- Moreover, although they invoke the 'target strength' of the post-nuclear syllable to explain the effects of stress clash on the alignment of the nuclear accent peak, they go on to explain the absence of effects of stress clash on the scaling of the same nuclear accent peak by saying that 'there is no real leftward push from the first post-focus syllable'. They do not comment on the apparent contradiction between this explanation and the previous point.

- They suggest that greater ‘target strength’ on a final stressed syllable will account for the differences we report in the alignment of the sentence-final rise. They do not make clear why the contour target on a sentence-final post-focus stressed syllable should yield lower F0 (their Fig. 7, right panel) while the level target on a non-final post-focus stressed syllable should have higher F0 (their Fig. 7, left panel), though this stipulation may help them more closely approximate our empirical data for medium-length utterances. They also say nothing about the fact that stressed syllables that are neither sentence final nor immediately post-focus have no effect on F0 whatever, as clearly shown in A&L09 Figs. 1c and 2.
- More generally, they make no attempt to model the stretches of low level F0 between the post-nuclear F0 fall and the sentence-final rise. Their simulation of the contour in their Fig. 7 (right panel) shows a simple slope from the nuclear peak to the onset of the final syllable, and they even speculate that Greek WH-questions may show a ‘progressive rise throughout the sentence’, which flatly contradicts the available literature on Greek WH-questions (e.g. Botinis 1989; Grice, Ladd & Arvaniti 2000, Arvaniti & Baltazani 2005, Alexopoulou & Baltazani 2012; Arvaniti, Baltazani & Gryllia 2014 – and A&L09).

We conclude by noting a more general problem with PENTA, which is that Xu and his colleagues talk about ‘prosody’ but really mean F0. We suggest that a narrow conception of prosody as F0 is an important motivation for a model in which F0 is specified syllable-by-syllable. In Mandarin, F0 does need to be lexically specified for every syllable if it is to be properly modelled phonetically, and PENTA provides an elegant and accurate model of Mandarin F0 contours. However, because they believe that PENTA captures something fundamental about how F0 functions in all languages, Xu and his colleagues assume that F0 in any language must therefore be controlled by syllable-by-syllable specifications. But the same assumption can just as plausibly lead us to the conclusion that voice quality must be specified syllable-by-syllable in all languages as well. In some Nilotic languages, every syllable has one of two distinctive voice qualities in addition to distinctive tone and quantity; in Vietnamese and some Chinese languages, the syllable tones typically involve both voice quality and F0 specifications. Models of speech production in any of these languages will therefore

necessarily involve a voice quality specification for every syllable. But since in all languages every syllable has voice quality, and since this is created by the mechanisms of speech production, PENTA's logic suggests that any model of voice quality in any language will also necessarily involve specifications for each syllable. As voice quality in most European languages is often a matter of long-term 'settings' (Laver 1980), any such syllable-by-syllable specification, no matter how successfully it modelled phonetic detail, would necessarily miss something fundamental about how voice quality is used. We believe that the same is true of PENTA's approach to F0 in languages with utterance-level F0 patterns. Xu et al.'s reply does nothing to address this issue.

References

- Alexopoulou Theodora, and Mary Baltazani. 2012. Focus in Greek *wh*-questions. In In Ivona Kučerová & Ad Neeleman (eds.), *Contrasts and Positions in Information Structure*, pp. 206-246. Cambridge: Cambridge University Press.
- Arvaniti, Amalia, and Mary Baltazani. 2005. Intonation analysis and prosodic annotation of Greek spoken corpora. In S.-A. Jun (ed.), *Prosodic Typology: The Phonology of intonation and phrasing* (Oxford: Oxford University Press), pp. 84-117.
- Arvaniti, Amalia, Mary Baltazani, and Stella Gryllia. 2014. The pragmatic interpretation of intonation in Greek *wh*-questions. *Proceedings of Speech Prosody 7*. Online at <http://fastnet.netsoc.ie/sp7/sp7book.pdf>
- Botinis, Antonis. 1989. Discourse intonation in Greek. *Lund University, Dept. of Linguistics Working Papers* 35: 5-23.
- Bruce, Gösta. 1977. *Swedish word accents in sentence perspective*. Lund: Gleerup.
- Grice, Martine, D. Robert Ladd, and Amalia Arvaniti. 2000. On the place of phrase accents in intonational phonology. *Phonology* 17: 143-185.
- Gussenhoven, Carlos. 1984. *On the grammar and semantics of sentence accents*. Dordrecht: Foris Publications.
- Ladd, D. Robert. 2008. *Intonational Phonology* (2nd ed.). Cambridge: Cambridge University Press.
- Laver, John. 1980. *The phonetic description of voice quality*. Cambridge: Cambridge University Press.
- Lindblom, Björn. 2004. "Emergent phonology", KIT Graduate School lectures, Helsinki University. Online at <http://www.ling.helsinki.fi/kit/tutkijakoulu/courses/lindblom.shtml>
- Pierrehumbert, Janet, and Mary E. Beckman. 1988. *Japanese tone structure*. Cambridge MA: MIT Press.

Pierrehumbert, Janet, and Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In P. Cohen et al. (eds.), *Intentions in communication* (Cambridge MA: MIT Press), pp. 271-312.

Steedman, Mark. 2014. The surface-compositional semantics of English intonation. *Language* 90: 2-57.

Steele, J. M. 1997. Solar eclipse times predicted by the Babylonians. *Journal for the History of Astronomy* 28: 133-139.

Xu, Yi. 1999. Effects of tone and focus on the formation and alignment of f0 contours. *Phonetica* 27: 55-105.

Xu, Yi, and Xuejing Sun. 2002. Maximum speed of pitch change and how it may relate to speech. *Journal of the Acoustical Society of America* 111: 1399-1413.

Xu, Yi, and Q. Emily Wang. 2001. Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication* 33: 319-337.