ON THE RELATIONSHIP BETWEEN PHONOLOGY AND PHONETICS (OR WHY PHONETICS IS NOT PHONOLOGY)

Amalia Arvaniti

Department of Linguistics, UCSD
amalia@ling.ucsd.edu

ABSTRACT

In this presentation, I argue that unifying phonetics and phonology in the grammar has undesirable consequences. Evidence for this position is provided from various sources, but focuses on intonation, an area of linguistic structure that has often been viewed as not requiring an abstract phonological representation.

Keywords: phonetics, phonology, intonation, gradience

1. INTRODUCTION

This presentation deals with the relationship between phonetics and phonology; in particular, it argues in favor of the "classic" view according to which phonetics and phonology are distinct components of the grammar. In what follows, I present a brief overview of the changes that have taken place in the past 40 years and which have led many to dispute the need for separate phonetics and phonology components. I then present general arguments in favor of the separation of the two levels, and use intonation as a case in point, since intonation has often been said not to fit the phonetics-phonology dichotomy.

2. THE RELATIONSHIP BETWEEN PHONETICS AND PHONOLOGY

The classic generative model of linguistics provides a straightforward view of the relationship between phonetics and phonology. According to Chomsky and Halle [7], phonological representations, seen as classificatory feature bundles, are converted by phonological rules into a phonetic representation in which features take on scalar values. Crucially, the "total set of features is identical with the set of phonetic properties that can in principle be controlled in speech; they represent the phonetic capabilities of man, and […] are therefore the same for all languages (p. 295).

Phonetic and sociolinguistic studies, however, have shown for some time that this view is simplistic: (a) repetitions of the same utterance by the same speaker can differ substantially from one another; (b) speakers of the same linguistic variety show systematic differences in the realization of the same category; (c) different realizations of the "same" category (say, segments transcribed as [s]) are also found cross-linguistically.

In phonetics, intra- and inter-speaker variation within the same linguistic variety has been typically viewed as noise to be stripped from the signal, as the use of normalization models clearly demonstrates (see e.g. [16] for a review of normalization models for vowels). On the other hand, cross-linguistic differences have often been said to have a functional basis. For example, Cohn [8] showed that vowel nasalization is much more extensive in English, in which it is not distinctive, than in French, which contrasts oral and nasal vowels. Eventually, an abundance of such language-specific findings led to the development of more sophisticated models of the relationship between phonetics and phonology, such as Keating's "window" model of coarticulation [17].

As far as I can tell, these early models do not *explicitly* claim that phonetics should be part of the grammar (with the exception of the work of Robert Port and colleagues), but this assumption is implicit, since language-specific patterns that do not follow from general phonetic principles must be acquired. Details aside, these models are based on the common assumption that phonology deals with abstract formal units, while phonetics deals with gradient phenomena [e.g. 30], and thus that the two are distinct components of the grammar. This division of labor between phonetics and phonology is evident in much of the laboratory phonology work and beyond [e.g. 8, 20, 30].

On the other hand, research within OT has led to a different conclusion, namely that the phenomena we describe as phonetic or phonological are not qualitatively distinct and thus are best treated as part of a uniform 'p' component of some sort, in which contrastive and non-contrastive properties co-exist [e.g. 10, 26, 31].

Proponents of this view have presented evidence from various phenomena that, they argue, are best explained if we do not assume that phonetics and phonology are separate. One main assumption of these models is that phonetic considerations (can) drive phonological patterns; e.g. both Flemming [10] and Zhang [31] have (re)formalized Lindblom's basic idea of perceptual distance [21]. Yet, as shown below, such functional explanations (which abound in this type of literature) cannot account for many attested patterns of variation (and are rarely put through rigorous testing by their proponents). Further, we know that production and perception do not always match, as in the cases of near-mergers and incomplete neutralization where speakers reliably produce distinctions they cannot reliably perceive [18]. Finally, there is at present no good metric of what a sufficient perceptual distance might be *cross-linguistically* (if one is in doubt, listening to, say, Nunngbuuyu coronals should convince them).

The above does not mean to denigrate the need to incorporate rich phonetic detail into our models of grammar. There is ample evidence that such information is stored and used in both production and perception. We know, for example, that many phonetic patterns stubbornly resist functional explanations: thus, if we explain the extensive nasalization of American English on the grounds that nasalization is not contrastive in this system [8], we have no explanation for the lack of extensive nasalization in British English. We also know now that listeners do not remove variation from the signal in order to understand it: rather, they use this detailed information in ways that affect both perception [5, 27] and production [15]. Yet, at the same time, speakers also operate on abstractions, e.g. when children correctly use inflection in forms they have not previously encountered or when adults create new lexical items (such as *irregardless*) by analogy.

Empirical evidence for the need for two levels of representation comes from several additional sources. Near-mergers suggest that speakers can make a fine-grained distinction in production but at a more abstract level categorize their productions as "the same" [14, 18]. The reverse applies in "covert contrasts," cases of children with phonological disorder (PD) who are said to substitute one phoneme for another but turn out, on close phonetic inspection, to use two different realizations (which are categorized as "the same" by the therapists) [25]. Children with PD also produce the same inaccuracies when learning new word forms with low or high probability diphones; in contrast, children with Specific Language Disorder (SLI) have greater difficulty learning new word forms if these involve low-probability diphones [3]. This difference between children with PD and SLI can be plausibly attributed to the inability of the former to create successful fine-grained token representations vs. the inability of the latter to make coarse-grained generalizations from word tokens to phonotactics [3].

Given such evidence, advocating a single phonetics-phonology component seems premature, to say the least. In what follows I show that even in intonation, the nature of which has often been disputed, it is easy to see why phonological abstraction is necessary.

3. INTONATION AND THE "INTERFACE"

In what follows I will use the term *intonation* to refer to the linguistically structured and *pragmatically* meaningful modulation of F0, leaving aside affective uses of fundamental frequency.

As mentioned, in intonation the boundaries between phonetics and phonology have often been blurred. This is largely due to the prevalent view that, unlike segmentals, F0 changes are gradient (see [12] for a review). Given the widespread view that gradient phenomena are not phonological, intonation has often been seen as not having phonological structure.

Yet it is hard to tell what this statement about the gradience of intonation really refers too. It is indeed the case that when we compare visual representations of speech—waveforms and spectrograms—to the output of any pitch tracker, the course of F0 is smooth compared to the changes evident at the segmental level. However, even if F0 presents problems of segmentation for researchers, there is no good evidence that these difficulties are also present in the production and perception of intonation. At present, it is not clear what exactly the native speakers of a language perceive as intonation, but there is no strong evidence that they hear a contour as a continuously modulated curve, while there is evidence to the contrary [9, 13]. As will be shown in some detail below, production also strongly suggests that speakers do not treat the F0 curve as continuous, but regulate particular points with precision [2], leaving others to vary.

What is perhaps a more valid reason for the view that intonation is gradient is that, unlike segments, intonation relies primarily on F0, and F0 modulations do play a large role in expressing phenomena that Ladd [19] has termed "paralinguistic".

How can we separate linguistic from paralinguistic aspects of F0 use? We could start by making a distinction between *meaning* and *significance. Meaning* is the product of the semantics of the words in the utterance, their syntax and the (pragmatic) context in which the utterance is used *in combination* with its intonation and information structure; this is the linguistic meaning of intonation. We can use *significance* to refer to anything else that F0 can convey, that is paralinguistic information. Additionally we could define as paralinguistic those aspects of F0 use that exhibit gradience both in form and in significance; e.g. excitement can be conveyed by greater range and span, and both range and span *and* their significance (more or less excitement) are gradient.

This view of the meaning of intonation is largely adopted in research within the autosegmental-metrical framework of intonational phonology (henceforth AM) [19]. This framework is based on the idea that the meaning of melodies is compositional and context-dependent [23]. (This of course is not a novel idea but applies to the meaning of an utterance whether intonation is taken into account or not.)

Crucially, the AM framework also views the form of intonation as comparable to other components of phonological structure. In AM, melodies consist of a series of tones at the phonological level; these tones are phonetically realized as tonal targets, with specific scaling and alignment with the segmental string; the alignment and scaling of the targets depends on the phonological association of the tones with the heads and boundaries of prosodic structure. As happens with all other phonological representations, a tonal entity is contrastive within the system it is part of, may be realized in different ways depending on context, and need not have identical phonetics across languages. In my view, the success of AM is largely due to this separation of phonetics and phonology, and all that it entails.

3.1. Phonetics, phonology and contour shape

Despite the success of AM, in the past few years theories that take a more surface-phonetic approach, albeit in two different directions, have also emerged [11, 29]. In the model developed by Grabe and colleages [11] the aim is to mathematically model entire melodies seen as rises, falls, rise-falls etc. In contrast, in the framework developed by Xu and colleagues [29] the emphasis is on modeling the F0 contour at a micro level: the F0 of each syllable is said to be specified, in contrast to the AM tenet that only a few syllables are tonally specified and the F0 of the rest is derived by interpolation.

These theories do not explicitly say that they make no distinction between phonetics and phonology, but their focus on surface phonetic forms and some of their formalisms suggest that this is so. They also share the assumption that there is a direct relationship between meaning and its phonetic expression through F0, a view that harks back to the naturalness of F0 [22], a tenet for which no good evidence has been found [4, 6]. In what follows, I address the challenges that these models face largely due to these assumptions.

Holistic models [11] rest on the traditional view that particular melodies are associated with specific meanings, attitudes or, more often, "discursive functions" such as "question" and "statement." But intonational meaning is not as straightforward as such descriptions suggest, because the relationship between melody and meaning is not one-to-one: the same melody may be used to express different meanings, depending on context, and the same meaning can be expressed by different melodies. For example, *That was really fun*! uttered with H* !H* L-L% will not persuade the hearer than the speaker had fun, but *That's five euros* uttered with the same contour is perfectly unremarkable. Similarly, particular functions, such as irony, are not expressed by some specific melody [6]. Thus, *Would you like some fries with your ketchup?* can be uttered with exactly the same intonation as *Would you like some ketchup with your fries?*, but most hearers would interpret only the former as ironic.

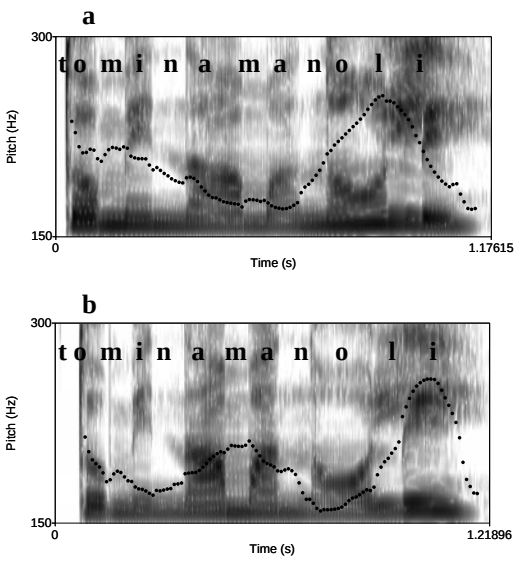
Even when context and discourse function remain stable, utterances will exhibit different contours, because contours depend partly on the length and number of words in the utterance, and the location of stressed syllables and boundaries.

A case in point is the default contour of Greek polar questions [2]. Greek polar questions with two content words are produced with one of two

intonation patterns: if the focus is on the first word, the stressed syllable of that word is low in pitch and is followed by a low-level stretch that ends with a rise on the last stressed syllable of the question and a subsequent fall (Fig. 1a); if the focus is the last word, its stressed syllable is low in pitch and is followed by a rise-fall that appears on the last syllable of the question (Fig. 1b).

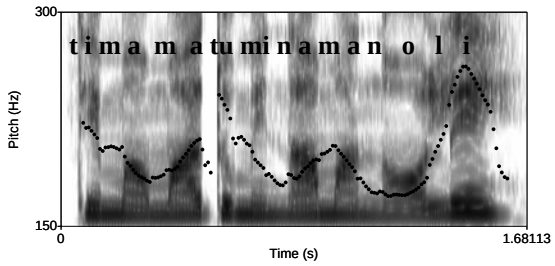
It is difficult to reconcile these data with an account that seeks to model the entire contour as a unit that expresses question intonation: the contour in Fig. 1a shows one clear rise-fall, while the contour in Fig. 1b shows two such movements (the first one due to a prenuclear pitch accent), making the same overall phonetic representation impossible. Yet, Greek native speakers feel that, despite the differences, both utterances have essentially the same pitch countour and the same pragmatic function: they are polar questions. Thus giving them distinct representations misses a significant generalization that is clearly part of the native speakers' grammar. The situation is in fact worse than Fig. 1 makes it appear. As can be seen in Fig. 2, questions with more content words can have even more complex contours. The differences increase further by the fact that questions with three content words, like the question in Fig. 2, can have focus on any one of these words, a change that radically alters the overall contour shape.

Figure 1: **(a)** [to mi'na ma'noli] (a Greek first and last name) with focus on [mi'na]; (b) [to mi'na ma'noli] with focus on [ma'noli].



In addition the final rise-fall movement appears earlier in Fig. 1a than in Fig. 1b; this difference is difficult to capture in a model that represents only the gross contour shape, yet without this distinction the contours are ill formed. These differences can be more elegantly captured by the AM phonological representation (L*+H) L* L+H-L%, which can account for all the contours shown in Figs. 1 and 2, and for the variation among them.

Figure 2: [ti ma'ma tu mi'na ma'noli] 'the mother of Mina Manoli' with focus on [ma'noli].



3.2. Phonology and phonetic detail

Detailed representations, as in Xu [29], might seem as an obvious solution to these problems. However, these models too run into difficulties.

First, a very detailed model would be as likely to miss a significant generalization about the contours because, if examined syllable-by-syllable, the course of F0 can be totally different in utterances such as those in Fig. 1a and Fig. 1b; e.g. the syllable [ma] has low F0 in Fig. 1a but high falling F0 in Fig. 1b. In AM this results from this syllable's lack of tonal specification, in combination with the tonal context in which it is found. Thus the phonetic differences are not expected to be meaningful. In [29] however such differences are meant to be part of the "symbolic representation" of intonation and thus, presumably, meaningful. Yet, it is clear that the meaning of these utterances is not affected by the exact F0 trajectory on the syllable [ma] or other similar syllables (see also [9] for supporting evidence).

Further, even if we can adjust various parameters within Xu's system (such as *strength*) in order to achieve the right F0 on syllables like [ma], such adjustments require more complex representations that those advocated by AM. But if we can generate these contours by simply specifying the location of a few tonal targets, instead of specifying the F0 of each syllable, this simpler solution should be preferred.

In short, a model in which the F0 of every syllable in an utterance is specified is clearly too powerful. As a result, the model cannot explain

why some syllables show different F0 in very similar contours; on the other hand, such a model can generate a variety of unattested and quite possibly uninterpretable contours.

3.3. Learnability issues

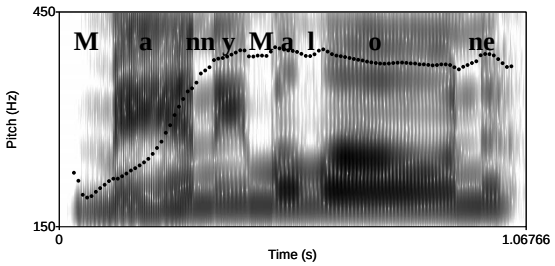
It is difficult to conceptualize how either of the models discussed above could account for the acquisition of intonation. In a model in which rich phonetic detail regarding F0 is stored (and no further abstraction is expected), it is hard to imagine how speakers would assemble and produce a melody, as this process would require some abstraction from individual word tokens. On the other hand, if the knowledge of the learner is limited to some overall melody it is hard to envision how she would correctly "fit" this overall melody to utterances of varying lengths without at least some more abstract knowledge of prosodic structure, number of accented words and the like.

3.4. Intonation and focus

One of the strongest claims in favor of a direct link between intonation and function is the finding that focus has concrete and predictable reflexes on F0 and affects particularly F0 range [29]. Furthermore, it has been suggested that languages manipulate not only the pitch of the word in focus, but also reduce the pitch of any following material, to make the focused item more salient [29].

This claim, however, is based on a small number of languages and on statements with narrow contrastive focus. Although such focus-related F0 changes are certainly attested, they are neither universal nor do they apply to all melodies. In English, for example, it is clear that there is no pitch range suppression when the contour is rising, as in the question shown in Fig. 3. Further, we know that not all languages use intonation to signal narrow focus. For example, Swerts et al. [28] show that this is not possible in Italian NPs, even though this constraint renders utterances such as *BLACK circle* and *black CIRCLE* ambiguous. On the other hand, Taiwanese, a tone language, manipulates duration rather than pitch range to show narrow focus [24]. Finally, the Greek question data (as well as the English example in Fig. 3) show that no general claims can be made about pitch rises and focus: in the Greek polar questions, the focus of the question shows low pitch, usually the lowest in the entire question. In addition, in longer utterances the focused word marks the beginning of a low-level stretch which can continue for several syllables [1]. In this case then, the focused word is not *tonally* set apart from the rest of the utterance in any way.

Figure 3: *Manny Malone* uttered as a question with narrow focus on *Manny*.



Taken altogether, these results suggest that it is not possible to claim a direct relationship between focus and F0, and by extension, I would argue, of any other function of intonation and F0 (as argued also in [20]). Even if a "natural" connection could be shown to exist between certain pragmatic meanings and their intonational realization, its existence would not disprove the view that this connection is mediated by phonological structure.

4. CONCLUSION

The discussion above has demonstrated that in the case of intonation, as with all other 'p' phenomena, models that do not distinguish between coarse-grained abstractions, which we could call phonological representations, and fine-grained phonetic detail cannot account for certain properties of intonation that have been shown to exist in a variety of languages. Further, even if they can explain certain aspects of intonation—such as narrow focus in declaratives—without the use of phonological representations, it does not follow that these representations can be dispensed with: if they are needed to account for at least some aspects of intonation—such as the shape of a contour and the expression of focus in a variety of contexts and languages—then it makes sense to assume that they are always present.

Now, if even intonation, the last bastion of the prosody-is-just-phonetics approach, requires phonological structure, it becomes much more difficult to argue that a phonological level of abstraction is unnecessary.

To conclude, it seems to me that progress in intonation research will be achieved if our working hypothesis is that intonation operates along the same lines as other 'p' phenomena (until we have

some evidence to the contrary at least): this means that we should expect compositionality of form, occasional lack of phonetic transparency both within and across systems, and contrastiveness of primitives within a system. This last point is the most important, as it suggests not only that we need to examine both the form *and* the meaning of intonation, but also—and most crucially—that the relationship between meaning and phonetics is going to be mediated by the type of coarse-grained abstraction that is usually known as phonology.

ACKNOWLEDGMENTS

I would like to thank Carlos Gussenhoven for giving me the opportunity to express my ideas in this session, and for his very helpful comments on an earlier draft. Thanks are also due Aditi Lahiri, Ryan Shosted, Alex Del Guidice and Cynthia Kilpatrick, for their suggestions, and to Laura Kertz and Rebecca Colavin for some very stimulating discussions during my 2007 Intonational Phonology seminar.

REFERENCES

[1] Arvaniti, A. in prep. On phrase accents and postnuclear pitch accents: Greek polar questions revisited.

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

[3] Beckman, M., Munson, B., Edwards, J. to appear. Vocabulary growth and the developmental expansion of types of phonological knowledge. In: Cole, J., Hulde, J. H. (Eds.), *Laboratory Phonology 9: Change in Phonology*. Berlin, New York: Mouton de Gruyter.

[4] Bryant, G. A., Fox Tree, J. E. 2005. Is there an ironic tone of voice? *LangSp* 48, 257-277.

[5] Clopper, C. G., Pisoni, D. Some acoustic cues for the perceptual categorization of American English regional dialects. *JPhon* 32, 111-140.

[6] Chen, A., Gussenhoven, C., Rietveld, T. 2004. Language-specificity in the perception of paralinguistic intonational meaning. *LangSp* 47, 311-349.

[7] Chomsky, N., Halle, M. 1968. *The Sound Pattern of English*. New York: Harper and Row, Publishers.

[8] Cohn, A. 1993. Nasalization in English: phonology or phonetics. *Phonology* 10, 43-81.

[9] Dilley, L. C., Brown, M. to appear. Effects of pitch range variation on f_0 extrema in an imitation task. *JPhon*.

[10] Flemming, E. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18, 7-44.

[11] Grabe, E., Kochanski, G., Coleman, J. to appear. Quantitative modelling of intonational Variation. *Proceedings of Speech Analysis and Recognition in Technology, Linguistics and Medicine 2003.*

[12] Gussenhoven, C. 1999. Discreteness and gradience in intonational contrasts. *LangSp* 42, 283-305.

[13] 't Hart, J., Collier, R., Cohen, A. 1990. *A perceptual study of intonation*. Cambridge: Cambridge University Press.

[14] Hay, J., Warren, P., Drager, K. 2006. Factors influencing speech perception in the context of a merger-in-progress. *JPhon* 34, 458-484.

[15] Harrington, J. 2006. As acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. *JPhon* 34, 439-457.

[16] Johnson, K. 2005. *Speaker normalization in speech perception*. In: Pisoni, D., Remez, R. E. (Eds), *The Handbook of Speech Perception*. Oxford: Blackwell, 363-389.

[17] Keating, P. 1988. The phonology-phonetics interface. In: Newmeyer, F. J. (Ed.), *Linguistics: The Cambridge Survey*, vol. 1. Cambridge: Cambridge University Press, 281-302.

[18] Labov, W., Karen, M., Miller, C. 1991. Near-mergers and the suspension of phonemic contrast. *LVC* 3, 33-74.

[19] Ladd, D. R. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.

[20] Ladd, D. R., Scobbie, J. M. 2003. External sandhi as gestural overalp? Counter-evidence from Sardinian. In: Local, J., Ogden, R., Temple, R. (Eds.), *Phonetic Interpretation: Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press, 164-182.

[21] Lindblom, B. 1986. Phonetic universals in language systems. In: Ohala, J., Jaeger, J. (Eds.), *Experimental Phonology*. New York: Academic Press, 13-44.

[22] Ohala, J. J. 1984. An ethological perspective on common-cross language utilization of F0 of voice. *Phonetica* 41, 1-16.

[23] Pierrehumbert, J., Hirschberg, J. 1990. The meaning of intonational contours in the interpretation of discourse. In: Cohen, P. R., Morgan, J., Pollack, M. E. (Eds.), *Intentions in Communication*. Cambridge, MA: The MIT Press, 271–311.

[24] Pan, H., 2007. Focus and Taiwanese unchecked tones, In: Lee, C., Gordon, M., Büring, D. (Eds.), *Topic and Focus: Cross-linguistic Perspectives on Meaning and Intonation.* Springer.

[25] Scobbie, J. M., Gibbon, F., Hardcastle, W. J., Fletcher, P. 2000. Covert contrast as a stage in the acquisition of phonetics and phonology. In: Broe, M. B., Pierrehumbert, J. B. (Eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, Cambridge: Cambridge University Press, 194-207.

[26] Steriade, D. 2000. Paradigm-uniformity and the phonetics-phonology boundary. In: Broe, M., Pierrehumbert, J. (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*. Cambridge: Cambridge University Press, 313-334.

[27] Strand, E. A. 1999. Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology* 18, 86-99.

[28] Swerts, M., Krahmer, E., Avesani, C. 2002. Prosodic marking of information status in Dutch and Italian: A comparative analysis. *JPhon* 30, 629-654.

[29] Xu, Y., 2005. Speech melody as articulatorily implemented communicative functions. *Speech Comm.* 46, 220–251.

[30] Zsiga, L. 1997. Features, gestures and Igbo vowels: an approach to the phonetics-phonology interface. *Lg* 73, 227-274.

[31] Zhang, J. 2000. Non-contrastive features and categorical patterning in Chinese diminutive suffixation: MAX[F] or IDENT[F]? *Phonology* 17, 427-47.