IS THIS IN THE PHONOLOGY? EXAMINING THE INTONATIONAL PHONETICS-PHONOLOGY INTERFACE WITH AMERICAN ENGLISH POLAR QUESTIONS

Z.L. Zhou & Byron Ahn

UCLA, Princeton
zlzhou@ucla.edu, bta@princeton.edu

ABSTRACT

In this paper, we discuss *a priori* unexpected pitch movements (spurious pitch movements; SPMs) preceding the L* in rising MAE polar questions. In experimental data, we find two non-canonical contours: a rise-fall SPM with a peak, and a steady-high-and-fall SPM with a plateau. In both, the alignment and scaling of SPMs is more variable than might be predictable with MAE_ToBI as currently defined.

We present a linear mixed effects model which shows that SPMs reflect fine-grained detail related to fluency and emotional state, but also the semantico-pragmatics of the discourse context: they appear to be correlated with fewer expectations about possible answers. Our results necessitate a model in which this type of phonetic variation can be understood as linguistically structured and motivated.

Keywords: Intonation, phonetics-phonology, intonational meaning, questions, phonetic variation

1. INTRODUCTION

Up to a certain point, any phonologically defined intonational contour may exhibit variation in scaling and alignment without crossing categorical boundaries. Phonetic variation of this sort is not always without meaning: it is well documented by investigations of segmental phonetics-phonology that variation within phonological categories can be indexed to particular meanings or discourse effects. The key questions for this paper are: How phonetically different can f0 contours for a particular phonological representation be? What kinds of variation do we find within a category, and is any of it meaningful?

We explore these questions in the domain of polar questions (PQs) in Mainstream American English (MAE). We take as our starting point the phonological model of MAE_ToBI [2], which establishes clear predictions about which f0 contours should be found. We find variation in PQ intonation beyond the predictions of this model, variation which suggests that f0 contours may contain pitch movements that are not directly specified in the phonological representation. Instead, this variation appears to be conditioned by both linguistic and non-linguistic factors.

2. BACKGROUND

MAE_ToBI is a phonological model of intonation in the Autosegmental-Metrical tradition [5]. Three core elements of this model are laid out in Table 1, along with their phonological associations.

Table 1: Some MAE_ToBI elements

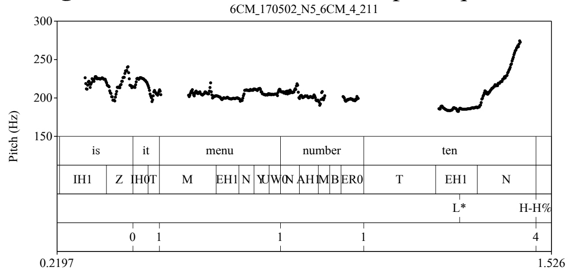
Phonological item	Phonological association
pitch accent (T*)	stressed syllable
phrase accent (T-)	ip right edge
boundary tone (T%)	IP right edge

MAE_ToBI is assumed to involve a relatively direct phonology-to-phonetics mapping, with all major points of inflection in the f0 contour corresponding to phonological elements in the underlying representation. Relatively few phonological processes are assumed, primarily downstep (and upstep).

In this way, the phonological association of a pitch accent (PA) or phrase accent is thought to directly map on to its phonetic alignment. For example, an L* phonologically associated with a syllable will be phonetically realized as low f0 within that syllable; no other phonetic pitch movements related to this PA are predicted. In addition, phrase accents and boundary tones correspond directly to a fixed maximum number of pitch movements. For example, an H- is realized as high f0 anchored to the ip edge that spreads leftward, space permitting; an H% is realized as (upstepped) high f0 within the IP's final syllable. Between specified pitch targets, f0 is derived by interpolating from one target to the next.

As such, MAE_ToBI and models like it predict that an L* H-H% contour (the core of the canonical polar question contour; [6]) will be realized as a gradual fall in f0 from the initial f0 of the utterance to the low f0 of the prominent syllable that realizes the L*. This gradual fall will be followed by a sharp rise to high, with a final (extra-)high on the IP's final syllable. An example of this is given in Fig.1.

Figure 1: A canonical L* H-H% polar question



However, as mentioned in §1, not all questions intonationally match this canonical tune. First, questions can occur with different PAs and boundary tones, depending on the question's meaning (e.g., [7]). Second, even within L* H-H% contours, there is some amount of phonetic variation across speakers and contexts. These facts lead to our research questions:

(i) How much phonetic variation is there in PQs?
(ii) Does phonetic variation affect interpretation?

3. METHODS

All participants were college-aged students attending Princeton University. In total, there were 20 participants, 9 male and 11 female. One participant was excluded for not being a native speaker of MAE.

In a sound-attenuated booth, participants played a modified version of *Guess Who?*, a board game where players choose a game object and ask polar questions about the properties of the other player's object. This task was chosen to maximally control for contextual and information-structural conditions between utterances while also eliciting a large number of naturalistic polar questions.

We modified the game board so that each game object was a menu of food items. These items were chosen such that each food item had sonorants or voiced consonants preceding the primary stress. A full list of items is presented in Table 2.

Table 2: Stimulus list

initial	peninitial		postpeninitial
raspberries	bologna	baguette	edamame
oyster	A1 sauce	romesco	marinara
marzipan	banana	meringue	amaretto

At the start of each round, participants were instructed to feel either excited or neutral in alternating rounds, in order to gauge the effects of emotional state. Within each round, participants were instructed after each turn to make a guess in the form of a polar question as to the identity of the experimenter's game object. Because of this, game rounds would occasionally end quickly if a lucky guess was made; to ensure adequate data collection, participants played rounds until they had played approximately 20 minutes of *Guess Who?*. Because of the rules of the game, each speaker produced a large number of naturalistic, unscripted PQs under similar, but different contexts.

The data was then transcribed in Praat and forced-aligned using the Montreal Forced Aligner [3, 4]. The data was intonationally annotated as best as possible with MAE_ToBI and further annotated for ABOUTNESS, which coded whether the utterance asked (a) about menu contents, (b) about the menu number, (c) about game rules, (d) a previously asked question, or (e) a non-game-related question.

To limit our scope, we focus on PQs with a final rise (e.g., H-H%) and a L* as the nuclear pitch accent (NPA), resulting in 1,011 total observations. Of these, 154 were disfluent and thus difficult to analyze. They were excluded, leaving a corpus of 857 PNQs from 19 speakers.

4. QUALITATIVE OBSERVATIONS

Addressing (i), we find that there is non-trivial variation within the set of PQs with final rises. Contra traditional assumptions about the canonical PQ contour (cf. §2), we regularly observed contours with additional pitch movements preceding the L* of the NPA syllable. These variants can be seen as falling into two types: a rise-fall pattern with a peak preceding the L* NPA (labelled as 'h' in Fig.2), and a steady-high-and-fall pattern with a high plateau preceding the L* NPA (labelled as '-h' in Fig.3). (See [1] for sound files and other examples.) For reasons that will become clear shortly, we call these peaks/plateaux spurious pitch movements (SPMs).

SPMs are regularly anchored to lexically unstressed syllables; as such, they cannot be analyzed as pitch accents, which must be linked to stressed syllables in an Autosegmental-Metrical analysis like ToBI; [5]). As such, using existing ToBI labels (e.g., a H* in Fig.2) would be inappropriate, as H*s must be associated with lexically stressed syllables. (The /nju/ of "menu" is unstressed.)

Figure 2: A spurious pointwise high

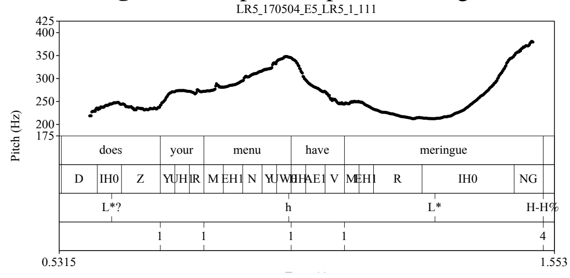
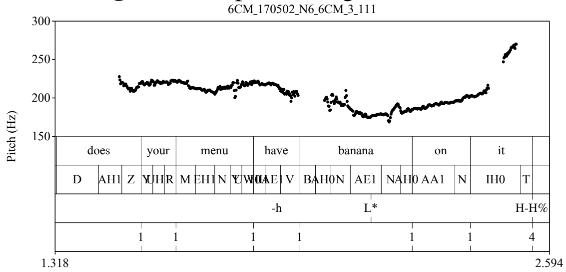


Figure 3: A spurious high cover-tone



Moreover, their alignment is not fixed — they were found to be anchored up to three syllables away from the NPA — so they cannot be described as the leading/trailing tone of a complex PA: e.g., there are two syllables between the peak and the L* in Fig.2. Finally, the scaling of these SPMs is unusual: SPMs are not consistently at the same height as identifiable phonological high targets like H* and H-.

Thus, the alignment and scaling of these SPMs in our data fall outside of the predictive domain of MAE_ToBI's currently established phonological inventory of intonational objects and processes. This is the sense in which the pitch movements are "spurious", though we also aim to understand what determines their distribution.

5. ANALYSIS

To address (ii), we conducted a logit linear mixed effects model in R, using the `lme4` package. To predict the likelihood of the presence of an SPM, this model used EMOTIONal state, QUESTION NUMBER, ROUND NUMBER, and question ABOUTNESS as fixed effects and SPEAKER as a random effect. We sequentially introduced fixed effects and then compared models using the `anova()` function, stopping once the addition of further variables no longer improved model fit. Using the `AIC()` function resulted in the same selection for best model.

Our model showed main effects of QUESTION NUMBER and ABOUTNESS for these final-rise PQs. We find that as QUESTION NUMBER increases during a round, the probability of a speaker producing an SPM decreases ($p = 0.04$). Each new round essentially resets the amount of information in the Common Ground (CG; see [8] for an overview) of the discourse, and the mechanics of the game require players to ask questions so that their opponent incrementally adds to the CG. Thus, an increase in the QUESTION NUMBER in this game generally tracks an increase in the amount of information a speaker had access to (in the CG) for their questions. SPMs occur most when there is less information in the CG about the domain of inquiry; thus SPMs are more likely when the speaker has less information about the relevant aspects of the state of the world.

As for question ABOUTNESS, questions about menu contents were most likely to be accompanied by SPMs ($p < 0.001$). Questions about game objectives were more likely to be accompanied by SPMs than others ($p = 0.01$): i.e., more than ones about menu number, ones about game rules, and previously asked ones. We interpret this as indicating that SPM likelihood depends on the type of pragmatic move a speaker is making with their question.

We hypothesize that when a player asked about menu contents or game objectives, the player had fewer expectations about the answer, because there was less relevant information in the CG. (The number of possible items on the menu was quite large [especially earlier in the game] or they were naive about the game.) On the other hand, when a player asked a question that served as a guess about the the opponent's menu number, they may have had clearer expectations about the outcome of the possible answer (even earlier in the game, when they may simply expect a 'no' answer). Thus, similar to QUESTION NUMBER, ABOUTNESS can also be seen as a proxy for the speaker's expectations about the possible/likely responses, given what they know about the context. Askers of questions about menu contents (e.g., Fig.2) were less likely to have developed expectations, and were the questions were therefore "genuniely" information-seeking. This adds further support to the view that SPMs occur less frequently when the CG contains more relevant information.

In contrast, questions about menu number (e.g., Fig.1) had the form of a PQ (recall that all PQs in this data set have final rises), but were not necessarily seeking out information in the same way. Instead, the speaker may have been using a PQ to have a pragmatic effect of suggesting a guess, meaning that PQs about menu number may have had categorically different semantic/pragmatic representations (it is not clear that they even had the illocutionary force of a question). In this way, SPMs may be seen as occurring more in utterances with interrogative force (i.e., "true questions").

Our model also showed a main effect of ROUND NUMBER: as the experiment continues and ROUND NUMBER increases, the probability of an SPM increases ($p < 0.001$). This is the opposite of the effect of QUESTION NUMBER. Unlike QUESTION NUMBER, which tracks the amount of information in the CG, we treat ROUND NUMBER as tracking familiarity of the task. As speakers gained experience playing the game, they were less distracted by game mechanics, and were better able to use more fluent, connected speech that encompasses a richer set of intonational cues. In this way, SPMs can be seen as occurring in more natural conversational contexts, as opposed to, e.g., reading tasks.

Finally, we found a main effect of EMOTION, showing that SPMs are most likely to appear when participants are told to be enthusiastic ($p < 0.001$). We also found a significant interaction of EMOTION×ROUND NUMBER: SPMs were less likely to appear in later parts of the task for rounds where speakers were instructed to be excited ($p < 0.001$). These findings appear to be paralinguistic in nature: higher emotional arousal leads to an increased probability that an SPM will occur. However, this effect diminishes over the course of the experiment; perhaps participants became fatigued by having to be excited over and over again.

6. DISCUSSION

According to these findings, SPMs are most likely to occur in a final-rise PQ when it is information-seeking, when the CG contains little information related to the question, and when the speaker is highly emotionally engaged and using fluent speech. Since the Common Ground and interrogative force are linguistic in nature, this suggests a linguistically-structured explanation for the appearance of SPMs. As such, we need to augment our model of intonation so that SPMs fall within the domain of predicted contours. This is not to say that SPM contours undermine the validity of a model like MAE_ToBI. We present here two analyses that adhere to the core aspects of the MAE_ToBI model.

One solution might be to augment MAE_ToBI's phonemic inventory. Perhaps MAE_ToBI needs a new (complex) pitch accent (e.g., HL*, H+L*) and/or a right-spreading cover tone (i.e., -H). However, it is not clear whether the presence/absence of these SPMs is a case of a categorical contrast. Empirically, preliminary results from a pilot perception task do not indicate that listeners hear SPMs are meaningfully contrastive. Theoretically, it is not clear that the alignment/scaling of SPMs would line up with any sort of phonological object under an Autosegmental-Metrical model (cf. §4). As such, we believe further empirical/theoretical work is necessary before determining whether ToBI's phonemic inventory ought to be changed.

Another analysis compatible with an MAE_ToBI model would be to treat SPMs as instances of acoustic strengthening, specifically of the cues to the L*. If L* conveys a lack of commitment by the speaker [6], strengthened acoustic cues to the L* may be interpreted as a strengthened cue to the speaker both seeking information and having fewer expectations about likely answers; QUESTION NUMBER and ABOUTNESS are analyzed as tracking exactly this.

Under this analysis, PQs with and without SPMs can have the same phonological representation of L* H-H%. With SPMs absent from the abstract representation, they would not yield fundamental changes in the interpretation of the PQ; this allows them to occur in all sorts of PQs, albeit at different rates. In addition, leaving SPMs out of the phonological representation *per se* is favored because their presence is conditioned by factors such as emotional engagement and fluency of speech. We do, however, note that our data does not show a phonetic gradiency; our model is one that predicts SPM *presence*, not, for example, SPM alignment or scaling.

7. CONCLUSION

MAE PQs serve as a case study in the intonational phonetics-phonology interface, as well as in intonational meaning. This study has uncovered important intonational variation within the category of polar questions — even within the set of PQs with final rises. We have also found that this variation is not without linguistic meaning.

Investigation of intonational phonetic variation is critical for phonological labelling systems: documenting within-category variation helps us to understand the nature of the categories. Additionally, given that SPMs have interpretive consequence, any analysis of intonational meaning must make reference to notions that are more complex or fine-grained than any single feature like 'inquisitive', 'polar question', or 'new/given'. More broadly, we should recognize that phonetic variation in intonation can be linguistically motivated and structured.

Finally, these findings suggest new ways of thinking about debates surrounding H* and L+H*: if not all f0 movements correspond to elements in the phonological underlying representation, perhaps some of what has been labelled as L+H* is in fact an H* plus a leading low SPM. This would mean that what has traditionally been labelled L+H* might reflect one of two different underlying representations (/L+H*/ or /H*/), and it may be that their surface representations could subtly differ too (cf. counterbleeding environments in the interaction between flapping and Canadian raising, or incomplete neutralization in German word-final devoicing).

8. REFERENCES

[1] Ahn, B., Zhou, Z. 2018. Spurious pitch movements in American English polar questions. Presented at Experimental and Theoretical Advances in Prosody 4. http://doi.org/10.17605/OSF.IO/4R79Q.

[2] Beckman, M. E., Hirschberg, J., Shattuck-Hufnagel, S. 2005. The original ToBI system and the evolution of the ToBI framework. In: Jun, S.-A., (ed), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press 9–54.

[3] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer [computer program]. Available at http://www.praat.org/.

[4] McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., Sonderegger, M. January 2017. Montreal forced aligner (version 0.9.0). Computer program.

[5] Pierrehumbert, J. 1980. *The Phonology and Phonetics of English Intonation*. PhD thesis MIT.

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

[7] Pruitt, K., Roelofsen, F. 2013. The interpretation of prosody in disjunctive questions. *Linguistic Inquiry* 44(4), 632–650.

[8] Stalnaker, R. 2002. Common ground. *Linguistics and Philosophy* 25, 701–721.