Conventions in Prosody for Affective Meanings: Non-Canonical Terminal Contours in English Polar Interrogatives

Sunwoo Jeong

Department of Linguistics
Stanford University, Stanford CA, USA
sunwooj@stanford.edu

Abstract

This paper investigates how prosody may work to signal discourse related speech acts on the one hand, and speaker related affective meanings on the other, highlighting the potential connection between the two types of meanings. Focusing on the intonation of English polar interrogatives, the paper reports on a perception experiment that used stimuli manipulated in terminal contours (rising, level, and falling) and embedded in different contexts. The results suggest that different terminal contours in *yes-no* questions reliably signal different types of affective meanings, and also, to a certain degree, different speech acts that are closely related to the respective affective meanings. Importantly, the results also show that the effect of prosody on speech acts holds only for pragmatically felicitous contexts, whereas the effect of prosody on affective meanings persists across different pragmatic contexts. These data support the claim that there is a stable convention for the association between prosody and affective meanings, crucially in conjunction with the sentence type.

Index Terms: question intonation, affective meanings, sentence types, speech acts

1. Introduction

Intonation serves a variety of linguistic and social functions in spoken language [1]. One important function of intonation is to signal the emotional and affective state of the speaker. Another is to signal the communicative intent, i.e. the illocutionary force [2], of the speaker, thereby assisting the speaker to perform different kinds of speech acts such as assertion, promise, invitation, command, etc. Although the two respective functions of intonation have mostly been studied independently of each other, a few studies have also hinted at a close connection between the two. For instance, [3] has argued that flouting the canonical association between a sentence type (e.g. interrogative, declarative) and intonation (e.g. rising, falling) induces the listener to attribute special kinds of enriched pragmatic interpretations to the utterance, which in turn elicits certain kinds of affective meanings. This hypothesized 'inferential process' of deriving affective meanings is particularly interesting, because the process seems to imply a mechanism that taps into a listener's tendency to actively exploit the associations between certain kinds of speech acts (e.g. request, command) and certain kinds of affects (e.g. dissatisfaction), blurring the boundary between prosody's two putatively distinctive functions of signaling illocutionary versus affective meanings.

Although the research mentioned above seems to allude to the possibility of prosody mediating a potential link between certain kinds of speech acts and certain kinds of affective meanings, the existence of an association between the two with respect to prosody has not yet been fully established via experimental data. Furthermore, how exactly prosody may mediate the connection between the two types of meanings has also remained an unresolved question. In light of these issues, this paper aims to establish a more systematic connection between illocutionary and affective meanings as conveyed by prosody, via an experimental design that explicitly addresses this question. The paper also aims to elucidate the exact nature of the relationship between the two types of meanings with respect to prosody, addressing the issue of whether there are intonational conventions for both types of meanings.

To this purpose, this paper focuses on one type of intonation, namely, non-canonical terminal contours in English polar interrogatives (*yes-no* questions). With respect to the illocutionary meanings associated with such contours, prior studies have noted that falling and level contours in English *yes/no* questions are often used to signal non-genuine questions such as assertions [4], [5]. With respect to the affective meanings associated with such contours, [6] have noted that the falling contours in English *yes/no*-questions signal negative affects such as being challenging. Crucially, [6] also showed that the same kind of falling terminal contours do not signal any such negative affects for English wh-questions, whose canonical contour is falling. Such findings suggest that the affective meanings signaled by falling contours are derived not from prosody alone, but from prosody in conjunction with a specific sentence type.

With these prior studies in mind, this study also has a more specific goal of examining whether there is a wider range of illocutionary and affective meanings that can be associated with non-canonical terminal contours in English *yes/no* questions than those that have been previously observed. Such potential findings are expected to create interesting contrasts with another similar intonational phenomena (in the sense that both seem to exploit the mismatch between tune and sentence type), namely, HRTs (High Rising Terminal contours) in declarative sentences, whose potential to signal diverse pragmatic, affective, and social meanings has been much more exhaustively covered in previous literature (e.g. [7]).

2. Method

To address the issues delineated above, a perception experiment was conducted. In the experiment, *yes/no* question stimuli with three different types of prosodic manipulations (three kinds of terminal contours) were used. In addition to the main prosodic manipulation, context manipulation was also included in the experimental design. This was done because pragmatic interpretation (and also possibly affective interpretation) is known to

be highly sensitive to not just prosody, but also interactional contexts (as well as to various other factors). Using different contexts in the experiment was expected to elucidate whether the effect of prosody continues to hold when the illocutionary and affective meanings signaled by prosody are infelicitous in a given context; or, whether it's cancelled out via the strong effect of contextual information. More details regarding the experimental design are provided below.

2.1. Stimuli

To create the stimuli for the experiment, a list of English polar interrogative sentences was recorded. The sentences were produced by a male native English speaker from the Midwest region of the U.S. Based on a separate text-based experiment, the sentence *Do you wanna do the laundry?* was chosen as the target stimulus among the recorded sentences, because it elicited the widest range of illocutionary interpretations in the text-based experiment (invitation, command, genuine information seeking question, etc.). Thus, the sentence chosen above was expected to maximally distinguish any subtle effects that prosody may have on illocutionary interpretations. The highly varied responses that had been elicited for the target sentence in the text-based experiment most likely occurred because in English, the burden of disambiguating between different speech acts that are associated with such a proposition falls heavily on intonation. It is worth noting at this point that other languages may have lexical and grammatical means (e.g. modal particles in German), in addition to intonation, to contextualize such a proposition as a command or an invitation.

After recording the target stimulus, two types of interactional contexts that would precede the main stimulus were recorded as well. These were produced by a female native English speaker from California in the U.S. The two contexts will henceforth be called the invitation context and the command context. The exact sentences used in the two contexts are given in Table 1. The expected effects of such contextual cues on listeners' interpretations of the speech acts - the former eliciting more invitation interpretations and the latter more command interpretations (other things being equal) - were again separately tested in a text-based experiment.

The final contour of the recording of the target sentence *Do you wanna do the laundry?* was then manipulated to produce stimuli with three different kinds of contours: rising, falling, and level. The specific procedure involved in the manipulation process was as follows. First, the pitch minima of the nuclear pitch accent L* (the first syllable in *láundry*) and the endpoint of the utterance were located from the original recording. The pitch tracks between these two points were then eliminated and three new target pitch points (low, mid, high) were set at the endpoint of the utterance. After this, the pitch point at the nuclear pitch accent and the three new endpoints were interpolated in a linear fashion, in order to produce stimuli with three types of final contours: falling, level, and rising, respectively. Following the conventions of ToBI [8], these three types of manipulated contours correspond roughly to !H* L-L%, !H* H-L%, and L* H-H%, respectively. In order to maximally control for the comparability of the three stimuli, no other parts of the utterance before the nuclear pitch accent was adjusted, bearing in mind that the manipulations in just the final contour could still shift the pitch range of the entire sentence, thereby potentially changing the identity of the nuclear pitch accent and affecting the pitch configurations of the sentence as a whole.

The magnitudes of the rise and fall in the rising and falling contours were then decided in the following manner. In order to determine respective target endpoints for each, the physical pitch value (Hz) of the nuclear pitch accent (103 Hz) was multiplied in either increasing (2.2 times the value of the nuclear pitch accent) or decreasing (0.6 times the value of the nuclear pitch accent) directions. The nuclear pitch accent value and the calculated endpoint values for the falling and rising contours were then converted into semitones to check if the differences between them were significantly higher than 1.5 st, which has been noted to produce reliable perceptual distinctions in pitch [9]. As for the level contour, the target endpoint of the utterance was the same value as the nuclear pitch accent.

All manipulations were done in Praat [10] using the built-in PSOLA pitch manipulation program and relevant scripts. The manipulated stimuli thus created were checked by 5 native English speakers to ensure the naturalness of the manipulation.

The three different types of stimuli were then crossed with three different contextual manipulations (no prior context, invitation context, command context) to produce 9 different conditions.

Table 1: *Different conditions used in the experiment*

conditions	stimuli
1-3 *no context*	*M: Do you wanna do the laundry?*
4-6 *invitation context*	*F: I just bought a new batch of fabric softener and it smells so good. I can't wait to try it out.* *M: Do you wanna do the laundry?*
7-9 *command context*	*F: I'm going outside to hang out with my friends.* *M: Do you wanna do the laundry?*

2.2. Experiment

The experiment was structured as a between-subjects design. Each subject was exposed to only one condition, with a single kind of intonation pattern embedded in a unique context. 437 native English speakers from the U.S. were recruited (45 to 50 subjects for each of the 9 conditions) from Amazon Mechanical Turk. In the first section, the subjects listened to the audio clip and were asked to rate the likehoods (from 0 to 100) of different kinds of pragmatic interpretations (invitation, command, and genuine question) that can be attributed to the utterance produced by the male speaker. In the second section, the subjects listened to the same audio clip once more and were asked to answer diverse questions that aimed to get at their affective interpretations of the male speaker's utterance, such as *How annoyed does the speaker sound?* (henceforth called the annoyance rating), *How polite does the speaker sound?* (henceforth called the politeness rating), *How authoritative does the speaker sound?* (henceforth called the authoritativeness rating), and *What kind of attitude (from very negative to very positive) do you think the speaker has towards the listener?* (henceforth called the stance rating).

3. Hypotheses

Before moving on to the results, a brief note about the paper's initial hypotheses seems in order. It was first hypothesized that there would be a close connection between certain kinds of illocutionary interpretations and certain kinds of affective interpretations (i.e. that certain kinds of speech acts would, to some

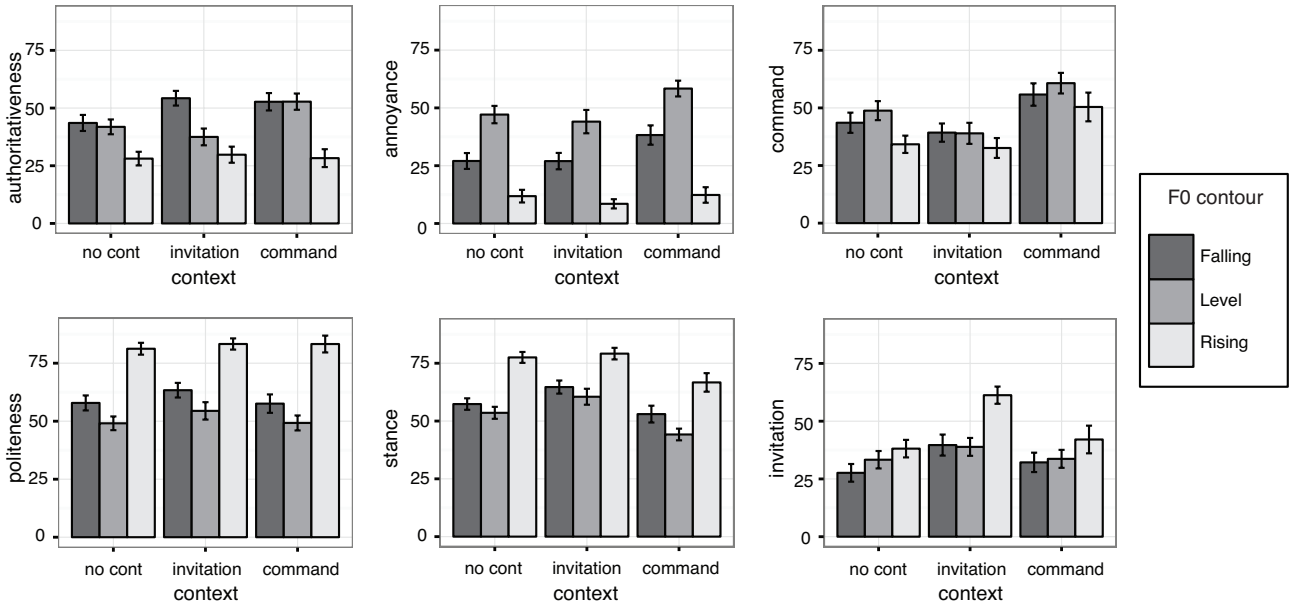

Figure 1: *Means and standard errors of responses in the experiment*

extent, belong together with certain kinds of affective meanings). More specifically, it was hypothesized that prosody that effectively signals command in terms of speech acts would also effectively signal authoritativeness, annoyance, etc. Similarly, it was also hypothesized that prosody that effectively signals invitation will also signal politeness and positive attitude towards the listener. No specific hypothesis was set in advance as to whether the effect that prosody has on one type of meanings can be more fundamental than the effect it has on other types of meanings. To put it differently, this study's initial question of whether prosodic convention exists only for one type of meanings (whereas the other type of meanings is simply derived from its association with the former) was left open at the outset. Finally, it was hypothesized that strong contextual cues will cancel out the effect of prosody on both illocutionary and affective meanings, and that the effect of prosody will manifest most clearly in the no context condition.

4. Results

With respect to the gathered data, a series of multiple linear regression analyses were conducted with each of the different kinds of illocutionary and affective interpretations employed as the main dependent variable in the respective models, and with context manipulation, intonation manipulation, and subject gender employed as independent variables. As mentioned below, an interaction between intonation and context manipulation was posited for one of the models with illocutionary interpretations (invitation, command) as the main dependent variables. Figure 1 visually captures the important data patterns.

4.1. Interpretations of affective meanings

There were clear effects of intonation on subjects' interpretations of affective meanings, and these effects were observable across all three contexts. First, level contours elicited significantly higher annoyance ratings than falling contours (β = .30, p < 0.001) and rising contours (β = .58, p < 0.001), and falling contours in turn elicited significantly higher annoyance ratings than rising contours (β = .28, p < 0.001, r^2 = 0.29 for the entire model). Second, falling contours elicited significantly higher authoritativeness ratings than level contours (β = .11, p < 0.05) and rising contours (β = .36, p < 0.001), and level contours in turn elicited significantly higher authoritative ratings than rising contours (β = .25, p < 0.001, r^2 = 0.12 for the entire model).

In contrast, rising contours elicited significantly higher politeness ratings than the falling contours (β = .40, p < 0.001) and the level contours (β = .55, p < 0.001), and falling contours in turn elicited significantly higher politeness ratings than the level contours (β = .16, p < 0.001, r^2 = 0.25 for the entire model). Similarly, rising contours again elicited significantly higher positive stance ratings than the falling contours (β = .33, p < 0.001) and the level contours (β = .43 p < 0.001), and falling contours in turn elicited significantly higher positive stance ratings than the level contours (β = 0.11 p < 0.05, r^2 = 0.23 for the entire model).

To summarize, negative affects (impoliteness, annoyance, negative attitude towards the listener) were consistently signaled by intonation in the following order: level $>$ falling $>$ rising, whereas positive affects (politeness, positive attitude towards the listener) were consistently signaled by intonation in the reverse order: rising $>$ falling $>$ level. Authoritativeness rating behaved slightly differently, and seemed to be signaled by intonation in a more linear order: falling $>$ level $>$ rising.

In addition to intonation, context also had weak effects on a limited number of affective meanings. First, context did not have any significant effects on politeness ratings and authoritativeness ratings. Second, the invitation context elicited significantly higher positive stance interpretations than no context (β = 0.11, p < 0.05) and the command context (β = 0.26, p < 0.001), whereas the command context elicited significantly higher annoyance ratings than the invitation context (β = 0.16, p < 0.001). Crucially however, the effect of context on affective meanings was much weaker compared to the consistent effect of intonation on affective meanings, as has been seen in their β values and as illustrated in Figure 1. Furthermore, there were

independent effects of intonation that persisted even when contexts became significant predictors of certain kinds of affective meanings. Finally, there was no significant interaction between intonation and context (and the interaction was not included in the regression models for affective meanings).

4.2. Interpretations of speech acts

Intonation's effects on subjects' interpretations of illocutionary meanings were much weaker and less graded than its effects on affective meanings. First, level contours (and to a marginal extent, falling contours as well) elicited significantly higher command interpretations than rising contours (β = 0.15, p $<$ 0.01, r^2 = 0.09). Second, rising contours elicited significantly higher invitation interpretations than falling (β = 0.32, p $<$ 0.001) and level contours (β = 0.34, p $<$ 0.001, r^2 = 0.10). Crucially, in certain cases, the effect of intonation on illocutionary meanings seemed shaped by its interaction with the relevant context. For example, the invitation interpretation rating in Figure 1 (bottom third) shows that the invitation interpretation is bumped up for rising intonations combined with invitation contexts. Indeed, the interaction between the two was significant (β = 0.19, p $<$ 0.05), and was thus included in the model. To recapitulate, the command interpretation seemed to be signaled by intonation in the following order: fall, level $>$ rise, and the invitation interpretation seemed to be signaled by intonation in the following order: rise $>$ fall, level, crucially via an interaction with the requisite context.

The effect of contexts on subjects' interpretations of illocutionary meanings, on the other hand, was much stronger than the effect of intonation. Context had expected effects whereby the command context, for example, elicited significantly higher command interpretations than the invitation context (β = 0.26, p $<$ 0.001).

5. Discussion

Several points of interest emerge from the data pattern delineated above. First, the results indicate that there is at least a three way distinction in intonation when it conventionally signals affective meanings. Also, the relationship between prosody and affective meaning seems to be non-linear in the sense that level contour, not falling, elicited the highest degree of negative affects (e.g. the highest annoyance rating), followed by the falling, then rising, contour. This significant, non-linear difference between the three types of contours in signaling affective meaning suggests that at least in terms of perception, explaining the effect of prosody on affective meanings via a simple, linear phonetic dimension would not be optimal, since no simple generalization such as 'lower F0 correlates with more annoyance' holds.

Combined with the observation that certain linguistic information such as sentence type (declarative, interrogative, etc.) is needed to predict the effect of prosody on affective meanings [6], such non-linear associations between prosody and affective meanings suggested by this experiment seem to further corroborate the argument in favor of the conventionalized nature of prosody in signaling at least certain types of affective meanings. This does not mean that non-conventional ways (e.g. more universal tendencies predicted by mechanisms such as the frequency code [11]) of signaling affective meanings don't exist. It merely suggests that not all kinds of affective meanings are derived from automatic, universal (i.e. non-conventional) tendencies of interpreting prosodic signals.

The second point of interest that appears to emerge from the data pattern is the observation that there does indeed seem to be a close relationship between certain kinds of speech acts and certain kinds of affective meanings. If affective meanings and illocutionary meanings were completely independent of each other with respect to their associations with prosody, there would not have been such a high degree of correlation between the kinds of prosody that effectively signal certain affective meanings and the kinds of prosody that effectively signal certain illocutionary meanings. As it turned out, the hypothesized correlations between affective meanings and speech acts (e.g. invitation and positive attitude towards the listener would pattern together, command and authoritativeness would pattern together, etc.) were mostly borne out in the actual experiment.

The final point of interest is the finding that the effect of prosody on affective meanings persists across different contexts (contra arguments made in works such as [12]), even when strong contextual cues may cancel any potential illocutionary forces signaled by the same prosody. This is ideally exemplified in Figure 1, where the effect of prosody on annoyance rating (second graph, top) for example, producing a tripartite prosodic distinction, is strikingly consistent across all three contexts (albeit with different baselines), whereas the effect of prosody on command interpretation (third graph, top) is much weaker. Instead, context seems to play a much stronger role than prosody, in determining the command interpretation ratings. In the previous hypothesis section, it was noted that no specific hypothesis had been posited regarding which of the two associations - namely between affective meanings and prosody, and between illocutionary meanings and prosody - is more primitive. The data patterns from the experiment seem to at least indirectly suggest that the convention is stable and much more straightforward for the former than the latter. This might at first glance seem counter-intuitive, in the sense that the inferential process assumed in works such as [3] implies that affective meanings are derived from illocutionary meanings rather than vice versa. Although additional data is needed to more definitively establish the claim, what the paper could argue at this point is that once the association between prosody and affective meanings is established and conventionalized, it can operate on its own and remain even in cases where the pragmatic interpretation that may have initially served as a bridge in the inferential process is cancelled due to strong contextual cues.

6. Conclusions

The perception study conducted in this paper shows that prosody, combined with a particular sentence type, may conventionally signal certain kinds of affective meanings. It also demonstrates that such affective meanings are closely related to certain kinds of speech acts, which in turn can be predicted, to some degree, by prosody. Finally, the paper shows that the effect of prosody on speech acts can be cancelled depending on the context, whereas the effect of prosody on affective meanings operates as a more stable convention and persists across a wide range of contexts.

7. Acknowledgements

I thank Rob Podesva, Chris Potts, Meghan Sumner, and Penny Eckert for their insightful comments. I also thank two anonymous reviewers for their helpful feedbacks. All remaining errors are my own.

8. References

[1] J. Vaissière, "Perception of intonation," in *Handbook of Speech Perception*, D. B. Pisoni and R. E. Remez, Eds. Oxford: Blackwell Publishing, 2005, pp. 236–263.

[2] J. L. Austin, *How to Do Things with Words*. Oxford: Oxford University Press, 1962.

[3] A. Wichmann, "Attitudinal intonation and the inferential process," in *Proceedings of Speech Prosody*, 2002, pp. 11–16.

[4] N. Hedberg and J. M. Sosa, "The prosody of questions in natural discourse," in *Proceedings of Speech Prosody*, 2002, pp. 375–378.

[5] N. Hedberg, J. M. Sosa, and L. Fadden, "Meanings and configurations of questions in English," in *Proceedings of Speech Prosody*, 2004, pp. 309–312.

[6] K. R. Scherer, D. R. Ladd, and K. E. A. Silverman, "Vocal cues to speaker affect: Testing two models," *The Journal of the Acoustical Society of America*, vol. 76, no. 5, pp. 1346–1356, 1984.

[7] J. Hirschberg and G. Ward, "The interpretation of the high-rise question contour in English," *Journal of Pragmatics*, vol. 24, no. 4, pp. 407–412, 1995.

[8] M. E. Beckman and G. Ayers, *Guidelines for ToBI labelling*. The Ohio State University Research Foundation, 1997, vol. 3.

[9] A. C. M. Rietveld and C. Gussenhoven, "On the relation between pitch excursion size and prominence," *Journal of Phonetics*, vol. 13, pp. 299–308, 1985.

[10] P. Boersma and D. Weenink, "Praat: doing phonetics by computer (version 5.3.51)," 2015. [Online]. Available: http://www.praat.org

[11] J. J. Ohala, "Cross-language use of pitch: an ethological view," *Phonetica*, vol. 40, no. 1, pp. 1–18, 1983.

[12] R. T. Cauldwell, "Where did the anger go? the role of context in interpreting emotion in speech," in *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*, 2000, pp. 127–131.