The Consistent Meaning of Intonational Tunes Across Sentence Type in American English

Kate Sandberg[1], Jennifer Cole[2]

[1]Northwestern University
[2]Northwestern University

katesandberg2024@u.northwestern.edu, jennifer.cole1@northwestern.edu

Abstract

Compositional accounts of intonational nuclear tune meaning propose that tunes have an impact on speech act meaning that is independent from the sentence type [1, 2, 3, 4]. We provide experimental support for this account through two intonation perception experiments on American English comparing the speech act inferences listeners draw based on rising (L*H-H%) and falling (H*L-L%) imperatives with those drawn based on rising and falling declaratives [5]. Native speakers of American English heard steep rising, shallow rising, and falling intonation on imperative sentences and performed a 2AFC task asking them to categorize each utterance as a Suggestion or a Command. Our results indicate that rising and falling tunes result in the same types of speech act meaning inferences when paired with imperatives (Commands vs. Suggestions) as they do with declaratives (Assertions vs. Questions) [5]. Additionally, we show that within-category variation for a single tune in terms of the pitch span (e.g. shallow rising vs. steep rising tunes) affects the likelihood of a given speech act response in the same manner for imperatives as for declaratives, although this pattern is limited to the non-default tune + type pairing.

Index Terms: intonational meaning, speech acts, speech perception

1. Introduction

Given the many-to-one mapping between sentence types and speech acts in English [6], listeners must rely on alternative linguistic resources, such as the intonational nuclear tune (the pitch pattern expressed over the final portion of the intonational phrase) that is paired with the sentence type in a given utterance, to narrow down the likely intended speech act. There are two main accounts of the impact tunes have on speech act inferences: compositional accounts, which propose that tunes have an independent impact on speech act inferences [1, 2, 3, 4], and non-compositional accounts, which propose that the effect of tunes is dependent on the sentence type they are paired with [7, 8, 5]. This paper provides support for a compositional account by exploring the way intonation guides speech act inferences across sentence type.

In her investigation of rising declaratives, Jeong [5] found differences in the inferences of speech act meaning for declaratives based on the pitch pattern of their nuclear tune as falling, shallow rising or steep falling. Adopting the ToBI representation of intonation, she claims that the H*H-H% tune (with a higher maximum f_o value for the nuclear pitch accent, along with a shallower rising f_o slope to the end of the intonational phrase) is associated with assertive rising declaratives (ARDs), while the L*H-H% tune (a lower maximum f_o value for the nuclear pitch accent, along with a steeper rising f_o slope) is associated with inquisitive rising declaratives (IRDs). Formalizing the effects of these sentence types within the Table model of discourse context [9, 8], she argues that falling declaratives and ARDs convey assertions because they commit the speaker to a single proposition $\{p\}$, while IRDs convey questions in part because they do not incur any commitments on the part of the speaker.

Rudin [2] proposes a compositional account of Jeong's [5] findings that argues that the intonational tune has an effect on speech act meaning that is independent of the declarative sentence type. Specifically, he argues that the L*H-H% tune is what is primarily responsible for calling off the speaker's commitment to the proposition, and that this effect carries over to alternative sentence types. For example, Rudin observes that, for imperatives, L*H-H% seems to result in ADDRESSEE-ORIENTED speech acts like suggestions, which describe actions that would presumably benefit the addressee, as opposed to SPEAKER-ORIENTED speech acts like commands, which describe actions that would presumably benefit the speaker [10]. This result is due to the fact that the tune calls off the speaker's commitment to a preference for the action stated in the imperative [11, 10], just as it calls off the speaker's commitment to a proposition for a declarative. The present study provides experimental support for Rudin's [2] proposal through speech comprehension experiments comparing Jeong's [5] findings for the speech act inferences listeners draw based on rising and falling declaratives with listener interpretations of falling and rising imperatives. We hypothesize that, just as falling declaratives convey SPEAKER-ORIENTED speech acts of assertions by committing the speaker to a proposition [5], falling imperatives will convey SPEAKER-ORIENTED speech acts of commands by committing the speaker to a preference [2]. In contrast, we hypothesize that, just as rising declaratives convey ADDRESSEE-ORIENTED speech acts of questions by calling off the speaker's commitment to a proposition [5], rising imperatives will convey ADDRESSEE-ORIENTED speech acts of suggestions by calling off the speaker's commitment to a preference [2].

The present study additionally builds on Rudin's [2] proposal by investigating whether shallow versions of L*H-H% (labeled by Jeong [5] as H*H-H%) also result in consistent distinctions in speech act inferences across sentence type. Jeong [5] proposes that the H*H-H% tune, which in her study is characteristic of ARDs, maintains the speaker's commitment to the proposition. However, while participants in Jeong's study more often interpret ARDs as assertions, they also sometimes interpret them as questions. A potential explanation for this response pattern is that the two rises Jeong investigates are not categorically distinct but represent variants of a single rising category. One reason to consider this alternative analysis is that the pitch values Jeong reports for the steep and shallow rises do not support a tonal distinction between a low (L*) and high

(H*) pitch accent, as she does not manipulate the f_o value of the nuclear pitch accent in the resynthesis of her stimuli, despite mentioning that the slope of the rise which is crucial to ARD vs. IRD interpretation is largely determined by the relative position of the nuclear pitch accent, and that H*H-H% should involve a higher position of the nuclear pitch accent *in addition to* a shallower rising slope. Therefore, we suggest that the rises Jeong [5] investigates are both versions of L*H-H%, since the speech act inferences listeners draw based on their pairing with declaratives seem to mirror those proposed by Rudin [2] for the L*H-H% tune on imperatives. In light of these findings from prior research, we hypothesize that there exists a continuum of rises within the L*H-H% category, as well as a gradient that maps this variation in rise slope onto variation in the speaker's commitment. In Exp. 1, we test this hypothesis by investigating whether steep rises result in a greater proportion of ADDRESSEE-ORIENTED speech act responses than shallow rises, due to the fact that they commit the speaker to a preference to a greater degree than shallow rises. In Exp. 2, we investigate whether the same type of mapping between variation in pitch and variation in commitment exists for the falling tune, resulting in a greater likelihood of a SPEAKER-ORIENTED speech act response for a steep fall as opposed to a shallow one.

2. Method

The stimuli for these experiments were 18 imperative sentences naturally produced with shallow and steep versions of L*H-H% and a single version of H*L-L% by the first author of this study, who is a female native speaker of American English. As shown in Fig. 1, the naturally produced rises differed based on the magnitude of the pitch movement from the nuclear pitch accent to the end of the intonational phrase, with steep rises exhibiting a greater pitch span than shallow rises. In Exp. 1 and Exp. 2, participants heard resynthesized versions of each of these tunes. In Exp. 1, following Jeong [5], the recordings were resynthesized so that the f_o value of the nuclear pitch accent of those contours was maintained, but the final f_o of the endpoints was -10 st for the fall, +6 st for the shallow rise, and +10 st for the steep rise. While the f_o value of the nuclear pitch accent for the resynthesized rises was not closely controlled, the natural productions did not differ greatly in their f_o minima, as shown in Fig. 1. In Exp. 2, the f_o value of the nuclear pitch accent was more closely controlled, as the stimuli were resynthesized using the single version of the naturally produced fall, with the final f_o for the falling stimuli being -10 st for the steep fall and -6 st for the shallow fall.

In Exp. 1, participants heard a single version of the fall (the steep fall) compared to both versions of the rise, while in Exp. 2, participants heard a single version of the rise (the steep rise) compared to both versions of the fall. The stimuli were presented auditorily following one of two written contexts, so as to probe the possible range of interpretations and improve participants' abilities to make an informed decision about the likely intended speech act. The contexts favored either a SPEAKER-ORIENTED or ADDRESSEE-ORIENTED response. Trials were blocked by context type. In ADDRESSEE-ORIENTED contexts, the addressee was described as someone of equal authority to the speaker, while in SPEAKER-ORIENTED contexts the speaker was described as having higher authority than the addressee. Example contexts are illustrated in Fig. 2. Participants were assigned to one of three lists, where each imperative sentence was paired with a different pitch contour across the lists, and with 6 sentences per contour in each list. Each imperative sentence
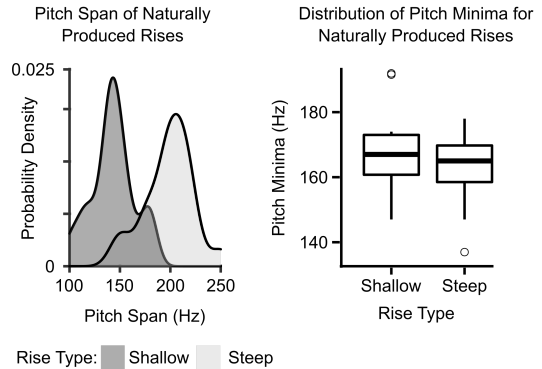

Figure 1: *The left plot shows the probability density of a given pitch span for both shallow and steep rises, with pitch span (in Hz) shown on the x-axis. The plot on the left illustrates the distribution of the nuclear pitch accent minima for both versions of the rise. Pitch (in Hz) is shown on the y-axis.*

was presented twice, once in each written context, (SPEAKER- and ADDRESSEE-ORIENTED), with the same tune in both contexts.

(1) (*ADDRESSEE-ORIENTED Context*)
George is making bread. The last time he tried making this loaf, it came out soggy and underbaked. This time, he wants it to have a better texture. When George asks his friend what they think he should do, they say to him:
Bake the bread for half an hour. *(audio only)*

(2) (*SPEAKER-ORIENTED Context*)
Scott's mom is making bread but has to run out to go pick up Scott's sister from school. She wants Scott to finish making the bread for her. On her way out the door, she says to Scott:
Bake the bread for half an hour. *(audio only)*

Figure 2: *Examples of a* ADDRESSEE-ORIENTED *(top) and* SPEAKER-ORIENTED *(bottom) preceding context.*

Participants were adult (ages 18-65), native speakers of American English who self-reported as having normal hearing, corrected-to-normal vision, and no speech or reading impairments. Additionally, to ensure that participants were familiar with American English, they all reported American English as their primary/dominant language and had learned English and lived in the US prior to college age. The participants for Exp. 1 were 45 individuals (20 male) who were recruited from Prolific and fit these criteria. Participants for Exp. 2 were 39 individuals (9 male) who were recruited from the Northwestern University Linguistics Subject Pool. A total of nine participants were excluded because they did not fit the exclusion criteria described above, resulting in a total of 30 participants (7 male).

Subjects participated in the experiment remotely using their own computers, and they were instructed to use headphones to listen to the stimuli. On each trial, participants were told to read the preceding context and then click on an icon of a speaker to hear an imperative utterance related to the context, realized with one of the three tested tunes, and presented without accompanying text. After listening to the audio clip, participants responded to four questions. In Question 1, participants identified the speech act inference they drew from the stimuli as SPEAKER-ORIENTED or ADDRESSEE-ORIENTED. Because the distinction between these two types of speech acts based on these la-

bels might be unclear to participants, they were relabeled with terms that would presumably be more familiar: namely, COMMAND (for SPEAKER-ORIENTED speech acts) and SUGGESTION (for ADDRESSEE-ORIENTED speech acts). Definitions of COMMANDS and SUGGESTIONS were provided following the category labels. Participants were told that a COMMAND response was felicitous if the speaker was expressing a preference for the addressee to perform a particular action, while a SUGGESTION response was felicitous if the speaker was supplying one possible course of action the addressee could take. Given these descriptions, Question 2 aimed to confirm participants' interpretation of the imperative as a COMMAND or SUGGESTION by asking them to identify which of the two interlocutors they perceived as being more invested in the action taking place—the speaker or the addressee, with the expectation that responses to this question would mirror those for Question 1 (COMMAND responses would correspond with speaker-invested responses and SUGGESTION responses with addressee-invested). Finally, Questions 3 and 4 were rating tasks assessing the social meaning inferences listeners drew based on the stimuli. However, as this paper will focus primarily on the responses to Questions 1 and 2, an analysis of Questions 3 and 4 will not be discussed.

3. Results

3.1. Experiment 1

Participant responses to Questions 1 and 2 were modeled using logistic mixed effects regression. The Command/Suggestion model included fixed effects for tune (Steep rise, Shallow rise, and Steep Fall), context (SPEAKER-ORIENTED, ADDRESSEE-ORIENTED), gender (Male, Female, and Non-Binary), list (List 1, List 2, and List 3), and the interaction of tune and context. Random factors included random intercepts for participant and item. Context was scaled sum-coded (SPEAKER-ORIENTED was mapped to -0.5 and ADDRESSEE-ORIENTED was mapped to 0.5). Gender and list were coded using scaled sum-coding, with Female and List 1 set to the reference levels. Finally, tune was coded using backwards difference coding, with comparisons made between each successive level (e.g. fall vs. shallow rise, fall vs. steep rise). The model was run twice, once with steep fall as the reference level, and once with shallow rise as the reference level, so as to assess comparisons between all levels. As shown in Fig. 3, a SUGGESTION interpretation was more likely for steep rises (β=1.85 z=11.43, p < 0.001) and for shallow rises (β=1.81, z=11.419, p < 0.001) than for falls. Furthermore, a SUGGESTION interpretation was more likely for steep rises than for shallow rises (β=0.35 z=2.2, p < 0.05). There was also a significant effect of context, such that the likelihood of a SUGGESTION response was significantly higher in the ADDRESSEE-ORIENTED condition than in the SPEAKER-ORIENTED condition (β=2.06, z=15.0, p < 0.001). All other predictors included in the model were not significant.

The finding described above is further supported by a logistic mixed effects regression model for Question 2 (speaker vs. addressee as most invested), which tests effects for the same fixed and random factors as the Command/Suggestion model. The model output indicates that a speaker-invested response (an indication of a COMMAND speech act interpretation) was significantly more likely for utterances with a falling tune compared to those with a steep rise (β=-0.87, z=-4.56, p < 0.001), and similarly was more likely for a fall compared to a shallow rise (β=-0.74595, z=-4.21, p < 0.001), but there was no significant difference in responses between utterances

with a steep rise and those with a shallow rise (β=-0.12520, z=-0.64, p = 0.5). Additionally, there was a significant effect of context, such that a speaker-invested response was significantly less likely in an ADDRESSEE-ORIENTED context than in a SPEAKER-ORIENTED one (β=-3.3, z=-19.04, p < 0.001). All other predictors included in the model were not significant, save for gender (males were significantly less likely to provide a speaker-invested response than females (β=-1.44, z=-2.49, p < 0.05)) and the interaction between the steep rise vs. shallow rise comparison and context, which indicates that a steep rise was even less likely to convey a speaker-invested response than a shallow rise in an ADDRESSEE-ORIENTED condition than in a SPEAKER-ORIENTED one (β=-0.43, z=-2.24, p = 0.05).

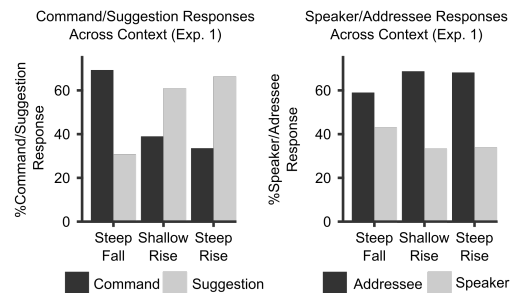Figure 3: *The left plot illustrates the frequency of* COMMAND *and* SUGGESTION *responses for each contour across contexts. The x-axis shows the three intonational tunes, while the y-axis represents the percentage of* COMMAND/SUGGESTION *responses for each contour. The right plot illustrates the frequency of speaker and addressee-invested responses across context for each of the three tunes.*

3.2. Experiment 2

The Command/Suggestion logistic mixed effects regression model for Exp. 2 included fixed effects for tune, context, gender, list, and the interaction between tune and context. Random factors included varying intercepts for both participants and items. As shown in Fig. 4, COMMAND interpretations were more likely for shallow falls (β =2.44, z=10.77, p < 0.001) and steep falls (β=2.72, z=11.79, p < .001) than for rises. However, there was not a significant difference between shallow and steep falls in terms of the likelihood of a COMMAND response (β=0.26, z=1.26, p=0.21). Furthermore, there was a significant effect of context, such that a SUGGESTION response was significantly more likely in the ADDRESSEE-ORIENTED context (β=2.78, z=14.45, p < .001). These findings are supported by the Invested model for Exp. 2, which included the same fixed effects and random effects as the Invested model in Exp. 1. The results of this model indicate that speaker-invested answers were more likely for shallow falls than steep rises (β=0.5, z = 2.37, p < .05). However, the difference between steep rises and steep falls was not statistically significant (β=0.27, z = 1.27, p = 0.2). There was also a significant effect of context, which suggested that speaker-invested answers were significantly less likely in ADDRESSEE-ORIENTED contexts than in SPEAKER-ORIENTED ones (β=-3.74, z=-18.24, p < 0.001), and a significant effect of list, which showed that a speaker-invested answer was significantly less likely for List 3 than for List 1 (β=-0.77, z=-2.03, p < 0.05).
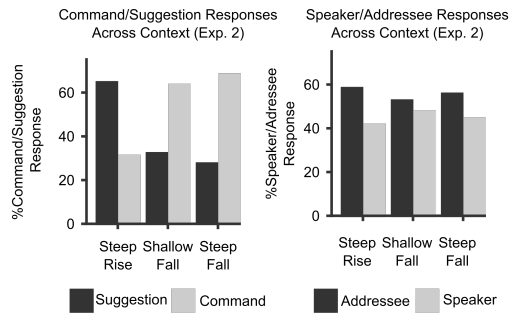
Figure 4: *The left plot illustrates the frequency of* COMMAND *and* SUGGESTION *responses for each contour across contexts. The x-axis shows the three intonational tunes, while the y-axis represents the percentage of responses for each contour. The right plot illustrates the frequency of speaker and addressee-invested responses across context for each of the three tunes.*

4. Discussion and Conclusions

Overall, the data support Rudin's [2] compositional account of tune + type meaning by illustrating that falling and rising tunes convey similar types of speech act meaning when paired with imperatives as they do when paired with declaratives. The results of Exp. 1 show that steep and shallow rises were more likely to result in SUGGESTION interpretations when paired with imperatives than falls, suggesting that the L*H-H% tune calls off the speaker's commitment to the preference stated in the imperative. Furthermore, the distinction between shallow and steep rising tunes in terms of the relative likelihood of SPEAKER-ORIENTED and ADDRESSEE-ORIENTED responses reported in Jeong's [5] study is also observed in the current study for imperatives: Shallow rises were more likely to result in SPEAKER-ORIENTED interpretations of COMMANDS than steep rises, just as shallow rises were more likely to result in SPEAKER-ORIENTED interpretations of assertions for declaratives. This result suggests that differences in pitch span for a single tune constrain speech act inferences in a similar way across sentence type. However, the results of Exp. 2 show that, while shallow rising imperatives were significantly more likely to bias listeners toward SPEAKER-ORIENTED responses than steep rising imperatives, steep falling imperatives were not significantly more likely to bias listeners toward SPEAKER-ORIENTED responses than shallow falling imperatives. Since falls are typically considered the default tune for the imperative sentence type [12], this finding suggests that the potential for variation in pitch span to associate with variation in meaning may be limited to the non-default pairing of tune and sentence type, and compositionally cannot be generalized to pitch span variation for all tunes across all sentence types. However, while the statistical models do suggest this difference in speech act interpretation for shallow vs. steep rising and falling imperatives, the descriptive results in Fig. 1 and Fig. 2 do not clearly reflect this difference. Therefore, the difference revealed in the models should be verified in future replication studies.

Both Jeong [5] and Rudin [2] formalize the effect of intonational tunes on speech act meaning within *dynamic models of meaning*, which are based on the proposal that sentence meaning is derived from information that has already been established as part of the interlocutors' shared background knowledge (the Common Ground), and that speakers add to the Common Ground by raising, resolving, and committing themselves to propositions [13, 9]. The fact that pitch span variation for the rising tune results in differences in the likelihood of a SPEAKER-ORIENTED/ADDRESSEE-ORIENTED response suggests that elements of these models should be re-structured to allow for internal gradience. For example, commitment may not be binary; rather, it might represent a continuum between full commitment and no commitment. This possibility receives further support in evidence from non-lexical intensification. For example, Beltrama [14] observes that the fact that intensifiers can target pragmatic attitudes suggests that these attitudes are scalar rather than discrete. Specifically, he writes that "both lexical and pragmatic orderings lend themselves to being "measured" and manipulated by intensifiers, suggesting that the two versions of scalar modification should be treated as ultimately related, though empirically distinct" [14, pp. 20]. The nature of these continuous pragmatic scales is still debated, although current proposals include the idea that intensifiers track the intensity of a preference for a proposition [15] or that they convey degrees of commitment [16]. The current study provides support for the idea that commitment might represent a pragmatic scale, and that linguistic structures like intonational tunes might operate over this scale. For example, variation in pitch span may commit the speaker/addressee to a proposition to varying degrees. Future studies should address exactly how within-category variation for intonational tunes might map onto continuous dimensions of pragmatic meaning.

We have suggested a re-analysis of Jeong's [5] rising tune stimuli as representing within-category variation rather than separate intonational tunes, and, if correct, this suggests different implications of her results for theories of intonational meaning. Viewed in this manner, her results show a pattern of co-variation between rise slope and the interpretation of speech act, which our data shows to extend across sentence type. Further research is called for to re-consider whether L*H-H% in American English represents a distinct intonational category with non-meaningful within-category variation, or whether it constitutes a continuum from level intonation to a steep rise, with variations in pitch conveying gradient meaning distinctions on some pragmatic scale. Neither the current study nor Jeong's [5] study test a true, multi-step continuum for pitch span, so future studies should look for graded response probabilities along the entire continuum to provide further support for this idea. If, as Jeong [5] suggests, the two rises we investigate here do represent separate intonational categories, then the current results suggest that categories of intonation can be distinguished based on pitch span even when the f_o value of the nuclear pitch accent remains the same.

5. Acknowledgements

6. References

[1] H. Truckenbrodt, "On the prosody of german wh-questions," in *Prosody and meaning*, G. Elordieta and P. Prieto, Eds. De Gruyter Mouton, 2012, pp. 73–118.

[2] D. Rudin, "Rising imperatives," in *Proceedings of the 28th Conference of Semantics and Linguistic Theory*, vol. 28, 2018, pp. 100–119.

[3] C. Gunlogson, "A question of commitment," *Belgian Journal of Linguistics*, vol. 22, no. 1, pp. 101–136, 2008.

[4] M. Westera, "Exhaustivity and intonation: a unified theory," Ph.D. dissertation, University of Amsterdam, 2017.

[5] S. Jeong, "Melodies in context: The semantics and pragmatics of english rising declaratives," Ph.D. dissertation, Stanford University, 2018.

[6] C. Bartels, *The intonation of English statements and questions: A compositional interpretation*. Routledge, 2014.

[7] D. F. Farkas and F. Roelofsen, "Division of labor in the interpretation of declaratives and interrogatives," *Journal of Semantics*, vol. 34, no. 2, pp. 237–289, 2017.

[8] S. A. Malamud and T. Stephenson, "Three ways to avoid commitments: Declarative force modifiers in the conversational scoreboard," *Journal of Semantics*, vol. 32, no. 2, pp. 275–311, 2015.

[9] D. F. Farkas and K. B. Bruce, "On reacting to assertions and polar questions," *Journal of semantics*, vol. 27, no. 1, pp. 81–118, 2010.

[10] C. Condoravdi and S. Lauer, "Imperatives: Meaning and illocutionary force," *Empirical issues in syntax and semantics*, vol. 9, pp. 37–58, 2012.

[11] D. Kimps and K. Davidse, "Illocutionary force and conduciveness in imperative constant polarity tag questions: A typology," *Text & Talk*, vol. 28, no. 6, pp. 699–722, 2008.

[12] D. Bolinger, *Intonation and its uses: Melody in grammar and discourse*. Stanford university press, 1989.

[13] R. C. Stalnaker, "Assertion," in *Syntax and Semantics: Pragmatics*, C. P, Ed. New York: Academic Press, 1978, vol. 9, pp. 315–332.

[14] A. Beltrama, "Bridging the gap: Intensifiers between semantic and social meaning," Ph.D. dissertation, The University of Chicago, 2016.

[15] A. Giannakidou and S. Yoon, "The subjective mode of comparison: Metalinguistic comparatives in greek and korean," *Natural Language & Linguistic Theory*, vol. 29, no. 3, p. 621, 2011.

[16] R. Bochnak and E. Csipak, "A new metalinguistic degree morpheme," in *Semantics and Linguistic Theory*, vol. 24, 2014, pp. 432–452.