Initial Pitch Cues in English Sentence Types

Matthew Patience, Olivia Marasco, Laura Colantoni, Gabrielle Klassen, Malina Radu, Olga Tararova

University of Toronto, Canada
matt.patience@utoronto.ca, olivia.marasco@utoronto.ca, laura.colantoni@utoronto.ca,
gabrielle.klassen@mail.utoronto.ca, malina.radu@mail.utoronto.ca, olga.tararova@mail.utoronto.ca

Abstract

Previous research has revealed that English speakers can differentiate between questions and statements after hearing an utterance's first pitch accent [1]. This suggests that initial F0 cues distinguishing questions from statements are present in the input. We examined this proposal by analyzing the first pitch accent in statements, absolute yes/no questions, and declarative questions. The production of these three sentence types was elicited from 10 Canadian English speakers who performed a sentence-repetition task. Results revealed that statements were produced with an initial H+, whereas both question types were almost exclusively produced with an initial L*+H. Statements were also produced with an earlier peak alignment, and a smaller F0 change. No differences were observed between absolute and declarative questions. The results are consistent with the stimuli analyzed in [1], and provide further evidence that initial pitch cues mark sentence type in Canadian English.

Index Terms: English prosody, English intonation, initial pitch cues, statements, questions, Canadian English

1. Introduction

In the present paper, we investigate the F0 realization of the first pitch accent to determine whether differences are observed between statements (Ss) and yes/no (Y/N) questions in English. While over 70% of languages differentiate between Y/N questions and Ss with a final pitch rise or fall [2], many typologically distinct languages also signal this contrast in the prenuclear region, including Spanish [3, 4], Northern German [5], Dutch [6], Swedish [7], Mandarin [8], Portuguese [9, 10], Hungarian [11], Cherokee [12], and Western Arabic [13], among others. The use of initial cues to mark sentence type is consistent with psycholinguistic research, which has demonstrated that listeners begin to plan their response before the interlocutor has finished speaking [14, 15]. The initial cues thus allow speakers to distinguish questions from Ss without having to hear the nuclear contour at the end of an utterance. Indeed, some scholars have even proposed that final question intonation is a redundant cue [16]. Moreover, the fact that Y/N questions are marked by inversion/particles in initial position further suggest that early cues to sentence type are important for processing.

While it is clear from previous work that sentence type is marked by initial pitch cues in many languages, to our knowledge, a detailed analysis comparing the production of initial cues of Ss and Y/N questions in North American English has not been conducted. Nevertheless, there is reason to believe that English also marks sentence type using initial cues. In a perception experiment investigating Canadian English, [1] demonstrated that adult speakers could distinguish Ss from declarative questions (DQs) after hearing only the first word. These results reveal that English speakers are sensitive to differences in the first pitch accent, which suggests that initial pitch cues marking sentence type are present in the input. Upon analysis of their stimuli, [1] found that the initial pitch accent was L*+H in DQs and L+H* in Ss. Given the occurrence of initial pitch cues cross-linguistically, the role of early cues in turn-taking, and the findings reported in [1], we should expect differences in the realization of English Ss and questions.

1.1. English statements and questions

English distinguishes between absolute Y/N questions (AQs) and Ss with either inversion or do-insertion. In contrast, DQs and Ss have identical surface structures. The sentences in (1) illustrate the three sentence types. Both types of Y/N questions are also marked with a final pitch rise [17, 18].

(1) S: Peter bought a piano.
  AQ: Did Peter buy a piano?
  DQ: Peter bought a piano?

Regarding the prenuclear region, as discussed previously, [1] found that their statement stimuli were produced with an L+H* initial pitch accent, compared to questions which were produced with an L*+H accent. Y/N questions have also been transcribed in previous work with primarily H* [19, 20, 21, 22] initial pitch accents. However, transcriptions of L*, L*+H, and L+H* have been reported as well [19, 20, 22]. Note that these studies did not compare statements and questions; thus, it remains to be determined if the initial pitch accents are different across sentence type.

The fact that DQs are syntactically identical to Ss in English provides an interesting domain in which to investigate the role of initial pitch markers. Both context and the final rise are generally considered to be the cues that speakers use to distinguish a DQ from an S [17]. However, the findings reported in [1] indicate that initial cues, if present, may be equally as important. Therefore, the primary objective of the current study is to determine whether differences in initial pitch are present between Y/N questions and Ss. An additional objective is to examine whether differences are observed in the initial pitch cues of the two Y/N question types. Given that AQs are marked syntactically, it is possible that initial pitch cues in DQs are more salient than in AQs.
2. Current Study

The production of Ss, DQs, and AQs was analyzed to determine whether Canadian English speakers mark sentence type by producing differences in the initial pitch accent. We present here the two questions that guided our study, along with their respective predictions.

RQ1: Are differences observed in the realization of the initial pitch accent in Ss compared to questions?

We expected to find phonetic and phonological differences in the initial pitch accent in Ss compared to questions. This prediction was based on previous research reporting that English speakers are sensitive to initial pitch cues and can use them to distinguish between Ss and questions [1]. Moreover, differences in prenuclear pitch are common cross-linguistically, including in languages that are typologically similar to English such as German [5] and Dutch [6].

RQ2: Are there differences in the first pitch accent of DQs compared to AQs?

Differences were expected in the phonetic realization of the two question types. This prediction was based on the fact that AQs, but not DQs, are marked by inversion or do-support. We therefore expected initial pitch cues to be more salient in DQs compared to AQs.

3. Methodology

3.1. Participants

Ten speakers (five male, five female) participated in the study. All participants were native speakers of Canadian English and were living in Canada at the time of testing. The average age of the participants was 22.5 years (SD = 3.51).

3.2. Tasks and stimuli

The present experiment's data was obtained from a larger study investigating the acquisition of English intonation by native speakers of different L1s. Participants completed a series of production and perception tasks across two testing sessions. Two of these tasks elicited the production of English sentence types: a repetition task in which participants listened to the target sentence and were subsequently asked to repeat it, and a contextualized task in which a context elicited the production of the target sentence type. We report here the findings from the repetition task only.

Stimuli consisted of short SVO sentences. There were eight stimuli per sentence type (N = 24), as in (1). All AQs began with Did. Thirty distractors were also included. The stimuli were produced by a female native speaker of Canadian English. She produced three repetitions of each sentence. These sentences were then rated for naturalness by three other native speakers of Canadian English, and the best productions were included in the experiment.

3.3. Procedure

The experiment was administered using PowerPoint. Participants listened to the target utterance and were instructed to repeat the utterance as closely as possible. Data were recorded using a Marantz PMD561 recorder and a unidirectional condenser microphone.

3.4. Data Analysis

The recordings were extracted and analyzed in Praat [23]. Phonological and acoustic analyses of the initial pitch accent were performed. For the phonological analysis, we labeled each pitch accent according to ToBI guidelines [24, 25]. We expected differences in the type of pitch accent, based on findings in [1], who reported that their S stimuli (produced by a Canadian English speaker) had an L+H* contour, whereas DQs had an L^3+H contour. AQs were not examined in their study.

If questions and Ss are indeed differentiated by L^*+H versus L+H* accents, we should expect a contrast in peak alignment, with questions displaying a delayed peak. Our acoustic analysis therefore examined the peak alignment with respect to the tonic vowel. Following [26], we marked two points: the end of the tonic vowel and the maximum pitch peak. Peak alignment was calculated as the difference in duration between these two points. A positive value reflects a peak that occurred after the end of the tonic vowel, whereas a negative value reflects a peak that occurred before the end of the tonic vowel.

We also examined the F0 change in the initial pitch accent. Previous work on Spanish and Dutch revealed a higher F0 in questions compared to Ss [3, 6]. We therefore included it in our analysis expecting potential differences between questions and Ss, or between AQs and DQs. To calculate the change in F0, the minimum and maximum F0 values associated with the stressed syllable were determined, and converted to ERB. A ratio of the maximum over the minimum was subsequently calculated.

Figure 1 illustrates a representative production of the three sentence types and how the data were coded. The beginning and end of the tonic vowel were marked (Tier 2), as were the beginning and end of the pitch movement associated with the stressed syllable (Tier 3).

Figure 1: Pitch contours of a representative S (top), AQ (mid), and DQ (bottom). Tier 2 indicates the beginning 'V' and end 'VV' of the tonic vowel. Tier 3 displays the ToBI label associated with the initial pitch accent.
Statistics were run in SPSS v. 23 using a significance level of .05. All generalized mixed effects models included ‘sentence type’ as a predictor, which was analyzed using a treatment coding with 'statement' as the reference variable. In all models, ‘participant’ and ‘item’ were included as random intercepts. Post-hoc analyses using Bonferroni corrections were run to compare differences between AQs and DQs.

4. Results

The results of the phonological analysis are presented first, followed by those of peak alignment. We finish by showing the results of the change in pitch.

4.1. Pitch Accent Type

Figure 2 displays the type of contour produced for the initial pitch accent across sentence types. Three different contours were observed: L*+H, L+H*, and H*. The two types of questions were produced almost exclusively with an L*+H contour. Ss, in contrast, were produced primarily with an H* contour, although some L+H* accents were also observed. Given the sporadic use of the L+H* accents, only the H* and L*+H pitch accent contours were included in the statistical analysis. A mixed-effects binomial logistic regression revealed that speakers were significantly more likely to use the L*+H contour in AQs ($\beta = -4.83; SE = 0.97; t = -4.96; p < .001$) and DQs ($\beta = -5.10; SE = 1.01; t = -5.06; p < .001$) compared to Ss. A post-hoc analysis comparing AQs to DQs revealed no differences in the use of the two contours ($\beta = 0.01, SE = 0.04, t = 0.25, p = .803$).

In sum, the phonological analysis revealed that both question types differed from Ss. However, no difference was observed between AQs and DQs.

4.2. Peak Alignment

The analysis in 4.1 revealed that questions were produced with an L*+H initial pitch accent, compared to Ss which had mainly an H* contour. We should thus expect to find differences in peak alignment. Recall from Section 3.4 that peak alignment was calculated as the difference in duration between the F0 peak associated with the pitch accent and the end of the tonic vowel. Therefore, a delayed peak alignment would have a positive value.

Figure 3 reveals that F0 peaks in both AQs and DQs occurred more than 100 ms after the end of the tonic vowel (AQ: $M = 119, SD = 61.3$; DQ: $M = 115, SD = 56.2$). In contrast, F0 peaks in Ss occurred around the end of the tonic vowel ($M = 22.1, SD = 49.6$). A generalized linear mixed-effects model indicated that these differences were highly significant (AQs: $\beta = 0.10; SE = 0.02; t = 4.85; p < .001$; DQs: $\beta = 0.09; SE = 0.02; t = 4.63; p < .001$). However, a post-hoc analysis revealed no differences between AQs and DQs ($\beta = 0.00, SE = 0.02, t = 0.22, p = .829$).

4.3. Percent change in F0

The second acoustic parameter we examined was the percent change in pitch (in ERB) that occurred between the minimum and maximum points of the F0 contour. Figure 4 reveals that questions had a larger change in pitch compared to Ss (AQ: $M = 0.35, SD = 0.29$; DQ: $M = 0.30, SD = 0.22$; S: $M = 0.12; SD = 0.12$). A generalized linear mixed-effects model indicated that these differences were highly significant (AQs: $\beta = 0.22; SE = 0.03; t = 6.88; p < .001$; DQs: $\beta = 0.17; SE = 0.03; t = 5.34; p < .001$). However, a post-hoc analysis revealed no differences between AQs and DQs ($\beta = 0.05, SE = 0.03, t = 1.54, p = .125$). Therefore, similar to the peak alignment results, differences were only observed between questions and Ss, but not between the two question types.

5. Discussion

5.1. Summary of results

In the current paper, we performed a phonological and acoustic analysis of the initial pitch accent in three sentence types: Ss, AQs, and DQs. Our first research question (RQ1) sought to determine whether differences were present in questions compared to Ss. Differences were found at both the phonological and phonetic levels of analysis. Our results revealed that the two question types were produced almost exclusively with an L*+H pitch accent, whereas Ss were produced with an H* pitch accent. The different pitch contours
corresponded to differences in peak alignment. Specifically, a more delayed peak was observed in both question types compared to Ss. We also found that questions were produced with a greater F0 change compared to Ss. We therefore conclude that questions and Ss were realized with differences in the initial pitch accent.

Our second research question (RQ2) sought to determine whether differences were observed between AQs and DQs. As previously mentioned, AQs are marked with salient syntactic cues, thus the potential need for initial pitch cues is reduced. Nevertheless, the results indicated that AQs and DQs were produced with very similar F0 profiles, at both the phonetic and phonological levels.

5.2. Comparison to previous studies

Our results demonstrated clear differences between questions and Ss, and are therefore consistent with the findings in [1]. In their perception study, [1] found that speakers can distinguish between Ss and DQs using initial pitch cues, which suggests that such cues are present in production. Moreover, the pitch contour distinction we observed was similar to the question versus statement distinction found in [1]'s stimuli. In both theirs and the present study, questions were produced with an L+H contour. The comparable results strongly suggest that English Y/N questions are partially characterized by a late pitch rise associated with the initial pitch accent. Note that [1] only compared Ss and DQs. We found that AQs showed the same F0 patterns as DQs.

One difference between our study and [1] was the pitch accent observed in Ss. We observed primarily H*, whereas in their study the stimuli were produced with L+H*. There are several possible explanations for these differences. First, both L+H* and H* are frequent pre-nuclear pitch accents for broad focus declaratives in North American English [27]; thus, neither should be unexpected, and any differences may simply have been due to standard variability in English intonation. Second, the different pitch accents may be partly due to different stimuli. For example, in contrast to our study, some stimuli in [1] had complex subjects (e.g., Donuts with sprinkles are tasty.), and auxiliary verb + verb sequences (e.g., Brendan is leaving town again.). Third, the divergent patterns may be due to differences in interpretation. The contrast between H* and L+H* in ToBI is minimal, and often open to interpretation [21], especially when occurring in sentence initial position (as was the case for the stimuli in [1], and our S and DQ stimuli).

It should also be noted that in previous corpus studies examining English Y/N questions, the initial pitch accents were transcribed primarily as H*, although L+H*, L+H*, and L* were also present [20]. The analysis in these studies focused on the nuclear contours, and thus the variable initial pitch accents were not discussed in detail. Moreover, these studies did not compare statements and questions. Consequently, it is not possible to know whether differences were present between sentence types. Nevertheless, the results of [20], combined with the results of the present study, suggest that a variety of accents are possible. Future work should examine whether differences in initial pitch accents are reliably present in more spontaneous speech.

5.3. Potential task effects

The consistent differences between questions and Ss could be due partly to a task effect, given the nature of the task (i.e., repetition). We therefore analyzed the stimuli to determine how closely the productions mirrored the task's stimuli.

Regarding pitch accent type, all question stimuli were produced with L+H, whereas Ss were produced with H* in all cases but one (produced as L+H*). The results of the present study's participants were therefore very similar to the stimuli. Regarding peak alignment, the mean for the statement stimuli was 26.9 ms (SD = 33.0), whereas the mean for the question stimuli was 113 ms (SD = 50.1). The present study's participants had peak alignment values of 22.1 ms (SD = 49.6) for statements and 117 ms (SD = 58.7) for questions. Regarding F0 change, the mean value observed in the statement stimuli was 0.17 ERB (SD = 0.092) compared to 0.33 ERB (SD = 0.077) in questions. In contrast, participants produced an average F0 change of 0.12 ERB (0.12) in statements, and 0.32 ERB (SD = 0.26) in questions.

As expected, the comparison reveals that, in general, the productions of the participants reflected the stimuli. Therefore, one of the limitations of the present study is the generalizability of the results. Nevertheless, we expect the results to be representative of English sentence types, for the following reasons. First, similar results were observed in the stimuli produced in [1]. Second, the stimuli were produced naturally by a Canadian English speaker, and identified as naturalounding by three other native speakers of Canadian English. Note that the stimuli were not originally created to analyze differences in initial pitch accents, and thus the four speakers (i.e., the one who produced the stimuli, and the three who selected the most natural productions) were naive to the objective of the study. Third, if the participants had different grammars than the stimuli they were asked to repeat, we might have expected a greater amount of variability, especially in the acoustic analysis.

In sum, the results suggest that differences are present in initial pitch cues in Canadian English. However, future research will have to determine how consistent the contrast between statements and questions is.

6. Conclusions

In the present analysis of initial pitch cues, we have demonstrated that speakers of Canadian English produce a different pitch accent in Ss (H*) compared to questions (L+H). Moreover, we have shown that the same pitch accent is found in AQs, which are syntactically marked, and DQs, which are not. We found no phonetic distinction between AQs and DQs. Our results thus suggest that the initial pitch cues found in Y/N questions are characteristic of these sentence types, and are present regardless of other redundant cues. The findings are consistent with previous work reporting initial pitch cues marking sentence type cross-linguistically, as well as psycholinguistic research on turn-taking [14, 15]. Our findings contribute additional evidence that prenuclear information should be considered when examining pitch contour meaning.

7. Acknowledgements

This research was supported by a SSHRC grant (#890-2011-0049) from the Government of Canada and a University of Toronto research completion fund.
8. References


