QUESTION INTONATION AS AFFECTED BY WORD STRESS AND FOCUS IN ENGLISH

Fang Liu [1], *Yi Xu* [2]

[1] Department of Linguistics, The University of Chicago, IL, USA
[2] Department of Phonetics and Linguistics, University College London, UK
liufang@uchicago.edu, yi.xu@ucl.ac.uk

ABSTRACT

The intonational realizations of statements and declarative questions in American English are studied by examining their interaction with focus and word stress. Results of F_0 analyses indicate that 1) statements and declarative questions start to diverge from the stressed syllable of the first content word, 2) the pitch range of post-focus syllables is compressed and lowered in statements, but compressed and raised in questions, and 3) the pitch target of the stressed syllable in a content word is high or falling in statements but rising in questions, depending on the focus condition and the stress pattern of the word.

Keywords: Stress, focus, sentence type, pitch target.

1. INTRODUCTION

Despite much research, the exact manner with which statements and questions are differentiated in English intonation remains unclear. In particular, three critical issues are unresolved: 1) the temporal scope of the acoustic contrast between the two sentence types, 2) the role of focus in their differentiation, and 3) whether there are any differences in terms of the underlying pitch targets of individual syllables.

The temporal scope of the acoustic contrast between statement and question intonations in English has been described differently by different studies. According to [3], the differences are mainly in the pitch movement of the last syllable, with statements showing a mid-falling movement and questions a rising movement. However, O'Shaughnessy [7] reported that the rising F_0 contours of the three critical regions (the first and last accented syllables and the syllable carrying sentence focus) in a question characterize its intonation pattern, and that F_0 in questions is generally higher than that in statements. Nevertheless, in the AM theory of English intonation [5, 6, 9, 10], a declarative intonation is described as having a high pitch accent, a low phrase accent and a low boundary tone (H* L-L% in the ToBI representation), and a yes/no question a low pitch accent, a high phrase accent and a high boundary tone (L* H-H%). Therefore, at least three different temporal scopes have been proposed in the distinction between statement and question intonations in English: the last syllable, the entire sentence, or on and beyond the nuclear accent. In addition, such distinction has been deemed to be realized through either a falling/rising or a high/low contrast.

When investigating focus in different sentence types, Eady & Cooper [2] noted that statements and questions with neutral or final focus differ only in the F_0 topline of the last key word, with peak F_0 of questions being significantly higher than that of statements. Nevertheless, there is no significant difference in peak F_0 on the focused word for statements and questions with initial focus. Rather, the two sentence types differ in the F_0 topline following the initial focused word, with that of statements dropping to a low value and that of questions remaining rather high. Pell [8] confirmed the above finding, and further noticed that peak F_0 of the on-focus (initial or final) vowels is increased in statements but not in questions. Xu & Xu [15:177] provided a more detailed account of the focus effect in statements: a narrow focus would "increase the size of the F_0 peak in the stressed syllable under focus, lower all the postfocus F_0, including that of the poststressed syllables in the focused word, and leave prefocus F_0 largely intact". This suggests that a more detailed account of the focus effect in questions is also needed, because what is affected seems to be in terms of the syllable – a smaller domain than word.

Pitch target, as proposed in [14], refers to the ideal underlying pitch trajectory assigned to a syllable as a production goal, which remains invariant despite extensive surface variations after different preceding tones. The contextual variability is described as due to articulatory inertia, and the manner of the variability in the surface F_0 trajectories is captured by the Target Approximation (TA) model [14], according to which the pitch target is articulatorily implemented by an asymptotic movement toward the production goal throughout the host syllable, and, as a result, is best

achieved near the end of the syllable. Thus, evidence of a pitch target can be found in the F_0 trajectory during and especially toward the end of the syllable.

For English, Eady & Cooper [2] reported that words in the sentence-final position have a falling F_0 contour in statements but a rising contour in questions. On the other hand, F_0 of the initial word in both statements and questions is rising, with the exception of the initial focused word in a statement, which exhibits a rising-falling (or generally falling) contour. Xu & Xu [15] found similar F_0 patterns for English statements, but treated them as evidence for the underlying pitch targets. In particular, a continuous F_0 rise in a non-final stressed syllable (focused or non-focused) or a non-focused word-final stressed syllable was viewed as due to a high pitch target, whereas the final F_0 fall in a focused word-final stressed syllable or a sentence-final stressed syllable (focused or non-focused) was viewed as due to a falling target. But little is yet known about the underlying pitch targets in English questions, as not enough detail was reported by [2], and only statements were investigated in [15].

This study thus aims to address three unresolved problems in English intonation: 1) What is the temporal scope of the acoustic contrast between statements and declarative questions? 2) How does focus affect the realization of this contrast? 3) Are there any pitch target variations involved in this contrast?

2. METHOD

2.1. Subjects

Three female and two male speakers, aged 18 – 30, served as subjects. They were raised in either California or the Midwest and spoke general American English. They had no self-reported speech or hearing disorders.

2.2. Materials

Test materials contain three sets of sentences, within which the final syllable of the last word is either stressed (e.g., *Elaine*) or unstressed (e.g., *Alan*). Each sentence was uttered with two sentence types (statement vs. declarative question) and with two focus conditions (medial vs. final). The sentence type (./?) and focus (in bold face) conditions were elicited by different leading sentences (in parentheses), as shown below. Each sentence was repeated eight times by each subject, resulting in 960 utterances in total.

1. (Not an **internship**./?)
   You want a **job** with Microsoft./?
   (Not an **internship**./?)
   You want a **job** with La Massage./?
   (Not **La Massage**./?)
   You want a job with **Microsoft**./?
   (Not **Microsoft**./?)
   You want a job with **La Massage**./?
2. (It's not **fate**./?) (It's not **you**/**me**./?)
   There is something **unmarriable** about **me/May**./?
3. (It's not **Sears**./?) (It's not **Elaine/Alan**./?)
   You're going to **Bloomingdales** with **Alan/ Elaine**./?

2.3. Procedure

Recordings were done in a sound-treated booth in the Language Labs at the University of Chicago. There were 8 sentence blocks (for 8 repetitions of each sentence) and a different randomization was applied to each block. The utterances were first recorded onto a digital memory card using a solid-state recorder and then transferred to a computer. The digitized sounds were re-sampled at 22.05 kHz for later analyses.

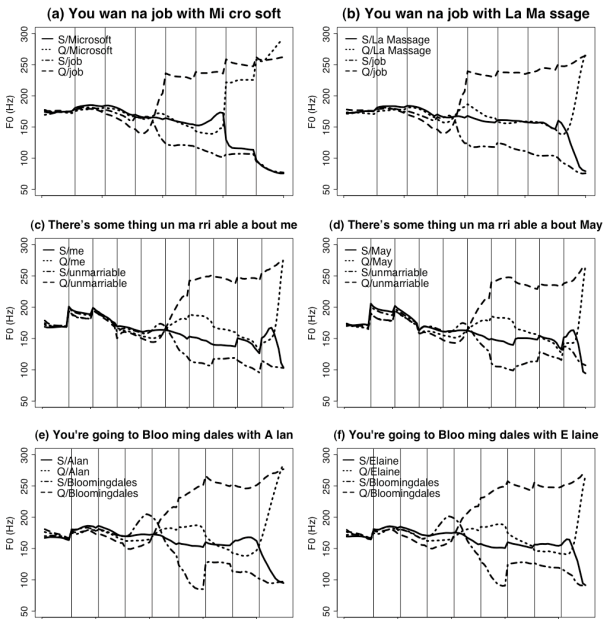
For graphical analysis, time-normalized F_0 contour of each syllable in each sentence was extracted by a custom-written Praat [1] script. Mean F_0 (indicating pitch height, in st) and final velocity — velocity (instantaneous rate of change of F_0, in st/s) of F_0 at 30 ms before syllable offset (as evidence of the underlying pitch target, cf. [4, 13], and will be further discussed later) were calculated for statistical analyses.

3. RESULTS

Fig. 1 displays mean F_0 contours (averaged across 40 repetitions by 5 subjects) of the test materials under different focus conditions and in different sentence types. The following patterns can be clearly seen. 1) The difference in pitch contours between statements and declarative questions in English seems to start from the stressed syllable of the first content word (focused or non-focused). 2) Focus has no effect on the pre-focus region in either statements or questions. In the on-focus region, the pitch range of the stressed syllable is expanded in both statements and questions. The post-focus pitch range is compressed and lowered in statements, but compressed and raised in questions. 3) The pitch target of the stressed syllable (e.g., *Bloo* in (e) and (f)) in a focused non-final-stressed word (e.g. *Bloomingdales*) seems to be high in statements, but rising in questions, and that in a focused final-stressed word (e.g., *May* in (d), *job* in (a) and (b), and *Elaine* in (f)) appears to be falling in statements, but rising in questions. In utterances with final focus (e.g., *You're going to Bloomingdales with **Alan*** in (e)), the pitch target of the stressed syllable (e.g., *Bloo*) in a pre-focus

content word (e.g. *Bloomingdales*) seems to be high in statements but rising in questions. In utterances with medial focus (e.g., *You're going to* ***Bloomingdales*** *with Alan/Elaine* in (e) and (f)), the pitch target of the stressed syllable (e.g., *A/laine*) in a post-focus content word (e.g. *Alan/Elaine*), despite the pitch range compression, is still recognizable: *A* is high in statements but rising in questions; *laine* is falling in statements but rising in questions. Statistical analyses were conducted using R [11] to verify the above observations.

Figure 1: Time-normalized mean F_0 (Hz) contours of test materials. Vertical lines indicate syllable boundaries. In the legend, "S" stands for statement, and "Q" for declarative question. "S/Microsoft" means a statement with focus on "Microsoft", and so on.



3.1. Focus effect

The grand mean F_0 (st) of each syllable (averaged across mean F_0 of 40 repetitions) was extracted to determine pitch range variations in each set of sentences under different focus conditions and in different sentence types. Paired t-tests indicate that a) post-focus syllables have significantly lower mean F_0 in statements than in questions, b) pre-focus syllables (except *some* ($t(9) = 2.87$, $p = 0.0186$) and *able* ($t(9) = -2.31$, $p = 0.0459$) in Fig. (1c) and (1d)) do not differ in mean F_0 in the two sentence types, and c) on-focus syllables may or may not have significantly different mean F_0, depending on their pitch targets related to sentence type (as discussed in 3.2). Therefore, with respect to pitch range, the distinction between statements and declarative questions mainly lies in the pitch height of post-focus syllables, with that of statements being low and that of questions high.

3.2. Pitch targets of stressed syllables

In light of the findings on the temporal interval of segments in [12], final velocity of the pitch contour of the stressed syllable in the focused word was extracted at 30 ms before the conventional syllable offset. The median of the velocity values from the 8 repetitions by each subject was calculated, and Table 1 lists the mean final velocities (averaged across the medians of 5 subjects) of the on-focus stressed syllables grouped by stress pattern and sentential position of the word. Because final velocities of on-focus stressed syllables do not vary with different focus locations ($F(1,3) = 5.14$, $p = 0.1082$), medial and final focus are not treated separately in Table 1.

Repeated measures ANOVAs with sex as between-subject factor and sentence type and stress pattern of on-focus word as within-subject factors indicate that on-focus stressed syllables have significantly different final velocities in statements and questions ($F(1,3) = 129.97$, $p = 0.0014$). Furthermore, although stress pattern of on-focus word has no significant main effect on final velocity, its interaction with sentence type is statistically significant ($F(2,6) = 6.48$, $p = 0.0317$).

Table 1: Mean final velocities (st/s) of on-focus stressed syllables, and the t-tests indicating whether they are significantly different from zero. Here and subsequently, "wfinal_nonsfinal" stands for word-final and non-sentence-final, and "wfinal_sfinal" for word-final and sentence-final.

Intonation	Non-final	W-final Non-s-final	W-final S-final
Question	29.05 $t(29) = 5.77$ $p < 0.0001$	61.95 $t(9) = 9.79$ $p < 0.0001$	41.19 $t(19) = 12.15$ $p < 0.0001$
Statement	-6.56 $t(29) = -1.59$ $p = 0.1225$	-39.64 $t(9) = -3.18$ $p = 0.0113$	-40.06 $t(19) = -5.31$ $p < 0.0001$

To determine the pitch targets of on-focus stressed syllables in different sentence types, t-tests were conducted to see if their final velocities are significantly different from zero. Approximately, a large positive velocity corresponds to a rising pitch target, a large negative velocity a falling pitch target, and a velocity close to zero a high or low pitch target [4, 13]. Results in Table 1 suggest that the pitch target of the on-focus stressed syllable is likely to be high (in non-final-stressed words) or falling (in final-stressed words) in statements, but rising in questions.

Table 2 shows the mean final velocities of stressed syllables in pre- and post-focus content words in statements and questions, and the t-test results

indicating whether they are significantly different from zero. Pre- and post-focus conditions are grouped together because final velocities of stressed syllables in content words do not differ according to their relative position with focus ($F(1,3) = 1.93$, $p = 0.2588$).

Table 2: Mean final velocities of pre/post-focus stressed syllables, and the corresponding t-test results.

Intonation	Non-final	W-final Non-s-final	W-final S-final
Question	9.75 $t(29) = 3.43$ $p = 0.0018$	17.78 $t(9) = 2.38$ $p = 0.0411$	7.75 $t(19) = 5.10$ $p < 0.0001$
Statement	-6.46 $t(29) = -1.91$ $p = 0.0667$	-1.02 $t(9) = -0.33$ $p = 0.7508$	-6.88 $t(19) = -3.45$ $p = 0.0027$

Repeated measures ANOVAs with sex as between-subject factor and sentence type and stress pattern of pre/post-focus content word as within-subject factors found that final velocities of stressed syllables in pre- and post-focus content words differ in different sentence types ($F(1,3) = 11.46$, $p = 0.0429$). The corresponding t-tests indicate that the pitch target of the pre- or post-focus stressed syllable in a content word is high (if the stress is non-final or word-final but non-sentence-final) or falling (if the stress is word-final and sentence-final) in statements, but rising in questions.

4. DISCUSSION AND CONCLUSIONS

Results of the present study suggest that the intonational contrast between statements and declarative questions in general American English involves variations of both pitch range of post-focus syllables and pitch targets of stressed syllables in content words (focused or non-focused). Specifically, the pitch range of post-focus syllables is compressed and lowered in statements, but compressed and raised in questions, which agrees with the general findings in [2, 8], but is verified by a more detailed syllable-by-syllable (instead of word-by-word) analysis. Statements and declarative questions also differ in underlying pitch targets of stressed syllables in content words, with the former having high or falling pitch targets (depending on the stress pattern, sentential position, and focus condition) and the latter having rising pitch targets. Eady & Cooper [2] also noticed pitch contour differences in content words between statements and questions. However, being concerned only with surface forms and without distinguishing words with different stress patterns, they did not reach a clear conclusion regarding the local pitch variations related to statement/question contrast.

In summary, the following conclusions regarding English intonation are reached: 1) The F_0 difference between statements and declarative questions becomes salient starting from the stressed syllable of the first content word, whether or not it is focused. This manifestation of the sentence type difference is due to a pitch target shift that changes the underlying pitch targets of the stressed syllables from high or falling in statements to rising in questions. 2) Focus expands the pitch range of the focused syllable, compresses and lowers (in statements) or compresses and raises (in questions) that of the post-focus syllables, and leaves that of pre-focus syllables largely unaffected. 3) The underlying pitch target of the stressed syllable in a content word varies systematically both with syllable position in word/sentence, and with focus and sentence type, as indicated by final velocity of the syllable.

5. REFERENCES

[1] Boersma, P., Weenink, D. 2006. *Praat: doing phonetics by computer*. http://www.praat.org/.

[2] Eady, S.J., Cooper, W.E. 1986. Speech intonation and focus location in matched statements and questions. *J. Acoust. Soc. Am.* 80, 402-416.

[3] Gårding, E., Abramson, A.S. 1965. A study of the perception of some American English intonation contours. *Studia Linguistica* 19, 61-79.

[4] Gauthier, B., Shi, R., Xu, Y. 2007. Learning phonetic categories by tracking movements. *Cognition* 103, 80-106.

[5] Hirschberg, J. 2004. Pragmatics and intonation. In: Horn L.R., Ward, G.L. (eds.), *The Handbook of Pragmatics*. Oxford: Blackwell, 515-537.

[6] Ladd, D.R. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.

[7] O'Shaughnessy, D. 1979. Linguistic features in fundamental frequency patterns. *J. Phon.* 7, 119–145.

[8] Pell, M.D. 2001. Influence of emotion and focus location on prosody in matched statements and questions. *J. Acoust. Soc. Am.* 109, 1668-1680.

[9] Pierrehumbert, J., Hirschberg, J. 1990. The meaning of intonational contours in the interpretation of discourse. In: Cohen, P.R., Morgan, J., Pollack, M.E. (eds.), *Intentions in communication*. Cambridge, Massachusetts, MIT, 271-311.

[10] Pierrehumbert, J. 1980. *The Phonology and Phonetics of English Intonation*. Ph.D. dissertation, MIT.

[11] R Development Core Team. 2005. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, http://www.R-project.org.

[12] Xu, Y., Liu, F. 2007. Determining the temporal interval of segments with the help of F_0 contours. *J. Phon*. 35, 398-420.

[13] Xu, Y., Liu, F. In press. Tonal alignment, syllable structure and coarticulation: Toward an integrated model. To appear in *Italian Journal of Linguistics*.

[14] Xu, Y., Wang, Q.E. 2001. Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication* 33, 319-337.

[15] Xu, Y., Xu, C.X. 2005. Phonetic realization of focus in English declarative intonation. *J. Phon.* 33, 159-197.