Working Papers of the Cornell Phonetics Laboratory
2007, v.16, pp. 57-100

Interrogative Intonation in North Kyungsang Korean: Language-specificity and Universality of Acoustic and Perceptual Cues*

Hye-Sook Lee

This paper investigates intonational characteristics of questions distinct from statements in North Kyungsang Korean. A variety of F0 targets are compared between the two sentence types, including pitch accent F0 peak, utterance-final edge, pitch range and F0 valleys. The acoustic analyses show that yes/no questions have significantly higher F0 values at the right edge and higher and earlier occurrences of the rightmost F0 valley. Perceptual salience of these acoustic cues is investigated through forced-choice identification tests, where it turns out that only the utterance-final edge F0 is a strong perceptual cue to distinguish between questions and statements. This study suggests that NK Korean conforms to the "universality" of question intonation in a broad sense, but the realization of acoustic cues and their perceptual salience are constrained by some kind of language-specificity.

1. Introduction

Across languages, it has been reported that declarative sentences are generally marked with low or falling pitch and interrogative counterparts are marked with high or rising pitch. The falling statement and rising question are summarized as the "Strong Universalist Hypothesis" (Ladd 1981, Gussenhoven and Chen 2000). Ohala (1983) argues that this intonational distinction should be accounted for with universal properties shared by human beings and even other primates. According to Ohala's "Frequency Code," "high pitch sounds vulnerable and submissive, while low pitch sounds protective and dominant…" and "high and high-ending utterances seem to sound dependent, appealing or questioning, etc. while conversely, low and low-ending utterances seem to sound authoritative, powerful, and assertive." (Ohala 1983 as quoted by Gussenhoven and Chen 2000: 2). However, as Ohala himself noted, the intonational distinction between statement and question is not only paralinguistic, but it has a denotational function as well.

* I would like to thank Abigail Cohn, Michael Wagner, Amanda Miller and John Whitman for their valuable comments on this study. I also thank two anonymous reviewers for their helpful feedbacks. All mistakes are mine.

Since intonational meaning is usually directly related to phonetic contours, e.g. a fall or a rise, intonation is sometimes considered as natural and paralinguistic. Such naturalness may have something to do with the universal patterns of intonation that are found in many different languages. However, there are still arbitrary aspects in intonation so it is said that intonation remains within a linguistic domain (Fox 2000). For instance, Belfast English declaratives are marked with rising pitch while interrogatives are marked with falling pitch (Gussenhoven and Chen 2000). Moreover, some lexical tonal languages e.g. Chinese (Yuan and Shih 2002) do not show robust intonational distinction between the two sentence types. Gussenhoven and Chen (2000) pointed out that there are both universal and language-specific properties about the two different intonation types, and argued that "universal and language-specific communicative effects derive from two different language components," which are "intonational lexicon" and "phonetic implementation module." These two components are reminiscent of the SPE-type modules of phonology and phonetics. According to them, the intonational lexicon contains the set of morphemes invested with intonational meanings and the mapping between phonology and meaning is arbitrary. The phonetic implementation *per se* is not universal either, but the articulatory mechanisms under the phonetic implementation are.

In this paper, I would like to examine both universal and language-specific properties of North Kyungsang Korean (NK Korean) question intonation, for this dialect of Korean does not seem to display a salient distinction between statements and questions. In particular, I would investigate whether NK Korean uses 'high' end pitch to consistently cue a question and if so, how high the 'high' pitch has to be in order for a sentence to be heard as question. For this, I perform both acoustic analyses and perception tests. Both production and perception reveal phonological and phonetic characteristics of the language.

Acoustic analyses can provide us with more accurate descriptions of intonation than impressionistic observations. It has been said that a NK Korean question is indicated with 'high' or 'rising' intonational tone at the right edge (Suh 1987 among others). In fact, there have been few instrumental investigations of NK Korean intonation so that the generalizations so far have been heavily dependent upon the authors' intuitive judgment.

Consequently, such impressionistic descriptions could not capture the language-specificity of the ‘high’ or ‘rising’ question intonation. In this sense, more systematic analyses with detailed acoustic characteristics are necessary. In the meantime, simple acoustic analyses may be limited as well. I believe that perception can be more decisive than production in the sense that the former is less subject to variability, such as speaker variability. Perception is strongly affected by language-specific phonology in that listeners are more sensitive to cross-categorical signals than intra-categorical signals. Therefore, both production and perception experiments are needed for identifying language-specific properties from which “universal” properties can be extracted.

Higher end pitch, or boundary tone, has been identified as one of the most important cues for questions across languages. ToBI analyses use boundary tones as utterance-level intonation markers, whatever their phonetic detail. For example, English H% (Beckman and Ayers 1994), Tokyo Japanese H% and LH% (Venditti 1997), Mandarin Chinese H% (Chan et al 2005) and standard Korean H%, LH%, HLH% and LHLH% (Jun 2000). These boundary tones may display very clear rising F0 excursions. However, as Yuan, Shih and Kochanski (2002) showed that question intonation after a falling lexical tone is not distinguished from statement intonation in Chinese, the question intonation may not always involve a salient F0 excursion. Jun (2005) argues that lexical tone languages such as Chinese and stress-accent languages such as English behave differently such that the latter demonstrates more flexible intonational patterns than the former, and that lexical pitch accent languages such as Japanese are in between.

Not a few studies used perception tasks to identify cross-linguistic intonational properties of questions. These studies share the findings that there is some kind of universal property but the perception is constrained by the language-specific properties of the listeners’ native languages. Makarova (2001) examined how Japanese and Russian listeners, whose native intonation systems are distinct, can distinguish different sentence types based on the perceived intonation pattern. Makarova (2001) wanted to examine whether two different language groups share certain kinds of perception patterns, which is closely related to language-universality of sentence type intonation. Makarova manipulated several different variables centering on pitch accent peaks, including peak height or slope of rising, for instance. Japanese and Russian listeners showed a similar pattern in perceiving declarative intonation whereas they were not categorically distinguishing questions or exclamatives. She attributed a part of the result to the different intonational properties of the two languages. It seems to me that Japanese puts more weight on utterance-final pitch than pitch accent peaks to distinguish questions from statements, but Makarova manipulated the latter only because she wanted to identify language-specific and universal patterns with a uniform set of variables. Gussenhoven and Chen (2002), in their perception experiments on three different language groups Dutch, Chinese and Hungarian, found a common effect among the three groups, which is that higher peaks, later peaks and higher end pitch seem to lead the listeners to perceive utterances as questions more often. There was a slight difference among the three language groups, such as Hungarian listeners being more sensitive to peak height than Chinese listeners. Their study demonstrates that both universal and language-specific characteristics are manifested in the intonation type perception.

The investigation of NK Korean intonation in the current paper will improve an understanding of the arbitrary nature of linguistic intonation, since this language displays some kind of 'high' tone for questions like in many other languages, but the phonetic details and weight of lexico-syntactic devices marking interrogation differ from other languages, even from the Standard Korean. NK Korean has not been described as fully as Standard Korean, especially in the area of intonation. There have been studies on the pitch accent types and tone sandhi (Kenstowicz and Sohn 2001, N. J. Kim 1997, Jun et al. 2004, among others) and phrasing (Kenstowicz and Sohn 1997, Sohn 2004 among others). However, relatively little has been done with respect to intonation. It is said that NK Korean is similar to Standard Japanese in that both the languages have lexical pitch accents and they both have Accentual Phrases led by L% (Venditti 1997, Jun et al 2004). However, it does not seem to be the case when it comes to question intonation. Comparing phonetic realizations of the two languages impressionistically, standard Japanese has much clearer rising pitch excursion for questions than NK Korean.

This paper will show that NK Korean questions are differentiated from statements mainly in terms of utterance-final F0 and relative height and timing of pre-final low target. The results are supported by acoustic measurements and perceptual tests. The structure of this paper is as follows. I introduce the prosodic systems in NK Korean in section 2. I present the production experiments and discuss the results in section 3. I discuss my perception experiments in section 4, and then I conclude my paper in section 5.

2. NK Korean prosodic system

NK Korean is a lexical pitch accent language. Like Standard Japanese, NK Korean tones are lexically determined, and each prosodic word has only one lexical tone, which plays both culminative and demarcative roles.

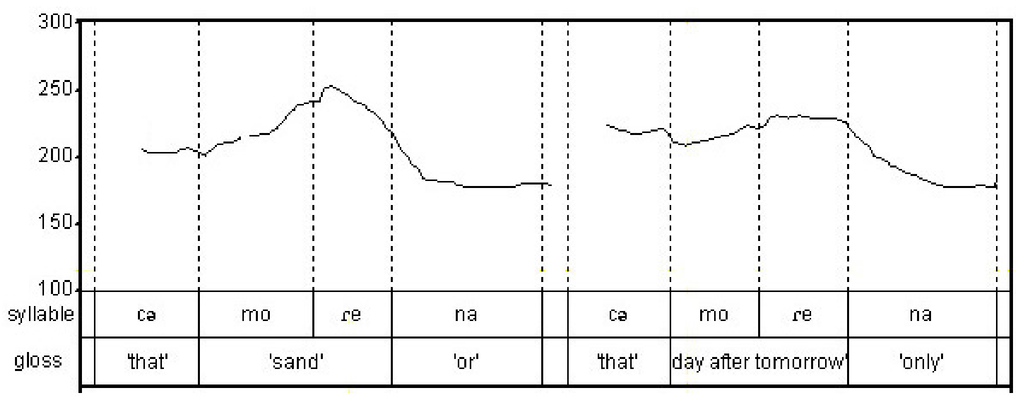

Figure 1: Minimal pair [cəmoɾe-na]: acute H (left; 'or that sand') and non-acute H (right; 'only two days after tomorrow')

NK Korean has two types of tones with respect to phonetic characteristics: acute H and non-acute H. I use these terms because acute H has a sharp F0 peak whereas non-acute H has a flatter F0 peak. Figure 1 displays pitch contours for these tones. I, as a native speaker of NK Korean, produced them.

Figure 1 shows a minimal pair [moɾe] where an acute H tone is lexically linked to *mo-* on the left while a non-acute H tone is lexically linked to both syllables of [moɾe]. In both the F0 contours, F0 peaks occur at the end of or slightly later than the syllable to which the H tones are lexically linked. If *mo-* has an acute H tone, [moɾe] means 'sand' whereas it means 'the day after tomorrow' if a non-acute H tone is associated with both syllables of [moɾe]. Acute H can be linked to different syllables, i.e. initial, penult and final syllables, whereas non-acute H is always linked to the first two syllables of a word.

There are disagreements in the literature about the phonological representation of the NK Korean lexical tones in the literature. Kenstowicz and Sohn (1997) used HL and LHL for acute H and non-acute H respectively, regardless of their phonetic realization, in order to explain an asymmetry between the acute and non-acute H tones: only acute H undergoes "upstep" which is triggered by a preceding acute H. In their analysis, the first L in LHL blocks upstep on LHL. Their representation is purely phonological since LHL for non-acute H does not reflect the intuition that non-acute H bear high pitch on both the first and the second syllables of a word. In their later work on loanword adaptation in NK Korean, Kenstowicz and Sohn (2001) named the acute H tone as *single H* and non-acute tone as *double H*. These terms may better reflect the descriptive nature of the two types of tones. Idsardi and Kim (1997) assumed three different types of input representations for NK Korean H tones in proposing a metrics-based analysis. The surface H tones are represented with two kinds, i.e. single x and double x for acute and non-acute H respectively. This is reminiscent of Kenstowicz and Sohn (2001)'s single H and double H. Idsardi and Chang (2003) classified NK Korean tone types as H and HH, while the single H is further classified into initial, penult and final H, depending on the tone placement. N. Kim (1997) used the terms Default H and Pre-linked H for acute H and Floating H for non-acute H. Despite the variations in notation across different analyses, it seems generally accepted that there are two types of lexical tones, acute and non-acute H, and acute H tones can be associated with initial, penult, and final syllable of a word. In the current paper, I control the type of lexical tones such that all the words that I use in the experiment bear acute H tones on their initial syllables.

Jun et al. (2004) claim that NK Korean demarcates a phonological word with a preceding %L, much like standard Japanese. Thus a phonological word is marked by an obligatory low target before the word and by the lexical pitch accent F0 peak. They also propose that NK Korean has Phonological Word-Accentual Phrase-Intonation Phrase as hierarchical prosodic structure, though they do not discuss Intonational Phrase boundary markers. The IP or Utterance level characteristics are the major concern of the current paper, i.e. the identification of melodic patterns for sentence types, whether or not such patterns are incorporated as IP "boundary tones."

Before moving on to the experiments, let me briefly discuss non-intonational question markers. Although prosody seems to be 'the one and only true universal' (van Heuven and Haan 2000) with respect to interrogativity, lexico-syntactic devices are also widely used across languages. Obvious examples are the use of wh-words and reversals of word orders in Germanic languages. The presence of wh-words clearly marks wh-interrogation of the sentence, and the reversal of subject and auxiliary verb, as in English for instance, overtly marks that the relevant sentence is a question. NK Korean also has wh-words to mark wh-questions but lacks Subject-Aux Inversion. However, NK Korean lexically marks statements and questions with sentence-final verbal particles, such as *–ta*, *-na*, and *–no* for declarative, yes/no interrogative and wh interrogative sentences respectively. Such lexico-syntactic factors are important because they may interact with intonation.[1] However, in the present study, I control the sentence-final particles in the current study to focus on intonational patterns without worrying about syntactic-semantic effects of those sentence particles. Specifically, I use only a type-neutral particle *–yo* in comparing statements and questions.

Now I turn to experimental works. In the following section, I first present acoustic analyses of a pilot study and then the full experiment.

3. Acoustic analyses

A general impression about NK Korean questions is that they are 'high' in F0, mostly toward the end. However, the 'highness' of question intonation seems to vary a lot, and such an impression has never been systematically investigated. In this section, I systematically measure a set of parameters and compare those of question and statement to establish how questions are different from statements with respect to intonation. In section 3.1, I discuss variations of question intonation rather impressionistically to show

[1] The interaction between particles and intonation is a very interesting research topic, but beyond the scope of this paper. It is investigated in my work in progress.

that there *is* some kind of 'high' target in questions. In section 3.2, I move on to the full experiments where I look into the possibility that there are more acoustic cues for questions than just the right-edge F0.

3.1 Pilot study: Variations at the right edge

I recorded two native speakers reading short sentences made of one or two words to find out general intonational patterns of statements and question. From this preliminary acoustic analysis, it has been found that there is a systematic variation with respect to the right-edge F0 of questions. In particular, the right-edge F0 of NK Korean questions seems to vary depending on the position of the rightmost pitch accent. The divergence between questions and statements is greatest when the pitch accent is placed on the ultimate syllable of a sentence. The farther away the pitch accent is placed from the right-edge, the smaller the divergence between questions and statements.

The greatest F0 difference between questions and statements is found when pitch accent is associated the sentence-final syllable in the absence of a sentence-final particle. For instance, Figure 2 shows representative F0 contours that correspond to the two-word sentence in (1). The sentence was read either as a question or as a statement in accordance to different punctuations on the written script. Therefore, the question and the statement are a minimal pair whose only difference is intonation.

(1) Yengmai-ne namwu[2]
Yengman-Poss tree
'Yengman's tree'

[2] Non-bracketed Korean sentences are romanized in accordance of the Yale Romanization Convention.

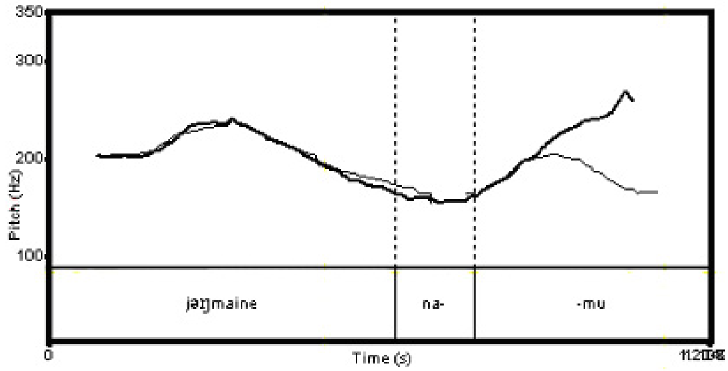

Figure 2: Pitch contours of [jəŋmaine namu]: statement (thin line) and question (thick line).[3]

The question contour in Figure 2 seems to rise continuously from the accent peak to the end, whereas the statement contour falls from the peak. The continuous rise in the question in Figure 2 makes the difference between the two F0 contours very salient. Figure 3 also shows two F0 contours where the question and statement contours look quite parallel in pattern. The sentences in Figure 3 have the rightmost pitch accent associated with the penultimate syllables of the sentences, as there is a sentence-final particle [-jo] as shown in (2). The question contour (thick line) in Figure 3 seems to end up being slightly higher than the statement contour (thin line), but the difference between question and statement is smaller than that in Figure 2.

(2) Yengmai-ne namwu-yo
Yengman-Poss tree-Polite
'Yengman's tree'

[3] The underlines mark the position of acute H tones in this paper.

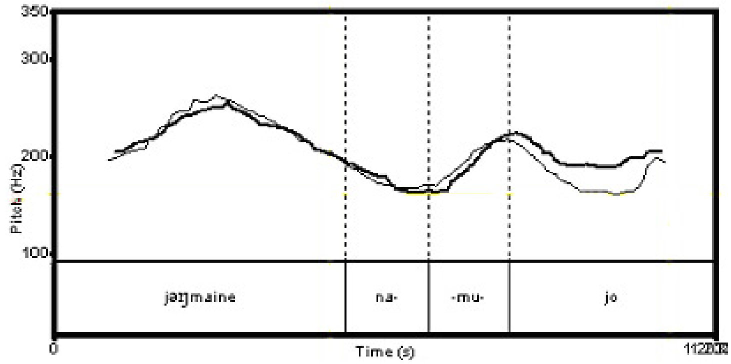

Figure 3: Pitch contours of [jəŋmaine na<u>mu</u>jo]: statement (thin line) and question (thick line).

Figure 4 shows very similar F0 contours for a statement and a question, where only a small divergence between the two utterances is seen at the right edge. In this case, the rightmost pitch accent peak is followed by two syllables, as shown in (3). The post-peak F0 is not as high for the question as in Figure 2 and Figure 3, although the question still seems sustained in F0 while the statement continuously falls to the end.

(3) Yengmai-ne namwu-re-yo
Yengman-Poss tree-be-Polite
'It's (Is it) Yengman's tree'

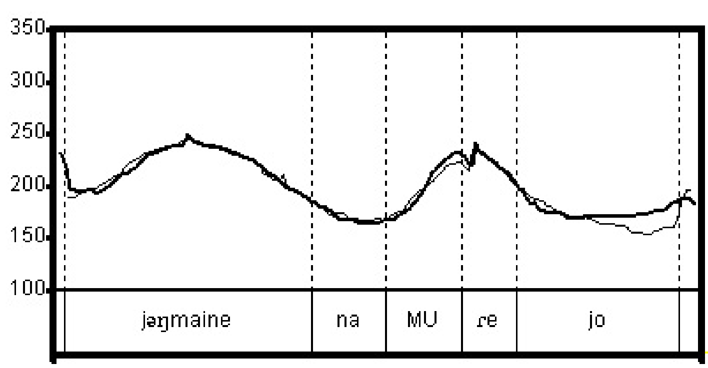

Figure 4: Pitch contours of [jəŋmaine na<u>mu</u>ɾejo]: statement (thin line) and question (thick line).

Since the F0 contours of questions and statements are very similar, it may be difficult to differentiate between the two sentence types solely by intonation, and the two intonational contours may be ambiguous in actual speech. If contextual and speaker variability are also considered, the differentiation of the sentence types may become even more problematic.

The pilot study shows that there are cases where questions are clearly distinct from statements solely in terms of the right-edge F0 and there are other cases where the difference is small. It raises two questions for the latter cases. First, is right-edge F0 still a major acoustic cue for distinguishing questions and statements? Second, if right-edge F0 is not crucial, are there any other intonational cues for questions? In the following section, I try to answer these questions, by conducting a full experiment with more and longer sentences.

3.2 Methodology

In order to resolve the questions raised in the previous section, I investigated more intonational parameters with which questions can be differentiated from statements in NK Korean.[4] The choice of parameters is determined by the general findings of the pilot study and by cross-linguistic findings about trends in F0. The parameters are presented in section 3.2.1, and experimental materials and procedures are introduced in sections 3.2.2 and 3.2.3, respectively.

3.2.1 Intonational parameters

The full experiment investigates the twelve parameters listed in (4) below. I assume that these intonational targets are possible candidates for distinguishing between statements and questions in NK Korean.

[4] I only compare yes/no questions and statements in the current study, but I had also looked at y/n questions versus wh questions, but the sentence-final particle type was not controlled in a consistent way, so I do not include the data and results of y/n versus wh questions in this paper. To summarize the results briefly, however, it turned out that there is no common intonational cue to distinguish between the two question types. A follow-up study is needed, with particle types being controlled.

(4) Parameters under investigation

a. Peak F0: F0 of nuclear pitch accent peak (Highest F0 point)

b. Edge F0: Utterance-final F0 at vowel offset of utterance-final syllable

c. Peak Time: Time (pitch accent peak - onset of accented vowel)

d. Pitch Range: Peak F0 – F0 of the immediately preceding valley

e. Early Trend F0: F0 (Valley 1 - Valley2)

f. Late Trend F0: F0 (Valley 1 - Valley3)

g. Valley2 Time: Time (Valley2 - Valley 1)

h. Valley3 Time: Time (Valley3 - Valley 1)

i. Early Slope: 100*(Early Trend F0/ Valley2 Time)

j. Late Slope: 100*(Late trend F0"/ Valley3 Time)

Figure 5 illustrates where those intonational target points are. The top panel in Figure 5 shows five F0 targets around the focused Word4 (Late Focus): Valley1, Valley2, Valley3, Peak and Edge. From the three Valleys, I calculate values of the parameters, Early and Late Trend F0s (4e-f), Valley2 and Valley3 Time (4g-h) and Early and Late Slope (4i-j): F0 and time differences between Valley1 and Valley2, F0 and time differences between Valley 1 and Valley3, and slopes of Valley2 and Valley3 relative to Valley 1. Further, Pitch Range (4d) is also calculated from Peak F0 and either Valley2 F0 for Late Focus or Valley1 F0 for Early Focus: i.e. F0 (Peak – Valley2) at Late Focus and F0 (Peak – Valley1) at Early Focus (see Figure 5).

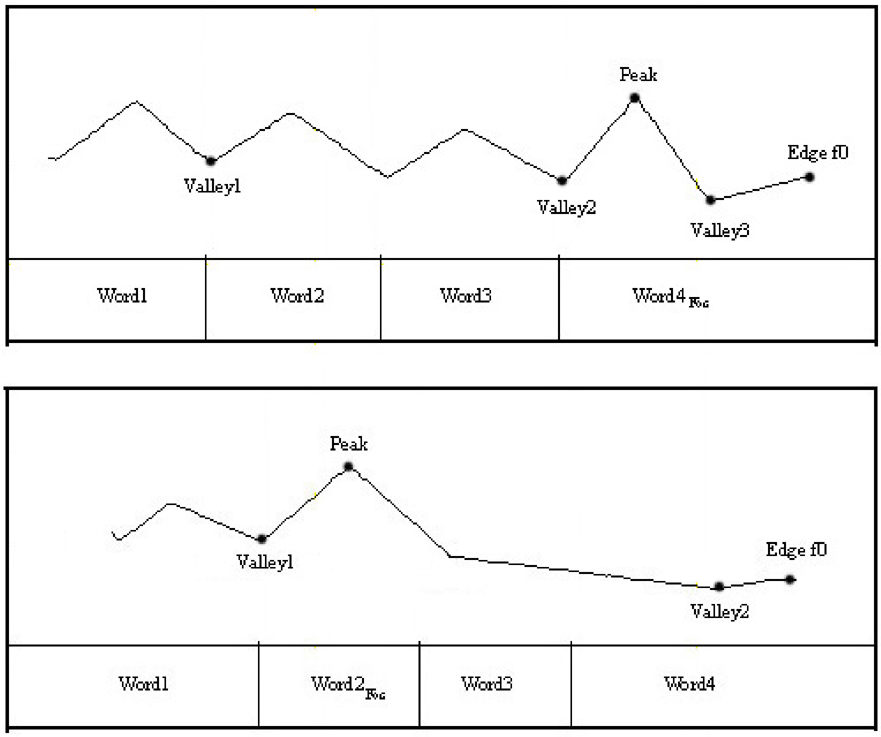

Figure 5: Target points to be measured: top panel for Late Focus (Foc4) and the bottom panel for Early Focus (Foc2)

The bottom panel of Figure 5 shows all four targets to be measured around the focused second word (Early Focus). Note that parameters Late Trend, Valley3 Time, and Early Slope are not applicable to the Early Focus cases, since the Early Focus yields only one Valley after the first valley.

Now let us move on to the experimental materials in the following sections.

3.2.2 Materials

In order to compare yes/no questions with statements in terms of the given parameters, I prepared experiment sentences in the following way. Each sentence is made up of four words, ending with an utterance-final 'informal polite' particle *-yo*, which can be used for both questions and statements. Since the utterance-final particle is neutral, each sentence is potentially ambiguous. Therefore, intonation is expected to be the only cue for distinguishing between statements and questions. Each sentence is placed in a context such that it could be read either as a question or a statement. Further, each sentence has two potential focus positions, which are second and fourth words. Focus was placed in two different positions to control the placement of the rightmost pitch accent peak, either close to or far from the right edge. Again, the intended focus is clear from context. Therefore, each sentence was produced in four different contexts and so is expected to exhibit four different intonation patterns. In addition, a pitch accent comes on the first syllable for every word, and most of the segments are sonorants. For instance, one sentence in four different conditions is shown in (5) below. Bold-faced words indicate foci and underlined syllables bear accents or lexical tones. The underlined tones are all acute H tones.

(5) One example sentence in four different conditions

a. Statement - Early Focus

<u>Mwun</u>ho-ney	**<u>nwu</u>na-ka**	<u>mo</u>ley-lul	<u>mi</u>le-yo.
Mwunho-Gen	sister-Nom	sand-Acc	push-Particle.

‘Mwunho’s **sister** pushes sand.’

b. Question – Early Focus

<u>Mwun</u>ho-ney	**<u>nwu</u>na-ka**	<u>mo</u>ley-lul	<u>mi</u>le-yo?
Mwunho-Gen	sister-Nom	sand-Acc	push-Particle.

‘Does Mwunho’s **sister** push sand?’

c. Statement – Late Focus

<u>Mwun</u>ho-ney	<u>nwu</u>na-ka	<u>mo</u>ley-lul	**<u>mi</u>le-yo**.
Mwunho-Gen	sister-Nom	sand-Acc	push-Particle.

‘Mwunho’s sister **pushes** sand.’

d. Question – Late Focus

<u>Mwunh</u>o-ney	<u>nwu</u>na-ka	<u>mo</u>ley-lul	**<u>mi</u>le-yo**?
Mwunho-Poss	sister-Nom	sand-Acc	push-Particle.

‘Does Mwunho’s sister **push** sand?’

On the written script, statements and questions were marked with different punctuations, and different foci were induced by different contexts. Five sentences were used, and 20 sentences (5 sentences x 4 conditions) were presented to each speaker. See Appendix A for the exhaustive list of the materials.

3.2.3 Participants and procedures

Two native speakers of NK Korean participated in the recording: one male and one female speaker. The two speakers were recorded at different places in Korea but both of them were recorded in a quiet setting. The male speaker is in his late teens and the female speaker is in her late twenties. The sentences were listed on two sheets of paper, and they were grouped by the conditions (punctuation and focus) so that there were four chunks of sentences. The sentences were not shuffled in a random way in order not to confuse the speakers. The male speaker read each sentence five times and the female speaker read them six times. They read all the twenty sentences once and then went back to the first sentence, reading them through again, and they repeated this process five or six times. One of the difficulties in producing expected patterns was that the sentences were so long that the male speaker failed to produce Early Focus (i.e. focus on the second word with depressed F0 afterwards). In addition, the male speaker often produced Seoul Korean intonations such as rising at the phonological word boundary. In these cases, I corrected it and had him produce those sentences all over again. The recording was done with Praat on my laptop computer.

The F0 and time measurements were made in Praat with reference to waveforms, and spectrograms. I also relied on my own perception when determining the syllable-offset. The perceptual judgment was crucial for the male speaker’s data because his voice was quite breathy and the breathiness was greater at the end of the sentence. Due to this breathiness, F0 tracks were truncated in tokens, and critically truncated tokens were excluded from the analysis.

In addition to the measurements mentioned above, accented syllable onsets and the utterance-final syllable offsets were annotated. Accurate labeling of vowel onset and offset of each relevant syllable is important since timing of peaks are calculated relative to the onsets of accented vowels. Praat scripts were used for systematically measuring F0 and timing. The targets in (4) were annotated while the scripts being run. Figure 6 demonstrated how words and F0 targets are annotated.

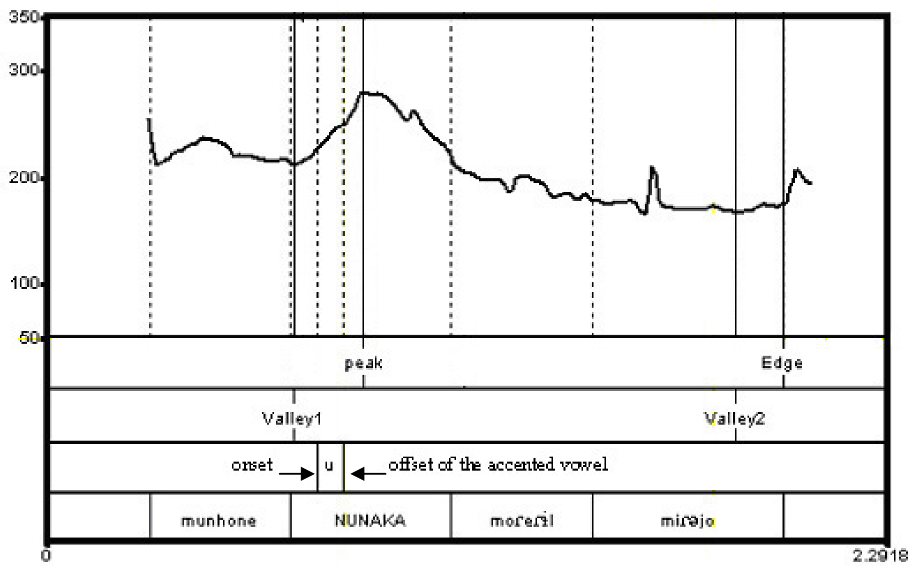

Figure 6: Annotation of words, accented vowel onset and offset, and F0 targets on an interrogative sentence [munhone nunaka moɾeɾɨl miɾəjo] ‘Munho’s **sister** pushes sand.’

On the first and second labeling tiers in Figure 6, F0 targets are marked and correspondent timings are marked with solid lines. The onset and offset of the accented vowel [u] of the focused word [nuna-ka] (‘Sister-Nom’) are marked on the third tier. On the fourth tier, lexical words are labeled. Again, waveforms, spectrograms, and intensity curves were also referred to in Praat for accurate detections of F0 targets. For waveforms, spectrograms and intensity curves, default settings were used in Praat, and for F0 contours, the pitch range was adjusted depending on the speakers.

Now let us move on to the results of acoustic analyses in section 3.3.

3.3 Results

Overall, it turns out that intonational parameters for distinguishing questions from statements are found toward the end of the sentence, such as Edge F0 and Late Trend F0. The acoustic analyses suggest that the two speakers use different strategies to differentiate statements and questions. The data of the two speakers are not pooled for the analysis because the speakers have different pitch ranges, and they seem to have different ways of marking questions. Further, the male speaker failed to produce the Early Focus intonation. Therefore, the results of the female speaker are presented first, and then the male speaker's Late Focus data are presented.

The female speaker shows the clearest distinction between questions and statements in Edge F0 in both the early and Late Focus conditions. Figure 7 shows I present boxplots of all the targets that were measured under Late Focus. Boxplots for question and statement are paired for each variable. As shown in the boxplots, the female speaker produced questions very differently from statements in terms of Pitch Range (top right panel), Late Trend F0 (top right panel), Valley3 Time (bottom middle panel), and Late Slope (bottom right panel). All the other parameters show that questions and statements overlap to a large degree.

For verifying the significance of the differences observed in Figure 7, multivariate ANOVA (MANOVA) was run on the data of Late Focus, with the independent variable *Sentence Type* (*SentType*), which is either question or statement. The results confirm that questions are significantly different from statements in the variables of Edge F0 [$F=56.05$, $p<0.001$], Pitch Range [$F=19.82$, $p<0.001$], Late Trend F0 [$F=13.36$, $p<0.005$], and Valley3 Time [$F=32.94$, $p<0.001$]. The difference between questions and statements is also statistically significant in Peak F0 [$F=10.52$, $p<0.005$], Valley2 Time [$F=10.46$, $p<0.005$], and marginally Late Slope [$F=4.62$, $p<0.05$].

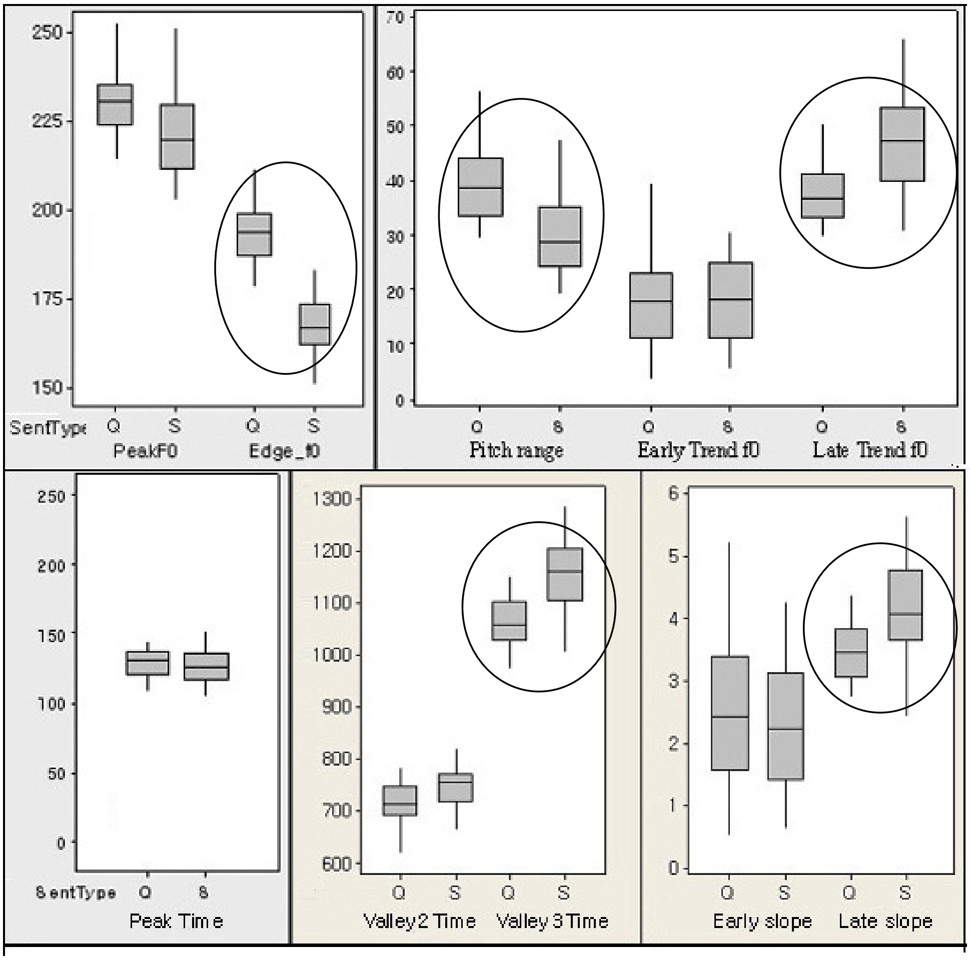

Figure 7: The boxplots of all the variables for the female speaker at Late Focus: Peak F0 and Edge F0 (top left), Pitch Range, Early Trend F0, and Late Trend F0 (top right), Peak Time (bottom left), Valley2 Time and Valley3 Time (bottom center) and Early Slope and Late Slope (bottom right). The horizontal line within the box indicates the median, the vertical length of the box the inter-quartile range, and the whiskers 1.5 times of the inter-quartile range. The units are Hz and ms for F0 and time respectively.

Overall, the female speaker seems to have produced questions with a *higher* pitch accent peak, *higher* Edge F0, a *larger* pitch range, a *higher* F0 for the rightmost valley (Valley3; Late Trend F0), an *earlier* rightmost valley (Valley3 Time), and an *earlier* penultimate valley (Valley2). The results are summarized in Figure 8.

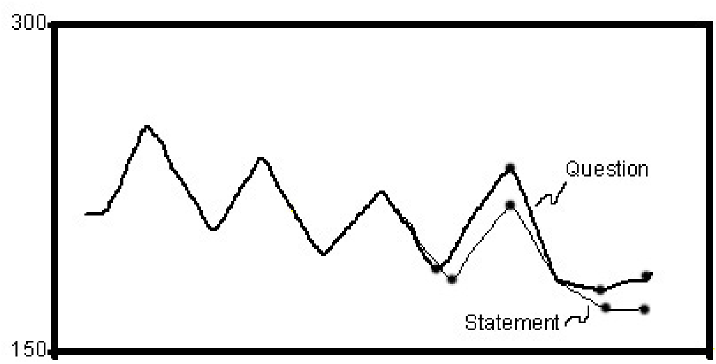

Figure 8: The schematic F0 contours of a question and a statement: the female speaker with Late Focus

Figure 8 shows schematic F0 contours for questions and statements to summarize the female speaker's production under the Late Focus condition. In the question contour (thick line), Valley2 is higher and earlier, the peak is higher, the penultimate and rightmost low target are higher and earlier, finally the right-edge is higher. The contours in Figure 8 are meant to show the differences at critical targets, so the differences are slightly exaggerated.

The Early Focus (Focus2) data show similar results for the female speaker. Figure 9 displays boxplots for all measurements. The clearest deviation between question and statement is found in the parameter Edge F0. In other words, questions with Focus2 seem to have clearly higher F0 at the right edge than statements for this speaker. Peak F0, Pitch Range, Trend F0, and Valley2 Time also seem to show quite different distributions between questions and statements. Like the Late Focus condition, questions have higher pitch accent peaks, larger pitch range, higher relative Valley2 F0 (smaller Trend F0), and earlier occurrence of Valley2. However, the boxplots do not show whether the two sentence types are *significantly* different, so MANOVA was run on the data of Focus2, with the independent variable *SentType*.

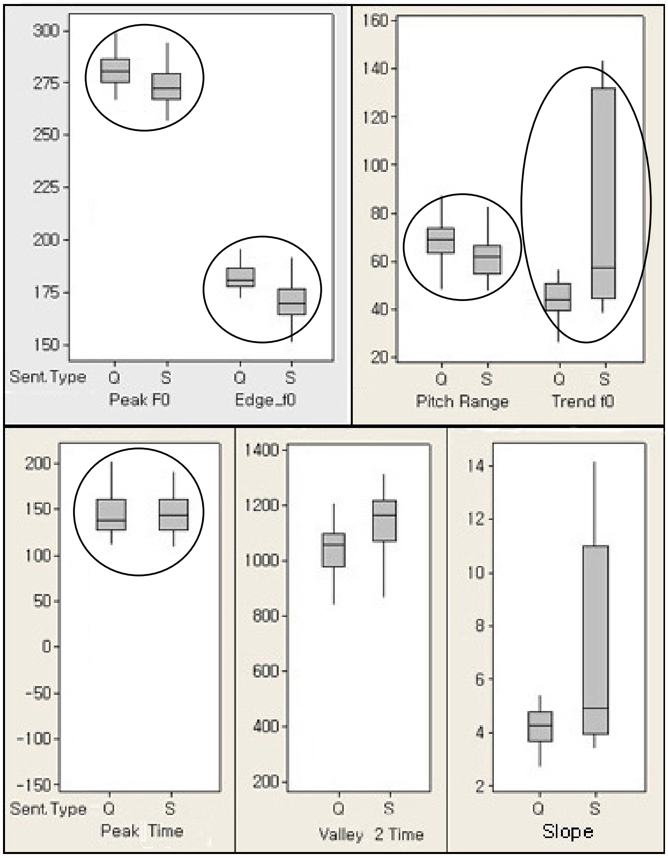

Figure 9: The boxplots of all the variables for the female speaker at Focus2: Peak F0 and Edge F0 (top left), Pitch Range and Trend F0 (top right), Peak Time (bottom left), Valley2 Time (bottom center) and Slope (bottom right). The units are Hz and ms for F0 and time respectively.

The MANOVA results show that Pitch Range, Edge F0 and Valley2 Time have strongly significant differences between the two sentence types: Pitch Range [F=11.76, p<0.005], Edge F0 [F=26.41, p<0.001], Valley2 Time [F=15.83, p<0.001], Trend F0 [F=8.39, p<0.01], and Peak F0 [F=7.20, p<0.01]. Figure 10 schematically summarizes these results.

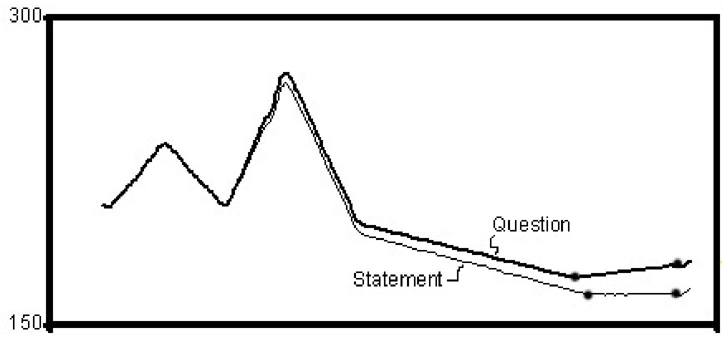

Figure 10: The schematic F0 contours of a question (thicker line) and a statement: the female speaker with Early Focus

As shown in Figure 10, the question F0 contour has greater Pitch Range, higher and earlier Valley2, and higher Edge F0. Note that the F0 contours are schematized so that the differences between the two contours may be exaggerated. The common cues for both Late and Early Focus are greater Pitch Range, higher and earlier final Valley, and higher Edge.

Absolute F0 differences vary according to focus positions. In order to see whether Focus2 and Focus4 are systematically different in terms of F0, I ran ANOVA on the Focus2 and Focus4 data, with independent variables *SentType* and *Focus*. Pairwise Comparisons of these independent variables show that Focus2 and Focus4 have significantly different Edge F0 for questions whereas they are not significantly different for statements. I ran *Simple Main Effects* (Winer et al. 1991; Yuan 2005), whose SPSS commands are borrowed from Yuan (2005). Table 1 displays the pairwise comparisons with the variable Edge F0.

SentType	(I) Focus position	(J) Focus position	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference [a]	
						Lower Bound	Upper Bound
Q	Focus2	Focus4	-11.620*	2.850	.000	-17.267	-5.974
S	Focus2	Focus4	3.726	2.801	.186	-1.822	9.275

Based on estimated marginal means, * The mean difference is significant at the .05 level.
[a] Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Table 1: Pairwise Comparisons with Dependent Variable *Edge F0*

As shown in Table 1, Focus2 and Focus4 are significantly different in questions (Q) [p<.001], but the focus difference is not significant in statements (S)[p=0.186]. The mean difference (I-J) of questions is -11.62, which means that Late Focus bear higher Edge F0 than Early Focus in questions.

The same Pairwise Comparisons were run with other dependent variables such as Pitch Range, Peak Time and Peak F0. See Appendix C for an exhaustive list of tables. With Pitch Range, first of all, both questions and statements show significant differences between Focus2 and Focus4 [p<0.001]. The mean differences between Focus2 and Focus4 are around 30 Hz for both questions and statements. In other words, the Pitch Range is larger for Early Focus than Late Focus in both questions and statements. Next, the Peak Time is also significantly different between Focus2 and Focus4 in both questions and statements ([p<0.005 and <0.001] respectively). The mean differences of Peak Time between Focus2 and Focus4 are 16 ms and 20 ms for questions and statements respectively. That is, focused pitch accent peak comes later relative to the onset of the accent vowel in Focus2 than in Focus4 in both the sentence types. Finally, Peak F0 is also significantly different for Focus2 and Focus4 in both questions and statements [p<0.001]. The mean differences of Peak F0 between Focus2 and Focus4 are 49 Hz and 53 Hz for questions and statements respectively. Namely, Focus2 induces a higher focused pitch accent peak for both questions and statements. In sum, Early Focus has a higher and later peak but lower F0 at the right edge than Late Focus.

Now let us move on to the male speaker. Recall that the female speaker showed significant differences between sentence types in terms of Edge F0, Pitch Range, Late Trend F0, Valley2 Time, Valley3 Time, and Peak F0. The male speaker’s intonation is flatter in the Pitch Range and absolute differences between sentence types are not as large as the female speaker’s. Moreover, he seems to use slightly different strategies to distinguish sentence types. Figure 11 shows boxplots of all the measurements.

Figure 11 contains boxplots for the Late Focus condition (Focus4) of the male speaker. In Figure 11, each parameter has a pair of boxplots for questions (Q) and statements (S), mean differences between questions and statements are large in such parameters as Edge F0, Late Trend F0, Valley2 Time and Valley3 Time.

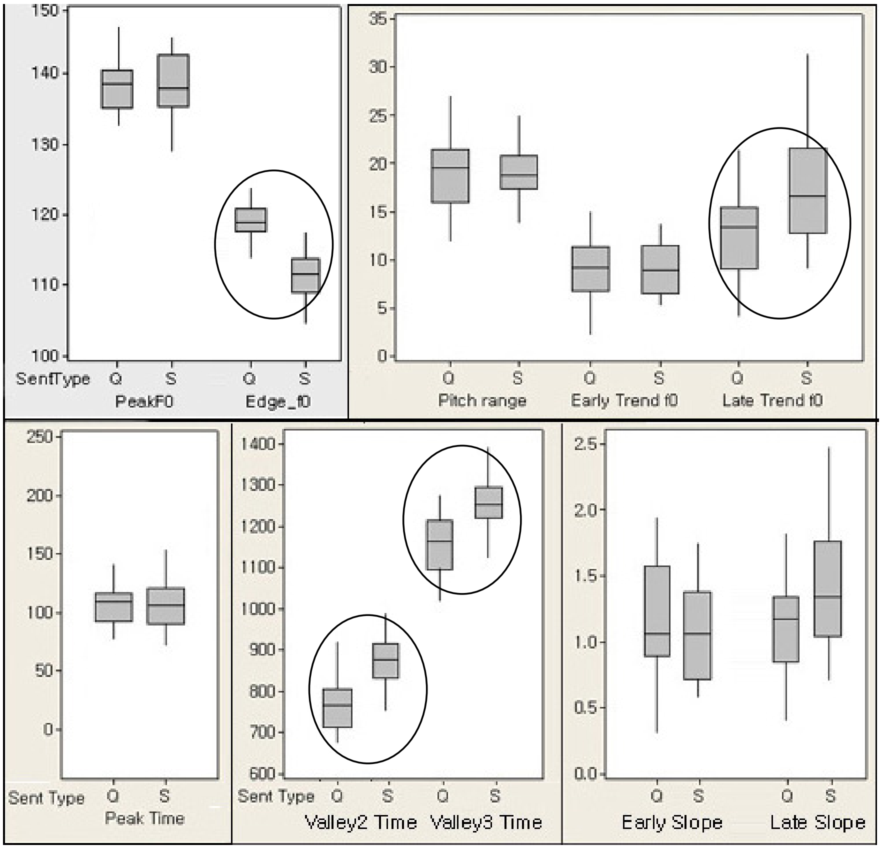

Figure 11: The boxplots of all the variables for the male speaker at Late Focus: Peak F0 and Edge F0 (top left), Pitch Range, Early Trend F0, and Late Trend F0 (top right), Peak Time (bottom left), Valley2 Time and Valley3 Time (bottom center) and Early Slope and Late Slope (bottom right). The units are Hz and ms for F0 and time respectively.

To test whether these variables show significant differences between the two sentence types, MANOVA was run with the independent variable *SentType*. It turns out that only Edge F0, Late Trend F0, Valley2 Time and Valley3 Time are significantly different for questions and statements: Edge F0 [$F=34.947$, $p<0.001$], Late Trend F0 [$F=10.441$, $p<0.005$], Valley2 Time [$F=19.999$, $p<0.001$], and Valley3 Time [$F=14.190$, $p<0.005$].

Overall, the male speaker had higher Edge F0, smaller F0 difference between Valley1 and 3 (Late Trend F0), and earlier occurrences of Valley2 and Valley3, when the focus is on the fourth word. These results are similar to the female speaker's. In fact, the male speaker used a subset of the parameters that the female speaker used in distinguishing questions from statements.

Note that these targets are all located toward the ends of utterances. Recall that Late Trend F0 is defined as F0 differences between Valley1 and Valley3, so the smaller Late Trend F0 means relatively higher Valley3. This can be interpreted as undershot in the anticipation of the final rise. Moreover, the smaller (i.e. earlier occurrences of) Valley2 and Valley3 Time also seem to be related to the final rise. Under Late Focus, Valley2 and Valley3 are the lowest F0 targets immediately preceding and following the rightmost pitch accent peak respectively. Unlike in statements, where the contour falls continuously from the rightmost accent peak, the final rise in questions seems to cause preceding low targets to occur early. Edge F0 is generally correlated with the other three variables (see Table 2). Table 2 shows that Edge F0 is strongly correlated with Late Trend F0 and Valley3 Time in the female speaker, whereas it is significantly correlated with all three variables in the male speaker. Edge F0 is negatively correlated with these variables, which means that as Edge F0 becomes higher, the other variables have smaller values, i.e. a higher Valley3 (smaller difference from Valley1), and earlier occurrences of Valley2 and Valley3.

Speaker: Female		Edge F0	Late Trend F0	Valley2 Time	Valley3 Time
Edge F0	Pearson Correlation	1	-.439**	-.224	-.468**
	Sig. (2-tailed)		.001	.097	.000
	N	58	56	56	56
Speaker: Male		Edge F0	Late Trend F0	Valley2 Time	Valley3 Time
Edge F0	Pearson Correlation	1	-.382*	-.447**	-.393*
	Sig. (2-tailed)		.021	.006	.018
	N	36	36	36	36

Correlation is significant at the 0.01 level (2-tailed) and 0.05 level (2-tailed) for ** and *.

Table 2: Correlations between variables (Late Focus)

This section has shown that not only F0 at the end but also other acoustic cues are involved in marking questions. Consistent parameters which were used to distinguish questions from statements were Edge F0, Valley2 Time and Valley3 Time. In particular, sentence-final F0 was higher, and the rightmost valleys were relatively higher (smaller Trend F0) and earlier in questions than in statements. It has turned out that strongly differential acoustic cues were clustered toward the end, and that these variables were significantly correlated with one another. In other words, these four parameters are not independent. As discussed earlier, NK Korean final rise for questions is relatively shallow, but the relatively shallow rise seems to be compensated for by the other acoustic cues. It has been shown that these acoustic cues have statistically significant differences between questions and statements. However, this does not necessarily entail that they are also perceptually significant. In order to understand whether these acoustic cues are also perceptually salient, identification tests are conducted in Section 5. Testing perceptual salience is important because acoustic distinctions, when present, might not directly reflect the phonology of the question intonation in this language, and the statistically significant differences may be only a byproduct of the small size of the sample populations. With semi-synthetic data, it is expected that the perception tests will support the significance of the acoustic cues and also demonstrate their perceptual salience.

4. Perceptual salience

In the previous section, various acoustic cues with which yes/no-questions can be distinguished from statements were identified. It was shown that some acoustic variables including Edge F0 have significant differences for those sentence types. To support the acoustic results, a perceptual identification test was performed. The identification test is expected to show whether acoustic variables are perceptually salient with respect to sentence types.

A subset of the parameters (variables) that were analyzed in the production experiment was manipulated for the perception test. From the perception test, some of the variables turn out to yield significantly different percepts of questions and statements, while others do not. I present the manipulation processes for all parameters regardless of the results from the previous section, but I restrict my discussion to those parameters that have shown statistically significant differences at the previous section. In the following section, the methods for the stimuli and perception test are introduced, and then the results are discussed in section 4.2.

4.1 Methodology

4.1.1 Material and manipulation

In order to create multiple semi-synthetic stimuli, natural speech was used for manipulation. Sentences to be manipulated were selected from the female speaker's data for the production experiment (see (5) above). Recall that there were five different sentences, and each sentence was produced with either focus condition and either punctuation. That is, one sentence was produced in four different ways. Among them, declarative sentences with two different focus conditions per variable were used for the manipulation.

Stimulus manipulation was done with Praat manipulation functions. Pitch contours were stylized to produce a minimal number of F0 targets. Selected F0 targets corresponding to the parameters were then moved around, and then the sentences were resynthesized with the Psola Resynthesis function. Similarly to the production experiment, seven variables were chosen: Peak F0 and Time, Valley1 F0 and Time, Valley2 F0 and Time, and Edge F0. Noteworthy is that Valley1 (L1) and Valley2 (L2) are defined slightly differently for the manipulation: Valley1 is a valley immediately preceding the final accentual peak and Valley2 is a low target immediately following the final accentual peak. Valley3 Time and Late Trend F0, which were significantly different for the two sentence types in the previous section, were not included for the perception test, because the tests were conducted before the analyses on these acoustic parameters were completed.

Figure 12 and Table 3 show how pitch targets were manipulated. Each panel in Figure 12 represents one variable. F0 variables manipulated at the Early Focus (Focus2) are shown in the left-hand panels, and time variables manipulated at Late Focus (Focus4) are shown in the right-hand panels. In the actual manipulations, however, each variable was manipulated at both Focus2 and Focus4. Six to ten stimuli were synthesized per variable and focus, as each "step" in Table 3 and Figure 12 is equivalent to a separate synthetic stimulus. F0 values were manipulated in 5 Hz steps and time values were manipulated in 10 ms steps, so the stimuli within each variable make up a continuum of F0 height or timing. Note that the variables were manipulated independently of each other. When one variable is manipulated, all the other acoustic values were kept constant.

Variables	**Manipulation**
Peak height (Peak F0)	Increase rightmost peak F0 by 5 Hz, yielding 10 steps - Focus2: 280 ~ 325 Hz - Focus4: 250 ~ 295 Hz
Peak timing (Peak Time)	Move rightmost peak relative to offset of accented vowel by 10 ms, ranging 40 ms before and 50 ms after the vowel offset. (10 steps)
Valley1 height (L1 F0)	Increase L1 F0 by 5 Hz, yielding 8 steps - Focus2: 210 ~ 245 Hz - Focus4: 180 ~ 205 Hz
Valley1 timing (L1 Time)	Move L1 relative to onset of accented vowel by 10 ms, ranging 50 ms before and 40 ms after the vowel onset. (10 steps)
Valley2 height (L2 F0)	Increase L2 F0 by 5 Hz, yielding 6 steps - Focus2: 190 ~ 215 Hz - Focus4: 165 ~ 190 Hz
Valley2 timing (L2 Time)	Move L2 relative to onset of word-final syllable by 10 ms, ranging 90 ms before and up to the syllable (vowel) onset. (10 steps)
Edge F0	Increase F0 at offset of the utterance-final syllable by 5 Hz, with F0 being constant at the syllable onset, yielding 8 steps - Focus2: 163 ~ 198 Hz - Focus4: 167 ~ 202 Hz

Table 3: Manipulation of variables

Variables	**Hypotheses**
Peak height (Peak F0)	The higher the peak, the more question responses.
Peak timing (Peak Time)	The earlier the peak, the more question responses.
Valley1 height (L1 F0)	The lower the L1, the more question responses.
Valley1 timing (L1 Time)	The earlier the L1, the more question responses.
Valley2 height (L2 F0)	The higher the L2, the more question responses.
Valley2 timing (L2 Time)	The earlier the L2, the more question responses.
Edge F0	The higher the Edge, the more question responses.

Table 4: Predictions

The predictions from the manipulations are listed in Table 4. Since the manipulations are deviations from the original *declarative* sentences, although directions of manipulation vary for different variables, those deviations may bias listeners toward questions if they are perceptually correlated with questions. Otherwise, they may not have effect on the percepts throughout the phonetic continuum.

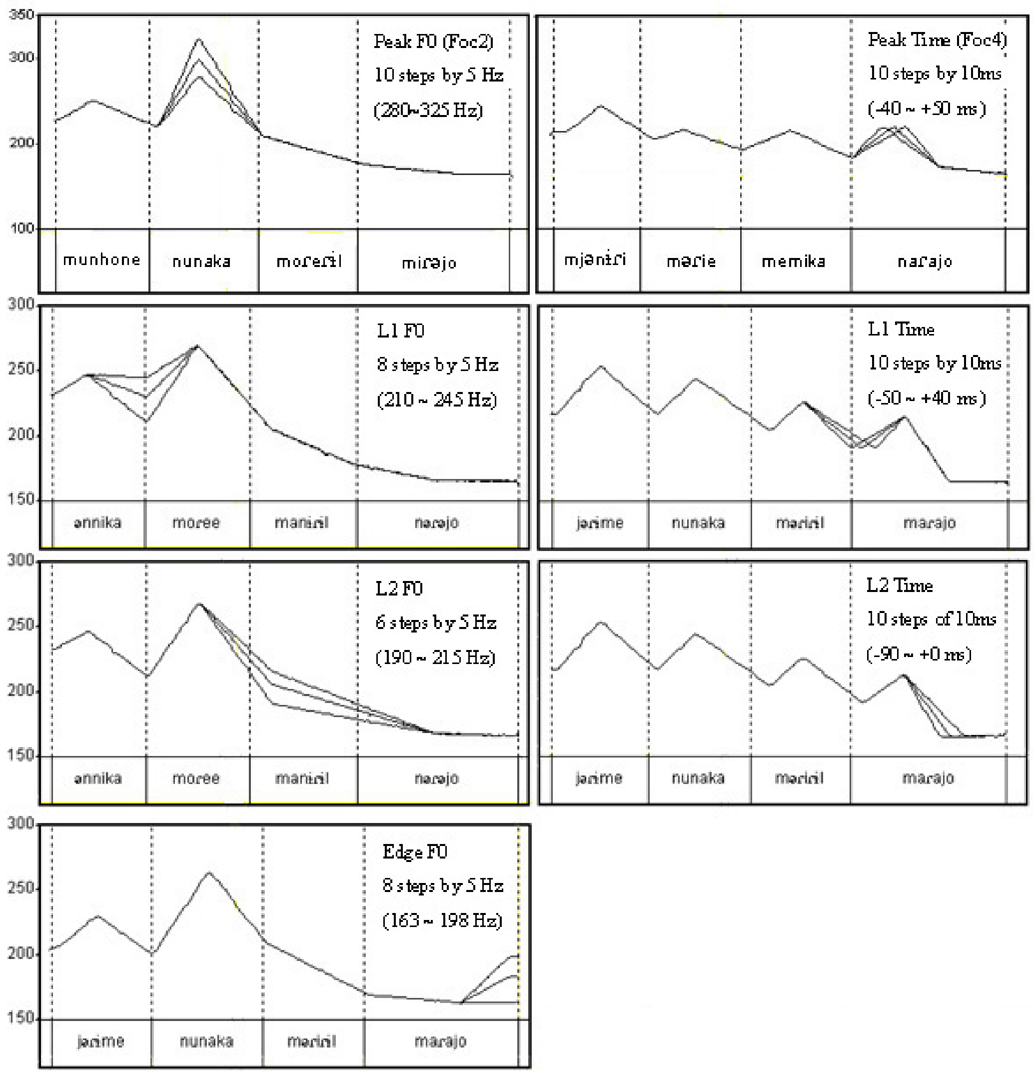

Figure 12: Manipulation of four variables: Focus2 on the left-hand panels and Focus4 on the right-hand panels

The synthetic tokens were produced as follows: 20 tokens for Peak F0 (10 steps x 2 foci), 20 tokens for Peak Time (10 steps x 2 foci), 16 tokens for L1 F0 (8 steps x 2 foci), 20 for L1 Time (10 steps x 2 foci), 12 for L2 F0 (6 steps x 2 foci), 20 for L2 Time (10 steps x 2 foci), and 16 tokens for Edge F0 (8 steps x 2 foci). The total number of the input tokens is 124.

The perception test was conducted with *E-Prime*. The stimuli were split into three E-Prime test files, and each file contained two sets. The very first file contained one more set which was an exercise set. Each set, except the exercise set, contained 18-22 tokens. The stimuli were split into several sets because it might be too tiring for participants to listen to similar-sounding tokens for too long without a break. Because the experiment could be held until a relevant key was pressed, the participants could take a break as long as they wanted between sets, although no participant took a break longer than five minutes. Finally, the three files were presented twice so the experiment consisted of 12 sets and 248 sentences.

4.1.2 Procedures

Eighteen native NK Koreans participated in the identification tasks. The experiments took place in the city of Andong, Kyeongbuk (North Kyungsang), Korea, but in several different places. The places where the test was performed were generally quiet, and all the participants wore headphones, so the outside noises did not affect the results of the experiments. My personal laptop computer was used such that the software E-Prime and all the synthetic tokens were installed in it. There were three *E-Prime Studio* files where all the tokens were contained, and each file contains two sets, except the very first file. The first file had one more set before the others, which was meant to be an exercise for the listeners. The exercise set is made up of five natural utterances, either questions or statements. This exercise was included to make the participants clearly understand what would be happening during the tasks, minimizing operational errors. The tokens were presented to the listeners in a random order within their own sets.

The tasks were forced-choice tests such that participants were to listen to each sentence and judge whether the presented sentence was a question or a statement. They were to press the key “1” for questions and “0” for statements. Each participant was told that no two sentences they would hear were identical. Each sentence was played twice and the next sentence was not played until the listener responded. Once the listener responded, the next sentence was played two seconds after the response. The participants could take as long a break as they pleased between sets. Finally, each participant completed the test twice without knowing that they were doing so, but the tokens were presented in a different random order.

The E-Prime logged several parameters including Responses (1 or 0) and Response Time. In the following section, I only discuss the results based on the Responses, regardless of Response Time. Sums of 1's are calculated and the ratios of question responses (1's) over all the responses are also calculated per sentence. I plot the results on graphs with continual steps on the X-axis and percent ratios of question response on the Y-axis. I discuss the results variable by variable.

4.2 Results

In section 3, it was shown that the female speaker produced questions and statements in significantly different ways with respect to Edge F0, Valley2 Time, Valley3 Time and Peak F0 among others, though not necessarily for both focus conditions.[5] The perception test, however, found that only Edge F0 was a perceptually salient cue for questions in the absence of other cues. Peak F0 and L2 time do, however, show a small gradient effect on the type identification. Those parameters that were not significantly different between questions and statements in section 3, also turned out to be not salient perceptually. In addition, the focus position, early or late, does not affect the type identification.

As eighteen participants completed the tests twice, the total number of responses per stimulus is 36. I calculate the ratios of question responses ("1"s) out of total 36 responses per stimulus. The threshold ratio for the sentence-type identification is assumed to be 50% in the present paper. However, strictly speaking, the threshold ratio could be lower than 50%. Because the stimuli were created based on declaratives, the question

[5] The manipulation windows in Figure 12 exhibit stylized F0 contours. Accordingly, L1 and L2 coincide with Valley2 and Valley3 for the Focus4 condition. For the Focus2 condition, only L1 coincides with Valley1. Therefore, it is possible to compare the perception results of L1 Time and L2 Time with the production results of Valley2 Time and Valley3 Time.

response ratio will be close to 0% if the modification of the base declarative sentence has no effect on perception at all. Even if the question response ratio is not as high as 50%, it does not necessarily mean that a relevant variable manipulated has no effect on the sentence-type identification. Nevertheless, if there is no systematic trend found along the F0 or temporal continuum, the effect will not be so meaningful.

Now let me discuss the results in more detail by referring to the response ratios of each variable.

4.2.1 Peak height and timing

It turns out that neither Peak F0 nor Peak Time had a strong effect on sentence-type identification. The left panel in Figure 13 displays the question response ratios for all ten steps (lowest 1 through highest 10) of Peak F0. The ratios are plotted in two groups, Early Focus (F2) and Late Focus (F4). Each point in the graph stands for the ratio of the question responses (= sum of '1's divided by 36). In both the panels of Figure 13, overall response ratios remain around 50%. The left panel of Figure 13 demonstrates a small tendency where the ratio of question responses tends to increase as focused accent peak goes higher. Still, the overall ratios are small, so Peak F0 had only a marginal effect on question identification.

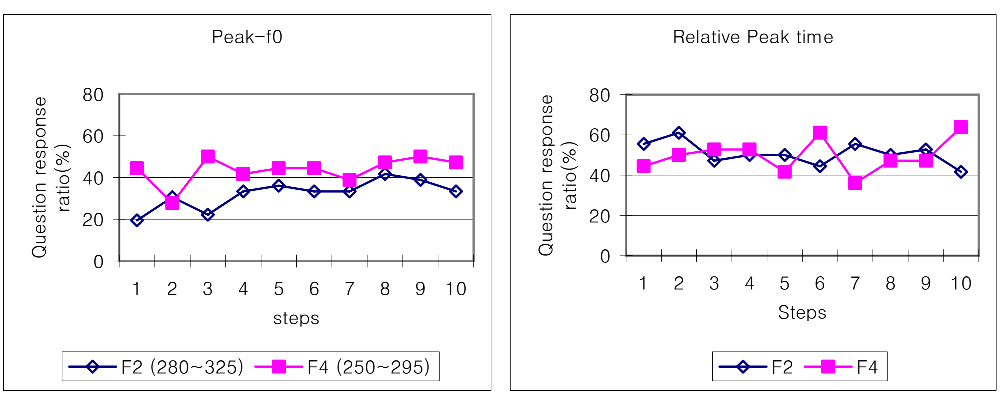

Figure 13: The ratios of question responses for Peak F0 (left panel) and Peak timing

The Late Focus condition ("F4" in the figure) triggers more question responses than the Early Focus condition ("F2"), even though the absolute peak F0 values are lower in Late Focus than Early Focus because of declination. This might have something to do with a stronger correlation between final peak height and the Edge_F0 in Focus4.[6] Alternatively, as hybrid sentences, not purely synthetic sentences with average values, were used, F0 values I chose might have happened to be beyond some kind of perceptual threshold. A more systematic study is required to resolve this focus effect.

Peak Time response ratios are plotted on the right panel in Figure 13. It shows that peak timing has no effect on sentence-type identification, although the overall ratios are slightly higher than those of Peak F0. That is, whether rightmost accent peaks occur relatively earlier or later, the sentence can still be either a question or a statement. Moreover, focus condition does not result in any difference in responses, as focus groups do not show any patterns.

I conclude that neither peak height nor peak time has as strong effect on question identification. Peak height (Peak F0) was significantly different for questions versus statements in the acoustic analyses, but it has only a small gradient effect on the sentence-type identification. Therefore, Peak F0 may not be a primary acoustic correlate of a question.

4.2.2 L1 F0 and timing

L1 F0 and L1 time did not have a strong effect on sentence-type identification. The left panel of Figure 14 shows the ratios of L1 F0 and the right panel L1 timing. The overall ratios are around 50% without any systematic tendency. Further, different focus conditions did not affect the identification systematically. Therefore, it can be concluded that L1 F0 and timing are irrelevant, when it comes to question identification.

---

[6] Bivariate Pearson Correlation tests were run on the female speaker's data from the production experiment, and Peak F0 is significantly correlated with Edge F0 in Focus4 [sig. (2-tailed) < 0.001], but not in Focus2 [sig. (2-tailed) = 0.716], for questions.

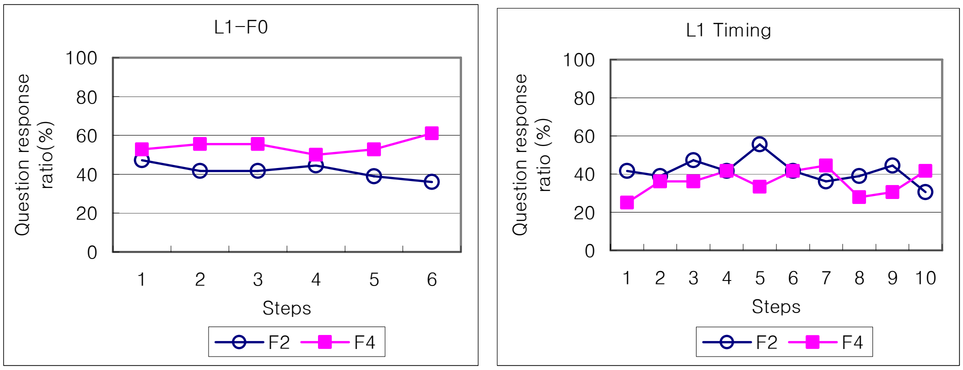

Figure 14: The ratios of question responses for L1 F0 (left panel) and L1 timing

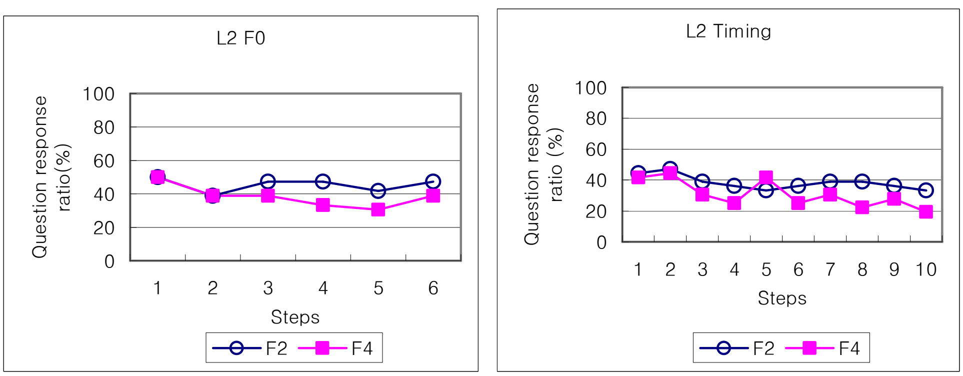

Figure 15: The ratio of question responses for L2 F0 (left panel) and timing

4.2.3 L2 height and L2 timing

L2 height and timing are not strong perceptual cues, either, for the overall ratios of question stay around 50 percent. However, they seem to have a small gradient effect on sentence-type identification. The graphs in Figure 15 demonstrate that higher L2 F0 and later occurrence of L2 triggers fewer question responses. The L2 F0 results are odd because penultimate low targets were higher for questions than statements in the acoustic analyses. The contradictory result may be interpreted as L2 F0 interacting with Edge F0 such that L2 F0 may be lower in questions when Edge is low-rising, whereas it may be higher in questions when Edge is high-falling.[7]

[7] I have not discussed what F0 contours patterns are there for questions, but simply compared Edge F0 as a static target. Impressionistically speaking, NK Korean questions are of two types: low rise and high fall. F0 patterns and the interaction between L2 F0 and Edge F0 are to be studied more systematically.

Early Focus seems to induce slightly higher ratios of question responses, which can be interpreted that a post-accent valley closer to the end of a sentence is more closely related to sentence-type identification. Remember that L2 under Focus4 is the penultimate low targets whereas L2 under Focus2 is the antepenultimate low target. Therefore, this result conforms to the production result in that most significant acoustic cues were located toward the end and, crucially, earlier occurrence and lower F0 of the rightmost valleys were marginal characteristics of questions.

4.2.4 Edge F0

Utterance-final F0 had a strong effect on sentence-type identification. As shown in Figure 16, the overall ratios of question responses are higher than 50% at and above the second step for Focus4 (F4) and above the third step for Focus2 (F2). In other words, if the final F0 is 10 Hz higher than the onset of the final syllable (step 2), the sentence is heard as question more than 50 times out of 100 in Late Focus. The Early Focus needs higher F0 at the edge in order to reach more than 50% in the question ratio. Step 3, which means that sentence-final F0 is 15 Hz higher than that of the final vowel onset, begins to be heard as questions more than 50 times out of 100.

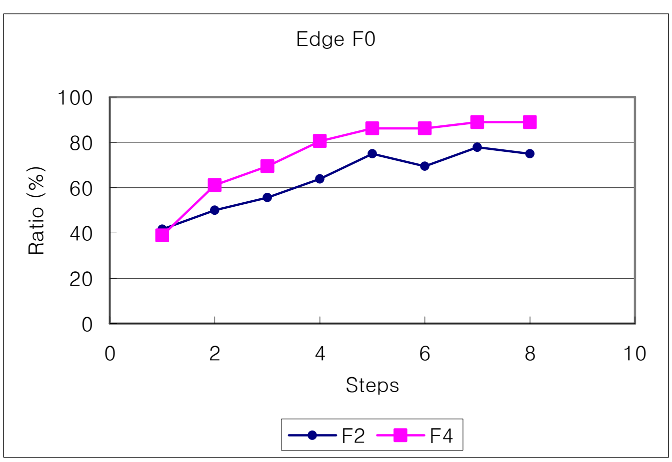

Figure 16: The ratio of question responses for edge F0: Early Focus (left) and Late Focus

The ratio becomes quite constant from step 5, though the ratios are still lower than 90%. That is, if Edge F0 reaches 25 Hz higher than the onset of the sentence-final syllable, the perception of questions becomes quite stable, and the edge F0 which is more than 25 Hz higher does not help increase the question ratio any more. If more steps had been included that are lower in Edge F0 than the onset of final vowels, it might have been possible to get a typical S-shape graph for the response ratio, but the other half of the S-shape does not appear because the lowest value of the continuum was the same as the vowel onset such that F0 was flat across the final syllable.

Recall that Late Focus had greater differences in Edge F0 between questions and statements than Early Focus, especially for the female speaker. The female speaker produced average 30 Hz higher pitch for questions in Late Focus, but she produced average 10 Hz higher for questions in Early Focus. This predicts that listeners should be more sensitive to F0 in Early Focus in order to distinguish questions from statements. However, my perception tests do not support the prediction. A small rise from the onset of the final syllable can be heard as a question in Late Focus, but a larger rise is required for a stable perception of questions in Early Focus. It means that acoustically more ambiguous questions are not compensated for by perceptual sensitivity.

In summary, the results of my perception experiment roughly correspond to the results of production experiment. A subset of the acoustic cues with statistic significance has turned out to be independently perceptually salient, and the acoustic cues without statistic significance did not show any perceptual salience, either. As in the production experiment, Edge F0 was the most consistent and reliable cue for sentence-type identification. Although the ratios are only around 50%, Peak F0 has a positive correlation with questions and L2 F0 and L2 Time have negative correlation with questions. L2 in the perception tests correspond to penultimate F0 target and antepenultimate F0 target in late and Early Focus respectively. Therefore, the fact that higher F0 and later occurrence of L2 induce fewer question responses is in agreement with the finding from production experiment in that rightmost F0 valleys toward the end and the sentence-final edge are the most important acoustic cues to differentiate questions from statements.

5. Discussion

This study shows that the NK Korean question intonation has 'universal' property, i.e. high or rising intonation at the sentence boundary. However, the rise at the end is not so salient and it only slightly rises from the onset of the final syllable, or remains flat at a higher level (Figure 3). The shallow final rise for most questions can be interpreted as language-specific properties of NK Korean; the extent of rising at the end is relatively small but the small differences are large enough for native listeners to distinguish between questions and statements. I hypothesize that the shallow final rise is constrained by a kind of phonological template which keeps intonation from rising dramatically. For instance, a Tokyo Japanese question ends with L% followed by (L)H% (Venditti 1997) on the final syllable. Due to this L% preceding the H%, F0 keeps falling from the peak up to the last syllable, and then rises back. Alternatively, it could be simply a language-specific phonetic implementation. Further studies are needed to identify the source of the language-specificity of the NK Korean question intonation.

The production experiment has shown that sentence-final pitch (Edge F0), and the relative timing and height of the rightmost F0 valleys are common acoustic properties to distinguish questions from statements in NK Korean. That is, questions bear higher edge F0, lower and earlier realization of F0 valleys. The two speakers who participated in this experiment did not share many strong acoustic cues. The female speaker seems to show statistically significant differences in more variables than the male speaker. Peak F0 and pitch range, which are often considered as interrogative acoustic cues, are not found to be significantly different for both the speakers.

The perceptual identification tests also demonstrate that edge F0, or boundary pitch, is a perceptually important cue to distinguish sentence types. Higher F0 at the right edge is more likely to be perceived as questions. Further, Edge F0 seems to show a categorical distinction between questions and statements. Besides, peak F0, L2 F0 and L2 timing seem to show a gradient effect on identification although the ratios of question responses remain around or lower than 50%. Peak F0 has a positive correlation with the question response ratio whereas F0 and time of L2, i.e. penultimate or antepenultimate F0 target for Early and Late Focus respectively, have negative correlations with questions.

This paper did not investigate whether there are non-pitch acoustic cues to mark sentence types, although those cues might have played a role in acoustic and perception results. Klatt and Klatt (1990) found evidence that voice quality plays a role in signaling sentence boundaries. According to them, declarative sentences often have breathy or creaky voice at the sentence boundaries (Klatt and Klatt 1990; Choi, Hasegawa-Johnson, and Cole 2005). If we assume that declarative sentences are unmarked while interrogative sentences are marked, the production of question intonation may require extra energy at the sentence boundary, which entails greater intensity at the boundary for questions than for statements. Such extra energy, presumably required for questions, may result in less frequent occurrences of breathy or creaky voices at the question boundaries as well. A more systematic experiment is needed to prove whether it is indeed the case in NK Korean as well.

My experiments are limited for the following aspects. First, I recorded only two speakers so that the acoustic and statistic analyses may not represent the whole population. Second, when I created synthetic tokens for my perceptual experiments, I used natural speech so that all possible factors, e.g. intensity and duration, were not controlled. Third, my listeners did not take a preliminary test to make sure that they clearly understood what 'questions' and 'statements' mean during the perception experiments. It was assumed that the concepts were straightforward.

NK Korean has a variety of type-specific final particles as well as type-neutral particles. The results of the comparison of yes/no- versus wh-questions are not included in this paper, because the types of sentence-final particles were not controlled. Specifically, the type-neutral particle –*yo* was used for statements vs. (y/n) questions, but the type-specific particles –*na* and –*no* were used for y/n questions and wh questions respectively. It turned out that there was no consistent acoustic cue to differentiate between the two types of questions. A hypothesis is that the type-specific particles might have overruled differential intonations. Therefore, it will be interesting to investigate the correlation between sentence-final particles and intonation.

.

References

Beckman, M. E. (1986). *Stress and Non-Stress Accent* (Netherlands Phonetic Archives No. 7). Foris.

Beckman, M. E. and Ayers, G. M. (1994). Guidelines for ToBI Labelling [WWW page]. URL http://www.ling.ohio-state.edu/phonetics/E_ToBI.

Chan, M., Peng, S., Tseng, C., Huang, T., Lee, O. J., and Beckman, M. E. (2005). Towards a Pan-Mandarin system for prosodic transcription. In S.-A. Jun (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford, UK: Oxford University Press.

Choi, J.-Y., Hasegawa-Johnson, M., and Cole, J. (2005). Finding intonational boundaries using acoustic cues related to the voice source. *Journal of Acoustic Society of America*, 118 (4), 2579-2587.

Cruttenden, A. (1981). Falls and rises: meanings and universals. *Journal of Linguistics*, 17.

D'Imperio, M. (2000). *The Role of Perception in Defining Tonal Targets and their Alignment*. Dissertation of Ohio State University.

Fox, A. (2000). *Prosodic Features and Prosodic Structure: The phonology of suprasegmentals*. Oxford University Press.

Gussenhoven, C. and Chen, A. (2000). Universal and Language-Specific effects in the perception of question intonation. *Proceedings of ICLSP 2000*.

Idsardi, W. and Chang W. (2003). Markedness and the Phonetic Implementation of Tone in North Kyungsang Korean. In P. M. Nowak et al (eds.), *Proceedings of the Twenty-Ninth Annual Meeting of the Berkeley Linguistics Society*.

Idsardi, W. and Kim, S.-H. (1997). The Metrical Tone in North Kyungsang Korean. In S. Kuno et al.(eds.), *Harvard Studies in Korean Linguistics*, 7.

Jun, J., Kim, J., Lee, H., and Jun, S.A. (2004). Prosodic structure of North Kyungsang Korean. *The Proceedings of Speech Prosody 2004*, 337-340.

Jun, S.-A. (2000). K-ToBI (Korean ToBI) labelling conventions: Version 3. *Speech Sciences*, 7, 143-169.

Jun, S.-A. (2005). *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford University Press

Kenstowicz M. and Sohn, H.-S. (1997). Phrasing and Focus in Northern Kyungsang Korean. *MIT Working Papers in Linguistics*, 30, 25-47.

Kenstowicz, M. and Sohn, H.-S. (2001). Accentual Adaptation in North Kyungsang Korean. In M. Kenstowicz (ed.), *Ken Hale*. The MIT Press.

Kim, N.-J. (1997). *The Optimality Theoretic Account of Tones, Segments, and their Interaction in North Kyungsang Korean*. Dissertation of Ohio State University.

Kirsner, R. S. and van Heuven, V. J. (1996). Boundary tones and the semantics of the Dutch final particles hé, hoor, zeg and joh. *Linguistics in the Netherlands*, 13, 133-46.

Klatt, D. H. and Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of Acoustic Society of America*, 87, 820-857.

Ladd, R. (1981). On Intonational Universals. In T. Myers et al. (eds.), *The Cognitive Representation of Speech*. Amsterdam: North Holland Publishing.

Makarova, V. (2001). Perceptual Correlates of Sentence-type Intonation in Russian and Japanese. *Journal of Phonetics*, 29, 137-154.

Ohala, J. (1983). Cross-language use of pitch: an ethological view. *Phonetica*, 40.

Sohn, H.-S. (2004). Wh-operator and phonological phrasing in North Kyungsang Korean. *Studies in Phonetics, Phonology and Morphology*, 10(2), 293-325.

Suh, C.-M. (1987). *Kwuke Uymwunmwun Yenkwu* (A study on the interrogative sentences in Korean). Tower Press: Seoul, Korea.

van Heuven, V. J. and Haan, J. (2000). When and how do we hear whether a Dutch speech utterance is a statement or a declarative question? *Linguistics in the Netherlands*. John Benjamins Publishing Company.

van Heuven, V. J. and Haan. J. (2002). Temporal distribution of interrogativity markers in Dutch: A perception study. In C. Gussenhoven and N. Warner (eds.), *Laboratory Phonology* 7.

Venditti, J. (1997). Japanese ToBI Labelling Guidelines. Papers from the Linguistics Laboratory. *Ohio State University Working Papers in Linguistics*, 50, 127-162

Winer, B. J., Brown, D. R., and Michels, K. M. (1991). *Statistical Principles in Experimental Design*. New York: McGraw-Hill.

Yuan, J. (2004). *Intonation in Mandarin Chinese: Acoustics, Perception, and Computational Modeling*. Dissertation of Cornell University.

Yuan, J. and Shih, C. (2002). Confusability of Chinese Intonation. *Proceedings of Speech Prosody 2004*. Nara, Japan.

Yuan, J., Shih, C., Kochanski, G. P. (2002). Comparison of Declarative and Interrogative Intonation in Chinese. *Proceedings of Speech Prosody 2002*. Aix-en-Provence, France.

Appendix A. Recording Materials for the Production Experiment

Statements – Early Focus

Mwunho-ney	**nwuna-ka**	molay-lul	mile-yo.	(Mwunho-ne	hyeng-i	aniko)
Monho-Gen	sister-Nom	sand-Acc	push-Part.	(Mwunho-Poss	brother-Nom	not)

‘Munho’s sister pushes sand.’ (Not Mwunho’s brother)

Yelum-ey	**nwuna-ka**	meli-lul	mal-ayo.	(Emma-ka	aniko)	
Summer-in	sister-Nom	hair-Acc	roll-Part.	(Mom-Nom	not)	

‘In summer (my) sister has her hair rolled.’ (Not my mom)

Yengtek-uy	**mwunho-ka**	nolay-lul	mul-eyo.	(Yengtek-uy	unhi-ka	aniko)
Yengtek-Gen	Mwunho-Nom	song-Acc	ask-Part.	(Yengtek-Gen	Unhi-Nom	not)

‘Yengtek’s Mwunho asks about a song.’ (Not Yengtek’s Unhee)

Enni-ka	**molay-ey**	manul-ul	nel-eyo.	(mengsek-ey	nel-ci	anhko)
Sister-Nom	sand-on	garlic-Acc	dry-Part.	(mat-on	dry-ing	not)

‘Sister dries garlic on sand.’ (Not drying it on a mat)

Myenuli	**meli-ey**	maymi-ka	nal-ayo.	(meynuli	elkwul-i	aniko)
Daughter-in-law	head-over	locust-Nom	fly-Part.	(Daughter-in-law	face-Nom	not)

‘A locust flies over Daughter-in-law’s head.’ (Not around her face)

Statements – Late Focus

Mwunho-ney	nwuna-ka	molay-lul	**mil-eyo**.	(molay-lul	phwu-ci	anhko)
Monho-Gen	sister-Nom	sand-Acc	push-Part.	(sand-Acc	dig-ging	not)

‘Munho’s sister pushes sand.’ (Not digging up sand)

Yelum-ey	nuna-ka	meli-lul	**mal-ayo**.	(meli-lul	phje-ci	anhko)
Summer-in	sister-Nom	hair-Acc	roll-Part.	(hair-Acc	straighten-ing	not)

‘In summer (my) sister has her hair rolled.’ (Not straightening hair)

Yengtek-uy	mwunho-ka	nolay-lul	**mwul-eyo**.	(nolay-lul	pwulu-ci	anhko)
Yengtek-Gen	Mwunho-Nom	song-Acc	ask-Part.	(song-Acc	sing-ing	not)

‘Yengtek’s Mwunho asks about a song.’ (Not singing a song)

Enni-ka	molay-ey	manul-ul	**nel-eyo**.	(manul-ul	peli-ci	anhko)
Sister-Nom	sand-on	garlic-Acc	dry-Part	(garlic-Acc	dispose-ing	not)

‘Sister dries garlic on sand.’ (Not throwing it away)

Myenuli	meli-ey	maymi-ka	**nal-ayo**.	(maymi-ka	ancaiss-ci	aniko)
Daughter-in-law	head-over	locust-Nom	fly-Part.	(locust-Nom	sit-ting	not)

‘A locust flies over Daughter-in-law’s head.’ (Not sitting on it)

Questions – Early Focus

Mwunho-ney	**nwunaka**	molay-lul	mil-eyo?	(Mwunho-ney	hyeng-i	aniko)
Monho-Gen	sister-Nom	sand-Acc	push-Part.	(Mwunho's	brother-Nom	not)

'Munho's sister pushes sand?' (Not Mwunho's brother)

Yelum-ey	**nwuna-ka**	meli-lul	mal-ayo?	(Emma-ka	aniko)	
Summer-in	sister-Nom	hair-Acc	roll-Part.	(Mom-Nom	not)	

'In summer (my) sister has her hair rolled?' (Not my mom)

Yengtek-uy	**mwunho-ka**	nolay-lul	mul-eyo?	(Yengtek-uy	unhi-ka	aniko)
Yengtek-Gen	Mwunho-Nom	song-Acc	ask-Part.	(Yengtek-Gen	Unhi-Nom	not)

'Yengtek's Mwunho asks about a song?' (Not Yengtek's Unhee)

Enni-ka	**molay-ey**	manul-ul	nel-eyo?	(mengsek-ey	nel-ci	anhko)
Sister-Nom	sand-on	garlic-Acc	dry-Part.	(mat-on	dry-ing	not)

'Sister dries garlic on sand?' (Not drying it on a mat)

Myenuli	**meli-ey**	maymi-ka	nal-ayo?	(meynuli	elkwul-i	aniko)
Daughter-in-law	head-over	locust-Nom	fly-Part.	(Daughter-in-law	face-Nom	not)

'A locust flies over Daughter-in-law's head?' (Not around her face)

Questions – Late Focus

Mwunho-ney	nwunaka	molaylul	**mil-eyo**?	(molay-lul	phwu-ci	anhko)
Monho-Gen	sister-Nom	sand-Acc	push-Part.	(sand-Acc	dig-ging	not)

'Munho's sister pushes sand?' (Not digging up sand)

Yelum-ey	nwuna-ka	meli-lul	**mal-ayo**?	(meli-lul	phje-ci	aniko)
Summer-in	sister-Nom	hair-Acc	roll-Part.	(hair-Acc	straighten-ing	not)

'In summer (my) sister has her hair rolled?' (Not straightening hair)

Yengtek-uy	mwunho-ka	nolay-lul	**mwul-eyo**?	(nolay-lul	pwul-ci	anhko)
Yengtek-Gen	Mwunho-Nom	song-Acc	ask-Part.	(song-Acc	sing-ing	not)

'Yengtek's Mwunho asks about a song?' (Not singing a song)

Enni-ka	molay-ey	manul-ul	**nel-eyo**?	(manul-ul	peli-ci	anhko)
Sister-Nom	sand-on	garlic-Acc	dry-Part.	(garlic-Acc	dispose-ing	not)

'Sister dries garlic on sand?' (Not throwing it away)

Myenuli	meli-ey	maymi-ka	**nal-ayo**?	(maymi-ka	ancaiss-ci	anhko)
Daughter-in-law	head-over	locust-Nom	fly-Part.	(locust-Nom	sit-ting	not)

'A locust flies over Daughter-in-law's head?' (Not sitting on it)

Appendix B. Statistics - Correlations

Speaker: Female		Edge F0	Late Trend F0	Valley2 Time	Valley3 Time
Edge F0	Pearson Correlation	1	-.439**	-.224	-.468**
	Sig. (2-tailed)		.001	.097	.000
	N	58	56	56	56
Late Trend F0	Pearson Correlation	-.439**	1	.276*	.360**
	Sig. (2-tailed)	.001		.039	.006
	N	56	56	56	56
Valley2 Time	Pearson Correlation	-.224	.276*	1	.736**
	Sig. (2-tailed)	.097	.039		.000
	N	56	56	56	56
Valley3 Time	Pearson Correlation	-.468**	.360**	.736**	1
	Sig. (2-tailed)	.000	.006	.000	
	N	56	56	56	56

Speaker: Male		Edge F0	Late Trend F0	Valley2 Time	Valley3 Time
Edge F0	Pearson Correlation	1	-.382*	-.447**	-.393*
	Sig. (2-tailed)		.021	.006	.018
	N	36	36	36	36
Late Trend F0	Pearson Correlation	-.382*	1	.153	.261
	Sig. (2-tailed)	.021		.332	.095
	N	36	42	42	42
Valley2 Time	Pearson Correlation	-.447**	.153	1	.883**
	Sig. (2-tailed)	.006	.332		.000
	N	36	42	42	42
Valley2 Time	Pearson Correlation	-.393*	.261	.883**	1
	Sig. (2-tailed)	.018	.095	.000	
	N	36	42	42	42

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Appendix C. Simple Main Effects (Yuan 2005) – SentType * Focus position

Pairwise Comparisons – Edge F0

SentType	(I) Focus position	(J) Focus position	Mean Difference (I-J)	Std. Error	Sig[a]	95% Confidence Interval for Difference(a)	
						Lower Bound	Upper Bound
Q	Focus2	Focus4	-11.620*	2.850	.000	-17.267	-5.974
	Focus4	Focus2	11.620*	2.850	.000	5.974	17.267
S	Focus2	Focus4	3.726	2.801	.186	-1.822	9.275
	Focus4	Focus2	-3.726	2.801	.186	-9.275	1.822

Pairwise Comparisons – Pitch Range

SentType	(I) Focus position	(J) Focus position	Mean Difference (I-J)	Std. Error	Sig. [a]	95% Confidence Interval for Difference(a)	
						Lower Bound	Upper Bound
Q	Focus2	Focus4	29.170*	2.170	.000	24.871	33.469
S	Focus2	Focus4	31.481*	2.170	.000	27.182	35.780

Pairwise Comparisons - Peak Time

SentType	(I) Focus position	(J) Focus position	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference(a)	
						Lower Bound	Upper Bound
Q	Focus2	Focus4	16.300*	5.093	.002	6.213	26.387
S	Focus2	Focus4	20.700*	5.093	.000	10.613	30.787

Pairwise Comparisons – Peak F0

SentType	(I) Focus position	(J) Focus position	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference(a)	
						Lower Bound	Upper Bound
Q	Focus2	Focus4	49.744(*)	2.575	.000	44.644	54.843
S	Focus2	Focus4	53.609(*)	2.575	.000	48.509	58.708

Based on estimated marginal means
* The mean difference is significant at the .05 level.
[a] Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).