말소리와 음성과학 제3권 제3호 (2011)
pp. 69~75

Phrase positional effects on F0 peak timing in Tokyo Japanese

Cho, Hyesun[1)]

ABSTRACT

This paper investigates phrase positional effects on the timing of F0 (pitch) peaks in Tokyo Japanese disyllabic words with varying accent type (HL or LH) and phrase position (final or non final). The F0 peak timing was normalized by the total word duration ('normalized H timing'). The normalized H timing was significantly affected by accent type and phrase position. The H timing was later in the LH accent type than in the HL accent type, and in non final positions than in final positions. In addition, to examine the validity of the quantitative results, different models of phrase position effects were compared by measuring H timing in two approaches: normalization versus relative distance measures. For the normalization measures, the H timing was measured as the time of the F0 peak divided by the total word duration or by the duration of the tone bearing syllable. For the relative distance measures, the H timing was measured as the distance in milliseconds from the end of the word or from the end of the associated syllable. The best model was the normalization by the total word duration, rather than by the duration of the tone bearing syllable. This means that phrase positional effects on the timing of F0 peaks in Japanese disyllabic words are best modeled in terms of proportion of the total word duration.

Keywords: F0 timing, Japanese, peak retraction, phrase position

1. Introduction

This paper investigates phrase positional effects on the timing of F0 (pitch) peak in disyllabic accented words in Tokyo Japanese. It has been observed that the timing of F0 peak is retracted (occurs earlier) in phrase final positions compared with phrase medial positions, due to the presence of the boundary tone at the phrase final positions (English (Steele 1986, Silverman and Pierrehumbert 1990), Mexican Spanish (Prieto et al. 1995), Greek (Arvaniti and Ladd 1995), Kinyarwanda (Myers 2003)). This paper will demonstrate that (i) Japanese also exhibits phrase positional effects on F0 peak timing, and that (ii) the retraction is best modeled in terms of a proportional measure relative to the total word duration, rather than relative to the associated syllable duration.

Japanese is a lexical pitch accent language, where pitch accent is lexically contrastive. Japanese pitch patterns are characterized by a L boundary tone on the initial mora unless the initial mora is accented (McCawley, 1968; Pierrehumbert and Beckman 1988). The L boundary tone is followed by a H phrase tone, and there is an accentual H tone on the accented mora if the word is accented (e.g. monomorai 'sty' has a LHHLL tonal pattern where the first H is a boundary tone and the second H is accentual). Disyllabic words can have two accent types: HL and LH. The examples in (1) show pitch accent patterns for disyllabic accented words: the words in (1a) are initial accented (HL), those in (1b) are second mora accented (LH).

(1)	a. HL		b. LH	
	hashi	'chopstick'	hashi	'bridge'
	ima	'now'	ima	'living room'
	mochi	'birdlime'	mochi	'durability'

Phonologically, the H tone in the HL accent type in (1a) is associated with the first mora of the word, and the H tone in the LH accent type in (1b) is associated with the second mora. Actual F0 peaks, however, do not necessarily occur within the mora where the H tone is phonologically associated. It is well known that the pitch peak of the initial H tone in a HL word is often found after the end of the accented (initial) mora. This is

1) Seoul National University, chohazel@gmail.com

Received: February 1, 2011
Revised: August 2, 2011
Accepted: August 26, 2011

known as late peak or ososagari (Sugito 1982, Hasegawa and Hata 1995, Ishihara 2006). A plausible explanation for this phenomenon would be that minimal time is required to reach a F0 peak from the beginning of the word. Autosegmental Phonology (Goldsmith 1976) is not sufficient to explain the discrepancy between the actual location of F0 peaks and the syllabic position of the underlying H tones. A phonological approach is to specify the 'secondary association' for a tone, which links the tone to the syllable where the F0 peak is actually found (Pierrehumbert and Beckman 1988:128, Grice et al.

In experimental studies, it has been discovered that F0 peak timing is affected by various factors, such as articulatory inertia (Xu, 2001, Xu, 2005:218), speech rate and tonal crowding (Caspers and van Heuven, 1993), phrase position (Arvaniti and Ladd 1995, Silverman and Pierrehumbert 1990, Myers, 2003). According to Xu (2001), the F0 peak of the level high (H) tone in Mandarin usually stays within the associated syllable, but the F0 peak of the rising (LH) tone is often found immediately after the offset of the end of the associated syllable. This is due to articulatory inertia that causes the larynx to terminate the sharp rise after the offset of the syllable. Distance to adjacent tones also affects tonal timing as well. That is, the F0 peak of a rising pitch accent tends to occur earlier if the following tone is adjacent (Dutch (Caspers and van Heuven 1993), English (Silverman and Pierrehumbert 1990), Spanish (Prieto et al. 1995)). For example, in the accent lending pitch rise in Dutch, the F0 peak of a rising pitch accent tends to occur earlier if the following fall is nearer (Caspers and van Heuven 1993: 169). In English, the F0 peak of an accented syllable tends to occur earlier if the following syllable is accented (Silverman and Pierrerhumbert 1990).

In many languages, phrase position significantly affects the tone timing, which is the concern of the present paper. In English, Steele (1986) showed that an F0 peak occurs earlier in a phrase final accented syllable than in a phrase medial accented syllable. Silverman and Pierrehumbert (1990) showed that in English the F0 peak of the phrase final nuclear accent comes earlier than the F0 peak of the phrase medial prenuclear accent. Also, in Spanish, an F0 peak tends to occur earlier in a phrase final word than in a phrase medial word (Prieto et al. 1995).

Myers (2003) examined the tonal timing in Kinyarwanda, a Bantu tone language spoken in Rwanda, by varying three factors: position in phrase (phrase final, non phrase final word), position in word (word final, or non word final syllable), and tone context (distance to the following H tone (including absence of the following H tone)). Kinyarwanda has three accent types: H, HL, and LH. The HL and LH tone types only occur within one syllable with a long vowel (e.g. HL: u**mwáa**na 'child', LH: u**mwaá**mi 'king', Myers 2003:72). The two accent types differ only in the timing of the F0 peak of the H tone. He measured 'relative peak delay', which means 'peak delay' (the interval from the beginning of the syllable to the F0 peak in milliseconds) divided by the duration of the tone bearing syllable. In both HL and LH accent types, non phrase final positions showed greater relative peak delay than phrase final positions for all speakers. In addition, he found that relative peak delay was greater for the LH accent type than for the HL accent type.

The present study examines the effects of phrasal position on the timing of F0 peaks corresponding to the H tones in Tokyo Japanese disyllabic words. As in Kinyarwanda, Japanese has contrasts between the HL and LH accent types, but a difference is that in Japanese, the two tones are realized over two syllables, instead of one. The H tone can be either immediately adjacent to the word boundary (the LH type) or one mora away from the word boundary (the HL type), so we can assess the magnitude of the effect of the distance to the word boundary and compare it with the result of Kinyarwanda.

2. Experiment

We tested the hypothesis that the F0 peak would occur earlier in phrase final position than in phrase medial (or non final) position. The speech materials consisted of 11 HL disyllabic words and 11 LH disyllabic words (See Appendix for the full list). The syllable structure was CVCV or VCV (i.e. the syllables were all monomoraic). Many of them were minimal pairs that have the same segmental materials but differ only in tonal patterns (e.g. ima (HL) 'now' vs. ima (LH) 'living room'). 6 out of the 11 HL words and 8 out of the 11 LH words consisted entirely of sonorants to minimize pitch tracking errors.

The words in isolation served as the phrase final context (F). The target words were embedded in a carrier sentence, which provided the non phrase final context (NF). The carrier sentences are as shown in (2). The sample sentences are shown in (3).

(2) a. watashiwa ________to itta.
I top _______comp said. "I said ___"
b. watashiwa ______ mo mitta.
I top ______ also saw. "I also saw ___"

(3) a. watashiwa umi to itta.
I top sea comp said. "I said, 'sea'"

b. watahiwa nana mo mitta.
I top seven also saw. "I also saw 'seven'"

Four native speakers of Tokyo Japanese, two male (J3, J4) and two female (J1, J2) were recorded reading the speech materials. The speakers were naive to the purpose of the experiment. J2 and J3 were recorded first with the carrier phrase in (2a). However, in (2a), the particle following the target phrase ('to') starts with an obstruent, so the pitch track during the post accentual consonant was invisible and the surrounding vowels are subject to perturbation due to segmental effects. Because of this, for J1 and J4, the carrier phrase was replaced with (2b), with the particle following the target word starting with a sonorant consonant so that continuous pitch tracks can be obtained. J4 read the non final condition sentences with both 'to' (2a) and 'mo' (2b).

The speech materials were written in Japanese transcriptions. Chinese letters were also provided wherever Japanese transcriptions were ambiguous. The sentences were presented on a sheet of paper. The speakers were asked to read the total of 22 words in isolation (F) and in the carrier sentence (NF), twice each. For the F condition, each target word was numbered, and speakers were asked to read the numbers as well, in order to put a sufficient pause between the words. Thus, 88 tokens were recorded for each speaker (22F×2+22NF×2=88), plus an additional set of NF sentences with a different particle for J4 (22NF×2=44), so the total number of tokens recorded were 396 (=88×4+44).

For two speakers (J3, J4), the recordings were made in a sound attenuated recording booth in the phonetics lab at the MIT Linguistics Department. The speakers wore a head mounted microphone (Shure SM10A), which was connected to the USBPre sound input device. Speech signals were directly digitized using Amadeus II (version 3.8.4, by HairerSoft) installed on an iMac, at a sampling rate of 44.1kHz, with a sampling size 16 bit. Speaker J1 was recorded in the same recording booth, but on a laptop (Mac OS X version 10.5.8) with a built in microphone using Praat (Boersma and Weenink, 2011). Speaker J2 was recorded on a digital recorder (Roland R 09) with a built in microphone, with the same sampling rate and sampling size as the other speakers.

Segment boundaries were manually labeled using Praat, and the timing of each segmental boundary was automatically collected using a Praat script. The pitch contours were smoothed by using a Praat smoothing function, and the timing and pitch level of F0 maxima (H) were measured. F0 maxima were collected only if they can be reliably located. It was difficult to reliably determine the location of F0 maxima near sibilant sounds or in glottalized vowels, so such tokens were discarded. After these procedures, the total number of tokens that were analyzed in the following section was 269.

3. Results

3.1 The effect of accent type

We first examined the effect of accent type on F0 peak timing. The H timing was measured as the proportion of the time of an F0 peak from the word onset to the total word duration ('normalized H timing'). This is a measure similar to Myers (2003) for Kinyarwanda in that the H timing is proportional rather than absolute, but different in that it is relative to the total word duration (i.e. two syllables) whereas in Myers (2003) it is relative to the duration of one syllable bearing the accent. In Section 3.3, we will show this measure of H timing yields the model that best fits the data for our Japanese data. Figure 1 shows the normalized H timing for the HL and LH accent types for each speaker. As expected, the F0 peaks occur earlier in the HL accent than in the LH accent for all speakers.

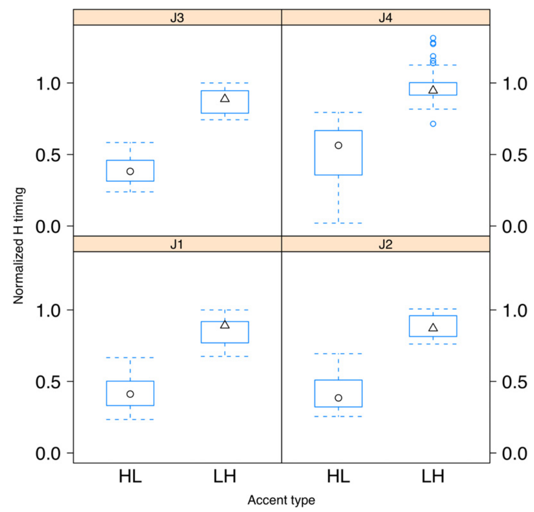

Figure 1. The normalized H timing for HL and LH accent types for each speaker. Normalized H timing = the time of H from the word onset divided by the total word duration.

For quantitative analyses, we fitted linear mixed effects models (LME) to the data, using the *lmer* function in the *lme4* package in the R software. Mixed effects models are appropriate when pooling across speakers because speakers can be treated as randomly chosen from a population. The dependent variable was normalized H, the fixed effect was accent type, and by speaker random intercepts were allowed. A log likelihood ratio test (LRT) showed that this model was better than the model without the accent type fixed effect [$\chi^2(1)=402.32$, $p<0.0001$]. Adding to this model by speaker random slopes for accent type was not significant [$\chi^2(2)=0.265$, $p=0.875$].

3.2 The effect of phrase position

Phrase position significantly affects the F0 peak timing. Figure 2 plots normalized H timing against phrase position (F: final, NF: non final), for both accent types together. Table 1 shows the mean of normalized H by speaker, accent type, and phrase position. We can see that normalized H timing is earlier in phrase final position than in non phrase final position across accent types and speakers.

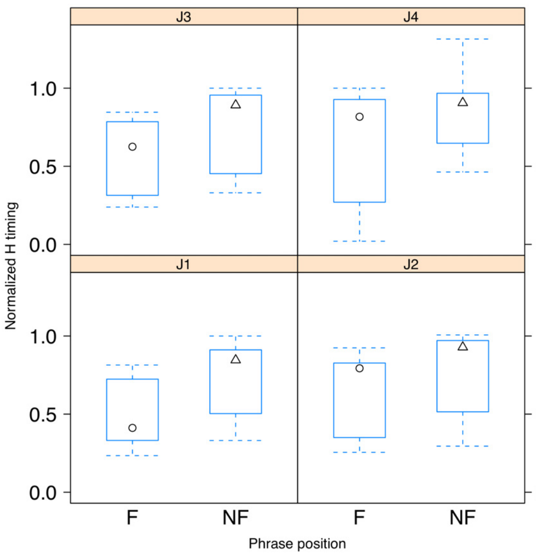

Figure 2. Normalized H timing for final (F) and non-final (NF) conditions for each speaker

Linear mixed effects models were fitted to the data to quantify the effect of phrase position on normalized H timing. The dependent variable was normalized H, the fixed effect was phrase position, and by speaker random intercepts were allowed. According to the log likelihood ratio test, the effect of phrase position was significant [$\chi^2(1)=100.19$, $p<0.0001$]. Adding by speaker random slopes for phrase position to this model was significant [$\chi^2(1)=118.83$, $p<0.0001$]. Adding the fixed effect of accent type and by speaker random slopes for accent type were also significant [$\chi^2(4)=541.07$, $p<0.0001$]. According to this best model, phrase position (non final) had a coefficient of 0.167, accent type (LH) had a coefficient of 0.458. The global mean (i.e. the intercept of the linear model) was 0.337. Conceptually, this means that normalized H timing is on average 33% into the total word duration, and this is adjusted by accent type and phrase position. If accent type is LH, H timing is additionally delayed by about 45% of the total word duration, and if non final position, H timing is delayed by about 16% of the total word duration. Combining the two effects, this model predicts that the normalized H timing will be 94% (=33+45+16) into the total word duration if non final position and LH accent type.

Table 1. The mean and standard deviation of normalized H by accent type and phrase position for each speaker (the numbers in the parentheses are standard deviation.)

Speaker	Accent	Final	Non-final
J1	HL	0.344 (0.06)	0.494 (0.08)
	LH	0.740 (0.05)	0.910 (0.04)
J2	HL	0.325 (0.04)	0.504 (0.10)
	LH	0.824 (0.04)	0.953 (0.03)
J3	HL	0.342 (0.07)	0.449 (0.08)
	LH	0.792 (0.02)	0.953 (0.03)
J4	HL	0.213 (0.12)	0.615 (0.08)
	LH	0.909 (0.08)	1.000 (0.11)

3.3 Alternative models of phrase position effects

In Section 3.1 and 3.2, phrase position effects were modeled with H timing measured by normalization. We can also consider another model of phrase position effects where H timing is normalized by the duration of the tone bearing syllable rather than by the total word duration. The tone bearing syllable is the syllable where the tone is phonologically associated. Thus, it is the initial syllable for the HL accent type (***na**na*, (HL) 'seven'), and the second syllable for the LH accent type (*ya**ma***, (LH) 'mountain'). Myers (2003) normalized the H timing by the duration of the tone bearing syllable, because in Kinyarwanda, the HL and LH accent types can only occur in a bimoraic syllable. Each mora in a bimoraic syllable bears a L or H tone. On the other hand, in our speech materials of Japanese, the L and H tones were phonologically associated with each syllable in a disyllabic word, so a pitch rise or a fall spans two syllables. Thus, it is plausible that H is timed proportionally with respect to its associated syllable duration, rather than the whole word.

This section compares models of phrase position effects with varying H timing measures: the normalization method by the tone bearing syllable duration and the normalization method by the total word duration. In addition, we test models where H timing is not normalized, but measured as the relative distance in milliseconds from the word final boundary, or from the end of the tone bearing syllable. To summarize, in this section we compare four models: (i) normalization by the total word duration, (ii) normalization by the tone bearing syllable duration, (iii) distance in milliseconds relative to the word boundary, and (iv) distance in milliseconds relative to the end of the tone bearing syllable.

In particular, we will compare the AIC (Akaike's Information Criterion) values in these models. These models cannot be directly compared using the LRT because the models are non nested, yet the AIC values can be compared to find which model best fits the data. The lower the AIC, the better the model. In particular, a rule of thumb is that there is basically no support for the model which has the higher AIC value by more than 10 (Burnham and Anderson 2002:70ff.).

The four models are applied to the HL and LH accent types separately, so there are eight models. Table 2 summarizes the AIC values of all eight models discussed in this section.

For the HL accent type, for model (i), the time of H was normalized by the total word duration. The best mixed model was found with normalized H as the dependent variable, phrase position as the fixed effect, and by speaker random intercepts. In this model, the effect of phrase position was significant (slope=0.21). The AIC of this model was 214.1, as shown in Table 2.

Table 2. Model comparison (AIC values)

	Normalization by		Distance relative to	
	(i) the total word duration	(ii) the associated syllable duration	(iii) the end of the word	(iv) the end of the associated syllable
HL	-214.1	70.76	1087	1072
LH	-327.9	-146.3	1283	NA

For model (ii), the time of the H of the HL accent type was normalized by the duration of the first syllable. In the best model, the dependent variable was the normalized H computed in this way, and the fixed effect was phrase position, and by speaker random intercepts were allowed. The effect of phrase position was significant (slope=0.538). The AIC of the best model was 70.76. This AIC value was much higher than the AIC of model (i), so we can conclude that normalization by the total word duration is a better measure than normalization by the associated syllable duration for the Japanese disyllabic HL words.

In the same way, the two different normalization criteria were compared with the LH accent type. For the LH accent type, the time of the H from the beginning of the second syllable was divided by the duration of the second syllable. As can be seen in Table 2, normalization by the word duration (model (i)) yielded the lower AIC value than normalization by the tone bearing syllable duration (model (ii)) (cf. the AIC value of the model in Section 3.2, where the HL and LH types were pooled together, was 486.7).

The right two columns in Table 2 show the AIC values of the models where the H timing is measured in milliseconds relative to the end of the word (model (iii)) or the end of the tone bearing syllable (model (iv)). For model (iv) of the HL type, the best mixed model had the following specifications: the dependent variable was the difference between the times of the F0 peak and the end of the first syllable (in milliseconds), the fixed effect was phrase position, and by speaker random intercepts and slopes for phrase position. Here, the effect of phrase position was also significant (slope=51). The AIC value is lower in model (iv) than in (iii), but the difference is small (only 15) and both AIC values are higher than the AIC values of normalization models.

A similar procedure was taken for the LH accent type. However, the end of the word and the end of the tone bearing syllable are the same, so this comparison is unavailable for LH words. Yet, in Table 2, we can see that the relative distance measure for the LH accent type has a higher AIC value than any other models of normalization.

Two conclusions can be drawn from the results above. First, normalization by duration gives rise to a better quantitative model of phrase position effects on H timing than distance in milliseconds relative to a segmental landmark (comparison of (i),(ii) vs. (iii),(iv)). Second, in the normalization model, the total word duration serves as a better reference domain than locally defined duration in modeling phrase position effects in disyllabic words in Japanese (comparison of (i) vs. (ii)).

4. Discussion

It is worthwhile to consider the implication of the H timing of the HL accent type normalized by the duration of the associated syllable (model (ii) in Section 3.3), though it was not the best model. The mean of the normalized H in this model was 1.089, which means that the H peak of the HL accent type tends to occur a bit later than the end of the first syllable. This is comparable to the previous findings in the literature reporting late peaks in Japanese (Sugito 1982, Ishihara 2006). Ishihara (2006) shows that when the initial accented syllable is monomoraic (***mi**nari* 'appearance', ***na**mida* 'tears'), the H peak occurs at the beginning of the vowel in the second mora. On the other hand, if the initial syllable is bimoraic (***nin**mu* 'duty'), the peak is usually found within the initial syllable. In our speech materials, the initial syllables were all monomoraic, so it is expected that the F0 peaks of the HL type in our data will be delayed after the end of the initial accented syllable, which is what is observed.

A difference can be found between the Japanese results in this paper and the Kinyrwanda results in Myers (2003). In Kinyawranda, phrase position effects were greater in the LH accent type than the HL accent type. Myers (2003:80) suggested that this difference could be due to the fact that the H in the LH accent is closer to the end of the phrase. However, in our data, the magnitude of the phrase positional effect was not significantly different in the HL accent type and in the LH accent type. In the normalization model (by total word duration), the slope of phrase position (NF) was 0.210 for the HL accent type and 0.137 for the LH accent type (with standard errors 0.065, 0.020 respectively). These slopes are not significantly different from each other [t(150)=1.058, p=0.295]. One might think that the effect of phrase position would be greater for the LH accent than the HL accent because the H tone is closer to the phrase end, as Myers (2003) suggested (and also supported by other studies such as Silverman and Pierrehumbert (1990) – F0 peaks on word final syllables were significantly earlier than peaks on non word final syllables in English). However, our results suggest that this is not always the case. The H peak of a HL word is in the middle of the word, but the H peak of a LH word is at the end of the word, so there may be more temporal space available for an H peak of a HL word to vary than for an H peak of a LH word.

We have highlighted the difference between Kinyarwanda and Japanese that the HL and LH accents are associated within one syllable in Kinyarwanda but they span two syllables in Japanese. Normalization was carried out by one syllable in Kinyarwanda but by two syllables in Japanese. However, these two domains are not different in that they both are the domains that carry the HL and LH tones. Thus, in both languages, the H timing is best explained by taking into account the whole domain that contains *both* L and H tones – in Japanese it was the whole disyllabic word, in Kinyarwanda it was one bimoraic syllable. This may further mean that the H timing is closely related to the other tone, L, so the L and H tones must be considered as one unit that comprises a rising or falling F0 movement, rather than tones that are separately linked to each mora (cf. Ladd 2004b:128).

Our result that normalization measures fit the data better than relative distance measures is comparable to the result of Ladd et al. (1999) for British English prenuclear pitch accents. Ladd et al. (1999) measured the distance in milliseconds between the F0 peak and several segmental landmarks (e.g. the onset of the post accentual vowel) to find the segmental position (the 'anchor') where the F0 peak is stably aligned. They also measured a proportional position of F0 peaks, relative to the duration of the post accentual consonant. Comparing these two measures (relative vs. proportional), the proportional measure better captured the regularity in the alignment of F0 peaks (Ladd et al., 1990: 1550). Similarly, in our results, the proportional measure (normalization by total word duration) fitted the data better than the relative distance measure.

Finally, in our experiment, isolated words were used for phrase final condition, so the two conditions were not directly comparable in terms of sentence length. While the current results should be understood with this limitation, various conditioning factors (e.g. sentence length, proximity to the prosodic boundary, speech rate) can be studied in future research to get a full picture of the timing patterns due to prosodic position.

5. Conclusion

In this study, we examined phrase positional effects on tonal timing in disyllabic words in Tokyo Japanese. The timing of H peaks was normalized by dividing the duration from the beginning of the phrase onset to the H time by the duration of the word. The normalized H timing was significantly affected by accent type and phrase position. The H timing was later in the LH accent type than in the HL accent type, and in non final positions than in final positions.

In addition, four models were compared to examine the validity of our quantitative results of phrase position effects. We compared models of phrase position effects with the normalization measures and the relative distance measures. These methods were tested with varying domains (the whole word or the associated syllable) and with varying reference points (the end of the word or the end of the associated syllable). The best model was found to be the normalization measure divided by the total word duration, rather than the duration of the associated syllable. This means that tonal targets in the Japanese rising accent are timed with respect to the total word duration, rather than with respect to each separate syllable.

Acknowledgement

I thank the Japanese speakers who participated in the experiment, Mafuyu Kitahara for finding a subject for me, and three anonymous reviewers for their helpful comments.

References

Arvaniti, A., and Ladd, D.R. (1995). "Tonal alignment and the representation of accentual targets", in *Proceedings of the 13th International Congress of Phonetic Sciences, Stockholm*, Vol. 4, pp. 220-223.

Boersma, P. and Weenink, D. (2011). *Praat: doing phonetics by computer*. Version 5.2.08, http://www.praat.org.

Burnham, K. P., Anderson, D.R. (2002). *Model selection and multimodel inference: a practical information theoretic approach.* 2nd Edition. New York, NY: Springer Verlag.

Caspers, J., van Heuven, V. (1993). "Effects of time pressure on the phonetic realization of the Dutch accent lending pitch rise and fall", *Phonetica*, Vol. 50, pp. 161-171.

Goldsmith, J. A. (1976). "Autosegmental Phonology", Ph.D. dissertation, MIT.

Grice, M., Ladd, D.R., Arvaniti, A. (2000). "On the place of "phrase accents" in intonational phonology", *Phonology*, 17, pp. 143-186.

Hasegawa, Y., Hata K. (1995). "The function of F0 peak delay in Japanese", in *Proceedings of the 21st Annual Meetings of the Berkeley Linguistics Society*, pp. 141-151.

Ishihara, T. (2006). "Tonal alignment in Tokyo Japanese", Ph.D. dissertation, University of Edinburgh.

Ladd, D.R. (2004b). "Segmental anchoring of pitch movements: autosegmental phonology or speech production?", in H. Quene and V. van Heuven (Eds.), *On Speech and Language: Essays for Sieb B. Nooteboom,* pp. 123-131. LOT.

Ladd, D.R., Faulkner, D., Faulkner, H., Schepman, A. (1999). "Constant 'segmental anchoring' of F0 movements under changes in speech rate", *Journal of Acoustical Society of America*, Vol. 106, pp. 1543-1554.

McCawley, J. (1968). *The Phonological Component of a Grammar of Japanese.* The Hague: Mouton.

Myers, S. (2003). "F0 Timing in Kinyarwanda", *Phonetica*, Vol. 60, pp. 71-97.

Pierrehumbert, J. and Beckman, M.E. (1988). *Japanese Tone Structure*. Cambridge, MA: MIT Press.

Prieto, P., van Santen, J., Hirschberg, J. (1995). "Tonal alignment patterns in Spanish", *Journal of Phonetics*, Vol .23, pp. 429-451.

Silverman, K., Pierrehumbert, J. (1990). "The timing of prenuclear high accents in English", in *Papers in laboratory phonology* I (J. Kingston & M. Beckman, editors). pp. 72-106. Cambridge: Cambridge University Press.

Steele, S. (1986). "Nuclear accent f0 peak location: effects of rate, vowel, and number of following syllables", *Journal of Acoustical Society of America*, Vol. 80, pp. S51-S51.

Sugito, M. (1982). *Nihongo Akusento no Kenkyuu* [Studies on Japanese Accent]. Tokyo: Sanseido.

Xu, Y. (2001). "Fundamental frequency peak delay in Mandarin", *Phonetica*, Vol. 58, pp. 26-52.

Xu, Y. (2005). "Speech melody as articulatorily implemented communicative functions", *Speech Communication*, Vol. 46, pp. 220-251.

• **조혜선 (Cho, Hyesun)**
서울대학교 언어학과
서울시 관악구 관악로 1
Tel: 02-880-6164, Fax: 02-882-2451
Email: chohazel@gmail.com
관심분야: 음운론, 음성학

Appendix: Speech materials

1. hashi	'chopstick'	12. hashi	'bridge'
2. ima	'now'	13. ima	'living room'
3. mushi	'disregard'	14. tsuyu	'heavy rain'
4. shiro	'white'	15. tabi	'trip'
5. tsuyu	'soup'	16. inu	'dog'
6. tabi	'socks'	17. ami	'net'
7. ame	'rain'	18. mimi	'ear'
8. imi	'meaning'	19. mono	'thing'
9. ani	'elder brother'	20. mune	'chest'
10. nana	'seven'	21. mura	'village'
11. umi	'sea'	22. yama	'mountain'