When does intonational transfer occur? A comparative study of interrogative rises in four groups of L2 Japanese learners

Aaron Albin[1], Ruilai Wang[2]

[1]Kobe University, Japan
[2]Nanjing Normal University, China
albin@people.kobe-u.ac.jp, wang_ruilai@yahoo.co.jp

Abstract

While studies of L2 intonation agree that cross-linguistic transfer is common, there is still no established way to predict which specific contrastive analysis predictions will (and will not) be substantiated. To collect more primary descriptive data along these lines, the present study compares 12 Tokyo Japanese native speakers to 70 L2 learners studying Japanese as a foreign language in terms of how they distinguish statements from questions – normally marked by the presence vs. absence of a final rise. Learners represented four typologically distinct L1 groups: Russian, Vietnamese, Korean, and Mandarin. Each speaker participated in a read-aloud task containing 12 verbs, each presented to learners in an alternating sequence of statements and questions, e.g. *neru* 'sleep' vs. *neru?* 'sleep?'. The resulting tokens were measured in terms of duration, F0 level, and F0 span, and cluster analyses were conducted to classify the F0 contours according to their shape. While learners differed systematically from native speakers in many ways (including with respect to lexical accent), many of the predicted cases of intonational transfer were either absent or only sporadically attested, highlighting the importance of non-transfer factors (such as developmental universals and interlanguage innovations) in explaining L2 intonation.

Index Terms: cross-linguistic influence, sentence type, typology, pitch accent, KNN

1. Introduction

As is the case for segmental aspects of phonology, second language learners have been reported to encounter difficulties in acquiring the intonational phonology of a second language (L2) when it is different from their first language (L1) [1]. Most problems stem from learners applying aspects of their L1 system when producing L2 utterances, in a phenomenon known as cross-linguistic 'transfer'.

While L2 intonation transfer is quite ubiquitous and can take many forms [2], it is still difficult to make *a priori* predictions about when and how transfer will appear. Indeed, in the field of Second Language Acquisition more generally, the results of studies from the 1960s working in the Contrastive Analysis framework [3] found that many types of transfer that would be predicted based on L1-L2 differences are in fact unsubstantiated, and some attested patterns are not predictable based on such differences. The discovery of the latter eventually led to the idea of learners creating their own idiosyncratic language system, known as interlanguage [4]. Since rigorous empirical studies of L2 intonation are still few in number and represent an emerging field, it is still important to check contrastive analysis predictions to determine where these fall short and where interlanguage aspects are most prevalent.

The present study does precisely this by working with perhaps the most fundamental intonational distinction: that between statements and yes/no questions. In many languages (such as English), yes/no questions are marked intonationally with an utterance-final rise, which is lacking in pragmatically unmarked statements (i.e., declarative sentences). However, not all languages follow this pattern. For example, in Mandarin Chinese, yes/no questions are marked morphologically with the sentence-final question particle *ma*. Moreover, Russian yes/no questions are marked with a final rise-fall contour. Thus, this is an area ripe for testing cross-linguistic transfer predictions. Specifically, the present study systematically explores where cross-linguistic transfer is observed in L2 Japanese yes/no question intonation by learners from four L1 backgrounds (Korean, Mandarin, Russian, and Vietnamese).

1.1. Intonation of yes/no questions cross-linguistically

In Tokyo Japanese, questions have a final rise (L%H%) whereas statements typically have no such rise (L%) [5, 6, 7]. The same situation also holds true for Seoul Korean [8, 9, 10, 11, 12]. However, the standard varieties of Russian, Mandarin, and Vietnamese are different in this regard. More specifically, Russian exhibits several contour types depending on the sub-type of question, with yes/no questions characterized by a sharp rise on the nuclear syllable followed by a fall over the following syllables [13, 14]. In Mandarin yes/no questions, while it is possible to have a final rise over the utterance-final particle -*ma* in pragmatically marked contexts [15], the contour shape of statements and questions is otherwise identical, reflecting the lexical tones in the utterance [16]. However, Mandarin questions have been reported to be realized in a higher register than declaratives [17]. The use of a higher register in questions has also been reported for Vietnamese [18].

1.2. Intonation of Japanese L2 learners' yes/no questions

There are relatively few studies focusing on the production of yes/no question intonation by L2 Japanese learners. Indeed, no previous study could be found on Japanese question intonation production by L1-Korean learners, perhaps due to the similarities between Korean and Japanese question intonation.

Regarding L1-Russian learners, Ayusawa [19] presented a case study of an advanced L1-Russian learner who produced echo questions (e.g., *are?* 'that one?') with final falling pitch. This result is in line with what would be predicted based on the non-rising yes/no question contour from L1 Russian.

Chen [16] examined the production of Japanese question intonation by three Mandarin learners. One learner (who had been studying Japanese for 1.5 years) produced yes/no questions with a final rise, whereas the other two learners (one with almost no prior experience studying Japanese and another who had been studying for 4 years) exhibited no final rise ("falling" or "flat" intonation, according to Chen). Again, this might be predicted based on the L1 system, in which a final rise is generally not obligatory in marking interrogativity.

Todoroki [20] investigated the production of Japanese echo questions on one-word utterances by two L1-Vietnamese learners. One learner (who had been studying Japanese for 9 months) produced all echo questions with global rising pitch over the whole word (even in cases where the contour should be rise-fall-rise), while another learner (Japanese learning length: 7 months) with final-rising pitch or rising-falling pitch. Moreover, both learners produced the echo questions in an expanded pitch range. If register modulations are the primary way yes/no questions are marked in Vietnamese, then such contour-shape confusions and modulations of F0 range would be predicted as a kind of 'strategy' transferred from the L1.

1.3. Present study

Two points are worth making about the studies just surveyed. First, these studies suffer from extremely small sample sizes (only 1-3 participants per study), calling into question the generalizability of these findings. Second, each study focused on a single L1 group, with no comparative study contrasting learners with different phonological systems regarding question intonation. The present study seeks to simultaneously fills both of these gaps. More specifically, a relatively large sample (N=13 or more) is collected and compared across four different L1 groups.

As established above, the over-arching research question for the present study is, "In what ways, and for which L1 groups, will predicted cases of L1 transfer be substantiated?". Based on the studies discussed in the literature review in the preceding section, four hypotheses were posited: that some Russian learners will show a rising-falling contour (#1), that some Mandarin and Vietnamese learners will have no final rise (#2) and instead show a raised register (#3), and that most Korean learners will pattern like native Japanese speakers with a final rise (#4). Instead of searching narrowly for these specific types of transfer, the analysis below will take the approach of broadly inventorying all kinds of deviation from native speaker norms. In particular, four analyses will be presented – namely, of utterance duration, F0 level, F0 span, and F0 contour shape. Then, in the general discussion, these four predicted cases of transfer will be re-examined in a new light.

2. Method

2.1. Materials

A total of 12 verbs were used in this study, structured as follows: 3 mora-counts (two vs. three vs. four) crossed with 2 accent types (unaccented vs. accented on the penultimate mora) crossed with 2 words per combination. The list of 12 verbs is presented in Table 1 below. Capitalized letters indicate the location of lexical accent, which is manifested as a H*+L pitch fall in Tokyo Japanese [21].

Table 1: *Verbs used in the present study.*

Moras	Accented words	Unaccented words
2	MIru 'see', NOmu 'drink'	iru 'exist', neru 'sleep'
3	taBEru 'eat', oRIru 'go down'	arau 'wash', owaru 'end'
4	yoroKObu 'be happy', gaNBAru 'try hard'	umareru 'be born', wasureru 'forget'

2.2. Participants

A total of 70 Japanese learners participated in this study, all of whom were undergraduate students studying Japanese as a foreign language. These learners were split into four groups: (1) 23 L1-Korean learners at a university in Cheonan city, South Chungcheong province, (2) 20 L1-Mandarin learners at a university in Jingzhou city, Hubei province, (3) 14 L1-Russian learners at a university in Novosibirsk city, Novosibirsk province, and (4) 13 L1-Vietnamese learners at two universities in Hanoi municipality.

Table 2 below summarizes three kinds of background information about these four L1 groups: the number of participants (N) broken down into females (F) and males (M), age (mean and standard deviation), and the duration of L2 Japanese study (mean and standard deviation).

Table 2: *Metadata on the four learner groups*

Group	N (F+M)	Age	L2 study
Korean	23 (17+6)	21.0 (1.4)	3.5 (1.8)
Mandarin	20 (15+5)	18.5 (0.5)	0.8 (0.0)
Russian	14 (11+3)	18.6 (0.9)	2.4 (1.6)
Vietnamese	13 (11+2)	19.8 (1.0)	1.8 (0.9)

In addition to these four learner groups, 12 Japanese native speakers (6 female and 6 male) also participated in this study as the control group. All of them were native Tokyo dialect speakers who were studying at two universities in Tokyo.

2.3. Procedure

Each participant completed a reading task which required participants to read the 12 verbs aloud first as statements, then (immediately thereafter) as questions. The question vs. statement distinction was cued visually with punctuation, i.e. a final period " 。 " vs. a final question mark " ? ". Thus, for example, a learner produced *neru.* 'sleep' and then *neru?* 'sleep?'. By pairing learners' productions in this way, learners were pushed to use whatever resources they felt appropriate for distinguishing the two. Due to widespread dropping of syntactic arguments that can be inferred from context, these utterances are completely normal (e.g., meaning "I will go to sleep" or "Will you sleep?", respectively). To ensure lexical knowledge was not a problem, literal English translations (e.g. 'sleep' vs 'sleep?') were displayed alongside the target words, and a phonetic guide (*furigana*) was put on all Chinese characters (*kanji*).

All participants were tested in a quiet room. Utterances were recorded at 44.1 kHz using an online recording system named "Online Voice Recorder" developed by the Minematsu & Saitō Laboratory at the University of Tokyo. Effectively, this system enables a PowerPoint-like interface for web-based collection of phonetic data.

3. Results

3.1. Duration

For each combination of speaker and word, the ratio of question vs. statement word durations was calculated using this formula: ((QuestionDuration / StatementDuration) - 1) *100. Thus, for example, a value of 20 means that a question token is 20% longer than the statement production of that same word by that same speaker. Figure 1 summarizes measurement's distribution for each L1 group and mora count.
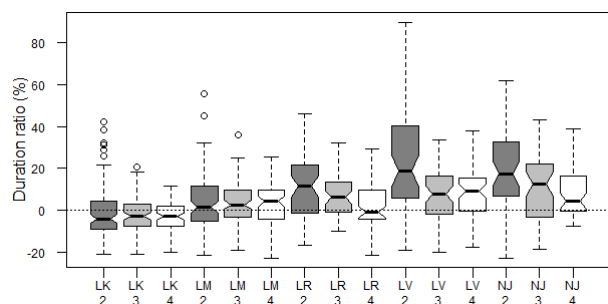

Figure 1: *Duration ratio by L1 group and mora count.*

First, note that in nearly all cases, the middle fifty percent of the distribution (i.e. the box) straddles 0 with both positive and negative values, although there is a global tendency towards the positive side. This suggests that lengthening a question may be one cue (albeit minor) to signal an utterance's status as a question. While this appears to be a common feature across the surveyed L1s, some L1-specific tendencies can be observed. For example, Russian learners (LR) and Vietnamese learners (LV) pattern like the Japanese native speakers (NJ) in that there is a slight tendency for shorter questions (e.g. 2 moras) to have a greater duration boost than longer questions (e.g. 4 moras). The same is not true, however, for the remaining two groups: Korean learners (LK) and Mandarin learners (LM).

3.2. F0 level

F0 level is operationalized here as the minimum pitch across an utterance (i.e. the pitch 'floor'), under the assumption that pitch range is expanded 'from the bottom up' [22]. For each combination of speaker and word, the distance was calculated between the F0 level of the question and the paired statement of the same word by the same speaker. To correct for nonlinearities in the F0 space and make males and females more comparable, such distances were calculated in semitones using the equation: $\log_{base=2}$(QuestionLevel / StatementLevel) * 12. Thus, positive values mean that the question has a higher level (i.e. utterance-level minimum pitch) than the statement.

Figure 2 shows how these 'level distances' are distributed across the various L1 groups. The results are also split by sex in order to illustrate the effect of the semitone normalization. (The horizontal width of each box indicates the relative sample size, such that wider boxes represent a more robust N.)

Note that the distributions for the Japanese native speakers (NJ) is right at 0, suggesting that they do not systematically modulate F0 level as a cue to interrogativity. To varying extents, the same is not true of the various learner groups, however. For Mandarin (LM), Russian (LR), and Vietnamese (LV) learners, one sex has an overall positive skew, and for Koreans, both sexes do.
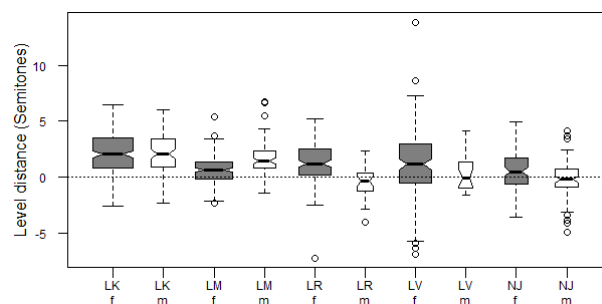

Figure 2: *Level distance by L1 group and gender.*

3.3. F0 span

Here, F0 span, another term based on [22], refers to the pitch range of an utterance, i.e. maximum minus minimum F0. Experimentation during the data analysis revealed that, compared to simply subtracting the difference in Hertz, gender differences are better neutralized if this is expressed as the semitone distance from the minimum to the maximum. Thus, F0 span was calculated as follows, analogous to the calculation of level distance above: $\log_{base=2}$(Maximum F0 / MinimumF0) * 12. The resulting values are always positive since the maximum is always higher than the minimum.

Next, the ratio between the F0 span of questions (numerator) vs. statements (denominator) was calculated for each combination of speaker and word. Thus, a value of 7, for example, means that the question has an F0 span 7 times as wide as the corresponding statement, and a value of 1 means the two are equally wide. Figure 3 below plots the distribution of these 'span ratios' according to not only L1 group but also accentedness as well (where 'a'=accented and 'u'=unaccented).
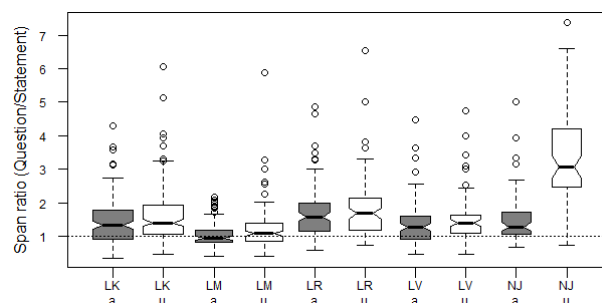

Figure 3: *Span ratio by L1 group and accentedness.*

Native speakers show a clear distinction between accented and unaccented words, with only the latter exhibiting a tripling of the F0 span in questions relative to statements. No learner group exhibits a similar effect, most likely due to the fact many learners lack the lexical specification for each word's pitch accent in the first place. Most learner groups are skewed to values above 1, however, suggesting that learners do employ an expansion in F0 range as a cue to interrogativity, although to the same extent regardless of the accented vs. unaccented distinction. The only exception is the Mandarin learners (LM), whose distributions are centered squarely at 0, indicating statements and questions are produced with similar F0 spans. When gender is incorporated into the analysis (not shown in Figure 3), males globally have larger values for span ratio, but this effect is relatively small in magnitude and independent of the observations above.

3.4. F0 contour shape

Each F0 contour was first segmented into 6 bins and the median F0 value within each bin was calculated. F0 values were normalized by expressing them as a percentage (0-100%) across that utterance's F0 span. Figure 4 shows the results of this analysis for the native speaker (NJ) data. The four lines represent the four major contour types: accented statements (aS), unaccented statements (uS), accented questions (aQ), and unaccented questions (uQ). The points in the contours represent the median of the normalized F0 values for each bin and each contour type.
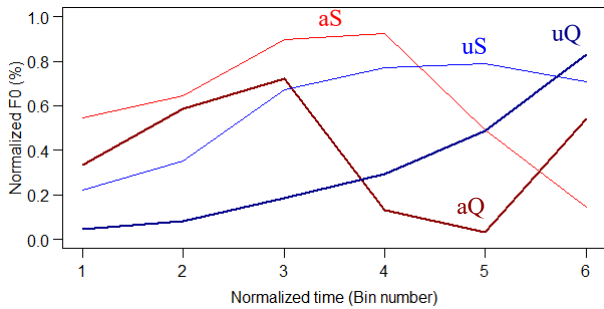

Figure 4: *Four median contours for native speakers.*

By applying this technique to all utterances in the data set (including those of learners), the result is that every utterance is represented as a 6-point vector in a space of normalized time and normalized F0. The resulting dataset was then input to the *k*-nearest neighbors (KNN) classification algorithm using the knn() function in the R package 'class' using k=9. The training data was the 6-point vectorization for all native speaker utterances, the labels were the four contour types (aS, uS, aQ, and uQ), and the test data was the 6-point vectorization for all learner utterances. Thus, this algorithm effectively uses the labeled native data (for which the 'true' contour type is known) and uses them as a reference point for automatically classifying the learner data (for which the 'true' contour type is unknown).

Table 3 shows the results of applying this algorithm to the four different L1 groups as test data. For comparison, the final column contains the results of using the native Japanese data as both the training data and the test data. The 'Mapping' column indicates which contour types are being dealt with in each row. For example, "aS→uQ" means that a word was expected to be produced with the 'accented statement' contour but was classified as actually being produced with the 'unaccented question' contour. A mapping with an equals sign (like "aS=aS") means that a word was classified as the expected contour type. The numbers themselves represent the percent of each expected category realized as the relevant contour type. For each group of 4 rows, these sum to 100%.

Mappings for expected 'aS' words were comparatively unproblematic, perhaps due to the rise-fall shape of the word-level contour being familiar and/or unmarked. In contrast, expected 'uS' words rarely showed the correct rise-only contour, and instead the modal response for all four learner groups was the aS (rise-fall) contour. As suggested above, confusion between accented and unaccented statement contours is most likely due to learners lacking a lexical specification for each word's accentuation. Thus, for statements, assimilation patterns were unidirectional: uS→aS. For expected aQ contours, accuracies were rather low, and a

significant portion of tokens were produced as uQ. Here, a similar confusion also occurs in the other direction, although not quite as strong. Thus, for questions, the confusion is bidirectional, perhaps because global rise contours (uQ) and rise-fall-rise contours (aQ) are both familiar and/or unmarked ways of marking interrogativity to these learners.

Table 3: *K-nearest neighbors classification results.*

Mapping	LK	LM	LR	LV	NJ
aS=aS	*65.2*	*68.3*	*72.6*	*65.4*	*88.9*
aS→uS	15.9	12.5	13.1	23.1	5.6
aS→aQ	18.8	4.2	13.1	6.4	2.8
aS→uQ	0	15	1.2	5.1	2.8
uS→aS	67.4	55.8	63.1	62.8	6.9
uS=uS	*15.9*	*18.3*	*8.3*	*28.2*	*77.8*
uS→aQ	15.9	4.2	25	6.4	2.8
uS→uQ	0.7	21.7	3.6	2.6	12.5
aQ→aS	4.3	15.8	2.4	7.7	2.8
aQ→uS	4.3	16.7	0	3.8	0
aQ=aQ	*51.4*	*26.7*	*28.6*	*35.9*	*97.2*
aQ→uQ	39.9	40.8	69	52.6	0
uQ→aS	3.6	11.7	1.2	6.4	0
uQ→uS	3.6	22.5	4.8	10.3	4.2
uQ→aQ	42	22.5	22.6	30.8	0
uQ=uQ	*50.7*	*43.3*	*71.4*	*52.6*	*95.8*

4. Discussion

For hypothesis 1, Russian learners were expected to use rising-falling contours (aS) for questions, i.e. high values for any incorrect mappings from Q to aS. However, such values were very low (2.4% for aQ→aS and 1.2% for uQ→aS). For hypothesis 2, Mandarin and Vietnamese learners were expected to lack a final rise for questions, i.e. high values for incorrect Q→S mappings (aQ→aS, aQ→uS, uQ→aS, and uQ →uS). While the values for Mandarin were fairly high (15.8, 16.7, 11.7, and 22.5 percent, respectively), this was not so for Vietnamese (7.7, 3.8, 6.4, and 10.3). For hypothesis 3, Mandarin and Vietnamese learners were expected to show a raised register, but this effect was only observed for Vietnamese females and Mandarin males. For hypothesis 4, Korean learners were expected to consistently have a nativelike final rise in questions, and indeed, incorrect Q→S mappings were very low (either 4.3% or 3.6%).

The results also uncovered various trends not predicted based on transfer alone: (1) Korean and Mandarin learners lacked native-like modulation of question-vs-statement duration ratio according to word length, (2) various learner groups (especially the Korean learners) raise F0 level as a cue to interrogativity not found in native speech, (3) all learner groups lacked the native-like tripling of F0 span in unaccented questions, and (4) various contour-shape confusions were observed, most notably uS→aS, aQ→ uQ and uQ→aQ. Thus, transfer must be treated as merely one among an inter-connected web of factors in explaining L2 intonation

5. Acknowledgements

6. References

[1] U. Gut, *Non-Native Speech: A Corpus-Based Analysis of Phonological and Phonetic Properties of L2 English and German*. Frankfurt: Peter Lang, 2009.

[2] A. Albin, "Typologizing native language influence on intonation in a second language: Three transfer phenomena in Japanese EFL learners," Ph.D. dissertation, Indiana Univ., Bloomington, Indiana, 2015.

[3] R. Lado, *Linguistics Across Cultures: Applied Linguistics for Language Teachers*. Ann Arbor: University of Michigan Press. 1957.

[4] L. Selinker, "Interlanguage," *International Review of Applied Linguistics in Language Teaching*, vol. 10(1-4), pp.209-232. 1972.

[5] T. Ayusawa, "Nihongo-no gimonbun-no inritsuteki tokuchō" [Prosodic features of Japanese interrogative sentences], *Nihongo-no Inritsu-ni Mirareru Bogo-no Kanshō: Onkyōonseigakuteki Taisyō Kenkyū* [Native Language Interference Observed in Japanese Prosody: Acoustic-Phonetic Contrastive Studies], vol.2, pp.1-20, 1991.

[6] T. Ayusawa and S. Taniguchi, "Nihongo onsei-no inritsuteki tokuchō [Prosodic features of Japanese speech]", *Nihongo-no Inritsu-ni Mirareru Bogo-no Kanshō: Onkyōonseigakuteki Taisyō Kenkyū* [Native Language Interference Observed in Japanese Prosody: Acoustic-Phonetic Contrastive Studies], vol.1, pp.1-24, 1990.

[7] S. Kōri, "Hantei yōkyū-no shitsumonbun-ni okeru gimongata jōshōchō-to sono onsei-no tokuchō" [Interrogative rising tones of judgement-requesting question sentences and their phonetic characteristics], *Gengo Bunka Kenkyū* [Language and Culture Research], vol.39, pp.221-244, 2013.

[8] C.-S. Suh, "Interrogatives and indefinite words in Korean: with reference to Japanese," Harvard Studies in Korean Linguistics, vol.3, pp.329-340, 1989.

[9] K.-J. Min, "Kankokugo gimonbun intonēshon-no onkyōteki bunseki-to gōsei onsei-ni yoru chikaku jikken: Nihongo-to-no: taisyō kenkyū-no tame-no kisoteki shiryō-toshite" [Acoustic analysis of Korean interrogative intonation and perceptual experiments with synthetic speech: As a basic resource for contrastive research with Japanese], *Onsei Gakkai Kaihō* [Phonetic Society Bulletin], vol.205, pp.29-33, 1994.

[10] S.-A. Jun and M. Oh, "A prosodic analysis of three types of *wh*-phrases in Korean," *Language and Speech*, vol.39, pp.37-61, 1996.

[11] S.-A. Jun, "Korean intonational phonology and prosodic transcription," in *Prosodic Typology: The Phonology of Intonation and Phrasing*, S.-A. Jun, Ed. Oxford: Oxford University Press. 2005, pp.201-229.

[12] H. Hwang, "Wh-Phrase Questions and Prosody in Korean," in Proceedings of the 17th Japanese/Korean Linguistics Conference, Stanford, USA, 2007, pp.295-310.

[13] E. A. Bryzgunova, Zvuki i intonatsiya russkoy rechi [Sounds and intonation of Russian speech], Moscow, Russkiy Yazyk [Russian Language], 1977.

[14] N. D. Svetozarova, "Intonation in Russian," in *Intonation Systems: A Survey of Twenty Languages*, D. Hirst and A. di Cristo, Eds. Cambridge: Cambridge University Press, 1998, pp.261-274.

[15] S. H. Peng, M. K. Chan, C. Y. Tseng, T. Huang, O. J. Lee, and M. E. Beckman, "Towards a Pan-Mandarin system for prosodic transcription," in *Prosodic Typology: The Phonology of Intonation and Phrasing*, S.-A. Jun, Ed. Oxford: Oxford University Press. 2005, pp.230-270.

[16] W. Chen, "Chūgokugo washa-ni yoru Nihongo gimonbun bunmatsu-no inritsuteki tokuchō" [Japanese interrogative sentence-final prosodic characteristics of Chinese speakers], *Nihongo-no Inritsu-ni Mirareru Bogo-no Kanshō* [Native Language Interference Observed in Japanese Prosody], vol.3, pp.1-26, 1992.

[17] X.-N. S. Shen, *The Prosody of Mandarin Chinese*. Berkeley, University of California Press, 1990.

[18] D. Hirst and A. D. Cristo, *Intonation Systems: A Survey of Twenty Languages*. Cambridge, Cambridge University Press, 1998.

[19] T. Ayusawa, "Intonēshon to Nihongo kyōiku" [Intonation and Japanese teaching], *Nihongogaku* [Japanese Linguistics], vol.10(7), pp.98-113, 1991.

[20] Y. Todoroki, "Betonamugo bogo washa-no toikaeshibun-no onchō-ni tsuite" [On the tones of echo questions by Vietnamese native speakers], STUDIUM, vol.19, pp.106-118, 1992.

[21] J. J. Venditti, "The J_ToBI model of Japanese intonation," in *Prosodic Typology: The Phonology of Intonation and Phrasing*, S-A. Jun, Ed. 2005, pp. 172-200.

[22] D. R. Ladd, *Intonational Phonology*, 2nd ed. Cambridge, Cambridge University Press, 2008.