International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:12, No:9, 2018

The Phonology and Phonetics of Second Language Intonation in Case of "Downstep"

Tayebeh Norouzi

***Abstract*—**This study aims to investigate the acquisition process of intonation. It examines the intonation structure of Tokyo Japanese and its realization by Iranian learners of Japanese. Seven Iranian learners of Japanese, differing in fluency, and two Japanese speakers participated in the experiment. Two sentences were used to test the phonological and phonetic characteristics of lexical pitch-accent as well as the intonation patterns produced by the speakers. Both sentences consisted of similar words with the same number of syllables and lexical pitch-accents but different syntactic structure. Speakers were asked to read each sentence three times at normal speed, and the data were analyzed by Praat. The results show that lexical pitch-accent, Accentual Phrase (AP) and AP boundary tone realization vary depending on sentence type. For sentences of type *XdeYwo*, the lexical pitch-accent is realized properly. However, there is a rise in AP boundary tone regardless of speakers' level of fluency. In contrast, in sentences of type *XnoYwo*, the lexical pitch-accent and AP boundary tone vary depending on the speakers' fluency level. Advanced speakers are better at grouping words into phrases and produce more native-like intonation patterns, though they are not able to realize downstep properly. The non-native speakers tried to realize proper intonation patterns by making changes in lexical accent and boundary tone.

***Keywords*—**Intonation, Iranian learners, Japanese prosody, lexical accent, second language acquisition.

I. Introduction

A. The Problem

MANY studies assume that Japanese has two patterns of lexical accent, with patterns that have a sharp fall from high pitch to low usually called *accented* and patterns that have no such fall usually called *unaccented* [1]-[4]. However, Japanese lexical pitch-accent is not typically covered in textbooks, and it is considered a difficult prosodic component for learners of Japanese to acquire [5].

Given the properties of Japanese lexical pitch-accent, earlier studies on the acquisition of Japanese prosody by learners of Japanese concentrated on the realization of lexical accent in isolation [6]-[9]. In other words, the fact that speech is a continuous articulatory stream and lexical accent realization may be affected by prosodic structure has been ignored. Furthermore, most previous research on second language speech acquisition has focused on the interaction between first (L1) and second (L2) language prosody. Only a few studies on the acquisition of L2 intonation have considered how the phonological and phonetic features of L2 prosody are acquired, which particular features are acquired, and when this acquisition occurs, based on instrumental evidence [10].

Studies on the evaluation of the pronunciation of foreign learners of Japanese assume that a proper production of intonation has more effect on the positive evaluation of foreign learners' speech [11]. Therefore, recent studies suggest that the teaching of intonation should be a higher priority than segmental or other super-segmental aspects of the Japanese language. Therefore, many studies have been conducted to invent simple but effective methods to facilitate Japanese prosody education [12]. However, current research on the acquisition of Japanese intonation is not sufficient.

B. Intonation Model

Many studies on Japanese intonation acquisition assume a model in which the speech system consists of two components, the phonological and the phonetic [13]. These components, which are independent aspects of intonation, have been strictly differentiated from one another.

For the phonological component, two hierarchical phrasing units at the word level have been defined: the intonational phrase (IP) and the AP [14]. The IP is the highest and the largest phrase, consisting of one or more than one AP. Each AP consists of one or more pitch accents.

In the phonetic component, abstract tones are interpolated following certain regularities such as declination (i.e., overall fundamental frequency (F0) downtrends toward the end of the utterance) and downstep (i.e., a stepwise lowering of F0 in high tones) [1], [10].

Ladd [13, p. 9] assumes that the phonological components have various phonetic properties both segmental and super-segmental.

C. Declination Realization in Japanese

Previous studies assume that when an accented word is followed by another accented word in Tokyo Japanese, the F0 peak of the following accented word tends to decrease over the course of the utterance. However, reports of F0 decrease are not limited to accented words, as has been observed by many phonologists and phoneticians [1], [15]. However, Kori [3], [4] suggests that there is also a systematic relation between the syntactic structure of words and downstep. He provides the examples of (1) *Na'ra de mo'miji wo Yu'mi to mi'ta* (I watched Nara's autumn leaves with Yumi) and (2) *Na'ra no mo'miji wo Yu'mi to mi'ta* (in Nara, I watched autumn leaves with Yumi), then indicates instrumentally that although the initial F0 peak of the AP *mo'miji wo* declines in both structure (1) and structure (2), the F0 peak of *mo'miji no* in (2) is more declined due to a semantic relation between the two APs. Kori [3], [4] suggested that in (2), the following PA *momji no* is restricted by *Nara*,

Tayebeh Norouzi is with the University of Tsukuba, Japan (e-mail: tyb_nrz@yahoo.com).

which is a modification for *momoji*. However, a similar restriction is not observed in (1). For this reason, he claims that what is known as *downstep* occurs only in structures that include a semantically restrictive modification.

II. Purpose of Study and Hypotheses

In contrast to previous studies that relied on conventional approaches focusing on lexical pitch-accent realization, this study focuses on the effect of syntactic structure on the realization of lexical pitch-accent. It aims to investigate the phonological and phonetic characteristics of Japanese lexical pitch-accent and intonation, as realized by Iranian learners of Japanese.

Based on previous studies [10], we contend that L2 learners acquire the intonational system of the target language in two steps. First, they internalize the phonological structure of L2 intonation, though they still implement these structures using their L1 phonetic habits. Later, learners master native-like phonetic regularities.

Our study focuses on downstep in restrictive modification environments, since downstep affects both the pitch realization and the F0 peak of the APs that follow. Specifically, we made the hypotheses below:

Hypothesis 1. Downstep is a phonetic property observed only in Japanese, not in Persian. However, in Persian, when two APs are in a modification relation they form an IP unit, a tendency similar to that of sentence structure (2), *XnoYwo*. In other words, since differences in IP formation (which is a phonological property) in restrictive modification structures like *XnoYwo* between Japanese and Persian are phonologically subtle, it is very likely that the Iranian learners will perceive the L1 and L2 IP formation as similar. Therefore, category formation for L2 IP formation is blocked and a single category will be used for both L1 and L2 IP formation. Given this assumption, we predicted that the Iranian learners may be able to combine the target APs of *XnoYowo* into one IP, but they would have difficulties in realizing downstep in *XnoYwo*.

Hypothesis 2. Persian is a stress-accent language, so accent usually falls on the last syllable of nouns, adjectives and most adverbs, and pitch accent contrasts do not exist [16]. In Japanese, downstep affects the initial F0 of the following AP and the degree of the initial F0 rise is affected by the lexical accent type [2], [3]. We thus predicted that Iranian learners of Japanese would have difficulties in producing Japanese lexical pitch-accents, and that lexical accent realization may therefore affect the quality and quantity of the downsteps they realized.

III. Experiment

A. Subjects

Two major speaker groups were compared: two female native speakers of Japanese (aged 24 and 25) as a control group, and seven female Iranian learners of Japanese (aged from 24 to 28, average age of 24.44) as an experimental group. Subjects in the control group were from Kanagawa and Tokyo, where Tokyo Japanese (standard Japanese) is spoken. Later, two other native speakers (different from the ones in the control group; both females, aged 25 and 35) were paid to evaluate the intonation patterns.

B. Materials

Test sentences were designated to test the degree of declination predicted by the theory of relation between modification restriction and downstep: F0 for a following AP declines due to a semantically restrictive modification by the preceding AP. We applied two sentences consist of similar words with the same number of syllables numbers (two syllables) and lexical pitch-accent (initial mora accented) but with different syntactic structures: Semantically restrictive modification and non-restrictive modification.

The data set for this production experiment is as below:

(1a) ˈ*Chiba de ˈchizu wo takusan katta.*
Chiba-PP map Case many buy-Past Tense
In Chiba, I bought many maps`.

(1b) ˈ*Chiba no ˈchizu wo takusan katta.*
Chiba-Case map-Case many buy-Past Tense
`I bought many maps of Chiba.

(2a) ˈ*Naha de ˈumi wo ooi ni tanoshinda.*
Naha-PP sea-Case much enjoy-Past Tense
In Naha, I enjoyed the sea much.

(2b) ˈ*Naha no ˈumi wo ooi ni tanoshinda.*
Naha-Case sea-Case much PP enjoy-Past Tense
I enjoyed the sea of Naha much.

C. Procedure

The order of sentences was randomized, and filler sentences were inserted pseudo-randomly to separate each target sentence. The subjects assigned task was to read each sentence t times. The subjects in the experimental group were recorded in a soundproof room at the phonetics laboratory of the Faculty of Foreign Languages at Tehran University in Iran, and those in the control group were recorded at the phonetics laboratory of Tsukuba University in Japan. Finally, in addition to the pitch tracks, to determine which utterances produced by the L2 learners are the most natural and the most native-like, two native speakers were asked to evaluate the nativeness of the target APs in all utterances and score them out of 5.

D. Measurement

For the analysis of phonetic properties, the F0 peak of the two target APs were measured for each sentence. Subsequently, we simply measured the F0 ratio by dividing the F0 peak of the following APs into the F0 peak of the preceding APs. Fig. 1 shows a schematic pitch pattern of the two target sentence types as well as the F0 peak measurement method.

The obtained value reveals the quality and degree of declination. As mentioned earlier, in Japanese, due to declination the following AP is realized with a lower F0 peak in both *XdeYwo* and *XnoYwo* structures. Thus, if the value is less than 1, it reveals that a declination has been realized and the possibility for the realization of a normal and native-like intonation increases. If the value is over 1, it shows that the overall intonation is abnormal and less native-like. Furthermore, as it was discussed in IC, since a downstep occurs in *XnoYowo,* it is expected that the initial F0 in the following AP of *XnoYowo*

is realized steeper and lower than that of *XdeYwo*.

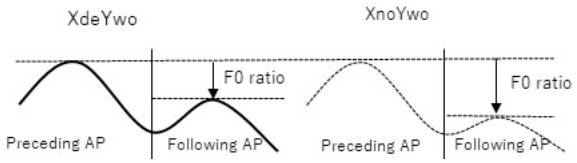

Fig. 1 Schematic pitch tracks of two heading target APs of sentence type *XdeYwo* (Solid line) and sentence type *XnoYwo* (dotted line)

IV. Results and Discussion

A. Phonology of L2 Intonation

The pitch tracks of representative speakers are shown in Figs. 1-6. These figures can be used to compare the phonological aspects of the speech of both groups, such as lexical pitch-accent and phrasal tones, as well as the phonetic aspects, such as the realization of declination in the following AP *Ywo* in *XdeYwo* and the realization of downstep in the following AP *Ywo* in *XnoYwo*. In the figures, there are two pitch tracks that show the first two target APs of each sentence. The pitch tracks of the solid line show the structure *XdeYwo*, and the pitch tracks of the dotted line show the structure *XnoYwo*.

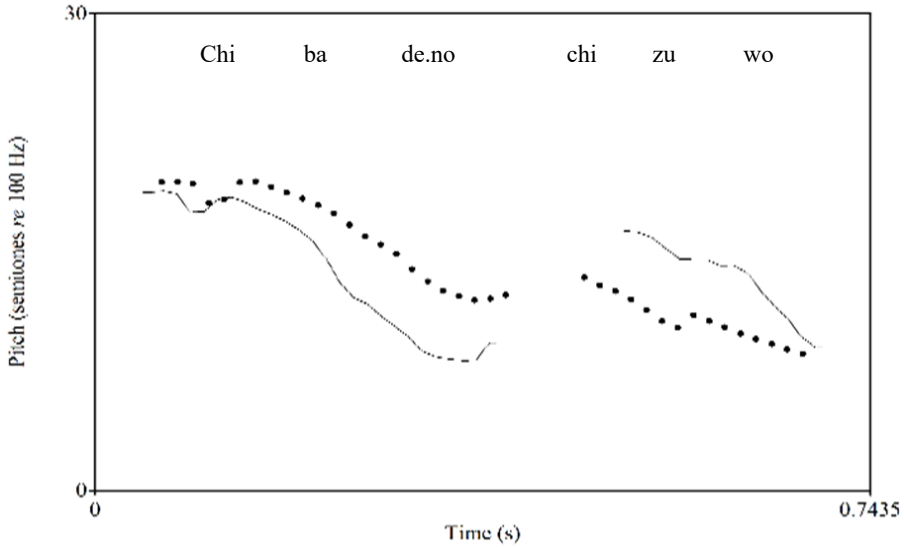

Fig. 2 Utterances produced by NS1. Solid line = *chiba de chizu wo* dotted line = *chiba no chizu wo*

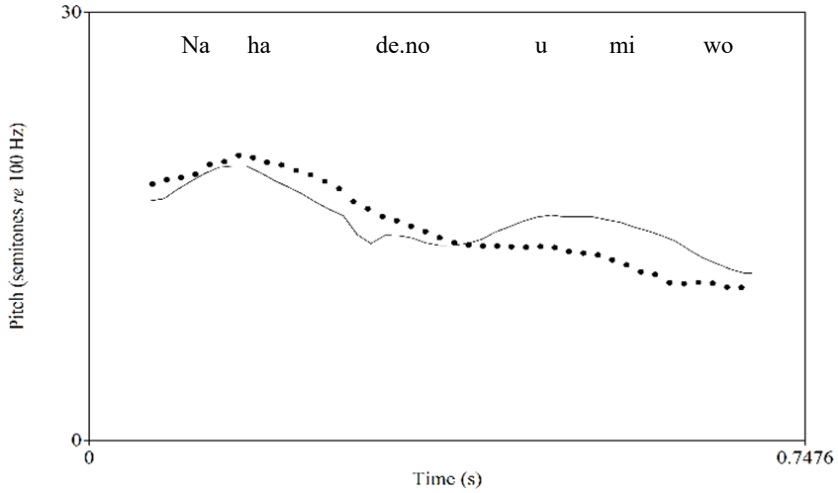

Fig. 3 Utterances produced by NS2. Solid line = *Naha de umi wo,* dotted line = *Naha no umi wo*

Figs. 2 and 3 show the pick tracks of native speakers. As we mentioned earlier and as these figures illustrate, the F0 peak in the following AP *Ywo* is not realized as sharply as the F0 peak in *Xwo,* neither in *XdeYwo* nor in *XnoYwo*. In other words, in both structures, a declination appears in the second AP. However, there are differences in the intonational phrasing of *XdeYwo* and *XnoYwo*: in *XdeYwo*, the preceding AP *Xde* and the following AP *Ywo* are realized in two separated IPs,

whereas in *XnoYwo*, the two APs form a unit IP. In both structures, the AP boundary is realized with an L boundary tone.

Figs. 4-7 show the pitch tracks of representative Iranian speakers. As these figures illustrate, in almost all learners' utterance of *XdeYwo*, each prosodic word and its past position are realized as one AP with a proper lexical pitch accent. However, each AP is realized with a rise in the AP boundary. Due to this rise, the F0 of the second AP is realized as high or higher than that of the first. Finally, the preceding AP, as well as the following AP, form separate IPs. In contrast, the overall intonation pattern of *XnoYwo* varies by speaker. The first pattern is shown in Fig. 4, in which the intonation of *XnoYwo* is realized similar to the manner of *XdeYwo* described above. This type of intonation, in which neither the IP formation nor the downstep is realized properly, is observed in the utterances produced by S1, S2, and S3.

The second pattern is shown in Fig. 5, in which the lexical pitch-accent of the preceding AP *Xno* is realized as if it were composed of Japanese unaccented words. It is also similar to the Persian lexical accent, in which the last syllable of the word is realized more prominently. Moreover, the two APs form an IP unit. However, the IP boundary is realized as an H tone.

The third pattern is shown in Figs. 6 and 7, where the overall intonation of *XnoYwo* is similar to that of the native speakers' intonation in Figs. 2 and 3. In other words, the lexical pitch-accent, the AP boundary tone, and the intonation formation for S6 and S7 are more native-like than the other speakers.

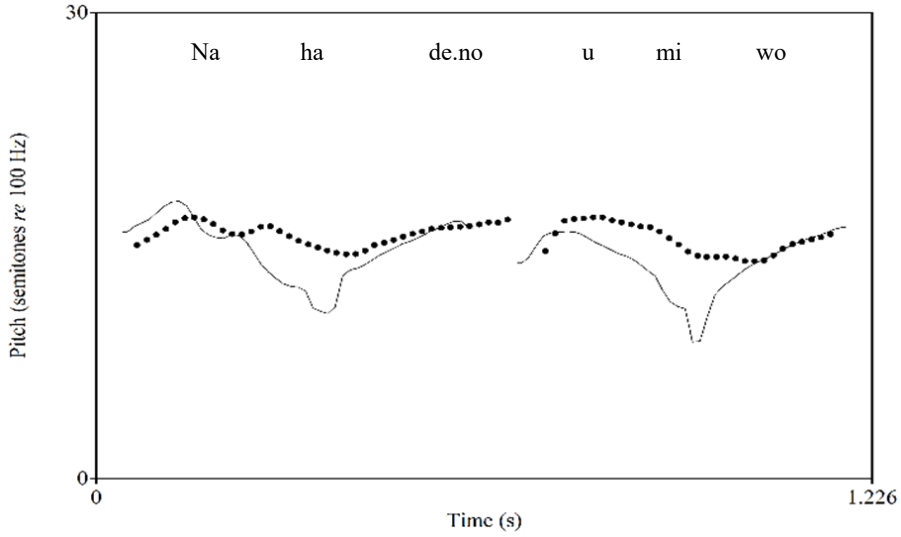

Fig. 4 Utterances produced by S2. Solid line= *Naha de umi wo,* dotted line= *Naha no umi wo*

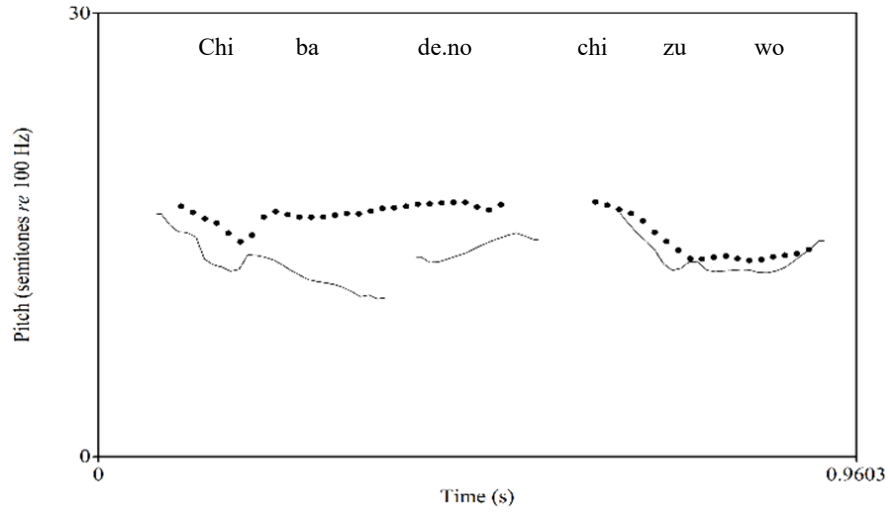

Fig. 5 Utterances produced by S5. Solid line= *chiba de chizu wo,* dotted line= *chiba no chizu wo*

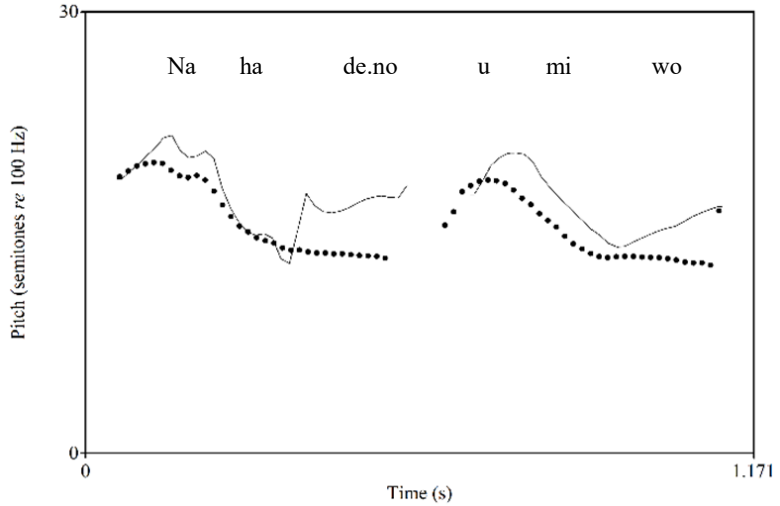

Fig. 6 Utterances produced by S6. Solid line= *Naha de umi wo,* dotted line= *Naha no umi wo*

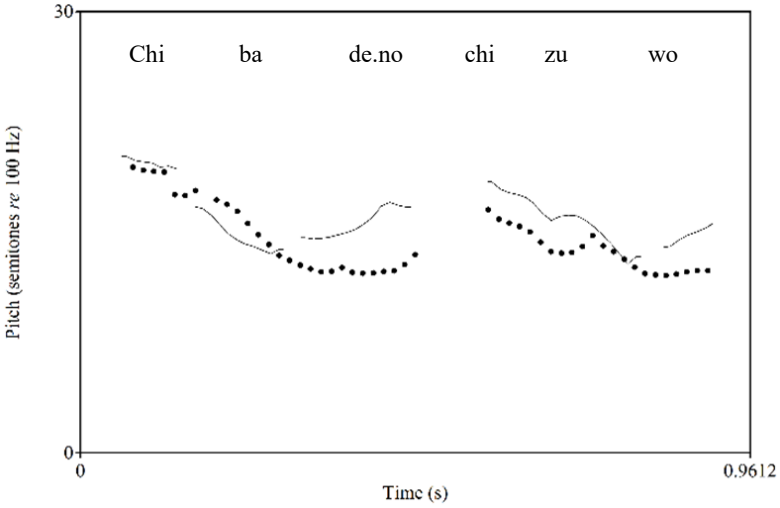

Fig. 7 Utterances produced by S7. Solid line= *chiba de chizu wo* and dotted line= *chiba no chizu wo*

B. Phonetics of L2 Intonation

For *XdeYwo,* almost all learners realized the lexical pitch-accent properly. They also properly realized each AP in separated IPs. However, they did not realize the AP boundary properly. It can be assumed that almost all learners acquired the proper phonology of *XdeYwo,* except for the AP boundary tone. In contrast, S4 and S5 realized the two APs as an IP unit in *XnoYwo* but tended to realize the preceding AP as an unaccented word. Moreover, S6 and S7 realized the lexical pitch-accent properly in all APs and their boundary tone was more native-like. It could therefore be suggested that S6 and S7 successfully acquired the phonology of *XnoYwo*. Does this mean that the acquisition of *XnoYwo* was complete? This remains to be seen. Further examination of the data with respect to phonetic properties may show more contrast between the learners and native speakers.

As mentioned earlier, in order to investigate the phonetic properties of native speakers' utterances, as well as those of Iranian learners of Japanese, we measured the F0 peak value of the preceding and following AP of each utterance. We also measured the F0 ratio of the two APs for both structures, *XdeYwo* and *XnoYwo*. The results are shown in Fig. 8.

As shown in Figs. 2 and 3, the pitch tracks of the native speakers suggest that, regardless of syntactic structure, the F0 peak in the following AP is always realized steeper than the F0 peak in the preceding AP. Therefore, the F0 ratio must be less than 1. As mentioned earlier, if the mean of the F0 ratio is less than 1, it can be inferred that the learners' utterance is more native-like. Moreover, if the value of the F0 ratio is around 0.67 (e.g., native speakers' F0 ratio for *XnoYwo*), it can be inferred that learners can realize downstep as successfully as native speakers. However, the results reveal that in *XdeYwo* the F0 ratio for S4 and S5 is greater than 1. Further, the F0 ratio of S1, S2, S3, and S6 is near 1. The F0 ratio for S7 is the lowest and is therefore most likely to be native-like.

In *XnoYwo*, though, it was expected that the mean F0 ratio

would be less than that of *XdeYwo*. This was not the case. Instead, there was an increase in the F0 peak of the following AP of *XnoYwo* for S1, S2, and S6. In contrast, the mean F0 ratio of *XnoYwo* for S4, S5, S6, and S7 indicates that the F0 peak of the second AP was realized steeper than that of *XdeYwo*. However, there is still a considerable difference between the F0 ratio of these speakers and that of the native speakers.

A paired-sample t-test was also conducted in order to compare the F0 ratio of *XdeYwo* and *XnoYwo* separately for each speaker. As shown in Fig. 9, there was a significant difference in the F0 ratio of *XdeYwo* (*NS1: M= 0.89, SD= 0.06; NS2: M= 0.84, SD= 0.03*) and *XnoYwo* (*NS1: M=0.67, SD=0.02; NS2: M= 0.69, SD= 0.02*) for the two native speakers (*NS1: t (5) =8.28, p > 0.001; NS2: t (5) = 4.11, p > 0.001*). Our results suggest that the F0 peak in the following AP of *XnoYwo* is realized lower than the F0 peak of *XdeYwo*. It assumes that in *XnoYwo*, a downstep is realized in the native speakers' utterance, as expected. However, considering the mean F0 ratio of all the learners except S7, no one can make such a reduction in the following AP of *XnoYwo*.

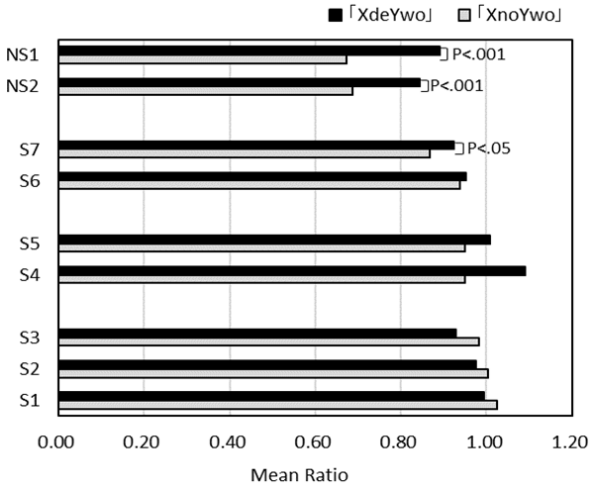

Fig. 8 Mean of F0 ratio between the first two APs of *XdeYwo* and *XnoYwo*

	Paired Differences					t	df	Sig. (2-tailed)	
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of Difference Lower	Upper				
S1	-0.031	0.106	0.043	-0.142	0.081	-0.705	5	0.512	
S2	-0.028	0.062	0.025	-0.093	0.037	-1.103	5	0.320	
S3	-0.054	0.073	0.030	-0.130	0.023	-1.811	5	0.130	
S4	0.145	0.249	0.102	-0.117	0.406	1.419	5	0.215	
S5	0.057	0.097	0.040	-0.046	0.159	1.424	5	0.214	
S6	0.013	0.032	0.013	-0.020	0.046	1.007	5	0.360	
S7	0.057	0.040	0.016	0.015	0.098	3.497	5	0.017	*
NS1	0.218	0.065	0.026	0.151	0.286	8.282	5	0.000	***
NS2	0.071	0.133	0.017	0.036	0.105	4.112	5	0.000	***

Fig. 9 Paired sample t test results for each speaker

C. Evaluation of the Native-Like Qualities of Speech

The results of the native speakers' evaluation of the speech by the two groups showed that type 3 is the most native-like intonation pattern. Surprisingly, although in intonation pattern type 1, two target APs were realized in two separate IPs, it was shown to be more native-like than intonation type 2, in which the two target APs were phrased and formed an IP unit. The native speakers claimed that none of these intonation patterns were natural. However, since type 1 was realized in two APs with a high AP boundary tone, it reminds them of a Japanese accent realized by a foreign learner. In contrast, type 2—in which the word is pronounced with an unaccented lexical pitch-accent—reminds them of an atypical Japanese person's accent. However, they were not sure where in Japan this accent originates.

V. Conclusion

In this work, we have proposed a framework for the development of instrumental studies about the realization of Japanese intonation, particularly downstep in the speech of Japanese language learners. Specifically, this experiment tested two hypotheses regarding the phonological properties of speech (such as the lexical pitch-accent and IP formation) as well as the phonetic properties of speech (such as downstep) in the speech production of Iranian learners of Japanese. We presented two syntactic structures, both similar in the number of syllables and word order but different in prosodic structure. The first two APs were not in a restrictive modification relation in the first type, but they were in the second type. Comparing these two prosaically different structures, we were able to investigate if the utterances of learners were realized with an appropriate F0 lowering (i.e., declination in *XdeYwo* and downstep in *XnoYwo*) in accordance with the prosodic structure. The results of S7 indicate that it is not impossible to acquire native-like downstep in one's L2. S7 exhibits F0 values which are similar to those produced by native speakers of Japanese. However, it is difficult to explain her success based solely on the data of this study.

Although only a very small number of the subjects produced a Japanese intonation with an almost native-like pitch track, our preliminary results could partially support the theory that the phonological properties of a speech system are acquired earlier than the phonetic properties. Moreover, the results show that at the phonological level most learners failed to produce the AP boundary tones of *XdeYwo* correctly, though they could all produce the lexical pitch-accent. All this suggests a few key findings: that different phonological aspects are acquired at different stages in the learning process, that lexical pitch-accent is acquired earlier than AP boundary tones, and that AP boundary is the most difficult phonological aspect of speech to acquire. Moreover, at the phonetic level the results show that most learners were not able to properly realize declination and downstep, as the F0 peak of the following AP was realized even more sharply than the preceding AP. One of the causes of this is that the AP boundary was raised and realized as a high tone. It can be assumed that since the AP boundary was not acquired and realized properly, the F0 peak for the following AP could not be realized steeper.

Another new finding of the study was that learners' grouping

phrases and making greater units is not necessarily equal to a natural and acceptable pronunciation, according to the evaluation of native Japanese speakers.

To our knowledge, this may be the first work that has pointed out that the lexical accent realization of Japanese language learners differs by sentence structure, even when the word order and number of syllables are the same. There have been many studies that focus on Japanese prosody education and claim that is not important to teach Japanese lexical pitch-accent, since it is so complicated [12]. However, this approach that focuses narrowly on intonation form is not very helpful for learners and can often lead to advanced students producing an uncommon accent of Japanese that is not typical of foreigners or native Japanese speakers. It can be theorized that the application of a more systematic prosody education that covers both lexical pitch accent and intonation properties would motivate learners to learn Japanese prosodic features more deliberately.

Our further work will be comprised of three stages. First, we will focus on extending the size of the data set for a more detailed analysis. We then intend to conduct the same experiment on learners who have acquired the correct intonation and observe if they are able to produce the right intonation consciously. Our final task will be to determine which of the two sentence types are different in degree of declination in the following AP.

ACKNOWLEDGMENTS

The author would like to give special thanks and appreciation to the author's supervisor, Dr. Hiroshi Matsuzaki, for his helpful advice and feedback on this research. The author would also like to thank the Japanese and Iranian participants for contributing their speech and for helping the author conduct this production experiment.

REFERENCES

[1] J. B. Pierrehumbert and M. E. Beckman, *Japanese Tone Structure.* Cambridge, MA & London: MIT Press, 1988.
[2] S. Kori "Restrictive modification and realization of lexical accents in Tokyo Japanese: A quantitative study," *Journal of Language and Culture Studies.*, vol. 38, 2012, pp. 1-22.
[3] S. Kori, "Accent reduction and restrictive modification in Tokyo Japanese," *Journal of the Phonetic Society of Japan.,* vol. 12(1), 2008, pp. 34-53.
[4] S. Kori, "A new analysis of accent in Tokyo Japanese: About handling rise of the beginning of the word," *Japanese Language Studies.,* vol. 55(2), 2004, pp. 16-31.
[5] N. Yamada, "Steps in the acquisition of Japanese lexical accent: The case of foreign learners," *Journal of Japanese Language Teaching*, vol. 85, 1994, pp. 108-120.
[6] E. Takahashi, "The essential factors for acquisition of Japanese lexical accent by Korean learners: A study of advanced learners," *Journal of Japanese Language Teaching.,* vol. 169, 2018, pp. 16-30.
[7] E. Takahashi, "The acquisition of Japanese accent in English speaking learners: The relationship of knowledge and production," *Hiroshima University Japanese Language Education Research*., vol. 24, 2014, pp. 1-8.
[8] Y. Nakato, "Pitch realization of Korean learners of Japanese in word reading," *Journal of Japanese Language Teaching*, vol. 109, 2001, pp. 80-89.
[9] Y. Sukegawa, "Pitch realization of 2/3-mora-words by Brazilian learners of Japanese," *Journal of the Phonetic Society of Japan.,* vol. 3(3), 1999, pp. 13-25.
[10] M. Ueyama, "The phonology and phonetics of second language intonation: The case of 'Japanese English," *Fifth European Conference on Speech Communication and Technology, EUROSPEECH,* Rhodes, Greece, 1997.
[11] T. Sato, "Comparison of the influence of monophonies and prosody on evaluation of Japanese speech," *Japanese Language Education in the World,* vol.5, 1995, pp. 139-154.
[12] C. Nakagawa, "An attempt to teach prosody focusing on the intonation type ∧," *Journal of Japanese Language Teaching.,* vol. 110, 2001, pp. 140-149.
[13] D. Ladd, *Intonational Phonology (Cambridge Studies in Linguistics),* Cambridge: Cambridge University Press, 1996.
[14] J. Venditti, "Japanese ToBI labelling guidelines," *Ohio State University Working Papers in Linguistics*, vol. 50, 1997, pp. 127-162.
[15] W. Poser, *The Phonetics and Phonology of Tone and Intonation in Japanese,* PhD dissertation, MIT, 1984.
[16] N. Sadat-Tehrani, *The Intonational Grammar of Persian,* Ph.D. dissertation, University of Manitoba, 2007.