GRIFFITH UNIVERSITY

Foreign accent matters most when timing is wrong

Author

Tsurutani, Chiharu

Published

2010

Conference Title

INTERSPEECH

Downloaded from

http://hdl.handle.net/10072/36647

Link to published version

http://www.interspeech2010.org

Griffith Research Online

https://research-repository.griffith.edu.au

Foreign accent matters most when timing is wrong

Chiharu Tsurutani

School of Languages and Linguistics, Griffith University, Australia

c.tsurutani@griffith.edu.au

Abstract

This study aims to investigate native speakers' perception of prosodic variation of Japanese utterances. The pitch contour above the word level is hard to determine due to individual variation or pragmatic and para-linguistic factors. Nevertheless, native speakers' intonation is relatively consistent as long as the context and intention of the utterance is predetermined. On the other hand, L2 speakers' intonation contains some prosodic deviation from the native speakers' model, and yet some deviations are treated as non-native production and some are not. By identifying the prosodic deviations that are tolerated by native listeners, we will have better understanding of crucial points necessary for the improvement of Japanese pronunciation and the reference for computer-based assessment tools. The study suggests that pitch errors affect the performance score, but not as significantly as do timing errors.

Index Terms: perception, foreign accent, pitch, timing

1. Introduction

For L2 learners the ultimate goal will be attaining native-like speech. However, in the multilingual society of the 21st century, accented speech is commonly heard and intelligibility or naturalness has been highlighted more than has the identical copying of native speakers' speech [4]. We need to know which phonological factor helps improve the quality of pronunciation in an economical way. It is known that the acquisition of prosody has more impact on learners' intelligibility, as judged by native speakers, than does the acquisition of segmental features [1] [9]. This point drew the attention of researchers involved in computer-based assessment programs, and some efforts to incorporate prosodic features in assessment have been attempted as well [6] [8]. Prosody has various aspects, such as pitch, duration and loudness, and they interact under phonological rules of each language. The significance of prosodic features that influence the judgement of native speakers could differ depending on the language. A study of L2 English reports the importance of temporal properties of speech [11]. Prosodic features are a key component of natural and intelligible speech, and thus need to be put under examination to find out exactly which feature strongly affect native speakers' judgements.

In Japanese, pitch accent is a lexical property of a word, and is not affected by the prosodic organization at the higher level. Japanese intonation is formed by incorporating the accent patterns of words into phrase pitch pattern [16]. Because Japanese is a pitch accent language, intonation, which is a feature of phrases or sentences expressed by pitch, is expected to have a major influence on the acceptability of Japanese utterance. At word level, it has been reported that pitch is the most dominant cue to accent patterns in the prosodic features, such as, pitch, duration, intensity and spectral coefficient patterns [2]. In other words, pitch is the most significant feature to determine the word accent for both English and Japanese. Does that mean that pitch is an important feature for assessing learners' performance both at word and phrase level? Sato (1995) reported the predominance of prosodic factors over segmental factors, and that pitch was most influential among prosodic factors in assessing L2 speech. However, the participants of his experiment were Korean and Chinese native speakers whose level of Japanese was intermediate. Due to their proficiency level, their speech could possibly have contained fewer timing errors than the speech of average L2 learners. The L1 background and L2 proficiency of learners influence the acceptability of their L2 speech. In this study, using L2 speech of English learners of Japanese as experimental stimuli, the influence of pitch errors on native listeners' judgement will be examined in order to determine their relative importance among prosodic factors in Japanese.

2. Experiment

This experiment was designed to test the influence of pitch on native speakers' judgements of learners' utterances.

2-1. Materials

The stimuli were extracted from a recording by Australian English speakers who had been studying Japanese for 160 hours at university. The recording was taken from a computer exercise that was developed as a self-assessment of pronunciation. Utterances that contained timing errors or pitch errors or both, with no obvious segmental errors, were chosen, together with a correct utterance from a large sample of speech data. The judgement of errors was made by the researcher and two other native speakers who had been teaching Japanese for many years. When two out of three speakers agreed, the judgement was accepted. The following four patterns were considered.

1. incorrect pitch, correct timing -- Px
2. correct pitch, incorrect timing -- Tx
3. correct pitch, correct timing -- O
4. incorrect pitch, incorrect timing -- X

Using the same sentence would be convenient for analysis but could wear listeners' concentration. Thus, six sentences that contained the listed error patterns were chosen and formed 24 stimuli. The four patterns, 1- 4, of the same sentence was played consecutively, but the order of error patterns was randomized. Since the stimuli were natural utterances, it was not possible to control the degree of incorrectness as accurately as in synthesized speech. Efforts were made to select speech with similar speech rate and number of errors. Errors at both word and phrase levels were counted inclusively. None of the utterances had pauses longer than 300ms and all sounded reasonably fluent. The following is the list of the sentences.

1) *Shachoo-no kekkonshiki-ni okyakusan-ga sennin kita.*
(1000 people attended the president's wedding reception.)
2) *Tsugino jugyoo-no suugaku-wa chotto muzukashii desu.*
(Mathematics in the next class is a bit hard.)
3) *Watashi-no kookoo-de isshoni shashin-o torimashoo.*
(Let's take a photo together at my high school.)
4) *Otooto-no okusan-wa ryokoo-ni ikuno-ga sukidesuyo.*
(My younger brother's wife likes cooking.)

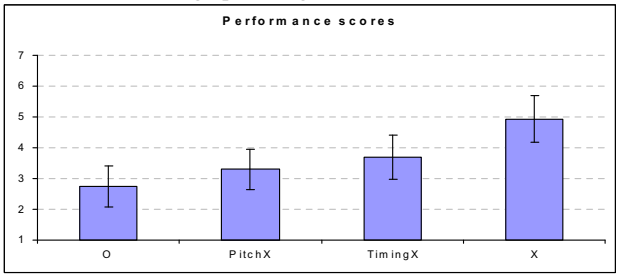
5) *Tanjoobi-ni tomodachi-kara kireena hana-o moratta.*
(I received beautiful flowers from my friend on my birthday)
6) *Shuumatsu-kara futarino hito-to shigoto-o suru yotee desu.*
(From the weekend I'm planning to work with 2 new people.)

2-2. Participants
Participants were 80 university students (22 females, 58 males) from Kansai region whose ages ranged from 18-20 years and who had not left Japan, apart from a short overseas trip. Kansai dialect employs pitch accents different from those in standard (Tokyo) Japanese. However, the young generation has been exposed to both Tokyo and Kansai dialects through mass media and school education, and is equipped with the standard perceptual criteria for common Japanese. Thus, participants' perception was regarded as the same as that of the average native listener. In past linguistic research, assessments were conducted by experienced language teachers who became accustomed to judging L2 speech for the necessity of the consistency and accuracy. However, there is a possibility that those assessors developed a level of tolerance towards L2 speech different from the tolerance of ordinary native listeners. L2 speech needs to be submitted to ordinary native listeners to see if it is acceptable for them. The current group of participants was chosen for this reason.

2-3. Procedure
To investigate participants' perceptual threshold towards pitch raising and final lengthening, their speech production and perception was analysed. Randomly selected participants took part in the recording.
-Recording of native speakers-
Twenty of the 80 native listeners participated in this task for a small payment. They formed pairs and talked about a fun experience or what they found interesting recently. Each speaker also individually read aloud the list of sentences spoken by the L2 speakers, so that their understanding of standard Japanese intonation could be examined. Their speech was recorded using the handy recorder Zoom H4n.
- Listening task-
The L2 learners' utterances that contained the 4 error patterns (24 in total) were played to the native listeners for naturalness judgement. The four versions of the same sentence were played consecutively for ease of comparison. Each stimulus was played twice with a four-second interval, and an inter-trial interval of eight seconds. Before the task, three practice sentences were played for listeners to become accustomed to the task and proficiency level of L2 speakers. The participants were asked to judge the naturalness of utterances using a Likert scale with potential responses ranging from 1 (native-like) to 7 (not at all native-like). Half of the participants listened to the version that started from the second half of the list (from No. 13) of the stimuli. The task was conducted at the end of class at their university. The listening task took about 15 minutes, which included time for instruction.

3. Results
3-1. Recording task
The researcher listened to the tape and confirmed that 19 out of 20 participants used pitch raising and lengthening at phrase final position frequently in their casual conversation. The one speaker who did not use them, however, used a severe pitch raising in the sentence reading. Sentence reading was examined in terms of pitch raising and phrase final lengthening, as there were no timing errors in native speakers' utterances. The participants were instructed to read the sentences as if they were presenting a model pronunciation for learners. One participant accidentally used Kansai pitch accent in a few words. However, the rest of participants read the sentences in standard accent pattern and showed their acquisition of it. All participants read the sentences formally and clearly without phrase final lengthening, but 10 out of 20 had pitch raising in some part of their reading despite the fact that Japanese speech does not normally present pitch raising at the phrase final position in formal style of reading. Thus, it was confirmed that this group of young people commonly uses pitch raising and lengthening at phrase final position in their every day conversation.

3-2. Listening task
The average and standard deviation of scores obtained from the judgement by the Likert scale were calculated for the 24 stimuli. First of all, four error patterns were compared using the average score for 480 sentences (6 sentences for 80 participants) in order to see if the four error patterns were distinguished by the listener participants. Standard deviations are listed below the graph in Figure 1.



SD 0.67 0.65 0.7 0.75

Figure 1: *The average score (+/- standard deviation) for four types of errors*
The average scores received were highest in sentences with X-type errors, followed by Timing X and Pitch X, and lowest in sentences with O-type errors. This means that the participants, on average, judged utterances with timing errors as worse than those with pitch errors. From the results shown in Figure 1 however, it is apparent that judgement of error types O and X will be different, but there appears to be likely overlap in the ranking of pitch and timing errors. The relative ranking of errors by individual participants was then looked at to further understand whether individuals more specifically distinguished between all four error types. Paired T-tests were undertaken to determine in particular if there were, at the individual level, significant differences in rankings between the 4 error patterns (see Table 1).

Table 1: *Results of multiple t-tests (2-tailed tests) between the four error groups*

Comparisons	t	p	Critical value
0 - X	-29.395	.000	.008
Timing - X	-23.799	.000	.01
Pitch - X	-22.939	.000	.0125
0 - Timing	-16.215	.000	.017
0 - Pitch	- 8.620	.000	.025
Pitch - Timing	- 7.065	.000	.05

This analysis showed significant differences not only between the most extreme assessments, but also between all other combinations, supporting the initial view of the error hierarchy presented in Figure 1.
In order to see the influence of error types, the stimuli were listed in the ascending order of their average score by 80 participants together with the number of errors.

Table 2: *Average scores of each stimuli*

ID No.	Error pattern	Number of errors			Score
		Word Timing	Word pitch	Phrase final	
10	O			1	1.61
17	O				2.14
19	Tx			3 (lengthening)	2.49
6	O				2.63
1	Px		2	1	2.74
3	O			1	2.76
18	Px		3		2.8
7	Tx			1 (lengthening)	2.93
24	Px		4	1	3.01
4	Tx	1		1 (lengthening)	3.03
14	Px	1		4	3.3
15	O				3.45
16	Tx	2			3.75
8	Px		2		3.88
22	O			1	3.88
9	Px		2	1 (sentence fin.)	4.11
11	X	2	1		4.2
5	X	4	3	1 (lengthening)	4.33
20	X	1	3		4.74
12	Tx	2	1		4.76
13	X	4	3		5.14
21	Tx	3			5.23
23	X	3	2	2	5.34
2	X	3		1	5.83

Overall stimuli at the bottom of the table had more timing errors. The top seven utterances do not contain any timing errors, but have some phrase final lengthening or pitch raising. Phrase final lengthening and pitch raising are seen across the ranking and do not seem to affect the native listeners' judgement. A linear regression analysis was conducted to see the impact of word- and phrase-level errors on the performance judgments (measured by score) shown in Table 2, with results presented in Table 3 below.

Table 3: *Results from the linear regression of score against different error types at the word and sentence level*

Model	Unstandardised coefficients B	Std. Error	Standardised coefficients β	t	Sig
Timing	.595	.110	.755	5.388	.000
Pitch	.071	.117	.086	.607	.550
Phrase	-.092	.152	-.085	-.608	.550

R2=.624, $F_{3,20}$ = 11.056, p<.000

A significant relationship between the number of timing errors and score was found .p<.000. However, the relationship between scores for utterances containing word pitch errors and errors at the phrase level was not significant. As can bee seen from Table 2, while a high incidence of timing errors was associated with higher scores (indicating lower acceptability), the incidence of pitch errors was less consistent, as are phrase-level errors.

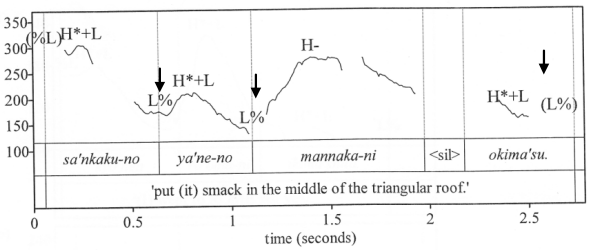
No particular type of timing error was associated with lower-scored utterances. That is, judgements of acceptability were not affected by whether the timing error involved lengthening rather than shortening, or vowels rather than consonants. This result also indicates that phrase final lengthening was irrelevant to naturalness judgement; it is suggested that this is because phrase final lengthening is a common phenomenon in speech. There were no particular types of pitch that affected native listeners' judgement, either. The correct pitch patterns for these words were the same both in Standard and Kansai Japanese. Thus, there is no possibility that pitch errors were overlooked by the Kansai speaker judges. Phrase final pitch raising errors occurred in several utterances, from the best to the worst, and showed no impact on the naturalness score. Thus, the results show that timing errors affect native listeners' judgement most significantly.

3. Discussion

When learners' speech is evaluated for assessment purposes, every deviation from the native speakers' model is pointed out for correction and feedback. However, even among native speakers, there were variations in pitch and intonation for one sentence [12]. When the goal of pronunciation improvement relates to intelligibility, we need to focus on prosodic factors that are most crucial for natives' judgement. While the research conducted by Sato (1995) reported that pitch is the most critical factor for improvement of pronunciation using one interrogative sentence as a model, the participants of this study, native speakers of Japanese who reside in Kansai, tolerated incorrect pitch pattern in a word even though the pattern was not in accord with their native pitch pattern. This study was conducted with English speakers whose Japanese utterances had noticeable timing errors as well as pitch errors, and used declarative sentences, which uniformly have declination towards the end of a sentence. Under this condition, pitch was not a crucial factor for naturalness judgement and timing errors were more important than pitch pattern.

In addition, phrase final pitch raising or lengthening was not regarded as a serious error. Final syllable lengthening functions as a boundary marker and appears to be a putative universal, but lengthening ratio varies depending on the language, and is higher in English and lower in Spanish [3]. Presumably, syllable-timed languages regulate timing more strictly than stress-timed languages. However, in this study only timing errors in words lowered the performance score.

Pitch raising at phrase final position is regarded as boundary tone, and a rising tone typically appears in English to signal non-finality. On the other hand, it is not common in Japanese sentences, as illustrated using ToBI annotation in Figure 2.

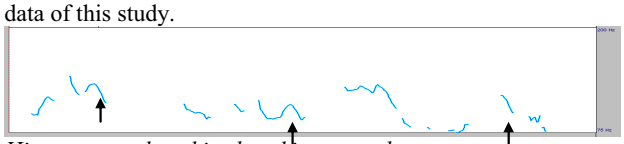


Sankaku-no yane-no mannaka-ni okimasu
The triangular roof-GEN in the middle put it
L% = Fall at the end of accentual phrase (Venditti et al:2008)

Figure 2: *Pitch contour of Japanese declarative sentence*

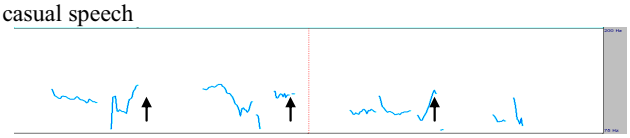
Figure 2 shows that pitch is lowered at the end of each phrase. However, pitch raising as well as final syllable lengthening was commonly observed in the casual speech of this group of native speakers and was not treated as negative points in their

judgement. Figures 3 and 4 show pitch raising observed in casual speech and formal reading respectively in the speech data of this study.



Hito-gaa *chuushin-dee* *kitarooro-do tenoga a-tee* *maa*
The person-Nom* plays a main role-and there is a road called "Kitaroo roodo"-and *Nominative particle

Figure 3: Pitch raising and final syllable lengthening in casual speech



Shuumatsu-kara *futarinohito-to* *shigoto-o suru* *yoteidesu*
Weekend –from two people-with work-Acc* do plan to

Figure 4: Pitch raising in one speaker's reading of Sentence 4

In both cases, pitch raising signals non-finality, but final syllable lengthening appears only in casual speech. Apparently this is because lengthening gives a sloppy impression [7] and makes the speaker's articulation unclear. When the final syllable is lengthened, it is also observed that the raised pitch goes down at the end (Figure 3), which is called *Gobiage, Shiriagari*-intonation by Inoue (1993).

In a previous study [12], when L2 speech was assessed by experienced teachers, pitch errors did not significantly reduce performance scores. It was suspected that the assessors who became accustomed to listening to L2 speech might have developed a tolerance towards it. In this study, ordinary Japanese listeners who assessed L2 speech showed the same tendency, which confirmed the Japanese listeners' tolerance toward pitch deviation. Pitch errors both at word and phrase levels were tolerated. Pitch raising and final syllable lengthening observed in the participants' production are known to be common speech behaviours in Japanese young people [5]. Although there are some generational and regional differences in the pitch pattern of the participants' Japanese, L2 speech should be evaluated in the light of average native listeners' judgements as closely as possible. This point should be taken into consideration when assessment tools are developed for Japanese pronunciation.

4. Conclusions

The importance of moraic timing has been recognized widely in Japanese phonology, whereas the importance of pitch has not been empirically tested. Due to the fact that Japanese is a pitch accent language, the importance of pitch in assessing pronunciation was anticipated. On the other hand, pitch accent patterns have regional variation in Japanese, and pitch raising at phrase final position is becoming common. It was hypothesized that Japanese native listeners might be more lenient about pitch deviation than previous studies suggest. In this study the influence of pitch errors on the performance judgement was not as strong as timing errors at word level. Items that had only pitch errors received relatively good scores. The findings of this study showed that moraic timing is the most urgent feature for L2 learners of Japanese to address. Pitch errors do affect the performance score, but not as significantly as do timing errors. Errors at phrase final positions, pitch raising and final syllable lengthening did not affect native listeners' judgement and were treated as natural speech behaviour.

Language instructors wish their learners' mastery of the target language and tend to point out every error of learners, and this teaching approach tends to be translated to and reflected in the production of assessment tools. However, not all prosodic features have to be brought to the level of a textbook model. The threshold of native listeners' acceptance requires mainly the accuracy of timing in Japanese. Their criteria, which count timing errors more heavily than pitch errors, should be taken into consideration for the development of computer assessment tools. This study was taken from native listeners from the region of non-Tokyo Japanese. The data from listeners in the Tokyo region would confirm the findings in future studies, although native speakers' perception towards foreign accents is not expected to differ significantly across regions. Even if such a difference exists, any native speakers' judgement should be treated as a valid assessment. This approach will be desirable in the global climate to accommodate intelligible L2 speech and foreign accent. Other languages with a different prosodic organization could prioritize a different prosodic component. Future study of prosodic judgement for L2 speech from other languages will create an interesting comparison.

5. Acknowledgement

This work was supported by Queensland Program for Japanese Education through Research Grant 2009.

6. References

[1] Anderson-Hsieh, J., Johnson, R. and Koehler, K. (1992) The relationship between native speakers judgements of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure *Language learning* 42: 4, 529-555

[2] Beckman, M. and Pierrehumbert, J..(1986) "Intonational structure in Japanese and English", *Phonology Yearbook.* 155-300, 1986.

[3] Cruttenden, A. (1986) *Intonation* Cambridge University Press

[4] Ingram, J. and Nguyen T. (2007) "Vietnamese accented English: Foreign Accent and Intelligibility judgement by listeners of different language backgrounds" TESOL in the internationalization of higher education in Vietnam"

[5] Inoue, F (1993) The significance of new dialects" Dialectologia et Geolinguistica (SIDG) 1 pp.3-27.

[6] Maier, A., et al. (2009) A language-independent feature set for the automatic evaluation of prosody Interspeech 2009, 600-603.

[7] Moriyama, T. (2004) On the extended pronunciation Speech and Grammar (ed.) Spoken Language Working Group, Kuroshio Tokyo

[8] Mouri, T., Hirose, K. and Minematsu, N. (2003) "Consideration on vowel durational modification for Japanese CALL system." *EUROSPEECH*, 3153-3156.

[9] Sato, T. (1995) "A comparison of phonemes and prosody in the evaluation of spoken Japanese" Japanese language education around the globe, vol.5, 139-154

[10] Sato, H. (1992) Production of prosody from phonetic point of view (Prosody no seisei) 58-65 *Gekkan Gengo*.

[11] Tajima, et al.(1997) Effects of temporal correction on intelligibility of foreign-accented English Journal of Phonetics, 25, 1-24

[12] Tsurutani, C. (2009). Intonation of Japanese sentences spoken by English speakers Interspeech 2009, 692-695

[13] Venditti, J., Maekawa, K. and Beckman, M. (2008) *Prominence marking in the Japanese intonation system – handbook of Japanese linguistics*