Foreign Accent of Brazilian Portuguese and Japanese –Perceptual Evaluation of Modified and Unmodified Speech–

Masahiko Komatsu [1] *and Fátima Tiemi Kimoto* [1,2]

[1] Department of Communication Disorders, Health Sciences University of Hokkaido, Japan
[2] Training Program for Japanese Descendants, Japan International Cooperation Agency
koma2@hoku-iryo-u.ac.jp; tiemifkimoto@gmail.com

Abstract

This paper examines the foreign accent in Brazilian Portuguese and Japanese. In the first experiment, Japanese speech samples of Japanese-descent Brazilians were compared to those of native Japanese speakers, as evaluated by Japanese, and in the second experiment, Portuguese speech samples of native Japanese speakers were compared to those of Japanese-descent Brazilians, as evaluated by Brazilians. In both experiments, detection of a foreign accent was more difficult in the stimulus with its spectral properties removed (representing prosodic features) than in the stimulus with its F0 flattened (representing segmental features), and was easiest in unmodified speech. Correlation between perception scores for modified and unmodified speech was found. Acoustic analysis suggested that speakers may bring their F0 characteristics of L1 into L2 speech, and showed that speaking rate was correlated with the perception scores. It concludes that foreign accents in both languages share some properties but also have some differences, possibly due to language differences, speakers' proficiencies, and listeners' backgrounds.

1. Introduction

This paper examines the foreign accent in Brazilian Portuguese (henceforth, Portuguese) and in Japanese. Foreign accents are thought to be caused by unnatural prosodic and segmental features. Although recent language pedagogy show increasing interest in prosody, it is not clear to what extent each of these features influences the perception of a foreign accent.

The importance of prosody in language and accent identification is known (see [1-3]). Experiments on various languages and accents have shown that people can identify languages and accents based solely on prosody. However, comparative research on foreign accents is limited.

Munro [3] applied the technique used in language and accent identification to the evaluation of the foreign accent in Mandarin speakers' English. He used lowpass-filtered speech to represent prosody, and found that people can detect a foreign accent based solely on prosody. He suggested that possible cues include speaking rates, F0 patterns, and the timing caused by the reduction of pronunciation. Interestingly, there was no correlation between the ratings of the foreign accent of lowpass-filtered speech and unmodified speech.

The present study extends the findings in [3] to other languages and speakers of different proficiency levels. On one hand, foreign accents in various languages are expected to share common properties because L2 speech in general tends to be slowly spoken and to have unnatural intonation and pronunciation, etc. On the other hand, the impacts of each of these elements may differ among languages because their difficulty in learning and tolerance differs. Also, speakers of different proficiency levels may have different characteristics. This study pays particular attention to Japanese descendants in Brazil (henceforth, Nikkei). Many of these people are native speakers of Portuguese but Japanese-accented, and are not fluent in Japanese. In this study, Nikkei's Portuguese and Japanese speech samples are compared to those spoken by native Japanese speakers.

2. Purpose and design of the experiments

The purpose of the experiments is two-fold. First, this study discusses whether or not the foreign accent in Portuguese and the foreign accent in Japanese have the same characteristics. The whole experiments consist of perceptual evaluation of Japanese speech and Portuguese speech. In Experiment 1, Japanese speech samples spoken by Nikkei people (L2, test group) were compared to samples spoken by native Japanese speakers (L1, control group). The perceptual evaluation was conducted by native Japanese speakers (Listeners). Experiment 2 primarily tested Portuguese speech samples spoken by native Japanese speakers (L2), but it also investigated whether a foreign accent is detectable in Nikkei's Portuguese speech (L1). They were evaluated by native Portuguese speakers, who are non-Nikkei and Nikkei Brazilians (Listeners).

Second, the study discusses whether or not the perceptual ratings of modified speech and unmodified speech are correlated. The correlation between acoustically manipulated and unmodified speeches is important because it indicates the impact of the controlled acoustic properties. Although no correlation was found in a previous study [3], it is not clear whether any correlation exists in other languages and speakers of different proficiency levels.

To this end, our experiments used three types of stimuli: *prosody*, *flat*, and *unmodified*, representing prosodic features, segmental features, and both of them, respectively. The modification of speech was based on the idea that prosodic and segmental features are acoustically related to the source and filter components, respectively (see [2,4] for theoretical and technical details). Roughly, in phonetic terms, the source corresponds to the laryngeal source sound, and the filter corresponds to the effect of the vocal tract shape. The *prosody* stimulus simulates the source and lacks spectral information. It sounds like muffled speech and is totally unintelligible. It represents the prosodic features contained in speech. The *flat* lacks F0 information in the source component. It is a monotonous sound whose pitch is flattened. It contains segmental features but loses prosodic features such as intonation, pitch accent, etc. The *unmodified* retains all information.

3. Experiments

3.1. Procedure of Experiment 1: Japanese speech

The recording of Japanese speech by 10 Nikkei (5 males and 5 females; mean age 24.5, range 22-28) and 4 native Japanese speakers (2 males and 2 females; mean age 23.3, range 19-30) was used in Experiment 1. The Nikkei speakers started learning Japanese at the age of 3-18 and learned it for 2-12

years, and their total period of living in Japan ranged 5 months - 4 years. Eight Nikkei's L1 was Portuguese. Although the other two declared that their L1 was Japanese, they were grouped with the other Nikkei, considering their language proficiency. The native Japanese speakers had never lived abroad.

These participants read "The north wind and the sun" in Japanese (8 sentences) [5]. Sentences that did not include clear epenthesis, syncope, repetition, and alteration of phonemes were selected. Four out of the 8 sentences (41-, 17-, 35-, 16-mora-long respectively) were selected. In total, 4 sentences × 14 speakers = 56 samples were used in the experiment.

These selected sentences were extracted from the recording (*unmodified*), from which two types of modified stimuli were made (*prosody* and *flat*). All the recordings were done using Marantz PMD670 with condenser microphones in quiet rooms. The stimuli were created using Praat (Version 4.6.34); CoolEdit2000 was used to suppress background noise before processing. All stimuli were sampled at 20 kHz, 16 bit.

The perceptual evaluation was conducted on the web, and listeners were asked to wear headphones if available. The experiment was divided into three sections of *prosody*, *flat*, and *unmodified*, and they were asked to proceed in this order. In each section, a training session of 5 samples, which were not used in the test, preceded the test session that consisted of the 56 samples. The presentation order of the stimuli was randomized within each section for each listener. The experiment took approximately 30 min.

For each stimulus, listeners first listened to the sound; they were allowed to listen repeatedly if they wished to. Then, they selected one from *Definitely foreign-accented*, *Probably foreign-accented*, *Probably spoken by a native speaker*, and *Definitely spoken by a native speaker* in the *prosody* and *flat* sections. In the *unmodified* section, listeners rated each stimulus on a 9-point scale: 1 (*Totally unnatural*), 3 (*Unnatural*), 5 (*Undecided*), 7 (*Natural*), 9 (*Totally natural*); and they also checked unnatural elements: "Pronunciation," "Intonation, etc. (height of voice)," "Rhythm," "Speed," "Dysfluency." The use of the 4-point and 9-point scales for the modified and unmodified stimuli respectively was decided upon based on the literature [3,6].

Forty-six native Japanese speakers (14 males and 32 females; mean age 25.3, range 19-51) participated as listeners. They were students and faculty in linguistics, communication disorders, and related fields.

3.2. Procedure of Experiment 2: Portuguese speech

Experiment 2 was carried out in the same manner. The Portuguese speech samples from the same 10 Nikkei who participated in Experiment 1 were used. Also, Portuguese speech samples spoken by 10 native Japanese speakers (5 males and 5 females; mean age 38.7, range 28-61) was recorded. The latter started learning Portuguese at the age of 12-28, learned it for 2 months - 6 years, and had lived in Brazil for 2-10 years.

The participants read "The north wind and the sun" in Portuguese (5 sentences) [5]. All 5 sentences (40-, 50-, 50-, 31-, 24-syllable-long respectively) were used, but some speakers failed to produce sentences without clear epenthesis, syncope, repetition, and alteration of phonemes. In total, 5 sentences × 20 speakers – 10 failed = 90 samples were used.

The perceptual evaluation was conducted on the web, but several listeners did it using a loud speaker and answer sheets due to network trouble. Some did the three sessions successively, and some did them at different times. In the training session, 5 samples were randomly selected and used, which were also used in the test session. The test session consisted of the 90 samples. Approximately an hour was needed for the experiment, but it often took longer due to network conditions.

Eleven people (4 males and 7 females; mean age 31.5, range 22-69) participated as listeners. Five of them were native Portuguese speakers who live in Brazil. The other 6 were Nikkei who knew the speakers personally; even some speakers were included in the listeners.

3.3. Results: Perceptual scores

Table 1 shows the scores of perceptual evaluation in the two experiments. It indicates mean scores along with SD in parentheses. The means and SDs were calculated among stimulus tokens: i.e. first, for each token, responses from all listeners were averaged, and then for each stimulus set, the mean and the SD of the tokens' scores were calculated. In *prosody* and *flat*, the scale ranges from 1 (*definitely foreign-accented*) to 4 (*definitely spoken by a native speaker*). In *unmodified*, the scale ranges from 1 (*totally unnatural)* to 9 (*totally natural*), with the scores linearly converted to 4-point displayed below. In either experiment, the difference between the scores of the two speaker groups gets greater as it goes from *prosody* to *flat* and *unmodified*.

Table 1: *Scores of perceptual evaluation*

Japanese speech

	Speakers			
Stimulus	Nikkei (*N*=40)		Japanese (*N*=16)	
Prosody	2.17	(0.54)	3.51	(0.38)
Flat	1.70	(0.55)	3.53	(0.30)
Unmodified	3.74	(1.47)	8.47	(0.37)
(4-point)	1.66	(0.65)	3.76	(0.16)

Portuguese speech

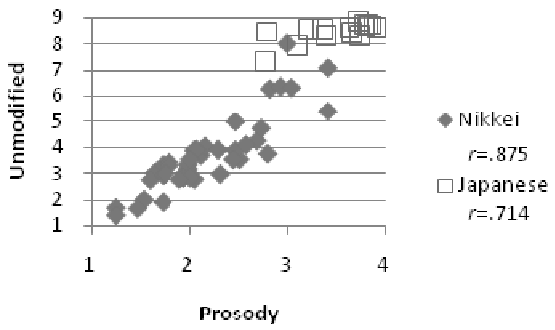
	Speakers			
Stimulus	Nikkei (*N*=47)		Japanese (*N*=43)	
Prosody	3.37	(0.44)	1.85	(0.60)
Flat	3.85	(0.21)	1.14	(0.17)
Unmodified	8.77	(0.29)	2.25	(0.65)
(4-point)	3.90	(0.13)	1.00	(0.29)

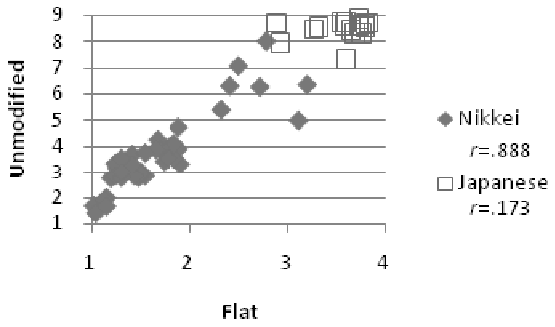
The relationships between the types of stimuli are shown in Fig. 1. Each mark represents a stimulus token's scores averaged over all listeners. In Fig. 1a, Nikkei's Japanese speech shows a strong correlation between the scores of *unmodified* and those of both *prosody* and *flat*. The correlation for native Japanese speakers in Fig. 1a is not reliable because the number of samples is small. Comparing the two graphs in Fig. 1a, Nikkei's scores are lower in *flat* than in *prosody* while Japanese' scores reside in the same range. As a result, the two speaker groups are better separated in *flat* than in *prosody*. In *unmodified*, they are the best separated.

Fig. 1b shows that there was no overlap in the ranges of Nikkei and Japanese in *flat* and *unmodified*. It is clear that Japanese' scores are lower and Nikkei's scores are higher in *flat* than in *prosody*. This means that the detection of a foreign accent is much easier in *flat* than in *prosody*. This tendency is clearer than in Japanese speech (Fig. 1a).

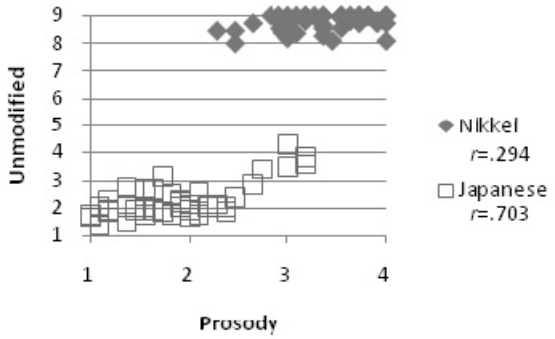
Table 2 shows the percentages of listeners who checked each unnatural element in the *unmodified* session. The indicated figures are the average over stimulus tokens. For example, 44.5 % of all listeners judged that Nikkei's Japanese speech included unnatural pronunciation on average. Overall, Nikkei's Japanese speech was less frequently perceived as unnatural than Japanese' Portuguese speech. In both Nikkei's Japanese speech and Japanese' Portuguese speech, pronunciation and intonation were the most frequently

a. Japanese speech





b. Portuguese speech



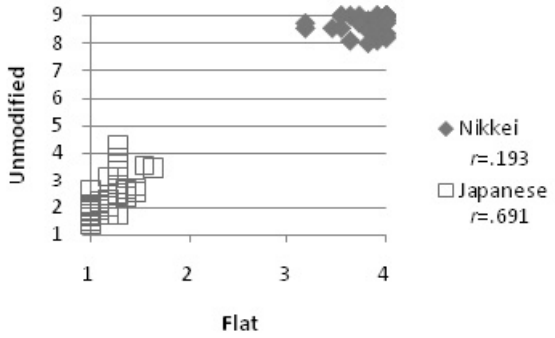


Figure 1: *Relationship between stimulus types.*

Table 2: *% unnatural elements.*

	N	Pron.	Into.	Rhythm	Speed	Dysflu.
Japanese speech						
Nikkei	40	44.5	61.8	31.6	18.0	20.1
Japanese	16	3.4	3.9	0.4	1.1	0.3
Portuguese speech						
	N	Pron.	Into.	Rhythm	Speed	Dysflu.
Nikkei	47	3.3	2.3	2.1	0.2	0.0
Japanese	43	79.5	72.9	55.6	25.2	30.7

Table 3: *F0 mean and SD in semitone.*

		N	Mean		SD	
Japanese speech						
Nikkei	M	20	1.3	(1.6)	2.1	(0.6)
	F	20	12.3	(1.5)	2.1	(0.9)
Japanese	M	8	6.0	(2.5)	2.7	(0.6)
	F	8	12.8	(1.9)	2.9	(0.7)
Portuguese speech						
		N	Mean		SD	
Nikkei	M	23	1.7	(1.7)	1.9	(0.5)
	F	24	12.6	(1.5)	2.1	(0.8)
Japanese	M	21	3.9	(1.4)	2.6	(0.7)
	F	22	12.0	(1.8)	2.4	(0.3)

Table 4: *Speaking rate.*

	N	Morae/s		*r*		
Japanese speech				Pros.	Flat	Unmod.
Nikkei	40	6.06	(1.37)	0.73	0.78	0.84
Japanese	16	7.94	(0.87)	-0.15	-0.62	0.06
Portuguese speech						
	N	Syllables/s		*r*		
				Pros.	Flat	Unmod.
Nikkei	43	6.70	(0.69)	0.01	-0.11	-0.03
Japanese	47	3.97	(0.62)	0.64	0.36	0.49

perceived as unnatural, rhythm was less, and speed and dysfluency were the least perceived as unnatural.

3.4. Acoustic analysis: F0 and speaking rate

Table 3 shows the F0 analysis of the sentences. First the mean F0 and SD in each sentence were calculated, and then their means and SDs over sentences and speakers were calculated (SDs are in parentheses). That is, "mean" represents the average F0 of a speaker group, and "SD" is related to the average F0 range. The unit is semitone re 100 Hz. Both in Japanese speech and Portuguese speech, Japanese males' mean F0 is slightly higher than Nikkei males', and Japanese' SD is larger than Nikkei's. This means that the speakers' characteristics do not change when they speak different languages.

Table 4 shows the mean speaking rate and SD in parentheses, followed by *r* of the speaking rate with the perception scores. In Japanese speech, Nikkei are slower than Japanese, and they show high correlations with scores for all stimuli. In Portuguese speech, Japanese are definitely slower, but their correlations are lower than Nikkei's Japanese speech, especially in *flat* and *unmodified*. This suggests that other factors are involved.

4. Discussion

The perceptual results of Japanese and Portuguese speech showed general tendencies in common. In both languages, the distinction between L1 and L2 speakers was clearer in the order of *unmodified*, *flat* (representing segmental features), and *prosody* (representing prosodic features). As for the L2 speech (Japanese speech spoken by Nikkei and Portuguese speech spoken by Japanese), both *prosody* and *flat* scores had a correlation with *unmodified* scores. These results indicate that segmental features have more impact on foreign accents than prosodic features although both have one. Elements

perceived as unnatural in L2 speech were common, too. In both languages, pronunciation and intonation were most frequently perceived as unnatural, rhythm was less, and speed and dysfluency were the least frequently perceived so.

The results also showed characteristics particular to specific languages and speaker groups. Segmental characteristics are thought to be influential in the foreign accent in Portuguese speech spoken by Japanese. In the perceptual test of Portuguese speech (Experiment 2), the listeners were able to distinguish L2 speakers (Japanese) from L1 speakers (Nikkei) better when presented with *flat* and *unmodified* samples, where segmental features was preserved, than when presented with unintelligible *prosody*, where segmental features were removed.

Auditory inspection of speech samples endorses this claim. Portuguese speech spoken by Japanese had some peculiar characteristics regarding not only the prosodic aspects such as intonation but also other differences especially regarding mispronunciation of unfamiliar consonants such as labiodental fricatives /f, v/, dental plosives /t, d/ followed by the vowel /u/, voiced alveolar fricative /z/ especially when followed by the vowels /i, u/, voiced palatoalveolar fricative /ʒ/, and alveolar and palatal lateral approximants /l, ʎ/. The neutralization of word-final post-stress vowels /e, o/ to the high vowels /i, u/ was observed in the spontaneous speech (recorded but not used for the experiment) but it hardly occurred in text reading, although they are characteristic of both natural Portuguese spontaneous speech and reading. These observations conform to what is generally known.

Japanese accent in Portuguese speech spoken by Nikkei was also detected in Experiment 2. In general, Portuguese spoken by Nikkei Brazilians is recognized as Japanese accented by non-Nikkei Brazilians. In the perceptual tests of Portuguese speech, while Nikkei speakers' *flat* and *unmodified* scores were high, their *prosody* scores spread to lower values. This suggests that prosodic features can be considered as one of the characteristics that determine the Japanese accent in Nikkei's Portuguese speech.

The perceptual scores for Japanese speech samples spoken by Nikkei showed different characteristics (Experiment 1). Although the *flat* scores were lower than *prosody*, the samples whose scores were above 5 in *unmodified* remained good scores both in *prosody* and *flat*. This suggests that segmental features are less influential in fluent speech.

Acoustic analysis showed that F0 was characteristic of speaker groups rather than languages. Bringing the F0 characteristics of L1 into the L2 speech may have resulted in foreign accents. It also showed that the speaking rate is lower in L2 speech than in L1 speech. Its correlation with perceptual scores was found in L2 speech, but the correlation was not strong in *flat* and *unmodified* Portuguese speech. The influence of segmental features may have overlaid the effect of speaking rate on low-scored L2 Portuguese samples.

Note that although the peculiarities of languages and speaker groups were found, we need to be careful in attributing it to specific languages or speakers. Considering the score distribution of *unmodified* samples and speakers' background, it can be assumed that speakers' proficiency level was higher in the order of Japanese speech spoken by Japanese, Portuguese speech spoken by Nikkei, Japanese speech spoken by Nikkei, and Portuguese speech spoken by Japanese. Obtained results may be interpreted as either of the effects of language or the speakers' proficiency. For example, the difference of L2 Japanese and Portuguese may be caused by the difference between languages or due to the higher proficiency of Nikkei's Japanese than native Japanese' Portuguese.

5. Conclusions

The foreign accent in Portuguese and the foreign accent in Japanese shared some properties. In both languages, the distinction between L1 and L2 speakers was clearer in the order of *unmodified*, *flat* (representing segmental features), and *prosody* (representing prosodic features). The results indicated a greater influence of segmental features than prosodic features in the perception of foreign accent, especially in less fluent speech. Elements that are frequently perceived as unnatural were common, too. Also, it was suggested that bringing the F0 characteristics of L1 into the L2 speech may have resulted in foreign accents. L1 speakers' speaking rate was faster than that of L2 speakers in both languages.

Meanwhile, characteristics peculiar to specific languages or speaker groups were found. They include the importance of segmental features in Portuguese speech spoken by Japanese, the importance of prosodic features in Japanese-accented Portuguese speech spoken by Nikkei, less importance of segmental features in Japanese speech spoken by Nikkei than in Portuguese speech spoken by Japanese, and lower correlations of speaking rate in Portuguese speech spoken by Japanese than in Japanese speech spoken by Nikkei. Some of these differences may be resolved into difference in speakers' proficiency rather than language difference.

In the present study, correlations between the scores of modified speech and unmodified speech were found. This suggests that both prosodic and segmental features contribute to foreign accents perceived in unmodified speech. This finding goes against the previous study [3]. Considering that the relationship pattern was different between Experiments 1 and 2, the correlation is thought to vary depending on the languages involved and the speakers' proficiency. Comparing the results of *prosody* and *flat*, correlation is thought to be lost when the L1 and L2 distinction is easy.

Another factor we may need to consider is the listener. Portuguese and Japanese speeches were evaluated by different listeners in our experiments. This may not be fair because some may be more sensitive to foreign accents than others. However, foreign accents are something decided upon by listeners: the same acoustic characteristics may be perceived as a foreign accent by the speakers of some languages but not by other language speakers.

6. References

[1] Komatsu, M., 2007. Reviewing Human Language Identification. In *Speaker Classification II,* Müller, C. (ed.). Heidelberg, Germany: Springer, 206-228.

[2] Komatsu, M., 2007. *Acoustic constituents of prosodic typology*. Tokyo: Yushodo.

[3] Munro, M. J., 1995. Nonsegmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition* 17, 17-34

[4] Komatsu, M.; Arai, T.; Sugawara, T., 2004. Perceptual discrimination of prosodic types. *Proceedings of Speech Prosody 2004,* 725-728.

[5] International Phonetic Association, 1999. *Handbook of the International Phonetic Association.* Cambridge: Cambridge University Press.

[6] Munro, M.; Derwing, T., 1994. Evaluations of foreign accent in extemporaneous and read material. *Language Testing* 11, 253-266.

* This research is supported by JICA's Training Program for Japanese Descendants and HSUH Kotaisa grant. The authors are grateful to the participants in the experiments, Tsuneyuki Oikawa, Mark Micallef, Magda Saiuri Kimoto, and the anonymous reviewer.