euronoise

Perception of Japanese consonants by native speakers of American English

Takeshi Nozawa

Ritsumeikan University, College of Economics, 1-1-1 Nojihigashi, Kusatsu, 525-8577 Shiga, Japan
t-nozawa@ec.ritsumei.ac.jp

Four native speakers of Japanese produced Japanese multi-syllable words and non-words, which include stops, fricatives, affricates, nasals, glides and the liquid. Twelve native speakers of American English, who had never learned Japanese, heard their utterances and spelled out what they heard in English alphabet. What was of interest was whether the listeners would perceive Japanese consonants in a way they are transcribed in English alphabet. The results revealed that Japanese stops were equated with generally the phonetically closest English stops, but voiceless stops were more likely to be equated with voiced stops of the same place of articulation than the other way around. Among voiced stops, /b/ and /d/ respectively were equated with /v/ and /l/. The word-initial /ts/, though words like "tsunami" have become part of English, was predominantly equated with /s/. The Japanese liquid, which is usually transcribed as /r/, was predominantly equated with /l/ or /d/ rather than /r/. This agrees with results of previous research that demonstrate that English /r/ is more dissimilar from Japanese /r/ than English /l/.

1 Introduction

This paper attempts to investigate perceived similarity of Japanese and American English consonants by looking at how Japanese consonants are mapped into American English consonant categories. Japanese has a smaller consonant inventory than English, but allophonic variations make the following possible.

stops	[p] [b] [t] [d] [k] [g]
fricatives	[ɸ] [β] [s] [ç] [ɕ] [h] [z]
affricates	[ts] [cɕ] [dz] [jʑ]
liquids	[ɾ]
nasals	[m] [n] [ŋ]
glides	[ɯ] [j]

Table 1: Japanese consonants

Some of these can be treated as variants of the same phonemes, but because in this present study, the perception of Japanese consonants to native speakers of American English is of first importance, whether [s] and [ɕ], for instance, are variants of a single phoneme or they should be described as /s/ and /š/ will not be discussed here. Mora consonants /Q/ and /N/ are not included in this study.

Japanese has a large number of borrowings from English. Some of the phonemic contrasts are lost when English words are imported to Japanese. Stops are usually transcribed as their respective counterparts, but pronunciation and transcription of English fricatives in Japanese are rather complicated. Both /b/ and /v/ are merged into a single category /b/. /s/, /ʃ/ and /θ/ are all pronounced as [ɕ] before /i/, in other words, "sin" "shin" and "thin" are all [ɕin]. Both /h/ and /f/ are pronounced as [ɸ] before /u/. In fact, Japanese [ɸ] is usually transcribed as /f/ in English as in *futon* and *tofu*. English liquids /r/ and /l/ are merged into a single category /ɾ/. These pronunciations and transcriptions may indicate Japanese speakers' perceptual and productive difficulty in differentiating English consonants

These pronunciations and transcriptions do not necessarily indicate how Japanese listeners perceive English consonants in terms of Japanese consonant categories. Guion, et al. (2000) performed a perceptual assimilation experiment of English consonants /b/, /v/, /w/, /θ/, /t/, /s/, /r/, /l/ uttered in a /Ca/ frame in terms of Japanese categories [1]. Their results show that /b/, /w/, /t/, /s/ were equated with their Japanese counterparts at a high mean percent identification, whereas /θ/ was equated with the Japanese /s/ category as a less ideal exemplar of /s/ and /ɸ/. /r/ and /l/ were equated with /ɾ/ and /wɾ/ at almost the same mean percentage with /l/ being the better exemplar of /ɾ/. /v/ was equated with /v/ (their choices included /va/), so their Japanese listeners discerned some difference between /b/ and /v/.

2 Experiment

2.1 Stimuli

Stimuli were extracted from utterances by four female native speakers of Japanese. They were all recruited in Shiga, Japan. They read aloud Japanese words and non-words on a word list in a carrier sentence "Tsugi wa ___ desu (The next is ____)." Their utterances were digitally recorded on a DAT recorder, and digitized at the sampling rate of 44100Hz, and edited and normalized at the peak intensity. The word list consisted of 70 words. Thus, 280 stimuli (70 words x 4 talkers) were created.

2.2 Subjects

Subjects were native speakers of American English recruited in Auburn, Alabama, USA. None of them had a self-reported hearing problem. None of them had been to Japan nor had learned Japanese.

2.3 Procedure

The stimuli were randomized for sequence in order to be presented to subjects. The stimuli were played on a lab computer, and each subject heard one stimulus per trial. The subjects were told that they would hear Japanese words and that they were supposed to spell out those Japanese words in English alphabet. Eight-page answer sheets were provided, and the subjects wrote their answers on them after hearing each stimulus. They were allowed to hear each stimulus as many as times as they wanted. When they were ready to hear the next stimulus, they clicked on a box, "NEXT", on the computer screen. The next stimulus was played 1,000ms after they clicked.

3 Results

The subjects' answers were analyzed according to how they were spelled out. English letters that represent consonants are consistent, so it was not difficult to decide which consonant they heard. For instance, if the letter "r" was included in their answer, then it was understood that they heard [r] in that stimulus. Some English letters represent more than one consonant. The letter "c" represents [k] and [s], but these two pronunciations are almost in a complimentary distribution. When it was followed by the letters "a", "o", "u" and other consonant letters, or it was in a syllable final position, it was counted as [k]. The letter "s" represents [s] and [z], but in most cases, it was counted as [s]. Only when it was associated with English words like *choose*, it was counted as [z]. The "th" is more problematic. It represents both voiceless [θ] and voiced [ð]. In such cases, it was counted as voiceless if the Japanese consonant that the subject attempted to transcribe was voiceless. The letter "n" was counted as [ŋ] if it was followed by "k" or "g". Sometimes handwritten answers were difficult to decipher, and sometimes the target consonant was missing, (The subjects may not have heard the consonant).

3.1 The liquid

A number of studies have been done on Japanese speakers' inability to differentiate English /r/ and /l/, and recent studies show that Japanese listeners find English /r/ more dissimilar from Japanese liquid (See Aoyama et al. 2004 [2]), which is usually transcribed as "r" in English.

Of interest here was whether American English listeners would perceive the Japanese liquid as /l/ as Japanese listeners perceptually assimilate English /l/ to Japanese liquid, or they perceive the liquid as /r/ as the conventional transcription shows. Japanese liquid is basically an apico-alveolar tap [ɾ]. Shirota (1995) [3] mentions that the Japanese liquid can be realized as a loose stop [ɹ] in a word-initial position. He also mentions that it can be realized as a lateral [l l'] in word-initial position, and as a trill [r r'] in vulgar speech. He states that it is realized as a flap when it occurs noninitially. Though Japanese speakers may not be normally aware of this, the liquid is realized as several different allophones, depending on individuals, speech level, and the position it occurs.

Of all the 70 words on the list, 18 words contained the liquid. Of the 18 words, nine contained the liquid word-initially, and the other nine noninitially. Table 4 shows major responses to the liquid in two different positions. In both positions, the listeners' responses are not consistent. The most frequent responses do not reach 50%.

/l/ seems to be the closest English consonant to /ɾ/. This agrees with Japanese listeners' responses in Aoyama et al. (2004)[2]. /r/ is the fourth choice in the word-initial position, and the third noninitially, so contrary to the common custom of transcribing the Japanese consonant as /r/ in English, /r/ is not perceptually the closest English consonant to Japanese /ɾ/. Large talker-to-talker differences are observed. We will discuss talker effects later.

Responses	word initial	noninitial
/l/	46.9 %	47.4 %
/d/	8.3 %	21.1 %
/r/	2.8 %	17.4 %
/b/	7.6 %	0.5 %
/t/	0.9 %	4.9 %

Table 4: Mean percent identification of Japanese /ɾ/

3.2 Voiceless stops

Vance (1987) [4] states that "the bilabials [p b] and the velars [k g] are essentially identical to their English counterparts, but [t d] are somewhat different (pp. 17-18)". Though Japanese [t d] are not articulated at exactly the same place as English counterparts, English [t d] are still the closest English consonants to Japanese [t d]. Japanese voiceless stops are, however, different from English counterparts in VOT. There seems to be a disagreement among researchers as to whether Japanese voiceless stops are aspirated or not, but they show far shorter VOT than English equivalents. Three words contain /p/, and one of the words *pompu (pump)* contains two /p/ tokens. Five words contain /t/ either initially or noninitially, and so 20 /t/ tokens were identified. Twenty-nine words contain /k/, and seven of them contain /k/ twice, so 144 /k/ (36 tokens x 4 talkers) tokens were identified. Table 5 shows three most common responses to /p/, /t/, /k/.

Responses to /p/	Responses to /t/	Responses to /k/
/p/ 58.6%	/t/ 56.3%	/k/ 78.7%
/b/ 14.7%	/d/ 20.4%	/g/ 11.6%
/f/ 6.8%	/k/ 1.7%	/tʃ/ 0.9%

Table 5: Percent identification of Japanese voiceless stops

Voiceless stops were generally equated with their respective equivalent American English stops and their corresponding voiced counterparts. /k/ was equated with the corresponding English stop at a higher rate than /p/ and /t/ probably because the consonant inventory is less crowded in the back of the oral cavity. Bilabial and alveolar consonants can be identified with dental, labiodental consonants. For instance, /p/ was identified with /f/ in 6.8% of all the responses.

3.3 Voiced stops

Japanese voiced stops were generally identified with their corresponding voiced stops in English, but they were equated with voiceless stops at the same place of articulation far less than the other way around. This is probably because Japanese voiceless stops do not have as large VOT values as their English equivalents. As far as VOT is concerned, they fall between English voiced and

voiceless stops, but Japanese voiced stops are often prevoiced and so it is unlikely that they sound close to aspirated English stops. As shown in Table 6, /b/ is identified more often as /v/ than as /p/, and /d/ was identified more often as /l/, /b/, /ð/ than as /t/. This may be an indication that the closure is not as tight as in English counterparts, and so they may not be perceived as stops by native speakers of English.

Responses to [b]	Responses to [d]	Responses to [g]
/b/ 67.4%	/d/ 56.3%	/g/ 60.4%
/v/ 27.7%	/l/ 18.8%	/k/ 20.8%
/p/ 5.7%	/b/ /ð/1.7%	/w/ 3.7%

Table 6: Percent identification of Japanese voiced stops

3.4 Voiceless fricatives

Five voiceless stops [s], [ɕ], [h], [ç], [ɸ] were involved. Twelve words include [s] and one of them includes it twice, so 52 [s] tokens were identified. Eleven words contain [ɕ], and so 44 [ɕ] tokens were identified. Only two words contain [h], and so eight [h] tokens were identified. Three words contain [ç] and so 12 [ç] tokens were identified. Five words contain [ɸ] and one of the words contain [ɸ] twice, so 24 [ɸ] tokens were identified. The mean percent identification is shown in Table 7. [s] and [h] are equated with /s/ and /h/ at a relatively high percentage, whereas the identification of the other three fricatives is more dispersed. This may indicate that [s] and [h] are perceptually closer to their closest English consonants than the other three fricatives are. Vance (1987) [4] mentions Japanese [s z] are essentially identical to their English counterparts, and the results here seem to support his account. [ʃ] may be the closest English consonant to [ɕ], but [ʃ] and [ɕ] are not phonetically identical, and there seem to be discernible differences between them. [ʃ] is a voiceless postalveolar fricative, but [ɕ] is a voiceless alveolo-palatal fricative. Also lip-rounding is involved in [ʃ], but not in [ɕ].

Fewer words that contain [h], [ç], [ɸ] were on the list, so it may not be safe to make a conclusion here, but still phonetic differences can explain the results. [h] in both languages is a glottal fricative, and so their perceived dissimilarity should be small. According to Vance (1987) [4], [ç] is heard in some dialects of English as in a word like *huge*, so it should be perceived as a (probably less ideal) exemplar of /h/. Of the three words that contain [ç], the consonant in *kyohi* (refusal) is equated with /h/ at 93.8%. In the other two words *Hyuga* (one Japanese region) and *hyaku* (hundred), [ç] occurs before a glide /j/, which may have affected the result. In 47.9% of responses to *Hyuga* and in 29.2% of responses to *hyaku*, [ç] is equated with /h/. In 39.6% of responses to *Hyuga* and in 20.8% of responses to *hyaku*, no consonant that corresponds to [ç] is found. The sequence of [çjV] must have affected the listeners' responses. In English /j/ can occur before all the vowels, but when another consonant precedes it, it can only occur before /u/. This is why words like *Tokyo* and *Kyoto* are pronounced [tóʊkioʊ] and [kioʊtoʊ] rather than [tóʊkjoʊ] and [kjóʊtoʊ]. Apparently, the presence of /j/ affected the results.

Responses to [s]	Responses to [ɕ]	Responses to [h]	Responses to [ç]	Responses to [ɸ]
/s/ 91.2%	/ʃ/ 58.3%	/h/ 80.2%	/h/ 56.9%	/h/ 45.5%
/z/ 3.8%	/s/ 30.9%	/k/ 5.2%	/k/ 6. 3 %	/f/ 21.2%
/ʃ/ 2.6%	/tʃ/ 4.9%	/n/ /j/ 2.1%	/ʒ/ 4.2%	/p/ 15.6%

Table 7: Percent identification of Japanese voiceless fricatives

[ɸ] is a voiceless bilabial fricative, but is usually transcribed as /f/ as in *Fuji*. As shown in Table 7, 45.5% of the responses are /h/, and /f/ accounts for 21.2% of all the responses. In the conservative variety, [ɸ] only occurs before /u/, but in recent borrowings, it is possible before all the other vowels, too. Three of the five words that contain [ɸ] are loan words from English. The listeners' responses are different, depending on whether [ɸ] occurs before /u/ or other vowels. [ɸ] is equated with /f/ only when it is followed by vowels other than /u/. Because these words are borrowings from English, it may be possible that they sounded like their original English words to the listeners, but very rarely are these words identified as their original English words, so apparently the results do not stem from lexical influence but from phonetic or auditory influence.

3.5 Voiced fricatives

There is no contrast between [z] and [dz]. Vance (1987) [4] states "the words *asa* 'morning' and *aza* 'village' contain alveolar fricatives [s z] that are essentially identical to their English counterparts." Many researchers state that [dz] is more common than [z]. Shirota (1995) [3], for instance, says [z] is only possible in an intervocalic position. Saito (1997) [5] states that in casual speech [z] is possible in an intervocalic position. There is no contrast between [ʑ] and [jʑ], either. Vance (1987) [4] states standard speakers find [ʑ] difficult or impossible to produce when asked. Saito (1997) [5] states [ʑ] is possible only intervocalically. Phonologically [z] and [ʑ] respectively are voiced counterparts of [s] and [ɕ], but [dz] and [jʑ] seem to be more commonly used. Five words contain [z] or [dz], and so 20 [z] or [dz] tokens were identified. Six words contain [ʑ] or [jʑ], so 24 [ʑ] or [jʑ] tokens were identified.

Responses to [z][dz]	Responses to [ʑ][jʑ]
/s/ 37.1 %	/z/ 37.2 %
/z/ 36.7 %	/dʒ/ 28.1 %
/d/ 6.3 %	/ʃ/ 9.4 %

Table 8: Percent Identification of [z]/[dz] and [ʑ]/[jʑ]

The majority of the responses to [z]/[dz] were fricatives /s/ or /z/. In no instance, was it identified as affricate [dz].

This may be because in English [dz] is not possible as a single morpheme. The combination is possible in such words like *cards* or *bids*. An affricate is a stop consonant followed by audible friction, and so it has the quality of a stop and a fricative. Acoustic analyses are necessary to carefully examine whether the consonant in each stimulus was produced as an affricate or a fricative, but the listeners in this study seemed to be affected by the frication rather than by the closure. /d/ accounts for 6.3% of all the responses, but in the word-initial position, /d/ accounts for 13.5% whereas in non-initial positions, it accounts for only 0.6%, so the closure seems to be tighter and more audible in the word-initial position, which agrees what Shirota (1995) [3] and Saito (1997) [5].

No clear reasons have been found why the consonant was perceived as /s/ as often as /z/. When a listener spelled out "s" for the consonant, the response was counted as /s/ unless it denoted a real lexical item like *deserve* in which the letter "s" stands for /z/. Because the letter "s" can stand for /z/ in English, a further study is necessary to confirm whether Japanese [dz] can be perceived as /s/ as often as /z/ by native speakers of English.

[ʑ]/[jʑ] is equated with /z/ more often than with /dʒ/. This is largely because in two of the words that contain the consonant, *shujin* (husband), *shuujin* (prisoner), it was predominantly equated with /z/. The second syllable of these two words was perceived as /zɪŋ/. Listeners may have associated Japanese words with English words they are familiar with. We will discuss the lexical effects later.

3.6 Affricates

There is no contrast between [z] and [dz], and between [ʑ] and [jʑ], but a contrast exists between [s] and [ts], and between [ɕ] and [cɕ], so [ts] and [cɕ] should be treated separately from [s] and [cɕ]. Five words contain [ts] and two words contain [cɕ].

In English, [ts] is usually spelled "ts" as in *tsunami*, and [cɕ] is spelled "ch" as in *hibachi* or *Ichiro*. Though words like *tsunami* has become part of English, [ts] is equated with /ts/ far less often than with /s/. Four of the five words that contain [ts] have [tz] in the word-initial position. [ts] in the word-initial position is equated with /s/ in 89.1% of all the responses. The other word, *katsu* (win), has /ts/ intervocalically. [ts] in *katsu* is equated with /ts/ in 47.9% of all the responses. In English /ts/ as a single morpheme is rare, and limited to borrowings like *pizza*, and it seems that /ts/ in the word-initial position is not established as much as /ts/ in other positions.

Responses to [ts]	Responses to [cɕ]
/s/ 71.7 %	/t/ 28.1 %
/ts/ 9.6 %	/tʃ/ 21.9 %
/z/ 9.2 %	/ʃ/ 13.5 %

Table 9: Percent Identification of [ts] and [cɕ]

Phonetic difference between English [tʃ] and Japanese [cɕ] is larger than between [ts] in the two languages. [tʃ] is voiceless palato-alveolar affricate, but [cɕ] is voiceless alveolopalatal affricate. This may be a reason that [cɕ] is equated with [tʃ] slightly more than 20% of all the responses. One of the two words that contain [cɕ], *chokin* (saving), has the consonant in the word-initial position, and the other word, *shachihoko* (statue of an imaginary animal) has it intervocalically. In both words, [cɕ] is equated with /tʃ/ in about 20% of all the responses.

3.7 Nasals

Japanese has two nasal phonemes /m/ and /n/, and [ŋ] is only possible as an allophone of /g/. Both /m/ and /n/ can only occur before vowels and /j/, so they never can be followed by velar consonants, which means they are not realized as /ŋ/. Listeners' responses are counted as /ŋ/ when the letter "n" is followed by "k" or "g". Four words contain /m/, and five words contain /n/ either initially or noninitially. As shown in Table 10, Japanese /m/ and /n/ are equated with their respective English counterparts in about 70% of the instances. /n/ in *nieyu* (boiled water) is identified as /n/ only in 29.1% of instances. As shown in Table 10, nasality is relatively easily identified.

Responses to /m/	Responses to /n/
/m/ 70.8 %	/n/ 67.9 %
/n/ 19.3 %	/ŋ/ 15.4 %
/h/ 2.1 %	/m/ 7.9 %

Table 10: Percent Identification of /m/ and /n/

3.8 Glides

Responses to /ɯ̟/	Responses to /j/
/w/ 87.5 %	/j/ 55.1 %
/v/ 6.3 %	/i/ 7.5 %
/f/ 1.0 %	/dʒ/ 0.2 %

Table 11: Percent Identification of /ɯ̟/ and /j/

According to Vance (1987) [4], Sakuma (1929) [6] says that "the Japanese back glide is articulated exactly like the high back vowel, and he transcribes it phonetically as [ɯ̟], that is, as a nonsyllabic [ɯ]." Saito (1997) [5] states the similar phenomenon (p.44). Lip-rounding is not as strong as the English counterpart. The front glide /j/ is articulated in different places, depending on the following vowels, like the English counterpart. Only two words on the list contain /ɯ̟/, and 13 words contain /j/, and two of them, *ryuhyo* (drift ice) and *yubinkyoku* (post office) have /j/ twice. When the letter "y" is included in a response, the listener is considered to equate Japanese /j/ to the English counterpart, but without the letter "y", the glide can be included in the letter "u" as in *human*. In such cases, the response is considered to include the glide. /ɯ̟/ is equated with the English counterpart in 87.5% of the responses, but /j/ is equated with its English counterpart in slightly more than 50% of all the responses. This may be largely because in English /j/ can only occur before /u/ when another consonant precedes the glide. /j/ in *kyohi* (refusal) is most

poorly identified. In 77.1% of the responses to *kyohi*, no consonant corresponding to /j/ is identified. But when /j/ occurs syllable-initially, it is equated with its English counterpart in 81.3% of the instances.

4 Discussion

The results of this experiment reveal that Japanese consonants are not necessarily perceived as the common practice of transcription of Japanese consonants into the English alphabet. The Japanese liquid /ɾ/, for instance, is commonly transcribed as "r" in English, but it is more often perceived as /l/ or /d/ than /r/ and /ts/ in the word-initial position is perceived as /s/ rather than /ts/. These differences can be, for the most part, explained by phonetic differences between Japanese and English consonants. In general, the larger the differences are, the more dispersed the listeners' responses are.

Contexts should also be taken into consideration. /ts/ is not equated with /ts/ in the word-initial position, but intervocalically it is equated with /ts/. [ɸ] is equated with /h/ before /u/, but before other vowels it is equated with /f/. There may be two reasons for this. For one thing, allophonic differences cause speech sounds to sound different to nonnative listeners. Frication is stronger in *fensu* (fence) than *fuufu* (couple). The other reason is nonnative speech sounds are mapped differently into the native speech sound categories, depending on where they occur. Certain sounds do not occur in certain positions. In English, /ts/ and /dz/ are not unusual as double morphemes as in *gets, bats, bids, cards*, but as a single morpheme, /dz/ is impossible, and /ts/ is possible in borrowings like *pizza*. In the word-initial position, /dz/ is impossible, and /ts/ is only possible in recent borrowings. /j/ is better perceived as a glide when it occurs syllable-initially. All these support Speech Learning Model (=SLM) by James Flege (Flege 1995)[7] . Flege (1995) [7] claims that "sounds in the L1 and L2 are related perceptually to one another at a position-sensitive allophonic level, rather than at a more abstract phonemic level (p239)."

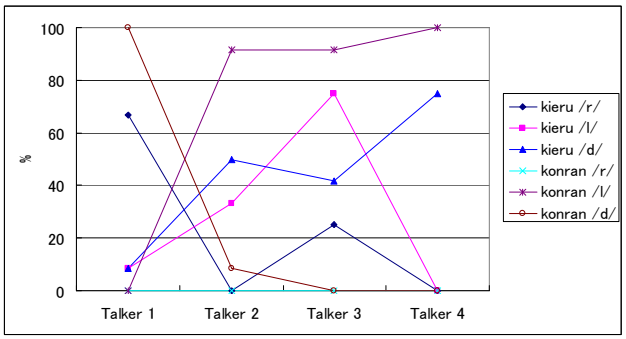

Figure 2: Percent Identification of /ɾ/ produced by four talkers in *kieru* and *konran*

Talker effects are also visible. Figure 2 shows the percent identification of four different /ɾ/ tokens in *konran* (confusion) and *kieru* (disappear). /ɾ/ in *konran* is, for the most part, equated with /l/, but Talker 1's /ɾ/ token is perceived as /d/. Talker 1's /ɾ/ token in *kieru* is perceived as /r/ at 66.7%, but the other Talkers' tokens are hardly not equated with /r/. Talker 4's /ɾ/ token in *kieru* is identified with /d/ at 75.0%. Interestingly, Talker 1's /ɾ/ token in *kieru* is not perceived as /d/ as often as the other talkers' even though the talker's /ɾ/ in *konran* is equated with /d/ at 100%. In other words, one talker does always produce /ɾ/ in the same manner. This is true with other talkers, too.

Native speakers of Japanese may perceive all these different tokens as exemplars of phoneme /ɾ/, but this shows a large allophonic variation of /ɾ/, and nonnative speakers' possible susceptibility to talker and token differences.

Categorical goodness rating and discriminability are not dealt with in this study. To shed light on how nonnative speech sounds are mapped into the L1 sound system, these should be included. Both /s/ and /ts/ in the word-initial position are equated with /s/, but it is not clear whether native English listeners cannot discriminate these two sounds in this position. Flege's SLM (Flege 1995[7]) states that "a new phonetic category can be established for an L2 sound that differs phonetically from the closest L1 sound if bilinguals discern at least some of the phonetic differences between the L1 and L2 sounds (p. 239)," but how large a phonetic difference has to be for bilinguals to establish a new phonetic category is yet to be investigated.

Acknowledgments

Work supported by Grant-in-Aid for Scientific Research (C) 18520338

References

[1] Guion, S.G., Flege, J.E., Akahane-Yamada, R. & Pruitt, J.C. An Investigation of Current Models of Second Language Speech Perception: The Case of Japanese Adults' Perception of English Consonants. *JASA*, *107*, 2711-2724 (2000)

[2] Aoyama, K., Flege, J.E., Guion, S.G., Akahane-Yamada, R. & Yamada, T. "Perceived Phonetic Dissimilarity and L2 Speech Learning: The Case of Japanese /r/ and English /l/ and /r/", *Journal of Phonetics, 32*, 233-250 (2004)

[3] Shirota, S. Nihongo no Oto (Sounds of Japanese). Tokyo: Hitsujishobo. (1995)

[4] Vance, T. J. *An Introduction to Japanese Phonology.* Albany: State University of New York Press. (1987)

[5] Saito, S. *Nihongo Onseigaku Nyumon (An Introduction to Japanese Phonetics)*. Tokyo: Sanseido. (1997)

[6] Sakuma, K. Nihon Onseigaku. Tokyo: Kazama Shobo. (cited in [4]) (1929)

[7] Flege, J. E. "Second Language Speech Learning Theory, Findings, and Problems", In Strange, W. (ed.). *Speech Perception and Linguistic Experience – Issues in Cross Language Research* (pp. 233-277). Baltimore: York Press. (1995)