ENGLISH CLUSTER PERCEPTION BY TAIWANESE MANDARIN SPEAKERS

Yueh-chin Chang[a], *Jiaqing Hong*[a], *Pierre Hallé*[b]
[a]National Tsing Hua University, Institute of Linguistics, Hsinchu, Taiwan,
[b]Laboratoire de Psychologie Expérimentale, CNRS-Paris V
ycchang@mx.nthu.edu.tw

ABSTRACT

Mandarin syllable structure does not allow consonant clusters. In this study, we investigated the perception of English initial consonant clusters by native speakers of Taiwanese Mandarin (TM). The results show that the factors that affect the perception of non-native clusters are the phonemic inventory of the native language, coarticulation within the cluster, articulatory command in producing consonant clusters, and native-language phonotactic constraints. However, these constraints are not the most important factor in the perception of non-native clusters by TM speakers.

Keywords: phonotactics, perception, English consonant clusters, Mandarin speakers

1. INTRODUCTION

Several theoretical models have been found in the study of speech perception. Native speech experience affects non-native speech perception, and non-native sounds may be assimilated to native phonemes, based on their acoustic, articulatory and/or phonetic similarity (cf. Best's perceptual assimilation model [1,2,3], Flege's speech learning model [4,5,6] and Kuhl's Native Language Magnet model [7,8,9]). Furthermore, some studies have claimed that native-language phonotactics affect the perception of native/non-native clusters [10,11, 12,13,14,15]: listeners tend to assimilate illegal sequences of phonemes to legal ones [12]. Japanese speakers perceived an illusory vowel inside consonant clusters [13,14]. Korean speakers reanalyzed the liquid in a /stop+liquid/ cluster as an intervocalic liquid, by vowel epenthesis [15]. Also, it has been shown that listeners are capable of predicting what the following vowel was based on the initial consonant alone.

Mandarin has reduced syllable types (i.e. (C)(G)V(N), where glides (G) are /j, w, ɥ/ and nasals are /n, ŋ/). There is a lateral /l/ and an English-like retroflex vowel in the phonemic inventory of Mandarin. English loan words containing /stop+liquid/ clusters are resyllabified by vowel epenthesis, ex. Blog → bulouge; Brandy → bailandi. The models described above could predict that: (1) speakers of Taiwanese Mandarin (TM) will perceive an epenthetic vowel in the English initial consonant cluster (hereafter, cluster); and (2) the cluster stop+/l/ will be identified more accurately than the cluster stop+/r/.

In the previous study [16], an epenthetic vowel /ə/ was found in the production of English clusters by TM speakers. In this study, we examine whether the TM listeners perceive an illusory vowel (/ə/) in the cluster and how the phonemic inventory of the native language affects the perception of English clusters by TM listeners.

2. PERCEPTION EXPERIMENT

An American phonetician recorded 18 pairs of CəCV disyllabic and CCV monosyllabic words (e.g. below (CəL) - blow (CL); berate(CəR) - brate (CR), cf. Appendix). Twenty-six students and two office staff members were paid to participate in both identification and discrimination tests at the phonetic laboratory of National Tsing Hua University. The students' age range was 19-22 years (mean=20.4 years), and the staff members' age range was 44-50 years (mean=47 years). They started learning English at the mean age of 10.8 years (range=8-13 years) and completed their study of English at age 22 in school. They speak Mandarin, Taiwanese, and English. Their self-judged English level was 2.9 (out of 5) for understanding and 2.6 for speaking. None had a hearing problem.

In the identification test, there were 144 stimuli (36 words x 4 repetitions) in the test phase, and 40 stimuli (10 pairs of words x 2 repetitions) in the training phase. The words were presented in random order and played twice at a time. The participants were asked to determine whether the word was dissyllabic or monosyllabic by pressing the number 1, 2, 3 or 4 on a keyboard for CəL, CL, CəR, or CR respectively. For the AXB discrimination test, 216 trials (4 trial types (AAB, ABB, BBA, BAA) x 18 pairs x 3 repetitions) were presented in random order. The interstimulus interval was 1 s, and the intertrial interval was 2.5 s.

2.1. Results of the identification test

Table 1 summarizes the English cluster identification results for the TM speakers with a confusion matrix. A series of three-way analyses of variance (ANOVA) was conducted, with Voicing (of the initial stop) (voiced, e.g. blow, voiceless, e.g. please)), Liquid (/l/, /r/) and Syllable (CəC, CC) as within-subject factors. The results show that the TM speakers could correctly identify the /CəCV–CCV/ contrast ($F(1,27)=0.57$, $p>0.05$). Only 20% of CL and 6% of CR clusters were perceived as having an epenthetic vowel in the cluster. However, 19% of CəL and 11% of CəR tokens were perceived as CL and CR, respectively. A significant effect was found for Liquid. The /r/ was perceived more accurately than the /l/. The interaction between Liquid and Syllable was significant. (Table 2)

Table. 1. Percent correct identification of English clusters by TM speakers, syllable such as /below/ identified as /bəl (CEL)/, as /bl (CL)/, as /bər (CER) or as /br (CR)/.

stimuli \ *response*	*CL*	*CEL*	*CR*	*CER*
CL	**75**	**20**	3	2
CEL	***19***	**76**	1	3
CR	3	2	**89**	**6**
CER	2	4	***11***	**82**

Table. 2. ANOVAS of Identification of English clusters

Source	df	F	Sig.
LIQUID	1, 27	18.95	0.00
SYLLABLE	1, 27	0.57	0.46
SYLLABLE*LIQUID	1, 27	5.15	0.03
VOICING	1, 27	47.60	0.00
VOICING*LIQUID	1, 27	3.25	0.08
VOICING*SYLLABLE	1, 27	6.09	0.02
VOIC*SY*LIQUID	1, 27	0.00	0.96

On examining the results of the identification test, we found that 20 subjects had good performance (88.5% correct, Group A hereafter). They perceived the /CəCV-CCV/ contrast well. The other 8 subjects (6 students and 2 staff members) performed above chance, 61.1% correct (hereafter, Group B). They had more difficulty in identifying the /CəCV-CCV/ contrast, especially in identifying CəL and CL. (Table 3)

Table. 3: English cluster identification confusion matrices

Group of Good Performance (Group A, n=20)

	CL	CEL	CR	CER
CL	**81**	***18***	1	1
CEL	***13***	**86**	0	0
CR	0	0	**95**	***5***
CER	0	1	***7***	**92**

Group of Poor Performance (Group B, n=8)

	CL	CEL	CR	CER
CL	**58**	***27***	9	6
CEL	**33**	**54**	4	10
CR	11	5	**73**	***11***
CER	7	11	***21***	**60**

2.2. Results of the discrimination test

Table 4 shows that the TM speakers discriminated the /CəCV-CCV/ contrast well (96.5% correct). An ANOVA with one between-subject factor (Group) and three within-subject factors (Voicing, Syllable, Liquid) was conducted. The results show that Group A performed significantly better than Group B on the discrimination task (98.1% vs. 92.6%, $F(1,26))=11.46$, $P<0.05$). The percent correct in discrimination was significantly higher for /r/ than for /l/ ($F(1,26))=12.62$, $P<0.05$). This is similar to the pattern found for the main effects (Liquid, Voicing) in the identification test.

Table. 4. Percent correct discrimination

	CL	CEL	CR	CER	Mean
Group A	**96.9**	**98.1**	**98.3**	**99.1**	**98.1**
Group B	**90.1**	**89.1**	**97.1**	**94.3**	**92.6**
Mean	**92.7**	**93.8**	**97.8**	**97.5**	**96.5**

This perception study shows that /r/ was perceived more accurately than /l/ by the TM speakers. These results do not support the prediction that the percent correct for stop+/l/ would be higher than that for stop+/r/ (prediction 2 above). We might suspect that the TM speakers assimilate English /r/ to the Mandarin retroflex vowel, since the Mandarin retroflex vowel and English /r/ have similar articulatory gestures and acoustic formants. However, Table 3 shows that Group B misperceived /CL/ and /CəL/ as /CR/ and /CəR/ and vice versa. This led us to examine production data from the 6 subjects who had participated in both the production and perception studies.

3. PRODUCTION EXPERIMENT

Participants were asked to produce 190 utterances. The initial word of each utterance contained a biconsonantal onset (e.g. bl, pl, gl, kl, sl, br, pr, gr, kr, dr). Using the phonetics software Praat, two of the authors transcribed and acoustically analyzed the initial word of each utterance.

The main results are as follows: (1) Three of the participants who had poor perceptual performance had poor production in English as well (Group PB hereafter). They made more production errors (58.7%) on clusters than the other three participants (Group PA hereafter) (4.8%). This seems to indicate that perception performance is correlated to production ability. (2) The main types of errors found in the production of Group PB were epenthesized schwa /ə/ in the clusters, the substitution of /r/ for /l/ (/l/->/r/), and the overlap of the range of F3 values between /l/ and /r/, due to the coarticulation between the initial consonant and the following liquid. (3) The duration of /l/ in the cluster produced by both groups of TM speakers was longer than that produced by the native English speakers (66.2 ms vs. 41 ms). This implies that the TM speakers needed more time to coarticulate the articulatory gestures in the production of the stop+/l/ cluster [17,18]. However, there was no significant difference between the duration of /r/ in the cluster stop+/r/ produced by the TM speakers (42.4 ms) and that produced by the native English speakers (33.6 ms). This suggests that even high-proficiency TM speakers of English cannot fully master the coordination of the articulatory gestures between the two consonants in the stop+/l/ clusters.

These production data might provide an explanation as to why the TM listeners were more accurate in perceiving stop+/r/ than in perceiving stop+/l/. Although Sheldon and Strange have pointed out that "perceptual mastery, although [it] often correlates with production accuracy, is not necessarily a prerequisite for, nor a consequence of articulatory mastery" [17], our study shows the influence of production ability upon perceptual ability in the acquisition of a non-native phonology.

4. Discussion

Dupoux et al. [14] showed that 79% of the consonant clusters transcribed by Japanese speakers contained an epenthetic 'u' between the consonants and concluded that the vowel epenthesis results from a prelexical process. Mandarin, like Japanese, does not allow clusters; however, the TM speakers identified the clusters better than Japanese speakers did. Moreover, the TM speakers tended to perceive CəCV as CCV. These results seem to put into question the importance of native-language phonotactic knowledge in speech perception.

The fact that Japanese speakers had difficulty in discriminating between CVCCV and CVCəCV might be due to the influence of the Japanese writing systems, such as the hiragana syllabary, in which each syllabary consists of a combination of a consonant and a vowel. The Japanese writing systems are unable to represent the individual sounds of Japanese. Therefore, Japanese speakers might use semi-syllable-sized or syllable-sized categories to parse the acoustic signal [13]. However, for TM speakers, the parsing mechanism is different from that of Japanese speakers. Each Chinese character represents a syllable consisting of the combination of an Initial/Onset plus a Final. Similar to Japanese writing systems, Chinese characters do not isolate the individual sounds of Mandarin. However, the TM speakers had learned to pronounce Chinese characters by using the Chinese Phonetic Symbol System called [pɤphɤmɤfɤ] (hereafter, CPSS). Each Chinese character is represented by 1-3 symbols, for instance, **gan 'dry'** represented by /kɤ/ and /an/; and **guan** 'close' by /kɤ/, /u/ and /an/. During their first stage of learning a foreign language, some TM speakers tend to transcribe the language by using CPSS, such as brandy → pɤ-lɤ-an-tɤ-i [p(ɤ)lanti]; Paris → pɤ-a-lɤ-i [pali]. Thus, it is difficult to define the perceptual unit size for the TM speakers, due to the fact that CPSS is a system in between a syllabary system and a phonetic alphabet system. But, our identification results show that there were 7%-13% of CəCV (in Group A's data) and 21%-33% of CəCV (in Group B's data) tokens identified as CCV. This implies that the TM speakers, especially Group B, classified [p] and [pɤ] as having the same perceptual unit size, because their within-category auditory impression forms a continuum with no clear demarcation. We suggest that the perceptual unit size is phoneme-like for TM speakers, but syllable-like for Japanese speakers. This explains why the TM speakers were able to discriminate the /CəCV-CCV/ contrast.

The second possibility might be that in Japanese, high vowel devoicing is very common and there is a short and long vowel contrast. Japanese speakers might therefore be very sensitive to the perception of short vowels [13]. By contrast, in Mandarin, vowels are resistant to the reduction process, because the vowel/rime is a prerequisite for the perception of tone.

To sum up, the present findings are not consistent with Dupoux's claim that the Japanese phonotactic constraint (i.e. not allowing clusters) affects the perception of non-native clusters. We have argued that syllable structure, as a phonotactic constraint, is not the decisive factor in the perception of non-native clusters by TM speakers. Since speech perception may be based on syllabic onset processing [12], the TM speakers may have perceived the English initial cluster as a whole rather than as two sequent consonants. They assimilated the English initial consonant cluster to a bad example of the Mandarin onset. According to PAM, Group A assimilated the /CəCV-CCV/ contrast to two different types of syllable structure in Mandarin (Two Category Assimilation). For Group B, who had poor perception performance and poor-production skills in English, the situation is more complicated. On one hand, they assimilated the cluster to a bad example of the Mandarin onset. On the other hand, since their perceptual unit size is phoneme-like (i.e. on the continuum between a phoneme and a syllable), they might assimilate /CəC/ to a bad example of the onset or to part of /CəCV/. Thus, they also showed Two Category assimilation, but their identification performance and discrimination performance are lower than those of the Group A.

Japanese and Korean speakers have more difficulty perceiving the English /l/-/r/ contrast in prevocalic clusters than in other positions, due to the coarticulation between the liquid and the preceding stops [17,18]. Our study supports the theory that coarticulation between two sequent consonants and the coordination of articulatory gestures do influence the perception of non-native clusters.

Since 6%-20% of the TM speakers reported hearing an illusory vowel in the cluster, native phonotactic knowledge clearly plays an important role, but not a decisive role, in non-native speech perception.

Appendix:

CəL-CL: belief-bleef; below-blow; police-pleese; galosh-glosh; collate-clate; collide-clide; saliva-sliva; select-slekt; galore-glore; galosh-glosh; polite-plite;

CəR-CR: **c**orrect-crect; parade-prade; piraff-praff; corona-crona; berate-brate; bereave-brieve; garage-grage;

Acknowledgement

We thank Arthur Abramson for recording the word stimuli. This research was founded by *NSC 93-2411-H-007-027.*

REFERENCES

[1] Best, C.T. (1994). Learning to perceive the sound pattern of English. In: Rovee-Collier, C., Lipsitt, L. (eds), *Advances in infancy research*. Vol. 8. Hillsdale, NJ: Ablex Publishers, pp.217-304.

[2] Best, C.T. (1995). A direct realist perspective on cross-language speech perception. In: Strange, W. (ed.), *Speech perception and linguistic experience: issues in cross-language speech research*. Timonium, MD: York Press, pp. 171-204.

[3] Best, C.T., McRoberts, G.W., and Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America* 109, 775-794.

[4] Flege, J, E. (1986). The production and perception of foreign language speech sounds. In: Winitz, H. (ed.), *Human communication and its disorders*. Vol. 2. Norwood, NJ: Ablex, pp. 224-401.

[5] Flege, J, E. (1991). Orthographic evidence for the perceptual identification of vowels in Spanish and English. *Quarterly Journal of Experimental Psychology* 43, 701-731.

[6] Flege, J, E. (1995). Second language speech learning: theory, findings and problems. In: Strange, W. (ed.), *Speech perception and linguistic experience: issues in cross-language speech research*. Timonium, MD: York Press, pp. 233-272

[7] Kuhl, P.K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics* 50, 93-107.

[8] Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., and Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255, 606-608.

[9] Iverson, P., Kuhl, P.K. (1996). Influence of phonetic identification and category goodness on American listeners' perception of /r/ and /l/, *Journal of the Acoustical Society of America* 99, 1130-1140.

[10] Massaro, D. & Cohen, M. (1983). Phonological context in speech perception. *Perception & psychophysics* 34, 338-348.

[11] Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (in press). What we know about what we never heard: Evidence from perceptual illusion. *Cognition*.

[12] Hallé, P.A., Segui, J., Frauendelfer, U., and Meunier, C. (1998). Processing of illegal consonant clusters: a case of perceptual assimilation? *Journal of Experimental Psychology: Human perception and performance* 24(2), 592-608.

[13] Dupoux, E., Kakehi K., Hirose Y., Pallier C., & Mehler J. (1999). Epenthetic vowels in Japanese: a perceptual illusion? *Journal of Experimental Psychology: Human perception and performance* 25(6), 1568-1578.

[14] Dupoux, E., & Pallier, C., Kakehi, K., Mehler J. (2001). New evidence for prelexical phonological processing in word recognition, *Language and cognition processes* 16(5/6), 491-505.

[15] Ingram, John and Park, S-G. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics* 25, 343-370.

[16] Chang, Yueh-chin (2006). Production and perception of English consonant clusters by Taiwanese Mandarin speakers. NSC research report. Ms. (in Chinese)

[17] Sheldon, A. & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics* 3, 243-261.

[18] Lively, S. E. , Pisoni, D. B., Yamada, R. A., Tohkuru, Y. & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America* 96, 2076-2087.