

SEGMENTATION STRATEGIES FOR SPOKEN LANGUAGE RECOGNITION: EVIDENCE FROM SEMI-BILINGUAL JAPANESE SPEAKERS OF ENGLISH

Kiyoko Yoneyama

Department of Linguistics, The Ohio State University
222 Oxley Hall, 1712 Neil Avenue, Columbus, OH 43210-1298 U.S.A.

ABSTRACT

The present study investigates speech segmentation of English by semi-bilingual Japanese speakers of English. Two experiments were conducted with forty subjects. The first experiment used a syllable-monitoring task and the results did not show a moraic segmentation strategy, which is a language-specific segmentation strategy for native speakers of Japanese. This suggests that they employed a general or segment-by-segment sub-lexical segmentation strategy rather than a language-specific moraic segmentation strategy for accessing the lexicon. The second experiment used a phoneme-monitoring task and here the results supported a language-specific moraic segmentation. Taken together, experiments suggest that semi-bilingual Japanese speakers have two segmentation strategies for spoken language recognition, supporting a model proposed by Cutler and colleagues. These results also suggest that these semi-bilinguals only partly suppress their native language-specific segmentation, which makes them comparable to the perfectly balanced French-English bilinguals in a similar study by Cutler and colleagues.

1. INTRODUCTION

The speech signal is a continuous stream of words, with no reliable cues to word boundaries. However, in order to understand speech, humans must be able to extract individual words effortlessly from the speech stream to use them to access lexical information. How is this possible?

Two types of model have been proposed to solve this paradox for lexical access. The first type of model presumes an explicit prelexical segmentation into rhythm-based constituents before lexical access occurs. Recent experimental studies have confirmed that this explicit segmentation differs across languages. Results from French suggest that listeners in that language segment speech at syllable boundaries [1]. English listeners, on the other hand, segment speech at the onset of strong (but not at the onset of weak) syllables [2]. Also, Japanese listeners segment speech at mora boundaries [3, 4]. However, these apparently different segmentation strategies can be unified in a universal model whereby segmentation exploits the rhythm-defined constituents of the language. The above studies exactly reflect the characteristic rhythm of each language. We call this strategy “the language-specific rhythm-based segmentation strategy” (hereafter “rhythm-based segmentation”).

The other type of model presumes the existence of a mature lexicon and a more general segmentation strategy that does not explicitly tap rhythmic constituents. We call this strategy “the

general segmentation strategy” (hereafter “general segmentation”).

In these models, lexical segmentation is achieved as an automatic consequence of the sub-lexical phoneme or feature recognition process so that for adult listeners with mature lexicons, speech can be mapped onto words, directly without reference to language-specific prosodic units [e.g., 5, 6, 7].

Cutler et al. [8] propose a universal lexical access model which allows for both of these segmentation strategies. In acquisition, rhythm-based segmentation is developed that is appropriate for the phonological structure of the ambient language and young children exclusively use this strategy to aid in the acquisition of lexicon. However, after acquiring a mature lexicon, general segmentation is also available and listeners can use both segmentation strategies to access to lexicon.

This dual strategy model presumes that listeners themselves must choose one of these segmentation strategies to access to lexicon. Monolingual and bilingual studies that tested for explicit prelexical segmentation show that monolinguals and bilinguals choose different strategies for segmenting the foreign-language input. Although monolinguals tend to use their rhythm-based segmentation whether or not it is appropriate for that language, bilinguals can inhibit their rhythm-based segmentation and employ the other general segmentation [1, 2, 3, 4, 8, 9].

However, is this true for semi-bilingual Japanese speakers of English? In this study, segmentation strategies were tested with two kinds of monitoring tasks (syllable-monitoring and phoneme-monitoring), to evaluate the Cutler et al. [8] model. This paper defined “Semi-bilingual Japanese speakers of English” as Japanese-English bilinguals who were native speakers of Japanese, and who were recognizably not native speakers of English by contrast to the “perfect” French-English bilinguals in [8].

2. EXPERIMENT I

2.1. Materials and Procedure

The materials were chosen as closely as possible to those in [3]. Sixteen English content words (nouns and adjectives) were chosen as stimulus words, beginning with a consonant-vowel-nasal sequence. They formed eight pairs, minimally contrasting in the affiliation of the nasal. A nasal in CV words was an onset to the following syllable, and that in CVN words was followed by another consonant, which was an onset to the following syllable. We used these to approximate Japanese

non-moraic versus moraic nasal. The stimulus words are shown in Table 1.

Table 1: The stimulus words for Experiment I

CV words	CVN words
tinny	tinsel
sunny	Sunday
panic	pansy
minute	Minster
penny	pension
tennis	tension
manner	mansion
senate	sensor

The stimulus words and 264 filler words were arranged into 64 lists. Each target word occurred twice, once in one of the first thirty-two lists and once in one of the last thirty-two lists. Of the thirty-two lists which did not contain target words, half contained no occurrence of the specified target, while the other half contained a dummy target. Two target orders were compiled, with type of target counterbalanced across order and first versus second half in the experiment for each item pair. The materials were recorded at a slow-normal rate by a male native speaker of British English.

Subjects were forty Japanese natives who have good communicative competence in English acquired in a minimum of three years of residence in English-speaking countries, but who were recognizably not native speakers of English.

They heard an auditory CV or CVN target, followed by a short list of words. The task was to listen for any word in the list that began with the target, and then to press a response key as quickly as possible when they detected the target. Stimuli and response pulses were recorded on the different tracks of a DAT tape and reaction times were measured by calculating the distance between the target onset in a stimulus word and subject's response on a Kay Sona-Graph 5500.

2.2 Results and Discussion

The mean response times (RT) for each condition are shown in Figure 1. Separate analyses of variance showed a main effect of word type; CV words were responded to faster than CVN words ($F(1, 38) = 13.01, p < .001$; subject analysis only). However, there was neither a significant main effect of target type effect nor a significant interaction between target type and word type.

Otake et al. [3] found the following patterns suggesting a language-specific moraic segmentation strategy. Firstly, CVN targets are difficult to detect in words where the nasal is an onset of the following syllable (CV words). Secondly, CV targets are equally easy to detect in both types of words. Finally, the RT for CVN is much longer than that for CV targets when they occur in words where a nasal is followed by another consonant, which is an onset to the following syllable (CVN words).

We regarded the nasals in English CV words in Table 1 as analogous to the non-moraic nasals in Japanese CV words which were used in [3]. However, this approximation is actually not very good because Takagi & Mann [10] reported that Japanese natives with greater exposure to English tend to

perceive ambisyllabic words as CVCCV if the first vowel is a lax vowel, and as CVCV if it is a tense vowel. In fact, six out of eight CV words in our materials contain a lax vowel as their first vowel. Contrary to our intention, semi-bilinguals might have perceived these six CV words as CVC words. Thus, we decided not to use the results of CV words.

The only way from our results of CVN words to know the segmentation strategy which semi-bilinguals employed is to investigate whether or not the results show a significant RT difference between CV and CVN targets in CVN words. However, they did not show this, suggesting that semi-bilinguals might have employed general segmentation rather than rhythm-based moraic segmentation.

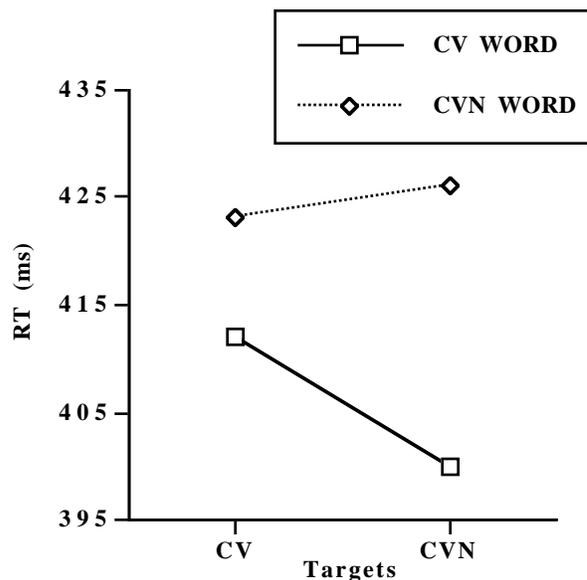


Figure 1: The mean target detection response time (RT) in ms. as a function of the size of an auditory target sequence and the phonological structure of a stimulus word.

Cutler and Otake [4] further tested their mora hypothesis with Japanese monolinguals, using phonemes which themselves can consist of a mora as targets. In Experiment II, we investigated speech segmentation by semi-bilinguals, using the same English materials in [4].

3. EXPERIMENT II

3.1 Materials and Procedure

The materials were taken from Experiments 3 and 4 of [4]. The stimulus words are shown in Table 2. Two visual targets were used: "O" and "N." "O" was always realized as [ɒ]. "N" was realized somewhat differently in syllable-initial ([n]) and syllable-final position (e.g., [ŋ] and [ŋ]). Table 2 also shows the expected equivalence to Japanese moraic versus non-moraic target.

Table 2: The stimulus words for Experiment II

"Moraic"			
Target O (C)V.V.C		Target N (C)V.N.C	
ion	eon	inlet	onward
iodic	ionic	incur	endear
kiosk	neon	fender	candy
dioxide	geography	conceal	confide
"Non-moraic"			
Target O (V)C.V.CV		Target N (C)V.NV	
cockerel	horrify	aniseed	enervate
torrid	dogged	enamel	anarchic
atomic	illogical	canopy	sanity
aloft	abolish	denial	canoe

(Dot indicates a potential mora boundary if these materials are equated with Japanese rhythmic patterns.)

The stimulus words and 245 filler words were arranged into 64 lists. Of the thirty-two lists which did not contain one of the experimental targets words, half contained no occurrence of the specified target, while the other half contained a dummy target. Materials were recorded at a normal rate by a male native speakers of Standard South British English.

Subjects were presented with a visual target consisting of a single Capital English letter (either "O" or "N") on a 15 x 17 cm card, immediately prior to the beginning of the list. The task was to listen for a word containing the sound associated with the target in the list of words, and then to press a response key as soon as they had detected the target sound. Reaction times were measured as in Experiment I.

3.2 Results and Discussion

The mean response times (RT) and the mean miss rates (%) are shown in Figure 2. The results showed that "N" targets were always detected more easily than "O" targets. The mean miss rate of 18.1% for "N" targets was lower than the mean miss rate of 38.7% for "O" targets ($F_1 [1,39] = 62.83, p < .001, F_2 [1,24] = 7.91, p < .01$). The mean RT of 655 ms. for "N" targets was faster than the RT of 693 ms. for "O" targets ($F_1 [1,39] = 41.12, p < .001, F_2 [1,24] = 14.58, p < .001$).

The results also showed an interaction between target type and the "moraic" condition ($F_1 [1,39] = 27.37, p < .001, F_2 [1,24] = 13.47, p < .005$ for RT; $F_1 [1,39] = 60.0, p < .001, F_2 [1,24] = 6.54, p < .02$ for miss rate). Figure 2 shows that "N" targets were detected faster and more accurately when they were "moraic" than "non-moraic." Also it shows that "O" targets were detected faster and more accurately when they were "non-moraic" than "moraic."

The results replicated Cutler and Otake [4]'s experiment with Japanese monolinguals: A moraic effect appeared only in "N" targets. We propose that the "anti-moraic" effect on "O" targets with semi-bilinguals might be caused by two kinds of mismatch as follows. The first mismatch might exist between the target sound in the target words and the sound which semi-bilinguals associated with the visual target. The second mismatch might come from the confound of stress assignment between (C)V.V.CV words and (V).CV.CV words in "O" targets. Even

though the stress was assigned to all target vowels in (V).CV.CV words, it is not assigned to half in the (C)V.V.C words (e.g., ion, kiosk, eon, and neon). These two kinds of mismatch can explain why we observed a significant difference in targets and why "moraic" targets were detected slower than "non-moraic" targets in "O" targets. The mismatch was worse for the (C)V.V.C words than for (V).CV.CV words.

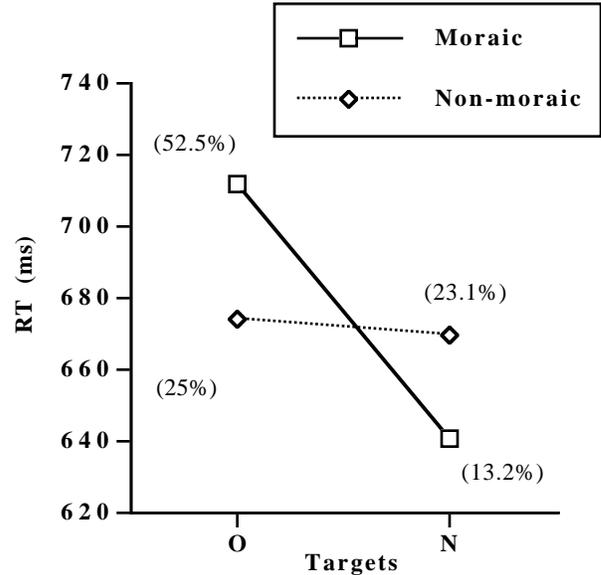


Figure 2: The mean target detection time (RT) in ms and the mean miss rate (%) in percentage as a function of vowel/consonant targets and moraic/non-moraic condition of a stimulus word, respectively.

4. GENERAL DISCUSSION

The results of our two experiments suggest that semi-bilinguals have two segmentation strategies: "rhythm-based segmentation" and "general segmentation." Thus, our experiments with semi-bilinguals provide supportive evidence for the Cutler et al. [8] model. However, Japanese semi-bilinguals seem to apply these segmentation strategies in a manner that is different from that employed by "perfect" French-English bilinguals in [8]. Even though French-English bilinguals applied general segmentation to the foreign-language input, semi-bilinguals employed these two segmentation strategies inconsistently.

Cutler et al. [8] propose that rhythm-based segmentation is the strategy that young children used first to solve the segmentation problem. Thus, even bilingual speakers can have only one rhythm-based segmentation, just like monolinguals. They also claim that bilingual speakers employ this rhythm-based segmentation when listening to the language from which they have acquired this strategy, and general segmentation when listening to a language with which they have extensive experience.

However, this is not the whole story with bilingual speakers. Bradley et al. [11] reported that their Spanish-English

bilinguals who learned Australian English after puberty did not employ rhythm-based syllabic segmentation to Spanish, their native language. Also Kearns [12] reported that French listeners in the process of learning English did not employ rhythm-based syllabic segmentation to French even though they employed general segmentation to English. As to the application patterns of two segmentation strategies provided by the three different types of bilinguals, Kearns [12] claims that only perfect bilinguals can perform the perfect separate application of two different segmentation strategies as the situation demands.

Our semi-bilinguals were native speakers of Japanese and were recognizably not native speakers of English by contrast to the "perfect" French-English bilinguals in [8]. Following Kearns [12], we assume that they might not be able to perform the perfect separate application of two different segmentation strategies as the situation demands, just like two groups of bilinguals in [11] and [12]. The present experiments showed that semi-bilinguals applied different segmentation strategies to their second language depending on experimental task. Even though they were able to employ general segmentation to the CV morae, they were not to the moraic nasals when listening to English. This suggests that semi-bilinguals only partly suppress their native rhythm-based segmentation as do perfectly the balanced French-English bilinguals in [8].

5. CONCLUSION

This study investigated segmentation strategies exploited by semi-bilingual Japanese speakers of English, to evaluate the Cutler et al. [8] model. The findings of two experiments suggested that semi-bilinguals use two segmentation strategies in processing spoken English: "rhythm-based segmentation" and "general segmentation." This confirmed the Cutler et al. [8] model. However, as Kearns [12] claims, semi-bilinguals cannot perform the perfect separate application of two different segmentation strategies as the situation demands. They only partly suppress their rhythm-based moraic segmentation.

6. ACKNOWLEDGMENT

This paper is based on the data of Experiments I and III in my M.A. thesis submitted to Dokkyo University in January, 1995 [13]. The earlier version of this paper was presented at the ninth Annual CUNY Conference on Human Sentence Processing in March 1996. I would like to express my sincere gratitude for valuable comments and advice provided by Takashi Otake and Akio Kamio at Dokkyo Univ., Anne Cutler and James McQueen at Max Planck Institute for Psycholinguistics, and Giyoo Hatano at Keio Univ. This work is also supported by Department of Linguistics, The Ohio State University. I would also thank Keith Johnson, Mary Beckman, Julie Boland, and the graduate students in our Department, for their valuable comments and a continuous support.

7. REFERENCES

- Mehler, J., Dommergues, J.-Y., Frauenfelder, U. & Segui, J. "The syllable's role in speech segmentation," *Journal of Verbal Learning & Verbal Behavior*, 20, 298-305, 1981.
- Cutler, A. & Norris, D. G., "The role of strong syllables in segmentation for lexical access," *Journal of Experimental Psychology: Human Perception & Performance*, 14, 113-121, 1988.
- Otake, T., Hatano, G., Cutler, A. & Mehler, J., "Mora or syllable? Speech segmentation in Japanese," *Journal of Memory & Language*, 32, 258-278, 1993.
- Cutler, A. & Otake, T. "Mora or Phoneme? Further evidence for language-specific listening," *Journal of Memory & Language*, 33, 824-844, 1994.
- Marslen-Wilson, W.D. & Welsh, A. "Processing interaction and lexical access during word recognition in continuous speech," *Cognitive Psychology*, 10, 29-63, 1978.
- McClelland, J.L. & Elman, J.F. "The TRACE model of speech perception," *Cognitive Psychology*, 18, 1-86, 1986.
- Norris, D.G. "SHORTLIST: A hybrid connectionist model of continuous speech recognition," *Cognition*, 52, 189-234, 1994.
- Cutler, A., Mehler, J., Norris, D. G. & Segui, J. "The monolingual nature of speech segmentation by bilingual," *Cognitive Psychology*, 24, 381-410, 1992.
- Otake, T., Hatano, G., & Yoneyama, K. "Japanese speech segmentation by Japanese listeners," in T. Otake & A. Cutler eds., *Phonological structure and language processing: Cross-linguistic studies*, 183-201, Mouton de Gruyter, Berlin, in press.
- Takagi, N., & Mann, V. "A perceptual basis for the systematic phonological correspondences between Japanese loan words and their English source words," *Journal of Phonetics*, 22, 343-356, 1994.
- Bradley, D.C., Sanchez-Casas, R.M., & Garcia-Albea, J.E. "The status of the syllable in the perception of English and Spanish," *Language and Cognitive Processes*, 8, 197-233, 1993.
- Kearns, R. K. *Prelexical speech processing in mono- & bilinguals*, Ph.D. Dissertation, University of Cambridge, 1994.
- Yoneyama, K. "A Speech Segmentation Procedure by Semi-Bilingual Speakers of Japanese and English," Unpublished M.A. thesis, Dokkyo University, 1995.