A cross-linguistic study of diphthongs in spoken word processing in Japanese and English

Kiyoko Yoneyama
Department of English
Daito Bunka University
yoneyama@ic.daito.ac.jp

Abstract
The current study conducted three phoneme-detecting experiments in order to (1) provide further evidence for the Japanese listeners' sensitivity to moraic structure and (2) treatment of diphthongs by three groups of language users (Japanese listeners, English listeners and semi-bilingual Japanese speakers of English). Japanese natives showed sensitivity to the moraic structure even when the moraic vowels are the second part of diphthongs. The results further showed that English listeners treat diphthongs as single units, while Japanese listeners treat them as two separate units. The semi-bilingual Japanese speakers of English treat diphthongs as single units in English but as separate units in Japanese.

1. Introduction
The rhythmic structure of the native language affects spoken language processing. The Japanese moraic structure also affects Japanese speakers' listening performance. For example, Otake and colleagues demonstrated sensitivity to the moraic structure by Japanese listeners through a series of syllable-detecting experiments [1]. Further, this sensitivity is also confirmed by the findings in phoneme-detecting experiments [2].

The vowel target used in [2] was a moraic vowel that was also a syllable by itself. However, a moraic vowel can occur tautosyllabically, following another vowel. Vance defines /ai/, /ei/, /oi/, /ou/ and /au/ as diphthongs in Japanese only if they occur in a single morpheme and if the second vowel segment does not bear an accent [3]. The second vowel of the diphthongs, of course, has a moraic status. If Japanese listeners' sensitivity to the moraic structure is robust, a second vowel of the diphthongs that has a moraic status (e.g., [i] in noizu) would be detected more quickly and accurately than a non-moraic vowel (e.g., [i] in bisu).

The current study has two aims. First, this study further investigates Japanese listeners' sensitivity to moraic structure using materials containing diphthongs. Second, this study also explores how three groups of language users (Japanese listeners, English listeners and semi-bilingual Japanese speakers of English) treat Japanese and English diphthongs. In order to fulfill these aims, three phoneme-detecting experiments with three groups of language users were conducted using the same sets of language materials (Japanese and English).

2. Experiment 1
Experiment 1 tests the Japanese listeners' sensitivity to moraic structure in spoken word processing. If Japanese listeners process speech mora by mora, the moraic vowels (e.g., [i] in noizu) should be detected more quickly and accurately than non-moraic vowels such as [i] in bisu.

2.1. Participants, Materials and Procedures
Eighteen Dokkyo University undergraduates participated in exchange for course credit. All were native speakers of Tokyo Japanese and reported no hearing difficulties. All had studied English only in school and none had ever stayed in an English-speaking country for more than three months.

Two sets of language materials (Japanese and English) were constructed. The Japanese materials, shown in Table 1, consist of 24 content words. Half contained [oi] (CV₁V₂ words) and the other half contained [i] (CV₁ words). They formed twelve pairs, contrasting in the occurrence of [o] in the first syllable. Each CV₁V₂ word contained a diphthong in Japanese. Similarly, the English materials were constructed as in Table 2. Half contained [ai] (CV₁V₂ words) and the other half contained [i] (CV₁ words). Note that the target sounds in question were different across language sets: a high front lax vowel ([i]) in English, a high front tense vowel ([i]) in Japanese.

Table 1: Japanese stimulus words.

<table>
<thead>
<tr>
<th>CV₁V₂ words</th>
<th>CV₁ words</th>
</tr>
</thead>
<tbody>
<tr>
<td>koin</td>
<td>hoiru</td>
</tr>
<tr>
<td>boiru</td>
<td>kika</td>
</tr>
<tr>
<td>koiru</td>
<td>join</td>
</tr>
<tr>
<td>noizu</td>
<td>roido</td>
</tr>
<tr>
<td>boisu</td>
<td>hoiru</td>
</tr>
<tr>
<td>doitsu</td>
<td>toire</td>
</tr>
</tbody>
</table>

Table 2: English stimulus words.

<table>
<thead>
<tr>
<th>CV₁V₂ words</th>
<th>CV₁ words</th>
</tr>
</thead>
<tbody>
<tr>
<td>join</td>
<td>voice</td>
</tr>
<tr>
<td>foil</td>
<td>void</td>
</tr>
<tr>
<td>toil</td>
<td>moist</td>
</tr>
<tr>
<td>coin</td>
<td>soil</td>
</tr>
<tr>
<td>boil</td>
<td>loin</td>
</tr>
<tr>
<td>coil</td>
<td>joint</td>
</tr>
</tbody>
</table>

The target words were mixed with filler words and were arranged into 48 word sequences in both sets of materials. In each set of materials, half contained one of the experimental target words and the other half did not. Each word sequence varied from two to six words. Of the twenty-four sequences that did not contain one of the experimental target words, half contained dummy targets that elicited participants’ responses. A target always occurred in the penultimate position in each word sequence.
A male American-English speaker of the North Kentucky dialect and a male Japanese speaker of the Tokyo dialect recorded the English and Japanese materials respectively. All materials, spoken at a normal rate of speech, were directly recorded onto DAT tapes.

Stimuli were presented to participants binaurally over headphones in a quiet room at a comfortable listening level from a DAT player. The participants’ responses were collected by them pressing a response button. The data collection was controlled by a PC computer. The English and Japanese materials were presented in separate sessions, with the English session coming first. In each test session, participants listened to 48 word sequences. For each sequence, they were instructed to think constantly of the target sound and to respond as soon as they detected it. A 10-sequence practice session preceded the 48 test sequences, which were presented in a fixed order in each test session. The entire experiment lasted approximately 50 minutes.

2.2. Results

Analyses of variance were conducted separately on missed data and response-time data. The mean miss rate are given in Figure 1. The results showed that targets were detected significantly more accurately in the Japanese materials than in the English ones ($F[1, 1860] = 12.916, p < 0.001$). The targets were detected significantly more accurately in CV$_1$V$_2$ words than in CV$_1$ words ($F[1, 1860] = 12.916, p < 0.001$). The effect of an interaction between material language and word type was not significant ($p = 0.076$). However, in English but not in Japanese, the targets were detected more accurately in CV$_1$V$_2$ words than in CV$_1$ words ($F[1, 430] = 12.991, p < 0.001$).

The means response times are given in Figure 2. The results showed that the targets were detected significantly more quickly in the Japanese materials than in the English ones ($F[1, 1860] = 1475.259, p < 0.001$). The targets were detected significantly more quickly in CV$_1$V$_2$ words than in CV$_1$ words ($F[1, 1860] = 18.290, p < 0.001$). There was an interaction between material language and word type ($F[1,442] = 35.442, p < 0.001$). Further analyses revealed that the targets were detected significantly more quickly in CV$_1$V$_2$ words than in CV$_1$ words in English, but not in Japanese ($F[1,50] = 8.604, p <0.05$).

2.3. Discussion

Experiment 1 provided contrasting evidence regarding whether or not Japanese listeners use general and rhythm-based segmentation strategies. There was no evidence of a moraic effect with the Japanese materials. The responses to English materials, conversely, suggest that Japanese listeners are sensitive to the moraic structure, which is consistent with the findings in [2]. Both the miss rate and reaction-time analyses showed a moraic effect.

The results of the reaction time analysis may suggest that the Japanese participants process speech phoneme by phoneme in Japanese whereas they process mora by mora when listening to English. Cutler and her colleagues have hypothesized that two types of listening strategies are available for processing of a native language [4]. Young children use a rhythm-based segmentation strategy to construct their lexicon. As their lexicon matures, however, they learn to segment speech directly into words, without phonological units intervening. Since all participants in this experiment have a mature Japanese lexicon, it is not surprising that they may have processed speech phoneme by phoneme in the processing of Japanese, their native language. This is also consistent with the findings in a recent study [5].

The results in the English materials are consistent with the findings in [2]: The Japanese participants are sensitive to moraic nasals even in English, as if they were listening to Japanese. The results here may suggest that the participants tried to find moraic structure even in English diphthongs.

Altogether, the results suggest that Japanese listeners treat diphthongs as sequences of two separate units.

![Figure 1: The mean miss rate as a function of stimulus word type and material language, Experiment 1](image1.png)

![Figure 2: The mean response times in ms, as a function of stimulus word type and material language, Experiment 1](image2.png)

3. Experiment 2

Experiment 2 investigates how English listeners treat diphthongs in English and Japanese. A prediction from previous studies is that English speakers do not show any sensitivity to moraic structure in English and Japanese and rather show a language-specific way of listening [1][2][4]. If this is the case, then sensitivity to moraic structure is a language-specific aspect of the Japanese language that is not acquired by English monolinguals through their acquiring the English language.
3.1. Participants, Materials and Procedures

Eighteen Ohio State University undergraduates participated in exchange for course credit. All were native speakers of American English and reported no hearing difficulties. None had ever either studied Japanese or stayed in Japan for more than 3 months.

The materials were the same as in Experiment 1. However, in this experiment, all the English and Japanese stimuli were digitally recorded onto computer disk (sampling rate of 10 kHz, low-pass filtered at 4.8 kHz) and edited and saved as separated sound files. Stimuli were presented to participants binaurally over headphones at a comfortable listening level. Responses were collected by participants pressing the button on a response button. Stimulus presentation and data collection were controlled by a PC/AT computer.

The procedures were generally the same as in Experiment 1, with a few minor changes. As many as four participants at a time were tested simultaneously in individual sound-attenuated booths. Also, the order of two test sessions was counterbalanced; half of the participants were presented with the English test session first and the latter half were presented with the Japanese test session first.

3.2. Results

Analyses of variance were conducted separately on missed data and response-time data. The mean miss rates are given in Figure 3. The results showed that the targets were detected in CV1 words more accurately than in CV1V2 words ($F[1, 860] = 289.864, p < 0.001$). An interaction between material language and word type was observed ($F[1, 860] = 7.182; p < 0.01$). Further analyses showed that the targets were detected more accurately in CV1 words than in CV1V2 words in both languages ($F[1, 430] = 217.703, p < 0.001$ for Japanese; $F[1, 430] = 92.851, p < 0.001$ for English).

The means response times are given in Figure 4. The results showed that the targets were detected significantly more quickly in the Japanese words than in the English ones ($F[1, 420] = 24.380, p < 0.001$). The targets were detected significantly more quickly in CV1 words than in CV1V2 words ($F[1, 420] = 17.854, p < 0.001$). An interaction between material language and word type was only marginally significant ($F[1, 420] = 3.678, p = 0.056$). Further analyses showed that the detection time difference between the word type conditions was significant in English, but not in Japanese ($F[1, 201] = 17.247, p < 0.001$).

3.3. Discussion

The miss rate analysis showed that English listeners treated diphthongs in English and Japanese in the same way. They consistently responded to targets in CV1 words more accurately than in CV1V2 words. The analysis of response times showed the same trend. If they were sensitive to moraic structure, the results should be the opposite: the targets should have been detected more quickly and accurately in CV1V2 words than in CV1 words. This not being the case, it appears that the English participants are not sensitive to moraic structure.

The result here confirmed the previous on-line studies with English listeners listening to Japanese: English speakers process Japanese as if they were listening to English [1][2]. The consistent patterns observed both in English and Japanese indicate a language-specific native way of listening by English listeners: diphthongs are treated as single units. Therefore, detecting the targets within diphthongs is extremely difficult for English listeners.

![Figure 3: The mean miss rate as a function of stimulus word type and material language, Experiment 2](image)

![Figure 4: The mean response times in ms, as a function of stimulus word type and material language, Experiment 2](image)

4. Experiment 3

Experiment 3 investigates how semi-bilingual Japanese speakers of English treat English and Japanese diphthongs. Of interest is whether knowledge of a second language can influence processing in the first language, or vice versa. The last experiment explores this possibility.

4.1. Participants, Materials and Procedures

Twenty-two Dokkyo University undergraduates participated in exchange for a nominal fee. The participants were native speakers of Japanese who had lived in an English speaking country for more than 3 years (semi-bilinguals). All had good communicative abilities in English but were recognizably not native speakers of English. They reported no hearing difficulties. Materials and procedures were the same as in Experiment 1.
4.2. Results

Analyses of variance were conducted separately on missed data and response-time data. The mean miss rates are given in Figure 5. The results showed an interaction between material language and word type ($F[1, 1052] = 7.204, p < 0.001$). Further analyses showed that the targets in CV1 words were detected significantly more accurately than those in CV1V2 words ($F[1, 526] = 5.192, p < 0.05$). No other main effect was observed.

The means response times are given in Figure 6. No significant main effect or interaction was observed.

![Figure 5: The mean miss rate as a function of stimulus word type and material language, Experiment 3](image)

![Figure 6: The mean miss rate in percentages as a function of stimulus word type and material language, Experiment 3](image)

4.3. Discussion

The results of the miss rate analysis further revealed that the semi-bilinguals treat English diphthongs as single units like English listeners. This pattern was not observed with Japanese monolinguals. This might suggest that diphthongs of English words stored in the semi-bilinguals’ English lexicon might be different from those of Japanese words stored in their Japanese lexicon.

Although the semi-bilinguals’ strategy treats diphthongs as single units in English, the semi-bilinguals and the English listeners performed differently. The semi-bilinguals performed much better than the English monolinguals when they detected the targets, even in English. Further, no effects of language material and word type in the reaction time analysis indicate not only that the semi-bilinguals process speech phoneme by phoneme, but also that the native and non-native speech is processed as equally quickly. Moreover, no effect of language materials in the miss rate analysis indicates that they can process both native and non-native speech without difficulties.

Altogether, the results may suggest that the semi-bilinguals have developed their own listening strategy in order to deal with English and Japanese effectively. The extensive knowledge of second-language enables the semi-bilinguals to modify the language-specific moraic listening strategy in order to accommodate with the non-native input.

5. Conclusion

This study satisfactorily confirmed sensitivity to moraic structure by Japanese listeners. This effect was observed even where the moraic vowels in question are the second part of diphthongs. Further, this study revealed that three language groups treat diphthongs differently in native vs. non-native input.

6. Acknowledgements

I thank Mary Beckman, Jose Benki, Anne Cutler, Keith Johnson, James McQueen, Takashi Otake, Katherine Smith and Lisa Shoaf for their help at various stages of this project. A preliminary analysis of the data used in this paper was published in [6].

7. References


