Effects of Consonantal Voicing on English Diphthongs: A Comparison of L1 and L2 Production

Kimiko Tsukada
School of Languages and Intercultural Education (SOLIE), Curtin University of Technology
e-mail: tsukadak@spectrum.curtin.edu.au

ABSTRACT

An acoustic comparison was made between native and non-native English diphthongs. We focused on the extent to which diphthong duration varied due to post-vocalic voicing. Talkers recorded three types of diphthongs /aɪ, eɪ, oʊ/ in /CVC/, /CVnC/ and /CVnC/ words in which the second C contrasted in voicing. The major difference between the talker groups was found for the diphthongs in /CVC/ and /CVnC/ words for which the voicing effect was larger in native than in non-native English. Two groups also differed clearly in the duration of the nasal in /CVnC/. While native talkers lengthened both diphthongs and nasals before /d/, non-native talkers lengthened only diphthongs in this position. In /CVC/, there was a greater similarity between the two groups for both the mean and variation of diphthong duration.

The data in this study indicate that transferring the L1 Japanese durational patterns to English would result in a positive transfer for /CVnC/, but a negative one for /CVC/. L2 learners need to acquire the knowledge and skills to pronounce the diphthong with distinct production strategies according to the syllable type in order to approximate to the phonetic norms of the native English speaker.

1. INTRODUCTION

It is well known that vowel duration is influenced by many linguistic factors. The presence or absence of the voicing feature in a neighbouring consonant is one of them and vowels tend to be longer before a voiced consonant than before its voiceless counterpart. While this effect is particularly pronounced in English for monosyllabic words in stressed position [2, 9, 14], it is reported to be small or non-existent in languages such as Arabic, Czech or Polish [11, 12]. Japanese is similar to the latter and it is expected that this temporal characteristics will be reflected in the speech production of Japanese L2 learners. In fact, previous work [8, 10, 17] showed that the extent of the voicing effect in Japanese-accented English (JE) speech was much smaller than that of native speakers of English. However, it is possible that the limited voicing contrast effect is characteristic of L2 speech in general [5, 6, 12] and not just of Japanese-accented English.

The objective of this study is to extend the acoustic comparison between Australian English (AE) and JE from monophthongs [17] to diphthongs with respect to the effect of consonantal voicing on vowel duration. This preliminary enquiry into the acoustic characteristics of AE vs JE diphthongs is motivated by the finding that native English listeners were sensitive to very subtle differences between the AE and JE vowels. They were capable of detecting a foreign accent even when the AE and JE monophthongs in /CVC/ presented to them in a perception study were carefully matched in F1 and F2 values at the vowels’ temporal midpoint and acoustic duration [17]. Although differences between AE and JE in the three parameters that were matched are undoubtedly crucial for the identification of vowels and are expected to contain cues to a foreign accent, the finding motivated us to search for residual cues. One of the possibilities that emerged as a result of the follow-up acoustic analyses is dynamic formant movement in the vowels. Since diphthongs are characterized by an extensive formant movement, the differences between AE and JE may be even more enhanced in their production of diphthongs as opposed to monophthongs. As a first step, we have decided to investigate English differential vowel duration in diphthongs. To be more specific, we are interested in verifying whether a smaller voicing effect in JE in comparison with AE for monophthongs and long vowels would be replicated for diphthongs. In order to control for the effect of syllable structure, both CVC and CVnC words were included. Since Japanese vowels predominantly occur in CV syllables, JE talkers need to make adjustments in producing the “new” CVC syllables in L2.

We also took it into consideration that consonantal voicing affects not only vowels but also the duration of the preceding nasal [15] and examined whether the same durational patterns would be observed for JE by including /CVnt/ and /CVnd/ words. Japanese nasal is somewhat unique in that when it occurs in syllable-final position, it is perceived as one mora /N/ on its own. Sato [17], in her cross-linguistic study, investigated the duration of syllable-final nasals in English, Korean and Japanese and confirmed that, in all three languages, nasals in this position are longer before a voiced consonant than before a voiceless consonant even across the syllable boundary. Her data showed that the average duration ratio of ’long nasal’ (before a voiced consonant) to ’short nasal’ (before a voiceless consonant) was highest in Japanese. This suggests that, in Japanese, the duration of syllable-final nasals, unlike that of vowels, may be used significantly to cue the voicing distinction in the following consonant.

2. METHOD
2.1. Talkers and Materials

9 AE talkers (4 male and 5 female) and 6 JE talkers (2 male and 4 female) participated in the study. Speech materials were recorded in sound-treated studios in Speech, Language and Hearing Research Centre (SHLRC) at Macquarie University and Curtin University of Technology. Talkers produced 3 English diphthongs /aɪ/, /eɪ/, and /oʊ/ in monosyllabic (CVt/CVd and CVnt/CVnd) and disyllabic (CVC) words. The first C varied across several consonants. The second C in disyllabic words included stops at different places of articulation. Most of the test words were minimal pairs in which the only difference was the voicing feature of the second C. The test words were embedded in simple carrier sentences such as ‘Say ___ again’ or ‘Say ___, please’ and they were presented on a computer screen one at a time in SHLRC. The list-reading method was used for the recording sessions at Curtin University. Talkers read the sentences from 2 to 7 times in random order. AE talkers in this study represent the general accent type on the scale of broad to cultivated which is conventionally adopted in describing AE accent types [3, 7]. JE talkers were all enrolled in tertiary education in Australia (4 in Sydney and 2 in Perth) at undergraduate or postgraduate levels when their speech was recorded.

2.2. Data processing and analysis

The speech data were digitized at 20 kHz and stored on the Unix workstations in SHLRC. Segmentation of the speech data were carried out using a signal processing package Xwaves+ and each utterance was phonetically labeled on the basis of waveform and spectrographic displays available in EMU speech database system [1]. Labeling criteria set out in [4] were followed. Statistical packages Splus and SPSS were used for data analyses and a graphic display.

3. RESULTS

Figure 1 shows the mean duration of all diphthongs averaged across all tokens by all speakers as a function of the voicing conditions for AE and JE data. Diphthongs in /CVC/ and /CVC\textsuperscript{ detached} words are plotted separately. Table 1 to 6 gives the mean duration for each diphthong type in different syllables. T-tests were used to compare the two means of the duration of diphthongs in minimal pairs. In testing the effects of the initial consonant and vowel height on diphthong duration, two-way ANOVAs were carried out.

3.1. AE dataset

The results from the AE dataset were in general agreement with the measurements reported in previous studies. When the diphthong occurred in /CVC/ (Table 1 and top left panel in Figure 1), it was considerably longer before /d/ than before /t/ for all three diphthong types without exception. Disregarding the diphthong type, the mean durational difference between the two voicing conditions was 71 ms (t(833) = -25.08, p < 0.01). Diphthong duration was also a good deal more variable before /d/ than before /t/. Although the effect of the initial consonant on the duration of the following vowel has been shown to be negligible in English [9, 14], we observed that /aɪ/ tended to be longer in the contexts of liquids. Of the 5 consonants (/h, l, r, s, t/) considered in the statistical analysis, /h/ induced the shortest /aɪ/.

<table>
<thead>
<tr>
<th></th>
<th>-voice</th>
<th>N</th>
<th>+voice</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>/aɪ/</td>
<td>189 (28)</td>
<td>190</td>
<td>263 (56)</td>
<td>189</td>
</tr>
<tr>
<td>/eɪ/</td>
<td>166 (19)</td>
<td>152</td>
<td>239 (39)</td>
<td>152</td>
</tr>
<tr>
<td>/oʊ/</td>
<td>152 (22)</td>
<td>38</td>
<td>220 (38)</td>
<td>38</td>
</tr>
</tbody>
</table>
Table 1: The mean duration (in ms) for diphthongs in /CVC/. Standard deviations are in parentheses. ** p < 0.01.

For /e/, the effect of varying initial consonants (/b, l, r, w/) was observed only when the diphthong preceded /t/. In this case, /t/ resulted in the shortest diphthong duration. The initial consonant effect was not tested for /o/, as there were only 2 consonant types (/b, t/ for /CVt/ and /t, r/ for /CVd/) preceding this diphthong.

Next, three sets of minimal pairs (‘write/ride’, ‘rate/raid’, ‘rote/road’) were examined to test the effect of vowel height on duration. /a/ with a low vowel as one of its components was significantly longer than both /e/ and /o/ when the consonantal contexts were fixed (F(2, 222) = 107.45, p < 0.01). The front diphthong /e/ was longer than the back diphthong /o/ when it was followed by /t/, but not /d/.

Table 2 shows the mean duration for diphthongs in /CVC/. In this position, the overall mean durational difference between the two voicing contexts decreased to 32 ms (bottom left panel in Figure 1), but remained statistically significant (t(425) = -15.6, p < 0.01). Even when the preceding and following consonants were kept identical, the diphthong in /CVd/ was much less variable than the same diphthong in /CVt/ in its duration. Also, in general, diphthongs in disyllabic words regardless of the voicing characteristic of the following stop.

<table>
<thead>
<tr>
<th></th>
<th>-voice</th>
<th>N</th>
<th>+voice</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>165 (16)</td>
<td>38</td>
<td>191 (21)</td>
<td>38</td>
</tr>
<tr>
<td>/e/</td>
<td>147 (16)</td>
<td>37</td>
<td>167 (18)</td>
<td>37</td>
</tr>
</tbody>
</table>

Table 2: The mean duration (in ms) for diphthongs in /CVC/. Standard deviations are in parentheses. ** p < 0.01.

As reported in [15], both nasals and diphthongs were influenced by the voicing feature of the following stop (Table 3). The nasals preceding /t/ were significantly shorter than the nasals preceding /d/ for both diphthongs considered. Proportionately, nasals lengthened more than did diphthongs in the context of /d/. This was interpreted to suggest that the nasal may be a stronger cue to the voicing feature of the following stop.

Table 3: The mean duration (in ms) for diphthongs in /CVnC/. Standard deviations are in parentheses. ** p < 0.01.

<table>
<thead>
<tr>
<th></th>
<th>-voice</th>
<th>N</th>
<th>+voice</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>183 (37)</td>
<td>90</td>
<td>206 (51)</td>
<td>89</td>
</tr>
<tr>
<td>/e/</td>
<td>171 (23)</td>
<td>72</td>
<td>184 (26)</td>
<td>72</td>
</tr>
<tr>
<td>/o/</td>
<td>171 (26)</td>
<td>18</td>
<td>187 (22)</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 4: The mean duration (in ms) for diphthongs in /CVC/. Standard deviations are in parentheses. ** p < 0.01.

<table>
<thead>
<tr>
<th></th>
<th>-voice</th>
<th>N</th>
<th>+voice</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>160 (23)</td>
<td>18</td>
<td>176 (19)</td>
<td>18</td>
</tr>
<tr>
<td>/e/</td>
<td>155 (18)</td>
<td>18</td>
<td>164 (27)</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 5: The mean duration (in ms) for diphthongs in /CVC/. Standard deviations are in parentheses. ** p < 0.01; * p < 0.05.

As in the AE dataset, the effect of initial consonant was observed, i.e., /a/ was longer when it was preceded by /l/ or /t/. For /e/, a two-way ANOVA showed the significant effect for the voicing factor (F(1, 136) = 9.77, p < 0.01), but not for the initial consonant factor (F(3, 136) = 1.19, p = 0.316). There was no significant interaction between the two main factors (F(3, 136) = 0.02, p = 0.995). The effect of vowel height on diphthong duration was present for this dataset, too, and /a/ was significantly longer than both /e/ and /o/, which did not differ from each other.

It is interesting to note that there was a noticeable durational increment in /CVnd/ words in comparison with /CVnt/ words, but that it was limited to the diphthongs and did not affect nasal duration (Table 6). All nasals had very similar duration at approximately 60 ms. While the lack of the morpheme boundary
effect was common to JE and AE data, the two data sets were in sharp contrast with each other in regards to the voicing effect on the preceding nasals.

<table>
<thead>
<tr>
<th></th>
<th>-voice</th>
<th></th>
<th>+voice</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>169 (24)</td>
<td>N</td>
<td>197 (27)</td>
<td>N</td>
</tr>
<tr>
<td>/I/</td>
<td>56 (12)</td>
<td>18</td>
<td>54 (17)</td>
<td>18</td>
</tr>
<tr>
<td>/e/</td>
<td>158 (25)</td>
<td>36</td>
<td>191 (34)</td>
<td>36</td>
</tr>
<tr>
<td>/u/</td>
<td>61 (12)</td>
<td>36</td>
<td>62 (21)</td>
<td>36</td>
</tr>
</tbody>
</table>

Table 6: The mean duration (in ms) for diphthongs in \(CVnC/\). Standard deviations are in parentheses. ** \(p < 0.01\); n.s. non significant.

### 3.3. Comparison of AE and JE

Although JE diphthongs showed varying degrees of lengthening before /d/, it was clear that the effect of consonantal voicing was considerably reduced in the JE production compared to the AE production. Quite a few JE diphthongs were longer than AE diphthongs before /I/, but mostly shorter before /d/. The same phenomenon was reported mainly for monophthongs in previous studies on foreign-accented English [5, 8, 10, 17]. There was also an interaction between the voicing and syllable type effects, as durational patterns in JE diverged from those observed in AE to a greater extent in monosyllabic than in disyllabic words.

There was a clear difference in the duration of the nasals between the two talker groups (Tables 3 & 6). While AE nasals were significantly longer before /d/ than before /I/ as were AE diphthongs, JE nasals were not influenced by the voicing in the following stop and their duration stayed constant around 60 ms. Only JE diphthongs showed slight lengthening in the context of the voiced stop. If JE talkers equated the post-vocalic nasals in \(CVnC/\) with the syllable-final nasals in their L1 Japanese, they might be expected to lengthen the nasals before voiced consonants, but not before voiceless consonants. However, such a contrast was not observed in the present study. Possibly, in the JE production, the lengthening of the nasals due to the voicing feature in the following consonant only affects those nasals that immediately precede a syllable boundary and not the nasals in a consonant cluster.

Although we found many differences between the two talker groups, we also noted some similarities. For instance, the initial consonant effect was observed for /a/ in both datasets and so was the vowel height effect. However, these effects were again of less magnitude in JE than in AE data.

### 4. SUMMARY

The results for AE dataset confirmed previous work on English vowels. We observed that 1) the voicing effect was much more pronounced in monosyllabic \(/CVCI/\) words than in disyllabic \(/CVC\) words, 2) the voicing effect was observed even when there was an intervening nasal as in \(CVnC/\), 3) voiced stops lengthened the preceding nasals as well as the diphthongs within the same syllable, and 4) the /a/ diphthong with a low vowel as one of its components was longer than the /e/ diphthong without a low vowel. With the exception of 3) above, JE talkers exhibited the same general pattern set by AE talkers, but to a much less extent. As for 3), there was a clear difference between the talker groups in their duration of the nasals. Only AE talkers produced the nasals significantly longer before /d/ than before /I/. The duration of JE nasals was not affected by the voicing feature of the following stop.

### 5. CONCLUSION

JE talkers in this study did not use vowel duration to cue the voicing distinction as much as AE talkers did. A smaller effect of the voicing contrast on vowel duration in JE than in AE was not limited to monophthongs and long vowels, but it was replicated for diphthongs, as well. It was observed that the major difference between the two talker groups lies in their production of monosyllabic \(/CVCI/\) and \(/CVnC/\) words. In \(/CVC\) words, the durational characteristics of JE speech was not so drastically different from the AE production (the bottom panels in Figure 1). JE talkers’ disadvantage with monosyllabic words is presumably due to the closed syllable which is ‘new’ in their L2. Since the voicing effect is considerably reduced in AE disyllabic words, transferring the L1 Japanese durational characteristics to L2 English would result in a positive transfer for \(/CVC\), but a negative one for \(/CVCI/\) words. L2 learners need to gain the knowledge and develop skills to pronounce the diphthong with distinct strategies depending on the syllable type in order to approximate to the AE phonetic norms. Whether AE talkers who have acquired Japanese as their L2 shows a similar lack of consonantal voicing effect in their L2 production is currently under investigation.

### 6. REFERENCES

ACKNOWLEDGEMENTS

I would like to thank associate professor Jonathan Harrington and the members of Speech, Hearing and Language Research Centre (SHLRC), Department of Linguistics, Macquarie University for the use of the facilities and technical support throughout the course of this study.