Production and Perception of Syllable Structure in Second-language Speech

Keiichi Tajima* and Reiko Akahane-Yamada*

* ATR Human Information Science Laboratories, Kyoto, Japan
** Hosei University, Tokyo, Japan
ktajima@atr.jp, yamada@atr.jp

Abstract

Production and perception of syllables in a second language (L2) that has relatively complex syllable structure, such as English, is expected to be difficult for native speakers of a language with relatively simple syllable structure, such as Japanese. A series of experiments investigated the perception, production, and learning of English syllables by native Japanese speakers. Results showed that Japanese speakers' ability to count syllables in spoken English words is closely related to the syllable complexity of the stimuli, while their production of English words showed more instances of epenthetic vowels for words containing voiced than voiceless consonants. In addition, improved skills in syllable perception did not lead to an improvement in production ability. Together, these results suggest that native Japanese speakers have considerable difficulties in both production and perception of syllables in spoken English, but the relationship between L2 syllable production and perception does not appear to be straightforward.

1. Introduction

Learning a second language (L2) entails not only mastering individual speech sounds but also learning the prosodic structure of L2 utterances. For example, acquisition of English syllable structure is expected to be difficult for native Japanese learners, given the relatively simple syllable structure of Japanese compared to English, and the role that moras instead of syllables play in Japanese [1]. Difficulties have been demonstrated in both perception of English syllable structure [2, 3] and their production [4, 5]. However, few studies have examined the extent to which the production and perception difficulties might be related to each other, and whether improved skills in one domain transfers to the other.

The relationship between speech production and perception has been a long-standing issue in speech research. In the context of L2 learning, studies on L2 phoneme perception and production have investigated which skills tend to be acquired first [6]. Training studies have demonstrated that improved skills in L2 phoneme perception transfers to L2 phoneme production [7].

The present paper is a preliminary attempt to address similar issues regarding the perception-production relationship in the context of L2 syllable structure. Results are reported from a series of independent studies investigating the production, perception, and learning of syllables in English words by native Japanese speakers. The first set of studies examined perception of English syllables by having listeners count syllables in spoken English words. The second study investigated production of English words, focusing on Japanese speakers’ alleged tendency to produce epenthetic vowels to break up consonant clusters (e.g., “stress” produced as /su.to.re.su/). Data from these studies were compared to examine whether the types of items whose syllable structure is difficult to perceive also tend to be difficult to produce, focusing on two stimulus-related factors, (1) syllable complexity and (2) consonant voicing. In addition, the present study explored whether perceptual identification training, which has been shown to improve listeners’ syllable perception accuracy [8], leads to an improvement in their production abilities.

2. Perception experiment

This study investigated how well native Japanese listeners can perceive syllables in spoken English words by using a syllable-counting task, in which listeners identified the number of syllables they heard in each stimulus. Stimuli were native-speaker productions of real English words (1–6 syllables in length) as well as English-like nonwords (1–2 syllables in length). Segmental content of the real words was not controlled carefully, but syllable complexity, as measured by the number of consonants in each item, was varied from relatively simple, e.g., “bee”, to complex, e.g., “scripts” (for further details of the experiment, see [8]).

Overall results from 23 native Japanese listeners indicated that, with minimal instruction and practice, listeners correctly identified the number of syllables only 57.1% of the time.

It has been previously shown using nonword stimuli that listeners’ accuracy declines as syllable complexity increases, and is lower for nonwords ending in voiced consonant clusters than voiceless clusters [9]. It has been speculated that the effect of syllable complexity arises due to listeners’ tendency to assign individual consonants to separate syllables, given that Japanese does not per-
To further examine the effect of voicing, Figure 1 shows data collected from a different group of Japanese listeners (n=21) using nonword stimuli that varied only in the voicing of initial consonant clusters (e.g., “spel” vs. “belp”) or final consonant clusters (e.g., “peps” vs. “pebs”). In initial position, accuracy for nonwords starting with /sp/ or with /bl/ were 60.7% and 66.7%, respectively, which was a non-significant difference (p = .06) in the opposite direction from that predicted above. In final position, accuracy for nonwords ending in /ps/ or in /bz/ were 65.5% and 59.5%, respectively, which was a significant difference (p < .05) in the expected direction.

Together, these results suggest that listeners’ syllable-counting ability is closely related to the complexity of the syllables. Consonant voicing appears to play only a secondary role if at all, suggesting that the likelihood of epenthetic vowels in production is not highly relevant in perception.

### 3. Production experiment

The production study investigated native Japanese speakers’ production of English syllable structure by examining the occurrence of epenthetic vowels within consonant clusters and following word-final consonants. A group of native Japanese speakers (n=26) produced 90 English words at two self-selected speaking rates, normal and slow. The productions were subjected to careful auditory and spectographic analysis, conducted by three phonetically trained judges, to determine the occurrence of epenthetic vowels. Specific criteria were used to identify epenthetic vowels; they were limited to those that showed clear phonation and clear vowel-like formant structure, containing at least two glottal pulses. Epenthetic vowels that were identified by all three judges were counted, and the percentage of tokens that contained epenthetic vowels was computed for each word (see [10] for further details).

Overall, 19.3% of the productions contained at least one epenthetic vowel, indicating that roughly one of out...
Table 2: Pairwise correlations among number of consonants, voicing index, and percentage of vowel epenthesis for monosyllabic English words (n=50).

<table>
<thead>
<tr>
<th>Num. cons.</th>
<th>Voicing</th>
<th>Epenthesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>-0.028</td>
<td>1.000</td>
<td>—</td>
</tr>
<tr>
<td>0.274</td>
<td>0.544***</td>
<td>1.000</td>
</tr>
</tbody>
</table>

***p < .001

Table 2: Percentages of tokens containing epenthetic vowels in words starting with or ending in voiced vs. voiceless consonant clusters. Bar graphs with different patterns are used to code epenthetic vowels occurring in different positions.

Assumes that native Japanese speakers apply vowel devoicing in L2 productions in much the same way that they would in Japanese. That is, Japanese speakers may have produced an epenthetic vowel within voiceless clusters such as /sp/ and /ps/, but it may not have been audible by virtue of vowel devoicing. Whether there are acoustic correlates of or perceptual consequences for such alleged “devoiced” epenthetic vowels remains an open issue.

When the production and perception data are compared, Figure 1 shows that word-initial /sp/ clusters are as difficult, if not more, to perceive as word-initial /bl/ clusters, but Figure 2 suggests that /sp/ clusters might be easier to produce than /bl/ clusters. That is, word-initial /sp/ clusters may be easy to produce without epenthetic vowels, but may nevertheless be difficult to perceive as belonging to the same syllable. Thus, it appears that perception and production difficulties do not always match each other, at least when syllable-counting accuracy and percentage of vowel epenthesis are taken as diagnostics of syllable perception and production difficulties.

4. Perceptual training and transfer to production

Previous studies have demonstrated that native Japanese listeners’ ability to count syllables in English words improve substantially as a result of perceptual training using feedback. Significant effects of training have been demonstrated for both young and elderly Japanese listeners [8], and have also been shown to be long-lasting [11].

As an additional way to assess the production-perception relationship, possible effects of perceptual training on production was examined. Production of English words (n=16) by 23 native Japanese listeners were recorded before and after five days of perceptual training during which they received feedback concerning their responses. The recorded words were also used during train-
ing; that is, listeners had multiple opportunities (15 times) during training to count syllables in these words and receive feedback. This was done to increase the likelihood of finding an improvement from pretest to posttest. The productions were used as stimuli in a paired-comparison task, carried out independently by three trained judges. In the task, two tokens of the same word, one recorded before and one recorded after training, were presented in either order, and the judge was to determine which version sounded “more native-like rhythmically”.

If training systematic modified listeners’ productions in perceptually relevant ways, then judges should be able to correctly identify the post-training member of each pair of productions at better-than-chance level. Across the 23 listeners, the mean percent-correct identification was 56.8% (s.d. = 12.2), which was only slightly above chance level (50%). Among individual listeners, there were only three listeners whose posttest productions were correctly identified more than 70% of the time. In addition, there was no strong correlation between individual listeners’ identification scores in the paired-comparison task and their degree of improvement in the syllable-counting task from the pretest to the posttest ($r = 0.34$; $p = .11$). These results do not provide evidence for a direct relationship between perception and production abilities, nor transfer of perceptual training to production.

5. Conclusion

Taken together, the results from these studies suggest, first, that native Japanese speakers have difficulties in both perception and production of syllable structure in English. Native Japanese listeners frequently miscount the number of syllables in English words, and erroneously produce epenthetic vowels in English words. Second, the results suggest that the relationship between production and perception is not straightforward in several regards. One aspect in that difficult-to-perceive words are not always difficult-to-produce words, at least when perceptual difficulty is measured by syllable counting accuracy and production difficulty by the occurrence of epenthetic vowels. Words whose syllable structure is difficult to perceive tend to be those that have complex syllable structure. In contrast, words that are difficult to produce are in general those that contain voiced consonants. If vowel epenthesis is taken as a diagnostic, then accurate perception abilities do not appear to be a necessary condition for appropriate production of English words. Whether there are acoustic or perceptually relevant differences between consonant sequences that allegedly contain “devoiced” epenthetic vowels and those that do not has yet to be resolved. The other aspect is that improved syllable perception skills do not seem to transfer straightforwardly to production. That is, having learned to perceive the correct number of syllables in L2 words is not sufficient for significantly improving productions skills.

6. Acknowledgments

This research was conducted as part of “Research on Human Communication” with funding from the Telecommunications Advancement Organization of Japan.

7. References


