PERCEPTUAL LEARNING OF SECOND-LANGUAGE SYLLABLE RHYTHM BY ELDERLY LISTENERS*

Keiichi Tajima, Reiko Akahane-Yamada
ATR Human Information Science Laboratories
2-2-2 Hikaridai, Seika-cho, Soraku-gun
Kyoto 619-0288, Japan
{ktajima,yamada}@atr.co.jp

Tsuneo Yamada
National Institute of Multimedia Education
2-12 Wakaba, Mihama-ku
Chiba 261-0014, Japan
yamada@nime.ac.jp

ABSTRACT

Past studies attempting to train second-language (L2) learners to accurately perceive L2 speech have focused primarily on non-native segmental contrasts, using young, college-aged listeners as trainees. To examine whether similar training methods improve perception of prosodic properties of L2, and whether older listeners can benefit from such training, the present study investigated the effect of auditory training using feedback on elderly native Japanese listeners’ perception of relatively complex syllables in spoken English words. Monolingual Japanese listeners aged 60–69 were trained to identify the number of syllables in spoken English words. Before training, listeners correctly counted syllables 56% of the time. After roughly three weeks of training, however, listeners correctly responded 91% of the time. Furthermore, listeners’ responses were less affected by syllable complexity of the stimuli after training than before. Results suggest that auditory training is effective for improving L2 syllable perception, even for elderly listeners.

1. INTRODUCTION

Learning phonetic contrasts that are not used in one’s native language (L1) is known to be difficult. However, it has been demonstrated that perception of L2 phonetic contrasts can be modified through appropriate training methods, even for adult L2 learners. For example, identification training using feedback has been shown to improve perception of the /t/-/l/ contrast by native speakers of Japanese [6]. The perceptual training generalized to novel words, was maintained after several months, and improved production of the contrast as well [5, 1].

Many previous studies on L2 speech learning have restricted themselves to addressing the perception of individual sounds, and it is not until recently that perceptual learning of prosodic aspects of L2 speech has been evaluated experimentally. One recent study [11] investigated perceptual learning of the four lexical tones of Mandarin Chinese by native American English listeners, and found that identification training using feedback improves American listeners’ ability to identify the tone of Chinese monosyllabic words.

Recently, the use of identification training has been further extended to include perception learning of non-lexical, global property of L2 speech, namely, the perception of syllables in spoken English words by adult native Japanese listeners [9, 8]. Perception of English syllables by Japanese was expected to be difficult because of substantial differences in typical syllable structures that occur in the two languages. While English syllables vary widely in structure, from simple consonant-vowel (CV) syllables, e.g., see, to complex syllables containing consonant clusters e.g., scripts (CCCCVCCC), Japanese allows at most one syllable-initial and one syllable-final consonant [10]. Moreover, the psychological “unit of speech” that appears to be dominant for Japanese speakers is the mora rather than the syllable [4, 7]. In fact, Japanese speakers are typically unfamiliar with the notion of a syllable. To evaluate perceptual consequences of such tendencies, a simple task was used in which Japanese listeners identified the number of syllables in spoken English words [9]. Training was carried out using the same task, but listeners received feedback about their responses and repeated trials until they entered the correct response. While the task was intuitively clear to native English listeners (mean accuracy of 93%), Japanese listeners did not perform well without any training (mean of 48%). After a brief, hour-long training session, listeners’ accuracy rose to 81% on average, suggesting that listeners could quickly learn to count syllables in spoken English words.

Generalization to untrained words in this study was only modest, however, showing an increase from 43% to 65%.

Aside from evaluating the effect of identification training on various segmental and prosodic properties of L2 speech, another set of factors that warrants close investigation pertains to listener-related variables, such as their age and amount of experience in L2-speaking communities. Studies have shown that these and other listener-related variables affect listeners’ ability to accurately perceive properties of L2 utterances [2, 12]. In one study on perceptual training of the English /t/-/l/ contrast by elderly Japanese listeners [3], training was found to result in smaller improvements in performance for relatively older listeners than for younger ones. Identification accuracy for Japanese listeners in their 40s, 50s, and 60s changed from 64% to 78%, 63% to 75%, and 61% to 64%, respectively, as a result of training. In comparison, younger Japanese listeners in their 20s improved their scores from 70% and 89% as a result of training [5], showing the greatest improvement of all age groups. These results suggest that aging interferes with perceptual learning of L2 sounds, although not completely.

The present study is part of a larger project investigating how aging affects perceptual learning of rhythmic properties of L2 utterances. It is conceivable that perceptual learning of different aspects of L2 speech (e.g., segments vs. syllables) is affected differently by aging. To test this, listeners of different age groups were trained to perceive English syllables using a syllable-counting task.

*The research reported here was supported in part by a contract with the Telecommunications Advancement Organization of Japan entitled, ‘Research on Human Communication’. Thanks are due to Terumi Fukuda for running the elderly listeners in the experiment.
As a way to ensure that the training led to more robust changes in performance than previously obtained [9], training was lengthened to 15 sessions and the number of training words were substantially increased. When such extensive training was conducted on college-aged Japanese listeners [8], accuracy rose from 53% in the pretest to 90% in the posttest, an improvement of 37 points. The present study addresses the extent to which elderly native Japanese listeners would benefit from a similar training method.

2. METHODS

2.1. Stimuli

A list of 358 English words was prepared, with 72 one-syllable words, 98 two-syllable words, 82 three-syllable words, 60 four-syllable words, 34 five-syllable words, and 12 six-syllable words. The words were divided into two overlapping sets: (1) 300 words used during training, and (2) 116 words used during pretest and posttest; 58 of these words belonged to both sets. An attempt was made to avoid words that may vary in syllable count (e.g., dictionary may have three or four syllables depending on the dialect). The segmental content of the words was not controlled carefully, but the number of consonants in each word, as it would appear in a typical phonemic transcription, varied so that the word set as a whole contained syllables of varying complexity (e.g., lie, conflicts, melody, development, economical, incomprehensible).

In addition to real words, a list of 32 nonwords was also prepared with varying degrees of complexity of initial and final consonants. The simplest nonword was the single vowel /AG/. To this, various initial consonants (none, /p/, /sp/, /spl/) as well as final consonants (none, /-pl/, /-ps/, /-mps/) were added in a factorial manner, yielding 16 one-syllable nonwords (e.g., /p/AG/ps/). Two-syllable nonwords were formed by adding /-ed/- immediately before the vowel /AG/ in each one-syllable word, and stressing the second syllable, yielding 16 two-syllable nonwords (e.g., /p/AG/ps/ /spl/AG/ps/).

Each item was read by two female and two male native American/Canadian English talkers between 31 and 46 years of age, residing near Kyoto, Japan. They were not fluent in languages other than English, and were naive to the purpose of the experiment. Recordings were made in an anechoic chamber at ATR Laboratories. Each word was saved into a separate audio file, which was amplitude-normalized so that the peak amplitude was the same across all files.

2.2. Listeners

A group of 12 monolingual Japanese listeners in their 60s, who took courses at the University of the Air, Chiba, Japan, were recruited for paid participation in the study. To test for age-related hearing loss, a brief hearing test was conducted before the pretest using an audiometer, at 500, 1000, 2000, and 4000 Hz. Of the 12 listeners, seven had hearing losses (20 dB HL or higher) in at least one of the frequencies tested in at least one ear. Of these seven, three showed relatively severe hearing losses (40–55 dB HL). As will be discussed in Section 4, listeners’ performance in the task did not appear to be affected by their hearing.

All the listeners had taken six years of English courses in junior and senior high school starting at age 12, and lasting until age 18–22 depending on whether they further took English courses in college. These courses focused on grammar, with very little time devoted to oral communications skills. Many listeners reported having had some exposure to spoken English after college, through courses on TV or radio, group lessons, or private lessons on English conversation. Only two out of the 12 listeners reported having no such exposure to spoken English. None of the listeners had stayed in an English-speaking community for more than three months.

2.3. Procedure

The experiment consisted of a pretest, a series of 15 training sessions, and a posttest. The pretest was conducted at the National Institute of Multimedia Education (NIME), located next to the University of the Air. The test consisted of two blocks of 148 trials each. In each block, 116 real words and 32 nonwords as spoken by one talker were presented once each in a random order. Two talkers were used in the pretest; the choice and presentation order of the talkers were counterbalanced across listeners.

Listeners sat in front of a laptop computer in a sound-treated room, and heard the stimuli through headphones at a comfortable listening level. On each trial, they were to listen to the stimulus, count the number of syllables in it, and use the mouse to click the appropriate button on the program window. Each stimulus was played back once, but listeners were able to click the “replay” button and listen to the stimulus again, although they were discouraged from doing so frequently. The trials were self-paced.

Since the Japanese listeners were expected to be unfamiliar with the notion of a syllable, they received a brief text describing through examples that a syllable consisted of a vowel that is optionally surrounded by consonants. Before the test trials, listeners were also given 10 practice trials with feedback; each practice trial was repeated until the listener entered the correct number of syllables. The practice words differed from the test items and were spoken by a different female American talker.

Following the pretest, listeners performed a series of 15 training sessions at their home. A laptop computer containing the training program as well as headphones (Sennheiser HD-265) were loaned to each listener to be used for training. Listeners were to perform no more than one session per day, and consecutive training sessions were not to be separated by more than three days. In each session, 300 real words as spoken by one talker was presented in a random order. Of the 300 words, 58 were words that also appeared during pretest and posttest. A different talker appeared on each session. Listeners cycled through three talkers in a fixed order for five times over the course of the 15 training sessions. The choice and presentation order of the talkers (one of whom also appeared in the test trials) were different for different listeners. Training trials were identical to pretest and posttest trials except that immediate feedback was provided concerning each response. If a response was incorrect, listeners repeated the trial until the correct response was entered.

Following the training sessions, the listeners returned to NIME for the posttest. The mean number of days from the pretest to the posttest among all listeners was 22.3 days. The posttest was identical to the pretest, using the same stimuli and the same number of trials. One of the two talkers in the posttest was a “trained talker” who appeared during training, while the other was an “untrained talker”. Likewise, 58 of the test words were “trained words” which appeared during training, while the other 58 words were “untrained words” (nonwords did not appear during training either).
3. RESULTS

Fig. 1 plots the percent-correct identification of the number of syllables in the stimuli in the pretest and posttest, averaged across all listeners. From left to right, scores are given for trained words ($n = 58$), untrained words ($n = 58$), nonwords ($n = 32$), and all the test stimuli as a group ($n = 148$). The data are collapsed with respect to the trained and untrained talkers, since no systematic effect of talker was found.

Fig. 1 shows that prior to training, the elderly listeners responded correctly 55–57% of the time, indicating that they miscounted syllables almost as often as they counted them correctly. However, after training, mean accuracy rose to 90–92%, an improvement of approximately 35 percentage points. Posttest accuracy on trained words (92.3%) was naturally very high, but accuracy on untrained words (89.9%) and nonwords (90.4%) were also remarkably high, suggesting that the training successfully generalized to items that the listeners were not trained with.

An analysis of variance (ANOVA) was carried out with Phase (pretest, posttest), Word Type (trained real words, untrained real words, nonwords), and Talker (trained talker, untrained talker) as within-subjects factors. The dependent variable was the arcsine-transformed values of the percent-correct responses. There was a highly significant main effect of Phase [$F(1,11) = 74.56; p < .001$]. No other main effects or interactions were found to be significant. This result indicates that the elderly Japanese listeners significantly improved in their ability to count syllables in the stimuli from pretest to posttest, and that their level of performance was not affected by exposure to a particular talker nor to particular words during training.

Analysis of the 32 nonwords was further carried out in order to examine how elderly Japanese listeners’ ability to count syllables was affected by the syllable complexity of the auditory stimuli. In Figs. 2 and 3, mean accuracies for the nonwords were plotted as a function of the initial and final consonants, for the pretest and posttest, respectively. The data are collapsed across one-syllable and two-syllable nonwords.

Fig. 2 reveals that the pretest data show considerable variation in accuracy among different types of nonwords. For example, accuracy is close to 100% for simple nonwords such as /p-/, but is lower than 25% for relatively complex nonwords such as /spl-mps/. There is a clear tendency for accuracy to decrease as the initial consonants become more complex, and also as the final consonants become more complex. While the line graphs in the figure are generally tilted to the right, there also appears to be a “floor effect” whereby the scores cease to decrease past certain values. For example, for nonwords beginning with /p-/ and accuracy gradually declines from left to right, but does not fall below approximately 45%. Similarly, for nonwords beginning with /spl-/, accuracy falls to about 30%, but does not decline much further even as the final consonants become more complex. An ANOVA conducted with Initial Consonants (four levels) and Final Consonants (four levels) as within-subjects factors revealed highly significant main effects of Initial Consonants [$F(3,33) = 21.84; p < .001$] and Final Consonants [$F(3,33) = 33.43; p < .001$], and a marginally significant interaction between the two factors [$F(9,99) = 2.14; p = .033$].

The posttest data in Fig. 3, by contrast, reveal generally high scores for all types of nonwords. The line graphs reveal no systematic effect of the final consonants, although there appears to be a small tendency for accuracy to decline as the initial consonants become more complex. An ANOVA conducted on the posttest data confirmed these observations; there was a significant main effect of Initial Consonants [$F(3,33) = 5.45; p = .004$], but the main effect of Final Consonants [$F(3,33) = 1.55; n.s.$] and the interaction
[\(F(9,99) = 0.86; \text{n.s.}\)] were not significant.

To sum up, these results reveal a strong effect of syllable complexity on elderly Japanese listeners’ ability to count syllables in the speech stimuli. However, identification training substantially improves listeners’ performance, virtually eliminating the effect of final consonants. The effect of initial consonants is still evident, but appears to be smaller in magnitude compared to the pretest.

4. DISCUSSION

Results obtained in this study for elderly native Japanese listeners are similar to the perceptual tendencies of English syllable perception observed for college-aged Japanese listeners. The improvement in accuracy obtained for college-aged listeners (53% to 90%) [8] is comparable to the improvement obtained here (56% to 91%). Also, training led to a substantial decrease in the effect of initial and final consonants on college students’ performance, but there remained a noticeable effect of initial consonants even after training [9, 8], as was also found in the present results. In contrast, elderly Japanese listeners trained on English /l/-/l/ perception appeared to benefit less from training than did younger listeners [3, 5]. According to these results, then, it appears that age does not affect perceptual learning of L2 syllables as much as it affects the learning of L2 speech sounds.

Performance of individual elderly listeners was also investigated. In the pretest, the scores ranged between 40% and 75% depending on the listener (SD = 11.4). In contrast, scores in the posttest were between 90% and 100% for 10 out of the 12 listeners. One of the two remaining listeners had a score of 83% in the posttest. Surprisingly, the other listener showed no improvement in performance even after 15 sessions of training; the listener scored 45% in the pretest and 46% in the posttest. According to the hearing test, this listener showed some hearing loss (25–30 dB HL) at 500 and 4000 Hz. However, it is unlikely that this is responsible for the lack of improvement in accuracy, since other elderly listeners with more severe hearing losses (40–55 dB HL) nevertheless showed scores above 90% in the posttest. These data, therefore, do not seem to reveal a straightforward relationship between the elderly listeners’ hearing and their performance at the syllable-counting task. Data from additional listeners are currently being collected to obtain a better picture of how these listener-related variables affect performance.

5. CONCLUSIONS

This study examined the extent to which perceptual identification training using feedback improves the perception of syllables in spoken English words by elderly native Japanese listeners. Results reveal that even listeners who are in their 60s can benefit from perceptual training and significantly improve their ability to count syllables in the stimuli, including nonwords and real words that they were not trained on. These results suggest that perceptual training using feedback is effective for training not just contrastive properties of spoken L2 utterances, but also global rhythmic properties as well. The results also suggest that aging exerts noticeable influences on perceptual learning of some properties of L2 speech, while it has relatively weak effects on perceptual learning of other properties.

6. REFERENCES