Differences of Pronunciation Practices:
A Study of “Repeat with me” and “Repeat after me”

Katsumi NAGAI*

1. Introduction

Teachers of foreign languages often have their students repeat their target pronunciation individually or in chorus. In radio language courses in Japan (NHK 2001), for example, French and English courses ask learners to repeat target words and phrases after a model reading by teachers. Korean and Chinese courses also require learners to repeat after teachers, but the teachers repeat the target phrases simultaneously with the learners. In the Korean course, a teacher slowly presents the model pronunciation first, and then the teacher repeats a faster model reading simultaneously with the learners. This type of pronunciation drill is done in the Chinese course too, but the teacher who pronounces the model reading, and the teacher who gets the learners to repeat target phrases simultaneously, are different. This example indicates that there are various transformed versions of these pronunciation drills. Tasks done in French and English courses (“Repeat after me”-type drills) are widely performed in a classroom setting, but it is not clear why teachers of Korean and Chinese courses repeat their model phrases with their learners, nor is it clear why teachers of Korean courses speak slowly in the first model presentation and then reduce their tempo the second time, nor why in the Chinese course different teachers read the first and second target phrases.

However, teachers of the radio courses recognize that imitating a native-speaker’s model pronunciation facilitates acquisition of desirable pronunciation. Computer-assisted instruction is useful for extracting acoustic parameters such as pitch, amplitude, formants, and their transitions, but essentially language learners must imitate the model pronunciation repeatedly. Teachers need the fundamental information about what type of pronunciation drills their students can do to improve pronunciation in the target languages. However, there are few quantitative analyses of teaching methods or commonly accepted opinions as to what type of pro-
nunciation tasks are more effective than others.

Celce-Murcia, Brinton, & Goodwin (1996, p.310) classify classroom tasks of pronunciation into “mirroring,” which requires imitation of teacher’s gesture, eye-movement, and posture, “tracking,” which concentrates on voice only, and “shadowing,” which allows a little time lag behind the teacher’s utterances. Because other terms, including “repeating,” are also used, it is necessary to define the relevant terms before further discussion. The present author introduces new terminology “a-repeat,” “w-repeat,” and the others to clarify the difference among repetitions in language classrooms in Table 1.

A-repeat is a basic task of pronunciation practice. It requires learners to imitate their teacher’s model pronunciation as precisely as possible immediately after they listen to the model. Learners try to reproduce target words and phrases by referring to their short-term (working) memory, which retains the teacher’s model pronunciation. It means that the learner’s own voice and its bone (ossaeous) conduction is the only simultaneous auditory feedback for the control of the learner’s speech. Because there naturally exists a time lag of some hundreds of milliseconds between the end of the model presentation and the beginning of the learners’ repetition, teachers need to set an equivalent span of pause between model readings. Although the length of pauses is considered to be under control of various psychological, biological, and physical factors, little quantitative research is available to decide how long is needed when the teachers tape their teaching materials. This question leads to the aim of Experiment 1. Reaction time, of which the origin is Helmholtz’s measurement of time between the stimulation of frog’s nerve and its muscular reaction (Brauzier 1959), is defined as the interval time between presentation of a stimulus and detection of the response. It corresponds to a time lag between the beginning of the teacher’s model pronunciation and the beginning of the learner’s repetition in classroom setting. Reaction time varies with the complexity of the stimuli and responses. Donders (1968) closely examined reaction times while changing the complexity. For example, (1) when subjects were requested to listen to a monosyllabic stimulus /ki/ and to repeat the same stimulus /ki/, the reaction time was the shortest (201ms). (2) When subjects were requested to listen to one monosyllabic stimulus out of /ki, ke, ka, ko, or ku/, and to repeat /ki/ only after they hear /ki/ (i.e., to repeat /ki/ after the subjects hear /ki/, but not to repeat any syllable after the subjects hear /ke, ka, ko, or ku/), the reaction time was slightly longer (237ms). (3) When the subjects were requested to listen to monosyllabic stimuli from /ki, ke, ka, ko, ku/, and to repeat each syllable as they hear it (to repeat /ki/ after subjects hear /ki/, and to repeat /ke/ after /ke/, etc.), the reaction time was much longer (284ms). According to Donders (1868, reprinted in 1969), the difference between (1) and (2) implies time to discriminate stimuli, and the difference between (2) and (3) suggests time for decision-making. Although this type of subtraction method was incompatible with the later framework of Gestalt psychology, which placed emphasis on the totality of the stimulus, his pioneering work has provided a vital clue for explaining the relationship between complexity of

<table>
<thead>
<tr>
<th>terms</th>
<th>examples of directions</th>
<th>mode of model presentation</th>
<th>presentation before imitation</th>
<th>presentation with imitation</th>
<th>learners’ imitation</th>
<th>notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>a-repeat</td>
<td>Repeat after me.</td>
<td>audio</td>
<td>yes</td>
<td>no</td>
<td>at learners’ pace</td>
<td>most basic task</td>
</tr>
<tr>
<td>w-repeat</td>
<td>Listen and repeat with me.</td>
<td>audio</td>
<td>yes</td>
<td>yes</td>
<td>slightly delayed</td>
<td>learners and teachers speak together</td>
</tr>
<tr>
<td>m-repeat</td>
<td>Repeat with me.</td>
<td>audio</td>
<td>no</td>
<td>yes</td>
<td>can precede teachers</td>
<td>repeat from memory</td>
</tr>
<tr>
<td>e-repeat</td>
<td>Repeat with me.</td>
<td>audio</td>
<td>no</td>
<td>yes</td>
<td>delayed by a few syllables</td>
<td>called “shadowing” or “echoing”</td>
</tr>
<tr>
<td>s-repeat</td>
<td>Read with me.</td>
<td>audio-visual</td>
<td>no</td>
<td>yes</td>
<td>can precede teachers</td>
<td>e-repeat with visual presentation</td>
</tr>
<tr>
<td>recitation</td>
<td>Recite from memory.</td>
<td>none</td>
<td>no</td>
<td>no</td>
<td>at learners’ pace</td>
<td>repeat from memory</td>
</tr>
<tr>
<td>buzz session</td>
<td>Practice freely.</td>
<td>none</td>
<td>no</td>
<td>no</td>
<td>at learners’ pace</td>
<td>masking by other learners’ voices</td>
</tr>
</tbody>
</table>
stimuli and reaction time. Reaction time is longer when auditory stimuli are weak in much the same way as those of visual and tactile stimuli are (Chocholle 1963, Welford 1980). It is also longer when subjects are not familiar with the stimuli (Connine et al. 1990, Amano & Kondo 1998). However, as for the modes of stimulus presentation, while some researchers conclude that reactions to audio stimuli are shorter than to visual stimuli (Forbes 1945, Goldstone & Lhamon 1975), other researchers report that there is no difference between reaction times to auditory stimuli and visual stimuli (Craig 1973, Bobko, Thompson & Schiffman 1977).

W-repeat is a task which gets learners first to listen to a teacher and then to repeat target phrases simultaneously with the teacher. Because teachers give the model pronunciation twice, learners are exposed to the model reading twice as many times. This task is an effective method for learners who are too shy to speak aloud in class, according to the author’s introspective view. The author’s unpublished feasibility study indicates that a-repeat is more effective than w-repeat. After the teachers’ first model reading, learners hold the auditory image in their working memory, concurrently planning how to articulate their target phrases. Then the learners start pronouncing the phrase almost at the same time with the teacher. Learners are required to control the place and manner of articulation with reference to the teacher’s model reading. They need to compare simultaneously (1) the teacher’s auditory remnants in their own memory, (2) direct auditory input of the teacher’s second model reading from their ears, and (3) auditory and bone-conducted feedback of their own voice. Unlike the a-repeat task mentioned above, learners cannot control the lag time before launching their repetition.

The following relevant forms of pronunciation practice are additionally introduced to restrict the definitions of a-repeat and w-repeat. M-repeat is a simple task of saying target phrases with a teacher. This lacks the first teacher’s presentation of w-repeat. Because no model reading precedes the learner’s imitation, learners lag behind the teacher unless the learners have memorized the target phrases. On the other hand, learners occasionally begin their pronunciation earlier than teachers in the case of knowing the target phrases by heart. E-repeat is a variation of m-repeat. It allows learners a delay of a few syllables, which enables them to imitate the target phrases without memorizing the task beforehand. This form of task is often called “shadowing” or “echoing,” and is popular among teachers who believe the effect of teaching is proportional to the fatigue they feel after a long shadowing task. S-repeat is a variation of w-repeat. It allows learners to look at a blackboard or a textbook during practice. This task can involve fundamentally different cognitive processes because it is tied to variability in the learners’ skills of reading. Pronunciation tasks involve such complicated psychological procedures that it is not practical to investigate all of the varieties defined in Table 1 above. For the sake of simplicity, only a-repeat and w-repeat are studied in the present report.

2. Experiment 1

2.1 Aim

The aim of Experiment 1 is to find practical criteria for determining pauses between the target phrases of language teaching. Table 2 sorts out the implications for language learners derivable from the previous research on reaction time. In order to elicit a quick response from the learners when recording their teaching materials, language teachers need to (1) raise their voice and repeat (2) simple and (3) already-learned phrases. Although the aim of a pronunciation drill is not shortening reaction time, it is necessary to examine it because it corresponds to the duration between the

<table>
<thead>
<tr>
<th>Previous research reveals:</th>
<th>In a classroom:</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) reaction times to strong stimuli are shorter (Chocholle 1963)</td>
<td>teachers’ louder phrases elicit a speedy reaction from learners.</td>
</tr>
<tr>
<td>(2) reaction times to simple stimuli are shorter (Donders 1868)</td>
<td>shorter phrases elicit a speedy reaction from learners.</td>
</tr>
<tr>
<td>(3) reaction times to familiar stimuli are shorter (Amano &amp; Kondo 1998)</td>
<td>already learned phrases elicit a speedy reaction from learners.</td>
</tr>
</tbody>
</table>
beginnings of teacher’s model reading and learners’ utterances. Since the complexity of test words increases as stimuli have longer phonological length, the relationship between reaction time and phonological length of the test words is examined in this experiment.

2.2 Method

Reaction times to Japanese and English test words were measured in this experiment. Subjects were presented with auditory stimuli differentiated on the basis of their phonological length. The independent variables were the languages of the test words (two levels: English and Japanese) and the phonological length in syllables (one to five syllables). The dependent variable was reaction time, which spanned from the end of the stimuli to the start of the subject’s repetition.

Japanese monosyllabic test words /be/, /ba/, and /bo/ were repeatedly spoken by a native speaker of Osaka Japanese, aged 38. The English set of test words was taped by a native speaker of British English, aged 36, who taught English at a University in Japan. Their pronunciations of Japanese and English syllables were digitally taped in a soundproof room with a condenser microphone and a DAT recorder (Sony ECM-MS957 and TCD-D7 at 48KHz/16bit sampling). With a digital-analog interface card (Creative SBLPT 1394) attached to a computer and sound edit software (Kay Multi-speech 3700 version 2.3), the experimenter made graphs of waveforms and intensity envelopes on a computer screen, and selected the most suitable test words that had the most similar length. The original syllables in both English and Japanese varied only slightly in fundamental and formant frequencies, vowel duration of all syllables was approximately 360ms. The digitized sounds were edited and re-synthesized to make nonsense test words listed as shown in Table 3. Figure 1 shows a waveform of the English test word /bæbæbə/. After a cue synthesized with different software (IBM Pro Talker V8) and a pause of 500ms, each test word was randomly set at intervals of 3000ms. The completed sets of stimuli were again recorded on an analog cassette tape for a multi-track recording.

The stimuli were presented to six female Japanese

![Fig. 1 Time domain waveform (top) and spectrogram (above) of English test word /bæbæbə/](image-url)
Differences of Pronunciation Practices: A Study of “Repeat with me” and “Repeat after me”

learners of English through a headphone (Sony DR-200 headset). The learners were speakers of Osaka dialect, aged 18 to 23. All participant learners in the present paper, who started learning English as a foreign language in Japan at a mean age of twelve, had normal hearing and were at either a pre-intermediate or intermediate level of English according to their placement test scores.

After a two-minute practice session, subjects were asked to repeat the test words after the cues. The instruction was to repeat the test words as naturally as possible. The subjects were not required to repeat quickly because the aim of this experiment was to investigate whether natural reaction time was affected by the complexity, i.e., phonological length of stimuli.

The stimuli and learners’ pronunciation practice were simultaneously taped on a multi-track recorder (Sony ER-8020) using a microphone attached to the headset. The recorded stimuli and pronunciation practice were again digitized for acoustic measurement. The time between the end of the stimulus and the beginning of a learner’s repetition was measured on the display.

2.3 Results

Mean reaction times of six learners are shown in Figure 2. Analysis of variance was conducted without logarithmic transformation because the experimental design does not require the subjects to react as quickly as possible. The result revealed that the number of syllables had a significant main effect ($F_{(4,35)}=14.399$, $p<.01$). According to Scheffe’s multiple comparison, monosyllabic English test words required shorter reaction times than words which had more than two syllables (monosyllabic words: $p<.05$, the others: $p<.01$).

For Japanese test words, however, two and five syllables needed longer reaction times than the words which had one, three, and four syllables (two syllables - four syllables: $p<.05$, others: $p<.01$). Perhaps the reason that reaction times for trisyllabic words are shorter than that for two or four syllables is related to the fact that bimoraic syllables play a phonologically important role in Japanese language (see e.g., Bekku 1977 and Poser 1990). That is, because Japanese bimoraic test words are self-contained, a longer reaction time is required to launch articulation of the additional third syllable.

Whether the test words were English or Japanese made no difference in reaction time ($F_{(1,4)}=0.814$, $p=.418$), but interaction of the difference between English and Japanese with the reaction time was observed ($F_{(4,170)}=6.562$, $p<.01$).

When making an audio recording for pronunciation practice, it is a common practice for recording technicians to set pauses among target words relying on the engineers’ own judgment. However, as clarified in this experiment, reaction time varies with the phonological length of target words and learner’s native language. For Japanese learners of English, it is better to allow more than 600ms for repetition of monosyllabic and bisyllabic words, and to allot about 800ms for words longer than two syllables.

3. Experiment 2

3.1 Aim

Pronunciation practice can be classified into many categories as illustrated by Table 1 of the preceding section. In Experiment 2, a-repeat and w-repeat techniques
are examined for simplicity. When learners do a-repeat, they listen to stimuli through headphones and repeat the test words after they hear the sound. When learners do w-repeat, they listen to stimuli through headphones and repeat the test words simultaneously with the second stimulus they hear. The principal difference between a-repeat and w-repeat is the teacher's participation (i.e., second model reading) during learners' repetition. While learners are doing w-repeat, their voice overlaps with the teacher’s voice. This enables the learners to compare their own pronunciation simultaneously with the teacher’s model. As a result, (or, if the learners make effective comparison,) w-repeat might lead to pronunciation which is evaluated as “very good” by native speakers of the target language. On the contrary, if overlapping prevents learner’s getting precise feedback of their own voice, w-repeat might result in poor reproduction of the target words. Experiment 2 is designed to test the hypothesis that, in terms of naturalness of production, there is no significant difference between a-repeat and w-repeat when Japanese learners of English do their pronunciation practice of English words varying in phonological length.

3.2 Method

The test words consisted of six nonsense words (/afa/ - /aαa/ , /aθa/ - /aθa/, and /aαa/ - /aαa/) and six English pairs (bathe - bays, best - vest, and major - measure). Labio-dental and dental consonants /θ/ , /s/ , /θ/ , and /θ/ are used only in English and often substituted with /θ/ , /b/ , /s/ and /z/ respectively by Japanese learners of English. Japanese alveolar approximant /r/ is also phonetically different from either English /r/ or /l/.

Recordings of nonsense words were taken from Wells and House (1995), which was a part of the set textbooks for the phonetics course taken by the learners. English test words were taped by a native speaker of British English using the same equipment described in Experiment 1. Both recordings were entered into a computer and edited digitally to make an audiocassette tape for pronunciation practice. All test words were randomized and repeated eight times. Duration between the beginning of the first stimulus and second stimulus for w-repeat was set at 800ms, and pauses between each token of the test words were set at two seconds.

Eight Japanese learners of English participated in the experiment with nonsense words. Five more learners, who made the total number of participants thirteen, took part in the experiment of English test words. All participants, aged 19 to 24, were native speakers of Japanese. After a brief explanation and practice for ten minutes, learners were told to a-repeat and w-repeat test words. Learners’ speech sounds were taped on a multi-track recorder simultaneously with the stimuli and digitized with the same computer as described in Experiment 1.

Learners’ second tokens of eight trials were used to make a recording for evaluation of naturalness. The tokens were randomized within each set of test words. Pauses between each word were set at one second. The stimuli were taped on a digital recorder (SONY TCD-D100) and presented to six native speakers of Southern British English aged 27 to 44. The number of native speakers who participated in the nonsense word experiment was five. The evaluation experiment was carried out in a quiet classroom at a British university and in the halls of residence. After a simple questionnaire about their birthplace and dialects, they were instructed to wear headphones (Senheisser HD-2SP) and to evaluate the naturalness of the test words. They listened to test words randomly for a few minutes as a practice session, and then listened to the digital audio tape to grade the naturalness of the learners’ pronunciation on a scale of one to seven. The total score was 70 for nonsense test words (5 native speakers x 7 points x 2 trials) and 84 for English test words (6 native speakers x 7 points x 2 trials).

3.3 Results

Scores of nonsense words and English words are plotted in Figure 3. Pearson’s correlation coefficient was calculated as shown in Table 4 under the assumption that the data fulfilled the condition of normal distribution and interval scale. It is assumed that the lower correlation of the test words /aθa/ , vest, and major was created from outliers in Figure 3, and therefore it is appropriate to conclude that there is a correlation between a-repeat and w-repeat.

Mean scores of naturalness of nonsense words and English words are shown in Figure 4. Repeated measure analysis of variance proved that scores of nonsense word /afa/ were better than that of /aαa/ (F(2,8) =453.08, p<.05) and that /aαa/ was better reproduced than /aθa/ (F(2,8) =4356, p<.01). Scores of the English test word best are significantly higher than those of vest (F(1,14) =14.82, p<.01), and major outperforms its counterpart, measure (F(1,14) =12.63, p<.01). Significant differences were found between a-repeat and w-repeat (p<.05) in English test words except for the best and major pairs. No significant difference between a-repeat and w-repeat was detected among the nonsense test words.
The result that Japanese learners’ /arə/ was more highly evaluated than /aɪə/ is consistent with the research that shows /r/ in American English is more rapidly acquired than /l/ by American children, in terms of both perception and production (Aoyama et al. 2004). With regard to the distinction between /l/ and /r/, it is known that American English as well as British speakers perceive /l/ mainly by the third formant (F3) being higher than 2KHz, whereas Japanese learners of English use the second formant (F2) as well as F3 (Yamada & Tohkura 1991, 1992). The most highly evaluated learner’s formant transitions of the test words /aɪə/ and /arə/ are shown in Figure 5; lowest scores are indicated in Figure 6. We assume that the learner’s high F3, as shown in Figure 5, enabled native speakers to perceive the English /l/ sound easily. In the data of Figure 6, on the other hand, a rapid transition of F3 enabled native speakers’ perception of British /r/. The differ-

---

**Table 4  Pearson’s correlation coefficient between a-repeat and w-repeat in Experiment 1**

<table>
<thead>
<tr>
<th>afa</th>
<th>ava</th>
<th>aθa</th>
<th>aθa</th>
<th>aθa</th>
<th>aθa</th>
<th>ala</th>
<th>bathe</th>
<th>boys</th>
<th>best</th>
<th>vest</th>
<th>major</th>
<th>measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.70</td>
<td>0.91</td>
<td>0.83</td>
<td>0.18</td>
<td>0.95</td>
<td>0.87</td>
<td>0.65</td>
<td>0.79</td>
<td>0.87</td>
<td>0.31</td>
<td>0.32</td>
<td>0.95</td>
<td></td>
</tr>
</tbody>
</table>

---

**Fig. 4** Mean scores of test words across all learners (a=a-repeat, w=w-repeat). Error bars represent the standard error of the mean.
ences between Figure 5 and Figure 6 imply that difference in formant transition can be a clue for perceiving English /l/ and /r/ by native speakers of British English. In Figure 5, the learner’s F2 of w-repeat changes more swiftly than that of a-repeat. Since a swift change in F2 can be an indicator of lip-rounding, perhaps w-repeat enables students to notice the lip-rounding necessary for a “good” /r/. This example suggests the possibility of applying the difference between a-repeat and w-repeat to finding learners who need special training.

Reaction time between end of stimuli and start of learner’s pronunciation was also examined as in Experiment 1. Figure 7 indicates the reaction time of nonsense test words and English test words. Because one-way analysis of variance revealed that the reaction time of nonsense test words had a significant main effect \((F_{(8,35)}=4.87, p<.01)\), Tukey’s multiple comparison test was used to determine if there was significant difference between latencies of each pair of test words. However, there was no significant effect between /afa/- /ava/, /əθa/- /əθa/, or /ara/- /ala/. English test words also had a significant main effect \((F_{(7,75)}=4.73, p<.01)\) and Tukey’s test revealed that reaction time for measure was shorter than that for major \((p<.01)\). These results coincide with the results of Experiment 1 that bisyllabic test words require latency of 600ms or more.

4. Experiment 3

4.1 Aim

This experiment examines differences between a-repeat and w-repeat using English test sentences. Japanese learners of English were presented two sentences which had strong-weak and weak-strong stress patterns. The learners were asked to repeat the test sentences
Differences of Pronunciation Practices: A Study of “Repeat with me” and “Repeat after me”

**Fig. 7** Reaction time of a-repeat and the standard error of nonsense test words in Experiment 2

**Table 5** Time structure of test sentences in Experiment 3 in ms

<table>
<thead>
<tr>
<th></th>
<th>he</th>
<th>wants</th>
<th>come</th>
<th>see-us</th>
<th>ho-</th>
<th>me#</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>−109</td>
<td>0</td>
<td>362</td>
<td>366</td>
<td>361</td>
<td>379</td>
<td>1555</td>
</tr>
<tr>
<td>leave</td>
<td>rest</td>
<td>food</td>
<td>lun-</td>
<td>ch#</td>
<td></td>
<td></td>
<td>total</td>
</tr>
<tr>
<td></td>
<td>−38</td>
<td>0</td>
<td>388</td>
<td>509</td>
<td>370</td>
<td>384</td>
<td>1689</td>
</tr>
</tbody>
</table>

after (a-repeat) and with (w-repeat) the model presentation. Their practices were taped and edited, then presented to native speakers of English for evaluation tasks. The null hypothesis is that no significant difference is observed between a-repeat and w-repeat.

**4.2 Method**

Two test sentences, “he wants to come and see us at home” and “leave the rest of the food for lunch” were taken from Roach (2001). The learners were quite familiar with the sentences because Roach (2001) was the main textbook of their phonetics course. Time structures of the test sentences are listed in Table 5. Figure 8 shows the time-domain waveform, its power and pitch contours, and the formant transitions. In Experiment 3, beginning of the test sentence is defined as a median of the two points on the amplitude contour set between the beginning and the peak of the first rise. Two test sentences were edited with the same equipment as used in Experiment 1, setting a pause time of 5 seconds for a-repeat and 2.5 seconds for w-repeat. The test sentences were recorded on a cassette tape and presented to 22 Japanese learners of English. After a practice of two minutes, learners listened to the tape through headphones and repeated the sentences after and with the models for about ten minutes each. Learners’ repetitions were taped simultaneously with the model sentences with a microphone attached to the headphones.

For the evaluation test, the second trials of the two test sentences were randomly rearranged and recorded.
on a digital tape with a five-second pause before each sentence. The test took two pages and each page consisted of six blocks. Each block contained fifteen scales of seven grades. Because the first and the last items of the evaluation experiment were fillers, 88 sentences were randomly presented to eight native speakers of British Southern Standard English aged from 27 to 41. The native speakers were asked to listen to the digital audio tape in a quiet room and to grade the naturalness of test sentences to seven ranks. Instruction given before the evaluation task was the same as used in Experiment 2. A full score for the learners is 56 points because eight native speakers evaluated the sentences on a scale of one to seven.

4.3 Results

Mean scores of 22 learners are shown in Figure 9. Mean score of all learners is 32.5 (S.D =10.66) for a-repeat of “he wants....” test sentence, and 30.05 (S.D=10.01) for w-repeat of the same sentence. Mean score of “leave the rest....” sentences is 32.09 (S.D.=7.37) for a-repeat, and 30.41 (S.D.=6.14) for w-repeat. Paired t-tests revealed that a-repeat achieved significantly better scores (p<.01) in both cases of “he wants....” sentence ($t_{(21)}=4.23$) and “leave the rest....” sentence ($t_{(21)}=4.23$). However, there was no significant difference in scores between “he wants....” and “leave the rest....” test sentences (a-repeat: $p=0.62$, $t_{(21)}=0.48$, w-repeat: $p=0.69$, $t_{(21)}=0.40$). The scores are plotted in Figure 10, which suggests the scores of a-repeat and w-repeat are positively correlated. Pearson’s correlation coefficient is 0.521 for the “he wants....” test sentence, and 0.401 for the “leave the rest....” test sentence. A-repeat and w-repeat of “he wants....” has a significant correlation ($p<.05$), but “leave the rest....” sentence does not ($p=0.189$).

Lag time of learner’s repetition was measured by the same method as used when the test sentences were synthesized on a computer. Latencies in Figure 11 show lag time measured from the beginning of 2.5-second frames, which was set as a half of 5 seconds intervals for a-repeat. The beginning of the test sentence indicates the onset of the total duration of the test sentences listed at Table 5. Figure 12 and Figure 13 show time structure of recorded sentences. The time labeled “wants,” “come,” “see us,” and “home” shown in each figure refer to duration between centers of amplitude peaks in the repeated sentence. Bottom bars labeled “org” in the figures are time structures of the native speaker’s original stimulus used in the experiment. Because learners were arranged in ascending order of their scores, the learner “F” at the bottom of the graph got the best score from native speakers, and the learner “K” at the top of the graph got the worst score. It may be inferred from the figures that similar intervals among the intensity peaks, as seen in the speech labeled “org.”
have a tendency to yield a higher score. It is also true that these isochronous peaks in intensity are not enough to explain the different order of naturalness in Figure 12 and Figure 13.

5. Experiment 4

5.1 Aim

Experiment 2 showed that there was no significant difference between a-repeat and w-repeat when nonsense words were repeated. Experiment 2 also indicated that a-repeat of some English test words outscored w-repeat. As for repetition of sentences examined in Experiment 3, it can be concluded that a-repeat of test sentences improved learners' scores of naturalness more than w-repeat did. The aim of experiment 4 is to examine whether the same advantage of a-repeat is seen in a-repeat of "nonsense" test sentences. Nonsense sentences in this experiment were synthesized reproducing stress patterns of the English test sentences used in Experiment 3. The stress patterns were formed with clicks at the middle points of intensity contours of the sentences. Learners were asked to a-repeat and w-repeat the click patterns by repeating the nonsense monosyllabic word "ta." The reproduced rhythm patterns were recorded and the precision of the timing was quantitatively measured.

5.2 Method

Intensity contours of the two test sentences ("he wants to come and see us at home" and "leave the rest of the food for lunch," both used in Experiment 3) were produced on a computer using the same equipment as in the previous experiments. On each contour, four points of each of the test sentences ("wants," "come," "see," "home," and "leave," "rest," "food," "lunch") were set at the middle points between the beginning and end of the four most salient rises. Clicks were generated from sine wave at 100Hz for 10ms and placed at the four points as shown in Figure 14. Intervals for a-repeat and w-repeat were set at 5 and 2.5 seconds, respectively.

The two stimuli were presented through headphones to sixteen Japanese learners of English. All learners also participated in Experiment 3. After a few minutes
of practice, they first a-repeated the beep patterns for ten minutes. Then they w-repeated the patterns for another ten minutes. Their repetitions of monosyllabic word “ta” were recorded by a multi-track recorder with the original stimuli and then were digitized using the same devices as in the previous experiments.

Intensity curves of learners’ repetitions were displayed on a computer screen with the spectrograms and pitch contours. Time lags between sinusoidal rhythm patterns and learners’ repetition of “ta” were measured on a computer display. Precision of timing was analyzed using a criterion for judging whether “ta” syllables reproduced by the learners fell chronometrically within a ten percent margin of error. This criterion was adopted because Saito and Ishio (1997) and Nagai (1999) succeeded in testing preciseness of rhythm reproduction with similar methods. Because one test sentence has four peaks and five trials per subject, the total score was 20.

5.3 Results

Results of Experiment 4 are shown in Figure 15. The mean score of all the learners is 13.75 (S.D.=2.35) for a-repeat of the “he wants...” pattern, and 12.68 (S.D.=3.66) for w-repeat of the same pattern. The mean score of the “leave the rest...” pattern is 10.5 (S.D.=3.26) for a-repeat, and 9.0 (S.D.=1.78) for w-repeat. A-repeat surpasses w-repeat in average scores. However, no significant difference was observed between scores of a-repeat and w-repeat for either sentences ($p=0.09, t_{(15)}=1.76$ for the pattern of “he wants...” sentence, $p=0.22, t_{(15)}=1.26$ for the pattern of “leave the rest...” sentence). Note that for repetition of English sentences (Experiment 3), a-repeat was significantly better than w-repeat; however, for repetition of rhythm

---

Fig. 14 Intensity contour (top), time-domain waveform (middle) of the test sentence and click patterns for w-repeat (bottom)

Fig. 15 Scores of rhythm reproduction in Experiment 4 (error bars indicate +/-2SE)

Fig. 16 Correlation between a-repeat and w-repeat of nonsense rhythm patterns in Experiment 4
Differences of Pronunciation Practices: A Study of “Repeat with me” and “Repeat after me”

(Experiment 4), no significant difference was found. This implies that repetition of a rhythm pattern is a different behaviour from repetition of sentences. Figure 16 suggests a correlation between a-repeat and w-repeat, but more data are needed in order to test for significance.

6. Discussion

Because repetition after a teacher and repetition with a teacher are the most widely used ways of pronunciation practice in a language classroom, the aim of this paper was set to measure the effectiveness of the two types of pronunciation practice, using four experiments which focus on the temporal factor. In Experiment 1, Japanese learners of English were asked to imitate English and nonsense syllables, and their reaction time was measured. Average latency before launching repetition fell between 600ms and 900ms, varying with the phonological length of the test words. In Experiment 2, pairs of nonsense and English words, which included difficult phonemic distinctions for Japanese learners of English such as between /l-r/ and /v-b/, were both a-repeated and w-repeated. In Experiment 3, pairs of English sentences, which had strong-weak and weak-strong stress patterns, were a-repeated and w-repeated. In both experiments, learners’ practices were recorded and evaluated by British teachers of English. In Experiment 2, the results of nonsense words revealed that there was no significant difference in naturalness between a-repeat and w-repeat. However, a-repeat surpassed w-repeat in Experiment 3. In Experiment 4, subjects were presented the identical stress patterns of test sentences but reproduced with short sinusoid waves. The learners a-repeated and w-repeated the patterns by pronouncing the monosyllable “ta.” The results showed no difference between a-repeat and w-repeat, which implies that repetition of sentences is different from that of nonsense words.

The naturalness of English sentences pronounced by learners was examined in Experiment 3, and the rhythm patterns, generated with spotted sine waves based on the power contours of the original test sentences, were investigated in Experiment 4. The results in Experiment 3 were based on naturalness scores given by native British listeners; the scores obtained in Experiment 4 were calculated from differences in time of speakers’ productions. Comparison of the timing structure of the test sentences repeated by the participants in Experiment 3 with those of the results of Experiment 4 was done in order to ascertain to what degree is the timing structure of the original test sentences maintained by learners.

The learners’ repetitions recorded in Experiment 3 were again digitized and intensity contours were drawn to obtain four peak points of the sentences. The mid point in time, from the beginning to end of rises at four major peak points (“wants,” “come,” “see,” “home,” and “leave,” “rest,” “food,” “lunch”), was decided from a computer display. Scoring learners’ repetitions from
Figure 18 indicates the correlation between Figure 15 and Figure 17. Pearson’s correlation coefficient between a-repeat and w-repeat of the test sentence “he wants....” (r=0.45, shown by “□”) and those of ta-repetition (r=0.47, shown by “△”) was higher than those of the “leave the rest....” test sentence (r=0.19, shown by “●”) and its rhythm pattern (r=0.26, shown by “×”). Note that some learners, who repeated ta precisely, could not repeat English sentences precisely. Figure 19 indicates correlations between ta-repetition and sentence-repetition both scored from the viewpoint of precision timing. The correlation is highest when the “he wants....” test sentence is a-repeated (r=0.52), and little correlation is observed between the “leave the...” sentence and its ta-repetition. This might indicate that repetition of a weak-strong stress pattern (i.e., the pattern of the “he wants...” sentence) has much in common with repetition of “ta”.

W-repeat requires learners to listen to their own voices during the practice. It forces them to process and manipulate input and output of information during their repetition. Therefore, language learners may tend to feel a sense of accomplishment and comfortable mental fatigue, and teachers are inclined to conclude that w-repeat is more effective than a-repeat. However, w-repeat is a task which hinders learners’ precise auditory feedback by its random and nonstationary noise including the learners’ own voices. Taking all results in this paper into consideration, it is quite likely that w-repeat is not the best way for a practice of natural sentences.

7. Conclusion

In the present paper, four experiments were done concerning English spoken by Japanese learners.
Experiment 1 showed that reaction time differed according to the number of syllables in test words; specifically, pauses of 0.5 second for teaching materials are too short, and that pauses of one second are too long. While Experiment 2 revealed that a-repeat and w-repeat of nonsense test words did not make a significant difference, Experiment 3 proved that a-repeat of test sentences produced significantly better scores. In Experiment 4, rhythm patterns of the test sentences were a-and w-repeated using “ta” syllables, and no significant difference was found between a-repeat and w-repeat. This suggests that repetition of sentences and that of nonsense words may involve different processes. The results of this paper based on this sample of learners is that a-repeat may be a better method for Japanese learners of English than w-repeat.

8. Acknowledgments

The author thanks two anonymous reviewers for their very useful comments on a previous version of this paper. This research was supported by Grants-in-Aid for Scientific Research #17530678 (Ministry of Education, Culture, Sports, Science and Technology, Japan). Part of the research was presented at the 17th Annual Convention of the Phonetic Society of Japan.

References


(Accepted May, 2007)