Training native English speakers to identify Japanese vowel length contrast with sentences at varied speaking rates

Yukari Hirata, a, b Elizabeth Whitehurst, and Emily Cullings
Colgate University, Department of East Asian Languages and Literatures, Hamilton, New York 13346

(Received 18 August 2006; revised 26 March 2007; accepted 26 March 2007)

Native English speakers were trained to identify Japanese vowel length in three types of training differing in sentential speaking rate: slow-only, fast-only, and slow-fast. Following Pisoni and Lively’s high phonetic variability hypothesis [Pisoni, D. B., and Lively, S. E., Speech Perception and Linguistic Experience, 433–459 (1995)], higher stimulus variability by means of training with two rates was hypothesized to aid learners in adapting to speech rate variation more effectively than training with only one rate. Trained participants identified the length of the second vowel of disyllables, short or long, embedded in a sentence of the respective rate, and received immediate feedback. The three trained groups’ abilities before and after training were examined with tests containing sentences of slow, normal, and fast rates, and were compared with those of a control that was not trained. A robust effect of slow-fast training, a marginal effect of slow-only training, but no significant effect of fast-only training were found in the overall test scores. Slow-fast and slow-only training showed small advantages over fast-only training on the fast-rate test scores, while effects for all three training types were found on the slow- and normal-rate test scores. The degree to which the results support the high phonetic variability hypothesis is discussed. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2734401]

PACS number(s): 43.71.Hw, 43.71.Es [ARB] Pages: 3837–3845

I. INTRODUCTION

This study examined the way non-native speakers perceive Japanese vowel length contrast in sentences produced at various speaking rates. The vowel length distinction (short versus long) is phonemic in Japanese, e.g., /kuro/ “black” versus /kuro/t/ “hardships,” and all five short vowels /i e a o u/ contrast with the corresponding long vowels (/iti et at ot ut/) (Vance, 1987). Long vowels are 2.2–3.2 times longer in duration than short vowels, whereas only small differences have been observed between the formant frequencies of short and long vowels (Kondo, 1995; Tsukada, 1999; Ueyama, 2000; Hirata, 2004a; Hirata and Tsukada, 2004). This vowel length distinction, as well as a consonant length distinction (e.g., /kata/ “shoulder” versus /katat/ “won”) has been found difficult for native English (NE) speakers to acquire (e.g., Landahl et al., 1992; Han, 1992; Landahl and Ziolkowski, 1995; Yamada et al., 1995; Toda, 1997; Oguma, 2000; Tajima et al., 2002; Hirata, 2004b). Native Japanese speakers’ primary perceptual cue to the Japanese vowel length distinction is vowel duration (Fujisaki et al., 1997), but a long vowel spoken quickly can be shorter than a short vowel spoken slowly (Hirata, 2004a). Speaking rate variation is thought to contribute to the difficulty of acquiring the length distinctions.

Research over the past 50 years has shown how a native speaker’s perception of a phonetic segment is affected by the rates of the segment and of the surrounding context (Denes, 1955; Port, 1977; Verbrugge and Shankweiler, 1977; Miller and Liberman, 1979; Port and Dalby, 1982; Johnson and Strange, 1982; Gottfried et al., 1990; Newman and Sawusch, 1996; Sawusch and Newman, 2000). For example, Pickett et al. (1999) found that native Italian speakers’ perception of Italian single and geminate stops was based on the stop closure duration relative to the preceding vowel. Listeners also use the speaking rate of a sentence to perceive native language (L1) contrasts of which the primary acoustic correlate is duration (e.g., Miller, 1987, and Wayland et al., 1994, for English /p-/b/ contrast; Magen and Blumstein, 1993, for Korean short and long vowels). These studies in L1 perception show rate normalization: native listeners do not identify a phonetic segment based on the absolute duration of the segment. Rather, their perception takes into account both the duration of neighboring segments and the segments that are contrastive in the language.

Consistent with the findings above, native Japanese (NJ) speakers use the rate of a sentence as a perceptual cue for distinguishing Japanese consonant and vowel length (Hirata, 1990a; Hirata and Lambacher, 2004). NJ listeners in Hirata (1990a) identified the pair of words /ita/ “stayed” versus /iita/ “went” based on the durational ratio of the stop closure to the preceding vowel /i/ when the words were presented in isolation. In a second experiment, an edited word [iti(t)ita] in which the durational ratio was identified as ambiguous by NJ listeners was embedded in a carrier sentence produced at different speaking rates. NJ listeners identified this ambiguous word as /ita/ or /iita/ based on the rate of the sentence. When the sentence was spoken slowly, the word was clearly identified as /ita/, and when the sentence was spoken faster,
the word was clearly identified as /itːa/. The perceptual cue present in the sentence rate superseded the ambiguous cue present in the isolated word.

Compared to rate normalization by native listeners, less is known as to how second language (L2) learners normalize speech rate and learn to make duration-based distinctions in sentences. In Hirata (1990b), when the Japanese words /ita/ and /itːa/ were presented in isolation, NE speakers distinguished the words, similarly to NJ speakers, using the durational ratio of the stop closure to the preceding vowel as a perceptual cue. However, when the words were embedded in a sentence spoken at different rates, NE speakers were unable to use the rate of the sentences as a perceptual cue as NJ listeners did. Thus, NE learners of Japanese have difficulty in normalizing rate over a sentence in Japanese, even though they can use the localized perceptual cue when isolated words are presented.

What method would most effectively enable NE speakers to normalize the rate of sentences and identify vowel length in an embedded word accurately? Given the results of Hirata (1990b), training L2 learners on isolated syllable or word contexts might not be the best method. The development of training methods for difficult L2 contrasts has been one of the most fruitful areas of L2 research (English /ʌ-/ː/: Strange and Dittmann, 1984; Logan et al., 1991; Bradlow et al., 1997; English /ɒ-/ː/: Jamieson and Morosan, 1986; Mandarin tones: Wang et al., 1999; Korean stops: Francis and Nusbaum, 2002; Japanese length contrasts: Yamada et al., 1995; Kawai and Hirose, 2000; Tajima et al., 2002). However, these studies provided L2 contrasts in isolated syllables or words. Investigation of effects of training using sentence stimuli has only recently begun (Hirata, 2004b; Hirata, 2004c; Kato et al., 2005). Studying the L2 vowel length distinction in a sentence context is particularly important, as research has shown that native listeners use the rate of sentences for making duration-based distinctions (Hirata, 1990a; Hirata and Lambacher, 2004). A recent study (Kato et al., 2004) showed that, after perceptual training with isolated words, NE speakers learned to perceive short and long vowels in sentences, indicating that their ability gained for the isolated word context generalized to the sentence context (see also Hirata, 2004b). However, Tajima et al.'s (2005) results showed that the accuracy of NE speakers' perception of Japanese length contrast was affected by the speaking rate of test stimuli, and that this effect did not decrease after training with isolated words.

Pisoni and Lively's (1995) high phonetic variability hypothesis provides a useful guide in the search for a method which would enable NE speakers to successfully identify Japanese vowel length. This hypothesis claims that the diversity of acoustic cues present in materials produced by multiple speakers helps rather than hinders non-native listeners in forming new perceptual categories. Traditionally, stimulus variability in voices and speaking rates was viewed as troublesome noise, extraneous to the encoding of linguistic units, and was deliberately excluded from experiments by the use of carefully controlled materials (Pisoni, 1997). However, evidence has accumulated that high stimulus variability in voices and phonetic contexts is informative, rather than distracting, and actually assists perceptual learning (Pisoni and Lively, 1995; Bradlow et al., 1997). In the present study, we tested this high variability hypothesis with regard to speaking rate variation. We hypothesized that providing variation in speaking rate helps non-native listeners to normalize rate and to learn to distinguish vowel length accurately.

The present study compared the relative effectiveness of three types of perceptual training: one with only a slow rate, one with only a fast rate, and one with both slow and fast rates. Training materials were sentences naturally produced by four NJ speakers at these rates. We chose this method rather than creating various speaking rates with a synthesizer because faster natural speech is not simply compressed slower speech, but involves changes in the size and velocity of various muscle contractions (Gay and Hirose, 1973; Gay et al., 1974; Gay, 1981). According to the hypothesis tested in the present research, the variations that exist in natural speech spoken at slow and fast rates are informative for non-native listeners. The prediction was that speech materials spoken at two rates (slow and fast), contain more diverse and rich acoustic cues than those spoken only at one rate (slowly or fast-only), and that this diversity aids non-native listeners in adjusting to different rates of speech and identifying short/long vowels accurately.

Effects of the three types of training were examined by a pretest and a post-test consisting of sentences produced by two additional NJ speakers at three speaking rates: slow, normal, and fast. These three rates were included in order to examine whether the ability gained through training with a given rate or rates would generalize to rate(s) to which participants had not been exposed. The present study compared each type of training with a control group that did not participate in training. Comparison with a control was important because our pilot results indicated effects of repetition and exposure, i.e., improvement of test scores simply by taking the tests twice without training. Thus, training is said to be “effective” if the amount of improvement from the pretest and the post-test was significantly greater for the trained than the control group. The study addressed the following questions:

1. How do NE speakers perform in identifying Japanese vowel length in carrier sentences spoken at three speaking rates (slow, normal, and fast)? Previous studies have shown that rate effects interacted with word type (Hirata, 2005a), vowel length (i.e., short or long; Hirata, 2005b), and context (i.e., in isolation or in sentences; Tajima et al., 2002 and Tajima et al., 2005). In these studies, identification accuracy in the sentence context is generally lower for faster rates. The present study will examine if this result is replicable.

2. Does perceptual training with sentences, as opposed to isolated words or syllables as in most previous studies, aid NE speakers in identifying Japanese vowel length accurately? This study will determine whether the present three types of training with sentences enable effective L2 contrast learning.

3. Does training with sentences spoken at two rates (slow
II. METHODS

A. Participants

Sixty-two monolingual native speakers of American English, whose ages ranged from 18 to 22, were randomly assigned to one of the following four groups: slow-only training group (n=16; mean age=19.2), fast-only training group (n=16; mean age=19.2), slow-fast training group (n=14; mean age=19.7), and a control group (n=16; mean age=19.3). None of the participants had previously studied Japanese or had extensive exposure to spoken Japanese. All but one participant had studied one or two foreign languages, usually Spanish and/or French, and one participant had studied five languages. None of the participants, however, had native-level fluency in any foreign language. No participant reported having any hearing problems. Participants were paid for their participation.

B. Training procedure

Participants assigned to one of the three training groups (slow-only, fast-only, and slow-fast) completed four training sessions in addition to a pretest and a post-test. Participants completed the experiment over a minimum period of 11 days and a maximum of 17 days. Participants had any two consecutive training sessions at least 24 hours apart, and no more than four days apart.

Training materials were a subset of materials used in Hirata (2004a). The target words were nonsense Japanese words in the form of /mV/mV/ and /mV/mV/ (V=ɪ, e, a, o, u; e.g., /mimii/ vs /mimii/), with the pitch accent always on the first vowel. These words were embedded in a single carrier sentence spoken by four native Japanese speakers (Male 1, Female 1, Male 2, and Female 2; Table I). For slow-rate and fast-rate training materials, the speakers were instructed to speak as slowly and as fast as possible, respectively. The four speakers’ mean speaking rate was 263 ms/mora for the slow rate and 105 ms/mora for the fast rate.1

Each type of training consisted of four sessions. Each session contained 160 stimuli by one speaker. As described in Table I, only slow-rate materials were used for slow-only training, and only fast-rate materials of the same speakers were used for fast-only training [5 vowels x 2 lengths x 2 repetitions x 8 blocks]. For slow-fast training, both of those slow- and fast-rate materials were used, but the total number of trials was the same as that in other training [2 rates x 5 vowels x 2 lengths x 2 repetitions x 4 blocks]. For slow-fast training, the rate of stimuli stayed constant within each block, either slow or fast. Slow-rate and fast-rate blocks were presented alternately (see Table I for the block structure).

Each session was broken into 8 blocks of 20 trials, and the first and the fifth blocks were preceded by three examples. When participants heard audio stimuli through headphones attached to a PC, the carrier sentence, soko wa___ to kaite arimasu (“___ is written there”) was displayed on the screen. Participants were asked to identify whether the second vowel of the disyllables inserted in the underlined part of the sentence was short or long. The participants received feedback on each response during training. The feedback consisted of a screen with the correct answer written out as either “short” or “long” and the target word spelled out in Romanized letters. Words with short vowels (containing two moras) had two dots over them, and words with a long vowel (containing three moras) had three dots. When participants answered correctly, the sign “Correct” appeared on the screen, and it enabled the participants to go on to the next block.

and fast) improve overall perceptual accuracy more than training with only one rate (slow-only or fast-only)? According to Pisoni and Lively’s (1995) hypothesis, the overall improvement should be higher for two-rate than one-rate training.

(4) How does the perceptual ability of NE speakers generalize from trained rate(s) to tested rates? For example, does slow-only training improve accuracy only for the slow rate, or improve it for other rates as well? Does slow-fast training improve accuracy only for these two rates, but not for the untrained (normal) rate?

II. METHODS

A. Participants

Sixty-two monolingual native speakers of American English, whose ages ranged from 18 to 22, were randomly assigned to one of the following four groups: slow-only training group (n=16; mean age=19.2), fast-only training group (n=16; mean age=19.2), slow-fast training group (n=14; mean age=19.7), and a control group (n=16; mean age=19.3). None of the participants had previously studied Japanese or had extensive exposure to spoken Japanese. All but one participant had studied one or two foreign languages, usually Spanish and/or French, and one participant had studied five languages. None of the participants, however, had native-level fluency in any foreign language. No participant reported having any hearing problems. Participants were paid for their participation.

B. Training procedure

Participants assigned to one of the three training groups (slow-only, fast-only, and slow-fast) completed four training sessions in addition to a pretest and a post-test. Participants completed the experiment over a minimum period of 11 days and a maximum of 17 days. Participants had any two consecutive training sessions at least 24 hours apart, and no more than four days apart.

Training materials were a subset of materials used in Hirata (2004a). The target words were nonsense Japanese words in the form of /mV/mV/ and /mV/mV/ (V=ɪ, e, a, o, u; e.g., /mimii/ vs /mimii/), with the pitch accent always on the first vowel. These words were embedded in a single carrier sentence spoken by four native Japanese speakers (Male 1, Female 1, Male 2, and Female 2; Table I). For slow-rate and fast-rate training materials, the speakers were instructed to speak as slowly and as fast as possible, respectively. The four speakers’ mean speaking rate was 263 ms/mora for the slow rate and 105 ms/mora for the fast rate.

Each type of training consisted of four sessions. Each session contained 160 stimuli by one speaker. As described in Table I, only slow-rate materials were used for slow-only training, and only fast-rate materials of the same speakers were used for fast-only training [5 vowels x 2 lengths x 2 repetitions x 8 blocks]. For slow-fast training, both of those slow- and fast-rate materials were used, but the total number of trials was the same as that in other training [2 rates x 5 vowels x 2 lengths x 2 repetitions x 4 blocks]. For slow-fast training, the rate of stimuli stayed constant within each block, either slow or fast. Slow-rate and fast-rate blocks were presented alternately (see Table I for the block structure).

Each session was broken into 8 blocks of 20 trials, and the first and the fifth blocks were preceded by three examples. When participants heard audio stimuli through headphones attached to a PC, the carrier sentence, soko wa___ to kaite arimasu (“___ is written there”) was displayed on the screen. Participants were asked to identify whether the second vowel of the disyllables inserted in the underlined part of the sentence was short or long. The participants received feedback on each response during training. The feedback consisted of a screen with the correct answer written out as either “short” or “long” and the target word spelled out in Romanized letters. Words with short vowels (containing two moras) had two dots over them, and words with a long vowel (containing three moras) had three dots. When participants answered correctly, the sign “Correct” appeared on the screen, and it enabled the participants to go on to the next block.

and fast) improve overall perceptual accuracy more than training with only one rate (slow-only or fast-only)? According to Pisoni and Lively’s (1995) hypothesis, the overall improvement should be higher for two-rate than one-rate training.

(4) How does the perceptual ability of NE speakers generalize from trained rate(s) to tested rates? For example, does slow-only training improve accuracy only for the slow rate, or improve it for other rates as well? Does slow-fast training improve accuracy only for these two rates, but not for the untrained (normal) rate?

II. METHODS

A. Participants

Sixty-two monolingual native speakers of American English, whose ages ranged from 18 to 22, were randomly assigned to one of the following four groups: slow-only training group (n=16; mean age=19.2), fast-only training group (n=16; mean age=19.2), slow-fast training group (n=14; mean age=19.7), and a control group (n=16; mean age=19.3). None of the participants had previously studied Japanese or had extensive exposure to spoken Japanese. All but one participant had studied one or two foreign languages, usually Spanish and/or French, and one participant had studied five languages. None of the participants, however, had native-level fluency in any foreign language. No participant reported having any hearing problems. Participants were paid for their participation.

B. Training procedure

Participants assigned to one of the three training groups (slow-only, fast-only, and slow-fast) completed four training sessions in addition to a pretest and a post-test. Participants completed the experiment over a minimum period of 11 days and a maximum of 17 days. Participants had any two consecutive training sessions at least 24 hours apart, and no more than four days apart.

Training materials were a subset of materials used in Hirata (2004a). The target words were nonsense Japanese words in the form of /mV/mV/ and /mV/mV/ (V=ɪ, e, a, o, u; e.g., /mimii/ vs /mimii/), with the pitch accent always on the first vowel. These words were embedded in a single carrier sentence spoken by four native Japanese speakers (Male 1, Female 1, Male 2, and Female 2; Table I). For slow-rate and fast-rate training materials, the speakers were instructed to speak as slowly and as fast as possible, respectively. The four speakers’ mean speaking rate was 263 ms/mora for the slow rate and 105 ms/mora for the fast rate.

Each type of training consisted of four sessions. Each session contained 160 stimuli by one speaker. As described in Table I, only slow-rate materials were used for slow-only training, and only fast-rate materials of the same speakers were used for fast-only training [5 vowels x 2 lengths x 2 repetitions x 8 blocks]. For slow-fast training, both of those slow- and fast-rate materials were used, but the total number of trials was the same as that in other training [2 rates x 5 vowels x 2 lengths x 2 repetitions x 4 blocks]. For slow-fast training, the rate of stimuli stayed constant within each block, either slow or fast. Slow-rate and fast-rate blocks were presented alternately (see Table I for the block structure).

Each session was broken into 8 blocks of 20 trials, and the first and the fifth blocks were preceded by three examples. When participants heard audio stimuli through headphones attached to a PC, the carrier sentence, soko wa___ to kaite arimasu (“___ is written there”) was displayed on the screen. Participants were asked to identify whether the second vowel of the disyllables inserted in the underlined part of the sentence was short or long. The participants received feedback on each response during training. The feedback consisted of a screen with the correct answer written out as either “short” or “long” and the target word spelled out in Romanized letters. Words with short vowels (containing two moras) had two dots over them, and words with a long vowel (containing three moras) had three dots. When participants answered correctly, the sign “Correct” appeared on the screen, and it enabled the participants to go on to the next block.
TABLE II. Test stimuli.

<table>
<thead>
<tr>
<th>Block and speaker</th>
<th>Carrier sentence</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) Male 3</td>
<td>sore ga ___ da to omoimasu</td>
<td>snf¹</td>
</tr>
<tr>
<td></td>
<td>('I think that is ___.')</td>
<td></td>
</tr>
<tr>
<td>(b) Female 3</td>
<td>asoko ni ___ to arimasu</td>
<td>snf</td>
</tr>
<tr>
<td></td>
<td>('It says ___ over there.')</td>
<td></td>
</tr>
<tr>
<td>(c) Male 3</td>
<td>koko wa ___ đa ra arimasen</td>
<td>snf</td>
</tr>
<tr>
<td></td>
<td>('This is not ___.')</td>
<td></td>
</tr>
<tr>
<td>(d) Female 3</td>
<td>hontō ni ___ wa kaitenai</td>
<td>snf</td>
</tr>
<tr>
<td></td>
<td>('It’s true that ___ is not written.')</td>
<td></td>
</tr>
<tr>
<td>(e) Male 3</td>
<td>soko de ___ to wa iwanakatta</td>
<td>snf</td>
</tr>
<tr>
<td></td>
<td>('I didn’t say ___ there.')</td>
<td></td>
</tr>
<tr>
<td>(f) Female 3</td>
<td>kito ___ de wa nai dejo:</td>
<td>snf</td>
</tr>
<tr>
<td></td>
<td>('It certainly won’t be ___.')</td>
<td></td>
</tr>
<tr>
<td>Post-test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) Male 3</td>
<td>kore ga ___ da to kikimafita</td>
<td>snf</td>
</tr>
<tr>
<td></td>
<td>('I heard that this is ___.')</td>
<td></td>
</tr>
<tr>
<td>(b) Female 3</td>
<td>koko ni ___ to arimusu ne</td>
<td>snf</td>
</tr>
<tr>
<td></td>
<td>('It says ___ here, right?')</td>
<td></td>
</tr>
<tr>
<td>(c) Male 3</td>
<td>are wa ___ đa ra nai desu jo</td>
<td>snf</td>
</tr>
<tr>
<td></td>
<td>('I tell you, that’s not ___.')</td>
<td></td>
</tr>
<tr>
<td>(d) Female 3</td>
<td>zettai ni ___ wa kakareteta</td>
<td>snf</td>
</tr>
<tr>
<td></td>
<td>('___ was definitely written.')</td>
<td></td>
</tr>
<tr>
<td>(e) Male 3</td>
<td>sojite ___ to ite kudasai</td>
<td>snf</td>
</tr>
<tr>
<td></td>
<td>('And then, please say ___.')</td>
<td></td>
</tr>
<tr>
<td>(f) Female 3</td>
<td>tabun ___ de wa arimasen</td>
<td>snf</td>
</tr>
<tr>
<td></td>
<td>('It’s probably not ___.')</td>
<td></td>
</tr>
</tbody>
</table>

¹ snf indicates that stimuli of slow, normal, and fast rates were presented randomly within the block.  
(a)–(f) indicate blocks.

trial. When participants answered incorrectly, the sign “Sorry…” appeared, and they were made to click “Play again” to hear the stimulus repeated three additional times. In each session, participants were asked to take a three-minute break after the fourth block.

C. Testing procedure

All of the four groups (slow-only, fast-only, slow-fast, and control) took a pretest and a post-test approximately two weeks apart. Each test consisted of 180 stimuli. Each stimulus used one of ten words drawn from five pairs of real Japanese words: /rubirubir/, /fiselfisel/, /rikakãrikak/, /katoíkatoí/, and /sajuí/sajui/ (meaning “agate”-“ruby,” “(name of a place)”-“opposite gender,” “science”-“liquor,” “transition”-“(surname),” and “hot water”-“left and right”), all accented on the first vowel. The stimuli were recorded by two NJ speakers who did not appear in the training sessions (Male 3 and Female 3; Table II). The NJ speakers recorded the words in three different carrier sentences, and each test had different carrier sentences (see Table II). Each sentence was spoken at three rates: slow, normal, and fast. Following Hirata (2004a) and Port (1977), speakers were given the following definition of the speaking rates: “tempo that is relaxed and comfortable” for the normal rate, “slowest tempo possible while keeping the sentence flowing together without obviously inserting breaks between words” for the slow rate, and “fastest tempo possible without making errors” for the fast rate.

Analyses on duration of sentences and target vowels confirmed (1) that speakers spoke with three distinct speaking rates, and (2) that they made a clear distinction between short and long vowels within each rate. The sentence speaking rates across speakers were 239 (SD = 24.8), 136 (SD = 15.2), and 105 (SD = 6.4) ms/mora for slow, normal, and fast rates, respectively, and the effect of rate was significant [F(2, 66) = 435.62, p < 0.001]. Comparisons of any two rates showed significant differences [p < 0.001]. The effect of speaker was not significant [F(1, 66) = 0.17, p = 0.686], although there was a significant interaction of rate and speaker [F(2, 66) = 4.29, p = 0.018]. The two speakers’ rates did not significantly differ for slow (244 vs 235 ms/mora) and fast rates (106 vs 103 ms/mora), but their normal rates were different (128 vs 145 ms/mora). The important point here, however, is that there were three distinct speaking rates across speakers. Analysis on target vowel duration also confirmed that the three speaking rates were distinct for both speakers and for both short and long vowels. Each speaker’s short vowels were significantly different among the three rates (Male 3: 152 vs 75 vs 64 ms; Female 3: 148 vs 83 vs 59 ms for slow, normal, and fast rates, respectively), and so were the long vowels (Male 3: 405 vs 165 vs 141 ms; Fe-
male 3: 341 vs 165 vs 109 ms) [p < 0.01]. These results were very similar to those of the materials in Hirata (2004a), a subset of which was used for the present training stimuli. An NJ speaker identified vowel length of these test stimuli with 98.9% accuracy.

The pretest and post-test each included 5 vowels × 2 lengths × 3 rates × 3 carrier sentences × 2 speakers for a total of 180 trials. Each test consisted of six blocks of 30 trials. Within a given block, every trial used the same carrier sentence, spoken by the same speaker (Table II). The stimuli were presented in a random order across rates. No words, sentences, or speakers from training were used in the tests. The pretest and the post-test used identical words but different carrier sentences. Each block was preceded by two examples using the same nonsense words used in training (e.g., /memel/ and /memell/). Immediately after each example, the answer, either “short” or “long,” was displayed on the screen. Participants were asked to take a three-minute break after the first three blocks.

The testing procedure was the same as the training procedure, except that participants did not receive any feedback on their responses.

D. Analysis

Raw percent correct test scores were transformed into rationalized arcsine units (RAU; Studebaker, 1985) to correct nonlinearity of the test score scale. A mixed-design Analysis of Variance (ANOVA) was conducted with these RAU scores. Group (slow-only, fast-only, slow-fast, control) was a between-subjects factor, and test (pretest, post-test) and rate (slow, normal, fast) were repeated-measures factors. If there is an overall effect of training, a significant group × test interaction should be found. Furthermore, if an effect of training is found only for specific rates, a significant group × test × rate should be found. Figures were made with the raw percent correct scale for ease of interpretation.

III. RESULTS

The results of ANOVA are summarized in Table III. With regard to the participants’ performance on the three speaking rates (question 1), a significant effect of rate was found. The mean test scores were significantly higher for the slow (75.6%) than normal rate (68.8%), and for the normal than fast rate (64.8%). It is important to note that the tests for the three rates included identical sets of words and sentences, and the only difference was that of the speaking rates. Thus, we can say that participants were less able to identify length of vowels in the same words spoken at faster rates than at slower rates.

A significant main effect was found for test (pretest: 65.8%; post-test: 73.6%), and a significant test × rate interaction was found in ANOVA (Table III), indicating that participants showed significant improvement from the pretest to the post-test, but that the amount of improvement differed across rates. The mean improvement, averaged over groups, was highest for the slow rate (from 69.9 to 81.1%) (11.2 percentage points improvement; below expressed as % improvement) [p < 0.001], less for the normal rate (7.8% improvement from 64.9 to 72.7%) [p < 0.001], and lowest for the fast rate (4.1% improvement from 62.7 to 66.8%) [p = 0.001].

A group × test interaction, which would indicate the overall effectiveness of training, was only marginally significant (Table III). The post-test scores were significantly higher than the pretest scores for all four groups (slow-fast training group: 65.5 vs 74.6% [t = 6.52, p < 0.001]; slow-only training group: 66.7 vs 75.3% [t = 6.86, p < 0.001]; fast-only training group: 65.5 vs 73.3% [t = 4.95, p < 0.001]; control group: 65.6 vs 69.9% [t = 2.43, p = 0.028]). Differences among the four groups in the overall amount of improvement between the pretest and post-test scores were small, as shown in Fig. 1.

![FIG. 1. Overall pretest (white boxes) and post-test (dark boxes) scores of the four groups. All test rates combined. The horizontal dashed line indicates the chance level performance.](image-url)
The overall improvement made by the control group was due to effects of repetition and exposure, since they took two tests. As mentioned in the introduction, training was said to be “effective” if the amount of improvement made by the individual trained groups was significantly greater than that made by the control group. Dunnett’s pairwise multiple comparison tests (which contrasted the three training groups against the control group) indicated that the amount of improvement made by the slow-fast training group (9.1% improvement) was significantly greater than that by the control group (4.3%) \(p=0.040\), indicating genuine effects of training. However, the slow-only training group’s improvement (8.6%) was only marginally greater than the control group’s improvement \(p=0.059\). The fast-only training group’s improvement (7.8%) was not significantly greater than the control group’s improvement \(p=0.116\). Thus, with regard to question (2), we found that training with sentences, particularly slow-fast training, was indeed effective in enabling participants to improve their perceptual ability to identify Japanese vowel length contrast. With regard to question (3), we found that, as far as the overall test scores were concerned, two-rate training was advantageous to one-rate training: slow-fast training was most effective, while only a marginal effect was found for slow-only training and no effect was found for fast-only training.

A significant group × test × rate interaction was found in ANOVA (Table III), indicating that the four groups’ differing amount of improvement from the pretest to the post-test depended on the speaking rate. Below, the four groups’ test scores were examined separately for each rate.

Robust effects of training were found for the slow rate, as shown in Fig. 2(a). All of the three training groups’ scores improved significantly from the pretest to the post-test: 12.3% improvement (means = from 70.4 to 82.7%) for the fast-only training group \(p<0.001\); 15.5% improvement (69.7 to 85.2%) for the slow-only training group \(p<0.001\); and 15.2% improvement (70.2 to 85.4%) for the slow-fast training group \(p<0.001\). In contrast, the control group did not significantly improve their slow rate scores (2.4% improvement from 69.4 to 71.8%) \(p=0.230\). Thus, significant effects of training were found for all three trained groups. It is interesting to note that the trained groups improved on the slow rate test scores regardless of the type of training they had received.

The results of the normal rate test scores [Fig. 2(b)] were similar to those of the slow rate scores. All of the three training groups improved their scores significantly from the pretest to the post-test. A 10.9% improvement (means = from 63.4 to 74.3%) was made by the fast-only training group \(p<0.001\), 7.4% improvement (65.5 to 72.9%) by the slow-only training group \(p=0.007\), and 8.0% improvement (64.4 to 72.4%) by the slow-fast training group \(p=0.001\). In contrast, the control group’s 4.9% improvement (66.2 to 71.1%) was not significant \(p=0.072\). Thus, effects of training on the normal rate test scores were found for all three trained groups. Compared to the slow rate test scores, however, the amount of improvement was less for the normal rate. It is noteworthy that the trained groups improved on the normal rate scores even though none of them had received normal rate materials in training.

The results of the fast rate test scores [Fig. 2(c)] differed from those of the other rates. The fast-only training group, rather unexpectedly, did not significantly improve on the fast rate test scores (1.6% improvement from 62.6 to 64.2%) \(p=0.524\), as was the case for the control group (5.6% improvement from 61.5 to 67.1%) \(p=0.061\). In contrast, the slow-only training group made a significant 4.0% improvement (64.7 to 68.7%) \(p=0.046\), and finally, the 5.5% improvement made by the slow-fast training group (61.9 to 67.4%) was statistically robust \(p=0.013\).

In summary, regarding question (4) from the introduction, we found that the perceptual ability of NE speakers generalized from trained rate(s) to tested rates. For example, slow-fast training showed the strongest effects not only on the participants’ overall scores, but also on all of the three rate test scores separately, including the normal rate which was not used in training. Slow-only training improved participants’ performance most prominently on the slow and normal rate test stimuli, and much less but still significantly
on the fast rate stimuli, even though their overall, cross-rate scores did not show a clear effect of training as a whole. Least effective was fast-only training where an overall effect of training was not found. Examining each rate separately, effects of fast-only training were found in the slow and normal rate test scores, but not in the fast rate scores.

**IV. DISCUSSION AND CONCLUSIONS**

This study examined the extent to which training with sentences of different rates helped native English speakers to learn to perceive Japanese vowel length contrast in the sentence context. One major finding, related to our first question, was that the speaking rate of sentences was a crucial factor in participants’ perceptual abilities. The participants’ test scores were lower for faster rates, and the amount of improvement shown by trained participants was also smaller for faster rates. These results are consistent with previous findings by Tajima *et al.* (2005) and Hirata (2005a and 2005b), and suggest that even if non-native speakers are able to perceive L2 contrasts at a slow rate, this does not guarantee the same ability at faster rates.

The second major finding, related to our second question, was that non-native speakers can learn to perceive L2 contrasts given only sentence input, supporting Hirata (2004b). The present slow-fast training, in particular, which provided only sentences but no words in isolation, was found to improve the identification of Japanese vowel length contrast in the sentence context. The overall improvement of the slow-fast training group was significantly greater than that of the control group, which did not participate in training. Slow-only and fast-only training also showed some limited effects (discussed further, below). This study provides additional evidence that adults are capable of learning difficult L2 contrasts with laboratory training, and not only with syllables or words in isolation (Jamieson and Morosan, 1986; Logan *et al.*, 1991; Bradlow *et al.*, 1997; Tajima *et al.*, 2002; Wang *et al.*, 1999), but also with sentences (Hirata, 2004b and 2004c).

It should be noted, however, that the overall effect of training in the present study was fairly small, as indicated by the improvement of 9.1 percentage points for slow-fast training. This is less improvement than that shown in some previous studies of L2 contrast training that used a two-alternative forced choice identification method. For example, Bradlow *et al.*’s (1997) subjects showed a significant improvement of 16 percentage points in distinguishing English /l/ and /l/. (The mean pretest score in Bradlow *et al.* (1997) was 65%, comparable to the pretest score of 65.8% in the present study.) On the lower end, Logan *et al.*’s (1991) subjects showed a significant improvement of 7.8 percentage points (though their mean pretest score was higher: 78.1%). The present study’s low level of improvement might partially be due to the fact that the test stimuli were embedded in sentences. This makes sense given that trained participants consistently showed less improvement in the sentence than word context in Hirata (2004b). However, the crucial role of sentence training cannot be overlooked. Hirata (2004b) found that participants who were trained with sentences made robust perceptual improvement in both the isolated-word and sentence contexts, but those who were trained only with isolated words made less improvement in the sentence context. Thus, even though training with sentences is more time consuming, there is a practical advantage to using sentences in training, if our ultimate goal includes enabling non-native speakers to perceive difficult L2 contrasts in fluent speech. Following Hirata (2004b), it would be interesting to examine whether the present sentence training would improve participants’ ability to perceive isolated words as well as words in sentences, as in the present experiment.

The third major finding of the present study relates to the differential effects of the three types of training, as addressed in our third question. The participants who received fast-only training improved least on their overall, cross-rate test scores. Their overall improvement (7.8 percentage points) was not significantly greater than that of the control group (4.3 percentage points). Slow-fast training was slightly more beneficial than slow-only training. The overall, cross-rate improvement made by the slow-fast training group (9.1 percentage points) was significantly greater than that made by the control group, but the slow-only training group’s improvement (8.6 percentage points) was only marginally greater than the control group’s. In summary, only the slow-fast training group improved significantly on the total test scores, as well as on the scores for each of the three rates. This result, that training with two rates was more effective than training with one rate (slow-only or fast-only), supports Pisoni and Lively’s (1995) high phonetic variability hypothesis. Their original hypothesis concerned the effectiveness of training with variability in talker and phonetic context. The present study provides support for the effectiveness of training with variability in speaking rate: the more variability in the rates of training stimuli, the more effective for non-native speakers’ perceptual learning. That slow-only training was slightly advantageous over fast-only training might also be consistent with the high variability hypothesis: variability of vowel as well as sentence duration in the present training was higher for the slow than fast rate (see Hirata, 2004a). This claim, however, should be examined in the future by comparing effects of high-variability fast rate training versus low-variability slow rate training.

The above conclusions in support of Pisoni and Lively’s (1995) high variability hypothesis merit further discussion. First, the difference among the three types of training was quite small (Fig. 1). Second, regardless of the type of training, participants showed a similar amount of improvement for the slow and normal rates (Fig. 2). Note that no participant heard normal-rate sentences in training, but that all trained participants improved their test scores for the normal rate. These results, relevant to our fourth question, imply that even beginning L2 learners are able to normalize different speaking rates fairly easily within the range of slow to normal speaking rates. The ability to normalize speaking rate in native languages has been examined extensively, as reviewed in the introduction, and this ability seems to be utilized for L2 speech within a limited range of speaking rates even at an early stage of L2 learning. This result is in line with Kato *et al.*’s (2004) finding that, with perceptual training, NE speak-
...ers were able to learn to adapt their categorical boundary positions of Japanese vowel length contrast according to different contexts.

Also worthy of discussion are the results of fast-only training, which perhaps are counterintuitive. Previous studies suggested that abilities acquired during L2 training are largely to deal with the kinds of stimuli that had been presented in training (e.g., synthetic versus natural speech, word versus sentence stimuli) (Strange and Dittmann, 1984; Morosan and Jamieson, 1989; Hirata, 2004c). This would predict that fast-only training should improve scores for the fast rate, but not the other rates. However, the opposite results were found: the fast-only training participants improved on slow and normal stimuli, but not on fast stimuli. In addition, this training was found to be least effective. Jamieson and Morosan’s (1986) perceptual fading technique, which successfully enabled Canadian francophones to identify the English /ð/-/ without compromise, is useful in understanding these results. Their method initially reduces stimulus uncertainty by providing the longest consonant frication in the /ðə/-/ə/ syllable continuum, and as training progresses, increases stimulus uncertainty by including syllables with progressively shorter frication. This method helps learners to initially focus their attention on critically relevant cues, and then to gradually learn to handle within-category acoustic variation. It is possible, then, that participants in the present fast-only training suffered from not having first been provided with critically relevant cues with little uncertainty. The durational difference between short and long vowels is about 50 ms for the fast rate, compared to about 150 ms for the slow rate (Hirata, 2004a) and, thus, fast-only training participants might have been unable to grasp the most critical cue, which is vowel duration, solely from the fast stimuli.2

Another possible interpretation of the ineffectiveness of fast-only training as compared to slow-only training might concern short-term memory effects. It is possible that attentional resources available in short-term memory are limited when learning to perceive unfamiliar non-native contrasts, and that there is a higher demand in processing fast than slow rate sentences. If so, then because of this higher demand, the fast rate speech signals might have degraded more quickly than the slow rate speech signals in short-term memory, and this might be why participants with fast-only training were unable to learn the critical acoustic cue of vowel duration as effectively. We expected from Pisoni and Lively’s (1995) high phonetic variability hypothesis that single rate training, whether fast-only or slow-only, would have the same effect on overall perceptual learning. However, the differential results of fast-only and slow-only training cannot be explained so simply by this hypothesis in regards to the issue of speaking rate variability. The present results may suggest that different processing and memory demands for slow versus fast rate stimuli are at play, and the amount and quality of information transmitted from short-term memory to long-term memory do matter. This factor may augment Pisoni and Lively’s (1995) high phonetic variability hypothesis in understanding how temporal characteristics of stimuli play a role in the ultimate formation of an L2 perceptual category.

On a final note, in order to achieve more robust perceptual learning, we might wish to train subjects on sentences spoken at a greater number of speaking rates. Introducing several more speakers would also create additional, possibly useful, speaking rate variation. A related issue is the distribution of stimuli among slow, normal, and fast rates. The speaking rates of the present test materials were 239, 136, and 105 ms/mora for slow, normal, and fast rates, respectively, showing a greater difference between the slow and normal rates than the difference between the normal and fast rates. It is possible that differential effects of training might appear more strongly if fast rate materials are even faster or if three speaking rates are distributed more equally. A future study should also examine whether non-native listeners benefit from randomizing rates at every trial instead of by block, as the former has been found to provide more difficulty than the latter (Tajima et al., 2005). Finally, the carrier sentence in the present training was held constant for the purpose of examining the precise effect of speaking rate, but a future study should also examine whether providing different carrier sentences would benefit learners. Although additional research is thus necessary, the present study reveals that different speaking rate is an important factor in L2 learners’ acquisition of duration-based distinctions.

ACKNOWLEDGMENTS

This study was supported by National Science Foundation Grant No. BCS0418246, given to Y.H. We thank Jon Bernard for his assistance in creating the training program and editing the manuscript, and Jacob Whiton, Connor Forbes, Lucia Livak, Jess Worby, Carol Glenn, Stephen Lambacher, and Spencer Kelly for their help in various stages of this project. We also thank reviewers for their insightful comments on the earlier version of this manuscript.

1See Hirata (2004a) for more details of duration measurements of these training materials. A subset of the materials was also used for NJ speakers’ perception in Hirata and Lambacher (2004). In that study, NJ speakers identified short and long vowels of these materials with 98.1% accuracy.

2Here, we used differences in absolute vowel duration as an example of speaking rate differences, but what the precise measure of variability should ideally be is an empirical question. One could argue that relational measures, such as ratios of short and long vowels, might be more appropriate than the measure of absolute duration. The present experiment showed generalizability of learning from one rate to another, which might suggest that relational measures are more suitable as a measure of variability. However, other evidence suggests that absolute duration plays an important role for non-native listeners (Hirata, 2005a; Tajima, 2006), as well as for native listeners when stimuli are impoverished (Hirata and Lambacher, 2004).


