This is to certify that the
dissertation entitled

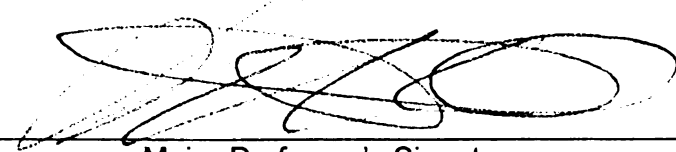
ACQUISITON OF GEMINATE CONSONANTS
IN JAPANESE BY AMERICAN ENGLISH SPEAKERS

presented by

MIKI MOTOHASHI

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Linguistics and Germanic, Asian and African Languages



Major Professor's Signature

12/20/06

Date

DATE DUE	DATE DUE	DATE DUE

ACQUISITION OF GEMINATE CONSONANTS IN JAPANESE BY AMERICAN ENGLISH SPEAKERS

By

Miki Motohashi

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Linguistics and Germanic, Asian and African Languages

2007

ABSTRACT

ACQUISITION OF GEMINATE CONSONANTS IN JAPANESE
BY AMERICAN ENGLISH SPEAKERS

By

Miki Motohashi

It has been pointed out that English-speaking learners of Japanese often exhibit timing problems in the perception and production of geminate consonants since durational contrast is a novel phoneme for English speakers. The present study has reported on data from the perception and production of geminate consonants in Japanese by American learners. Based on these data, an effective way to train learners to identify geminate consonants was developed and tested.

Four experiments were conducted. Two experiments collected perception and production data of geminate consonants by American learners of Japanese to investigate the way that the learners perceived and produced geminate consonants and examine whether there were any particular phonetic contexts and identities of geminate consonants which were particularly more difficult for learners. The conditions considered were types of consonants; /s/, /t/ and /k/; preconsonantal segments; /sa/ and /a/, postconsonantal segments; /u/ and /a/, and comparison between words in isolation and carrier sentences. The results showed that the learners' performances were affected by the phonetic contexts and identities of geminate consonants. Specially, a combination of fricative geminate consonant /s/ and low sonority vowel /u/ was found the most difficult to perceive, while there was no such tendency for production.

The other two experiments considered a method of training to improve perception

of such difficult geminate consonants. In addition, another issue to consider is the modality of training. The training program was developed to investigate whether audio-visual (AV) training would be more beneficial than auditory-only (A-only) training to improve the learners' perception of geminate consonants. The previous training studies generally used auditory-modality cues; however, Hardison's studies (e.g., 2003, 2004) reported that L2 learners benefit from visual cues as well as auditory cues in perception training. The present study used visual displays of waveforms of geminate consonants as aids for learners to identify difference of mora weight between singleton and geminate consonants. The result indicated that AV-training showed its superiority in producing perception improvement over the A-only training. Further, the AV-training group data showed a transfer effect of perception training to their improvement of production. This result suggests that there is a close link between perception and production development processes.

Further, the present study emphasizes the importance of collecting data from learners' performances and aims to develop an effective training program. The stimuli used for the training were selected based on the findings from the data collected. The effectiveness of high-variability stimulus demonstrated in the present study is compatible with previous studies (e.g., Pisoni et al., 1999), which supported a multiple-trace memory theory in which each event or input is encoded in memory as a trace, rather than prototype. Through the training, all attended perceptual details were stored in memory and modify an attention weighting scheme to perceive distinctive features in L2. It is assumed that the bimodal training used in the present study would facilitate this process.

ACKNOWLEDGMENTS

My deepest appreciation goes to my dissertation committee members. Since I left the United States three years ago to pursue my teaching career in Japan, communication became difficult. I am grateful that the committee members always—and very patiently—answered my questions promptly by e-mail.

I would like to thank my co-chairs, Dr. Dennis Preston and Dr. Susan Gass. Dr. Preston always encouraged me to go on. His academic advice (as well as his sense of humor) was always helpful. Dr. Gass always gave me helpful feedback and patiently read my dissertation.

The influence of other committee members is also appreciated. I thank Dr. Barbara Abbott for being generous with her time whenever I had questions. Dr. Mutsuko Endo Hudson has been supportive by advising me to complete this dissertation, as well as giving me precious opportunities to expand my teaching career. I thank Dr. Shawn Roewen for his willingness to read my work, ask important questions, and provide feedback and helpful suggestions.

Last but not least, I am grateful to Dr. Debra Hardison for her patience—and a tremendous amount of feedback. The topic of this dissertation originated from a project for her class. She helped me with the entire process of revising the dissertation by giving me very specific comments on many points—not only about SLA, but also about phonology, statistics, thesis format, etc. in a very timely manner.

I want to thank all of my friends with whom I studied at the University of Wisconsin-Madison, and Michigan State University. I am especially grateful to Emiko

Magnani for her patience in reading my dissertation and giving me very helpful feedback.

I also extend many thanks to my students and colleagues at Kansai Gaidai University in Japan for their support for collecting data. I would especially like to thank Dr. Jeffrey Rasch for patiently proofreading my dissertation and giving me many suggestions and comments as a linguist, which contributed to my preparation for the defense.

I thank Hideki Saigo, as my colleague, best friend and husband, for sharing the difficult time during the last years of completing the dissertation, giving me advice about finishing a dissertation based on his own experience, and doing all the household chores (especially cooking). I appreciate his patience and support. Finally, I want to thank my parents and grandmother for always helping and encouraging me to continue my study in the U.S. This work is dedicated to them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The present study will address issues in the acquisition of second language (L2) phonology by English speakers, focusing on geminate consonants in Japanese. Specifically, the main experiment that provides data for the present study was conducted to investigate whether adult learners of Japanese could be trained to perceive and produce geminate consonants accurately. In order to determine effective training materials and methods, the present study also examined what kinds of geminate consonants were most difficult for learners to perceive and produce. The rationale for choosing geminate consonants is the notorious difficulty which many learners of Japanese as a second language have with durational contrasts. Japanese is a mora-timed language, and duration is contrastive, while such contrasts do not exist in English. It has been reported that learners whose first language (L1) is English often have problems with the perception and production of geminate consonants, as well as with other timing morae in Japanese (Toda, 2003). These problems often lead to miscommunication.

In the present study, participants were given perception training in order to examine the possibility of improvement due to such training in their L2 performance. A number of studies have reported that intensive laboratory training resulted in improvement in learners' performance (e.g., Jamison & Morosan, 1986, 1989; Logan, Lively & Pisoni, 1991; Pisoni, Aslin, Perey & Hennesy, 1982; Yamada, Akahane-Yamada & Strange, 1995, Hardison, 2003, 2004). Further, it has also been reported that the effects of training in perception were transferred to ability in production (e.g., Hardison, 2003;

Bradlow, Pisoni, Akahane-Yamada & Tohkura, 1997; Catford & Pisoni, 1970; Rochet, 1995). Through investigation of the effect of perception training in modifying foreign accented production, we will also examine issues of the relationship between perception and production.

The present study also emphasizes the importance of collecting data from learners' performances and aims to develop an effective training program to train English-speaking learners' perception as well as production of geminate consonants in Japanese. Many of the previous studies of geminate consonants in Japanese limited their focus to stop consonants, and few studies have referred to the types of phonetic contexts in which geminates occur. Therefore, it is necessary to examine whether there are any particular phonetic contexts that make perception and production of a geminate consonant more difficult for learners. Based on the outcome of such research, we could find ways to focus training more specifically and to develop effective materials for training.

Although previous training studies have been shown to be effective, the methodologies have not been evaluated thoroughly enough. For example, traditionally the dominant method for examining learners' development of the ability to perceive new, difficult nonnative contrasts was laboratory auditory training, by using a two-alternative identification or discrimination tasks involving minimal pairs in isolation (e.g., Akahane-Yamada, Tohkura, Bradlow & Pisoni 1996; Ingram & Park, 1998; Ziolkoski, Usami, Landahl & Tunnok, 1992), but few studies provide details of why the particular methods themselves were adopted. The present study aims to suggest a more effective training method by actually collecting and examining in detail the perception and

production data of learners, determining in greater detail where errors occur.

Another issue to consider is the modality of training. The above mentioned previous training studies generally used auditory-modality cues; however, Hardison's studies (e.g., 2003, 2004) reported that L2 learners benefit from visual cues as well as auditory cues in perception training, and further showed that bimodal training was especially effective on the more phonologically difficult segments from the point of view of the learners' L1. The psychological evidence supports the claim that information from one modality helps to reinforce another's sensory pathway, and the combination of information from different modalities enhances the development of the learning process (de Sa & Ballard, 1997). The present study examined the effect of combining visual cues with auditory cues to train the learners to identify geminate consonants, and the results were compared with auditory-only training.

Hardison (2005a) used visual displays of pitch contours of French as visual cues to train English speakers and reported their effectiveness as visual input. The present study used visual displays of waveforms of geminate consonants as aids for learners to identify difference of mora weight between singleton and geminate consonants. A growing number of language teaching programs have been utilizing computer-assisted instruction for perception and pronunciation teaching to enhance self-monitoring skills by learners. Recent developments in technology allow researchers to display formant frequency graphs, waveforms, or spectrograms on computers to teach both suprasegmental (stress, rhythm and intonation) and segmental features (e.g., Anderson-Hsieh, 1994, 1996; Chun, 1989, 1998; de Bot, 1983; Hardison, 2004, 2005; Levis & Pickering, 2004; Molholt, 1988; Weltens & de Bot, 1984). Such visual displays

have been used mainly as production training to give learners feedback on their own production. As a potential alternative training method, the present study proposes computer-assisted perception training. Visual information which consists of graphs of the waveforms of geminate consonants is expected to facilitate learners' sensitivity to mora timing in Japanese and improve their perception and, subsequently, production as well. As shown in several studies mentioned above, the effect of perception training can be carried over to production ability without additional explicit production training. The present study aims, therefore, to examine whether this transfer effect can be observed in the acquisition of geminate consonants. The training method of the present study was developed on the analysis of actual data which were able to pinpoint the difficulties that learners of Japanese a second language have.

There are also other advantages of computer-based instruction. For example, it appears that computer-delivered materials are helpful in reducing the nervousness that students may feel in the classroom, and easy access may encourage learners to use a computer program on a daily basis. The present study suggests that pronunciation training should be incorporated into everyday classroom teaching, in addition to intensive laboratory training, which has also been shown to be highly effective in previous studies.

In sum, the present study was motivated by the following research questions:

1) How do the learners perceive and produce geminate consonants? Is there any particular phonetic context of geminate consonants, which makes perception and/or production more difficult for learners?

2) Are audio-visual instruction and training using visual displays of waveforms of geminate consonants more beneficial than auditory-only information?

3) Does perceptual training improve production ability without any additional explicit production training?

To examine the first research question, Experiments I and III described below were conducted to collect data on the perception and production, respectively, of geminate consonants by American English speakers. As reviewed in the following section, previous studies have shown that learners perceive and produce geminate consonants in different ways from native speakers of Japanese. However, the results of these studies vary according to many factors including the data collection methods and the focus of the analysis. The present study focused on the phonetic form and context of geminate consonants, that is, the types of consonants and the preceding and following segments; few previous studies have examined these factors. Research questions 2 and 3 are tested by Experiments II and IV. Experiment II was conducted to test the effect of electronic visual input on the perception of geminate consonants, and Experiment IV was a pretest and posttest of the experimental training study using visual input in perception training, to examine whether the training is effective and whether the effects of such training transfer to improvement in production.

The organization of the remainder of this dissertation is as follows. Chapter 2 reviews the relevant literature to explore the above research questions, mainly regarding 1) perception and production of geminate consonants by native and nonnative speakers of Japanese; 2) the relation between perception and production and the effects of training to improve ability in these two domains; and 3) the effects of electronic visual feedback on acquisition of L2 phonology. Chapters 3 through 6 discuss the methodology and results for Experiments I through IV, respectively. Chapter 7 provides a general discussion of the

results of the experiments and the pedagogical implications.

CHAPTER 2

BACKGROUND

2.1 Previous studies of geminate consonants

2.1.1 Mora in Japanese

The duration of vowels and consonants is a contrastive feature in Japanese while it is not in English. A mora is a unit of timing, and each mora is supposed to take about the same length of time to pronounce (Ladefoged, 1993). As the number of morae increases, the total duration of the syllable increases proportionately. Syllable duration, and thus total word length is attributable to the number of morae. There are three kinds of special morae in Japanese; geminate consonants, moraic nasals, and those resulting from long vowels, and they are called '*tokushu-haku*' (special timing morae). To perceive and produce these special morae, sensitivity to timing is necessary, and it is a difficult task for learners of Japanese whose native language does not have durational contrasts to acquire native-like perception and production. Below are examples of minimal pairs of both *tokushu-haku* and non-*tokushu-haku*:

Table 2.1

Japanese special morae "tokushu-haku"

Long vowel	Geminate consonants	Moraic nasal
/kiite/ "Listen." (3 morae; 2 syllables)	/kitte/ "stamps" (3 morae; 2 syllables)	/kiNka/ "gold coins" (3 morae; 2 syllables)
/kite/ "Come." (2 morae; 2 syllables)	/kite/ (2 morae; 2 syllables)	/kika/ "vaporization" (2 morae; 2 syllables)

Although basic Japanese syllables are open, consisting of /CV/, syllables with geminate consonants and moraic nasals are exceptionally closed ones. According to Shibatani (1990), geminate consonants consist of a non-nasal consonant coda followed by a homorganic consonant onset in the following syllable. This homorganic geminate consonant adds one mora. For example, a word containing a geminate consonant like *kitte* "stamps" is considered a three-mora word, while its single consonant counterpart is counted as a two-mora word, e.g., *kite* "come," and again this difference in duration is contrastive as shown in Table 2.1.

By actually measuring the duration of words produced by native speakers, research has shown that each mora has equal duration (e.g., Port, Dalby & O'Dell, 1987; Sugito, 1999). For example, Port et al. (1987) measured the duration of a number of words which contain different numbers of morae, including words with geminate stops and long vowels, and found that the duration of words with an increasing number of morae increased by nearly consistent increments. It would appear that native Japanese speakers discriminate between short and long vowels, as well as single and geminate consonants, by the relative duration of the target vowel or consonant.

With morae of relatively equal length, Japanese therefore has isochronous timing of morae, while English has syllables and/or morae of different lengths, most notably as a result of stress; i.e., stressed syllables are longer. Duration of units, however, is not systematically contrastive in English. It has been pointed out that English-speaking learners of Japanese often show timing problems in the production and perception of long vowels and geminate consonants, since durational contrast is novel for them. Thus, not only learners, but also instructors need a clear understanding of how native Japanese

speakers make durational contrasts.

2.1.2 Research on native speakers' production of geminate consonants

A number of studies have conducted acoustic analyses to see how native speakers of Japanese actually produce a geminate as opposed to a single consonant. It has been found that one of the most important acoustic cues for producing the distinct duration of a geminate stop consonant is the closure duration of the first part of the geminate. The results of these previous acoustic studies are mostly consistent; the total duration of a syllable with a geminate consonant is approximately 50% longer than that of its single consonant counterpart, though there is a discrepancy in actual measurements as to whether the single/geminate duration ratio is exactly 2:3 or not.

Homma (1981) measured word duration of two and three mora words with geminate stops (/pp/, /tt/, and /kk/) and their singleton consonant counterparts produced by native Japanese speakers. She found that the ratio duration between words with single stops and those with geminated stops was about 2:3, confirming that the morae of geminates are isochronous timing units. Such duration was not affected by the phonological context of types of preceding and following consonants and vowels.

Sugito (1999) measured the duration of one- to five-mora words with geminate stops and found that words with an increasing number of morae increase in duration by nearly constant increments. She also conducted the same experiment with English native speakers, having them read English words, and pointed out that English syllables are, unlike morae in Japanese, inconsistent in their duration, as shown in Figure 2.1.

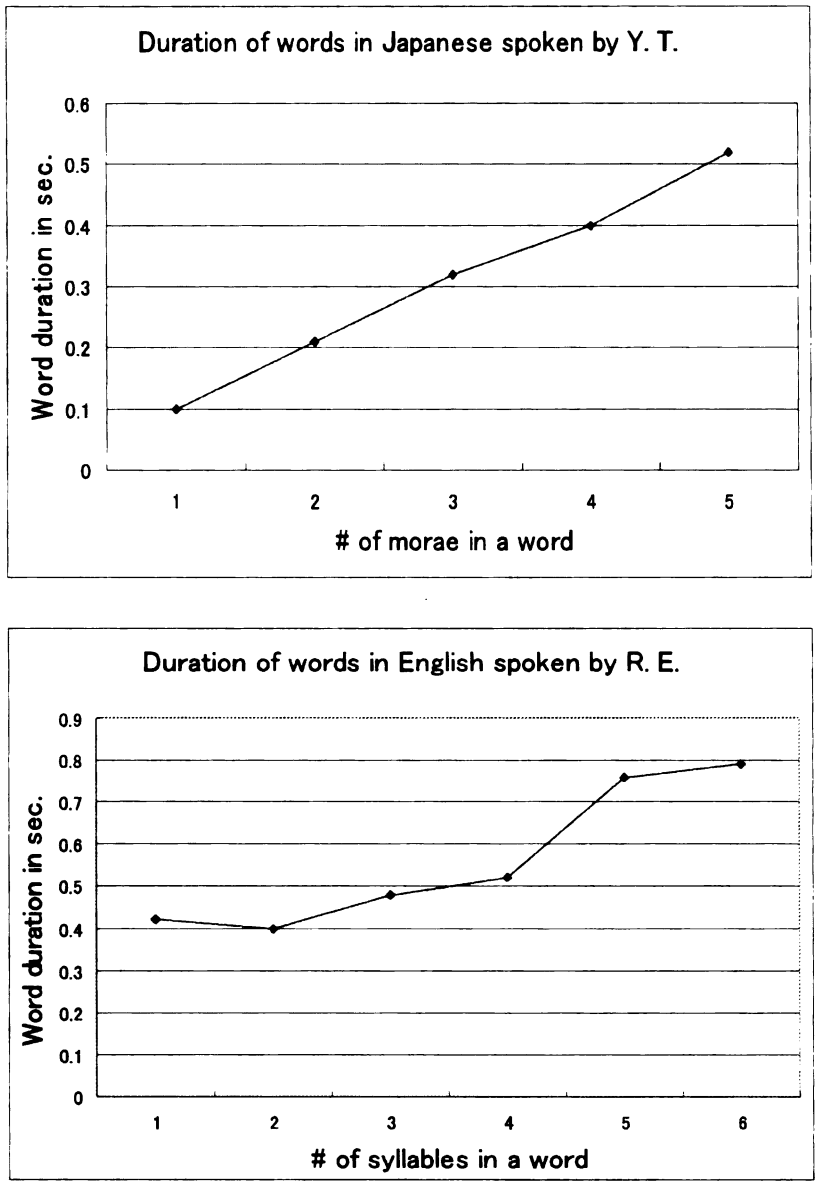


Figure 2.1. Duration of words in Japanese and English (Sugito (1999); reproduced from p.69)

2.1.3 Research on native speakers' perception of geminate consonants

Researchers have also been intrigued by the question of what acoustic cues native Japanese speakers use to discriminate durational contrasts. It is generally agreed

that native Japanese speakers use closure duration of the first part of a geminate consonant as a primary cue in discriminating between two- and three-mora words.

For example, Min (1987, 1993) used digitally edited stimuli made from two-mora words and their three-mora geminated counterparts. Using a stimulus such as /ita/, the stop closure duration between /i/ and /ta/ was gradually lengthened in 10 ms steps from 110 ms to 250 ms (15 different lengths altogether). The participants were then asked to tell whether they heard /ita/ or /itta/. The results showed an apparent perceptual categorical boundary at 160-180 ms among native speakers. Fujisaki and Sugito (1977) also used synthesized stimuli manipulating the closure duration of the first part of a stop geminate consonant and had native Japanese speakers discriminate between their perceptions of geminate and single consonants. They reported that the closure duration played the most important part in discrimination, and the perceptive boundaries for the native Japanese speakers to distinguish a single from a geminate consonant were categorical. Many other studies indicate similar findings (e.g., Fukui, 1978; Hirato & Watanabe; 1987; Toda, 1998).

Besides closure duration, there have been some studies which show that pitch accent also influences native speakers' perception of durational contrasts. Ofuka (2003) also used synthesized stimuli to examine the relationship between pitch accent location and the perception of durational contrasts. Pitch accent is contrastive in Japanese, e.g., minimally contrastive /ame/ (H(igh)-(L(ow)) and /ame/ (LH) are different words which mean "rain" and "candy," respectively. She examined the perceptual boundary between a singleton consonant word /kata/ "shoulder" (HL) and its geminated counterpart /katta/ "won" (HLL) and another pair /kata/ "form" (LH) and /katta/ "bought" (LHH), by

manipulating the closure duration of stimuli by 10 ms in ten steps between /kata/ and /katta/ in both pitch accent patterns. She found that native Japanese speakers were affected by the location of pitch change so that there is a significant difference between the two pitch accent patterns in the placement of the perceptual boundary; the LHH pattern in /katta/ "bought" required a longer closure duration than the HLL "won" pattern to be perceived as containing a geminate.

Furthermore, Hirata (1990a) distinguished word-level from sentence-level perception. In her experiment, native speakers seemed to use different acoustic cues at different levels. In word-level perception, preceding vowel length as well as stop closure duration were utilized by native speakers as acoustic cues to discriminate single and geminate consonants. She concluded that the ratio of the closure duration of the consonant to the duration of the preceding vowel was a crucial acoustic cue. If this ratio is short, the consonant is perceived as single, but if it is long it is perceived as a geminate. In the sentence-level perception, the distinction is made based on the speed of the units following geminate consonants. A single consonant can be perceived as a geminate consonant if the following parts of an utterance are read fast, and geminates can be heard as single consonants if the following parts of an utterance are read slowly.

In sum, for native speakers, the acoustic cues for durational discrimination are not limited to a single factor; the cues may include the duration of the preceding vowel, the closure duration of the stop, and the speed of the following elements, depending on whether perception occurs at word level or sentence level. Despite such interacting conditions, however, native speakers have clear and consistent perception of durational contrasts.

2.1.4 Research on nonnative speakers' production of geminate consonants

It has been reported by many researchers that the timing control of geminate and single stop closures differs significantly between native speakers (NS) and nonnative speakers (NNS), which contributes to the characterization of an accent as "foreign." Han (1992) reported that her American subjects' closure duration of stop geminate consonants was consistently shorter than that of the NS subjects. On the other hand, Toda (1994) claimed that it was not only the shorter duration of geminate consonants and long vowels, but also the longer duration of single consonants and short vowels, which made learners sound like they were producing geminate consonants and long vowels. As a result, her Australian subjects tried to produce geminate consonants which were even longer than their already lengthened single counterparts, and this adjustment resulted in a noticeable foreign accent.

2.1.5 Research on nonnative speakers' perception of geminate consonants

As discussed so far, previous studies have generally agreed that the absolute closure duration of a stop consonant is one primary cue for native speakers of Japanese for discriminating between durations of consonants. Researchers are also interested in seeing whether learners of Japanese use the same acoustic cue.

Most of the previous studies of the perception of geminate consonants by nonnative speakers have been carried out to determine the categorical boundaries of perception of contrasts using synthetic stimuli. Inaccurate perceptual boundaries of closure duration for the single vs. geminate discrimination will cause faulty perception by nonnative speakers. The results agreed that nonnative speakers would perceive the stimuli

as geminate consonants when they had shorter closure duration than was required by native speakers, but in general, such a perceptual boundary was not categorical, but rather blurred and continuous (e.g., Hirata, 1990b; Min, 1987; Nishibata, 1993). In Min (1987), Korean speakers' results were compared with those of native participants. The results showed an apparently categorical boundary among native speakers, while nonnatives did not have such a clear boundary. Min further found that, while native speakers used closure duration as an acoustic cue for perception, some Korean learners, though too small a number to generalize from, tended to depend on additional phonetic characteristics such as tenseness and aspiration of the consonant as acoustic cues.

Korean and Chinese speakers in Minagawa and Kiritani's (1996) study, and Thai speakers in Minagawa's (1996) study, were found to be affected by pitch accent types when discriminating single and geminate consonants. In a High-Low (HL) accent context, the error pattern of C→CC (mishearing a single as a geminate consonant) was significantly higher than that of CC→C (mishearing a geminate as a single consonant), but in the Low-High (LH) accent context there was no difference in error types. Since the acoustic measurement of the stimuli revealed durational differences of postconsonantal vowel duration between the HL and LH contexts, that is, the average postconsonantal vowel duration in an HL accent context is shorter than in an LH accent context, closure duration to postconsonantal duration ratio was suggested as a possible acoustic cue for Korean and Chinese speakers in judging the single vs. geminate contrast. In contrast, according to Hirato and Watanabe (1987), the perception of the single vs. geminate stop contrast by native Japanese speakers is not affected by postconsonantal vowel length. In addition, Toda (1998, 2003) reported that, while NS were affected by preconsonantal

vowel duration, NNS did not show such an influence.

Another interesting finding is in Yamagata and Preston's (1999) study. Their study showed that the learners (L1 was English) often perceived long vowels in Japanese loanwords, which native Japanese speakers spelled with geminates instead. Although the learners failed to geminate, they were successful in giving the target words the correct number of morae.

Enomoto (1989) and Toda (2003) further reported a learning effect through formal instruction, in which advanced level learners came to acquire a clearer perceptual boundary for geminate consonants, compared to the beginning learners.

In summary, there may be different acoustic cues which learners of Japanese might depend on. Although the findings on perceptual cues used by native speakers are consistent among researchers, there has been little agreement and no clear generalization on what acoustic cues nonnative speakers use to distinguish geminate/single consonants. This is because each research program is different, for example, in the subject's L1, the phonetic contexts of the stimuli, the data collection method, or the levels of proficiency of the subjects.

The amount of research which focuses on the correlation between learners' perceptual ability and the types of phonetic contexts of the stimuli is specially limited. Most studies have examined stop consonants (e.g., Minagawa, 1996; Nishibata, 1993), except Toda (1998) and Hayes (2002) which also included fricatives (/ss/).

Hayes (2002) conducted an experiment to examine the relative perceptibility of the contrasts based on durational differences among particular types of single/geminate contrasts. To analyze the duration of single/geminate consonants acoustically, a fricative

/s/ and two stops /t/ and /k/ were chosen for test items. She hypothesized that differences in durational contrasts between single and geminate consonants belonging to different natural phonological classes would affect the learners' perception of the singleton-geminate contrast.

The subjects of the study were to listen to minimal pairs, 60 of the same word and 60 with different words, and then to tell whether the pair that they had just heard was the "same" or "different." As seen in Table 2.3, since the difference in duration between a single /t/ and a geminate /tt/ is larger than the difference between both the /s/ and /ss/ pair and the /k/ and /kk/ pair, she hypothesized that the discrimination of /t/ and /tt/ should be the easiest. On the other hand, there should be no difference in difficulty between /s/ contrasts and /k/ contrasts, since they have little durational difference.

Table 2.2

Stop closure and frication duration of single/geminate consonant pairs (Hayes, 2002, p.32)

	t/tt	k/kk	s/ss
Single duration	95.7	81.7	136.1
Geminate duration	276.1	223.6	270.1
Difference (geminate duration minus single duration)	180.3	141.9	134.0 (in msec)

Her hypothesis was supported by the results of the experiment. However, learners do not usually encounter such situations, in which they can compare single and geminate counterparts for discrimination. Therefore, it is hard to say that this result

reflects the reality of learners' perception in other context.

Further, the geminate consonants used as stimuli in previous studies differed in phonetic contexts, i.e., they were of various consonant types and had various preceding and following elements. One of the objectives of this present study is to examine more comprehensively how learners perceive and produce geminate consonants and whether their perception as well as production is affected by such phonetic conditions. In addition, we have seen that nonnative speakers' perception and production of geminate consonants are different from native speakers'; however, there have been few studies of how these two abilities are related, except Akahane-Yamada (1999), whose study was limited to stops. It is important to explore this issue further in order to clarify the details of the fundamental problems which learners might have in acquiring geminate consonants. The next section will review general views on the development process of perception and production by adult L2 learners.

2.2 Development of speech perception and production

The view that perceptual development comes before production development is consistent with the results of a number of experiments which have been concerned with the relationship between L2 perception and production in the course of L2 acquisition. Many such studies suggest that perception plays an important role in production, and production problems result not only from motoric difficulties but also from perception problems.

Flege, Munro, and MacKay (1995) suggested that production inaccuracy of the Italian learners in their study might have been due to a perceptual problem; they argued

that an L2 phone must be perceived in a fully native-like fashion if it to be produced in a fully native-like fashion. Thus, they argue, perception should come before production; although correct perception does not guarantee correction production, it is a prerequisite for it. Rochet (1995) also observed the role of perception in foreign-accented pronunciations of L2 sounds. His study examined perception and production of the French high front rounded vowel [y] by untrained Portuguese and English speakers, whose native languages contain only two high vowels (/i/ and /u/). In the perception test to identify vowels along a synthetic high vowel continuum, native French speakers identified a stimulus with the F2 values between 1300 and 1900 Hz as /y/, but Portuguese speakers identified it as /i/ and English speakers as /u/. Based on this result, Rochet hypothesized that an imitation task would indicate a similar tendency; when /y/ was produced incorrectly, Portuguese speakers tended to produce it more /i/-like, whereas English learners produced more /u/-like vowels. The results supported his hypothesis; therefore, he claims that foreign-accented pronunciation by untrained speakers may be perceptually motivated. This perception precedence idea can be also observed in several other studies (e.g., Aslin, Pisoni, Hennesy & Perey, 1981; Barry, 1989; Bohn & Flege, 1990).

However, there have been reports which showed opposite tendencies. Sheldon and Strange (1982), replicating Goto (1971), collected data from Japanese learners of English regarding the English liquids /r/ and /l/, which are not contrasted in Japanese. The data showed that the subjects performed better and more accurately on the production of /r/-/l/ contrasts than in perception. The data included perception test materials involving minimal pairs with /r/ and /l/ and the subjects' judgments regarding their own productions

of the pairs. According to Sheldon and Strange, "perceptual mastery of a foreign contrast does not necessarily precede adult learners' ability to produce acceptable tokens of the contrasting phonemes" and "may lag behind production mastery" (p. 254). Flege and Efting's (1987) experiment with Dutch speakers of English showed that their subjects were able to produce a substantial voice onset time (VOT) difference between the /t/ phonemes in Dutch and English, but they did not show such good discrimination in perception. Further, Mack (1989) also conducted studies which showed that production can be more accurate than perception. Gass (1984) examined the perception of L2 learners of English of the VOT of /b/ and /p/ in initial positions by using a forced-choice task with synthesized stimuli, and the learners' production data were also collected. Perception data showed an unclear, continuous distinction between the segments, compared with the native speakers' clear categorical boundaries. As opposed to this nonnative-like perception, the learners could produce /b/ and /p/ in native-like fashion. Thus, in this study, nonnative speaker production was in advance of perception.

However, as Flege (1991) and Mack (1989) pointed out, these results have to be interpreted carefully. For example, the data from the Japanese learners of English in Sheldon and Strange's (1982) study may have been influenced by the formal English training in production which Japanese school students had received, i.e., instruction to use articulatory strategies such as "to say /l/, combine the features of the Japanese X and Y sounds" (Flege, 1991, p. 265). Thus, the types of input which the subjects have been given should also be considered cautiously to determine precisely how the data collected could point to a specific process of L2 development.

One of the findings in Sheldon (1985), which reanalyzed the Korean learners'

data in the U.S. reported by Borden, Gerber and Milsark (1983), was that the precedence of production by perception decreased as the Korean learners' time residing in the U.S. increased. This could be interpreted as an effect of instruction as Sheldon and Strange (1982) argued above. Sheldon hypothesized that a functional perceptual level in an L2 learner might be enough for communication purposes, while heavily accented productions are socially less accepted, with the consequence that L2 speakers would feel more pressure to improve production than perception. Bohn and Flege (1990) also agreed that speech production was more subject to social control than perception, and as a result, the perception of a new contrast showed more resistance to L2 experience than the production of the contrast did.

Although no conclusive determination has been made, we can assume that speech perception and production capacities of individuals have great overlap. It is important to consider both of the areas simultaneously. To explain the relation between L1 and L2 in perception and production, and predict difficulties that learners tend to have, Flege (1995) and Best (1995) proposed the following L2 developmental models.

Best (1995) proposed the "Perceptual Assimilation Model (PAM)," which hypothesizes that L2 speakers perceive nonnative sounds based on similarities to or discrepancies with the L1 phones which are closest to them in terms of the manner of articulation. The model predicts an L2 discrimination ability that depends on the degree to which an L2 contrast can be associated with L1 categories. Thus, for L2 learners, certain contrasts in L2 are easier to discriminate than others, while some are more difficult. For example, /r/ and /l/ present the most difficult contrast for Japanese learners of English to master since these two phones are identified as the same Japanese phoneme,

a situation which is called 'single category contrast.'

Similarly, Flege (1995) developed the 'Speech Learning Model (SLM),' which hypothesizes that L2 sounds that are perceptually similar to sounds in L1 are more difficult to acquire accurately than sounds that are dissimilar to any sounds in L1, and L2 speakers try to assimilate a new L2 phone to a close L1 phone although the two phones are acoustically different. This indicates that such L2 learners have not detected the phonetic differences between an L2 sound and the most similar L1 sound, which results in foreign accents. The greater the phonetic distance between an L1 phone and the closest L2 phone is, the more easily the L2 learner can detect the difference. Greater phonetic distance facilitates the eventual establishment of a phonetic category.

Both models assume that perceptual learning occurs first but the perception and production skills develop in parallel, although this prediction that perception development is followed by production is not always true. However, as we have seen, production depends on perception in certain ways, although its development may not always follow perception development. We can therefore assume that production difficulties may be associated with perception difficulties. As the SLM and PAM suggest, since learners are language-specific perceivers of speech sounds and tend to adjust their perception to the phonetic characteristics of speech segments found in their L1s, nonnative-like perception often occurs during the course of L2 acquisition. The previous linguistic experience with L1 might influence the way L2 sounds are perceived, at least in the early stages of L2 perceptual category development.

Jusczyk (1993, 2000) proposed a model of the development process of infant speech perception, which can be applicable to the L2 acquisition process, too. Infants

have an innate auditory analyzer which can process any potential L1 at the initial stage of processing of speech signals. A set of auditory analyzers provide a preliminary description of the spectral and temporal features present in the acoustic signal. Once language is acquired, the output of the auditory analyzers is weighted to give prominence to those features that are the most critical to distinctive phonological features. This "weighting scheme" is a way of directing attention to features critical for recognizing and distinguishing words in a particular native language. For example, information from auditory analyzers concerning aspiration in syllable-initial voiceless and voiced stops would receive heavy weighting in the acquisition of English, but not of French. Therefore, to acquire a new language, a listener must learn a new weighting scheme in order to be attuned to the target language. Many studies of first language acquisition reported that children's linguistic ability to learn to discriminate between new contrastive features decreases after a certain age. This is not a loss in auditory capability, but rather a reorganization of the perceptual space optimal for L1 (Guion & Pederson, 2002). Since L2 learners tend to fall back on the weighting scheme used for the native language, a new weighting scheme must be developed. They must learn to alter the focus of attention, which affects the way in which speech sounds are perceived (Jusczyk 1993).

Based on the idea that perceptual space is modified by experience (Nosofsky, 1986), a number of training studies have been conducted in order to examine how adult learners can alter such focus of attention; they are reviewed in the following section.

2.3 Training studies

Earlier researchers have postulated that the poor performance observed in adult

learners' perception and production was due to a permanent change in the perceptual or sensory mechanisms as a result of selective early experience (Pisoni, Aslin, Perey & Hennesy, 1982). On the other hand, training experiments have been conducted based on the assumption that it is possible to train adult learners to perceive and/or produce novel L2 phonemes. This implies that adult learners' perceptual and/or productive systems can be modified. Such training studies generally aim to 1) find the cause of difficulty in acquiring new L2 phones; 2) discover how capabilities of the adult perceptual system are modified; 3) show that linguistic experience has a substantial effect on speech perception; 4) find an effective way for L2 learners to acquire difficult sounds; and/or 5) examine further the relationship between perception and production.

In the early studies which showed the effectiveness of training, researchers were interested in the perception of voicing contrasts in stops. Pisoni et al. (1982) trained monolingual English speakers to identify and discriminate VOT contrasts that are not phonemically distinctive in their native language. For the experiment, synthetic VOT stimuli based on measurements of natural speech were used to train the subjects to identify -70, 0, and 70 ms VOT synthetic stimuli (voiced, voiceless unaspirated and voiceless aspirated stops, respectively). Whereas English has only a two-way contrast of voiced and voiceless, and the features aspirated and unaspirated are not contrastive, the results showed that adult learners could perceive an additional perceptual contrast easily in the laboratory after a short training period (1 hour a day for 4 days). Thus the adult subjects were successful at modifying their perception of VOT. Pisoni et al. also argued that the key to this successful training was to provide immediate feedback during training tasks. Further, McClaskey, Pisoni and Carrell (1983) also showed that knowledge about

VOT perception gained from laboratory training was genuinely acquired in that the result of discrimination training on one place of articulation (e.g., labial) was transferred to another place of articulation (e.g., alveolar) without any additional training.

Another speech contrast that has been investigated in a great detail by a number of studies is the /r/-/l/ contrast in English. The contrasts are harder for learners to acquire than VOT distinction. In order to distinguish between /r/ and /l/ in various phonetic environments, processing of complex temporal and spectral changes is required, although the stop voicing involves only a temporal difference. Voicing may be more discriminable to listeners than the acoustic cues that underlie other speech contrasts, since it is psychophysically more distinctive or robust (Pisoni, Lively & Logan, 1994).

A series of studies was conducted by Pisoni and his colleagues to address the problems experienced by L1 Japanese learners of English as a second or foreign language. Japanese does not have the /r/-/l/ contrast (Bradlow, Akahane-Yamada, Pisoni & Tohkura, 1999; Bradlow, Pisoni, Akahane-Yamada & Tohkura, 1997; Lively, Logan & Pisoni, 1993; Logan, Lively & Pisoni, 1991; Pisoni, Lively & Logan, 1994). The training procedure and stimuli used in their experiments were designed to avoid some of the problems found in Strange and Dittmann's (1984) study. In Strange and Dittmann, although discrimination performance improved gradually over the training sessions, the effects of discrimination training did not generalize to naturally produced stimuli. One of the causes of their failure to train learners' linguistic ability was that the variability of the stimuli was too limited to generalize, since the training stimuli consisted of only one /r/-/l/ minimal pair produced by one synthetic voice. Based on this observation, Pisoni and colleagues used a wider variety of training stimuli, which consisted of natural speech

tokens instead of synthesized speech, and minimal pairs in different phonetic environments produced by five different talkers. In doing so, they considered the important role of stimulus variability in perceptual learning. In addition, a two-alternative forced-choice identification task was used instead of a discrimination task. An identification task encourages classification of stimuli into categories, while a discrimination task focuses perception only on fine within-category acoustic differences. This high-variability training approach for perceptual learning contributed to generalization to novel stimuli and talker's voice. Lively, et al. (1993) showed that increasing the stimulus variability during learning was effective in the development of robust phonetic categories. The training was also effective in promoting long-term retention of learning in both perception and production; the Japanese subjects in Japan maintained their improved levels of performance three months after the perception training (Bradlow et al., 1999).

Another training technique was described in Jamieson and Morosan (1986). Their study examined the ability to identify the American English fricatives /θ/ and /ð/. Training was given to Canadian francophone speakers by using a perceptual fading technique, in which stimuli were presented sequentially from the most acoustically distinct stimuli to the least distinct stimuli. In a more recent study, McCandliss, Fiez, Protoppapas, Conway, and McClelland (2002) used a similar technique called adaptive training to train Japanese learners to acquire the English /r/-/l/ contrast through synthetic stimuli, which maximally exaggerated the acoustical difference between the contrasts, and gradually minimized the difference to approximate that found in natural exemplars. They also investigated the effect of feedback, comparing the presence and absence of

feedback in combination with the different types of training. It was found that combination of adaptive training and feedback facilitated learning the most by calling a subject's attention to the critical cues that distinguish the training stimuli. In addition, the result of the training experiment indicated that the exaggeration effect would increase the likelihood that the subject would be able to generate consistent labels of contrasts even in the absence of feedback. However, they also suggested that, as similarly implied by Jamieson and Morosan's result, what the subjects have learned is a very general phonological discrimination, and it could not apply to all instances of /r/ and /l/ spoken in all possible contexts by all speakers. They assumed that the fixed training with a large number of various stimuli in combination with feedback such as was used by Pisoni and colleagues in the study mentioned above would lead to more robust generalization and contribute to mitigating the difficulty learners have in acquiring the target contrast from natural experience, although their adaptive technique would provide more rapid learning.

While many training studies predominantly used auditory presentation methods, Hardison (2003) also used and extended the high-variability training approach to include training in combined auditory and visual modalities. Her study was the first published study that investigated auditory-visual vs. auditory-only training for L2 learners. Using a talker's face, including articulatory gestures, as a visual cue, and locating the sound in various phonetic contexts and positions within the word, Hardison examined the effect of speaker and context variability on the perception of the English /r/-/l/ contrast by Japanese and Korean learners of English. The result demonstrated significant interactions of these variables and indicated both generalization to novel stimuli and production improvement. The effectiveness of multimodal training in addition to high variability of

stimuli was also observed in the earlier identification of words beginning with /r/, /l/, /p/ and /f/ by Hardison (2005b). As another effective way of providing audio-visual training, a real-time computerized pitch display was used to provide prosody training for English-speaking learners of French (Hardison, 2004) and Chinese-speaking learners of English (Hardison, 2005a). The learners could visualize their own pitch contours in utterances in the target language and compare it with native speakers'. This is another example of effective training utilizing visual input along with auditory input to facilitate learning.

According to Hardison (2000, 2003), these results, that a multi-modal, high-variability perceptual training approach facilitates learning and generalization, indicate that language learners store detailed individual instances as memory traces, rather than creating abstract prototype categories, and use these stored detailed episodes for memory encoding.

Traditional view of the learning process, known as the abstractionist view, assumes that any representations of the sound patterns of words are stored as abstract prototypes and are normalized with respect to variables affecting the sounds, such as the talker's voice, speaking rate, and so on. It is assumed that these variables, which were not necessary for processing the meanings of any given utterances, were discarded as noise somewhere during speech processing.

The alternative view, the episodic view, does not assume such normalization or prototype formation, but assumes that listeners store specific instances or tokens in memory. During processing, they evoke specific instances, rather than abstract representations of the sound patterns of words, and try to match new instances to these.

This view is supported by empirical studies of adult learners (e.g., Goldinger, 1997; Johnson, 1997) and the investigations regarding adults' and infants' retention of specific details of particular instances of perceptual experience, e.g., recognition of particular voices (Jusczyk 2000). Multiple-trace theory incorporates the above-mentioned prototype and episodic views of perceptual processing, such as in the Minerva 2 model by Hintzman (1986), and explains how repetition affects episodic memory. The model assumes that each experience event has it own memory trace as an episodic trace and stores specific events in primary or short-term memory (PM) as collections of primitive properties that include perceptual details, context, affect, semantic connotation, and so on. When retrieving a memory trace, a retrieval cue or "probe," which is an active representation of experience, is simultaneously sent to communicate with all stored dormant traces in secondary or long-term memory (SM). When the probe is sent from PM to all traces in SM, PM receive a single reply or "echo." Repetition of the same experience produces multiple traces of an item but does not cause strengthening of a single memory trace. Each trace reacts more or less intensely depending on its similarity to the probe, and the contribution of traces which are the most similar to the probe is greater because they produce a more intense response. If the information in the representation is more detailed, the probe becomes more specific, which produces a smaller set of highly activated traces. Thus the responses or echoes to the probes vary in their intensity and content. Whenever several traces are very strongly activated, the intensity of content of the echo is very strong and reflects their high level of common properties; therefore, if a new instance is very similar to previously stored traces, the intensity of the echo reflects more common properties. A strong echo reflects greater

degree of similarity in activated traces and familiarity to the experience. However, if the probe resembles only a few of the previously stored traces, the returned echo should reflect more idiosyncratic properties of those activated traces. Thus, the specificity of the probe and, the number of strongly activated traces will determine whether the echo content is ambiguous or clear.

Jucszyk's (1993) development model reviewed earlier is also based on the episodic view. During the course of perceptual development, the output of the innate auditory analyzers at an earlier stage of development is weighted to give prominence to the critical distinctive features in the target language to enable the learner to recognize words. Through this attention weighting scheme, sound pattern extraction is made, and then the matching process occurs. The representation obtained through linguistic experience and by the weighting scheme serves as a probe that will try to be matched against existing representations, or traces previously analyzed and stored in SM. If a close match is obtained between the probe and the stored items, the input is recognized; if not, the input is stored as a new item. It is also assumed that representations of the sound structure of a word are not stored in the form of abstract descriptions such as abstract prototypes; rather, the sound properties of items actually encountered in different contexts, in other words, multiple traces of individual instances of the item are stored.

The above episodic views on the learning process are consistent with the results from the previous L2 training studies whose results indicate that repetition of high-variability stimuli and immediate feedback are indispensable factors in effective training. Hardison (2000) proposed a scenario of bimodal L2 speech processing and the role of training, based on multiple-trace theory, Jusczyk's model of child L1 development,

and results from her auditory-visual resulted in Hardison (1998). The following is her proposal: at the first stage of L2 acquisition, auditory and visual inputs are preliminarily analyzed through different pathways. In the next stage, a new weighting scheme for L2 must be developed, so that learners can alter their attentions from the optimal setting to perceive distinctive features in L1 to the optimal setting for L2 by learning to attend to new sources of information obtained through the auditory analyzers. For example, to hear the distinction between /r/ and /l/, attention has to be shifted auditorily to the F3 transition and visually to the articulatory gestures in order to distinguish between the sounds. Learning occurs through copying the features of an experience into a trace. Probes, or signals processed in PM, activate dormant stored traces in SM, and the weighting process occurs according to the trace's similarity to the features of the probe, which ultimately will return an echo to PM. Attention to auditory and visual attributes of the stimulus will determine the features of the probe. Training with multiple exemplars and immediate feedback enhance learning; repetition and feedback can direct attention to within-category similarities and between-category distinctions in L2, adding traces to memory and modifying the memory system. Old traces are not altered, but new traces are added. As the result of learning, new L2 memory traces become less ambiguous and less confusable. Thus, the objective of training is "to create a situation in which the echo from an aggregate of L2 traces acting in concert overshadows the echo from L1 traces" (Hardison, 2000, p. 321). The advantage of prototype is its long retention, while exemplars may be forgotten over time, but decaying multiple-traces of each exemplar with redundancy can also be reduced. Through many new exemplars in perceptual learning, learners store multiple traces which mach the probe; these multiple traces share

common features, thus functioning like a prototype.

Another important feature of Hardison's (2000) model of L2 development, a weighting scheme is required to direct learners' attention to critical distinctive features in L2. Multiple-trace theory is based on the assumption that all items that are attended to are stored in memory; learners must be able to attend to critical features of input for categorization and identification of new L2 contrasts. Multiple exemplars, immediate feedback, and repetition add traces and increase the salience and information value of important features to focus on, and consequently enhance learning.

Not all tokens in the target language are equal candidates for incorporating into the phonetic category, and only those tokens that are perceived during a "signal-oriented" mode can be collected for incorporation and subsequent modification of a phonetic category (Lindblom, Guion, Hura, Moon, & Willerman, 1995). Signal orientation, which is the cognitive mechanism of attention, helps to create novel categories in addition to the modification of existing categories. Nosofsky (1986) argues that in categorization and identification of newly encountered stimuli, selective attention process is assumed to operate, which leads to systematic changes in the structure of the perceptual space and changes inter-stimulus similarity relations. Attention weights act to shrink or expand the perceptual space; the psychological space is stretched along the dimension that is selectively attended to, maximizing within-category similarity, and is shrunk along the other dimensions, minimizing between-category similarity, so that learners are optimizing similarity relations for the given categorization problem. If selective attention properly modifies similarity relations across the identification and categorization paradigms of stimuli, the probe to memory will provide good matches to stored L2 traces, returning

less ambiguous echoes to PM, and categorization will be enhanced. Therefore, it is necessary to direct learners' attention to focus on the critical properties.

Based on Nosofsky's proposal, Pisoni, Lively & Logan (1994) examines adult phonetic processing and concludes that the cognitive structures created by attentive processes are adjusted from prior linguistic experience and can be modified through training for better discrimination of non-native phonetic contrasts. As we have reviewed, training programs with high variability and multimodality have shown their effects in the shifting of learners' focuses, which leads to generalization to new tokens they encounter in the real world. Empirical studies have reported that different sensory areas affect other classification learning in the individual modalities. Bimodal speech recognition reported by McGurk and MacDonald (1976) showed that a pair of auditory and visual stimuli (the visual stimulus being a speaker's lip movement) can affect each other and produce a sensory effect different from either the actual auditory input or the visual input. de Sa and Ballard (1997) argued that responses of cortical cells in the primary sensory modality would respond to features from other sensory modalities. They then proposed a computational model using the information in one modality to modulate learning in another, instead of merging the outputs from different pathways. In perceptual learning in SLA, not only auditory input but also visual input in AV-training, such as described in Hardison (2003), facilitates such processes. Based on these observations and the exemplar-based theory of learning, the current study also aims to give beginning learners of Japanese effective training to accurately perceive geminate consonants through multimodal (not only auditory, but also visual) training, with a variety of stimuli and immediate feedback, expecting better improvement than that resulting from auditory-only

training, as well as generalization to novel stimuli.

Some studies have shown that there is a close link between perception and production through demonstrating transfer of training, in which the effect of training on one domain was transferred to another. In Bradlow, Pisoni, Akanae-Yamada, and Tohkura (1997), 11 Japanese learners of English received 45 sessions (30 minutes each) of perceptual identification with feedback over 15 days. The stimuli consisted of minimal pairs for /r/ and /l/. Although the training was designed only for perception, the pretest and the posttest included assessment of production ability. The result showed that the subjects improved not only in perception but also in production. In Rochet (1995), native speakers of Mandarin Chinese received perception training for French voiceless stops, and the result also showed that improvement in perception performance could carry over to improvement in production. In addition, a similar transfer effect was found in the studies on phonologically delayed children conducted by Jamieson and Rvachew (1992). Their studies also showed that speech production treatment for the children benefited from perception training.

A very early training experiment in production showed a similar transfer effect; i.e., the effect of production training carried over into perception performance. Catford and Pisoni (1970) compared the performance of subjects who received production and articulation training involving unfamiliar or "exotic" sounds and that of those who were trained only in perceptual discrimination. The results of production and perception tests showed that those who received articulation training in addition to perceptual training performed better. This finding implied, as they suggested, "some kind of carry-over from productive competence to auditory discriminatory competence" (p. 481); thus,

improved production abilities may contribute to better discrimination of L2 sounds.

Leather (1990) conducted two experiments in parallel with two different groups of Dutch learners of Mandarin Chinese; perception tests were given to the subjects who had been trained only in the production of Chinese tones, and production tests were given to those who had received only perception training for the same tones. The progress that the two groups made was compared, and the results showed no difference in their progress. Both groups improved at the same rate. He argued that his subjects "did not need to be trained in production to be able to produce, or in perception to be able to perceive, the sound patterns of the target system" and "training in one modality tended to be sufficient to enable a learner to perform in the other" (p. 95).

The bimodal (audio and visual) training of Hardison (2003) also showed improvement in subjects' production ability. Hardison suggests that L2 learners may "attempt to coordinate information about perception and production in category development" (p. 516), a claim similar to that made by Jusczyk's model (1993) of L1 development. Interactions between the developing perception and production systems may affect the way learners acquire knowledge of L1 sound patterns. Learners are under pressure to coordinate the way that these systems function and to relate the perceptual representations of words to the articulatory representations for production, so they may reach an abstract representation to capture generalizations that apply to both systems, which is phonology. It is the coordination of perceptual and productive representations that may lead the language learner from a more global representation of sound patterns of words to one that is structured with respect to phonetic segments. According to Jusczyk, when infants start babbling, they are very attentive to the distinctive features of the

language. Production development lags behind perception since infants have to wait until they gain control and coordination over their jaw movements; it also takes time to coordinate information from both modalities. Adults learners do not have to wait for the development of their articulatory system, but it is observable that they also need some time for the coordination of both modalities. At the same time, it should also be noted that production is more easily altered through formal instruction, as has already been mentioned.

On the other hand, differences in the rate of development in perception and production were found in Bradlow et al. (1997), who reported little correlation between degrees of learning in perception and production after perception training. The learners who improved the most in perception did not necessarily improve the most in production. There was variation in learning; degrees of learning in perception are different from the transferred learning in production. They noted that "learning in the perceptual domain is not a necessary or sufficient condition for learning in the production domain; the processes of learning in the two domains appear to be distinct within individual subjects" (p. 2397). This claim is compatible with the results of Akahane-Yamada (1999). As Bradlow et al. (1997) indicated, their study did not support Flege's (1995) SLM. The SLM assumed that improvement in speech production as a consequence of perceptual learning is due to a reorganization of the underlying system used for both speech perception and production and hence, predicts that changes in perception will transfer to changes in production, and these changes will proceed in parallel. However, the SLM does not account for the results of Bradlow et al. (1997), which indicated the presence of individual variations in learning and the lack of correlation between degrees of learning

in the two domains.

The specific relationship between production and perception is not clear; they might differ according to sound types, phonetic contexts, methods of data collection and training, and so on. However, most of these studies agree on the following; 1) perception ability and production ability are closely related: though the degree of correlation is not clear, the abilities do not appear to develop independently; 2) training experiments bring apparent improvement to adult learners, either in perception or production, or both. Therefore, it is possible to train adult learners to perceive and/or produce novel phonemes in the L2, though training methods and data collection processes in the above-cited studies varied.

The above-reviewed training studies demonstrate the adaptability of the adult perceptual system through training, and there is a certain relationship between perception and production. There have been a number of studies involving various L1s and L2s, as well as various kinds of segments (vowels and consonants) and suprasegmentals (e.g., Chinese tones); however, very few studies have been conducted in this context to examine geminate consonants in Japanese. The present study took as one of its principal objectives the investigation of the relationship between the acquisition of the perception and the acquisition of the production of geminate consonants, in particular, the contribution of perceptual training to productive ability.

2.4 Making visual information available to L2 learners

2.4.1 Electronic Visual Feedback (EVF)

The previous section reviewed some previous laboratory training studies. In this

section, alternative ways of improving learners' perception of geminate consonants will be considered. A growing number of language programs have been utilizing recent developments in the technology available as computer-assisted instruction for perception and pronunciation training to enhance self-monitoring skills by learners. For example, electronic visual feedback (EVF) is a type of computerized training which utilizes software (e.g., Cool Edit by Syntrillium Software, Wavesurfer, and Praat) or hardware (e.g., Computerized Speech Lab (CSL) and Visi-Pitch by Kay Pentax and IBM Speech Viewer) to perform an acoustic analysis of a target sound. Chun (2002) used Speech Tools, downloadable web-based software provided by SIL, for the images of intonation in her book. Speech Analyzer, a component of Speech Tools which offers visual analyses such as waveform, pitch plot, spectrogram, spectrum and various F1 vs. F2 displays. All these programs and devices involve the digitization of speech and its subsequent visual representation on a video screen. Such technology allows learners to measure and visualize intensity, duration, frequency range, etc. of the target sounds. Researchers have reported the effectiveness of such training in improving learners' perception and production.

2.4.2 Effects of instruction on production of segments and suprasegmental features

Molholt (1988, 1990) reported effective use of EVF when teaching difficult consonants and vowels to Chinese ESL students in laboratory sessions using Kay Pentax' Visi-Pitch and Speech Spectrographic Display (SSD). With Visi-Pitch, students can see simultaneously both an instructor's and their own spectrograph and waveform of a target sentence to practice. In general, the energy concentration of Chinese consonants has a

higher frequency than that of American consonants. The differences in the duration of American /v/ and /b/ are new to Chinese speakers. Molholt (1988) introduced EVF as an effective way to teach segments through the visual representation of frequency (including voicing), aspiration, and duration of such difficult consonants. For example, as for frequency, since Chinese has no voiced stops and only one voiced fricative, the language in general has a higher frequency range than English. Therefore, it is important at the beginning of pronunciation lessons for Chinese students to start building more sensitivity to sounds in the low-frequency range. EVF enables teachers to provide students with visual instruction on how to control frequency, such as in a minimal pair for /s/ and /z/. The visual display provides an objective measure that helps students focus their attention on the exact features of their pronunciation that need to be changed. This technique is also used in teaching vowels.

Many researchers have reported that EVF has been used by ESL learners for teaching various aspects of suprasegmental features, such as stress, rhythm, and intonation (e.g., Anderson-Hsieh, 1994, 1996; Chun, 1989, 1998, 2002; de Bot, 1983; Hardison, 2004, 2005b; Levis & Pickering, 2004; Molholt, 1988, 1990; Weltens & de Bot, 1984). Anderson-Hsieh (1994, 1996) also reported advantages of EVF in teaching suprasegmental features. On listening to spoken discourse, her ESL learners only focused on individual lexical items, and they tended to ignore the accompanying rhythm of utterances. In addition, a more serious problem was that they did not notice the importance of perceiving these suprasegmental features, so that they tended to have difficulty producing them.

By providing visual information about suprasegmental features in real time, it

becomes possible to raise learners' awareness of such speech characteristics, as well as providing an effective training procedure. For example, one of the typical problems that Japanese ESL learners have is transfer of their L1 rhythm, which is a "mora-timed rhythm," and their failure to highlight stressed syllables sufficiently because they use pitch accent instead of stress. Anderson-Hsieh (1996) used EVF in her classroom instruction. While EVF provided visualizations of the difference between the native speaker model's and students' own speech, the students were encouraged to repeat the words, make greater differentiation in length between stressed and unstressed syllables, and use higher pitch on the stressed syllables. She also reported that EVF was effective not only for word-level stress, but also sentence-level stress and intonation. Levis and Pickering (2004) also reported the use of speech visualization technology in teaching intonation at the discourse level. They claimed that providing practice with discourse-level intonation features is the next step in using technology for the teaching of intonation, so that learners can learn to use intonation for real communicative needs. For teaching prosody, Hardison (2004) also used a computer assisted speech training program by Real-Time Pitch (RTP) along with Kay Pentax Computerized Speech Lab (CSL), which displays simultaneously both of an instructor's and a learner's pitch contours for comparison, to teach French prosody to English speakers. In addition to RTP, in Hardison (2005a), Anvil, a web-based annotation tool integrating the video of a speech event with its pitch contour display was used to teach English prosody to Chinese speakers. For learners of Japanese, Landahl, Ziolkowski, Usami and Tunnock (1992) and Hirata (1999, 2004) reported effectiveness in teaching Japanese pitch contours using Visi-Pitch with CSL.

2.4.3 Experimental studies of EVF

Although their number is limited, several reports on studies of EVF have provided relevant experimental information concerning the number of subjects, statistical analysis of data, etc. de Bot (1983) conducted an experiment to assess the influence of auditory-visual feedback vs. solely auditory feedback on the learning of English intonation. The subjects in the experimental group were presented with a sentence through headphones. The F0 contour, i.e., pitch, of this sentence was plotted on a display, and then they had to imitate the sentence as their own F0 contours appeared on the display for comparison. In the experiment, practice time was another factor: one group received only one training session of 45 minutes while two sessions were provided for the other group. The control group followed the same procedure, but without visual feedback. The result of the experiment showed that visual feedback produced a significant effect on the learning L2 of intonation, whereas practice time was not a critical factor. In other words, optimum imitation of a sentence was reached sooner with auditory-visual feedback than with auditory feedback only. One of the advantages that de Bot pointed out was that the use of this kind of equipment tends to increase the subjects' motivation to try harder to achieve their learning target.

In Hardison (2004), 16 American learners of French received three weeks of training in French prosody using computerized displays of pitch contours as visual feedback. The results revealed significant effects of training in the acquisition of prosody. In addition, generalization to segmental accuracy and novel sentences was also found. Thus, the effect of training is apparent not only in the immediate focus of the visual feedback but also in novel tokens. Hardison's observation of the learners during sessions

suggested that there appeared to be a hierarchy of the learners' awareness, from more global elements such as the pitch contour, which was the focus of visual feedback, to more local elements such as individual sounds.

Further, Hardison (2005a) conducted prosody training with Chinese learners of English using a web-based annotation tool integrating the video of a speech event with visual displays showing the pitch contours and examined the effects of discourse-level input versus sentence-level input. The presence of video was more helpful with discourse-level input than with individual sentences. Here again, high variability of the stimulus was effective in combination with auditory and visual input sources.

However, as Anderson-Hsieh (1996) pointed out, EVF has some drawbacks, too. The major disadvantage of EVF is that the commercial hardware may be too costly to use in language laboratory settings and for individuals, e.g., it may be too costly to purchase Kay Pentax products. It is also not convenient for use in large classes except for demonstration. However, there are a number of free or low-cost programs available for use as "e-learning" tools (e.g., Praat, SIL Speech Analysis software WaveSurfer). Another point that should be considered is that instructors need to acquire technical knowledge to read some types of visual displays, and their careful control of the information in guiding students is indispensable.

As we have seen, there are many studies reporting on the use of EVF to improve learners' production of segments (vowels and consonants) and of suprasegmental features (e.g., tone, stress, and intonation). This present study examined the possibility of using EVF for enhancing the learning of durational contrasts, mainly related to geminate consonants in Japanese, through the display of waveforms, which make duration visible,

as discussed in the following chapter (Experiment II). Further, EVF so far has been used mainly to train learners' production ability, and few reports have addressed perception improvement. As seen in the previous section, a number of studies concluded that gains from perception training in L2 contrasts can transfer to productive ability. In light of the previous studies of the effects of EVF, the present study explored the potential of visual input for both perception training and production learning of Japanese geminate consonants by American learners of Japanese.

2.5 Research questions and hypotheses

The present study was motivated by the following research questions and hypotheses:

1) How do L2 learners perceive and produce geminate consonants? Is there any particular phonetic context of geminate consonants, which makes perception and/or production more difficult for learners?

Many previous studies of geminate consonants have been conducted on native and nonnative speakers, but few studies have focused on the effect of the types of consonants and of phonetic contexts. I hypothesized that the learners' perception and production would be affected by phonetic environments, and this might be a cause of difficulties in acquiring the contrasts. The present study aimed to find if there are any particularly difficult contexts for learners. While previous studies which examined learners' perception and production used only words in isolation as stimuli, the present study also examined whether there was any difference between word-level and sentence-level performance, as either of these levels might constitute a difficult context

for the accurate perception and/or production of geminate consonants.

2) Are audio-visual instruction and training using visual displays of waveforms of geminate consonants more beneficial than auditory-only information?

Coupled with the results of the above research question, this study aimed to find an effective method of perceptual training. The effectiveness of visual input in addition to audio input in perceptual training has been reported by previous studies (e.g., Hardison 2003), so it was suggested that it was also effective in training learners in the perception of geminate consonants. Based on the results of the previous studies, and as a possible application of the theory of episodic memory (Hintzman 1986), I hypothesized that the perceptual training with visual information would also be successful in guiding learners' attention to critical durational contrasts; thus, the training would be more effective.

3) Does perceptual training improve production ability without any explicit production training?

Previous studies have suggested the relationship between perception and production and reported production improvement through perceptual training of /r/ and /l/ contrasts (e.g., Akahane-Yamada et al., 1996). This research hypothesized that the perceptual training with visual input would also lead to development in the ability to produce geminate consonants; therefore there was a close link between perception and production would be demonstrated. The participants in this study were given only perceptual training, but they were given production tests to examine whether their ability to produce geminate consonants improved at the same time.

CHAPTER 3

Experiment I

3.1 Overview of the experiments

In the present study, a total of four experiments were conducted. The subjects were all native speakers of American English who were studying Japanese at the university level. Experiments I through III concerned the difficulties that the learners encountered with regard to geminate consonants. Experiment I and Experiment III were conducted to obtain perception and production data, respectively. Considering the learners' difficulties found in Experiment I, Experiment II was conducted as a pilot study to test electronic visual input as a method to improve their perception. Based on the findings of these three experiments, a training method was explored, and Experiment IV was conducted to test the effect of the training.

3.2 Objectives of Experiment I

Many of the previous studies of geminate consonants limited the test items to stops and did not refer to the types of geminates tested and their phonetic contexts. This study aimed to examine if there are any particular phonetic contexts for and identities of geminate consonants that make perception more difficult for learners. Through detailed examination of such conditions, the research question of Experiment I is thus to find what causes the learners' difficulty in perceiving a particular type of geminate consonant. As discussed above, the closure duration of stop geminate consonants produced by native speakers varies depending on the identity of the consonant itself, and this may affect

non-native speakers' perception. Furthermore, consonant types other than stops should be considered to see if there are any particular consonant types and phonetic conditions in which learners find it difficult to distinguish a singleton from a geminate consonant.

First, I hypothesized that one of the causes of difficulty in acquiring accurate perception of geminates was related to the sonority of the target segments. Previous studies reported that there was no effect on the perception of Japanese geminates of the following vowel for native speakers (Hirato & Watanabe, 1987). However, there may be some effect on learners' perception resulting from the identity of the vowel (/a/, /i/, /u/, /e/, or /o/) that follows a geminate consonant (Minagawa & Kiritani, 1996). In order to see if there was any effect of the following vowel on the perception of a geminate consonant, /a/ and /u/ were selected. These two vowels have different levels of sonority according to a scale which is considered universally applicable. Sonority is a ranking on a scale that reflects the degree of openness of the vocal apparatus during production, or the relative amount of energy produced during the sound (Goldsmith, 1990). The sonority hierarchy is generally described as having the organization shown in Figure 3.1. Japanese has a five-vowel system, which consists of /a/, /i/, /u/, /e/, and /o/. Between the two vowels selected for this experiment, /a/ has the highest sonority and /u/ has the lowest sonority.

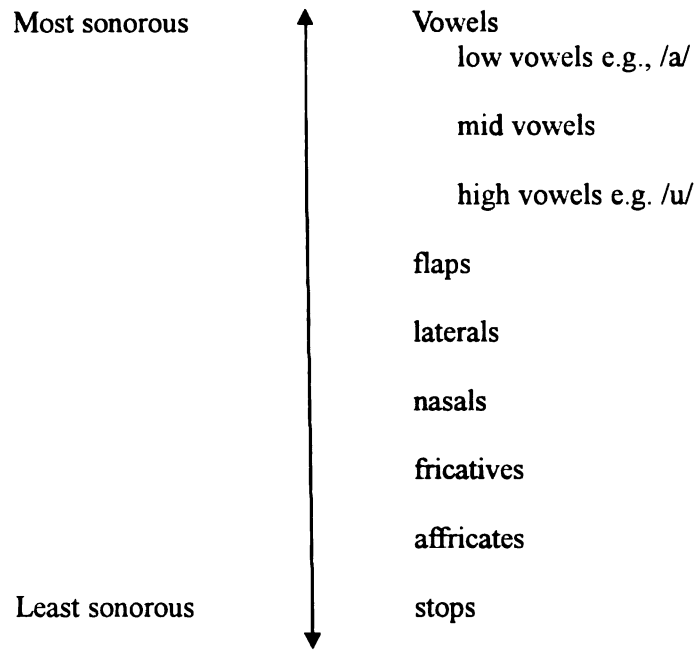


Figure 3.1. Sonority hierarchy (from Goldsmith, 1990; p.110)

Thus, it was also examined how the sonority of the following vowels would affect the learners' perception of geminate consonants. Since hierarchies do not indicate an actual degree of distance, Selkirk (1984) proposed the quantification of sonority in a Sonority Index as shown in Figure 3.2. The higher the number is, the greater its sonority.

10	9	8	7	6	5	4	3	2	1	0.5
a	e, o	i, u	r	l	m, n	s	v, z,	f,θ	b, d, g	p, t, k

Figure 3.2. Sonority index (Selkirk, 1984, p. 112)

In addition, according to this index, in bisyllabic words, the sonority distance between a

geminate consonant and the following vowel is closer in a fricative (e.g., in *sassa*) than when it is a stop (e.g., in *sakka*). The bigger difference might help to perceptually highlight the boundary between the geminate consonant and the following vowel, aiding speech perception (Kenstowicz, 1994), while the closer difference might obscure the boundary between the consonant and the vowel. Highlighting the boundary between the geminate consonant and the following vowel would make the precise duration of the geminate easier to perceive. Thus, it could be predicted that the learners would have more troubles with perceiving words containing an /ss/ fricative geminate consonants than those containing a /kk/ stop geminate.

Another hypothesis is that English, the learners' L1, may play a role in determining their ability to perceive a geminate consonant in Japanese to some extent. English has a constraint called the Maximum Onset Principle in syllabification; it says that intervocalic consonants should be syllabified into the onset of the second syllable rather than the coda of the first syllable. Thus consonants are preferred in the onset position, while no coda consonants are preferred except in the word final position (Goldsmith, 1990). According to this principle, the preferred syllabification of VCCV is V.CCV rather than the syllabifications VC.CV or VCC.V. Since the L1 of all the participants of this study is English, they might determine a syllable boundary by following this principle. It could be predicted that if a learner failed to perceive the mora weight (two morae for a vowel plus a geminate consonant) correctly, s/he might have a bias toward assigning the consonant as part of the onset, so that s/he might perceive a geminate consonant as a singleton as CV.CV. If the entire geminate is syllabified as part of the onset, then it cannot have moraic weight (Hayes, 1989), that is, it cannot be a

geminate.

For ease of exposition, following Kenstowicz' (1994) and Hayes' (1989) description of moraic syllable structure, geminate and nongeminate consonants are represents as follow (syllable= σ ; mora= μ). For example, in a CVC monosyllabic word in English, the vowel in nucleus is assigned one mora and consonants in onset and coda positions are nonmoraic as shown in Figure 3.3.

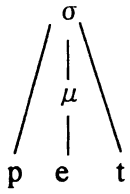


Figure 3.3. CVC word (e.g., "pet" in English)

On the other hand, the first part of a geminate consonant is moraic. For example, the Japanese word /sakka/ containing a geminate consonant is syllabified as a trimoraic bisyllabic word as shown in Figure 3.4.

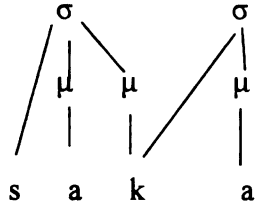


Figure 3.4. CVC.CV word (e.g. "*sakka*" in Japanese)

It has been suggested that the syllable plays a important role in the processing of speech

sound segments (e.g., Derwing, 1992; Ishikawa, 2002; Mehler et al., 1981; Schiller, Meyer & Levelt, 1997). In Japanese, morae as timing units have to be processed in addition to syllables, which may cause difficulties for L2 learners.

As another possible source of difficulty for learners, the question of whether there is any difference in identification accuracy for geminates in words in isolation as opposed to those embedded in sentences was examined. Traditionally, the dominant method for examining learners' development of the ability to perceive new, difficult nonnative contrasts has been to use a two-alternative identification or discrimination task with minimal pairs in isolation. For example, many studies use a two-alternative identification task; the learners are presented with stimuli consisting of minimal pairs for /l/ and /r/ in isolation (Bradlow et al., 1997; Bradlow et al., 1999; Hardison, 2003; Lively, Logan, & Pisoni, 1993; Pisoni, Lively, & Logan, 1994). With regard to the present study, Minagawa (1996), Hayes (2002), and Min (1993) all used a two-alternative identification task, having the learners identify minimal pairs for single/geminate consonants in isolation. A question raised here is whether a two-alternative forced choice task using minimal pairs in isolation is sufficient. Would the results reflect the overall perception ability of learners in a variety of phonological contexts? In actual conversations, learners have to identify phones or sounds and syllables in a flow of sounds, a longer and more complicated context than that of words in isolation. Is there any difference in learners' perception at the sentence level?

3.3 Method

3.3.1 Participants

All participants were undergraduate students at a large university in the U.S., and all were native speakers of American English whose ages ranged 19 though 22. There were no heritage learners. They were divided into three groups on the basis of the level of the Japanese courses in which they were enrolled at the time of the data collection. The 101 level group was made up of students in first-year Japanese language courses at the university (n=28; 7 females and 21 males). The students enrolled in the 101 level had almost no previous knowledge of Japanese, and it had been about three months since they had begun to study the language. The 201-group of students in the second-year Japanese language courses (n=42, 17 females and 25 males) was composed of those students who had passed the first-year class in Japanese. It was the third semester for these students. The 401-group of students in fourth-year Japanese (n=15; 5 females and 10 males) was in their seventh semester of studying Japanese. All the 401-level learners had studied in Japan for one or two semesters. Generally, it can be said the 401-level students had had more interactions with native Japanese speakers than had the students in the other lower levels, although this did not guarantee that they had become proficient proportionally, since the learning opportunities, motivations, and L2 uses of Japanese varied among the students. In the regular introductory Japanese classes, the first- and second-year courses, the students met for 50 minutes, 5 times per week. There were substantial oral drills and communicative activities in class, and the instructor sometimes corrected the students' inaccurate pronunciations. However, there was no special training for discriminating particular phonemic contrasts in class.

3.3.2 Materials

Stimuli consisted of 30 bisyllabic Japanese real words and non-words, which included 12 singletons, with the segmental form /(C)V.CV/, and 12 geminate counterparts, with the segmental form /(C)V.C.CV/, where the first CV and the final CV were identical, and 6 fillers which were bisyllabic words consisting of three morae, but including no geminate consonant. As defined in 2.1.1, a mora is a unit of timing, and each mora has approximately the same duration in production. Long vowels (e.g., /kiite/ 'listening') and geminate consonants (e.g., /kitte/ 'a stamp') take twice as long to produce as a short vowel or singleton counterpart (e.g. /kite/ 'coming').

Two tests were conducted; in Test 1, the words were heard in isolation, but in Test 2, the following carrier sentences were used.

watashi	*wa*	_____	*to*	*iimashita*
I	topic marker		that	said

'I said _____.'

kore	*wa*	_____	*desu*
this	topic marker		is

'This is _____.'

Note that as the word-for-word gloss shows, in Japanese word order, the stimuli come in the middle of the sentence, instead of the sentence final position seen in the English translations. The same set of 30 words was used for both tests, but the order of presentation was randomized.

As for the target consonants, the stops /t/, /k/ and a fricative /s/ were the same sounds as were used in Hayes' (2002) study (cf. Ch. 2.1.4). In Hayes' study, learners were

presented with a set of two words in each trial, either geminate-geminate, geminate-singleton, or singleton-singleton combination, and the learners were given a same-different discrimination task, i.e., they were asked to determine whether the two words were same or different. Such a discrimination task might not directly reflect learners' linguistic perception ability since it is rare to encounter a comparison of two sounds in a natural setting. In the present experiment, the learners were presented with only one token in each trial, and they were given an identification task to identify whether the token was a geminate or a singleton.

Previous studies revealed that the duration of the vowel preceding a geminate consonant plays a role as an acoustic cue for native speakers, but not for non-native speakers (Min, 1987), and that variable was therefore excluded from consideration in the present study. To test for the effect of the difference between /CV/ and /V/ as preceding segments, /sa/ and /a/ were chosen as preceding contexts. The vowels following a geminate consonant were a high vowel /u/ and a low vowel /a/. Since Experiment I did not consider the effect of pitch accent, the accent patterns of all the stimuli were kept consistent; they were H(igh)-L(ow) for singletons (i.e., two mora stimuli) and H-L-L for geminate consonants and long vowels (i.e., three mora stimuli). The stimuli consisted of 14 non-words and 16 real-words. In Table 3.1 below, translations are given for real-words; non-words are indicated with '*.'

Table 3.1

Examples of test items by phonological structure

Singleton	Geminate
/t/	/tt/
/s/	/ss/
/k/	/kk/

Structure of items		Example V_2=/u/		Example V_2=/a/	
(C)V_1.tV_2	(C)V_1t.tV_2	satu 'volume'	sattu *	sata 'trouble'	satta 'went'
(C)V_1.sV_2	(C)V_1s.sV_2	sasu 'stab'	sassu 'guess'	sasa *	sassa 'quickly'
(C)V_1.kV_2	(C)V_1k.kV_2	saku 'tear'	sakku 'sack'	saka 'refreshments'	sakka 'composer'

3.3.3 Procedure

The above stimuli were recorded by a female native speaker of Japanese (Tokyo dialect) using a SONY MD MZ-RH10-S with a SONY ECM-CS10 microphone. The stimuli were presented on the same mini-disk player in a classroom setting. Each recorded token was played only once.

A three-alternative forced choice identification procedure was used for the word-level and the sentence-level tests. The learners were asked to choose from one of three choices on their response sheets, which consisted of minimal triplets including a singleton /(C)V.CV/, a geminate consonant /(C)V.C.CV/ and a long vowel /(C)V.V.CV/, where the first CV and the final CV were identical in each item in the triplet (in the case of vowel-initial words, the first vowel was the same in each item).

Below are the examples of the test items:

The participants heard: /sassu/

Choices given (instructed to circle one):

a. sasu
b. saasu
c. sassu

The participants heard: /aka/

Choices given (instructed to circle one):

a. aka
b. aaka
c. akka

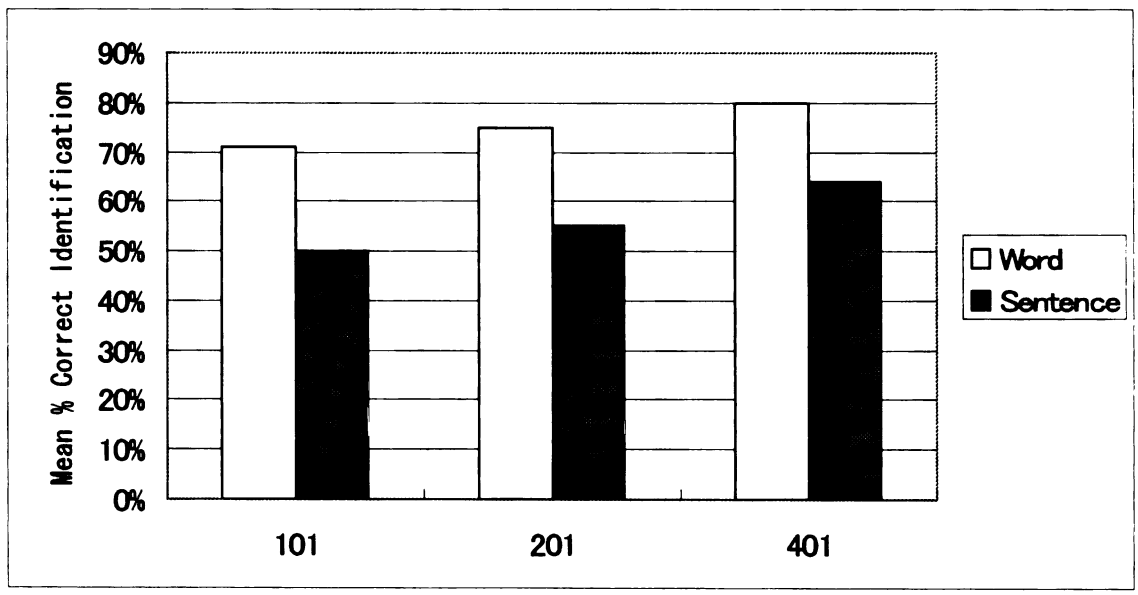
The answer sheets were collected, and correct answers were tabulated for each participant; if an answer was correct, one point was given, but no point was given for an incorrect answer. Two types of test were given to each participant, words in isolation and words in frame sentences as previously described. There was a total of 24 points in each test.

3.4 Results

The first set of data was scored by totaling the number of correct responses as in Figure 3.5. A mixed design 2 x 3 ANOVA was used with test (word level, sentence level) as between-group variable, and with level (101, 201, 401) as within-group variable. Test and level had significant main effects, $F_t(1, 102) = 181.15, p = .000$; $F_l(2, 102) = 4.085$, $p = .020$. The words in carrier sentences produced more errors (56%) than the words in isolation (75%). The interaction of Test x Level was not significant ($F(2, 102) = 1.035, p = .359$), which indicated the difficulty of the sentence-level test as opposed to that of the

word-level test was compatible throughout the levels.

Comparison among the levels in a post-hoc test (Tukey's HSD) revealed that there was no significant difference between the 101 level (61%) and the 201 level (65%). However, the 401 level (72%) was better than the 101 level at significant levels (p = .015) and the 201 level (p = .045). It is assumed that the 401 level students' superior performance could also be due to much more exposure to Japanese language through actually studying in Japan. Over the course of Japanese language study and exposure, native English speakers learning Japanese develop increased sensitivity to consonant duration. This result supports the findings of Hayes (2002).

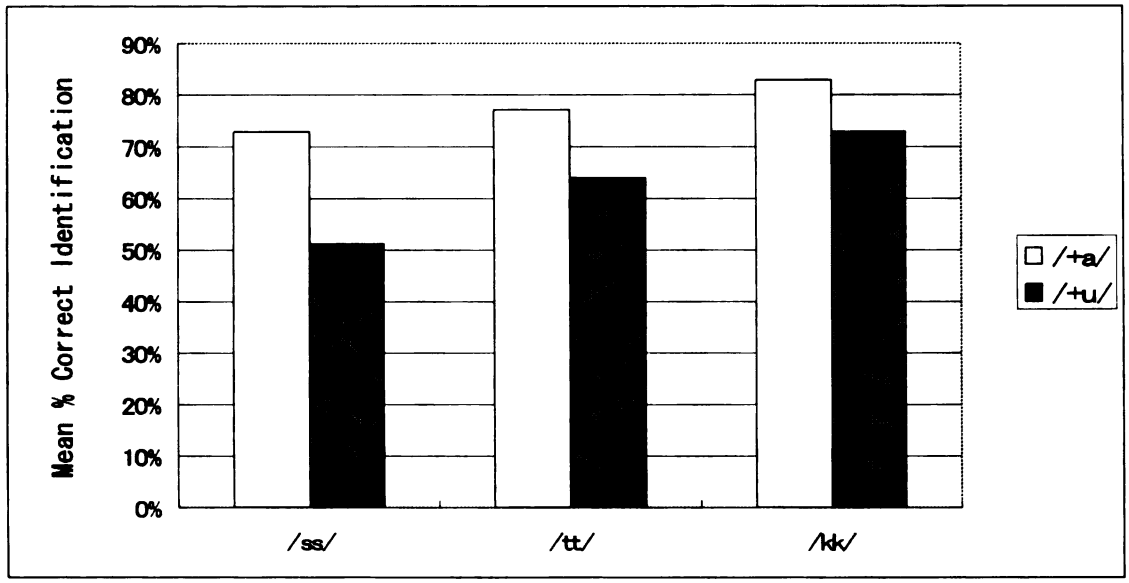


	101	201	401
Word level	71%	75%	80%
Sentence level	50%	55%	64%

Figure 3.5. Mean percent correct identification by level of proficiency

Next, word-level perception and sentence-level perception of geminate

consonants were examined in detail separately. A detailed analysis was made of the data from the 201 level students, which was the largest group of the three. Figure 3.6 shows the data of word-level perception of geminate consonants followed by /a/ (N=6, *M*=4.57, *SD*=1.70) and /u/ (N=6, *M*=3.95, *SD*=1.63).



	/ss/	/tt/	/kk/
/+a/	73%	77%	83%
/+u/	51%	64%	73%

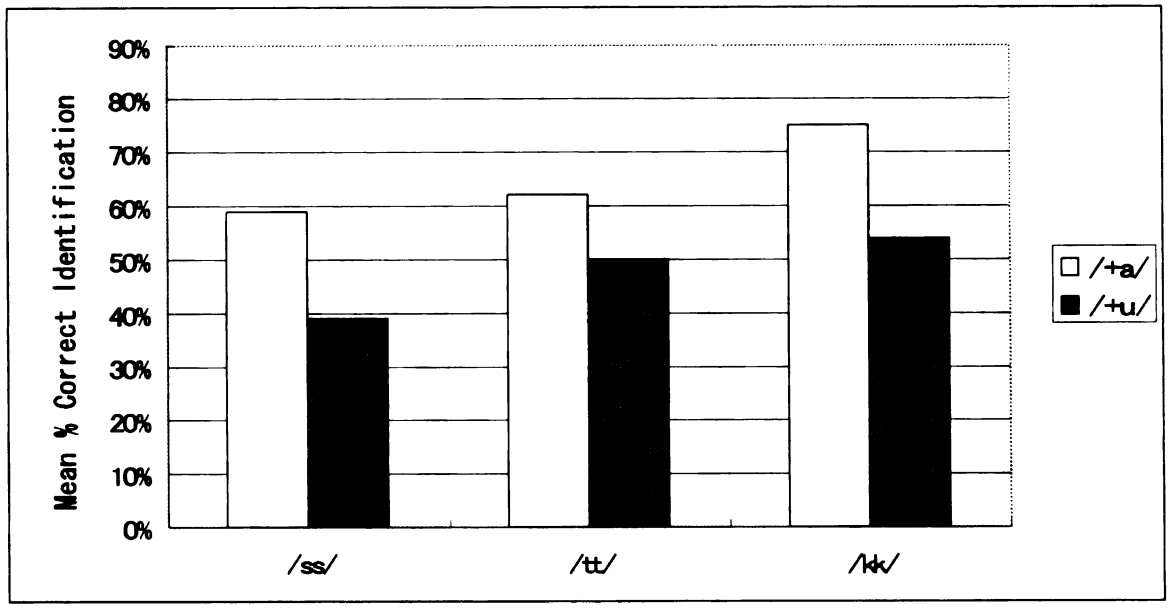
Figure 3.6. Mean percent correct identification at word-level by item condition (Japanese 201 students)

A two-factor repeated measures ANOVA was conducted. Variables were consonant (/s/, /t/, k/) and vowel (/a/, /u/). Both had significant main effects, F_c (2, 198) = 38.500, p = .000; F_v (1, 99) = 46.718, p = .000. The interaction of Consonant x Vowel was also significant (F (2, 198) = 7.579, p = .001). /s/ consonants were the most difficult

to perceive as geminate (73%), while /k/ (83%) was the easiest, and /t/ was in the middle (77%). As for the vowel following a geminate consonant, the geminates preceding /u/ in the final position (63%) were more difficult to perceive than those preceding /a/ (78%). This result may indicate that, as predicted, the sonority of a vowel following a geminate consonant played an important role in the learners' perception. Thus, of all the geminate consonant types, /ss + u/ was the most difficult for the learners to perceive (51%) while /kk + a/ was the easiest (83%), as shown. In addition, the difference between the effects of /a/ and /u/ was the biggest in following /ss/, while the perception of /tt/ and /kk/ showed almost parallel effects. This result does not support Hayes' (2002) result, which showed that there was no difference between /kk/ and /ss/, and that /tt/ was the easiest to perceive. This discrepancy may be the result of a difference between a discrimination task, as in Hayes' study, and the identification task used in the present study.

In addition, there was no significant difference between the scores with /sa/ and /a/ as the preceding segments (t (299) = -.218, p = .828). That is, the difficulty in perception was not affected by the difference between preceding segments with consonant + vowel or with vowel only.

Figure 3.7, which is also a detailed analysis of 201 students' data, shows sentence-level perception of geminate consonants followed by /a/ (N=6, M=3.16, SD=1.37) and /u/ (N=6, M=2.36, SD=1.47).



	/ss/	/tt/	/kk/
/+a/	59%	62%	75%
/+u/	39%	50%	54%

Figure 3.7. Mean percent correct identification at sentence-level by item condition (Japanese 201 students)

Again, a two-factor repeated measures ANOVA was conducted, and variables were consonant (/s/, /t/, k/) and vowel (/a/, /u/). Both had significant main effects, F_c (2, 198) = 32.575, p = .000; F_v (1, 99) = 43.190, p = .000. As in the word-level perception, the geminates preceding /u/ in the final position (48%) were more difficult to perceive than those preceding /a/ (65%). As for the consonant, /s/ consonants were the most difficult to perceive (49%), while /k/ (65%) was the easiest, and /t/ was in the middle (56%) as in the word-level data. However, the interaction of Consonant x Vowel was not significant (F (2, 198) = 1.740, p = .178). Similar difficulty of /u/ compared to /a/ was observed across the three consonants at the sentence level; /s/+/u/ was the most difficult

combination.

In the sentence level, too, there was no difference between the scores with /sa/ and /a/ as the preceding segments (t (299) = -.1738, p = .083). Again, the difficulty in perception was not affected by the preceding segments at the sentence level.

The data in Figures 3.8 and 3.9 show how the learners perceived a geminate consonant when they did not perceive it correctly, at word-level and sentence-level, respectively. In order to enable a closer examination of the most difficult item, i.e., a geminate + /u/, the data are broken down by error pattern. Some of the learners who could not perceive a geminate consonant correctly tended to think that the word contained a long vowel. This tendency was especially strong in the /ss + u/ geminate sequences.

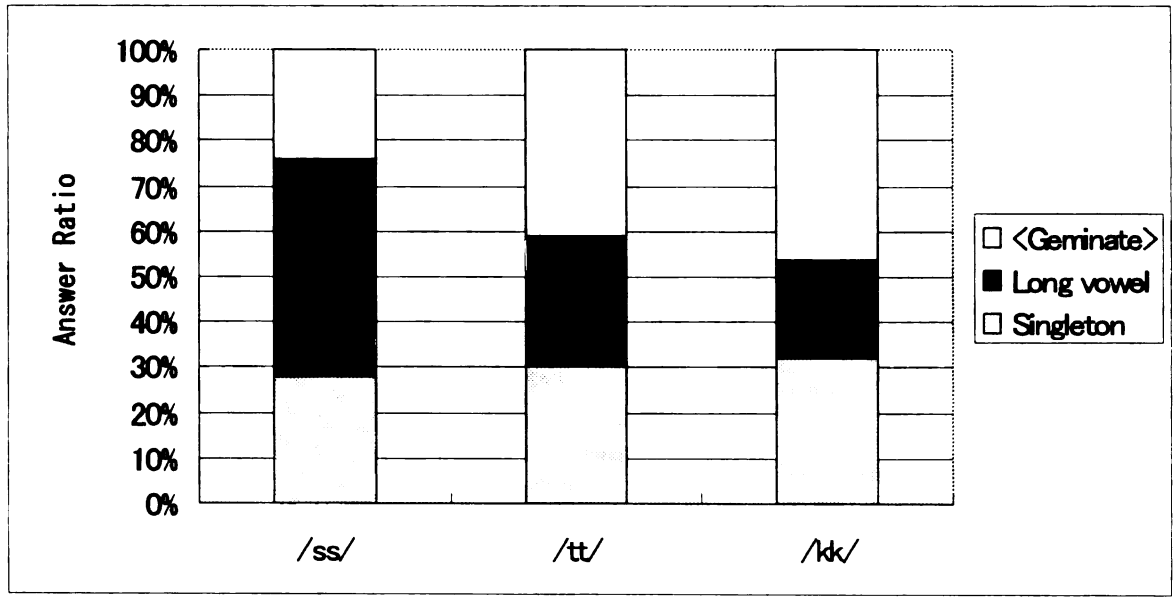


Figure 3.8. 201 students' perception of /CC+u/ at word-level

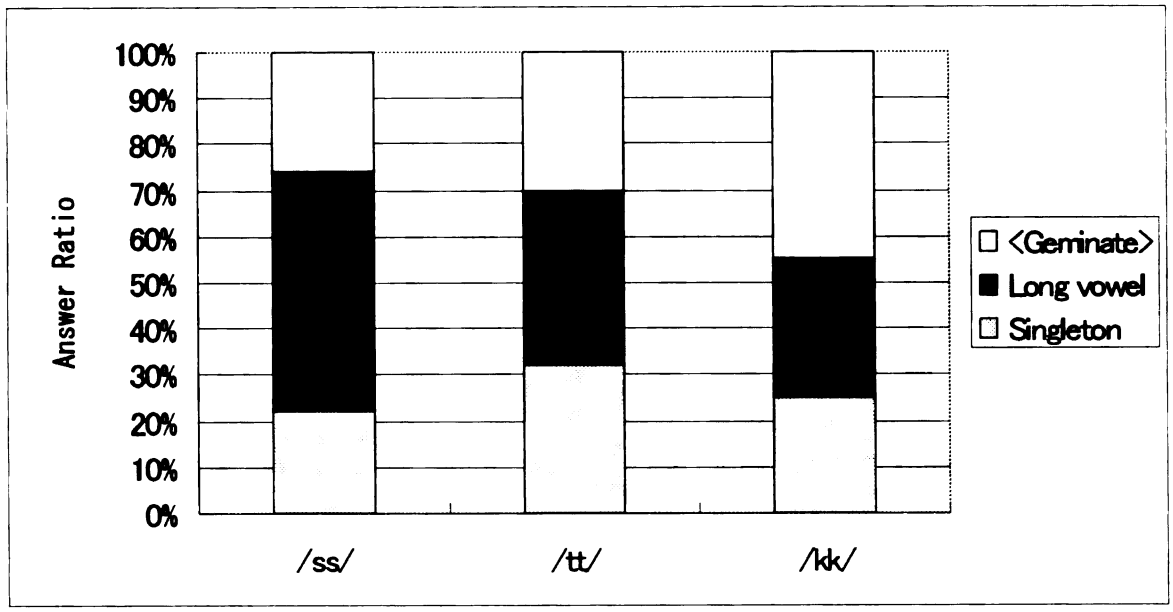


Figure 3.9. 201 students' perception of /CC+u/ at sentence-level

Yamagata and Preston's (1999) study shows an interesting correlation with this result. They conducted a study to see how English-speaking learners of Japanese acquired the spelling of English loanwords in Japanese. English loanwords are converted into Japanese spelling and generally follow the phonological system. This conversion is made very systematically, and gemination plays an important role. For example, monosyllabic words of CVC syllable structure with a lax vowel are systematically realized with gemination of the coda consonant (e.g., pot [*potto*]), and that is basically how native speakers of Japanese perceive the English word.

Yamagata and Preston had the learners spell some English words in Japanese to see how they perceived the English words in terms of the Japanese phonology they were acquiring and precisely how they conformed them to Japanese phonological rules. The results showed that the learners often lengthen vowels where native Japanese speakers spelled the words with geminates. Yamagata and Preston concluded that, although the

learners failed to geminate, they still felt the demands of giving the target word the same number of morae which it would have had if gemination had occurred. Since the spelling may or may not reflect the learners' actual production or perception, these results should be treated with caution. However, this result is compatible with the present findings: even if the learners in the present study failed to perceive geminate consonants correctly, they often perceived the geminate consonants as long vowels; hence, they could perceive the correct number of morae. Since long vowels add the same number of morae as geminate consonants, it can be assumed that the learners can perceive the mora weight correctly, particularly in the /ss/ condition.

3.5 Discussion

In Experiment I, it was found that certain types of geminate consonants were more difficult for the learners to perceive. First, a geminate consonant followed by the high vowel /u/ was more difficult to perceive than one followed by the low vowel /a/. The lower sonority of /u/ versus /a/ suggests that the learners may depend on the perceptibility of the boundary between the second part of the geminate consonant and the following vowel. Another observation was that a geminate consonant read in a sentence frame was more difficult to perceive than one read in isolation. This result indicates that the ability to perceive geminate consonants in isolation does not always guarantee the ability to perceive them in fluent speech, which the learners will encounter in natural settings.

The learners with more Japanese language experience exhibited a better ability to identify durational contrasts of single versus geminate consonants. The difference between adjacent instructional levels (101 and 201 levels) was not significant, but the

general pattern of improvement over time is apparent through the upward slope of the identification accuracy rate. The results are compatible with those of Enomoto (1989), who also reported that advanced level learners had a clearer perceptual boundary for identifying geminate consonants, compared to beginning learners. These results also suggest the possibility of improving learners' perceptual accuracy.

This result, which showed the subjects' performance over length of exposure to Japanese language study, confirmed the result of Hayes' (2002) study. However, the data in Figures 3.6 and 3.7 do not support Hayes' result, which showed that /tt/ was the easiest to perceive and that there were no differences between the /ss/ and /kk/ conditions. In this study, /ss/ shows the lowest correct rates – it was the most difficult to perceive. With regard to the stop consonants /kk/ and /tt/, the sonority distance between the consonant and the following vowel is bigger than with /ss/, and as predicted, it is found that the perception of consonant-vowel boundary was easier than with /ss/. Although the results differed from those of Hayes (2002), the present study also showed that the learners' perception varied, depending on the phonetic context in which the geminate consonants appeared.

Previous studies on single/geminate contrasts mainly used a two-alternative forced choice task, to characterize a consonant as either singleton or geminate, so it was perhaps assumed that if learners could not perceive a geminate consonant correctly, they must have perceived it as a singleton, i.e., they could not perceive the correct number of morae. However, as shown in Figures 3.8 and 3.9, many of those who could not perceive /ss/ geminate consonants thought that they perceived a long vowel instead. This result indicates that they at least perceived the correct duration of the relevant portion of the

word, i.e., having two morae instead of one. However, considering the fact that the type of consonant does not generally affect native speakers' perception of a geminate consonant, one might conclude that the learners rely on a different acoustic cue from native speakers when they perceive geminate consonants. Based on this result, at least for some phonological conditions, it is safe to assume that, in order to correctly perceive geminate consonants, the learners have to be able to accurately identify not only a single/geminate minimal contrast, but also a geminate/long vowel minimal contrast.

Let us consider what acoustic cues the learners might have used to identify geminate consonants. Figures 3.10 and 3.11 are the waveforms of /akku/ and /assu/, recorded by a female native speaker of Japanese. The speech-waves were created with Praat (http://www.fon.hum.uva.nl/praat/).

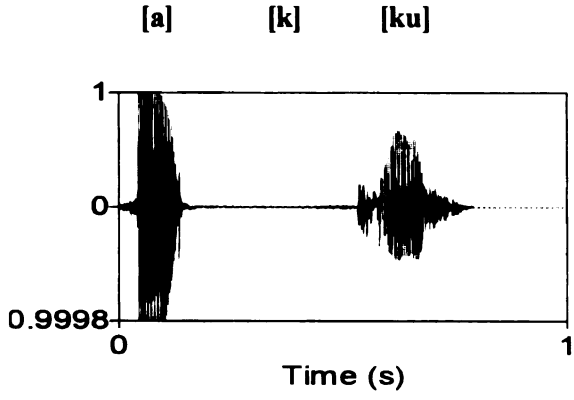


Figure 3.10. Waveform of /akku/

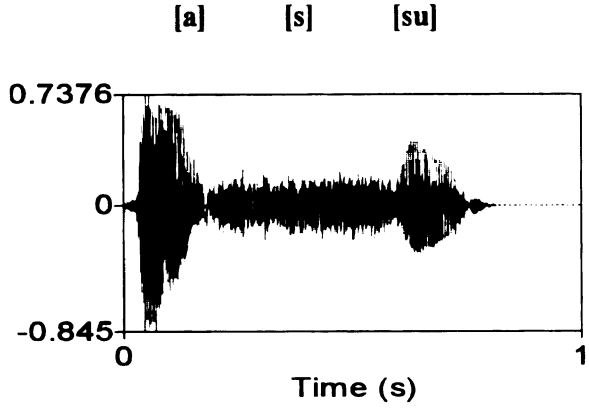


Figure 3.11. Waveform of /assu/[1]

As shown in the native speaker's utterance, the first part of the geminate stop consonant is silent, as shown in Figure 3.10. However, as shown in Figure 3.11, the first part of a fricative geminate consonant is filled with frication noise, which continues into the second part of the geminate. A possible explanation for the learners' difficulty in perceiving a geminate fricative consonant is this frication; the frication might have prevented the learners from perceiving the duration of the geminate consonant correctly. As observed in Figure 3.8 and 3.9, if a stimulus containing a geminate consonant was misperceived as one containing a long vowel, the learners could not tell accurately which segment the length should be attributed to; the length was incorrectly placed on the first vowel of the stimulus so that they misperceived the frication of the /ss/ geminate consonant as part of the vowel length. On the other hand, some learners could allocate length to the correct segment within the stimulus. When they correctly perceived a geminate consonant, length was attributed to a consonant.

[1] In the Tokyo dialect, /u/ is often devoiced between voiceless obstruents or in the word-final position, unless the vowel is in the position to receive an accent. However, this waveform showed that the test items containing /u/ were not devoiced. This may be due to the fact that the test items were read very carefully, since the speaker was aware of being recorded for the experiment.

This error pattern, shown in Figure 3.8 and 3.9, occurred most often when /u/ was the vowel following a geminate consonant. Thus, another difficulty found in the present study was perception of the stimuli with a low sonority vowel /u/ following a geminate consonant, compared to a high sonority vowel /a/. Clearly, perceptibility of the vowel affected learners' perception of mora weight of a geminate consonant.

In perception of a fricative consonant, there might be two scenarios regarding how the learners actually processed the stimuli containing geminate consonants: 1) they might have taken this longer frication for the onset of the second syllable; or 2) they might have perceived the frication noise as a coda of the first syllable. Considering the result that the phonetic conditions influenced the difficulty of perception of geminate consonants, principles of syllabification in English might influence which strategy was taken. As mentioned in Ch. 3.2., English L1 speakers generally tend to follow the Maximum Onset Principle for syllabification so that the consonants are preferred in the onset position rather than in the coda (Goldsmith, 1990). At the same time, we have observed that the learners' perception was also influenced by the types of vowels following geminate consonants. When they could perceive the following vowel clearly but could not allocate the length correctly, the learners might have followed the principle so that they syllabified the stimulus by the onset strategy as CV.CV, losing one mora as shown in Figure 3.12.

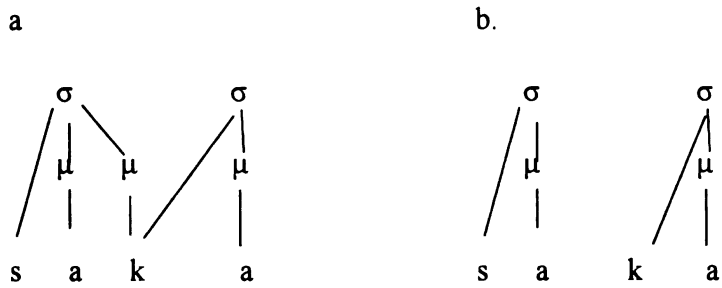


Figure 3.12. (a) The syllabification pattern of a geminate consonant when it is perceived correctly; (b) The pattern when a geminate is misperceived as a singleton

In such case, a mora might not be placed on the onset of the second syllable, so that the stimulus was perceived as a singleton. This assumption conforms to Hayes (1989), which argues that onset consonants should be non-moraic.

When they failed to hear clearly the vowel following a geminate, which was /u/ in most cases, it is assumed that they might have chosen to syllabify the consonant as a coda since the Maximal Onset Principle did not come into play. It is assumed that in such cases, the geminate consonant is syllabified as part of the coda of the first syllable. Specially in cases with /ss + u/, although many learners could allocate the correct mora weight, they failed to attribute it to the second part of the geminate consonant but attributed it wrongly to the vowel. This is probably because relatively sonorous fricative consonant and frication noise might interfere with the perception of /u/, and the second syllable might not be clearly heard. Thus, the learners misplaced the mora weight as shown in Figure 3.13.

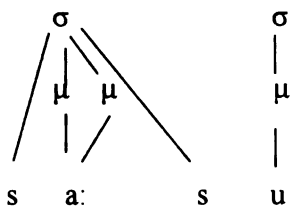


Figure 3.13. Syllabification pattern of a geminate consonant when the word is misperceived as containing a long vowel

Considering the result that /ss + u/ caused the most misperceptions, resulting in the perception of a long vowel, this might be the scenario which most frequently occurred. Thus, we have observed that an English syllabification strategy might be competing with the Japanese one. Other previous studies have also reported examples in which L2 learners were influenced by L1 syllabification strategies (e.g., Derwing, 1992; Ishikawa, 2002; Schiller, Meyer & Levelt, 1997). Further research is necessary since other factors, such as the learners' developmental pattern or their L1 (not only English but also other languages which have similar syllabification patterns) should be considered.

In this experiment, the question of whether the learners' perception of geminate consonants would be affected by some phonetic contexts and the identity of the geminate itself was investigated, based on the hypothesis that such difficult contexts and identities would be causes of learners' difficulty. The result revealed that the learners had difficulty especially in contexts with fricative geminate consonants and low sonority vowels. Hence, their performance, as predicted, was affected by phonetic contexts and identities. In some phonetic contexts, the learners relied on different processing cues from those used by

native speakers, which prevented accurate perception. The processing cues might have been associated with the English syllabification strategy, and it seems that for those students who misperceived geminate consonants, the processing strategy was competing with the Japanese mora processing strategy. On processing an /ss + u/ geminate consonant-vowel sequence, it was likely that learners were not aware of appropriate processing cues for the processing of morae.

The next chapter will examine a proposal for effective instruction, which would facilitate learners to identify durational contrasts, based on the assumption that the frication of a fricative consonant disrupts the perception of it as a geminate consonant.

CHAPTER 4

Experiment II

4.1 Objectives of Experiment II

As I have outlined in Chapter 2, electronic visual input is a type of computerized input which displays the acoustic analysis of an utterance and allows learners to visualize one or more features, such as duration, frequency range, etc. Researchers have reported the effectiveness of such training in improving learners' perception and production (e.g., de Bot, 1983; Hardison, 2003, 2004; Lambacher, 1999, 2001). The research question of Experiment II was to evaluate the potential benefit of instruction with visual information to enable learners to identify a geminate consonant, based on the assumption that the frication of a fricative consonant disrupts the perception of it as a geminate.

In Experiment I, it was found that the frication of a fricative consonant might prevent learners from perceiving /ss/ geminates correctly. Experiment II examined whether electronic visual input that displayed this frication noise would help learners improve their perception of geminate consonants. In addition, the stimuli were presented in different phonetic contexts (followed by five different vowels). As reviewed in Chapter 2, in multiple trace theory, all attended perceptual details of stimuli are stored as traces in memory (Goldinger, 1997), and a weighing scheme determines which phonetic features to be attended to in the perception of speech signals (Jusczyk 1993). Therefore, it was hypothesized that the visual input could successfully facilitate learners' shift of attention to mora weights of geminate consonants. If successful learning occurs, the information is stored and gathered as a composite of the traces that constitute episodic memory as

Hintzman (1986) and Hardison (2000) argue. Experiment II aimed to examine whether visual input was effective in altering the learner's weighting scheme and preserving the salient word length in memory. Such a process can be accounted for within the episodic model framework.

4.2 Experimental design

The experiment consisted of the following sequence: a pretest, a lecture on using electronic visual input, and a posttest. Two groups, an experimental group and a control group, participated in the study. The experimental group received the instruction with electronic visual input, while the control group received the instruction without visual input. The pretest and posttest involved a three-alternative forced choice task in the perception of geminate consonants in Japanese.

4.3 Method

4.3.1 Participants

The experimental group consisted of 33 learners, all native speakers of English, who were taking a beginning Japanese course (Japanese 101) at a large university in the U.S. They had received less than three months of Japanese instruction. Experiment II was limited to beginning learners so that variables in terms of learning experience could be controlled to some extent. The control group consisted of 31 learners, who were also taking the same Japanese 101 course. Classes were held five times a week, for 50 minutes each, and lectures were given that emphasized the grammatical use of linguistic structures in Japanese, followed by practice drills. Though some general lectures on

sounds in Japanese were given, there was no special perceptual training in durational contrasts during the class.

4.3.2 Materials

4.3.2.1 Pretest and posttest

The stimuli consisted of 20 non-words and real Japanese words in isolation. Ten words included fricative /ss/ geminates and singletons in five vowel environments (/a/, /i/, /u/, /e/, and /o/). For each token, there were three choices given, which constituted minimal triplets including a geminate consonant (e.g., /sassu/), a singleton (e.g., /sasu/) and a long vowel (e.g., /saasu/). The basic format of the task was similar to that of Experiment I. The participants were to identify the word they thought they heard. The other ten words, which contained stop geminate consonants and long vowels, served as fillers. The list of all 20 tokens was recorded by a female native speaker (the same speaker as in Experiment I) of Japanese (Tokyo dialect) using a SONY MD MZ-RH10-S with a SONY ECM-CS10 microphone, at a natural rate, with accent on the first syllable. Each token was read only once.

4.3.2.2 Instruction by electronic visual input

All the participants of the experimental group gathered and received a brief lecture in the classroom setting between the pretest and posttest. The instructor (the researcher, a different person from the speaker who recorded the stimuli) demonstrated how geminate consonants appear in waveforms. The instructor's model input of stop/fricative geminate contrasts, such as /sakku/ vs. /sassu/, were recorded real-time via

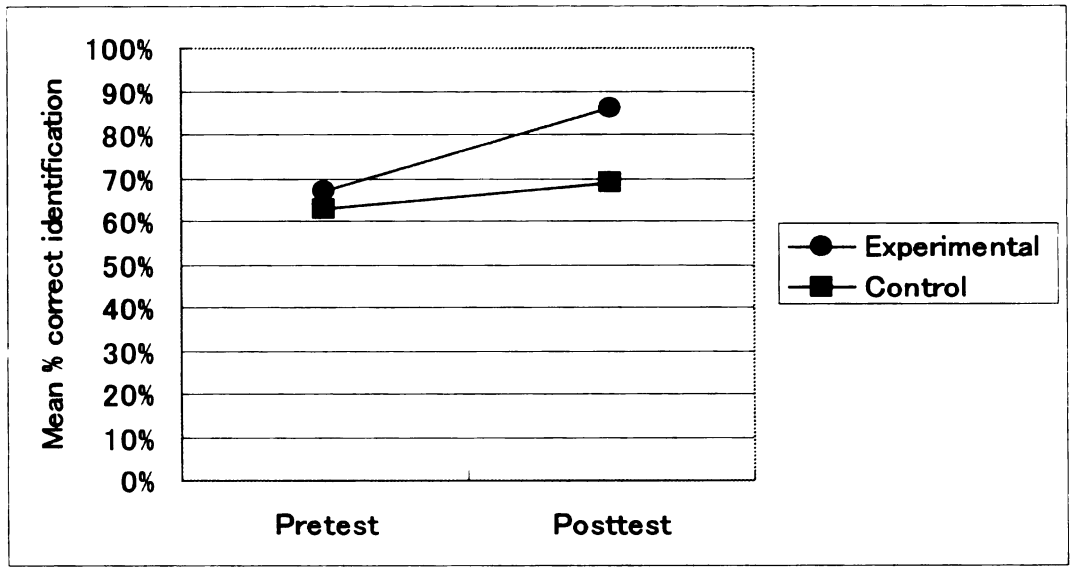
a microphone into the Praat acoustic analysis system and projected onto the screen in the classroom, so that the participants were able to see the waveforms. The focus of this instruction was to help learners understand the relation between the acoustic signal they were receiving and the electronic visual representation of different types of geminate consonants. This instruction was given for about five minutes, between the pretest and posttest. The control group received the same lecture on the difference between stop and fricative geminate consonants, but with no visual information.

4.3.3 Procedure

First, the pretest was conducted. The stimuli were presented separately to the experimental and the control groups, using a SONY mini disk MZ-RH10-S in a classroom setting. The participants listened to each stimulus only once. After the pretest, the instruction with visual input was given to the experimental group. The posttest was given in the same manner as the pretest, but the order of the stimuli was randomized. The whole procedure took about 15 minutes.

4.4 Results

The percent correct rates on perception of /ss/ geminate consonants are shown below for both groups.



	Experimental Group (M, SD)	Control Group (M, SD)
Pretest	67% (6.7, 1.25)	63% (6.3, 1.01)
Posttest	86% (8.6, 0.95)	69% (6.9, 1.19)

(Test items N=10)

Figure 4.1. Mean percent correct identification of /ss/ geminate consonants

There was no significant performance found between the two groups on pretest performance; t (18) = -.726, p = .477. To see the effect of the training, a mixed design 2 x 2 ANOVA was used with group (experimental, control) as the between-subjects variable and time (pretest, posttest) as the within-subjects variable. The results showed that the main effects of time and group (F_t (1, 18) = 15.918, p = 0.001, F_g (1, 18) = 7.694, p = 0.013) and the Time x Group interaction (F (1, 18) = 4.818, p = 0.042) were all significant. Improvement over time in the perception scores differed significantly between these two groups. The improvement for the group receiving training with visual information was greater than that for the control group.

4.5 Discussion

From the result of Experiment I, I hypothesized that the learners could not perceive a fricative geminate consonant correctly due to its frication. Experiment II demonstrated a significant effect of visual information on the learners' perception of fricative geminate consonants. The explanation of fricative sounds and geminate consonants with the aid of electronic visual input was given to the learners so that they might be able to pay special attention to fricative noise when listening to the test tokens again. They might be able to visualize the length of words even in the absence of waveform presentation during the task. The result of the posttest suggests that it is indeed possible to improve learners' perception through visual display.

Taken together, Experiments I and II suggest that English-speaking learners of Japanese use a different acoustic cue from native speakers in order to identify duration. This result is compatible with other studies done with learners whose L1 is other than English (e.g., Min, 1987 with Korean L1 learners; Minagawa & Kiritani, 1996 with Thai, Korean, and Chinese L1 learners). Treatment which focuses on those cues that the learners actually use can be effective in improving their perception; waveform displays of geminate consonants helped learners to identify geminate consonants in terms of mora attribution. It is assumed that visual information with waveform gives salience to mora length, and shift their attention to this distinctive feature. Through training with immediate feedback, learners add new traces which are matched against the ones already stored in memory.

Studies by Hardison (e.g., 2003) reported that L2 speakers can be effectively trained to perceive and produce new sounds through the auditory and visual modalities.

Experiment II also suggested that visual information of geminate consonants might have succeeded in refocusing learners' attention and altering their weighting schemes (Jusczyk 1993) by storing new traces and composing episodic memory (Hintzman 1986). These changes were reflected in the higher score of the posttest. However, Experiment II of this study cannot be called "training," considering the fact that its focus was too small (only fricative consonants) and too brief (comprising only five minutes of instruction). Although limited, the results of Experiment II suggest that instruction with electronic visual input can be incorporated into a more extended, effective training program.

The visual information used in Experiment II focused on frication, and it succeeded in making the learners pay attention to the focused item. However, as shown in the findings of Experiment I, the learners had greater difficulty in perceiving a geminate consonant in a frame sentence, so a training program should include practice to allow learners to become accustomed to perceiving a target phoneme in the natural flow of conversation, not only in an identification task involving words in isolation.

In Experiments I and II, the learners' perception of geminate consonants was examined and the possibility of electronic visual information for instruction was suggested. As hypothesized, the visual input could successfully facilitate learners' attention shift to the duration of geminate consonants. The visual input succeed in making the information stored and gathered as a composite of such traces constitute episodic memory as Hintzman (1986) and Hardison (2000) argue.

In Experiment III, the learners' production of geminate consonants was examined. Experiment IV explored the potential benefits of a training program using electronic visual information.

CHAPTER 5

Experiment III

5.1 Objectives of Experiment III

Experiment I examined how learners of Japanese perceive geminate consonants. The analysis included effects of the phonetic (specifically the following vocalic) environment of geminate consonants on the learners' error patterns. After Experiment I, Experiment II was conducted to examine whether electronic visual input was effective in mitigating the special difficulties that the learners showed in the perception data collected in Experiment I. Experiment III turned to production to see how learners of Japanese produce geminate consonants.

As in the studies of perception, previous studies of production of geminate consonants limited the test items to stops and did not refer to the types of geminates and their phonetic contexts. Experiment III investigated whether there were any phonetic contexts and identity of geminate consonants that affect learners' production. Thus, the research question was to find what causes the learners' difficulty in producing a geminate consonant through detailed examination of such conditions.

Based on the results of these three experiments, Experiment IV was then conducted to test a training method for learners of Japanese to perceive geminate consonants. The training method was developed by extending the results of Experiment II, which examined the effectiveness of electronic visual input, to a larger training program. In addition, based on the perception and production data from Experiments I and III, the training was developed and studied further in order to determine whether it was

specifically effective in the difficult phonetic contexts.

5.2 Method

5.2.1 Participants

The participants of Experiment III were 32 college students in an elementary Japanese course in a university in Japan. All of them were in Japan on study abroad exchange programs from their home institutions in the US. They were all native speakers of American English. Before coming to Japan, they had received no Japanese instruction. At the time of data collection, the students had received Japanese instruction for about two months at the university. The course met every day (five days a week) for 50 minutes. The class was designed to emphasize development of oral communication skills, but was not especially focused on sounds and pronunciation. In addition to the Japanese language course, the students took lecture courses in their major, such as economics, culture, and politics related to Japan. Those lecture courses were conducted in English by native English-speaking lecturers. The students who participated in the exchange programs had a choice of residence: they could stay with a volunteer Japanese host family or in the student dormitory. Since the number of students who used the homestay program was very small, the participants in Experiment III were limited to those who stayed in a student dormitory for the purpose of controlling the amount of Japanese to which the participants were exposed outside of class. They spoke primarily English in the dorm, which is understandable given that they were very beginning learners and their roommates were also international students whose common language was English. I also recruited five native speakers of Japanese, who were Tokyo dialect speakers and had no

sustained experience with English speakers. They served as a panel of judges who evaluated the learners' production.

5.2.2 Materials

Two types of production tests were used. One consisted of reading a list of words in isolation, and the other consisted of reading a list of words in frame sentences as shown below:

watashi	*wa*	_____	*to*	*iimashita*
I	topic marker		that	said

'I said _____.'

kore	*wa*	_____	*desu*
this	topic marker		is

'This is _____.'

The structures containing geminate consonants used for the test were identical to those of the perception test in Experiment I. Both tests included 12 minimal pairs of single/geminate consonants and 6 fillers, a total of 30 Japanese real words and non-words. The learners would not have known the real words, considering that those words had not been taught in class and they were very early beginners. In Table 5.1, the test items are categorized by phonetic structure.

Table 5.1

Examples of the test items

Structure of items		**Example**	**V_2=/u/**	**Example**	**V_2=/a/**
(C)V_1.tV_2	(C)V_1t.tV_2	satu 'volume'	sattu *	sata 'trouble'	satta 'went'
(C)V_1.sV_2	(C)V_1s.sV_2	sasu 'stab'	sassu 'guess'	sasa *	sassa 'quickly'
(C)V_1.kV_2	(C)V_1k.kV_2	saku 'tear'	sakku 'sack'	saka 'refreshments'	sakka 'compose'

* indicates non-words.

5.2.3 Recording procedure

The recording of production tests was carried out individually on cassette tape recorders in the language laboratory of the university. The participants were given a list of words and sentences and asked to read them at a normal speaking rate. The word list was given to the students before the recording so that they could become familiar with the words.

5.2.4 Judgment of production

The 32 cassette tapes submitted by the participants were transformed into WAV format. The data from all 32 participants were stored on a CD-ROM using NEC Lavie PC-LT 700AD, and copies were given to the five native Japanese participants for the judgment task in the language lab. They were provided a headset (Panasonic RP WH 5000-5) in a quiet room and listened to all the tokens of each learner. Their task was to write down the words that they thought they had just heard in standard Japanese orthography. For evaluation of sentence-level tests, the judges were given the frame

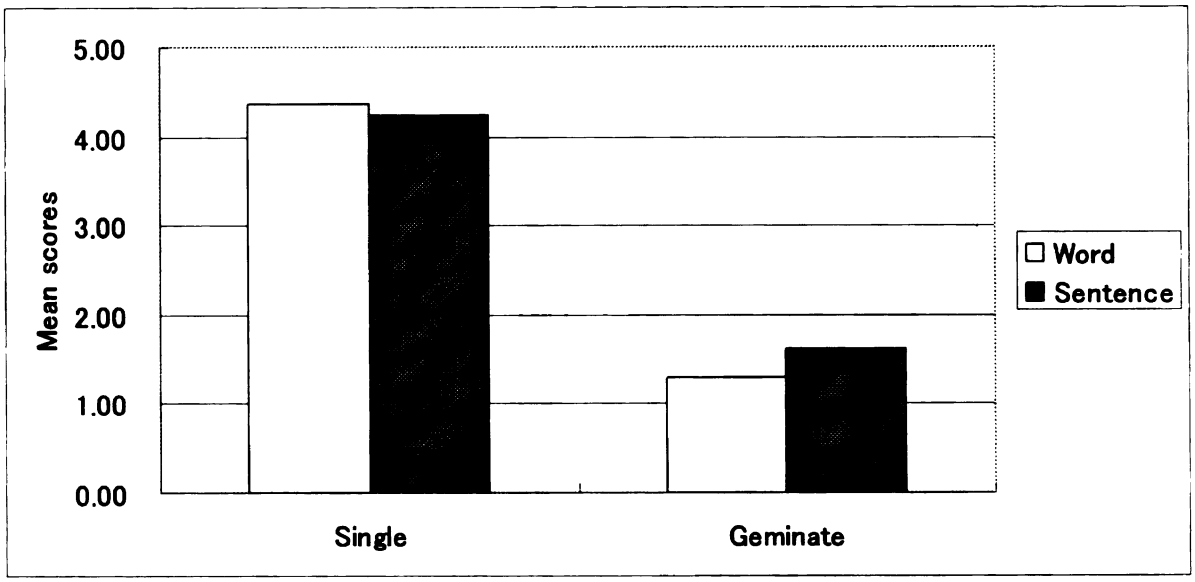
sentences on the evaluation sheets and asked to write in only the target words. The stimuli for each participant were blocked in the presentation given to the judges for ratings. The entire procedure took about 2-3 hours. At the end of the task, each judge was offered a gift card worth 1500 yen (about US$18) for their participation.

5.3 Results

5.3.1 Results by phonetic conditions

Correctly identified items were tabulated for each item and for each subject. If a judge correctly recognized the token the subject pronounced, 1 point was given; so if all five judges agreed in identifying pronounce word as the correct token, a maximum score of five was possible. For all ratings by the five judges, inter-rater reliability was determined (using Pearson correlation) and ranged from .73 to .94.

Figure 5.1 shows comparisons between mean scores for judgment of singleton and geminate consonants. The score for "Single" indicates the number of occasions on which the judges perceived a singleton pronounced correctly as a singleton; the score for "Geminate" indicates the number of occasions when a geminate consonant was pronounced correctly and thus perceived as such by the judges.

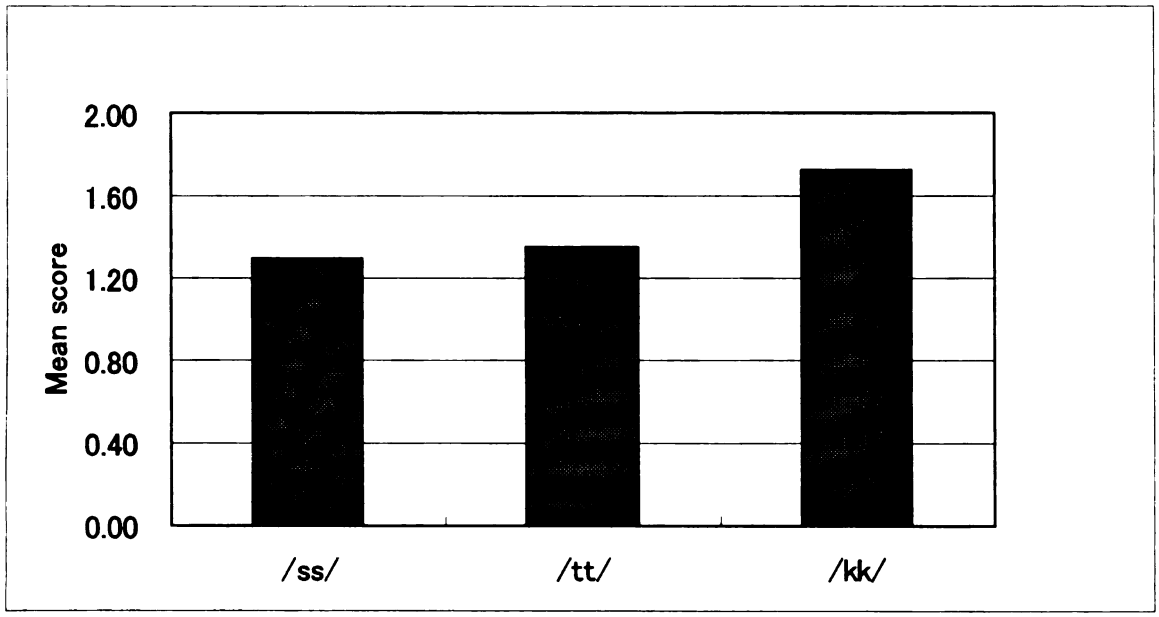


	Single (SD)	Geminate (SD)
Word level	4.35 (.26)	1.29 (.51)
Sentence level	4.23 (.31)	1.62 (.64)

Figure 5.1. Comparison of judges' mean scores between word level and sentence level for single and geminate production

First, a two-factor repeated measures ANOVA was conducted. Variables were test (word-level, sentence-level) and item (singleton, geminate consonants). Although the effect of test was not significant, F (1, 139) = 2.611, p = .108, a significant main effect was found for item, F (1, 139) = 488.663, p = .000, and the interaction of Test x Item was also significant, F (1, 139) = 10.327, p = .002. As in the perception data, producing geminate consonants was more difficult than producing singletons. Further, the mean accuracy of producing geminate consonants was higher at the sentence level than at the word level though the difference seems to be small, t (139) = -2.947, p = .004, while the production of singletons was comparable in both contexts. This result was opposite to the perception test in Experiment I, which showed a greater difficulty at the sentence level.

In order to examine which consonant types induced more errors in geminate consonant production, the data for each consonant were separated and tabulated separately as in Figure 5.2. The mean scores for the three consonant types were analyzed with a repeated measures ANOVA, and the pairwise comparison revealed a significant difference between /ss/ and /kk/, and between /tt/ and /kk/; F (2, 278) = 16.227, p = .000, but not between /ss/ and /tt/, (p = .471). This result is consistent with that of the perception test, which indicated that /ss/ was the most difficult but /kk/ was the easiest, although the difference between /ss/ and /tt/ was not significant.

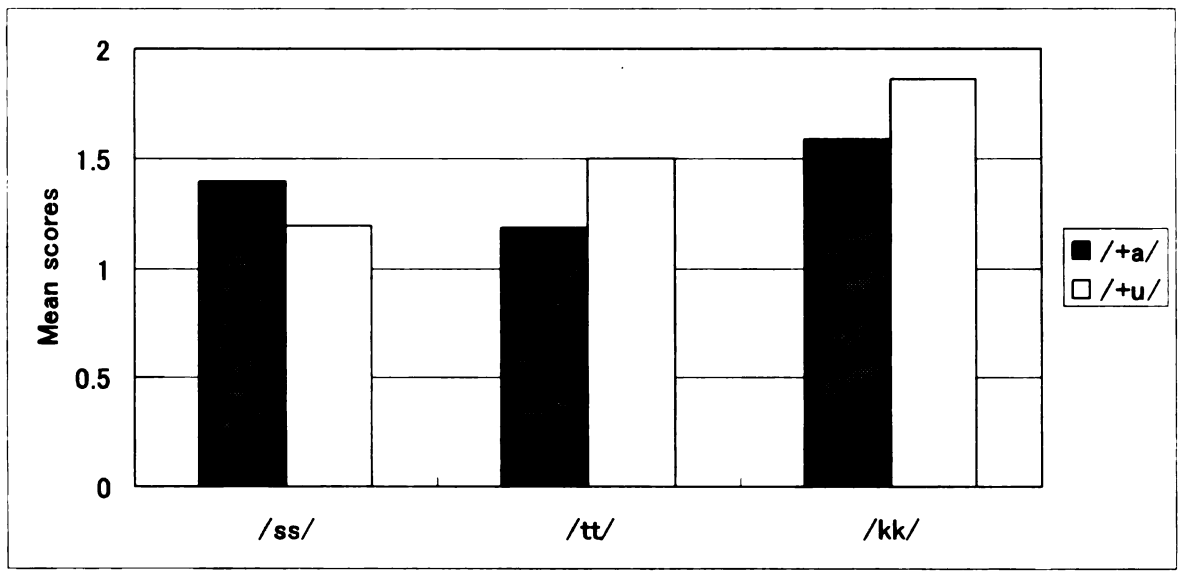


	/ss/ (SD)	/tt/ (SD)	/kk/ (SD)
Mean score	1.29 (.32)	1.35 (.46)	1.72 (.51)

Figure 5.2. Mean production scores by consonants (n=1,576)

Further, the effect of the phonetic environment on the three consonants was

examined. First, the segments following the geminate consonants were also separately examined as shown in Figure 5.3. There were two vowel conditions following the geminate consonants: /a/ and /u/. A two-factor repeated measures ANOVA was conducted with variables of vowel (/a/, /u/) and consonant (/s/, /t/, /k/). Both showed significant main effects, F_v (1, 278) = 4.948, p = .028; F_c (2, 278) = 16.227, p = .000. The interaction of Vowel x Consonant was also significant, F (2, 278) = 12.481, p = .000. When /u/ was the following vowel (1.52), fewer errors were produced than with /a/ (1.38).



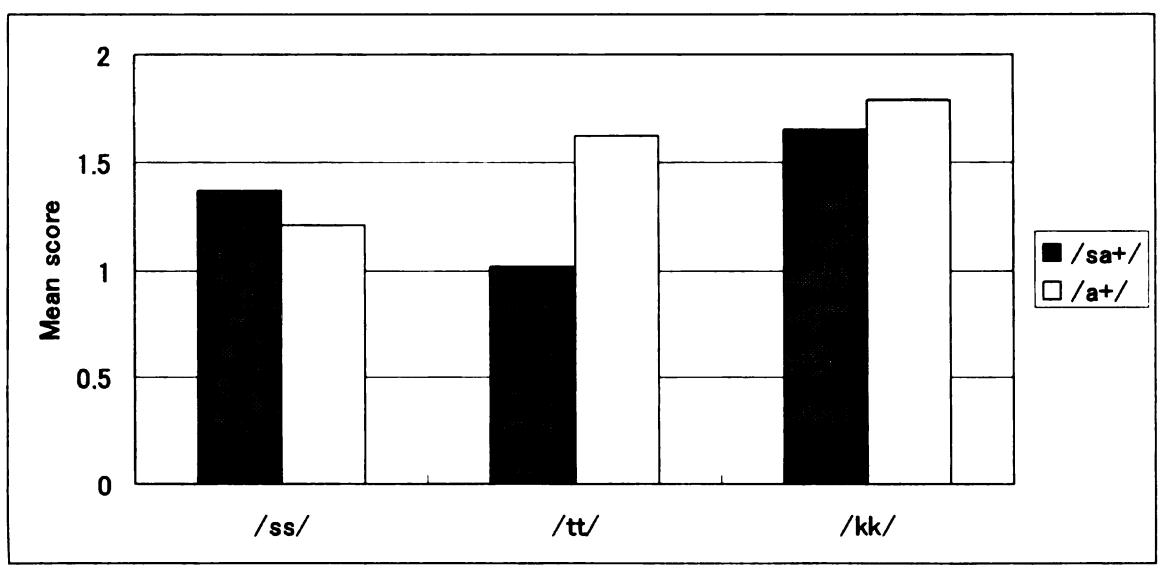
	/ss/ (*SD*)	/tt/ (*SD*)	/kk/ (*SD*)
/+a/	1.39 (.34)	1.18 (.32)	1.59 (.76)
/+u/	1.19 (.32)	1.5 (.43)	1.86 (.46)

Figure 5.3. Mean production scores by vowels following geminate consonants

Among consonants, pairwise comparison showed significant results; when the following vowel was /u/, the difficulty hierarchy was the same as that shown by the perception test

result; the /ss/ fricative consonant was the most difficult among the three, and the /kk/ stop consonant was the easiest. However, when the following vowel was /a/, the /tt/ + /a/ combination was the most difficult among the three consonants, and there was no significant difference between /ss/ + /a/ and /kk/ + /a/.

Next, Figure 5.4 showes a comparison made between the two different preceding segments: /sa/ and /a/. A two-factor repeated measures ANOVA was conducted. Variables were preceding segment(s) (/sa/, /a/) and consonant (/s/, /t/, /k/). Both showed significant main effects, F_{ps} (1, 278) = 8.035, p = .005; F_c (2, 278) = 16.227, p = .000. The interaction of Preceding Segment x Consonant was also significant, F (2, 278) = 14.245, p = .000.



	/ss/ (*SD*)	/tt/ (*SD*)	/kk/ (*SD*)
/sa+/	1.37 (.35)	1.01 (.59)	1.65 (.46)
/a+/	1.21 (.31)	1.62 (.35)	1.79 (.79)

Figure 5.4. Mean production scores by preceding segment types

The mean scores for the environment with consonant /s/ + vowel /a/ as preceding segments (1.34) were significantly lower than those with only the vowel /a/ (1.54). However, pairwise comparison showed that only for /tt/ geminate consonants was the difference between /sa/ and /a/ significant; there was no difference between /sa/ and /a/ with relation to the other two consonants. The cause of these results in production might be the judges', that is, the native Japanese speakers', perspective. Previous studies agree that the length of preceding vowels will affect native speakers' perception of a geminate consonant (Toda, 1998); if the duration of the preceding vowel is longer, the perceptual boundary dividing a geminate consonant and a singleton will become bigger; in other words, native speakers would tend to perceive the token as a geminate consonant; this difference is applicable to either fricative or stop consonants. In the present experiment, the learners' pronunciation of the fricative /s/ of /sa/ might be perceived as part of the duration of the following segment, making not just long vowels but sequences of continuant consonants plus short vowels another salient feature producing the NS perception of gemination.

5.3.2 Error patterns

The perception test in Experiment I revealed that many learners inaccurately perceived geminate consonants as long vowels. For comparison, the error patterns in the production of geminate consonants (/CC+u/) were also examined, as shown in Figure 5.5.

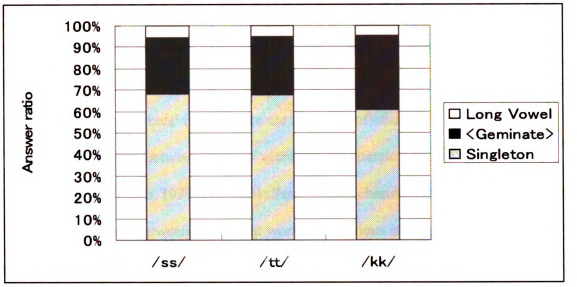


Figure 5.5. Ratio of errors: Production of geminate consonants (/CC+u/)

The misproduction of singletons as geminate consonants did not occur as frequently as in the perception test in Experiment I (Figure 3.9). Misproducing a geminate consonant as a long vowel was limited throughout the three consonant types, and there were no differences among them.

5.4 Discussion

In Experiment III, learners' production of geminate consonants was studied by having them read the test words in isolation and in frame sentences. The learners showed greater difficulty in producing geminate consonants than singletons. The majority of errors were made by misproducing a geminate as a singleton.

Producing geminate consonants at sentence-level showed higher correct score than producing them at word-level. The previous studies of L2 production of long vowels (Koguma, 2001) and geminate consonants (Hirata, 1990b) in Japanese demonstrated a

greater accuracy at word level, and the researchers suggested that this was because less attention might have been paid to the target tokens (e.g., long vowels, geminate consonants) in sentence-level production since the other elements in the sentences would have distracted the learners' focus. There are a number of other studies suggesting a significant correlation between learners' accuracy and attention in L2 acquisition (e.g., Dickerson & Dickerson, 1977; Lin, 2001; Sato, 1984; Schmidt, 1987; Tarone, 1982). One important difference between the current work and previous studies (e.g., Koguma, 2001; Hirata, 1990b) is the proficiency level of the participants. While the participants of these previous studies had experienced learning Japanese for some time, the participants of the present experiment were at the very beginning level. For such learners, reading sentences might have required much more attention, and reading words might have been much easier, requiring less attention, which might have been the cause of the higher accuracy level in sentence-level production in my data. It is assumed that learners' proficiency level might affect the direction of learners' attention in a reading task. On the other hand, the perception data in Experiment I showed an opposite tendency; sentence-level perception of geminate consonants were more difficult. Directing attention to items in larger contexts in perception might be more difficult than in production at the beginning level. Further research will be necessary to fully explain these results.

Comparison among the three consonant types revealed that producing /kk/ geminate consonants was significantly easier than producing the other two consonants, /ss/ and /tt/. /ss/ geminate consonants were the easiest, although there was no significant difference between /ss/ and /tt/. It seems that the result is comparable with the perception data to some extent, although any comparison of the perception and production data

should be made with caution because of the different subject groups.

Detailed examination showed that the phonetic environment affected learners' production. As in the perception experiment, /a/ and /u/ were chosen as the vowels following the geminates. Comparison between these two conditions revealed that the production of geminates + /u/ was easier than geminates + /a/. As noted in the previous chapters, the cause of difficulty in perceiving a geminate consonant + /u/ could be due to the lower sonority of the vowel, but such a sonority difference did not appear to affect the learners' production, at least not in the same way. This result adds to the literature demonstrating that perception does not necessarily parallel production in L2 phonological learning. With regard to consonant type, the /tt/ + /a/ combination was significantly more difficult than the other two consonants in production, and there was no difference between /ss/+/a/ and /kk/+/a/. This, of course, suggests an interaction between consonant and vowel identity and will require further study.

An analysis of the types of error patterns was also made. In the perception data of Experiment I, misperception of the stimuli including a geminate consonant as including long vowels was found most frequently in the /ss/ + /u/ stimuli. This result indicated that the learners could allocate the correct mora weight but had troubles with attribution of length in the stimuli. However, production data showed no such tendency for any of the three consonant types. The most frequent error pattern was misproducing the stimuli with geminate consonants as having singletons, so the learners had trouble with allocation of the length itself. The misperception of words with /ss/ geminates as having long vowels indicated that the learners were sensitive to mora weight. Since the production data did not indicate such a tendency, the influence of durational contrasts on

production seems to be a more difficult issue and will require further study. In Broselow and Park (1995), Korean learners at a certain stage of proficiency showed a tendency for mora conservation, which is a resistance to the loss of morae, by inserting an extra vowel in production. The production data in Experiment III showed that the learners had perhaps not reached that stage yet, while the learners in the perception study in Experiment I showed a tendency for mora conservation although they had troubles with the attribution of the mora.

CHAPTER 6

Experiment IV

6.1 Objectives of Experiment IV

Based on the results of Experiments I-III, Experiment IV was then conducted to test a training method to improve the ability of learners of Japanese to perceive geminate consonants. The training method was developed by extending the results of Experiment II, which examined the effectiveness of electronic visual input as part of a larger training program. In addition, based on the perception and production data from Experiments I and III, the training was developed in such a way as to permit investigation into whether it was specifically effective in the difficult phonetic contexts.

The main research question of Experiment IV was to the effectiveness of training using electronic visual input in the perception of geminate consonants; the effectiveness was assessed through a comparison of auditory-visual (AV) and auditory-only (A-only) training. The visual materials involved the speechwave displays of the stimuli created by Praat, which is the software that creates visualizations of speech signals and that was used in Experiment II. In addition, by conducting a production test as well as a perception test, the relationship between perception and production was also examined.

Previous studies demonstrated that perceptual training with visual input was effective for L2 learners (e.g., Hardison, 2004, 2005a), therefore, it was hypothesized that the training with speechwave displays as visual information would also be effective to train Japanes2 L2 learners. As reviewed in Chapter 2, to establish new L2 perceptual categories, learners must develop a new attentional weighting scheme for optimal

processing of the distinctive features of the L2 speech input. Experiment II demonstrated that the use of speech waveforms of geminate consonants as visual input might be effective for the development of such processing. In multiple-trace theory, all attended perceptual details of stimuli are stored as traces in memory (Hintzman, 1986). The information stored and gathered as a composite of such traces constitute episodic memory (Goldinger, 1997; Hardison, 2000, 2003; Hintzman, 1986). Training with visual input could direct learners' attention to the salient timing features (mora lengths) to be preserved in memory. Thus, another objective of Experiment IV was to account for the result of the AV training within the episodic model framework.

The experiment was carried out in a pretest-posttest design. The posttest was followed by a generalization test to examine whether the result of the training could be transferred to novel stimuli. Following the generalization test, a follow-up interview was conducted to collect qualitative data on the effectiveness and efficiency of the training from the learners' perspective. They were asked questions such as whether perceiving the difference between a singleton and a geminate consonant had been difficult before the training, and how difficult it became after the training; what they thought caused such difficulties in perception; whether they had any special strategy to identify a geminate consonant; whether there were any items that were particularly difficult to perceive; and whether they thought the training was effective. They were also asked to comment on the web-delivered training format.

6.2 Method

6.2.1 Participants

A total of 30 students were recruited from among the participants in Experiment III, the purpose of which was to examine their production of geminate consonants (two students from Experiment III did not participate). Those 30 participants were divided into two groups based on the results of the perception pretest, which will be explained later, so that the average test scores of both groups were comparable prior to training. In addition, 10 students participated in the study as a control group; they took only the pretest and posttest, but received no training.

The AV group received auditory and visual (electronic visual input) training and consisted of 15 students. The other 15 students formed the A-only training group, which received similar training (i.e., the same oral instructions and auditory input) but were not presented with visual information.

6.2.2 Pretest

6.2.2.1 Production test

The data from Experiment III were used as the pre-training production data.

6.2.2.2 Perception test

Before the training sessions started, the participants gathered and took a perception test in a classroom setting, which was the same perception test as was used in Experiment I. As explained below, the same stimuli, recording materials, and test procedures were used.

6.2.2.2.1 Materials

The recorded material was the same as in Experiment I. The stimuli were recorded by a female native speaker of Japanese (Tokyo dialect) using a SONY MD MZ-RH10-S with a microphone SONY ECM-CS10. Two tests were conducted: Test 1, the words in isolation; and Test 2, words in the frame sentences that follows:

watashi	*wa*	_____	*to*	*iimashita*
I	topic marker		that	said

'I said _____.'

kore	*wa*	_____	*desu*
this	topic marker		is

'This is _____.'

In each test (word-level, sentence-level), stimuli consisted of 30 bisyllabic Japanese real words and non-words, which included 12 singletons, /(C)V.CV/, and 12 geminate counterparts, /(C)V.C.CV/, where the first (C)V and the final CV were identical between each word in a minimal geminate-singleton pair, and 6 fillers which were bisyllabic words consisting of three morae, but including no geminate consonant. The types of test items are as follows:

Target geminate consonants: /ss/, /tt/, /kk/

The segments preceding the target consonant: /sa/, /a/

The segments following the target consonants: /a/, /u/

Table 6.1 shows some examples:

Table 6.1

Examples of the test items

Structure of items	Example	
	V_2=[u]	V_2=[a]
(C)V_1.tV_2 vs. (C)V_1t.tV_2	satu vs. sattu	sata vs. satta
(C)V_1.sV_2 vs. (C)V_1s.sV_2	sasu vs. sassu	sasa vs. sassa
(C)V_1.kV_2 vs. (C)V_1k.kV_2	saku vs. sakku	saka vs. sakka

The same set of 30 words was used for both tests, but the order of presentation was randomized. The maximum score for each test is 24, so the total score is 48.

6.2.2.2.2 Procedure

The stimuli were presented in a classroom setting on the same MD player as the one used for recording. A three-alternative forced choice identification task was used for the word-level and the sentence level tests. The learners were instructed to choose from one of three options given, which consisted of minimal triplets including a singleton consonant /(C)V.CV/, a geminate consonant /(C)V.C.CV/, and a long vowel /(C) V.V.CV/, where the first and final CV were identical between the members.

6.2.3 Perception training

6.2.3.1 Training materials

The basic structures of words and sentences used for the training followed the test materials. Although the stimuli of the perception test in Experiment I included only minimal pairs for singleton and geminate consonants, considering the result which

showed that many students inaccurately perceived geminate consonants as long vowels, it also appeared to be necessary for learners to be able to distinguish long vowels from geminate consonants. Therefore, the training words were divided into three categories, forming minimal triplets of bisyllabic words: (1) words containing singleton consonants, (2) words containing geminate consonants, and (3) words containing long vowels. The tokens were either in isolation or in such sentence frames as the following:

watashi	*wa*	_____	*to*	*iimashita*
I	topic marker		that	said

'I said _____.'

kore	*wa*	_____	*desu*
this	topic marker		is

'This is _____.'

As a form of training, web-delivered material was selected. Compared to laboratory training sessions, which have been traditionally used for L2 perception training, on-line training material is easier for learners to access, and it allows them to learn at their own pace. The on-line format might also reduce the learner's anxiety or stress. As described above, the learners were instructed to take 10 sessions for about 15-20 minutes each, but it was left to the learners to decide what time in the day and where to take the training. Thus, easy access and a less stressful environment are some of the advantages of web-delivered training. However, in exchange for such flexibility, there is a drawback for research purposes; without the presence of an observer, there is no way to watch and control the amount and frequency of training learners really take.

The training material was created by transforming the speechwave displays created by the Praat data by using Macromedia Flash MX 2004. Flash is a multimedia authoring program, which enables a user to create animation with sound in a very simple way. The stimuli created by Praat were converted to SWF files when saved in Flash. Using the SWF file, on the webpage, learners could listen to the sound and simultaneously view the waveform and the movement of the cursor through the waveform (see Figure 6.1).

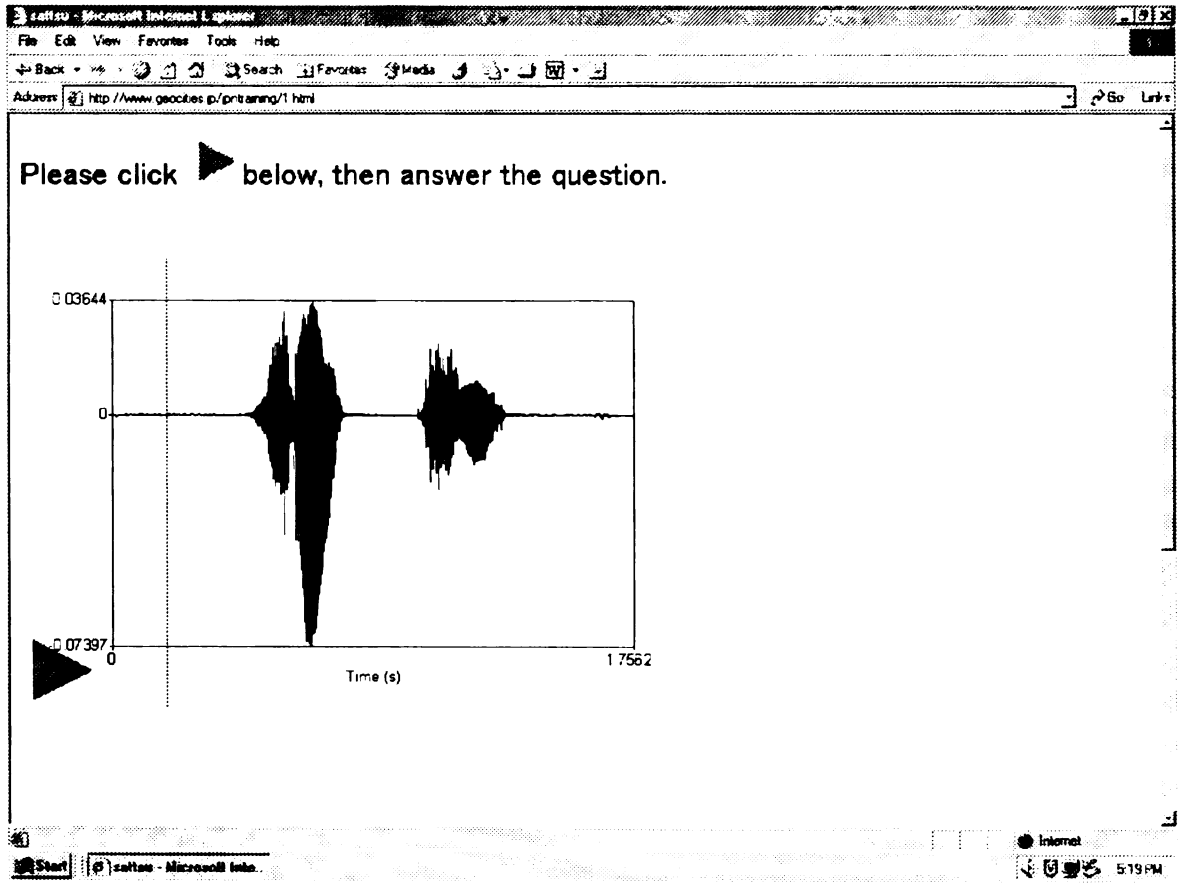


Figure 6.1. Waveform display

For the A-only group, this waveform display was not given; only the auditory stimuli were played. Below the waveform display, in a dialogue box that learners opened by moving the curser down, the learners were asked to identify the word they heard from three alternatives that appeared on the computer screen as in Figure 6.2. Such multiple choice exercises can be easily created using free downloadable e-learning tools which are abundant now on the web.[2]

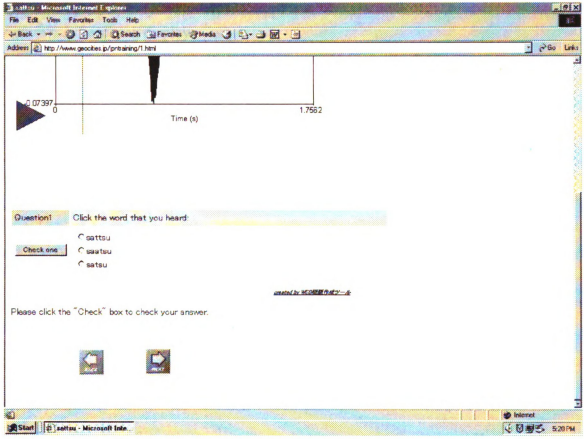


Figure 6.2. Exercise questions display

[2] The present study used "Web Quiz Creating Tool" provided at http://www.fureai.or.jp/~irie/webquiz

Immediately after responding by clicking an answer on the screen, they received feedback; if they chose an incorrect answer, "Wrong!" appeared, and they had a second trial. If they answered correctly, the screen showed "Correct!" as in Figure 6.3.

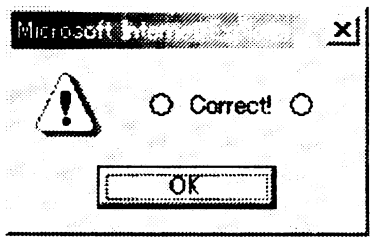


Figure 6.3. Feedback display

Either a word-level token or sentence-level token appeared randomly. The selection of the words and sentences were changed every day, while all types of stimuli were presented equally every day.

6.2.3.2 Training procedure

As described above, based on the result of the perception pretest, the 30 learners were divided into two groups: the AV group and A-only group. Before the training began, the two groups met separately for an instruction session about the experiment. To the AV-group, a lecture about geminate consonants and a demonstration of speech waveforms was given for about five minutes, similar to the one given in Experiment II, so that the learners could see how differently some acoustic features of geminates and singletons were realized graphically. The A-only group also received the same lecture about geminates, without the waveform demonstration. In this instruction session, a trial web-session was also given in order to teach the learners how to access the materials,

which they would have to do on their own, and in order to observe how many minutes it would take the learners to go through a training session.

The appropriate amount of training depends on the type of training used. In Logan, Lively, and Pisoni (1991), about 280 stimuli were presented to participants. Each session lasted 40 minutes, and a total of 15 sessions were undergone. This type of training represented relatively implicit learning, and that type of learning generally requires more training stimuli and sessions than training involving explicit instruction about the target. While Logan et al. (1991) and Yamada et al. (1995) conducted 8 sessions and 15 sessions, respectively, the present study provided 10 training sessions; this number chosen considering the explicit nature of the instruction. The amount of time for each session was about 15-20 minutes. This approximate time was averaged from among all the learners observed in the trial sessions. The participants were instructed to take 10 sessions, one session per day for two weeks, by the deadline date.

The learners began each session by accessing the researcher's website, where the training program was uploaded. They clicked the "start" button, and the stimuli played. For the AV group's training session, the sound (e.g., '*attsu*') and visual (e.g., the waveform of '*attsu*') stimuli appeared on the display as shown in Figure 6.1. They were also instructed to report any problems with the training session, and also to make a report if it took them too short a time (less than 10 minutes) or too long a time (more than 15 minutes) to complete the training session; however, no such report was brought throughout the training period.

6.2.4 Posttest

6.2.4.1 Perception test and production test

After all 10 perception training sessions, the learners gathered and took a posttest in a classroom setting. The test materials included the same set of stimuli as the pretest. In order to see the effect of perception training on production improvement, a production test was given before the perception test, in the same procedure as the pretest.

6.2.4.2 Generalization test

In order to examine whether the effect of training could be transferred to unfamiliar tokens pronounced by a different speaker, a generalization test in perception was also conducted, immediately following the posttest. The test consisted of words containing singletons, geminate consonants, and long vowels; the latter had not been included in the training stimuli. A new talker, who was also a female native speaker of the Tokyo dialect of Japanese, recorded the stimuli. A three-alternative forced choice task was used as in Experiment I. In addition to /u/ and /a/, /i/, /e/, and /o/, i.e., all the Japanese vowels were included following the consonants to examine whether effects of the training had been transferred to the untrained vowel contexts as well. These vowels can occur following any type of consonant. The stop consonant /p/ was also included, and long vowels were added to the singleton and geminate counterparts. A total of 24 tokens (12 tokens x 2 levels [word-level and sentence-level]) and six distracters were presented. Examples of words with geminate consonants are presented in Table 6.2. Singletons and long vowels with identical combinations of consonants and vowels were also included in the test as distracters.

Table 6.2

Examples of the geminate consonants

	Vowel type				
consonant	**/a/**	**/i/**	**/u/**	**/e/**	**/o/**
/s/	-	/sassi/	-	/sasse/	/sasso/
/p/	/sappa/	/sappi/	/sappu/	/sappe/	/sappo/
/t/	-	/satti/	-	/satte/	/satto/
/k/	-	/sakki/	-	/sakke/	/sakko/

6.2.4.3 Follow-up interview

To obtain insight into the training approach from the learners' point of view, a follow-up interview was conducted with each of the learners. The interview was carried out individually in the researcher's office for about ten minutes. The learners were asked questions such as whether they found it difficult to identify geminate consonants; what they thought the causes of such difficulty were; and how the difficulty changed after the training. They were also asked for overall comments on the training sessions: how they found the training; what effects they thought it brought; and when they used the training during a day.

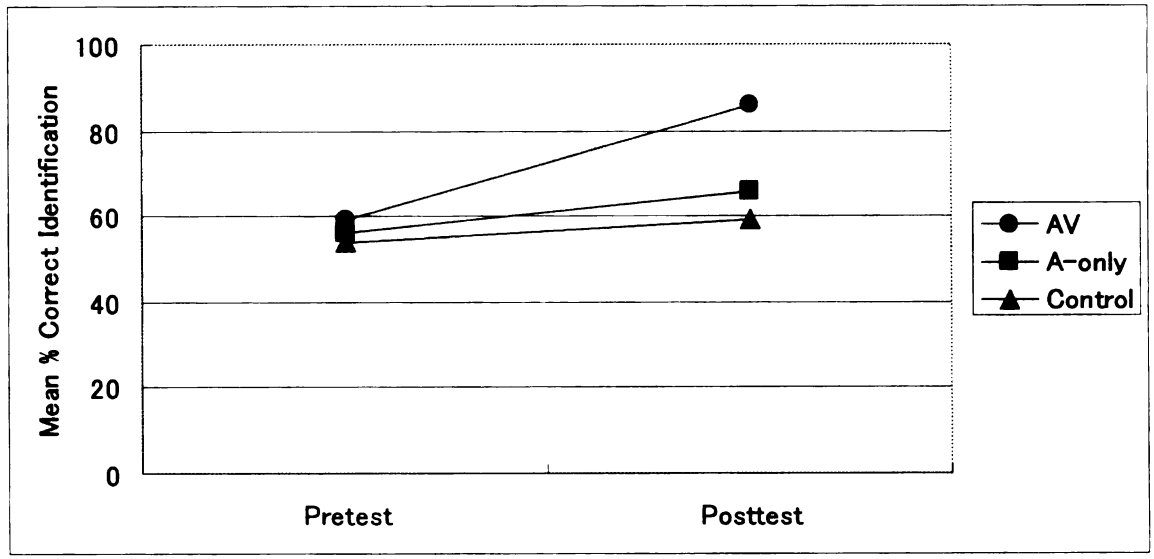
6.3 Results

The mean percent of correct answers for pretest and posttest was tabulated. The results are discussed in the following order: a) comparison of training types (AV, A-only, Control), b) a comparison of posttest and generalization tests between the training groups (AV, A-only), c) a comparison of word-level and sentence-level test between AV and

A-only groups, d) the effect of consonant types and the following vowels in the word-level test (AV group), e) the effect of consonant types and the following vowels in the sentence-level tests (AV group), and f) the error pattern of the AV group by consonant type. Finally, production test results (per group, and by consonant type), and a comparison of individual results within the AV group (perception and production in pretest and posttest) are presented.

6.3.1 Comparison of training types (AV, A-only, Control)

In order to examine the effects of training, both group (AV-group and A-only group) and time (pretest and posttest) were considered, as shown in Figure 6.4. A one-factor ANOVA indicated that the mean percent correct identification scores before training among the AV-group (59%), the A-only group (56%), and the Control group (54%) were comparable, $F(2, 538) = .612, p = .552$. Though the scores for both training groups showed an increase in the posttest, the score of the AV-group (86%) was higher than that of the A-only group (66%). A mixed design 2 x 3 ANOVA was used with time (pretest, posttest) as the within-subject variable and group (AV, A, Control) as the between-subject variable. The results showed significant main effects of time and group, $F_t(1, 538) = 135.985, p = .000$, $F_g(2, 538) = 8.211, p = .000$. The Time x Group interaction was also significant, $F(2, 538) = 847.750, p = .000$. The post-hoc test (Tukey's HSD) revealed that gains were significantly greater for the AV-group compared to the A-only and Control groups; the latter two groups showed no significant difference.



	Pretest (*M*, *SD*)	Posttest (*M*, *SD*)
AV	59% (14.26, 4.44)	86% (21, 2.49)
A-only	56% (13.44, 3.82)	66% (16, 3.21)
Control	54% (13, 3.74)	59% (14.13, 3.36)

(Test items N=24)

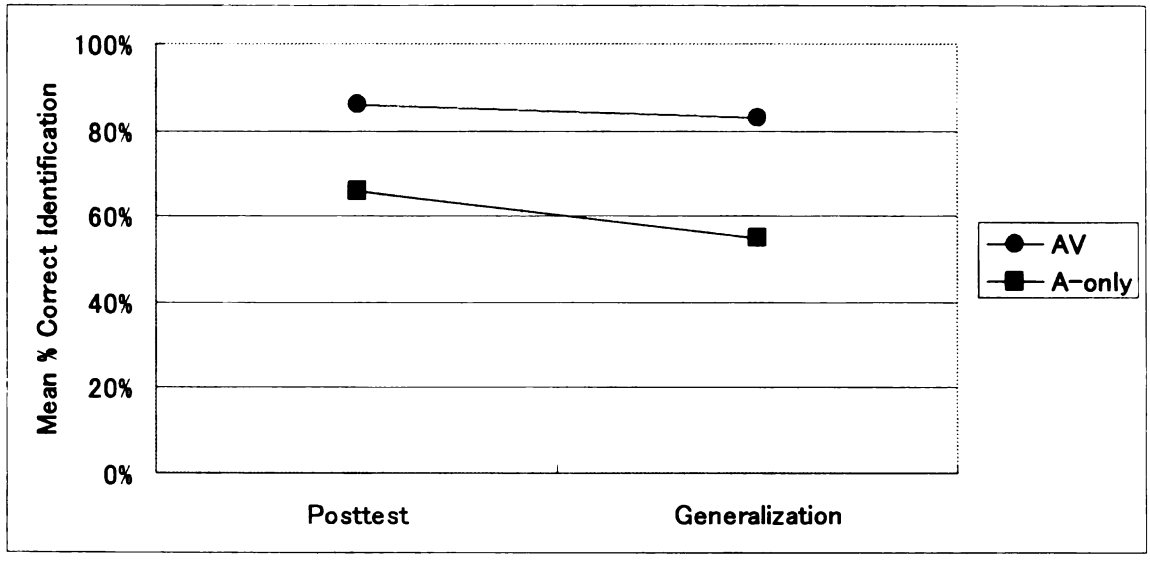
Figure 6.4. Mean percent correct identification for perception pretest and posttest

Thus, the greatest improvement was observed for the AV training group. This result is consistent with Hardison (2003), which compared AV-training and A-only training. In Hardison's study, talker's face with visible articulatory gesture was used as a visual cue to train Japanese and Korean learners to identify /r/ and /l/. The result showed a significant advantage for AV-training.

6.3.2 Results of generalization test

Next, the results of the generalization test were analyzed. This test included 24 tokens (12 tokens x 2 levels) involving the untrained geminate consonant /p/, and the bisyllabic words containing long vowels. Figure 6.5 shows the effectiveness of the

training on the students' ability to correctly apply the training to novel tokens.



	Posttest (*M*, *SD*)	Generalization(*M*, *SD*)
AV-Group	86% (21, 2.49)	83% (19.80, 2.96)
A-only Group	66% (16, 3.21)	55% (13.3, 3.22)

(Test items N=24)

Figure 6.5. Mean percent correct identification for posttest and generalization test

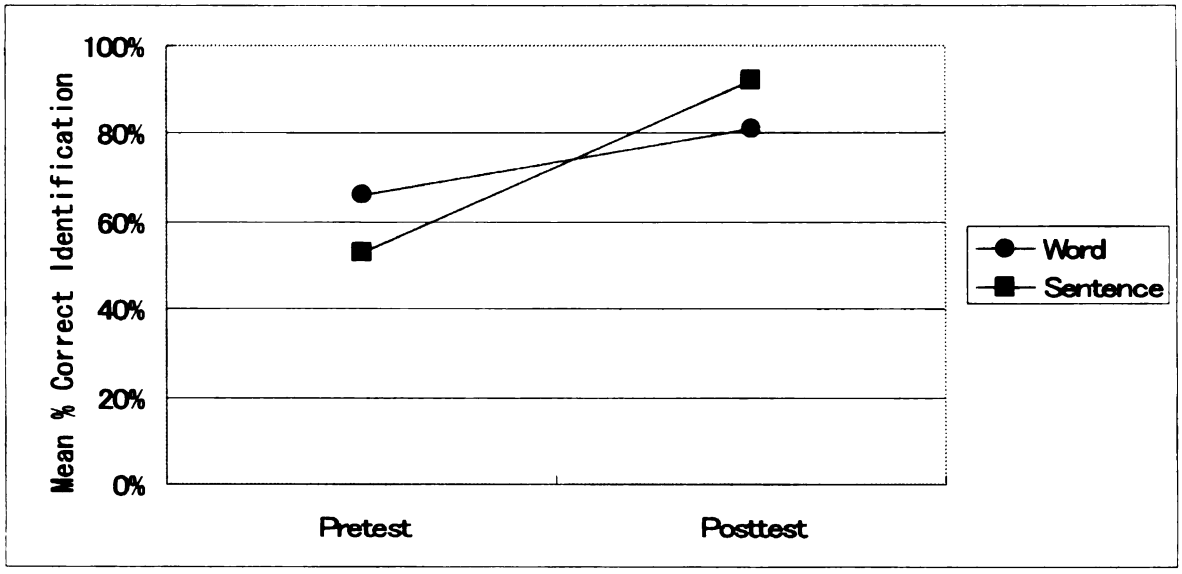
A mixed design 2 x 2 ANOVA was conducted. Variables were group (AV, A-only) and test (posttest, generalization test). Both had significant main effects, F_g (1, 358) = 13.377, p = .000; F_t (1, 358) = 82.071, p = .000. The interaction of Group x Test was also significant, F (1, 358) = 13.385, p = .000. The decrease in accuracy between the two tests was greater for the A-only group. This result indicated that the effect of AV training could be transferred to unfamiliar tokens and a new voice.

6.3.3 Results of the AV-group

6.3.3.1 Word level vs. Sentence level

Since the interest for the present study is in examining the effect of AV training in

detail, further analyses were conducted on the results of the AV-group. First, we have already observed that there was a significant improvement from the pretest to the posttest, but the result was further analyzed at the two different levels of the test, word vs. sentence, shown as in Figure 6.6. Recall that the results of Experiment I showed that the learners found geminates more difficult to identify at the sentence versus word level. Therefore, it was considered important to see whether there was any difference in improvement between these two levels after training.



	Pretest	Posttest	Gains
Word level	66%	81%	15%
Sentence level	53%	92%	39%

Figure 6.6. Mean percent correct identification in pretest and posttest for word-level and sentence-level geminate consonants

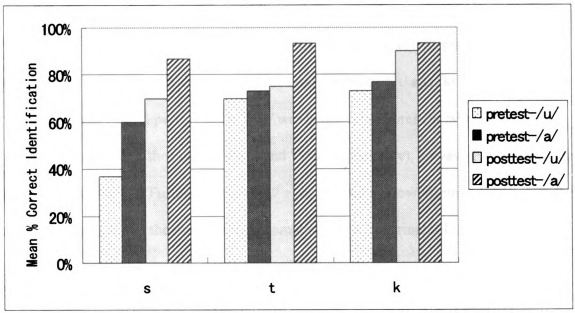
First, comparison was made between word-level and sentence-level perception in the pretest. The result showed that sentence-level (53%) was significantly more difficult than

word-level (66%); $df = 11$, $t = 2.277$, $p = .044$.

Next, a two-factor repeated measures ANOVA was conducted on the AV group data. Variables were time (pretest, posttest) and test (word level, sentence level). The effect of test was not significant, but there was a significant main effect of time, $F(1, 14) = 17.742$, $p = 0.001$, and the interaction of Time x Test was also significant, $F(1, 14) = 14.253$, $p = 0.002$. The interaction indicated improvement over time differed significantly between word-level and sentence-level perception. As Figure 6.6 indicates, the gain in scores as a result of the training differs in that the sentence-level test showed gains of more than twice those of the test using word-level stimuli. The effectiveness of AV-training is more pronounced in sentence-level perception, which the learners found more difficult in the pretest.

6.3.3.2 Word-level perception

Further analysis focused on the word-level and sentence-level data from the AV group. Figure 6.7 shows the data of word-level perception. A comparison was made using the identification accuracy scores in the pretest and posttest for the three different consonants, /s/, /t/, and /k/, with the two different post-consonantal vowels, /a/ and /u/.



	/s/		/t/		/k/	
	/u/	/a/	/u/	/a/	/u/	/a/
Pretest	37%	60%	70%	73%	73%	77%
Posttest	70%	87%	75%	93%	90%	93%

Figure 6.7. Mean percent correct identification for each consonant (/s/, /t/, /k/) and each post-consonantal vowel (/a/, /u/) in pretest and posttest for word-level geminate consonants

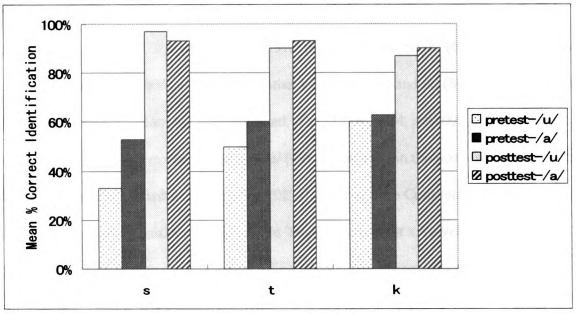
To assess the effects of training, a three-factor repeated measures ANOVA was conducted. The variables were time (pretest, posttest), consonant (/s/, /t/, /k/), and vowel (/a/, /u/). All had significant main effects, F_t (1, 29) = 41.308, p = .000; F_c (2, 58) = 22.929, p = .000 and F_v (1, 29) = 33.305, p = .000. The means for /s/ was significantly lower than that for the other two consonants, while there was no significant difference between /t/ and /k/.

Although Time x Vowel interaction was not significant, the Time x Consonant interaction was significant, F(2, 58) = 10.119, p = .000, indicating the increase in scores

varied significantly by consonant. In addition to the interaction between Consonant and Vowel, F (2, 58) = 18.125, p = .000, the Time x Consonant x Vowel interaction was significant, F_t (2, 58) = 5.236, p = .008. The improvement of /t/ and /k/ was almost parallel, though overall performance of /k/ was better. The scores for /s/ were the lowest in both the pretest and the posttest (48% and 79%, respectively), however, /s/ showed the greatest improvement. Further, although /s/ + /u/ showed the lowest mean accuracy in the pretest, and was found the most difficult combination in Experiment I, it showed the greatest improvement — a percentage rise of 34%.

6.3.3.3 Sentence-level perception

Separate analyses were conducted on the sentence-level data. Figure 6.8 shows the sentence-level perception accuracy.



	/s/		/t/		/k/	
	/u/	/a/	/u/	/a/	/u/	/a/
Pretest	33%	53%	50%	60%	63%	63%
Posttest	96%	93%	90%	93%	87%	90%

Figure 6.8. Mean percent correct identification for each consonant (/s/, /t/, /k/) and each post-consonantal vowel (/a/, /u/) in pretest and posttest for sentence-level geminate consonants.

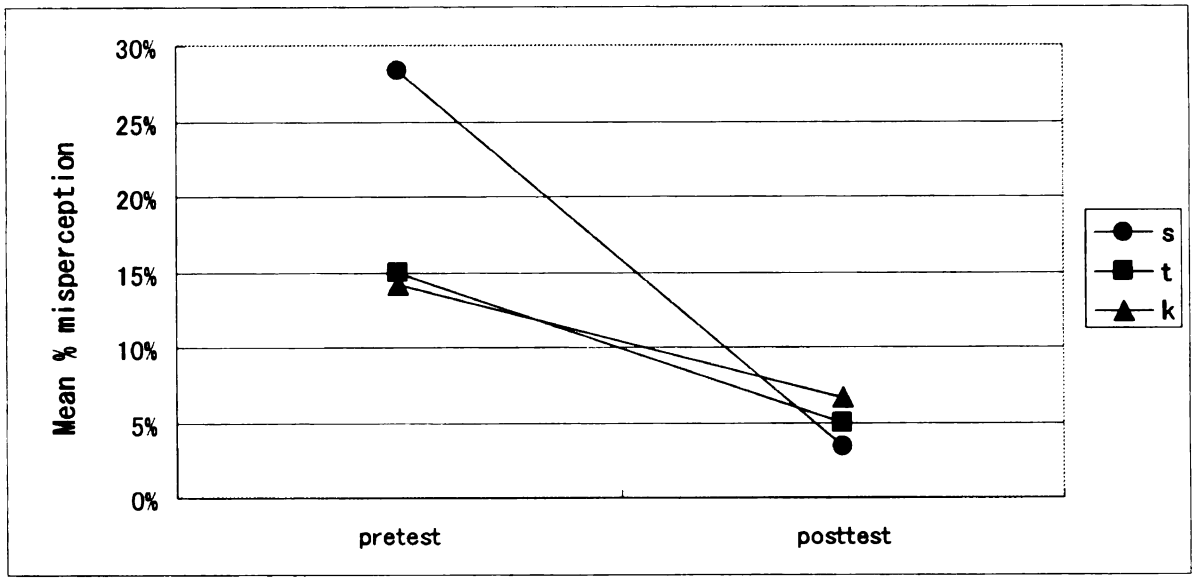
A three-factor repeated-measures ANOVA was conducted on the results of the sentence-level test, with time (pretest, posttest), consonant (/s/, /t/, /k/), and vowel (/a/, /u/) as variables. Again, all variables had significant main effects, F_t (1, 29) = 65.862, p = .000; F_c (2, 58) = 5.342, p = .007 and F_v (1, 29) = 20.714, p = .000. There was also no significant difference in the means between /t/ and /k/, while that for /s/ was significantly lower than the other two consonants.

A significant Time x Consonant interaction was found, F (2, 58) = 45.877, p

= .000, again indicating the increase in scores varied significantly by consonant. The pretest score indicated /s/ was the lower (43%) than /t/ (55%) and /k/ (63%); however, the posttest score for /s/ (95%) exceeded the other two, /t/ (92%) and /k/ (89%). A significant Time x Vowel interaction was also observed, $F(1, 29) = 7.864$, $p = .009$; in the pretest, the score for /a/ (59%) exceeded that for /u/ (49%), however, in the posttest the score for /u/ (91%) was almost tied with that of /a/ (92%). Although the Consonant x Vowel interaction was not significant, a significant Time x Consonant x Vowel interaction was observed. The improvement in the perception of geminate consonants with /s/ + /u/ combination was the largest; in the pretest, the score was the lowest among the three consonants (33%), but at posttest, it was the highest (96%). The effect of the training at the sentence level was most evident for /s/ + /u/, which was the most difficult consonant-vowel sequence in Experiment I. Taken together with the results at word level, these findings suggest that the AV-training contributed the most to improve perception of the most difficult condition for the learners. This result is consistent with previous training study (Hardison, 2003). In Hardison's study on Korean and Japanese learners of English, visual input contributed the most to the perception of the /r/-/l/ contrast sounds in the more difficult environments, which are utterance-initial for the Japanese subjects and syllable-final for the Koreans. These difficulties were based on pretest identification accuracy and were predicted by their L1 phonology. Considering the observation in 6.3.3.1, which showed that the sentence-level accuracy in the posttest exceeded the word-level accuracy, it is also suggested by the present study that perceptual training with visual information was particularly effective in difficult environments.

6.3.4 Error pattern (AV group)

In Experiment I, it was found that the learners tended to perceive words containing a geminate consonant as containing a long vowel, and this tendency was especially strong in the perception of the fricative geminate consonant /s/. To assess the effect of the training in reducing such errors, the error rate of inaccurate perceptions of words containing geminate consonants as containing long vowels was analyzed with an ANOVA. The variables were time and consonant. Main effects were found for both time, $F(1, 119) = 28.876$, $p = .000$, and consonant, $F(2, 238) = 9.318$, $p = .000$, and the interaction of Time x Consonant was also significant, $F(2,238) = 21.591$, $p = .000$. Tukey's HSD reveals that in the pretest, the error rate of /s/ was significantly higher than that for /t/ and /k/ while these latter two were not statistically different. However, in the posttest, all three consonants showed no significant difference. The result of comparison is shown in Figure 6.9.



	/s/	/t/	/k/
pretest	28%	15%	14%
posttest	3%	5%	7%

Figure 6.9. Mean percent misperception of geminate consonants as long vowels for each consonant type in pretest and posttest

As Experiment I and the pretest of Experiment IV indicated, the fricative geminate consonant was perceived less correctly than the other two stop geminate consonants, but the fricative geminate consonant showed the highest improvement rate of the three. In sum, the training contributed the most to improvement in the perception of /s/ geminate consonants.

6.3.5 Production test

Experiment IV also aimed to examine whether production would be improved as a result of perceptual training. The Japanese judges' pretest and posttest scores were totaled, and the means were calculated for each learner in the training groups and the

control group.

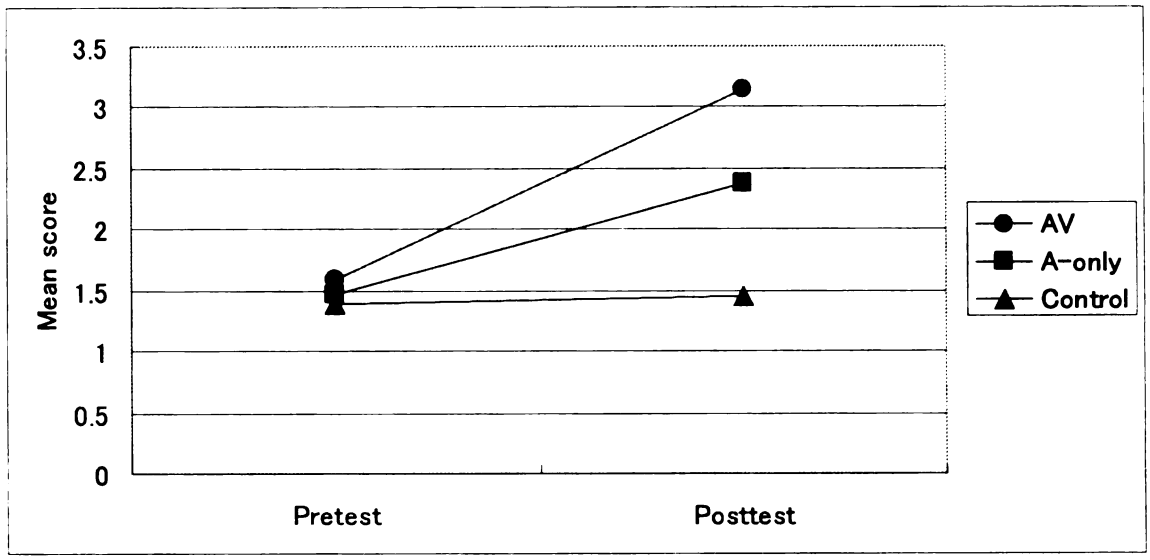
6.3.5.1 Judgment of production

The contents of the cassette tapes submitted by the learners were transformed into WAV format. The data from all 40 learners were stored on a CD-ROM, and copies were given to the five native Japanese raters for the judgment task in the language lab. They were provided with a headset in a quiet room and listened to all tokens produced by each learner. Their task was to write down the words they thought they had just heard using standard Japanese orthography. The stimuli provided to the learners were blocked from the presentation to the judges for their ratings.

6.3.5.2 Results

Correctly identified items were tabulated for each item and for each subject. If a judge correctly recognized the token the subject pronounced, 1 point was given; therefore, if all five judges made correct judgments of the token, a maximum score of five was possible. For all ratings by the five judges, inter-rater reliability was determined (using Pearson correlation) and ranged from .72 to .96 by category of token (consonant /s/, /t/, /k/; pre-consonantal segment /sa/ and /a/; post-consonantal vowel /u/, /a/). The highest coefficient was found in judgment of /a/ + /kk/ + /a/, while the lowest was /sa/ +/tt/ + /u/, which was also found to be relatively difficult to perceive in the production data of Experiment III.

Figure 6.10 shows the result of the production tests of the three groups.



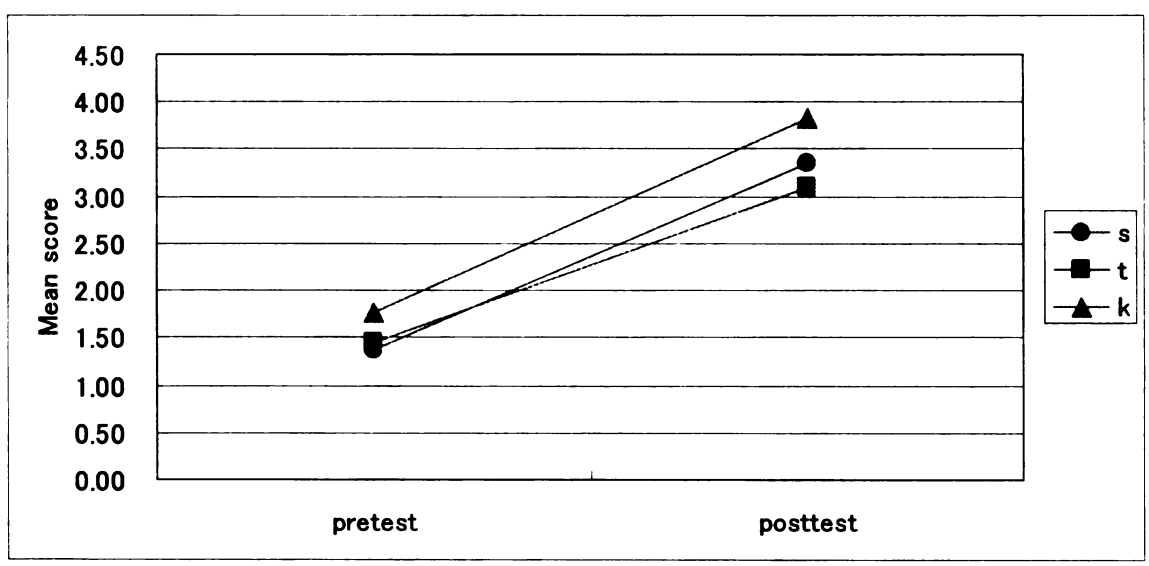
	Pretest (*SD*)	Posttest (*SD*)
AV Group	1.59 (1.04)	3.14 (.70)
A-only Group	1.46 (.72)	2.38 (.68)
Control	1.39 (.61)	1.45 (.49)

Figure 6.10. Mean ratings for the production test

A mixed design 2 x 3 ANOVA was conducted with time (pretest, posttest) as the within-subject variable, and with group (AV, A-only, Control) as the between-subject variable. The results showed significant main effects of time and group, F_t (1, 219) = 21.473, p = .000; F_g (2, 219) = 18.974, p = .000; and a significant Time x Group interaction, F (2, 219) = 8.837, p = .000. Tukey's HSD revealed that scores for the three groups, as shown in Figure 6.10, were significantly different. While the mean scores before training between the AV-group (1.69), A-only group (1.46), and Control group (1.39) were comparable, which was statistically confirmed by a one-factor ANOVA, F (2, 219) = 1.921, p = .149, the perceptual training improved production for the AV and A-only groups to a significantly different degree. The mean posttest score for the A-only

training group (2.38) was significantly lower than that for the AV group (3.14). Again, AV training contributed to greater improvement.

Further analysis of the AV-group data was done to examine how production of each consonant type was improved. For analysis, a repeated-measures ANOVA was used. The variables were time and consonant. The main effects of these two variables, F_t (1, 74) = 151.038, p = .000, F_c (2, 148) = 26.342, p = .000, and the interaction of Time x Consonant, F (2, 148) = 3.757, p = 0.026 were significant. Improvement over time in the production scores differed significantly according to the consonant as shown in Figure 6.11.



	/ss/ (*SD*)	/tt/ (*SD*)	/kk/ (*SD*)
pretest	1.36 (1.24)	1.44 (1.13)	1.76 (.98)
posttest	3.34 (.74)	3.08 (.57)	3.82 (.81)
gain	1.98	1.64	2.06

Figure 6.11. Mean production rating for each consonant type in pretest and posttest for the AV training group.

Although the three consonant types showed significant improvement in the posttest, Tukey's HSD showed that improvement of /tt/ geminate consonants was significantly lower, while /kk/ was the highest. /ss/ was significantly lower than /kk/, but the actual gain of /ss/ (1.98) was close to that of /kk/ (2.06). As we have seen, the greatest perception improvement was also found for /ss/. This result suggests a close link between perception and production.

6.3.6 Individual development for the AV training group

Tables 6.3 and 6.4 show individual development in perception and production respectively, by comparing percentages between pretest and posttest scores. The improvement rate was obtained by dividing the difference between the pretest score and the posttest score by the pretest score.

In the perception data, the learners whose score was higher than average (62%) in the pretest showed little improvement (Participants 3, 8, 10). On the other hand, several of the learners whose score was lower than average in the pretest showed a greater than average improvement rate (Participants 2, 4, 12, 13, 14).

Table 6.3

Mean percentage of perception accuracy in pretest and posttest, and improvement rate for participants in the AV-group

Participants	Pretest (%)	Posttest (%)	Improvement rate (%)
1	59	86	46
2	50	86	72
3	79	85	9
4	41	88	115
5	64	75	17
6	50	68	36
7	73	92	26
8	90	96	7
9	71	77	8
10	86	92	7
11	59	63	7
12	27	92	241
13	50	96	92
14	32	79	147
15	64	96	50
Average	62	84	35

In the production test of Table 6.4, a similar pattern was found.

Table 6.4

Mean percentage of production accuracy in pretest and posttest, and improvement rate for participants in the AV-group

Participants	Pretest (%)	Posttest (%)	Improvement rate (%)
1	12	43	258
2	3	52	1617
3	74	81	9
4	23	71	210
5	23	52	125
6	32	40	25
7	17	74	334
8	58	64	11
9	19	79	315
10	43	65	51
11	58	75	29
12	3	51	1583
13	33	71	115
14	21	84	298
15	42	71	69
Average	31	64	106

Participants 3, 8, 11 showed higher scores in the pretest than the average, (31%), but a small improvement rate. The learners whose scores were lower than the average in the pretests showed considerable improvement (Participants 1, 2, 4, 5, 7, 9, 12, 14). This pattern was observed more often and the improvement rate was much greater in production than perception.

The relation between perception and production was compared with Pearson's correlation, and a significant correlation was observed between the rates of perception and production improvement (r=0.636, p=0.015). However, we observed that considerable individual variation was found in the degree of perception and production

improvement.

Two categories of improvement were observed: 1) Learners who showed perception precedence (high score in the perception pretest, but not in production (Participants 5, 7, and 9)), and 2) Learners who showed balanced abilities and improvement (Participants 4, 12, and 14). The latter group started lower than average in both perception and production, but improved substantially in both domains. There were no learners who showed a production precedence pattern (high score in production pretest, but not in perception). Generally the perception test scores were higher than production ones both in pretest and posttest. This result is consistent with the prediction SLM makes; production will follow perception, although this is not always the case as discussed in Chapter 2. In addition, the observed individual variation of development showed that these two skills did not always develop in parallel, in partial contradiction to the prediction of SLM. Production improvement with perception training in the absence of explicit production training supports a close link between perception and production, and is consistent with other studies (Bradlow et al., 1997, 1999; Hardison, 2003).

6.3.7 Follow-up interview responses

At the conclusion of the training sessions, interviews were conducted with learners to assess the training from their point of view. First, 9 out of 15 learners said that fricative consonants were the most difficult to identify as geminates. Learners were aware of the difficulties of distinguishing fricative geminate consonants from singletons.

All the learners mentioned their greater awareness of the difference between singletons and geminate consonants as a result of the training sessions. Some learners

said that if they had not participated in this training they probably would not have known that 'sasu' and 'sassu' are pronounced differently and might have thought these were the same word.

Thirteen students said that they did the training with their own computer at home at night, between dinner and sleep, and said that they liked having the choice to do the training either at home or at school. Eleven learners responded that the training with the visual information was helpful. One learner explicitly stated, "the picture of the sounds gave me an idea of what I should try to listen to." Along with the learners' comments on the effectiveness of the training and the results of the posttest, their remarks indicated that electronic visual input was helpful to draw the learners' attention to the durational contrasts used in the training. Attention shifts are important theoretically and help to account for the fact that the most problematic sound for perception (i.e., /s/) showed the greatest improvement.

Three students said they would become better if they had more training, having realized that they still had not achieved native-like proficiency. At the same time, these three students (Participants 4, 12, 13) had actually shown substantial improvement. Considering the fact that all learners had volunteered for the training, it can be said that they were very motivated to improve their Japanese; however, these three students were sufficiently motivated to express willingness to participate in additional training sessions if they were offered. This may indicate a correlation between learners' motivation and development, as discussed in the literature.

6.4 Discussion

the results of Experiment IV showed significant improvement in perception accuracy with AV training using electronic visual input. In addition, this improvement generalized to novel stimuli and a new talker. The A-only group showed less improvement from the pretest to the posttest than the AV-group. This result, showing the effectiveness of visual information in improving L2 phonology, replicated the results of several previous studies (e.g., de Bot, 1983; Hardison, 2003, 2004).

The detailed examination of the data of the AV-group revealed that perception of geminate consonants improved under various phonetic contexts. The training was effective in improvement of the perception of all three consonant types at both the word- and sentence-level. More importantly, the data showed a greater improvement under the conditions which were most difficult in the pre-training perception of geminates, which were examined in Experiment I; learners showed the most difficulty in perceiving /ss/ geminates, especially followed by the lower-sonority vowel /u/. This combination, /ss/ + /u/, showed the lowest score in the pretest, but a much greater improvement was noted for this condition in the posttest. Further, the perception data in Experiment I showed that the learners had more difficulty in sentence-level than word-level perception, but the score for sentence-level perception exceeded that for word-level perception in the posttest. The results indicated that the AV-training was especially effective in improvement of perception of the geminate consonants under the conditions that are the most difficult for the learners.

Hardison (2003) also reported a similar finding. In her study of /r/-/l/ contrasts in English, visual input contributed the most to the perception of the sounds in the more

difficult environments, which were utterance-initial for the Japanese subjects, and syllable-final for the Koreans. These difficulty ratings were based on pretest identification accuracy and predictions from analyses of their L1 phonology. The present study can be added to the evidence that shows that training with visual input is effective, especially with sounds that are difficult for learners.

Another important finding is that the scores for production intelligibility also improved from pretest to posttest. Since no explicit production training was provided, this improvement is most likely due to the perception training. Both the AV-group and the A-only group showed significant improvement in production. The improvement in production for the training groups was consistent with previous findings (e.g., Akahane-Yamada, 1999; Bradlow, Pisoni, Akahane-Yamada & Tohkura, 1997; Hardison, 2003; Rochet, 1995). The results of these studies indicated that training in speech perception could transfer to improvement in the production domain. However, the mean score of the AV-group in the present experiment was significantly better than that of the A-only group. Thus, we could observe a greater effectiveness of perception training with electronic visual input on this transfer effect. It was also important that the improvement of /s/ was very high; almost tied to /k/, which was the highest. Since the greatest perception improvement was also found for /s/, a close perception-production link was indicated.

As expected from the results of previous studies, the result of this experiment also indicated a close connection between the speech perception and production domains, in this case, in the process of the acquisition of Japanese geminates by English speakers.

As hypothesized, the result of the training was compatible with the learning

process captured by an episodic model framework. As reviewed in Chapter 2, multiple trace theory assumes that all attended perceptual details of events are stored as traces in memory (Hardison, 2000, 2003; Hintzman, 1986). In Experiment IV, the visual input of speech waveforms of geminate consonants succeeded in drawing learners' attention, and the repetition of the stimuli in various phonetic contexts encountered through the two-week training sessions formed the bases of the episodic memory (Hintzman, 1986). The attended feature was mora weight of singletons and geminate consonants, which was presented to the learners through the speech waveforms, and as a result, a weighting scheme was developed to accommodate detection of such features in Japanese. The result of the generalization test also indicated that learning had taken place in terms of the formation of more clearly defined memory traces to which new input could be matched for identification.

In sum, the results of the present experiments indicate that 1) computer-based perceptual training with visual information (i.e., waveforms of geminate consonants) showed a greater advantage over auditory-only training, 2) adult learners can be trained in the perception of geminate consonants in Japanese, and 3) this ability is closely related to their L2 speech production. A further pedagogical implication is that web-delivered perceptual training, as examined in the present experiment, can be practically incorporated into classroom teaching.

CHAPTER 7

CONCLUSION

7.1 Summary of findings

The present study has reported on data from the perception and production of geminate consonants in Japanese by American learners. Based on these data, an effective way to train learners to improve their perception and production was developed and tested. The following is a summary of the findings with reference to the research questions proposed for this study.

The first research question addressed the issue of the way that the learners perceive and produce geminate consonants. The question of whether there any particular phonetic contexts and identities of geminate consonant which make perception and/or production more difficult for learners was investigated. The findings of Experiments I and III, both in the perception and production data, showed that the learners were affected by the phonetic contexts and identities of geminate consonants, and the degree of difficulty in perception and production varied depending on the types of consonants and the phonetic environments.

The second research question investigated whether audio-visual training would be more beneficial than auditory-only training by comparing audio-visual training using visual display of waveform of geminate consonants to audio-only training. Experiments II and IV conducted a comparison of pretest and posttest scores, and the result indicated that the learners could be trained to improve perception by means of a training program with electronic visual input as well as auditory-only training; however, the effect of

AV-training showed its superiority in producing such improvement over the A-only training.

The third research question examined the relationship between perception and production development. It was investigated whether perceptual training would improve production ability without any explicit production training. In Experiment IV, the learners did not receive any explicit production training; however, the learners improved in production ability. Overall, the AV-training group data showed a significant correlation in their improvement of perception and in their improvement of production. However, data from individuals revealed that learning rates varied considerably among the learners. Some learners showed greater development in perception and others showed balanced improvement, in which perception and production improved equally. This result indicates that these two skills do not always develop in parallel, but we can still conclude that there is a close link between perception and production development processes.

7.2. General discussion

The perception results of Experiment I showed that the learners had difficulties in perceiving geminate consonants, particularly in certain phonetic contexts: geminate consonants /CC/ + high vowel /u/ were more difficult to perceive than geminate consonants /CC/ + low vowel /a/. It is suggested that the lower sonority of /u/ made perception of the geminate more difficult since the difference between a consonant (the geminate) and a more sonorous environment would be more salient, highlighting the boundary between the consonant and the vowel and thus making the precise duration of the consonant easier to perceive. On the other hand, the preceding segment, either a

consonant + vowel or a vowel alone, did not affect perception of geminate consonants. It was also revealed that the learners found a geminate consonant read in a sentence frame more difficult to perceive than one read in isolation. This result indicated that it was necessary to include sentence-level stimuli in both research and training, while many of the previous studies of geminate consonants in Japanese used only words in isolation (e.g., Enomoto, 1992; Fukui, 1978; Minagawa & Kiritani, 1996).

The types of consonants also had an effect: fricatives (/ss/) were more difficult to perceive as geminates than stops (/kk/ and /tt/). Further, there were many learners who misperceived fricative geminate consonants as long vowels. Previous studies indicated only that the learners' errors in perceiving geminate consonants were failure to allocate mora weight correctly, taking geminates for single consonants. However, the present results indicated a possibility that even the learners who made errors can at least perceive the correct number of morae. In other words, they could allocate the correct mora weight, but had trouble with attributing it to the correct segment.

In Experiment II, an alternative to traditional laboratory auditory training as a way to train learners was tested. Based on the perception data in Experiment I, electronic visual input which displayed the frication noise of fricative geminate consonants visually on a computer screen was tested to examine whether it could improve the learners' perception of geminate consonants. Many studies have been conducted to examine if visual feedback was effective in improving production of new L2 phones at both word and sentence levels. The studies showed that such feedback was successful in improving the learners' production. Experiment II, however, focused on perceptual training by displaying waveforms, and the result indicated a significant effect for such visual input in

improving learners' perception.

Experiment III showed that the learners' production was also affected by some phonetic conditions. The type of the preceding segment brought about a significant difference; when a consonant + vowel (/sa/) came before a geminate consonant, the learners had more difficulty in production than when only a vowel (/a/) preceded. The effect of sonority was not observed in production. As for the consonant types, production of /kk/ was significantly easier than that of the other two consonants (/ss/ and /tt/), which seemed to be equally difficult. Comparison between word-level and sentence-level production showed slightly better accuracy for the latter. For the learners in the present study, who were at the very beginning levels of learning Japanese, reading sentences might be more difficult and have required much more attention, which might have led to a better score.

Experiment IV, which used display of waveforms as visual input, involved a two-week training program. An AV-group was given instruction on geminate consonants with accompanying waveform displays, and an A-only group was given only auditory training in the pretest and posttest design. The result indicated that both training groups showed improvement, but the AV-group's improvement was significantly greater than that of the A-only group; this result could be attributed to the electronic visual input.

The detailed analysis of the AV-group data revealed that the training had an effect on the learners' ability to perceive geminate consonants under the conditions which were found to be specially in the perception data of Experiment I. Before the training, the lower accuracy scores indicated that the learners had particular problems with perceiving geminate consonants in frame sentences, geminate fricatives, and geminates before /u/.

However, the posttest score showed that perception of these geminates improved even more dramatically than the others. In addition, the effect of training seems to have been generalized to tokens with combinations of new consonants and vowels spoken by an unfamiliar voice which the learners had not encountered in the training sessions.

Another important finding of the present study is that the effect of the perceptual training transferred to improvement in production. No explicit production training was given to the learners, but the production test indicated a significant improvement after the training. In the pretest, the learners apparently had difficulty with timing control since most of the errors were made by misallocating mora weight and misperceiving a geminate consonant as a singleton. After the training, such errors were reduced. The learners were able to produce durational contrasts to some extent. Some previous studies also found such transfer effects of perception training on production, for example, the /r/-/l/ contrast in English by Japanese speakers (Bradlow et al., 1997) and the VOT of the /bu/-/pu/ contrast in French by Chinese speakers (Rochet, 1995). Hardison (2003) also reported that such an effect was found in the training of /r/-/l/ contrasts with AV input including a talker's facial cues. The present study revealed that a transfer effect could also be found in the acquisition of durational differences, which no previous studies had referred to. Together with other previous studies, we can conclude that development of phonetic categories is achieved through intricate communication between the perception domain and the production domain.

The results of our training studies are compatible with the view that developmental change and associated perceptual reorganization in speech perception is primarily due to modification of selective attention to linguistically related mechanisms.

The importance of attentional process in identification, categorization, and discrimination of speech signals has been acknowledged by many researchers in first and second language acquisition. According to Jusczyk's (1993) model of the development process of infant speech perception, infants have an innate auditory analyzer which provides a preliminary description of the spectral and temporal features present in the acoustic signal. The output of the auditory analyzers is weighted to give prominence to the most critical distinctive features. This weighting scheme is a way of directing attention to critical features for recognizing and distinguishing words in a particular native language. L2 learners must reorganize the weighting scheme which is used for the native language so that a new weighting scheme optimal for L2 can be developed. They must learn to alter the focus of attention, which affects the way in which speech sounds are perceived. Nosofsky (1986)'s study, based on an experiment using multidimensional visual stimuli, showed that selective attention to specific stimulus dimensions can modify the learners' psychological space, which leads two novel phones to be represented as similar (a shrink in dimensions) or dissimilar (a stretch) in the psychological space. For example, when comparing two objects that differ in size, weight, and color, if size is the critical dimension for comparing the two objects, then a learner will give more weight to fine gradations along this dimension and less weight to the other dimensions. Thus, in the case of Japanese, it is necessary to give weight to the duration dimension to identify the difference between a singleton and a geminate consonant.

In the field of SLA study, based on Nosofsky's study, Pisoni et al. (1994) showed that adult L2 learner's attentive processes can be modified based on their prior linguistic experience, which will lead to better discrimination of phonetic contrasts not present in

their L1. Successful training as linguistic experience will affect perception by modifying learners' attentional processes so that reorganization of psychological dimensions and changes in the weighing scheme of different acoustic cues appropriate to the novel categories can take place (Logan et al., 1991). Guion and Pederson (2002) examined the role of attention in phonetic category formation by manipulating attention shift. Participants oriented to phonetic information demonstrated greater learning of trained novel phonetic categories, whereas participants oriented to semantic information demonstrated greater learning of meaning. This result indicates that 'different instructions gave different "weighting" to the importance of orienting to phonetic or semantic features of the stimuli during training presentation' (p. 17). They assumed that these different weightings could result in either greater signal detection or greater manipulation in working memory. Apparently the orientation to phonetics increases successful detection of the relevant features of the sounds being trained.

As we have observed in Experiment I, the perception data indicated that the beginning level learners were at the stage where their L1 processing cues (the syllabification strategy of English) were competing with L2 timing cues (mora weight assignment in Japanese). Many of the learners had troubles with allocation of length, and even those who could allocate the length had troubles with attributing it to the appropriate segment. Explicit instruction through visual input of speech waveforms raised the learners' awareness of their troubles with length allocation and attribution and further facilitated the shift of their attention to the critical cue to identify contrasts, in this case, singletons and geminate consonants, through visualization during the training. It is assumed that the learners' psychological space was modified favoring the identification

of durational contrasts.

When modification of the perceived psychological spacing among dimensions occurs, there are also changes in the memory representations with regard to the psychologically more salient dimensions (Pisoni et al., 1994). Evidence from previous studies has shown that L2 learners were encoding very specific and detailed contextual information about the acoustic cues of speech sounds in the different phonetic environments, which is consistent with the episodic view of memory encoding and categorization. Unlike the traditional abstractionist view, which assumes that the speech signals are filtered and irrelevant information for processing such as voice details is discarded as noise, the episodic view holds that all tokens of a category are equally available for processing and learning, as seen in multiple-trace theory (e.g., Hintzman, 1986), which assumes that no such normalization process before memory storing. Evidence showed that listeners store detailed memories which include seemingly irrelevant information, such as the exact wording of sentences (Begg, 1971) and exact word format (upper case or lower case in writing; Hintzman, Block & Inskeep, 1972).

In multiple-trace theory, traces of the individual episodes are stored, and aggregates of traces activate in concert at the time of memory retrieval (Hintzman, 1986). This aggregate represents the category as a whole. A number of repetitions of the stimuli presented as various tokens through the training are stored as the traces of each of all the experiences, which they heard and saw as the stimuli in the training; these traces were preserved as primitive properties in PM at the time of stimulus input. After this preservation, the probe or retrieval cue is sent from PM to all traces in SM and PM can receive a single reply or echo back from SM.

As reviewed in Chapter 2, Hardison (2000) proposed scenario of bimodal L2 speech processing and the role of training based on multiple-trace theory. After auditory and visual inputs are preliminarily analyzed through the innate auditory analyzer, a new weighting scheme for L2 must be developed, so that learners can alter their attentions from the optimal setting to perceive distinctive features in L1 to the optimal setting to perceive distinctive features in L2 by learning to attend to new sources of information obtained through the auditory analyzers. Through training, the features of an experience are copied into a trace. Probes, or signals processed in PM, activate already stored traces in SM, and the weighting process occurs according to the trace's similarity to the features of the probe, which ultimately will return an echo to PM. Attention to auditory and visual attributes of the stimulus will determine the features of the probe. Training with multiple exemplars and immediate feedback enhances learning; repetition and feedback can direct attention to critical features of the speech signal and modify the memory system. As the result of learning, new L2 memory traces become less ambiguous and less confusable. Thus, the objective of L2 perceptual training is 'to create a situation in which the echo from an aggregate of L2 traces acting in concert overshadows the echo from L1 traces' (p. 320), since training should direct attention to the distinctive features of the L2 input in order to distinguish them from those of the distinctive features of the L1 system. A great similarity between the L2 input and stored L1 traces causes an ambiguous echo to be returned to PM; therefore, the greater opportunity to create a weighting scheme that shifts learners' attention to the critical auditory and/or visual stimulus properties makes it possible to encode less ambiguous L2 traces in memory.

The visual input of the present study serves as an explicit instruction to orient the

learners' attention to allocation and attribution of length. Through the trainings, new traces were created and stored. The traces were copies of the detailed information about geminate consonants in different contexts, instead of forming a prototype as a single representation of a geminate consonant. A new episodic trace in PM send a probe to search for a good match in SM. When a match is made, an echo is sent back from SM through analogy between the new trace and the stored traces. A better echo or a better match can be achieved if the content of the echo is accurate and strong. Through encountering the stimuli in various contexts many times through training, the echo becomes sharper and stronger and enhances a better match. It is assumed that in general, good learners might get better at searching for a good match through such training. On the other hand, weak learners have weak memory, and analogy is difficult, so that a good match cannot be sent back.

Works such as Pisoni and colleagues (e.g., 1991) and Hardison (e.g., 2003) used varied stimuli with regard to voice, vocalic context, and word position in training, thus exposing the learners to the variability of stimulus which could be encountered in the real world. The results of the training support the hypothesis that memory encoding of speech involves storage of individual episodes, preserving both contextual variability and indexical properties of speech, rather than abstract prototypes. Strange and Dittman (1984) showed that the subjects who received low variability training failed to generalize the effect of the training to new tokens, but high variability stimuli produced successful results. The present study also used geminate consonants in various phonetic contexts as stimuli; different consonants were used with different preceding and following vowels, in different grammatical constructions (words in isolation and in sentence frames).

Perception data in Experiment I indicated that the learners' perception ability varied depending on the phonetic context and identities of geminate consonants; therefore, the result of the perception test indicated that geminate consonants are not all the same in terms of perceptibility. When the phonetic context, such as surrounding vowels differed, and when different types of consonants were used, learners perceived each geminate consonant in different ways, so that it is necessary to consider all these various contexts.

de Sa and Ballard (1997) argue that the co-occurrence of multi-sensory signals can assist processing. They showed through psychophysical studies that information from one modality can assist classification in another and the physiological evidence supports this finding in showing that input to one of the modalities can influence processing in another sensory pathway, in spite of neurophysiological and psychophysical evidence which indicated each modality has its own processing stream. Different modalities have access to each other's output at a higher level at the early stage of processing but can reach the lower levels within each processing stream through mutual feedback, which suggests that this information is coming top-down through feedback pathways from multi-sensory areas. It is also suggested that this integration may not only affect the properties of developed systems but also play an important role in the learning process itself. The visual and auditory stimuli are able to interact to produce a unified perception, which is assumed to be copied to a trace.

Taken together, the visual input and the auditory stimuli used for the present study enhanced learning by raising learners' attention to the distinctive feature of mora weight in geminate consonants, which eventually modified the learners' weighting scheme. The experience or episode through training was enhanced through auditory and

visual sensory feedback, stored as trace, which returned a clearer, less ambiguous echo through training. The effectiveness of the training can be accounted for by an episodic view of learning such as that suggested by multiple-trace theory.

7.3 Limitations of the study and further study suggestions

One of the limitations of the present study is the difference between the learners' groups in the perception data (Experiment I) and the production test (Experiment III); the perception data was obtained from the learners in the U.S. and the production data was obtained from the learners who were studying in Japan. Although we reached some generalization that the learners' production as well as perception was affected by phonetic contexts and identities of geminate consonants, a direct comparison between the perception and production data was not conducted due to this difference. While Experiment IV showed some links between the learners' perception and production abilities by demonstrating the transfer effect from perceptual training to production improvement, further research needs to be conducted on the same subject group to investigate the relationship between these two abilities more closely, and see whether there would be any common different contexts and types of geminate consonants, specially because there is no previous study of such comparison regarding geminate consonants in Japanese.

Second, the participants in the training study were limited to beginning learners to avoid previous knowledge of the stimulus; however, it is also important to examine the interaction between learners' knowledge of the meaning of words containing geminate consonants and perception ability of them. While most of the previous studies of the

perception of geminate consonants focused on acoustic factors such as absolute stop closure duration, no studies have been done in this respect. Since this may be another important factor besides acoustic cues for learners to perceive geminate consonants in natural settings, further investigation would be necessary.

There are many other possibilities to expand the training studies. For example, the training in the present study showed the effect on improvement in perception of geminate consonants, but it is worth examining whether the same effect would be found on other types of difficult durational contrasts in Japanese, i.e., long vowels (/kite/ vs. /kiite/) and moraic nasals (/kona/ vs. /koNna/). Considering how the visual input successfully guided the learners' attention and promoted a new weighting scheme for the distinctive features of the L2 input, it is assumed that the perceptual training with visual input as the one used in this study might bring about the similar effect.

Furthermore, it may also be interesting to explore whether a focused audio-visual presentation of another sort would achieve similar results as the present study. While in the present study, the contribution of visual input was brought by a presentation of speechwave forms, it is assumed that visual input to direct learners' awareness to mora weight contrast might be similarly effective. For example, the spelling system in Japanese is an effective tool to indicate the numbers of mora. Geminate consonants are spelled with small "tsu (っ)" in Japanese orthography, so that /kitte/ is spelled as "ki tsu te (きって)", while singleton /kite/ is spelled as "ki te (きて)". Thus, in Japanese orthography, the number of letters show a difference in the number of mora (3 and 2, respectively), which might be another effective visual input to promote their awareness of durational contrasts.

Another interest for future research is to investigate the long-term retention of

learning in both perception and production after the training. Bradlow et al. (1999) reported that three months after training of /r/-/l/ contrasts in English, the Japanese participants maintained their improved levels of perception performance. They concluded that the high-variability training was effective for not only the acquisition but also retention of new segmental contrasts. In addition, the transferred improvement on production performance through perception training was also retained. It is highly likely that the present training program may also show a similar effect, considering the high variability stimulus and transfer effect of another modality; hence it is worth further study.

While the traditional training programs in previous studies were typically carried out in a laboratory setting, the on-line training program of the present study could be easily integrated into classroom teaching. One of the advantages of the on-line training was easy and flexible access for the learners. Teachers do not have to spare class hours for training sessions, and students can access the training materials after school, anytime and anywhere they want to. In the post-hoc interview, many of the learners said that they accessed the training material between 10:00 and 12:00 at night, since that was the time when they spent leisure time in front of their computers after they did school work required for the following day. Such easy access could reduce the learners' psychological burden, such as nervousness and pressure, and they do not have to make extra effort to go and sit in a language lab at a designated time. However, in exchange for such flexibility, there is a drawback for research purposes; without the presence of an observer, there is no way to watch and control the amount and frequency of training learners really take. Login systems with a password and tracking systems to monitor students' attended training

duration might be a solution, though it might require more specialized knowledge for the teacher to create such a program. This might risk losing another advantage of the training program of the present study; creating on-line materials such as the ones used in the present study does not require any special equipment, while the laboratory setting is costly and requires special skills and knowledge to manage it. Hence, additional research is required to avoid sacrificing such advantages, while developing more advanced program to monitor learners.

Further, the results indicated that the effect of training in perception transferred to improvement in production. Some studies have also reported a transfer effect, but with the opposite tendency; that is, production training brought about an improvement in perception, as reviewed in Chapter 2. According to Rochet (1995), either perception training or production training could be equally effective to improve ability in the other skill. However, considering the difficulty in creating automated feedback programs to respond to learners' production, perception training would be more realistic for instructors. Akahane-Yamada, Tohkura, Bradlow, and Pisoni (1996) also suggest that since computer-based training in production is more difficult than that in perception, 'tuning the trainees' speech perception' (p. 609) will facilitate learning in production. However, since it has been reported that production ability improved by perception training lagged behind perception and required more time to achieve parity with perception ability (Akahane-Yamada, 1999), it is worth exploring to what extent productive ability can be trained through perceptual training and vice versa.

In conclusion, the results of the present experiment demonstrate that the perception of geminate consonants by English speakers can be improved using a

perceptual training with visual input which displays speechwave forms of the target segments. An alternative and effective way is offered to train learners of Japanese to perceive geminate consonants as well as produce them. The result of the perceptual training brought about improvement in the learners' perception, which was not limited to the tokens used in the training; the result of the generalization tests indicated the effect could be carried over to new tokens and pronunciations by a new voice which the learners were not familiar with. It is assumed that visualization with waveforms helped the learners grasp a clearer picture of durational contrasts. The present study raises many interesting and potentially important questions about the nature of stimulus variability in perceptual learning and its role in training nonnative listeners to perceive phonetic contrasts that are not distinctive in their language.

APPENDIX A

Perception test

Part I

You will listen to some Japanese words. Each question has 3 options. Circle the word you're hearing. You will hear each word only once. If you are not sure which word you heard, make your best guess. Please do not leave any numbers unanswered. Don't be concerned if you don't understand the words.

Examples:

You will hear: kitte

Circle: a kite
b. kiite
©. kitte

1) a. ata
b. aata
c. atta

2) a. atsu
b. aatsu
c. attsu

3) a. kaze
b. kazee
c. kaaze

4) a. aka
b. aaka
c. akka

5) a. taachu
b. tachuu
c. tachuu

6) a. saku
b. saaku
c. sakku

7) a. asu
b. aasu
c. assu

8) a. aku
b. aaku
c. akku

9) a. sasu
b. saasu
c. sassu

10) a. asa
b. aasa
c. assa

11) a. saka
b. saaka
c. sakka

12) a. satsu
b. saatsu
c. sattsu

13) a. sechi
b. seechi
c. secchi

14) a. sasa
b. saasa
c. sassa

15) a. kaki
b. kaaki
c. kakki

16) a. atsu
b. aatsu
c. attsu

17) a. sasa
b. saasa
c. sassa

18) a. saka
b. saaka
c. sakka

19) a. aka
b. aaka
c. akka

20) a. kate
b. kaate
c. katte

21) a. sasu
b. saasu
c. sassu

22) a. asu
b. aasu
c. assu

23) a. saku
b. sakku
c. saaku

24) a. satsu
b. saatsu
c. sattsu

25) a. ata
b. aata
c. atta

26) a. atsu
b. aatsu
c. attsu

27) a. kakite
b. kakiite
c. kakkite

28) a. asa
b. aasa
c. assa

29) a. aku
b. aaku
c. akku

30) a. sata
b. saata
c. satta

Part II

You will listen to Japanese words in frame sentences. The frame sentence is given as in the examples. Each question has 3 options. Circle the word you're hearing. You will hear each word only once. If you are not sure which word you heard, make your best guess. Please do not leave any numbers unanswered. Don't be concerned if you don't understand the words.

Examples:
You will hear: kitte

Circle:
watashi wa _____to iimashita
a kite
b. kiite
©. kitte

1) watashi wa _____to iimashita
a. ata
b. aata
c. atta

2) kore wa _____desu
a. asu
b. aasu
c. assi

3) watashi wa _____to iimashita
a. saka
b. sakka
c. saaka

4) kore wa _____desu
a. sasu
b. saasu
c. sassu

5) watashi wa _____to iimashita
a. kakite
b. kaakite
c. kakkite

6) kore wa _____desu
a. sata
b. saata
c. satta

7) kore wa _____desu

14) watashi wa _____to iimashita

a. aka
b. aaka
c.akka

8) watashi wa _____to iimashita
a. taachu
b. tachuu
c. tachuu

9) kore wa _____desu
a. saku
b. saaku
c. sakku

10) watashi wa _____to iimashita
a. tatu
b. taatsu
c. tattsu

11) kore wa _____desu
a. aku
b. aaku
c. akku

12) watashi wa _____to iimashita
a. aka
b. aaka
c. akka

13) kore wa _____desu
a. aku
b. aaku
c. akku

a. satsu
b. saatsu
c. sattsu

15) kore wa _____desu
a. sasa
b. saasa
c. sassa

16) watashi wa _____to iimashita
a. ata
b. aata
c. atta

17) kore wa _____desu
a. keta
b. keeta
c. ketta

18) watashi wa _____to iimashita
a. sasa
b. saasa
c. sassa

19) kore wa _____desu
a. asa
b. aasa
c. assa

20) watashi wa _____to iimashita
a. atsu
b. aatsu
c. attsu

21) kore wa _____desu

28) kore wa _____desu

a. sasu
b. saasu
c. sassu

22) watashi wa _____to iimashita
a. atsu
b. aatsu
c. atssu

23) kore wa _____desu
a. kaze
b. kaaze
c. kazze

24) watashi wa _____to iimashita
a. asu
b. aasu
c. assi

25) kore wa _____desu
a. saka
b. sakka
c. saaka

26) watashi wa _____to iimashita
a. asa
b. aasa
c. assa

27) kore wa _____desu
a. kate
b. kaate
c. katte

a. saku
b. sakku
c. saaku

29) watashi wa _____to iimashita
a. satsu
b. saatsu
c. sattsu

30) watashi wa _____to iimashita
a. sata
b. saata
c. satta

APPENDIX B

Production test

Part I

Please read the following words only once and record on the provided cassette tape.

1. satta さった	2. assu あっす	3. saka さか
4. aka あか	5. tachuu たちゅう	6. sakku さっく
7. taatsu たあつ	8. akku あっく	9. sasu さす
10. assa あっさ	11. sakka さっか	12. kakisu かきす
13. akka あっか	14. aku あく	15. sassu さっす
16. asu あす	17. kazeki かぜき	18. sasa ささ
19. takata たかた	20. sassa さっさ	21. atta あった
22. attsu あっつ	23. saku さく	24. satsu さつ
25. atta あった	26. atsu あつ	27. sattsu さっつ
28. asa あさ	29. kakute かくて	30. sata さた

Part II

Please read the following sentences only once and record on the provided cassette tape.

1. watashi wa sata to iimashita
   わたしは　さたといいました

2. kore wa kakute desu
   これは　かくてです

3. watashi wa asa to iimashita
   わたしは　あさといいました

4. kore wa atta desu
   これは　あったです

5. watashi wa atsu to iimashita
   わたしは　あつといいました

6. kore wa kakisu desu
   これは　かきすです

7. watashi wa satsu to iimashita
   わたしは　さつといいました

8. kore wa saku desu
   これは　さくです

9. watashi wa attsu to iimashita
   わたしは　あっつといいました

10. kore wa sasa desu
    これは　ささです

11. watashi wa sassa to iimashita
    わたしは　さっさといいました

12. kore wa takata desu
これは　たかたです

13. watashi wa atta to iimashita
わたしは　あったといいました

14. kore wa aku desu
これは　あくです

15. watashi wa asu to iimashita
わたしは　あすといいました

16. kore wa sassu desu
これは　さっすです

17. watashi wa kazeki to iimashita
わたしは　かぜきといいました

18. kore wa akka desu
これは　あっかです

19. watashi wa sattsu to iimashita
わたしは　さっつといいました

20. kore wa sakka desu
これは　さっかです

21. watashi wa assa to iimashita
わたしは　あっさといいました

22. kore wa sasu desu
これは　さすです

23. watashi wa akku to iimashita
わたしは　あっくといいました

24. kore wa taatsu desu
これは　たあつです

25. watashi wa sakku to iimashita
わたしは　さっくといいました

26. kore wa tachuu desu
これは　たちゅうです

27. watashi wa aka to iimashita
わたしは　あかといいました

28. kore wa saka desu
これは　さかです

29. watashi wa assu to iimashita
わたしは　あっすといいました

30. kore wa satta desu
これは　さったです

REFERENCES

Akahane-Yamada, R. (1999). Learning non-native speech contrasts: Perception and production relationship. *Technical Report of IEICE, 99*(23), 37-42.

Akahane-Yamada, R., Tohkura, Y., Bradlow, A., & Pisoni, D. (1996). Does training in speech perception modify speech production? *Proceedings of the 1996 International Conference on Spoken Language Processing*, 606-609.

Anderson-Hsieh, J. (1994). Interpreting visual feedback on suprasegmentals in computer assisted pronunciation instruction. *CALICO Journal, 11*(4), 5-22.

Anderson-Hsieh, J. (1996). Teaching suprasegmentals to Japanese learners of English through electronic visual feedback. *JALT Journal, 18*(2), 315-325.

Aslin, R. N., Pisoni, D. B., Hennesy, B. L., & Perey, A. J. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. *Child Development, 52*, 1135-1145.

Barry, W. (1989). Perception and production of English vowels by German learners: Instrumental-phonetic support in language teaching. *Phonetica, 46*, 155-168.

Begg, I. (1971). Recognition memory for sentence meaning and wording. *Journal of Verbal Learning and Verbal Behavior, 10*, 176-181.

Best, C. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues* (pp. 171-204). Timonium, MD: York Press.

Bohn, O. -S., & Flege, J. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics, 11*(3), 302-328.

Borden, G., Gerber, A., & Milsark, G. (1983). Production and perception of the /r/-/l/ contrast in Korean adults learning English. *Language Learning, 33*(3), 499-526.

Bradlow, A., Akahane-Yamada, R., Pisoni, D., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics, 61*, 977-985.

Bradlow, A., Pisoni, D., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101*, 2299-2310.

Broselow, E & Park, H-B. (1995). Mora conservation in second language prosody. In J. Archibald (Ed.), *Phonological acquisition and phonological theory* (pp. 151-168). Hillsdale, NJ: Erlbaum.

Catford, J. C., & Pisoni, D. (1970). Auditory vs. articulatory training in exotic sounds. *Modern Language Journal, 54*, 477-481.

Chun, D. M. (1989). Teaching tone and intonation with microcomputers. *CALICO Journal, 7*(1), 21-46.

Chun, D. M. (1998). Signal analysis software for teaching discourse intonation. *Language Learning & Technology, 2*(1), 74-91.

Chun, D. M. (2002). *Discourse intonation in L2: From theory and research to practice*. Amsterdam: John Benjamins.

de Bot, K. (1983). Visual feedback of intonation I: Effectiveness and induced practice behavior. *Language and Speech, 26*(4), 331-350.

de Sa, V.R., & Ballard, D. (1997). Perceptual learning from cross-modal feedback. In R. L. Goldstone, P. G. Schyns, & D. L. Medin (Eds.), *The psychology of learning and motivation*, (Vol. 36, pp. 309-351). San Diego, CA: Academic Press.

Dickerson, L., & Dickerson, W. B. (1977). Interlanguage phonology: Current research and future directions. In S. P. Corder & E. Roulet (Eds.), *The notion of simplification, interlanguages and pidgins and their relation to second language pedagogy* (pp. 18-30). Geneva, Switzerland: Droz.

Derwing, B. L. (1992). A 'pause-break' task for eliciting syllable boundary judgments from literate and illiterate speakers: Preliminary results for five diverse languages. *Language and Speech, 35*, 219-235.

Enomoto, K. (1992). Interlanguage phonology: The perceptual development of durational contrasts by English-speaking learners of Japanese. *Edinburgh Working Papers in Applied Linguistics, 3*, 25-35.

Flege, J. (1991). Perception and production: The relevance of phonetic input to second language phonological learning. In T. Huebner & C. Ferguson (Eds.), *Crosscurrents in second language acquisition and linguistic theories*. Amsterdam: John Benjamins.

Flege, J. (1995). Second language speech learning theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp.233-269). Timonium, MD: York Press.

Flege, J., & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics, 15*(1), 67-83.

Flege, J., Munro, M., & MacKay, I. (1995). Effects of age of second-language learning on the production of English consonants. *Speech Communication, 16*(1), 1-26.

Fujisaki, H., & Sugito, M. (1977). Onsei no butsuriteki seishitsu. [Physical property of sounds]. In M. Hashimoto (Ed.), *Iwanami Koza Nihongo* (Vol. 5, pp. 63-106). Tokyo: Iwanami Shoten.

Fukui, S. (1978). Perception for the Japanese stop consonants with reduced and extended durations. *The Bulletin of the Phonetic Society of Japan, 159*, 9-12.

Gass, S. (1984). Development of speech perception and speech production abilities in adult second language learners. *Applied Psycholinguistics, 5*, 51-74.

Goldinger, S. D. (1997). Words and voices: Perception and production in an episodic lexicon. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 233-277). San Diego: Academic Press.

Goldsmith, J. A. (1990). *Autosegmental and metrical phonology*. Cambridge, MA: Blackwell.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds /l/ and /r/. *Neuropsychologia, 9*, 317-323.

Guion, S. G., & Pederson, E. (2002). The role of orienting attention for learning novel phonetic categories. *ICDS Technical Reports, 02-5*. Retrieved June 23, 2006, from http://hebb.uoregon.edu/02-05tech.pdf

Han, M. (1992). The timing control of geminate and single stop consonants in Japanese: A challenge for nonnative speakers. *Phonetica, 49*, 102-127.

Hardison, D. M. (1998). Acquisition of second-language speech: Effects of visual cues, context and talker variability. *Dissertation Abstracts International. 59* (05), 1546 (UMI No. 9834598)

Hardison, D. M. (2000). The neurocognitive foundation of second-language speech: A proposed scenario of bimodal development. *Social and Cognitive Factors in Second Language Acquisition*, 312-325.

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics, 24*, 495-522.

Hardison, D. M. (2004). Generalization of computer assisted prosody training: Quantitative and Qualitative findings. *Language Learning & Technology*, *8*(1), 34-52.

Hardison, D. M. (2005a). Contextualized computer-based L2 prosody training: Evaluating the effects of discourse context and video input. *CALICO Journal*, *22*(2), 175-190.

Hardison, D. M. (2005b). Second-language spoken word identification: Effects of perceptual training, visual cues, and phonetic environment. *Applied Psycholinguistics*, *26*, 579-596.

Hayes, B. (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry*, *20*, 30-253.

Hayes, R. L. (2002). The perception of novel phoneme contrasts in a second language: A developmental study of native speakers of English learning Japanese singleton and geminate consonant contrasts. *Coyote Papers*, *12*, 28-41.

Hintzman, D. L. (1986). `Schema abstraction' in a multiple-trace memory model. *Psychological Review*, *93*, 411-428.

Hintzman, D. L., Block, R. A., & Inskeep, N. R. (1972). Memory for mode of input. *Journal of Verbal Learning and Verbal Behavior*, *11*, 741-749.

Hirata, Y. (1990a). Perception of geminated stops in Japanese word and sentence levels. *Journal of the Phonetic Society of Japan*, *194*, 23-28.

Hirata, Y. (1990b). Perception of geminated stops in Japanese word and sentence levels by English-speaking learners of Japanese language. *Journal of the Phonetic Society of Japan*, *195*, 4-10.

Hirata, Y. (1999) Acquisition of Japanese rhythm and pitch accent by English native speakers. *Dissertation Abstracts International. 60* (08), 2892 (UMI No. 9943077)

Hirata, Y. (2004). Computer assisted pronunciation training for native English speakers learning Japanese pitch and durational contrasts. *Computer Assisted Language Learning*, *17*(3-4), 357-376.

Hirato, N., & Watanabe, S. (1987). The relation between the perceptual boundary of voiceless plosives and their moraic counterparts and the length of vowels following closure. *Studies in Phonetics and Speech Communication, 2*, 99-106.

Homma, Y. (1981). Durational relationship between Japanese stops and vowels. *Journal of Phonetics*, *9*, 273-281.

Ingram, J., & Park, S.-G. (1998). Language, context and speaker effects in the identification and discrimination of English /r/ and /l/ by Japanese and Korean listeners. *Journal of the Acoustical Society of America, 103,* 1161-1174.

Ishikawa, K. (2002). Syllabification of intervocalic consonants by English and Japanese speakers. *Language and Speech, 45,* 355-385.

Jamieson, D., & Morosan, D. (1986). Training nonnative speech contrasts in adults: Acquisition of the English /ð/ - /θ/ contrast by Francophones. *Perception & Psychophysics, 40*(4), 205-215.

Jamieson, D., & Morosan, D. (1989). Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology, 43*(1), 88-96.

Jamieson, D., & Rvachew, S. (1992). Premeditating speech production errors with sound identification training. *Journal of Speech-Language Pathology and Audiology, 16,* 201-210.

Johnson, K. (1997). Speech perception without speaker normalization. In K. Johnson & J. W. Mullenix (Eds.), *Talker variability in speech processing* (pp. 145-165). San Diego, CA: Academic Press.

Jusczyk, P. W. (1993). From general to language-specific capacities: The WRAPSA Model of how speech perception develops. *Journal of Phonetics, 21,* 3-28.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Kenstowicz, M. (1994). *Phonology in generative grammar*. Cambridge, MA: Blackwell.

Koguma, R. (2001). Production of Japanese short and long vowels by English-speaking learners. *Journal of Japanese Language Takushoku University, 11,* 79-88.

Ladefoged, P. (1993). *A course in phonetics* (3rd ed.). Orland, FL: Harcourt Brace Javanovich, Inc.

Lambacher, S. (1999). A CALL tool for improving second language acquisition of English consonants by Japanese learners. *Computer Assisted Language Learning, 12*(2), 137-156.

Lambacher, S. (2001). The taming of English vowels. *Proceedings of the PC3/JALT 2001 Convention,* 1-5.

Landahl, K. L., Ziolokowski, M. S., Usami, M., & Tunnock, B. K. (1992). Interactive articulation: Improving accent through visual feedback. *The Proceedings of the Second International Conference on Foreign Language Education and Technology*, 283-292.

Leather, J. (1990). Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers. In J. Leather & A. James (Eds.), *New Sounds 90* (pp. 72-89). Amsterdam: University of Amsterdam.

Levis, J., & Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System*, *32*, 505-524.

Lin, Y. (2001). Syllable simplification strategies: A stylistic perspective. *Language Learning*, *51*(4), 681-718.

Lindblom, B, Guion, S., Hura, S., Moon S.-J., & Willerman, R. (1995). Is sound change adaptive? *Rivista Di Linguistica*, *7*(1), 5-37.

Lively, S., Logan, J., & Pisoni, D. (1993). Training Japanese listeners to identify English /r/ and /l/ II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, *94*(3), 1242-1255.

Logan, J., Lively, S., & Pisoni, D. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, *89*(2), 874-886.

Mack, M. (1989). Consonant and vowel perception and production: Early English-French bilinguals and English monolinguals. *Perception and Psychophysics*, *46*, 187-200.

McCandliss, B., Fiez, J., Protopapas, A., Conway, M., & McClelland, J. (2002). Success and failure in teaching the [r]-[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, & Behavioral Neuroscience*, *2*(2), 89-108.

McClaskey, C., Pisoni, D., & Carell, T. (1983). Transfer of training of a new linguistic contrast in voicing. *Perception and Psychophysics*, *34*, 323-330

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746-748.

Min, G. J. (1987). Kankokujin no nihongo no sokuon no chikaku ni tsuite. [Perception of Japanese geminate consonants by Korean]. *Nihongo Kyoiku*, *62*, 179-193.

Min, G. J. (1993). Nihongo sokuon no kikitorihandan ni kansuru kenkyu. [A study on perception of Japanese geminate consonants]. *Sekai no Nihongo Kyoiku, 3*, 237-249.

Minagawa, Y. (1996). Sokuon no shikibetsu ni okeru akusentogata to shiinshu no yoin. [Accent patterns and consonant types as factors for perceiving geminate consonants]. *Proceedings of the Spring Meeting of the Society for Teaching Japanese as a Foreign Language*, 97-102.

Minagawa, Y., & Kiritani, S. (1996). Discrimination of the single and geminate stop consonant in Japanese by five different language groups. *Anmual Bulletin, RIPL, 30*, 23-28.

Molholt, G. (1988). Computer-assisted instruction in pronunciation for Chinese speakers of American English. *TESOL Quarterly, 22*(1), 91-111.

Molholt, G. (1990). Spectrographic analysis and patterns in pronunciation. *Computers and the Humanities, 24*, 81-92.

Nishibata, C. (1993). Heisokujikan o hensuu to shita nihongo sokuon no chikaku no kenkyu. [A study of perception of geminate consonants with stop closures as variables]. *Journal of Japanese Language Teaching, 81*, 128-140.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39-57.

Ofuka, E. (2003). Perception of a Japanese geminate stop /tt/: The effect of pitch type and acoustic characteristics of preceding/following vowels. *Journal of the Phonetic Society of Japan, 7*(1), 70-76.

Pisoni, D., Aslin, R., Perey, A., & Hennessy, B. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology, 8*, 297-314.

Pisoni, D., Lively, S., & Logan, J. (1994). Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception. In J. Goodman & H. Nusbaum (Eds.), *The development of speech perception: The transition from speech* (pp. 121-166). Cambridge, MA: MIT Press.

Port, R. F., Dalby, J., & O'Dell, M. (1987). Evidence for mora timing in Japanese. *Journal of the Acoustical Society of America, 81*(5), 1574-1585.

Rochet, B. L. (1995). Perception and production of L2 speech sounds by adults. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 379-410). Timonium, MD: York Press.

Sato, C. (1984). Phonological processes in second language acquisition: Another look at interlanguage syllable structure. *Language Learning, 34*, 43-57.

Schiller, N.O., Meyer, A.S., & Levelt, W.J.M. (1997). The syllabic structure of spoken words: Evidence from the syllabification of intervocalic consonants. *Language and Speech, 40*, 103-140.

Schmidt, R. (1987). Sociolinguistic variation and language transfer in phonology. In G. Ioup & S. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 365-377). Rowley, MA: Newbury House.

Selkirk, E. O. (1984). On the major class features and syllable theory. In M. Aronoff & R. T. Oehrle (Eds.), *Language sound structure* (pp. 107-136). Cambridge, MA: MIT Press.

Sheldon, A. (1985). The relationship between production and perception of the /r/-/l/ contrast in Korean Adults learning English. A reply to Borden, Gerber and Milsark. *Language Learning, 35*, 107-113.

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: evidence that speech production can precede speech perception. *Applied Psycholinguistics, 3*, 243-261.

Shibatani, M. (1990). *The languages of Japan*. Cambridge: Cambridge University Press.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics, 36*, 131-145.

Sugito, M. (1999). Nihongo onsei no onseigaku teki tokuchoo. [Phonetic characteristics of Japanese]. In M. Sugito (Ed.), *Kyoiku e no teigen* (pp. 59-71). Osaka: Izumi Shoin.

Tarone, E. (1982). Systematicity and attention in interlanguage. *Language Learning, 32*, 69-82.

Toda, T. (1994). Interlanguage phonology: Acquisition of timing control in Japanese. *Australian Review of Applied Linguistics, 17*(2), 51-76.

Toda, T. (1998). Nihongo gakushusha ni yoru sokuon choon hatsuon no chikakuhanchuka. [Categorical perception by Japanese learners]. *Bungei Gengo Kenkyu, 33*, 65-82.

Toda, T. (2003). *Second language speech perception and production: Acquisition of phonological contrasts in Japanese*. Lanham, MD: University Press of America.

Weltens, B., & Bot, K. de (1984). Visual feedback of Intonation II: Feedback delay and quality of feedback. *Language and Speech*, *27*(1), 79-88.

Yamada, T., Akahane-Yamada, R., & Strange, W. (1995). Perceptual learning of Japanese mora syllables by native speakers of American English: Effects of training stimulus sets and initial states. *Proceedings of the XIIIth International Congress of Phonetic Sciences*, *1*, 322-325.

Yamagata, A., & Preston, D. (1999). English learners' acquisition of the Katakana spelling of English loan-words in Japanese. In M. Wysocka (Ed.), *On language theory and practice. In honor of Janusz Arabski on the occasion of his 60th birthday* (Vol. 2). Katowice: University of Silesia.

Ziolkowski, M. S., Usami, M., Landahl, K. L., & Tunnock, B. (1992). How many phonologies are there in one speaker? Some experimental evidence. *Proceedings of 1992 International Conference on Spoken Language Processing*, 2, 1315-1318.