The Acquisition of the /u/ and /l/ by Japanese Learners of English

By Jason Parry

For David Wood

August, 2012
The Acquisition of the /ɾ/ and /l/ by Japanese Learners of English

Introduction

Japanese is strikingly different from English in a variety of ways. A look at Japanese structures for formality and politeness for example, reveals a language which is extremely sensitive to context. Unlike English, Japanese is so meticulously classified that an eavesdropper who overhears just a fragment of a Japanese conversation is able to tell the age, sex, relationship, and relative status of its participants with ease. As a result, the meaning of an utterance in Japanese is dramatically affected by even the minutest changes of lexis or grammar. These features of the language have an impact on the way Japanese speakers perceive the world around them, giving them “an amazing ability to hear the unspoken word and to sense changes in atmosphere and human relationships” (Swan and Smith, 2001, p. 297).

In addition to the lexico-grammatical idiosyncrasies which make Japanese unique are phonological characteristics which further isolate it from other languages. A look at the language’s r sound for example, reveals a striking difference between Japanese and English. In Japanese, the r sound has been identified as the tap /ɾ/, and is “produced with lateral articulation, usually with a tendency towards retroflexing” (Sheldon and Strange, 1982, p. 244). In lateral articulation, the air travels along the sides of the tongue, but is prevented by the tongue from going through the middle of the mouth. Retroflexing involves “curling the tongue tip back so that the underside of the tongue tip taps the alveolar ridge (or by bunching the tongue upwards and back in the mouth in order to get the tongue tip to tap the back of the alveolar ridge)” (Dobrovolsky and Katamba, 1996, p.32). Although this sound is referred to as the Japanese r, it is actually phonetically more similar to the American English /l/ sound than it is to the American

The American English r, on the other hand, is quite different from the Japanese tap /ɾ/ (henceforth, referred to as the ‘Japanese /ɾ/’). The IPA symbol for the American English r is /ɹ/, and it is described as a postalveolar approximant. Approximants are consonant sounds which are produced by narrowing but not completely blocking the vocal tract so that air can still escape from the mouth. In the case of /ɹ/, the tongue is “brought close to the area just past the alveolar ridge (hence the label postalveolar) without making contact with the roof of the mouth” (Dobrovolsky and Katamba, 1996, p.32).

Although they vary in their articulation, the Japanese /ɾ/, English /ɹ/, and English /l/ all belong to a class of consonants known as liquids. Liquids are common to most languages of the world; however, their articulation and quantity vary (Dobrovolsky and Katamba, 1996). Much to the detriment of Japanese learners of English, Japanese contains only one liquid (Japanese /ɾ/), while English contains two (/l/ and /ɹ/). Studies have demonstrated that this feature of Japanese phonology has a negative influence on how Japanese speakers experience English. Since their first language (L1) only contains one liquid, they perceptually assimilate both English liquids to the Japanese /ɾ/ (Miyawaki, Strange, Verbrugge, Liberman, Jenkins and Fujimura, 1975).

Beyond the articulatory differences between Japanese /ɾ/, and /ɹ/ and /l/ are differences between their formant values. When a person produces a sound (such and Japanese /ɾ/ or /ɹ/), the vocal cords begin opening and closing at an average of anywhere between 125 to 250 times per second, causing sound to be produced in bursting peaks of energy at varying frequencies. The production and subsequent subsiding of these amplitude peaks are known as formants (Carmell,
In English, the most essential cue for differentiation between the liquids (/ʃ/ and /l/) is the frequency of their third formant. While the third formant of /ʃ/ is produced at as low as 2000 Hz, /l/ is produced at approximately 2400 Hz. (Bradlow, 2003, p. 292). In the production of the Japanese /r/ on the other hand, the third formant is not necessary for comprehension of the sound. As a result, Japanese speakers lack experience in giving auditory attention to the third formant of liquid consonants. Since the third formant plays an essential role in the accurate perception and production of the English liquids, Japanese learners of English experience a great deal of difficulty learning to differentiate between them (MacKain, Best and Strange, 1981).

This is well illustrated by a study conducted by Iverson, Kuhl, Akahane-Yamada, Diesch, Tohkura, Kettermann and Siebert (2003), which contrasted the perceptual ability of 24 native speakers of Japanese with that of 12 native speakers of German and 19 native speakers of English. In the study, the participants were subjected to multiple tests using a series of 18 /ra/ and /la/ stimuli which were synthesized to model speech recorded by a female native-American English speaker. Three tests were completed, including an identification and goodness test, a similarity scaling test, and a discrimination test. The researchers found that while the American English and German speaking group were able to accurately categorize the stimuli as /ʃ/ or /l/, the Japanese speaking group could not. In the first test (the identification task), the subjects were asked to identify which sound from their language each of the stimuli represented. Although nine of the 18 stimuli represented English /l/, the Japanese group determined 17 of the 18 stimuli as representations of the Japanese /r/ and the remaining stimulus a representation of the Japanese /w/. Since however, the Japanese language does not have an /l/ sound in its phonology, this is not surprising.
In the second test, the participants were asked to rate how perceptually different the stimuli were from each other. For the American and German group, the stimuli sounded perceptually ‘further’ from each other as they approached the /d/-/l/ boundary, and ‘closer’ to each other the further they were from the /d/-/l/ boundary. Likewise, the American and German groups showed the highest sensitivity during the discrimination test closest to the /d/-/l/ boundary. Conversely, the Japanese speaking group did not experience an increased perceptual “distance” between the stimuli as they approached the /r/-/l/ boundary. Instead, distance was perceived within each of the categories. Similarly, they were unable to differentiate between the stimuli close to the /d/-/l/ boundary, but were able to differentiate between the stimuli within each of the categories more effectively than the American English or German speaking group. As a result of these findings, the researchers concluded that Japanese speakers are more sensitive to the second formant (F2) of the English liquids, but are not sensitive to the third formant (F3) (Iverson, et al., 2003), which as noted above, is considered to be the primary cue for the contrast in English (MacKain, et al., 1981).

Zhang, Kuhl, Imada, Kotani, and Tohkura (2005) found similar results in a study involving American and Japanese native speakers who were tested on their ability to perceive /d/-/l/ and /b/-/w/ contrasts. The latter were chosen because they are native to both languages, whereas the former were chosen because as stated above, they are native only to English. In addition to asking the study participants to correctly differentiate between the phonemes, the researchers used magnetoencephalography (MEG) to record neural electrical activity throughout the tests. Results from the MEG showed that Japanese listeners were less sensitive to the phonemic /r/-/l/ difference than American listeners. Furthermore, processing “non-native speech sounds recruited significantly greater brain resources in both hemispheres and required a
significantly longer period of brain activation in two regions, the superior temporal area and the inferior parietal area” (Zhang, et al., 2005, p. 703).

This paper will attempt to elucidate the perceptive difficulties facing Japanese learners of English, first by outlining several variables which have been shown to further complicate the perception of /ɹ/ and /l/, then by exploring theoretical models which attempt to explain the perceptual difficulties experienced by Japanese learners of English, and finally by offering some suggestions for phonemic training in the classroom.

**Variables affecting the perception of /ɹ/ and /l/ phones by native speakers of Japanese**

The perceptual difficulties faced by Japanese learners of English have been exhaustively researched over the years. Findings from these studies illuminate several factors which influence a learner’s ability to successfully discern between /ɹ/ and /l/. In this section, the age of the learner, the learner’s previous interaction with English, the quality of the previous interaction with English, and the nature and position of the English liquid will be discussed.

**Age of the learner**

The idea that age could be related to the potential proficiency of language learners has been debated for over half a century. In 1959, Penfield and Roberts proposed that there was a Critical Period (CP) where it was biologically possible to achieve native-like competence in a second language. This hypothesis was then substantiated by Lenneberg in 1967, with evidence stemming from data on recovery from language loss after brain trauma. Lenneberg speculated that the critical period for acquisition began at two years of age, and closed with the onset of puberty (defined as 15 years old) (Hakuta, Bialystok, and Wiley, 2003). This was believed to be
the result of a cessation in the neural plasticity of the brain (Kuhl, Conboy, Padden, Nelson, and Pruitt, 2005).

Although there is little disagreement in the literature regarding the existence of a critical period for first language acquisition, the existence of such a period in second language acquisition is still contested. While some researchers claim that the acquisition of an L1 does not neurologically inhibit further language acquisition, others believe that universal grammar (UG) atrophy and changes in the neurological mechanisms of the brain make native-like attainment of a target language (TL) impossible (Hakuta, et al., 2003).

In a study (Hakuta, et al., 2003), 1990 census data from 2.3 million immigrants to the United States with either Spanish or Chinese language backgrounds was used to test the Critical Period Hypothesis (CPH). These immigrants had all arrived in the United States at different ages, thus varying their age of initial exposure to English. Census data included the participants perceived language proficiency at the time of the survey, their educational background, and age of arrival to the United States. Using the data, Hakuta, et al. (2003) were able to test immigrants with an arrival at both the ages of 15 and 20 years for a sudden change in their ability to acquire English. Upon analysis of the data, however, the researchers were unable to find a sudden decline in language learning ability due to age. Instead, they concluded that although the ability to find success in second language acquisition declines throughout the life span, its decline is gradual and not abrupt (Hakuta, et al., 2003). Similarly, while a Japanese learner’s ability to acquire the phonology of English may decline with age, CPH cannot fully account for the perceptual difficulties experienced by adult Japanese learners of English.
**Amount of interaction with English**

Several studies on Japanese learners of English seem to demonstrate a positive relationship between the amount of exposure to the TL and the accurate perception of the /u/ and /l/ phonemes. For example, in 1981 MacKain, et al. contrasted the perceptive abilities of inexperienced and experienced Japanese learners of English. The inexperienced group had an average of 8.7 months of conversational English experience with native English speakers, whereas the experienced group had an average of 27.6. As a control, ten native English speaking undergraduates from Yale University also participated in the study.

By using a series of ten synthetic /rok/-/lok/ stimuli in a battery of tests (forced-choice identification, AXB discrimination and oddity discrimination), the researchers were able to see a positive relationship between the participants’ exposure to English and their ability to perceive the English liquids. While those with an average of 8.6 months of conversational English experience performed only a little better than chance in the tests, the experienced group was able to categorically identify the /u/ and /l/ sound at the same rate as English speaking controls had. The experienced group’s performance level on the discrimination tests, however, was somewhat lower than the control group. That being said, performance levels “on both the AXB and oddity tests, were more similar to those of the Americans than those of the NOT Experienced Japanese,” and “[a]lthough their discrimination performance was not as high as that of the Americans, both discrimination tests revealed an increase in correct performance near the /u/-/l/ category boundary” (MacKain, et al., 1981).

These results are in line with a similar study conducted by Flege, Takagi and Mann (1996), where Japanese adults who had lived in the United States for an average of 21 years were
able to identify the English liquids more accurately than Japanese adults who had lived in the United States for just two years.

It is interesting to note that in Miyawaki et al. (1975), results were contrary to the aforementioned findings. Here, 21 native Japanese speakers were only able to perform slightly higher than chance on discrimination tests, despite having an average of 10 years of formal English training. These findings indicate that interaction with a language in itself is not enough to predict acquisition of its phonology. Perhaps in this case, the quality of the interaction with the language played a role in the poor acquisition of the phonemes. Unlike in the previously mentioned studies, these subjects had not lived in an English speaking country, and there is no mention of any conversational English training.

It is also interesting to note that in Miyawaki et al. (1955), the researchers set out to determine whether the difficulty was related to the nature of the original stimuli. They proposed that by eliminating cues which were related to speech sounds, the subjects would be able to correctly differentiate between the stimuli. Researchers removed the first and second formant from the stimuli and re-administered the tests with only the third formant present. This caused the stimuli to sound not like speech, but “rather light high-pitched glissandos followed by a steady pitch” (Miyawaki et al., 1975, p. 333). Surprisingly, even though the subjects were unable to differentiate between the original stimuli, when they were only subjected to the third formant they were able to discern between them as well as a native-English control group. Keeping in mind that the third formant is considered to be the most important cue in categorizing the English liquids (MacKain, et al., 1981), and that the previously mentioned studies have indicated that Japanese speakers are less sensitive to it than the first two formants (Iverson, et al., 2003, MacKain, et al., 1981), these results are quite perplexing.
/ɹ/ vs. /l/

Research has also shown that Japanese learners of English are more likely to acquire the English /ɹ/ than /l/ phoneme. In Sheldon and Strange (1982), four Japanese speakers were asked to record themselves reading 46 words which contrasted the English liquids. A month following this, the recordings were played for the participants and they were asked to identify which phoneme they had heard. Although each of the phonemes was identified well above chance, exemplars of /ɹ/ were successfully identified more often than /l/ (96% vs. 82%). Flege, et al. (1996) also found that study participants correctly identified /ɹ/ over /l/. In their study, Japanese subjects who had lived in the United States for an average of two years were able to correctly identify /ɹ/ 87 percent of the time, but only were able to identify /l/ 53 percent of the time. These results were compared with those of Japanese speakers who had lived in the United States for an average of 21 years who were still unable to perform at the same level of accuracy as native-English speakers. Despite their prolonged English experience, they still were only able to successfully identify /ɹ/ 96 percent of the time, and /l/ 81 percent of the time.

Position of the English liquid

It seems that along with the previously mentioned variables, the position that the English liquid appears in a word also has an effect on a Japanese speaker’s ability to perceive it. In Sheldon and Strange (1982), the 46 words that the study participants had attempted to identify were further analyzed according to which position they appeared in. It was found that errors were highest when liquids appeared in consonant clusters, intermediate in the initial and medial position, and lowest in the final position. Mochizuku (1991) also found that /ɹ/ and /l/ perception varied depending on the position of the liquid in the words. While accuracy was greater than 95 percent in the final position, it was less than 65 percent for /ɹ/ in an initial consonant cluster.
Considering these findings (Sheldon and Strange, 1982, Flege, et al., 1996, and Mochizuku, 1991), it can be assumed that a Japanese learner of English would have the most difficulty understanding a word with an /l/ in an initial consonant cluster, and the least difficulty identifying a word with an /ɹ/ in the final position. Words such as ‘bleed’, ‘bloom’, ‘crown’, or ‘fly’ would be most challenging, whereas words such as ‘poor’, ‘bar’, ‘deer,’ or ‘steer’ would be the least.

The previous studies demonstrate that the perceptual difficulties faced by Japanese learners of English can be further confounded by several factors. It seems that although age cannot precisely predict when a Japanese learner’s perceptual abilities will decline, a negative relationship between age and the ability to master differentiating between the English liquids can be assumed. Moreover, it appears that once a learner has undertaken learning English, a great deal of exposure to the TL is necessary before the liquids can be perceived at native-like levels. Where these liquids appear in a word, however, and which liquid is being produced are also likely to influence a learner’s ability to correctly identify them.

Over the years, researchers have brought forward several theories which attempt to explain why Japanese learners experience difficulty perceiving the English liquids. In the following section, the most relevant of these theories will be discussed.

**Theoretical models which account for the difficulties faced by Japanese learners of English**

According to contrastive analysis (CA), differences between the first language and the second language result in learning difficulty. CA claims that L2 learners have more difficulty learning sounds which have no equivalent in their first language than sounds which resemble sounds in their first language (Vanpatten and Williams, 2008. p. 21).
For a variety of reasons, however, CA is not able to account for the phonological difficulties experienced by Japanese learners of English. As noted before, the English /l/ is more similar to the Japanese /r/ than the English /ɹ/ is. CA would then predict, that since the English /ɹ/ is more dissimilar from the L1 than the /l/ is, that Japanese learners would have more difficulty acquiring the /ɹ/ sound than the /l/ sound. However, as illustrated by Sheldon and Strange (1982), and Flege, et al. (1996), Japanese learners of English acquire the English /ɹ/ with more ease than the English /l/.

CA is also unable to predict transfer related errors due to the position of the English liquid. In the Japanese language, the Japanese /r/ occurs in the initial and medial position, but does not occur in consonant clusters or in the final position. CA would then predict that Japanese speakers would have more difficulty perceiving the English liquids in consonant clusters and in the final position. As discussed above, while this holds true for the perception of the phonemes in consonant clusters, it is unable to predict intermediate problems in the initial and medial positions. Furthermore, CA wrongly predicts that learners will have difficulty perceiving the English liquids in the final position, when in actuality the phonemes are easiest to identify in this position (Mochizuki, 1981; Sheldon and Strange 1982).

The Speech Learning Model (SLM), which addresses the limitations of CA was proposed by Strange (1995). Unlike CA, in SLM an increased distance between a phoneme in the L2 and the closest phoneme in the L1 will make a TL sound more learnable. The hypothesis is supported by findings from Sheldon and Strange (1982), and Flege, et al. (1996). In a similar study, Aoyama, Flege, Guion, Akahane-Yamada and Yamada (2004) set out to evaluate the validity of the SLM hypothesis by comparing a Japanese speaker’s ability to perceive and produce the English /l/, /ɹ/, /w/, /b/, /s/ and /o/ phonemes. These sounds were chosen because the /l/-/ɹ/, /r/-
/w/, /s/-/ø/ and /b/-/v/ contrasts were considered to be difficult for Japanese speakers. The /b/-/s/ contrast was also included in the study as a control. The participants were tested twice on their ability to differentiate between and produce these contrasts, with a 1.1 year gap between the two tests. Results demonstrated more improvement in identifying and producing the /ɹ/ over /l/ sound, supporting the SLM hypothesis. At the same time, however, results also demonstrated a significant increase in production ability of the English /w/ phoneme. Since Japanese phonology has a similar equivalent to the English /w/, SLM would predict its acquisition to be similar to the English /l/. This contradiction, however, could be due to the fact that phonemes which have a similar equivalent in the TL (English /w/) are acquired differently from phonemes which have more than one equivalent in the TL (English /ɹ/ and /l/) (Aoyama, et al., 2004).

This could be explained by Best’s (1991) Perceptual Assimilation Model (PAM), which states that as an infant, an L1 learner creates categories for its native sounds as it is learning to produce them. Thereafter, when the learner encounters a non-native sound it will be learned with more ease if it can be assimilated to an existing native category. Best defined this type of acquisition as ‘Two Category’ (TC) contrasts or ‘Category Goodness’ (CG) contrasts, and predicted it to be easier to acquire than other more difficult ‘Single Category’ (SC) contrasts. In SC contrasts, the TL contains a contrast where both sounds are identified as the same category in the L1 (Best, 1991). This would explain why /w/ was learned with relative ease in Aoyama, et al., (2004), while SLM was only able to accurately predict the unequal acquisition of /ɹ/ over /l/.

Similar to Best’s Perceptual Assimilation Model is Kuhl’s Native Magnet Model (NMM)(1992, as cited in Kuhl, et al., 2008). The NMM attempts to explain the process by which infants learn to master their native language. In this framework, Kuhl claims that in infancy L1 phonetic prototypes are developed which perceptually interfere with non-native sounds.
According to the model, this occurs in three phases. In the first phase, which lasts for the first six months of life, infants are considered to be receptive to the phonology of all human languages. At this time, they are able to discriminate between all of the sounds of all languages. The second phase begins in their sixth month, when it is thought that neurological changes allow infants to statistically absorb the phonological distribution of their first language. As these distributional representations of the native language stabilize, phonetic prototypes or categories are developed which statistically represent the phonology of their language. Language input in this second phase is thought to warp the phonetic perceptive abilities of the infants by decreasing their perceptive abilities near these phonetic prototypes and increasing their perceptual sensitivity near the boundaries between these categories. This second phase, described as the critical period for sound development, is believed to last from six to eight months of age (Kuhl, et al., 2008). Following the critical period, instead of taking statistics about the phonology of language input, phonetic perception of input is governed by the representations in memory which were formed earlier in development (Kuhl, 2011).

Best’s PAM, and Kuhl’s NMM provide an explanation for the difficulties experienced by Japanese learners of English. In Iverson, et al., (2003) for example, NMM explains why 17 of the 18 /l/ and /ɹ/ stimuli were heard as the Japanese /ɾ/. According to NMM, these sounds were magnetized to the already existing sound in the L1. Furthermore, this explains why subjects were able to acquire /ɹ/ over /l/ in Sheldon and Strange (1982) and Flege, et al. (1996). Since /l/ is phonetically closer to the Japanese /ɾ/ than /ɹ/, the /ɹ/ sound would be less likely to suffer from magnetic interference from the L1.
Empirical findings in support of Kuhl’s Native Magnet Model

Since its inception, several studies have been able to substantiate Kuhl’s NMM. In Tsao, Liu and Kuhl (2005), 37 American and 32 Taiwanese infants were tested on their ability to discriminate between the Mandarin /tɕʰ/ and /ɕ/ sounds. The participants were organized into groups according to their age, resulting in 19 American infants between six and eight months, 18 American infants between ten and 12 months, 16 Taiwanese infants between six and eight months, and 16 Taiwanese infants between ten and 12 months.

In their respective countries, the infants from each group participated in a Head Turn (HT) test. In such a test, the infants are trained to turn their heads when a sound changes. This is accomplished by sitting the infants on a parent’s lap and playing contrasting sounds over a loudspeaker. If the infants turn their heads when a contrasting sound is played, they are rewarded with a visual of a panda bear hitting a drum (Kuhl, 2011).

Between the ages of six and eight months, both the American and Taiwanese infants were able to discriminate between the two sounds at a success rate of 69 percent. The older group (10 to 12 months), however, discriminated between the two sounds at varying degrees of success. While the Taiwanese infants had improved their ability to differentiate between the sounds at a success rate of 74 percent, the American infants demonstrated a decreased level of proficiency with a success rate of 64 percent (Tsao, et al., 2005).

This diverging point validates the NMM. Prior to the critical period, the American infants could discriminate between the phonological sounds of Mandarin at the same rate as the Taiwanese infants. They were able to ‘hear’ the difference between the sounds, even though they hadn’t been living in a Mandarin speaking environment. Just two months later, however, their
ability to do so had already started to decline. It can be assumed that after these two months, the American infants were no longer taking phonological statistics on their language input, and were beginning to perceive sounds based on the phonology of their L1 (Tsao, et al., 2005).

**Training Japanese learners of English to improve their perception of /ɹ/ and /l/**

The findings in Tsao, et al., (2005) raise many questions regarding the learnability of the English liquids for Japanese learners. If the critical period for sound development has already ended at eight months, then what can be done for a learner who begins studying English at the age of five or 20?

As noted earlier, empirical evidence (MacKain, et al., 1981; Flege, Takagi and Mann, 1996) has demonstrated that despite a critical period for sound development, Japanese learners can improve their perception of the English liquids over time. Likewise, Flege, Takagi and Mann (1995) have demonstrated that with time, Japanese speakers of English are able to improve their production of the English liquids as well. In a study, the researchers compared the productive ability of two groups of Japanese learners of English by recording them pronouncing minimally paired English words beginning in /l/ or /ɹ/ (such as ‘read’ and ‘lead’). The first group was comprised of 12 native Japanese speakers who had lived in the United States for two years, and the second consisted of 12 native Japanese speakers who had lived in the United States for an average of 21 years. The recordings produced by the two groups were subsequently played for a group of native English speakers, and categorized as either /ɹ/ or /l/. While the recordings produced by the experienced Japanese group were almost always categorized correctly by the English speaking judges, the recordings produced by the inexperienced group were not. In their case, native speakers were only able to correctly identify /ɹ/ words 80 percent of the time and /l/
words 90 percent of the time (Flege, et al., 1995). The question then is not whether the English liquid phonemes can be acquired by Japanese learners of English, but what can be done to accelerate this process.

Strange and Dittman (1984) were the first to explore this question by attempting to train eight adult Japanese women living in the United States to correctly differentiate between a series of synthetic /rok/ - /lok/ stimuli. The subjects participated in 14 to 18 training sessions, where they were asked to identify whether pairs of stimuli were the same or different. In one training session, 126 of these pairs were presented. Immediately after each pair was categorized as ‘same’ or ‘different’ by the subjects, a light indicated whether they had chosen correctly or not. Over the 14-18 training sessions, results revealed gradual improvement by all of the participants but one. For these seven participants, they were able to discern between the /rok/-/lok/ series as well as or better than a native English control group. Moreover, in a synthetic /rake/-/lake/ posttest, five of the participants demonstrated an ability to transfer what they had learned in their previous training to the new stimulus. None of the participants, however, were able to transfer their skills to natural words contrasting /ɹ/ and /l/ (Strange and Dittman, 1984).

These results were later replicated by a similar study (Jamiesan and Morosan, 1986) which attempted to train French Canadian learners of English in the perception of the English /θ/ and /ð/. The study varied slightly from the previous by adopting a training technique which has come to be known as Perceptual Fading (Iverson, Hazan and Bannister, 2005). Like Strange and Dittman (1984), the participants in Jamiesan and Morosan (1986) were asked to differentiate between a synthetic series which contrasted the target sounds. In this case, however, the subjects were presented with minimal pairs which exaggerated the contrast at the outset of their training, and over time the perceptual contrast was gradually decreased. Using this technique, the
researchers found that after just 90 minutes of training, the participants had demonstrated improvement in their ability to discern between the target sounds. Moreover, unlike in Strange and Dittman (1984), results also indicated that the subjects were able to generalize this ability to natural speech. In 2002, however, McCandliss, Fiez, Protopapas, Conway and McClelland compared these two training methods in a study involving forty adult Japanese learners of English residing in the United States. Despite the claims made by Jamieson and Morosan (1986), the researchers found no evidence to support Perceptual Fading over the technique employed by Strange and Dittman (1984). Jamieson and Morosan (1986) also conceded that the generalizability experienced by their trainees may have been the result of the ubiquitous nature of the English language in French-Canada.

Following their research, both Strange and Dittman (1984) and Jamieson and Morosan (1986) concluded that training would be more effective if it were administered using a wider range of stimuli. Furthermore, the latter predicted that training would be more effective if it were “presented in an acoustic context that is appropriate for normal speech, rather than in isolation”, it should “involve identification with feedback rather than practice at discrimination” (Jamieson and Morosan, 1986, p. 214).

These suggestions were implemented by Logan, Lively and Pisoni (1991), in a study involving six adult Japanese subjects living in the United States. Rather than responding to synthetic stimuli, 207 minimal pairs were recorded by five native speakers of English which were exemplars of the /l/-/l/ contrast in English. As suggested by Jamieson and Morosan, the stimuli were played individually and the subjects were asked identify which minimal pair they had heard. Each training session involved recordings from just one of the five native speakers, and immediate feedback was given to the subjects each time they made a selection regarding the
stimuli. Following 15 training sessions lasting approximately 40 minutes each (completed over three weeks), posttests demonstrated significant perceptual improvement, as well as some generalization to untrained stimuli and a sixth untrained speaker (Logan, et al., 1991).

These results were subsequently replicated in a follow-up study by Lively, Logan and Pisoni (1993). The researchers hypothesized that by focusing exclusively on stimuli which contained the English liquids in the positions which were perceptually most difficult for Japanese learners, they could produce results which were similar to Logan, et al., (1991). In this, the study was recreated with the exception that only stimuli with exemplars of /ɹ/ and /l/ in the initial position, in initial consonant clusters, and in the medial position were used in the training tasks. Stimuli with exemplars of /ɹ/ and /l/ in the final position were not included. As noted above (Sheldon and Strange, 1982, and Mochizuku, 1991), English liquids in this position are the easiest to perceive by Japanese learners of English. Despite the absence of exemplars of /ɹ/ and /l/ in the final position during the training tasks, pre and posttests data showed improved perceptive ability among the participants across all positions of the English liquids. Likewise, in generalizability tests the subjects were able to differentiate between new exemplars of the English liquids in the final position with increased accuracy, even though they had not been prepared for them in the training tasks (Lively, et al., 1993).

To further substantiate the findings in Logan, et al. (1991), a second follow-up study was performed which involved 19 monolingual speakers of Japanese living in Kyoto Japan. The purpose of this study was to determine whether the results from Logan, et al., (1991) could be replicated with a larger number of participants who had never been subjected to the phonology of the TL before. Once again, after just 3 weeks of training, the participants demonstrated an increase in their perceptual sensitivity to the /ɹ/- /l/ contrast. By the posttests, the participants had
improved their accuracy by an average of 12 percent (from 65-77 percent), despite having never studied English before (Lively, Pisoni, Yamada, Tohkura and Yamada, 1994). Again in Japan, Yamada (1993) found that by extending the aforementioned training to 45 sessions (instead of 15), the participants experienced nearly perfect generalization to new stimuli and new speakers.

Finally, researchers set out to determine what impact the training methods introduced by Logan, et al. (1991) had on retention and productive ability. Following the procedure in Logan, et al., (1991), Lively, et al., (1993), Lively, et al., (1994) and Yamada (1993), researchers found that after 15 training session (3-4 weeks), participants did not experience any reduction in their perceptive abilities in follow-up tests performed three months after their training had been completed (Bradlow, Yamada, Pisoni and Tohkura, 1999). Moreover, following training, monolingual Japanese speakers were able to produce English words which demonstrated the /r/-/l/ contrast with increased accuracy. According to native English speaking listeners, the trainees were able to produce the English /r/ with 79 percent accuracy and the English /l/ with 63 percent accuracy after training. These scores represented an increase of four and seven percent, respectively. These results suggest that in addition to improving perceptual ability for a sustained period, perceptual training also has the capacity to transfer to the productive domain (Bradlow, Pisoni, Akahane-Yamada and Tohkura, 1997).

Studies which have been modelled after Logan et al. (1991) represent a turning point in training perceptual training methods. While studies prior to Logan et al. (1991) (Strange and Dittman, 1984; Jamieson and Morosan, 1986) have been coined ‘low variability phonetic training’ (LVPT), those modelled after it are named ‘high variability phonetic training’ (HVPT) (Iverson, et al., 2005). Naturally, research has also compared HVPT and LVPT to determine which is most effective. Unfortunately, despite the promising findings in the research
surrounding HVPT, a recent study was unable to uncover any significant differences between the results produced by LVPT and HVPT. Nonetheless, the researchers concluded that “[r]egarding the applied goal of aiding L2 phoneme learning, it thus appears that training with natural speech [HVPT] is currently the best method, because the signal processing techniques used here [LVPT] are more labor intensive and offer no additional gains in performance” (Iverson, et al., 2005, p. 3275). HVPT has also been criticised for its inability to motivate learners due to the time required to see results, and due to the monotony of repeatedly completing identification tasks. To overcome this, Handley, Sharples and Moore (2009) have suggested including oddity discrimination tasks in HVPT. In a preliminary study, the researchers were able to find some evidence to support the use of oddity discrimination training as a method of perceptual training for learners of foreign languages (Handley, Sharples and Moore, 2009, p. 5).

Discussion

Several studies have demonstrated that the perceptual difficulties experienced by Japanese learners of English are unavoidable. Evidence seems to suggest that after the first year of life, learners have already passed the critical period for sound development, making it no longer possible to acquire the phonology of a foreign language naturally. Best’s Perceptual Assimilation Model and Kuhl’s Native Magnet Model assert that this occurs because as an infant is learning its first language, the language’s phonology is categorized accordingly. Thereafter, foreign phonemes are perceived as either good or poor representations of the phonology of the first language. Those which are similar to the first language are ‘magnetized’ to representations already present in the mind, whereas those which are dissimilar are not. As a result, dissimilar phonemes are more easily acquired because they do not suffer from L1 interference.
In the case of Japanese speakers who are learning English, the /l/-/ɹ/ contrast causes difficulty because the Japanese /r/ sound is perceptually somewhere between the two English liquids. Since Japanese phonology only carries one liquid, and since the major cue for this liquid is in the second formant, Japanese learners have difficulty mapping the contrasting phonemes to the existing categories in their minds. Since the English /l/ is perceptually closer to the Japanese /r/ than /ɹ/ is, it is especially difficult to acquire. Regardless, since the cue for accurate perception of each liquid is in the third formant, both cause a great deal of difficulty. This is further confused by several factors, including age, previous interaction with English, the nature of the stimuli and its position in a word.

A variety of training methods have demonstrated success in training Japanese learners to perceive and produce the English liquids with increased accuracy, yet to this point the most accurate method has not been realized. Nonetheless, when preparation time is taken into account, High Variability Phonemic Training seems to be the most efficient choice.

Moreover, this method seems to be the most adaptable to language learning classrooms. In a classroom adaptation of HVPT, learners could attend daily classes for a minimum of three weeks and a maximum of nine weeks, and participate in identification tasks with immediate feedback after each stimulus is presented. For variety, oddity discrimination tasks could also be included in the training. Since HVPT calls for natural stimuli, teachers could simply read from a list of words which exemplify the phonemic contrasts and provide feedback after the students have made their selections. Since HVPT is generalizable to new words, as well as new positions in the words, only words exemplifying the most difficult positions need to be trained. Since HVPT calls for several speakers to provide the stimulus, in a language school at least five teachers would need to alternate teaching the HVPT course. Finally, prior to enrollment in the
course, learners could be counselled regarding the obstacles which are preventing them for accurate perception of the TL’s phonology. This would help to manage their expectations and increase motivation for learning.
References


