CHAPTER 10

Training non-native language sound patterns
Lessons from training Japanese adults on the English /l/-/l/ contrast

Ann R. Bradlow
Northwestern University

Introduction

During native language acquisition the infant progresses from a language-general to a language-specific state. The task of the native language learner can be characterized as a “tuning” of the learner’s phonetic system to the distributional patterns of sounds in the ambient language resulting in a self-reinforcing match between native talkers and native listeners (for extensive and up-to-date discussions of native language phonetic and phonological acquisition, see Peperkamp 2003, and accompanying articles). In contrast, during non-native language acquisition the learner must progress from a monolingual to a bilingual state. The task of an adult non-native language learner can be characterized as shift from a system that is tuned uniquely to the sound structure of the native language (and therefore mis-tuned to the sound structure of the to-be-acquired non-native language) to a flexible system that can be tuned to the sound structure of both the native and the non-native languages (Iverson et al. 2003). While the tuning required for native language speech perception and production acquisition develops spontaneously in response to exposure to the ambient language, the flexibility and “re-tuning” required for the acquisition of non-native language perception and production is usually rather effortful and could presumably benefit from explicit instruction. Accordingly, the goal of non-native training programs is to identify the conditions under which the most general and linguistically functional phonetic and phonological learning can be achieved by adult second language learners.

An important premise of the entire non-native language sound structure training enterprise is that the monolingual adult speech perception and production capabilities are sufficiently plastic to support the acquisition of non-native
language sound patterns. Indeed, a major goal of early training studies was to test the hypothesis that sensitivity to acoustic features that are not reinforced by linguistic experience is permanently lost over the course of normal language development (e.g., Pisoni, Aslin, Perey, & Hennessy 1982; Tees & Werker 1984; Werker & Tees 1984; see also Chapter 6 by Strange & Shafer, this volume). Within this theoretical context, numerous non-native language sound structure training studies were conducted on various sound contrasts with listeners from various native language backgrounds. These relatively early training studies tended to adopt auditory training methods that had been developed in the speech and hearing sciences and which focused on increasing sensitivity to fine-grained acoustic differences. Examples of these studies include training English listeners to perceive an “extra”, nonphonemic category along a voice onset time continuum (Pisoni, Aslin, Perey & Hennessy 1982), training Canadian French speakers on the English /θ#/–#/contrast (Jamieson & Morosan 1986, 1989; Morosan & Jamieson 1989) and training Chinese speakers on word-final /t/ and /d/ in English (Flege 1989). These studies achieved some success in modifying the listeners’ responses to the trained stimuli and, in some cases, to untrained stimuli that differed minimally from the trained stimuli thereby providing evidence against a strong interpretation of the hypothesis that the adult speech perception system is no longer plastic. However, at the same time, they began to reveal some limitations on adult abilities to acquire non-native speech sound contrasts. Most noteworthy in this regard is the exceptional difficulty encountered by studies that attempted to train Japanese listeners on the English /p#/–#/l/ contrast (e.g. Strange & Dittman 1984) using the auditory training techniques that had proved successful in the training studies described above.

Due to its unusual resistance to acquisition, the case of English /p#/–#/l/ contrast learning by adult Japanese speakers has been particularly well-studied and has effectively served as a testing ground for different non-native speech sound training approaches. Therefore, the goal of this chapter is to compare and contrast training approaches to this notoriously difficult case. This examination of a well-studied case of non-native contrast acquisition will serve as a base from which we will attempt to derive some general principles of non-native speech sound training that can be applied to a wide range of cases.

The chapter will begin by considering the nature of the problem that the English /p#/–#/l/ contrast poses for adult Japanese speakers. Although the focus of this discussion will be on a particular non-native contrast for learners from a particular native language background, it will serve as a convenient vehicle for pointing out the parameters that need to be considered when developing an adequate description of the learners’ “initial state,” prior to any training, for all cases of non-native language sound structure learning. Studies that tested various approaches to training Japanese speakers to perceive the English /p#/–#/l/ contrast will then be presented in the next section. This particular case serves as an effective means of comparing
and contrasting training approaches since it has been the subject of investigation for multiple training studies using different approaches, thereby setting the stage for an unusually well-controlled evaluation of different training procedures. Finally, the last section will present some general lessons that we can extract from this case and raise some additional questions that future research should address.

The object of training: What needs to be learned?

Over the course of the past two decades numerous empirical and theoretical developments have made it possible for us to describe in detail the nature and extent of the difficulties encountered by native speakers of one language in response to speech sounds from another language. These advances have consequently made it possible for us to provide adequate descriptions of the task of any second-language learner trying to acquire the sound structure of any non-native language, thereby allowing us to clarify the object of training and to understand exactly what needs to be learned in any particular case. Here we consider in detail the case of perception and production of the English /p/–/l/ contrast by Japanese speakers. This case provides a convenient illustration of the parameters of cross-language comparison that need to be considered in order to understanding the nature of non-native speech sound learning.

It has long been noted that native speakers of Japanese have extreme difficulty perceiving and producing the English /p/–/l/ contrast. Several studies have provided experimental data that identify the precise conditions under which the perceptual difficulty of Japanese speakers with English /p/ and /l/ is manifested. The broad conclusion to be drawn from these empirical studies is that under controlled laboratory conditions, Japanese listeners generally exhibit great difficulty identifying and/or discriminating stimuli that exemplify the English /p/–/l/ contrast, but there is considerable variability in perceptual accuracy across individual listeners and across stimulus types. For example, Miyawaki, Strange, Verbrugge, Liberman, Jenkins, and Fujimura (1975) showed that American English listeners exhibited categorical perception along a synthetic /α/–/l/ continuum in which only the third formant varied; but Japanese listeners showed continuous perception along this speech continuum.¹ That is, the Americans exhibited a peak in

¹. Categorical perception is the phenomenon according to which listeners perceive sounds that differ from each other in terms of equal steps along a continuum as belonging to either one or another category. In contrast, continuous perception of sounds along a continuum is observed when listeners’ conscious perception of the sounds is analogous to their physical difference, that is, all differences are perceived and the sounds are not “forced” into one or another category. Categorical perception is typically assessed by testing (a) the consistency with which subjects
discrimination accuracy for stimulus pairs that straddled the category boundary as determined from an identification test, but the Japanese showed uniformly poor (though above chance) discrimination for all pairs along the synthetic speech continuum. In this same study, the American and Japanese listeners performed virtually identically on a discrimination task with non-speech stimuli that consisted of the isolated third formant (F3) component, suggesting that the effect of language background is limited to complex, synthetic speech stimuli and does not extend to relatively simple, non-speech stimuli. Similarly, Iverson et al. (2003) demonstrated that Japanese listeners have some degree of sensitivity to the acoustic differences between English /t/ and /l/ exemplars even if they tend to classify them all as members of a single phoneme category. This pattern of results demonstrates that, rather than having “lost” sensitivity to the acoustic features that cue the English /t/-/l/ contrast at a basic auditory perceptual level, Japanese listeners have instead learned to effectively “ignore” this difference during speech perception resulting in a perceptual space that is “mis-tuned” to the English /t/-/l/ contrast (Iverson et al. 2003).

In response to naturally produced words exemplifying the English /t/-/l/ contrast, Mochizuki (1981) found varying levels of identification accuracy by Japanese listeners depending on the position of the /t/ or /l/ in the words. Identification accuracy ranged from greater than 95% for /t/ and /l/ in word-final position to less than 65% for /t/ in a word-initial consonant cluster. Although the American English listeners also showed some variability in performance as a function of position in the word, the native listeners showed consistently more accurate /t/ and /l/ word identification than the non-native Japanese listeners. Finally, substantial individual listener differences in /t/-/l/ contrast perception have been observed even across native Japanese listeners with apparently comparable language backgrounds (Yamada & Tohkura 1992; MacKain, Best & Strange 1981). Together, these findings indicate that under certain circumstances Japanese listeners are sensitive to the

2. Formants are amplitude peaks in the spectra of vowel and other sonorant sounds, including /r/ and /l/. Formant frequencies are directly related to the articulatory configuration of the vocal tract during speech production. The third formant frequency (F3) is a major cue for the /r/-/l/ distinction with a low F3 frequency providing a strong indicator of the presence of an /r/ articulation.
acoustic differences between English /s/ and /l/; however, in general, their perceptual responses to this linguistic contrast of English are substantially less accurate from a linguistic functional point of view than the responses of native English listeners.

The most obvious source of the Japanese listeners' trouble English /s/ and /l/ perception is at the level of phoneme inventory structure. Whereas English has four contrasting approximant categories (/s, j, w, l/), Japanese has just two contrasting approximants, /j/ and /w/ (Handbook of the IPA 1999; Vance 1987). Thus, when a native Japanese speaker is presented with the English system of sounds, English /s/ and /w/ can be quite well mapped onto Japanese /j/ and /w/, respectively. However, the two English alveolar approximants, /s/ and /l/, do not map well onto any contrasting Japanese approximant pair. Instead, by virtue of similarity on other features (voicing, place of articulation), both of these English phonemes are identified by Japanese listeners rather unsystematically as the Japanese apico-alveolar tap /ɾ/, the Japanese labio-velar approximant /w/ or the Japanese high back unrounded vowel /u/ (Best & Strange 1992; Yamada & Tohkura 1992; Mochizuki 1981; Guion, Flege, Akahane-Yamada & Pruitt 2000). Thus, in order to acquire the sound structure of English, a Japanese speaker must learn to organize a poorly distinguished pair of sounds into two contrasting phoneme categories.

Current models of non-native language perception (Perceptual Assimilation Model (PAM): Best 1994, 1995; Best et al. 1988, 2001; Native Language Magnet (NLM) model: Grieser & Kuhl 1989; Kuhl 1991, 1992; Kuhl & Iverson 1995; Speech Learning Model (SLM): Flege 1995, 1999, 2002, 2003) all offer formalizations of this basic conceptualization of the English-Japanese alveolar approximant mapping at the level of phoneme inventory structure (for additional discussion of these models, see also Chapter 2 by Ioup, Chapter 6 by Strange & Shafer, and Chapter 8 by Zampini [the latter for Flege's SLM]). In particular, these models capture the important insight that non-native contrasts are not uniformly poorly perceived. Instead, the difficulty with which a particular non-native contrast is perceived by listeners from a particular native language background depends on the relationship between the phoneme inventories of the two languages in question. All three models agree that the case of Japanese speakers and the English /s/-/l/ contrast is an example of the most difficult kind of non-native contrast to acquire due to the fact that the organizing perceptual framework of the native language (Japanese) results in both English /s/ and English /l/ being identified with the same Japanese category (or categories). Best’s Perceptual Assimilation Model (PAM) is explicit in identifying this kind of contrast, a “Single Category” (SC) contrast, as the most difficult kind of contrast for non-native listeners to acquire. According to PAM, SC contrasts are predicted to be more difficult than “Two Category” (TC) or “Category Goodness (CG) contrasts in which the members of a contrasting pair are
assimilated by non-native listeners into two separate native categories or into a single native category with different degrees of goodness-of-fit, respectively.

Furthermore, in the production of English /s/ and /l/, the primary acoustic difference between the realization of these phonemes is in the higher formants. For /s/, the third formant frequency can dip below 2000 Hz; whereas, for /l/ the third formant frequency is in the neighborhood of 2400 Hz. Additionally, for /l/ (but not for /s/), the higher formants are substantially reduced in intensity. The exceptionally low F3 frequency for /tm/ is related to simultaneous constrictions in the pharyngeal and velar regions of the vocal tract as well as lip rounding. (For additional information regarding the acoustic properties of English liquids see Stevens 1998; Johnson 2003; Ladefoged 2003). In Japanese, the phonemes that are closest to the English liquids, /s/ and /l/, in terms of their acoustic features are the apico-alveolar tap, /ts/, the palatal approximant, /j/, the velar approximant, /w/, and the high back unrounded vowel, /u/. None of the contrasts represented by this group of phoneme categories (or for that matter, any of the Japanese phonemes) requires auditory attention to the combination of frequency and intensity features that cues the English /s/–/l/ contrast. Therefore, as a consequence of Japanese listeners’ lack of experience attending to this particular combination of acoustic-phonetic features, Japanese listeners can be expected to have great difficulty in tasks that require sensitivity to the distinguishing acoustic features of English /s/ and /l/.

It is important to note here that not all novel phoneme contrasts require the same degree of modification at the auditory-perceptual level as the case of Japanese listeners acquiring the English /s/–/l/ contrast. For example, Best et al. (2001) report that the plosive versus implosive voiced bilabial stop contrast of Zulu was treated by the majority of American English listeners in their study as a clear single category (SC) contrast: the American English listeners generally classified both members of the contrasting pair as belonging to the single English /b/ category and showed poor discrimination of the two phones. However, acoustic analyses showed that the primary acoustic differences between these contrasting phonemes in Zulu were that the implosives had higher pitch and F1 frequencies in the early part of the following vowel, higher-amplitude bursts, and substantial pre-voicing in contrast to the small positive VOT for the plosives. This combination of acoustic cues is not entirely unfamiliar to American English listeners and the acquisition of this contrast would require modifications to the category boundary locations along a constellation of dimensions that are already functionally significant for the American English listeners. This situation stands in contrast to the required atten-

3. Note that the lip rounding feature of English /t/ production can be a useful characteristic to stress when teaching English pronunciation.
tion to a new constellation of speech signal dimensions for the acquisition of the English /p/-/l/ contrast by native Japanese speakers.

In terms of speech production, native Japanese speakers have little or no experience with the precise articulatory configurations required for English /p/ and /l/ production (see Gick et al., Chapter 11 of this volume, for discussion of these articulatory configurations). While the separate articulatory gestures involved may be represented in the inventory of native Japanese sounds, including retroflexion, lip rounding and even lateralization, the exact constellation of gestures for English /p/ and /l/ are likely to be novel for native Japanese speakers. Indeed, several experimental studies have demonstrated that native Japanese speakers generally have difficulty producing /p/-/l/ minimal pairs accurately enough for native American English listeners to identify them with a high degree of accuracy (Goto 1971; Sheldon & Strange 1982; Mochizuki 1981). However, somewhat surprisingly, it appears that for many Japanese learners of English, their ability to produce the English /p/-/l/ contrast exceeds their ability to perceive the contrast, particularly in the early stages of acquisition (Yamada, Strange, Magnuson, Pruitt, & Clarke 1994).

In summary, the English /p/-/l/ contrast presents great difficulty for native Japanese speakers due to extensive mismatches between the underlying systems of contrasting approximant categories of the two languages, the particular acoustic-phonetic features that listeners of the two languages have learned to attend to, and the articulatory configurations that talkers of the two languages have learned to produce. The available data on Japanese speakers’ perception and production of English /p/ and /l/ clearly demonstrate that this difficulty is general across individuals and is apparent in a range of speech perception and production tasks. However, despite this rather stark contrast between American English and Japanese listeners, Japanese listeners exhibit some sensitivity to the English /p/-/l/ contrast in terms of both perception and production. That is, Japanese listeners are apparently not entirely insensitive to the acoustic and articulatory dimensions that English speakers use to cue this contrast. Thus, the task of learning this novel contrast for a Japanese second language learner is a matter of developing a new organizational framework along existing phonetic dimensions rather than a matter of (re)acquiring sensitivity along acoustic and/or articulatory dimensions that were previously completely unattended to or ignored.

For the reasons described above, it is not surprising that the case of training Japanese speakers to acquire the English /p/-/l/ contrast has been met with remarkable resistance. Few other cases are likely to be as difficult to train since few other cases are likely to require such extensive modification by the learners. Indeed, the early successes of training studies all involved cases that differed from the /p/-/l/ case in some significant way. For example, the early training studies that focused on introducing an extra voicing category (e.g. Pisoni, Aslin, Perey & Hennessy 1982) were probably quite successful with very relatively little training due
to the fact that the listeners already had experience with categorization along the relevant acoustic-phonetic dimension (i.e. voice onset time). Similarly, Canadian French speakers being trained on the English /θ/-/ʌ/ contrast could potentially take advantage of their native language experience with a voicing contrast for other fricatives (Jamieson & Morosan 1986, 1989; Morosan & Jamieson 1989) and Chinese speakers being trained on word-final /t/ and /d/ in English could potentially take advantage of their native language experience with this contrast in other word positions (Flege 1989). Thus, in general, when designing or evaluating a training procedure it is important to first consider the nature of the learners’ task in terms of the relevant aspects of the phonetic and phonological structures of the native and the target languages.

Approaches to training: What can be learned?

The first indication that learning should be possible for this difficult case came from reports that Japanese listeners with extended immersion in an English speaking environment generally performed better on English /θ/-/ʌ/ perception and production tasks than inexperienced Japanese listeners (McKain et al. 1981; Flege, Takagi & Mann 1995, 1996; Best & Strange 1992; Yamada et al. 1994; Yamada 1995). Although there are virtually no reports of native-like performance, the fact that performance varies with amount of exposure to English even amongst individuals whose first exposure is at a relatively late stage (beyond childhood) is strong evidence that experience-dependent learning is possible even for this difficult case. Indeed, because of the well-documented difficulty of this particular case in both laboratory and natural settings, it has been upheld as the “gold standard” for proposed training approaches, and has served as a productive testing ground for general principles of learning and claims about adult neural plasticity. The focus here is on perception training procedures; however, it should be noted that the development of production training procedures is an active area of research as well (e.g., Catford & Pisoni 1970; Akahane-Yamada, Adachi & Kawahara 1995; Kewley-Port & Watson 1995; Dalby, Kewley-Port, & Sillings 1998; Dalby & Kewley-Port 1999).

In a seminal study that laid the groundwork for future non-native phoneme contrast training, Strange and Dittman (1984) attempted to train Japanese speakers on the English /θ/-/ʌ/ contrast using a training procedure that had proved successful in auditory training studies that aimed to increase listeners’ sensitivity to small differences between sounds. In particular, Strange and Dittmann (1984) adopted a training strategy that was used in a study demonstrating that American English listeners could be trained to discriminate within-category differences along a voice-onset time continuum (Carney, Widin & Viemeister 1977).
Chapter 10. Training non-native language sound patterns

The objective of this general training approach is to explicitly draw attention to the acoustic parameters that vary from one end of a synthetic speech continuum to the other and in so doing to enhance discrimination between items along the continuum.

A crucial feature of the overall design of this study was that following the discrimination training (with stimuli along a rock-lock continuum), subjects were tested on a different synthetic continuum (rake-lake), as well as on a minimal pair identification task using naturally produced /t/ and /l/ words. For example, subjects heard “rock”, and identified it as either “rock” or “lock.” Thus, this training study assessed the extent of any learning on the trained stimuli and task as well as the generalization of this learning to novel stimuli (i.e., stimuli not included in the training set) and a novel task (i.e., a task that was different from the training task).

Subjects were native speakers of Japanese who were recruited from an English as a Second Language program at the University of Minnesota. The subjects ranged in age from 25 to 33 years and had lived in the USA from 5 to 30 months. Although their levels of English proficiency varied widely at the pretest phase, all subjects reported difficulty with English /t/ and /l/ and all were highly motivated to improve their English skills. At the pretest and posttest phases, the subjects performed a minimal pair identification test with naturally produced stimuli (16 pairs of words produced by an adult male native speaker of American English), as well as identification and discrimination tests with the rock-lock stimuli (from the training phase) and a novel (i.e., untrained) rake-lake stimulus continuum. The training task was a same-different discrimination task in which subjects were presented with pairs of stimuli from the synthetic rock-lock continuum and required to respond by labeling a pair as either S (same) or D (different). Immediate feedback after each trial was provided during training. Subjects completed 14–18 training sessions conducted over the course of 3 weeks.

During training the Japanese subjects generally improved in their ability to discriminate stimuli along the synthetic rock-lock continuum. This improvement during training was evident in the posttest phase by a change towards greater categorical perception along the rock-lock continuum for seven of the eight Japanese subjects. However, it is important to note, that the Japanese subjects still differed from American English listeners in terms of their identification consistency and discrimination accuracy for stimuli along this rock-lock continuum. The Japanese subjects also showed more categorical perception along the rake-lake continuum at posttest than at pretest; however, the Japanese subjects exhibited considerably less categorical perception along this untrained rake-lake continuum than along the trained rock-lock continuum. In contrast to this move towards greater cat-

4. See Note 1 above.
egorical perception in response to discrimination training, the Japanese subjects showed no improvement in their ability to identify naturally produced /l/-/l/ minimal pairs from pretest to posttest. In other words, while the discrimination training in this study modified the Japanese subjects’ responses to synthetic stimuli, this change did not generalize to naturally produced words.

The Strange and Dittmann (1984) study is an example of a “low variability” training approach since training involved the presentation of stimuli representing only one /l/-/l/ minimal pair as produced by only one synthetic “talker.” In a further test of this general, low variability training approach, a recent study investigated whether Japanese listeners would acquire the English /l/-/l/ contrast through initial exposure to maximally differentiated, or exaggerated, category exemplars (i.e. exemplars of the English /l/-/l/ contrast in which the acoustic difference between /l/ and /l/ is maximized) followed by exposure to increasingly natural exemplars (McCandliss, Fiez, Protopapas, Conway, & McClelland 2002). The rationale behind this training procedure is as follows: provided that the exaggerated exemplars are discriminable at the start of training, and that the discrimination of exaggerated exemplars generalizes to less exaggerated exemplars, then by slowly decreasing the acoustic distance between the training stimuli, listeners should eventually be able to discriminate natural exemplars.

The stimuli for this study came from a synthetic /l/-/l/ continuum (road-load or rock-lock) that was created by editing samples of the words as produced by a male native speaker of American English. The continuum was constructed by calculating the spectral distance between the members of the minimal pair (based on a linear predictive coding (LPC) analysis at intervals of approximately 10 msec) and then adjusting the LPC coefficients to interpolate between and extrapolate beyond the two endpoints, yielding a well-sampled, extended /l/-/l/ continuum. The training task was an identification task in which the subject had to identify the initial segment of the test word as /l/ or /l/. The general design of this study tested the effects of two training variables: “adaptive” (i.e. training that begins with exaggerated stimuli and ends with more typical stimuli) versus “fixed” (i.e. training with typical stimuli only), and with feedback versus without feedback during training.

Following training, subjects in the adaptive training group showed more native-like identification and discrimination functions along the trained continuum than subjects in the fixed training group or subjects in the untrained control group. However, there were no significant differences between the two groups of trained subjects (adaptive vs. fixed) nor between either of these trained groups and the untrained control group when tested on a novel continuum (rock-lock for subject trained on road-load, or road-load for subjects trained on rock-lock.) The most dramatic effect revealed by this study was that, regardless of whether the subjects were initially exposed to exaggerated stimuli (adaptive vs. fixed training procedures), subjects who were provided with feedback during training made
Chapter 10. Training non-native language sound patterns

substantial gains towards establishing distinct /s/ and /l/ categories along both the trained and the generalization continua. Unfortunately, this test of generalization was severely limited in that it did not test generalization to a novel (i.e., untrained) talker or to a novel phonetic environment. It is therefore impossible to determine at this point whether the learning that results from this type of low-variability identification training with feedback is stimulus-specific or stimulus-general.

Although the results of the low-variability, discrimination training procedure of Strange and Dittmann (1984) and those of the low-variability, identification training procedure of McCandliss et al. (2002) showed some success in modifying the Japanese learners’ responses to synthetic /s/–/l/ continua, neither provided evidence that laboratory-based training could induce improved recognition of novel, naturally-produced English /s/ and /l/ words. Other studies have tested an alternative, “high variability” training approach that attempts to achieve this goal by exposing subjects to the full range of stimulus variability within each of the contrasting categories that the learner can expect to encounter in the real world. The first attempt at implementing this training approach (Logan, Lively & Pisoni 1991) began as a follow-up to the suggestion of Strange and Dittmann (1984) to expand the training procedure to cover a wider range of training stimuli (see also Jamieson & Morosan 1986 for a similar suggestion). Logan et al. (1991) also noted that this suggestion was consistent with work in visual stimulus classification demonstrating that training on highly variable stimuli promoted more accurate classification of novel, untrained stimuli than training with a low variability stimulus set (Posner & Keele 1968). Moreover, in a departure from the focus on categorical versus continuous perception of /s/–/l/ continua for American English versus Japanese listeners, respectively, the high variability training procedure involves a training task that more closely matches the task of word recognition that occurs in real-world spoken language processing. Specifically, the training task and stimuli require the listeners to classify a wide range of naturally produced and highly variable words exemplifying English /s/ and /l/ into broadly defined categories.

The overall design of the first test of the high variability approach (Logan et al. 1991) included pretest, training, and posttest phases. In all phases, the subjects performed a minimal pair identification task in which they heard a single word and had to identify it as either the /s/ or /l/ word from an /s/–/l/ minimal pair. For all tests, the stimuli were naturally produced words by native American English talkers that placed /s/ or /l/ in various positions in the word (e.g., right-light, pray-play, bear-bell, bard-bald). At pretest, the subjects performed the minimal pair identification test without feedback using the word list from Strange and Dittmann (1984) as produced by one male talker. During the training phase the subjects performed the minimal pair identification task with stimuli produced by 5 talkers (3 males and 2 females). The training stimulus set included 68 minimal pairs (a total of 136 stimuli) none of which were included in the pretest. During training, the subjects
were provided with immediate feedback. At posttest, all subjects performed the same minimal pair identification test as at the pretest phase. In addition, a subset of the subjects also performed two tests of generalization: the first presented a novel set of words produced by one of the talkers that produced the training stimuli, while the second presented a novel set of words produced by a novel talker. Subjects were 6 native speakers of Japanese who were students at Indiana University. They had lived in the USA from 6 to 36 months.

The results of this training study showed significant minimal pair identification improvement from the pretest phase to the posttest phase for all subjects, and for those subjects who performed the generalization tests, this learning showed some generalization to novel, untrained stimuli and a novel, untrained talker. In a follow-up study Lively et al. (1993) demonstrated similar learning when the extent of the phonetic context variability of the training stimuli was reduced to include only words that place /s/ and /l/ in the most difficult positions in the word (i.e., in pre-vocalic positions where pretest performance is poorest.) Stimuli with /s/ and /l/ in post-vocalic positions were eliminated from the training stimulus set due to high identification accuracy for such words even at pretest. However, this same study demonstrated that a reduction in the extent of talker variability in the training stimulus set to just one talker (instead of five talkers) did not lead to substantial improvements in identification accuracy (however see Magnuson et al. 1995 for evidence that training on some individual talkers can be as effective as multiple talker training). These results suggest that exposure to multiple talkers during training is effective for achieving general, rather than stimulus-specific, learning. However, training may be optimized by focusing only on phonetic environments that are known to be difficult at the pretest phase.

This pattern of results was replicated and extended to monolingual Japanese subjects who had never lived in an English-speaking country, and the perceptual learning that resulted from the high variability training procedure was shown to be retained for a period of at least 6 months with no additional training (Lively et al. 1994). A subsequent study showed that if the training period continued to the point were the average learning curve “leveled off,” (i.e. 45 rather than 15 sessions of approximately 30 minutes each), essentially perfect generalization of the perceptual learning to novel, untrained words and to a novel, untrained talker could be attained (Yamada 1993). Finally, the generalized perceptual learning that was induced by the extended (45 session) high variability training procedure transferred from the perceptual domain to improvements in /s/–/l/ contrast production by the Japanese trainees (Bradlow et al. 1997, 1999). While all of the above-mentioned high variability training studies used the same stimulus set and training procedure, a separate study using new stimuli but the same overall high-variability approach, replicated and extended these learning patterns for the English /s/–/l/ contrast to
training in combined audio and visual modalities and to both Japanese and Korean speaking adults (Hardison 2003).

Taken together, the series of studies on training Japanese speakers to identify English /p/ and /l/ using a high-variability training approach proved conclusively that robust, linguistically-functional learning can be achieved under laboratory training conditions even for this unusually difficult case. Provided that the training phase continued to the point of saturation, the learning demonstrated by these studies was not specific to the training items, was resistant to decay over time, and extended beyond the perception domain to the production domain. This high degree of success of the high variability training approach stands in contrast to the limited success of the low variability training approach which, as far as could be determined, was stimulus and task specific (i.e., did not generalize to stimuli and task that were not part of the training procedure). It is important to note here that the high variability approach could not have been devised without the groundwork laid by prior low variability training studies. In particular, the overall design of Strange and Dittmann (1984), which emphasized the importance of testing the generalization of training-induced learning beyond the specific stimuli and task used in training, was a critical step in the development of non-native contrast training approaches.

Following the success of the high-variability training approach with the difficult case of training Japanese speakers on the English /p/-/l/contrast, several other non-native contrast training studies adopted the high variability approach and showed similar learning patterns. These studies include training of English listeners on Chinese lexical tone contrasts (Wang, Spence, Jongman & Sereno 1999; Wang, Jongman & Sereno 2003), training of English and Japanese listeners on Hindi dental and retroflex stops (Pruitt 1995), training English listeners on Japanese vowel length contrasts (Yamada, Yamada & Strange 1996), training Chinese listeners on English word-final /t/ and /d/ (Flege 1995), and training English listeners on various German vowel contrasts (Kingston 2003). These studies have all demonstrated substantial learning in response to high variability training and, in those cases that tested generalization, results showed good to excellent generalization to novel, untrained talkers and stimuli. While all of these studies involved a high variability approach with respect to the stimuli used during training (multiple words produced by multiple talkers), they differed somewhat with respect to the training task (identification versus discrimination) and sequence of stimulus presentation during training (unstructured versus gradual introduction of more various and challenging stimuli). In particular, in the training of Chinese

5. For a more recent and direct comparison of training methods for this particular case, see Iverson, Hazan & Bannister 2005.
speakers on English word-final /t/ and /d/, Flege (1995) directly compared two training tasks, identification and categorial discrimination (in which the stimuli presented for discrimination are always members of contrasting categories rather than members of the same phoneme category). Both training tasks resulted in significant learning and generalization to novel stimuli. In the training of English and Japanese listeners on Hindi dental and retroflex stops, Pruitt (1995) adopted a “fading” stimulus presentation scheme, in which training began with a limited set of easily identified stimuli and, as the subject’s performance improved, additional and more challenging stimuli were gradually introduced. In this study, while the listeners from the two native language backgrounds showed different levels of performance at all stages (the Japanese listeners always performed better than the English listeners), both showed significant improvements in response to training and this perceptual learning generalized to novel, untrained stimuli.

General lessons and future directions

As in the study of many physical and psychological systems, it is often highly instructive to consider the extreme cases. The case of Japanese speakers’ difficulties with English /s/ and /l/ has in many respects served this purpose in the development of non-native speech sound training programs and has therefore been the major focus of this chapter. We conclude by identifying three general lessons to be learned from the rich history of research on training Japanese speakers on the English /s/-/l/ contrast.

Lesson 1

Laboratory-based training can lead to successful non-native contrast learning even for the most difficult cases. Even though native-like performance may be an unattainable goal for non-native language sound structure training programs, robust and highly generalized improvements in speech perception and production can be attained by adult learners with extremely limited prior exposure to the target language. This now well-established fact contributes an important line of evidence against a strict interpretation of the hypothesis that, in the absence of early language exposure, certain sensorineural sensitivities are permanently lost. Instead, it appears that the ability to modify speech perception and production patterns is retained well into adulthood. This claim is made explicit in the Speech Learning Model (Flege 1995) and argued for extensively in much of Flege’s recent writings (Flege 1995, 1999, 2002, 2003).

The open questions that current research on the issue of neural plasticity for speech learning should continue to address are: (1) What levels of processing
and representation are shaped by early language exposure? That is, where along the pathway from lower level sensorineural encoding to higher level, linguistic processing does the effect of linguistic experience become evident? (See for example, Iverson et al. 2003, Cheour et al. 1998; Bent, Bradlow & Wright 2006) (2) Is early exposure necessary and sufficient to induce native-like speech perception and production of a non-native language, and how does continued native language exposure and use interact with non-native language acquisition? (See for example, Flege & MacKay 2004; Flege, Frieda & Nozawa 1997; Pallier, Bosch & Sebastián-Gallés 1997; Pallier, Colome & Sebastián-Gallés 2001; Mayo, Florentine & Buus 1997) (3) What is the relationship between initial non-native language speech perception abilities and training-induced learning? This question pertains both to the causes and consequences of individual differences across learners from the same native language background, as well as across learners from different native language backgrounds in response to a given non-native language. In the case of the former, variables such as age and conditions of initial exposure are of interest. In the case of the latter, learning patterns generated from models of non-native contrast perception (Flege’s Speech Learning Model, Kuhl’s Native Language Magnet Model, and Best’s Perceptual Assimilation Model) can be identified and tested (see for example, Polka & Bohn 1996; Guion et al. 2000; Bohn & Polka 2001; Kingston 2003).

Lesson 2

The essential goal of non-native language contrast acquisition is accurate recognition of words that exemplify the contrast in the target language rather than native-like patterns of categorization along acoustic-phonetic continua. Although the effect of training with naturally produced words on categorization along the relevant acoustic dimensions has not been examined, it is possible (even probable) that non-native listeners develop functional, non-native language category representations for the purposes of word recognition in the absence of native-like sensitivity to specific acoustic features of the speech signal. Conversely, as demonstrated by the low-variability training reviewed earlier in this chapter, non-native listeners may develop more native-like patterns of categorization but still show highly inaccurate word recognition (e.g. Strange & Dittmann 1984). It is very likely that a key to successful non-native sound structure acquisition is to focus on the perception of even more contextualized speech samples than isolated words, such as full sentences and larger discourse units (e.g. Hirata 2003, 2004). However, the cost of introducing greater processing requirements must be examined in relation to the benefit of presenting the non-native contrast under acquisition in the context of a meaningful linguistic unit instead of in isolation.
Lesson 3

Exposure to highly variable training stimuli promotes, rather than interferes with, non-native contrast acquisition. In particular, exposure to multiple talkers appears to be a highly effective means of ensuring that perceptual learning generalizes to novel talkers. This general principle of the high variability training approach has received further support from studies of speech learning at a more global level than the level of phoneme category contrasts. For example, native English listeners exposed to multiple talkers of Chinese-accented English during training were able to generalize their learning of this particular accent to a novel (i.e. never before heard) talker of Chinese-accented English; whereas, listeners exposed to a single talker during training showed only talker-specific learning (Bradlow & Bent 2003). Similarly, in an American English dialect classification task (in which listeners are asked to identify the region in the USA from which the talker comes) a group of native English listeners who had been exposed to multiple talkers from each region were better able to categorize a set of novel talkers than a comparable group of listeners who had been exposed to just one representative talker from each region (Clopper & Pisoni 2004). This positive effect of the high variability training approach on speech category learning is consistent with exemplar-based models of speech perception (Goldinger 1996; Johnson 1997; Pierrehumbert 2001, 2002, 2003a, 2003b) in which item-specific acoustic-phonetic variability is encoded in the cognitive representation of experienced speech samples. Moreover, the patterns of learning and generalization revealed by dialect/accent and non-native phoneme contrast training studies such as those discussed in this chapter provide crucial information regarding the dimensions over which linguistic and paralinguistic generalizations are formed, and about the structure of an exemplar-based phonetic category system. Thus, speech training studies represent an area of research with unusual importance in both theoretical and practical arenas.

Suggested readings


Chapter 10. Training non-native language sound patterns


References


Chapter 10. Training non-native language sound patterns


Chapter 10. Training non-native language sound patterns


