TECHNIQUES FOR TRAINING DIFFICULT NON-NATIVE SPEECH CONTRASTS

D.G. Jamieson
Hearing Health Care Research Unit
Elborn College, University of Western Ontario
London, ON CANADA, N6G 1H1

Abstract
This paper examines the prospects for applying systematic procedures to train adult second language (L2) learners to perceive new speech contrasts. Four techniques that can produce generalized improvements in such speech perception are reviewed. Future research to optimize such training is then discussed.

Introduction
For adult second language (L2) learners, speech perception and production are strongly influenced by one's first language (L1). Early reports that training did not improve performance substantially, were seen as evidence that the capacity to learn new speech contrasts declined irreversibly, with age.

Within the past decade, new studies applying systematic approaches to train non-native speech contrasts have demonstrated that there is considerable ability to learn to perceive new speech contrasts, at least until early adulthood [1,2]. While the body of this work remains quite limited, the progress already made is extremely encouraging. This paper summarizes four successful training approaches that can facilitate the acquisition of L2 speech contrasts, and attempts to identify promising directions for future work in this field.

Effects of L1 on L2 Performance
When substantial exposure to an L2 is delayed until adulthood, one's ability to perceive and produce certain L2 speech sounds will be limited. These effects involve complex interactions among the learner's age when substantial exposure to the L2 begins, the sound patterns of the L1 and L2, the amount and type of exposure to L2, and the listener's individual perceptual skills and learning abilities [3,4,5]. For example English "th" voicing distinctions cause difficulty for native speakers of French; English /r/ and /l/ are difficult for native speakers of Japanese, Korean, and Cantonese; and the Hindi dental-retroflex consonant and French [u]-[y] vowel distinctions are difficult for English speakers [1,6,7,8].

Such difficulties may remain even after many years in the L2 environment, and they appear to be little influenced by traditional language training. When measured in terms of the accuracy with which the target sounds can be identified in high-quality recordings of isolated L2 words, under favourable listening conditions, error rates of 20% to 40% or more are common even for adults with several years of L2 experience; native speakers achieve virtually 100% accuracy under these circumstances. Moreover, the performance of L2 learners declines rapidly in more typical real-life listening environments, such as in multi-talker noise or reverberation [9].

More detailed consideration of the complex interactions among L1, subject, experience, and L2 variables is provided in the other papers from this session. Relevant data are also provided in [3,10,11] and useful theoretical perspectives are provided in [7,12].

Development of Speech Perception Abilities
Developmental studies of speech perception abilities have established the general principals that: 1) very young infants can discriminate most phonetic contrasts; 2) the ability to discriminate non-native contrasts declines rapidly within the first year of life; and 3) reduced discrimination ability reflects a change in attention to acoustic cues, rather than a loss of sensitivity to the acoustic cues [13]. Developmental studies therefore encourage the notion that many difficulties with L2 speech perception may be overcome with appropriate training techniques.

Characteristics of Successful Training Approaches
Much work remains to be done to optimize training techniques for new speech contrasts. However, we do know that training is more likely to be effective when the training task is designed to:

1) **focus the listener's attention on how acoustic patterns are mapped into phonemic categories in the L2.** That is, training tasks should direct attention to the acoustic factors that define each L2 category, while suppressing attention to acoustic cues that are irrelevant for classification in L2 [1,2]. This focus can be achieved by requiring listeners to identify (categorize) target sounds from among a set of candidate tokens. Discrimination tasks will not normally improve listeners' abilities to categorize sounds using L2 categories, as such tasks encourage attention to even phonetically-irrelevant differences between stimuli.

2) **provide prompt, unambiguous feedback concerning the L2 category appropriate to each training token.** Effective feedback can simply inform the listener about the accuracy of each judgement as soon as it is made. For example, the correct response can be indicated using a light, or by otherwise highlighting the display. One or more repetitions of the correct signal may also accompany this indicator. Mere listening to a sequence of samples from a specified category may also be effective [14].

3) **expose subjects to an adequate range of acoustic variation during training.** Subjects must learn not only about the acoustic cues that define and differentiate categories in the L2, but also about the range of variation that is tolerated within each category [1,15]. If enough is known about the relevant acoustic cues, synthetic speech signals may be used to direct listener attention to certain important cues from the outset of training. Alternatively one may use multiple tokens of the target sounds, spoken by several talkers, to provide a range of variation in the training set. Failure to provide sufficient variation restricts learning and/or reduces transfer outside the training environment [15, 16].

Measuring Performance of L2 Learners
A fundamental consideration for studies of L2 acquisition is how the L2 learner's speech perception and/or production performance will be measured. Assessment of performance typically focuses on the subject's ability to identify the presence of sounds from each of a few target categories, when listening to isolated L2 words. Production is assessed more rarely, and typically involves rating the quality of the learners' utterances, by native speakers.

Training might certainly be expected to improve performance when the test task and conditions are the same as those used in training. It is therefore important to ask **to what extent does training generalize to new and different tasks and conditions?** For example, to what extent does learning generalize to new speakers, to new words containing the target sounds, to words in which the target sounds occur in new (untrained) phonetic environments, to target sounds that share a feature with the training sounds, to other listening conditions (e.g., in conversation, in fluent speech, or when listening to degraded speech), and to the production of target sounds under various conditions? It is also important to establish the extent to which performance changes endure once formal training has been discontinued.

This review considers training techniques that have demonstrated such generalization of learning.

Successful Training Techniques

A fundamental finding of cross-language speech research is that many production difficulties have an underlying perceptual difficulty. The major effort has therefore been directed to improving the learner's speech perception abilities. At least four training techniques are known to produce relatively rapid improvements in listeners' abilities to perceive non-native speech contrasts. Studies have included several L1 groups and L2 targets.

Subject and specific language variables aside, the approaches can be differentiated in terms of several important variables: 1) the training task (eg., identification training); 2) the type and sequencing of training signals (eg., tokens from multiple talkers); and 3) the form of informational feedback used. These approaches are discussed in turn, below. Table 1 summarizes results from studies with these successful approaches.

Table 1
Summary of Training Studies Demonstrating Significant Improvement with Transfer to New Speakers

	L1 Group	L2 Target	Task	# Sess	Total Time (h)	# Trials	# Wks	Init. Perf. (%)	Final Perf. (%)	(%) Chg.	New Spk (%)
J&M '86	Fr.	/θ/-/ð/	F	2	~1.5	~720	1	62	92	48	-14
M&J '86	Fr.	/θ/-/ð/	F	2	~1.5	~720	1	61	92	52	-11
J&M '92	Fr.	/θ/-/ð/	F	4	<4	~720	1	70	89	27	-8
J & Y '95	Kor.-y	E /r/-/l/	NT	15	~4	1500	3	69	90	32	3
J & Y '95	Kor.-o	E /r/-/l/	NT	15	~2.5	720	3	63	75	20	2
Logan '91	Japan.	E /r/-/l/	NT	15	~10	4080	3	78	85	10	2
Pruitt '94	Engl.	H den-retr.	Lis & NT	30	~2	840	1	56	84	49	-17
Yamada '93	Japan.	E /r/-/l/	NT	45	~25	1224	9	70	89	28	-2
Flege '95	Mand.	E /t/-/d/	CD	7	~3.5	1680	3	66	77	16	-8
Flege '95	Mand.	E /t/-/d/	NT	7	~3.5	1690	3	67	83	24	-7

1. The Fading Technique. The first study demonstrating generalization beyond the training situation used synthetic stimuli to train native speakers of Canadian French to hear the English /θ/-/ð/ distinction [1]. Subjects were asked to identify each of a sequence of sounds as containing voiced or voiceless "th". Subjects received accuracy feedback immediately after each response.

Speech synthesis was used to create a sequence of signals, varying systematically in the amount of voiced or voiceless frication. At the start of training, listeners heard just two signals, one containing an exaggerated amount of the voiced target frication, and the other containing an exaggerated amount of the voiceless target frication. These "super-fricative" signals were designed to help the subjects attend to the target contrast immediately, and without making errors. As training progressed, additional signals with reduced amounts of frication were included in the training set, so that subjects gained experience with signals having more typical amounts of frication.

This approach improved the identification of tokens of natural speech; training with synthetic speech modelled on a male speaker generalized to natural tokens spoken by women as well as by men. Training with only a pair of "prototypical" speech signals was less effective than training with the full set of synthetic signals [16]. Moreover, training in word-initial position did not transfer to tokens containing the target sounds in other positions, nor to a task requiring subjects to identify sounds as containing one of *four* possible target sounds -- /d/ and /t/ and well as /θ/-/ð/ [17].

Targets in syllable-medial position contain cues common to those in word-final and word-initial position. However, training with syllable-medial tokens did not transfer substantially to target sounds occurring in word-final or word-initial position [18]. Thus, while fading with synthetic signals can improve perceptual ability substantially, learning seems specific to the phonetic environment in which the training sounds occurred.

2. Multiple natural tokens. Another successful technique for training new speech contrasts in adults involves the identification of multiple natural tokens of the target sounds. The first successful use of this technique used tokens from several talkers to train native speakers of Japanese to hear the English /r/-/l/ distinction [2].

Listeners identified each of a sequence of these tokens as being one of two words from a minimal pair, one containing an "r" and the other an "l". Incorrect responses were indicated by illuminating a light associated with the correct response, and then repeating the stimulus. Three weeks of such training improved identification performance by approximately 10%. When training used several talkers, learning generalized to novel words produced by a familiar talker and to a lesser extent to novel words produced by an unfamiliar talker (~8%; [2]). When training used a single talker, learning did not generalize to novel words produced by an unfamiliar talker [15]. Extending training to 9 weeks further improved performance [19].

The basic approach and stimulus set from [2] have also been applied to improve English /r/-/l/ identification for native speakers of the Korean language [20]. Training improved performance substantially for young Koreans, who had recently arrived in Canada; older Koreans who immigrated to Canada as adults received less benefit.

3. Alternating Listening & Identification Sets. A third successful approach was used to train English-speaking adults to identify Hindi dental and retroflex consonants [14]. This task cycled between sets of 50" listening trials" using a sequence of natural tokens of either dental or retroflex consonants, and sets of ten identification trials, each containing two repetitions of one of several possible dental or retroflex consonant tokens, followed by the subject's identification response, followed by accuracy feedback. This combination of listening and identification sets was repeated six times on each training day. The type of sound presented during each of the listening blocks was selected by the subject immediately prior to the block. Just one week of such training (~2 hrs) produced a 48% increase in average identification accuracy. Training also transferred to new words spoken by a new talker (with a 30% improvement from pretest performance).

This performance improvement is impressive, and it strongly encourages further work using this approach. Further research is required to evaluate the separate effects of, and interaction between, the listening task and the identification training task, used in [14].

4. Categorial Discrimination. Requiring listeners to categorize sounds in terms of the phonetic categories of the L2 is thought to be critically important for successful transfer of learning beyond the

training task. Normally, this focus on categorization is achieved by using an identification task. However, a **categorial** discrimination task (CD [21]) in which listeners must decide whether or not a pair of acoustically different signals are members of the same L2 phonetic category also focuses attention on linguistically-relevant categories. Thus, the CD task may also be effective for training new contrasts [22].

Recently, CD training was evaluated with native speakers of Mandarin for whom perception of unaspirated, English-language, word-final /t/ and /d/ is difficult [23]. Some subjects were trained with an identification task using multiple tokens of natural speech. Other subjects were trained in a categorial same/different task, in which two *different* tokens were presented on each trial, with the listener being required to indicate whether both were tokens from the same English-language category or whether the two tokens were from different English-language categories.

Three weeks of identification training improved identification accuracy by about 24%, while an equivalent amount of categorial discrimination training improved identification accuracy about 16%. Both types of training generalized to new tokens produced by a novel speaker; however, categorial discrimination training showed better retention, so that the performance of subjects from the two groups was more equivalent after a 2 month period without further training.

Training to Improve Speech Production.

Data relating systematic training to improved speech production abilities are presently very limited with adult L2 learners. However, training studies with young children who have difficulty producing sounds in their native language are encouraging. Many such children have correlated perceptual difficulties that identification training [24,25] improves.

These studies trained children in a "Category Inclusion" task, with feedback. Each child heard a series of speech sounds related to a specific production error manifested by the child. For each sound presented, the child indicated whether or not the sounds belonged to the target L1 category. Thus, a child who misarticulated /s/, would hear a sequence of misarticulated and correctly-produced utterances containing a target word containing /s/. For example, such a child could hear the word "shoe", spoken correctly and incorrectly, by many different speakers. The child touched a cartoon picture of a shoe, if a token was judged to have been pronounced correctly or a cartoon "X" when a token was judged to have been pronounced incorrectly.

The Category Inclusion task requires the subject to make an explicit judgement about whether or not a sound is appropriate for a particular linguistic category. Many more "inappropriate" sounds are included than in the standard "forced-choice" identification procedure. Importantly, such perceptual training not only improves children's identification performance, but it also transfers to speech production performance. However, this technique has not yet been evaluated with adult L2 learners.

Summary and Conclusions

When adults are trained using appropriate protocols, their abilities to perceive non-native speech contrasts can improve substantially. Appropriate protocols have the following characteristics: 1) they induce the learner to attend to cues relevant to assigning speech sounds to a phonemic category in the L2; 2) they provide prompt and unambiguous information about the appropriate categorization of each speech signal; 3) they use a set of speech tokens containing sufficient variability to permit subjects to learn about acoustic cues that are relevant to defining category membership in the L2 -- both cues relevant to inclusion of signals and cues irrelevant to classification in L2.

At least four training techniques that meet these criteria have now been demonstrated to produce effective training. All four can lead to improvements in speech perception that generalize beyond the training situation.

Such improvements have the following characteristics: 1) learning occurs quickly; 2) there is at least some transfer of training to the identification of novel (untrained) words, novel talkers, and novel phonetic environments; 3) there is minimal or no transfer to production; and 4) individual L2 listeners differ greatly in how much they benefit from training, even when age, linguistic background, and other factors are considered.

These results confirm that even in mature humans, those portions of the auditory system that are required to perceive and identify speech sounds remain relatively plastic. The speed with which such sizeable performance changes can be acquired suggests that training redirects attention more than inducing fundamental auditory system changes.

Optimizing Training: Directions for Future Research

For the L2 learner, the achievement of fluent, unaccented conversational speech is the ultimate objective. Performance clearly falls far short of this objective, even with extended training. However, there is much more reason to be optimistic, than there was a decade ago, when many speech researchers had reached the gloomy conclusion that adults had rather limited opportunities to learn new L2 contrasts because of permanent and irreversible neural changes.

Such a conclusion is no longer viable. However, it seems likely that we are still well short of what may be able to be achieved through systematic training. As impressive as the demonstrations reviewed here may seem, much can be done to further optimize training procedures. Answers to several questions are still required:

1). **How are listeners' L2 difficulties related to their L1 backgrounds?** Already, there is a substantial and growing body of work directed towards improving our understanding of this question. Good empirical work on this topic is very time consuming, but such work is continuing to appear [3,10]. Two significant theoretical positions which recently appeared on this topic [4,7] have helped to consolidate our understanding of empirical results. As this understanding improves, protocols for helping specific L1 groups to acquire specific L2 contrasts can be expected to become more effective.

2). **How can the differences between individual listeners, who have apparently similar linguistic backgrounds, auditory capabilities, etc., be understood?** This topic remains a challenge for many areas of speech perception. A particular question is how training can be personalized, through improved assessments of how a particular listener uses cues in the L1, and better targeting of the training stimuli and task for each listener. A relevant approach is provided in [26].

3) **How can protocols be structured to optimize learning and retention?** L2 training research remains largely in the "demonstration" phase, having an emphasis on determining whether or not some technique helps some L1 group to acquire some L2 contrast. Few studies have compared alternative training approaches and little protocol refinement has yet been attempted. Work may now be approaching a point of consolidation and refinement [8].

One very positive step has been the use of the stimulus set and training approach from [2] in several laboratories and across different subject populations.

Such sharing facilitates comparison across studies and should lead to more rapid advancement of knowledge.

4) **How should training be structured to optimize production ability?** The transfer of perceptual training to speech production has received such limited attention that few conclusions can be drawn. However, results with young children [24,25] suggest that such transfer may well be possible.

It seems unlikely that an "optimal" protocol can be created from a single one of the approaches examined to date, or indeed that any one protocol will be optimal for all listeners and situations. Rather, protocols will need to target subject needs, and components of each of several protocols may be used briefly at different points in time. For example, at the beginning of training, perceptual fading with synthetic signals may be used to direct the listener's attention to specific acoustic cues, without allowing others to vary [1]. Training with a structured sequence of natural tokens in a single phonetic context may then help the listener to classify sounds appropriately while ignoring irrelevant, naturally-occurring variation. The categorical inclusion task may then help the listener to further refine the L2 category. Training in additional phonetic environments may be appropriate. Finally, categorial discrimination may be used to consolidate learning and improve retention over the longer term.

ACKNOWLEDGEMENTS

*Supported by grants from NSERC and Starkey Canada. Direct e-mail to jamieson@audio.hhcru.uwo.ca.

REFERENCES 205-215.

[1]Jamieson, D.G. & Morosan, D.E. (1986), "Training non-native speech contrasts in adults: Acquisition of the English /θ/-/ð/ contrast by Francophones", *Perception & Psychophysics, 40(4)*, 205-215.

[2]Logan, J.S., Lively, S.E., & Pisoni, D.B. (1991), "Training Japanese listeners to identify English /r/ and /l/: A first report", *Journal of the Acoustical Society of America, 89*, 884-86.

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

[4]Best, C.T., McRoberts, G.W., & Sithole, N.M. (1988), "Examination of perceptual reorganization for non-native speech contrasts: Zulu click discriminations by English-speaking adults and infants", *Journal of Experimental Psychology: Human Perception and Performance, 14(3)*, 345-360.

[5]Flege, J.E. (1992), Speech learning in a second language. In Ferguson, C.A., Menn, L. & Stoel-Gammon, C. (Eds.) *Phonological development: Models, research, implications*, Timonium, MD: York Press.

[6]MacKain, K.S., Best, C.T., & Strange, W. (1981), "Categorical perception of English /r/ and /l/ by Japanese bilinguals", *Applied Psycholinguistics, 2(4)*, 369-390.

[7]Flege, J.E. (1991), Speech learning in a second language. In Ferguson, D., Mann, L., and Stoel-Gammon, C. (Eds.) *Phonological Development: Models, Research, and Application*, Parkton, MD: York Press.

[8]Rochet, B.L. (1994), "The efficient use of the computer in L2 pronunciation instruction", In Proceedings of the CALICO 1994 Annual Symposium on Human Factors, 178-182.

[9] Takata, Y. & Nábělek, A.K. (1990), "English consonant recognition in noise and in reverberation by Japanese and American listeners", *Journal of the Acoustical Society of America, 88(2)*, 663-666.

[10]Yamada, R.A. (1995), Age and acquisition of second language speech sounds: Perception of American English /r/ and /l/ by native speakers of Japanese.

In W. Strange (Ed.) *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research*, Timonium, (in press).

[11]Yamada, R.A. & Tohkura, Y. (1992), "The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners", *Perception & Psychophysics, 52(4)*, 376-392.

[12] Best, C.T. (1992), The emergence of language-specific phonemic influences in infant speech perception. In Nusbaum, H.C. & Goodman, J. (Eds.) *The transition from speech sounds to spoken words: the development of speech perception.* Cambridge, MA: MIT Press.

[13]Werker, J.F. and Pegg, J.E. (1992), Infant speech perception and phonological acquisition. In C.A. Ferguson, L. Menn, and C. Stoel-Gammon (Eds.) *Phonological Development: Models, Research Implications*, Timonium, MD: York Press. (pp 285-311).

[14]Pruitt, J.S. (1994), "Identification of Hindi dental and retroflex consonants by native English and Japanese speakers", *Journal of the Acoustical Society of America, 95(5) pt 2*, 3011.

[15]Lively, S.E., Logan, J.S., & Pisoni, D.B. (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*, 1242-1255.

[16] Jamieson, D.G. & Morosan, D.E. (1989), "Training new, non-native speech contrasts: A comparison of two techniques", *Canadian Journal of Psychology, 43(1)*, 88-96.

[17]Morosan, D.E. & Jamieson, D.G. (1989), "Evaluation of a technique for training new speech contrasts: Generaliztion across voices, but not word-position or task", *Journal of Speech and Hearing Research, 32*, 501-511.

[18]Jamieson, D.G. & Moore, A.E. (1991), "Generalization of new speech contrasts trained using the fading technique", *XII Internation congress of Phonetic Sciences, 5*, 286-289.

[19]Yamada, R.A. (1993), "Effect of extended training on /r/ and /l/ identification by native speakers of Japanese", *Journal of the Acoustical Society of America, 93*, 2391.

[20]Jamieson, D.G. and Yu, K. "Perception of English /r/ and /l/ by adult native speakers of Korean", (unpublished)

[21]Polka, L. (1991), "Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions", *Journal of the Acoustical Society of America, 89(6)*, 2961-2977.

[22]Strange, W. (1994), "Speech perception by second language learners" *Journal of the Acoustical Society of America, 95(5)* pt 2, 2998.

[23]Flege, J.E. (1995), "Two procedures for training a novel second-language phonetic contrast", *Applied Psycholinguistics* (in press).

[24] Rvachew, S. & Jamieson, D.G. Learning new speech contrasts: Evidence from adults learning a second language and children with speech disorder. In Strange ,W. (Ed.) *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research.* Timonium, MD: York Press (in press).

[25] Jamieson, D.G. & Rvachew, S. (1994), "Perception, production and training of new consonant contrasts with children having a functional articulation disorder", *Proceedings of the Third International Conference on Spoken Language Processing*, Yokohama: Acoustical Society of Japan, 1199-1202.

[26] Best, C.T. and Strange, W. (1992), "Effects of phonological and phonetic factors on cross-language perception of approximants", *Journal of Phonetics, 20*, 305-330.