PHONETICS OF SECOND-LANGUAGE ACQUISITION: PAST, PRESENT, FUTURE

Winifred Strange
Communication Sciences & Disorders
University of South Florida
Tampa, FL, USA

ABSTRACT

This paper summarizes the recent history of research on three issues in second-language (L2) phonetics: a) predicting the relative perceptual difficulty of L2 phonetic categories, b) describing the relationship between perception and production of L2 phones, and c) optimizing perceptual training to improve perception of L2 phonetic contrasts.

INTRODUCTION

The following three papers of this semi-plenary session report on three areas of research on the perceptual phonetics of second-language (L2) acquisition which have had a long and productive history. In this paper, I summarize the empirical progress on these topics over the last 25 years and discuss briefly how the theoretical and methodological issues have evolved.

THE PAST

The field is indebted to Arthur S. Abramson & Lee Lisker, who reported their seminal findings on cross-language differences in the perception of voice onset time (VOT) at the VIth ICPhS [1]. That study demonstrated that the ability to *discriminate* differences in VOT which underlie voicing and aspiration contrasts in initial stop consonants in many languages was predictable from language-specific patterns of phonetic labeling of the synthetic stimulus continuum. This finding fit well with the theoretical claims of Motor Theory [2] that the perception of speech sounds was accomplished via special processes that were intimately related to speech production. The *categorical perception* (CP) paradigm, which compared performance on tests of (physical identity) discrimination and (phonemic) identification of synthetically-generated acoustic continua, provided a rigorous methodological tool to examine these language-specific patterns of perception. Problems in perceiving non-native vowels were largely ignored in early cross-language research because (synthetic, steady-state) vowels were not perceived categorically and did not show language-specific patterns of perception [3].

Cross-language CP studies with adults in the 1970s replicated and extended the finding that discrimination of acoustic continua underlying voicing and place contrasts among consonants was determined by the phonemic significance of the stimuli in the listeners' native language [4,5]. At the same time, developmental research demonstrated that 2- to 6-month-old infants *could* discriminate place and voicing contrasts in consonants, whether or not they had been exposed to a language in which the contrasts occurred [6,7,8]. Thus, it was concluded that there was some *loss* in discrimination ability as a function of age or learning one's native-language phonology, or both.

This conclusion was reinforced by two additional kinds of data on the perception of non-native consonant contrasts by adults. One finding, which was first reported by Goto [9] see also [10], demonstrated that native Japanese speakers of English who had learned to produce /r/ and /l/ correctly nevertheless failed to distinguish these liquids perceptually when presented their own or native English speakers' productions, i.e., production preceded and exceeded (auditory) perception in L2 learning. Second, early training studies (using synthetic stimuli) which attempted to improve the perception of non-native contrasts met with limited success [11, 12,13]. While subjects' performance improved on training materials, generalization to new tasks and novel stimuli, including natural speech utterances, was limited. Thus, adult L2 learners' perceptual problems appeared to be serious and very long lasting, if not permanent. These results provided empirical support for the strong Critical Period Hypothesis proposed by Lenneberg [14].

In retrospect, these conclusions, based on a very limited number of phonetic contrasts, experimental paradigms, and subject groups, were premature, overstated, and in some respects, incorrect. Cross-language research in the 1980-1990s, which expanded the investigation to additional contrasts and subject groups using new stimulus materials and testing techniques, improved our understanding of the phenomena in all three areas of research.

Relative Perceptual Difficulty

Questions about the perceptual difficulty of an extended set of non-native contrasts were explored using carefully constructed natural speech materials (as well as synthetic stimuli) and an expanded variety of perceptual tests. For instance, Gottfried [15] demonstrated that both monolingual English speakers and experienced L2 learners of French had difficulty perceptually differentiating French front rounded vowels in a *categorial* (name identity) discrimination task (see Jamieson's paper). English listeners also had difficulty distinguishing French /e-ɛ/, which constituted a native phonemic contrast but whose members differed in phonetic detail (see also Bohn's paper). Werker and Tees [16,17] reported that adult English listeners had more difficulty categorizing a non-native place contrast than a non-native voicing contrast in Hindi stops; difficulty with the place contrast persisted even after one year of Hindi instruction [18]. Polka [19] further demonstrated that the Hindi place contrast differed in perceptual difficulty (for English listeners) as a function of the voicing context in which it occurred. Best and her colleagues [20] reported that both voicing and place contrasts among Zulu clicks were well discriminated by native English speakers, despite their being unlike any native phonetic categories. Thus, (initial) difficulty in perceiving both consonant and vowel contrasts ranged from minimal to extensive.

Experiments on the effects of L2 experience suggested that perceptual differentiation of non-native contrasts improved with immersion experience or intensive conversational instruction [21]. However, perception of some contrasts did not reach native-like levels even after years of experience. Furthermore, experiments using synthetic speech in which multiple acoustic cues for a contrast were manipulated independently indicated that L2 learners based their perceptual responses on different acoustic cues than native listeners. For instance, relatively inexperienced Japanese L2 learners of English appear to base their perceptual differentiation of (syllable-initial) /r-l/ more on temporal differences and on F2 spectral cues, than on the F3 spectral cue that is considered the primary differentiating parameter for native listeners [22,23]. Flege [24] reported that inexperienced Arabic learners of English assigned more perceptual weight to vowel duration than to consonant duration cues for voicing contrasts in final fricatives, whereas more experienced learners showed a native-like trading relation.

Developmental cross-language research continued to produce significant insights

with regard to the ontogeny of language-specific patterns of perception. Werker and her colleagues [25] published some remarkable experiments demonstrating that between 6 and 12 months of age, English-learning infants showed a decline in their ability to differentiate non-native Hindi and Salish consonant contrasts. More recently, Polka & Werker [26] reported the emergence of language-specific patterns of vowel perception at an even earlier age. It thus appears that native-language patterns of phonetic perception are formed in the first year of life. In addition, one study suggests that exposure to an L2 before the age of 2 years old has lasting consequences for later perceptual learning [18].

The data are not consistent concerning whether children between the ages of 2 and 13 years old have any advantage over adolescents and adults in the perception of non-native contrasts. Flege and Eefting [27] found that many (but not all) Puerto Rican children who started learning English at the age of 5-6 years had English-like perceptual boundaries on a VOT continuum, whereas older learners displayed perceptual boundaries that were a compromise between native Spanish and native English locations. However, other studies [28, 29, 16] failed to show better perception of non-native contrasts by preadolescent L2 learners.

Perception/Production Relationships

Although it is often assumed that perceptual difficulties lead to incorrect or accented production of non-native phonetic categories by L2 learners, until quite recently, there have been few studies that directly assess the relationship between perception and production in L2 learning (see Llisterri's paper). Rochet [30] demonstrated that the perceptual assimilation patterns of speakers of different languages are predictive of L1 substitution patterns in the production of French /y/. Portuguese speakers assimilated the non-native /y/ to their /i/ category while native English speakers assimilate the same stimuli to their /u/ category. These differences in perception accounted for production patterns, suggesting a causal relationship between perception and production of L2 phonetic categories, at least in the early stages of L2 acquisition. However, recent research by Yamada and her colleagues has shown that perception and production may proceed independently in L2 learning. In a study of a large group of Japanese learners of English with different amounts of immersion experience, perception generally lagged behind production such that perception mastery was a good predictor of production mastery, but the reverse relation did not hold.

With respect to questions about the effects of age-of-learning on L2 phonetics, it has been well documented that "earlier is better" with respect to learning to produce non-native phonetic segments with little or no accent. However, as mentioned earlier, the same advantage has not been demonstrated convincingly for L2 perception.

Perceptual Training of L2 Contrasts

Studies conducted in the 1980-90s have demonstrated that short-term intensive training can improve perception of non-native consonant contrasts when the appropriate stimuli and tasks are employed (see Jamieson's paper). Non-native voicing contrasts appear to be easier to learn than place contrasts [18,31]. However, Pisoni and his colleagues [32] have demonstrated that Japanese performed significantly better on the difficult /r-l/ contrast after completing 15 sessions of identification training with a large corpus of natural speech minimal pairs. Yamada [33] further demonstrated that performance continued to improve over 45 training sessions, and for some subjects, reached native-like levels of performance.

An interesting finding of recent training studies is the significant role that syllable context plays in limiting the extent of generalization. Morosan & Jamieson [34] reported that while training native French speakers on the English /ð-θ/ contrast in synthetic CV syllables improved perceptual differentiation of natural speech CV utterances, there was no significant transfer to the contrast in VCV or VC contexts. Apparently, subjects learn to differentiate position-specific allophones of phonetic categories, rather than context-free phoneme categories.

THE PRESENT

Results of research in the 1980-90s increased our understanding of the phenomena of L2 phonetics. Several conclusions can now be drawn, although many questions remain unanswered. (See [35] for further reviews.)

Conclusions from Recent Research

1) Both children and adult L2 learners have significant difficulties perceptually differentiating some, but not all, vowels and consonants that are not functionally distinctive in their native language. They may also have difficulty differentiating phonetic categories that are phonemic in their native language, but differ in their phonetic realization in the L2.

a) These perceptual difficulties are *not* due to a loss of sensory capabilities, but rather reflect *perceptual attunement* to phonetic information that is phonologically relevant in the native language. Language-specific patterns of selective perception are formed very early in L1 acquisition.

b) Since all non-native phonetic contrasts are not equally difficult; contrastive analysis of phoneme inventories cannot accurately predict perceptual problems of L2 learners. Perceptual difficulty varies as a function of the phonotactic and phonetic context in which the non-native contrasts occur. It may be that temporally cued contrasts are easier to perceive than spectrally-cued contrasts (but see below).

2) selective perceptual patterns are modified in adult L2 learners (as well as children) through immersion in the L2 environment or intensive conversational instruction. Perception of L2 contrasts may continue to improve for several years. However, some perceptual difficulties may persist, even after production of non-native phonetic segments is mastered. Thus, while L1 substitution patterns in production by inexperienced L2 learners are predictable from perceptual assimilation patterns, perception and production mastery may be uncorrelated in more experienced L2 learners.

3) Short-term training using stimuli and tasks that emphasize *equivalence classification* (rather than discrimination of physical differences) can lead to significant and lasting improvement in the perception of non-native contrasts. Such training has been shown to transfer to novel talkers and stimuli (i.e., new phonetic contexts) but, to date, generalization across different phonotactic contexts has not been demonstrated.

Current Theories of L2 Phonetics

Current research on the phonetics of L2 learning focuses on several remaining questions about the nature of the language-specific patterns of perception, the relationship between L2 perception and production, and the effects of perceptual training on L2 perception (and production) patterns. While good descriptive studies are still being conducted (and provide very valuable data), more current experimentation is theory-driven. Current theoretical debates center on some basic questions regarding how to characterize phonetic categories, L1 and L2 categorization processes, and *what* is learned during perceptual training.

Two working models have been offered that attempt to predict (and explain) the relative perceptual difficulty of non-

native phonetic categories. They complement each other in that Best's Perceptual Assimilation Model (PAM) [36] focuses on initial perceptual difficulties, while Flege's Speech Learning Model (SLM) [37] proposes an account of perceptual reorganization in both L2 and L1 as a function of L2 experience.

According to PAM, non-native phonetic segments are perceptually assimilated to native phonetic categories according to their articulatory-phonetic (gestural) similarity to native *gestural constellations*. If the non-native phones are very discrepant from any native phonetic gestures, they may be assimilated as uncategorizable speech or even as non-speech sounds. Perceptual difficulty in differentiating a non-native contrast is predictable from these assimilation patterns. If the contrasting phones are both assimilated as good exemplars of a single native category, perceptual differentiation is extremely difficult; if the contrasting phones differ in their "goodness of fit" to a single native category, then perception will be somewhat easier. If the two phones are assimilated to two different native categories, they will be differentiated with ease. Finally, non-assimilated phones will be perceptually differentiated on the basis of their psychoacoustic distinctiveness.

In a recent version of his SLM, Flege proposes that L1 and L2 *position-sensitive allophones* are related along a continuum of interlingual phonetic similarity, defined in acoustic-phonetic terms, such as the F1/F2 formant space for vowels or the VOT continuum for voicing in stop consonants. He hypothesizes that beginning L2 learners perceptually assimilate most L2 segments to native categories; however, if the L2 segment is sufficiently dissimilar from any L1 segment that L2 learners can discern the difference perceptually, then a new L2 perceptual category will be established over time. For less dissimilar L2 segments, separate L2 category formation may continue to be blocked because of equivalence classification of L1 and L2 segments. In these cases, a single perceptual category subsumes both L1 and L2 segments, leading to persistent accented production of the L2 segment and even to shifts in production of the native segment away from the monolingual norm.

Although these two models differ in the emphasis placed on acoustic vs. articulatory specification of phonetic similarity, they both take context-dependent phonetic segments as the appropriate level of analysis rather than the more abstract phonemes or distinctive features of traditional linguistic analysis.

Other theorists are concerned with the nature of phonetic category organization and how it affects the perception of native and non-native phonetic segments. Pisoni [38] argues that an exemplar-based model can best account for several phenomena in speech perception, including why training with a large and variable corpus is successful in reorganizing phonetic categories. According to this model, perceivers store detailed information about individual phonetic segments, including speaker-specific and context-specific information. Categorization involves matching an incoming signal on the basis of its overall physical similarity to previously stored exemplars. Thus, native phonetic categories are represented as clusters of exemplars that share certain critical (acoustic) parameters, while varying on other, non-criterial characteristics. For L2 perceptual learning to be successful, training must be conducive to the formation of (new) equivalence clusters.

In Kuhl's *Perceptual Magnet* model [39] native-language phonetic categories are organized around best cases or *prototypes* (established within the first year of life) which distort the phonetic perceptual space. Acoustic variations around these prototypes come to be perceived as more similar to each other. This warping of the perceptual space around native-language prototypes accounts for the failure to differentiate phonetic variants that are distinctive in the L2 but constitute within-category variations in the L1. L2 perceptual learning would require the reorganization of the phonetic perceptual space around newly established prototypes.

Each of these four theorists makes somewhat different claims about the nature of the stored representations of L1 phonetic categories. However, all depend on one or another definition of *phonetic similarity* as an organizing schema. An important task for future research is to characterize the notion of phonetic similarity in explicit and non-circular ways.

THE FUTURE

Research on the phonetics of L2 acquisition is a vibrant and productive area of endeavor. While there have been great advances in our understanding of the basic phenomena in the last 10 years, important unanswered questions about the very nature of phonetic categories and categorization processes involved in the perception of speech remain. In my remaining comments, a few suggestions for future research are made.

1) In describing assimilation patterns in L2 perception, it is important that experiments be conducted with stimuli and tasks that tap perceptual processes at appropriate levels of analysis. Thus, when investigating the relative salience of temporal cues *vs* spectral cues, it must be remembered that both kinds of information are imbedded in a larger context in continuous speech. Temporal cues for phonetic contrasts are defined relative to other gestural timing characteristics (stress, rate of speech) which are specified over larger stretches of speech than single syllables. Spectral cues also vary as a function of coarticulation and timing patterns. So, for instance, similarity of L1 and L2 vowels ascertained from judgments of vowels produced in isolation or citation-form syllables may not predict perceptual assimilation patterns in more naturally produced utterances [43]. Further, the phonotactic contexts in which phonetic contrasts are investigated influence the results profoundly. Language-specific knowledge of allophonic variation and syllable structure rules interacts with (language-universal) constraints to determine listeners' expectations about how phonetic segments influence each other in speech utterances. There is a need for research that investigates how this aspect of L1 phonology affects L2 perception/production patterns.

2) Research on L2 perception and production suggests that their interrelationship may change in complex ways over a relatively long period of time. More research is needed which traces these changes over sufficiently long time periods. Because studies of L2 perception and production show large individual differences, long-term longitudinal studies are needed.

3) Perceptual training studies have concentrated primarily on L2 consonant contrasts where members of the contrasting pair sound "the same." On the other hand, L2 vowel contrasts are usually discriminable even from the outset. Thus, the perceptual problem is one of learning which of the discriminable differences are critical for the contrast, and which others constitute within-category variations. With respect to production/perception relationships, vowels and consonants may also differ. Whereas consonant gestures involve contact of articulators (with concommitant tactile feedback), vowel articulation requires spatial positioning of the tongue in a relatively open vocal tract. It may be the case, therefore, that production of vowels is more dependent on auditory feedback. Training studies

that assess effects of perceptual training on L2 vowels and consonants will provide important insights into these differences. [Work supported by NIDCD]

REFERENCES

[1] Abramson, A.S. & Lisker, L. (1970) "Discriminability along the voicing continuum: Cross-language tests." *Proc. VIth ICPhS* 569-573.

[2] Liberman, A. M., et al. (1967). "Perception of the speech code." *Psychological Review*, 74, 431-461.

[3] Stevens, K.N., et al. (1969) "Cross-language study of vowel perception." *Lang. Speech* 12, 1-23.

[4] Miyawaki, K., et al. (1975) "An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English." *Percept. Psychophys.*, 18, 331-340.

[5] Williams, L. (1977) "The perception of stop consonant voicing by Spanish-English bilinguals." *Percept. Psychophys.*, 21, 289-297.

[6] Lasky, R. E., et al. (1975) "VOT discrimination by four to six and a half month old infants from Spanish environments." *J. Exper. Child Psychol.*, 20, 215-225.

[7] Streeter, L. A. (1976) "Language perception of two-month old infants shows effects of both innate mechanisms and experience." *Nature*, 259, 39-41.

[8] Trehub, S.E. (1976) "The discrimination of foreign speech contrasts by infants and adults." *Child Development*, 47, 466-472.

[9] Goto, H. (1971). "Auditory perception by normal Japanese adults of the sounds "L" and "R"." *Neuropsychologia*, 9, 317-323.

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

[11] Lisker, L, (1970) "On learning a new contrast." *Haskins Laboratories : SRSR*, SR-24, 1-17.

[12] Strange, W. (1972) "The effects of training on the perception of synthetic speech sounds: Voice onset time." Unpubl. Ph.D. dissertation, U. Minnesota.

[13] Strange, W. & Dittmann, S. (1984) "Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English." *Percept. & Psychophys.*, 36, 131-145.

[14] Lenneberg, E. (1967) *Biological foundations of language*. NY: Wiley.

[15] Gottfried, T.L. (1984) "Effects of consonant context on the perception of French vowels." *J. Phonetics*, 12, 91-114.

[16] Werker, J.F., & Tees, R.C. (1983). "Developmental changes across childhood in the perception of non-native speech sounds." *Canadian J. Psychol.*, 37, 278-286.

[17] Werker, J.F., & Tees, R.C. (1984) "Phonemic and phonetic factors in adult cross-language speech perception." *J. Acoust. Soc. Am.*, 75, 1866-1878.

[18] Tees, R.C. & Werker, J.F. (1984) "Perceptual flexibility: Maintenance or recovery of ability to discriminate non-native speech sounds." *Canad. J. Psychol.* 38, 579-590.

[19] Polka, L. (1991). "Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions." *J. Acoust. Soc. Am.*, 89, 2961-2977.

[20] Best, C.T., et al. (1988) "Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants." *J. Exp. Psychol.: Human Percept. Perform.*, 14, 345-360.

[21] MacKain, K.S., et al. (1981) "Categorical perception of English /r/ and /l/ by Japanese bilinguals." *Applied Psycholing.*, 2, 369-390.

[22] Underbakke, M., et al. (1988) "Trading relations in the perception of /r/-/l/ by Japanese learners of English." *J. Acoustic. Soc. Am.*, 84, 90-100.

[23] Yamada, R. A., & Tohkura, Y. (1992) "Perception of American English /r/ and /l/ by native speakers of Japanese." In Y. Tohkura, E. et al. (Eds.) *Speech perception, production and linguistic structure*. Tokyo, JAPAN: OHM Publishing Co. Ltd. 155-174.

[24] Flege, J.E. (1984) "The effect of linguistic experience on Arabs' perception of the English /s/ vs. /z/ contrast." *Folia Linguist.*, 18, 117-138.

[25] Werker, J.F. & Tees, R.C. (1984) Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behav. Develop., 7, 49-63.

[26] Polka, L. & Werker, J.F. (1994) "Developmental changes in the perception of nonnative vowel contrasts." *J. Exper. Psychol.: Human Percept. Perform.*, 20, 421-435.

[27] Flege, J.E. & Eefting, W. (1987) "Production and perception of English stops by native Spanish speakers." *J. Phonology*, 15, 67-83.

[28] Snow, C.E., & Hoefnagel-Hohle, M. (1977) "Age differences in the pronunciation of foreign sounds." *Lang. Speech* 20, 357-365.

[29] Cochrane, R.M. (1980) "The acquisition of /r/ and /l/ by Japanese children and adults learning English as a second language." *J. Multiling. Multicultural Develop.*, 1, 331-360.

[30] Rochet, B.L. (in press) "Perception and production of L2 speech sounds by adults." in [35].

[31] McClaskey, C.L., et al. (1983) "Transfer of training of a new linguistic contrast in voicing." *Percept. Psychophys.*, 34, 323-330.

[32] Logan, J.S., et al. (1991) "Training Japanese listeners to identify English /r/ and /l/: A first report." *J. Acoustic. Soc. Am.* 89, 874-886.

[33] Yamada, R.A. (1993) "Effect of extending training on /r/ and /l/ identification by native speakers of Japanese." *J. Acoustic. Soc. Am.* 93, 2391.

[34] Morosan, D.E., & Jamieson, D.C. (1989) "Evaluation of a technique for training new speech contrasts: Generalization across voices, but not word position or task." *J. Speech Hear. Res.* 32, 501-511.

[35] Strange, W. (ed.) (in press). *Speech perception and linguistic experience: Issues in cross-language speech research.* Timonium MD: York Press.

[36] Best, C. (in press) "A direct realist view of cross-language speech perception." in [35].

[37] Flege, J.E. (1991). "Speech learning in a second language," in Ferguson, D. et al. (eds.) *Phonological development: Models, research, and application.* Parkton, MD: York Press.

[38] Pisoni, D.B. & Lively, S.E. (in press) "Variability and invariance in speech perception: A new look at some old problems in perceptual learning." in [35].

[39] Kuhl, P.K. & Iverson, P. (in press) "Linguistic experience and the "perceptual magnet effect"." in [35].

[40] Strange, W. et al. (1993) "Consonant context affects perceived similarity of North German and American English vowels." *J. Acoustic. Soc. Am.*, 94, 1866.