Working together
www.rcis.ro

Revista de Cercetare si Interventie Sociala

ISSN: 1583-3410 (print), ISSN: 1584-5397 (electronic)

ANALYSIS OF SECOND LANGUAGE ACQUISITION (SLA) SPEECH PERCEPTION MODEL & THE PERCEPTION OF SECOND LANGUAGE PROSODY

Jing CHU, Chunsheng YANG, Guofa LIU

Revista de cercetare și intervenție socială, 2019, vol. 64, pp. 334-351

https://doi.org/10.33788/rcis.64.25

Published by:
Expert Projects Publishing House

expert projects
publishing

On behalf of:
„Alexandru Ioan Cuza" University,
Department of Sociology and Social Work
and
HoltIS Association

REVISTA DE CERCETARE SI INTERVENTIE SOCIALA
is indexed by Clarivate Analytics (Web of Science) -
Social Sciences Citation Index
(Sociology and Social Work Domains)

Analysis of Second Language Acquisition (SLA) Speech Perception Model & the Perception of Second Language Prosody

Jing CHU[1], Chunsheng YANG[2], Guofa LIU[3]

Abstract

This paper provides a critical review on the major models of speech perception in second language (L2) acquisition. It is argued that some new models, such as L2LP and ASP, have more explanatory power for L2 speech perception. However, due to the different theoretical frameworks, objectives and hypotheses in these models, it is difficult to integrate these models into one which is universally applicable. Although most of these models were proposed for accounting for the perception of L2 segments, they can also be applied in the perception of L2 prosody. When these models are used in examining L2 speech prosody, the prosodic systems of both L1 and L2 should be thoroughly investigated first.

Keywords: second language, speech perception, perception models, prosodic perception,

Introduction

When learning a second language (L2), learners process L2 through their native linguistic systems. It is widely attested that L2 learners have difficulty in achieving native-like phonology, as can be seen in the phenomena of the so-called "foreign-accented speech". The first language (L1) sound system has long been observed to exert impacts on the perception of L2 phonemes (Polivanov, 1931). Trubetzkov (1969) believes that the inadequate production of L2 sounds has a perceptual basis, arguing that the L1 sound system acts as a "phonological filter" through which L2 sounds are perceived and categorized. In the same vein, Escudero (2005) proposes that L2 learners also have "perceptual foreign accent", in that they perceive the L2 on the basis of their L1 perceptual system (Strange,

[1] Bohai University, jinzhou City, Liaoning Province, 121000, CHINA. E-mail: feierandclub_bhu@163.com (Corresponding author)

[2] University of Connecticut, Storrs, CT, USA. E-mail: yangchunsheng_uc@163.com

[3] Beijing Century Internet Software co. LTD, Beijing, CHINA. E-mail: liuguofa_bjcls@163.com

1995). Escudero (2005: 2) attributes the origin of a foreign accent to the use of language-specific perceptual strategies that are entrenched in the L2 learner and can not be avoided when encountering L2 sound categories. L2 perception is closely related to L2 production. As for the interplay between L2 production and L2 perception, most studies confirm the proposition that perception precedes production and that a perceptual difficulty is likely to underlie the widely observed difficulty adults have in producing L2 sounds (Flege, 1993; Leather, 1999; Llisterri, 1995; Neufeld, 1988). However, there are also some studies (Flege & Eefting, 1987; Goto, 1971; Sheldon & Strange, 1982) that challenge the precedence of perception over production in L2 phonology. For example, both Goto (1971) and Sheldon & Strange (1982) find that the perception of [l] and [r] by Japanese learners of English does not necessarily precede or may even lag behind acceptable production. However, Escudero (2005:111) argues that these studies have some methodological drawbacks, such as the controlled nature of the production tasks and the articulatory training undergone by the learners, which elicited highly monitored and unnatural L2 production, as well as the problematic nature of their data analyses.

In spite of the uncertainty with respect to the relationship between L2 production and L2 perception, there is no denying that perception plays a significant role in the acquisition of L2 phonology, and in many cases the incorrect perception leads to incorrect production. Thus, it is of great importance to examine how L2 speech perception works and how linguistic and non-linguistic factors influence the perception of non-native sounds in the process of acquiring L2 phonology.

Models of speech perception in second language acquisition (SLA)

Different from the L1 phonetic perception, naïve L2 listeners or L2 learners, when confronting a non-native language, have difficulty in perceiving many non-native contrasts (Strange, 1995). Various theories have been advanced to explain and predict the possible difficulties that L2 listeners might face when perceiving non-native sounds. These theories differ in their theoretical frameworks: phonological, phonetic, or a combination of both. In the following sections, some major perception models will be discussed, followed by a summary of them.

The Ontogeny Phylogeny Model (OPM)

The OPM (Major, 2001; 2002) aims to specify the principles involved in the formation of L2 phonological systems (an inter-language system), the change in an L1 resulting from exposure to an L2, and the language contact phenomena such as bilingualism and multilingualism. The OPM posits that the L2 inter-language is comprised of three components: L1 transfer, L2, and language universals (U). U includes Universal Grammar (UG) and many other universal factors that occur in L2 as well, such as learnability principles, markedness, underlying representations, rules, processes, constraints, and stylistic universals (see Major, 2001: 41-52). According to the OPM, the inter-language roughly develops in the following pattern: Over time and as style becomes increasingly formal, L2 increases, L1 decreases, and U increases and then decreases. In addition, the relative proportions of U and L1 depend on whether the linguistic categories are normal, similar, or marked, determined by the nature of the linguistic categories to be learned: L2 increases more slowly in similar and marked situations than the normal scenario. In similar situations, L1 is relatively more important than U, whereas in marked situations U is relatively more important than L1 (Major, 2001: 157).

The OPM is mainly a model of L2 production. However, Major argues that the intermediate performance in L2 perception, as reported in Caramazza *et al.* (1973) and Williams (1977), can be accounted for by the OPM's proposal of partially merged L1 and L2 systems (Major, 2002: 82), in that each of the bilingual's phonological systems is proposed to have a component of the other system. In addition, Major seems to imply that the development of L1, L2, and U in an inter-language applies to both production and perception.

The OPM only makes general claims concerning the interaction and development of L1, L2 and U, without formulating the detailed claims concerning specific phenomena. Major (2001, 2002) argues that this lack of specifics is not necessarily a weakness, because in so doing, the claims can survive the challenge of any linguistic framework. Even so, it seems that if a model intends to be adequately explanatory, it needs to provide a more explicit proposal regarding the role of perception in the development of phonological competence, namely the learning mechanisms that trigger the decrease and increase of the three components should be explicitly spelt out.

The Perceptual Assimilation Model (PAM)

The Perceptual Assimilation Model (Best, 1995) focuses on the perception/ discrimination of the non-native contrasts by naïve non-native listeners. The PAM is founded on a direct realist approach to perception, in which articulatory gestures are assumed to be the perceptual primitives for speech perception. The PAM hypothesizes that listeners assimilate non-native sounds to the closest sound in their native system in accordance with the articulatory similarity. There are three possible patterns of perceptual assimilation of non-native segments: (1) assimilated to a native category, or to a cluster or string, in which case the fit can vary from good exemplar to a notably deviant exemplar of the category; (2)assimilated to an uncategorized speech sound: the segment is perceived as speech sound but can not be assigned to any native category, and (3) not assimilated to speech: the segment is not assimilated into native phonological space at all; heard, instead, as some sort of non-speech sound. These three possible assimilation patterns result in the difference in the knowledge that listeners employ when listening: if a sound is assimilated to a native category, the linguistic-phonetic knowledge will be employed; for sounds that are assimilated as uncategorizable speech sounds, listeners will process the sounds using a combination of linguistic and auditory processing; for unassimilable non-speech sounds, listeners can only resort to the auditory processing.

The assimilation patterns and the degree of perceptual differentiation for non-native contrasts can be predicted from the assimilation of each member of the contrast. The six possible pairwise assimilation patterns and the associated predicated patterns of discriminability are as follows (Best, 1995: 195): (1) Two-Category Assimilation (TC type): each non-native segment is assimilated to a different native category; discrimination is expected to be excellent; (2) Category-Goodness Difference (CG type): both non-native segments are assimilated to the same native category, but one segment fits the native category better than the other. Discrimination is expected to be moderate to very good, depending on the magnitude of the difference in category goodness for each of the non-native sounds; (3) Single-Category Assimilation (SC type): both non-native sounds are assimilated to the same native category. Discrimination is expected to be poor; (4) Both Uncategorizable (UU type): both non-native sounds are perceived as speech but can not be assimilated to any native speech category. Discrimination is expected to range from poor to very good, depending on the similarity to each other and to other native categories; (5) Uncategorized versus Categorized (UC type): One non-native sound is assimilated to a native category, while the other is not. Discrimination is expected to be very good; and (6) Nonassimilable (NA type): both non-native sounds are perceived as non-speech sounds. Discrimination is expected to be good to very good. The predictions of this model have been borne out, though most studies focus on the consonant contrasts in isolated syllables without considering the allophonic variation across languages (Bent, 2005: 8). Even though the PAM predicts the perceptual patterns based on the assimilation of the non-native sounds into the native categories, the PAM also acknowledges that experience with L2 leads to the reorganization of the perceptual assimilation patterns, and that the perceptual learning continues into adulthood, even though the mechanism of how to incorporate experience factors into the PAM has yet to be investigated. Another issue to be tackled in the PAM is the development of a detailed, objective means for predicting assimilation patterns and discrimination of particular non-native contrasts (Best, 1995: 198).

The latest development of the PAM is its extension to predict patterns of speech perception by L2 learners (PAM-L2) (Best and Tyler, 2007). Best and Tyler argue that L2 perception involves three levels of attentional focus, namely the gestural level, the lower-order phonetic level, and the higher-order phonological level; the specific focus in perception depends on the perception task and the stimuli types. Best and Tyler describe several patterns of cross-language assimilation at the phonetic (allophonic, dialectal) level, and at the phonological level (lexical minimal pairs), as well as the perceptual development with the increase of L2 experience. For example, L2 phonetic segments can be assimilated as more or less "good" exemplars of L1 phonological categories (CG type in PAM), based on the differences in the details of their articulator-phonetic realization in the two languages, or on the basis of similar phonological functions (e.g., phonotactic distributions, as the French /r/ and the American /r/). Or if only one L2 phonological category is perceived as equivalent to a given L1 phonological category, a TC type of assimilation or a UC type of assimilation will occur. However, the probability that an L2 contrast will come to be perceptually differentiated with L2 experience is dependent upon the patterns of phonetic/phonological assimilation, as well as the functional load in L2, i.e., whether the minimally contrasting words are high frequency or low frequency ones.

In the PAM-L2, L2 learners are assumed to be active learners and their perception involves both phonetic and phonological levels. Best and Tyler (2007: 32) point out that, in addition to exposure to the target language, L2 learners differ from naïve listeners in many other dimensions, especially the effects of establishing lexical items, which is argued to lead L2 learners to "re-phonologize" their perception of the target language.

The PAM-L2 is different from the SLM (for experienced L2 learners, see below). On the one hand, the PAM-L2 covers the initial stage of L2 learning (PAM) to the advanced bilingual stage of L2 learning (SLM). In this sense, the PAM-L2 serves as a bridge to connect the PAM and the SLM. On the other hand, the PAM-L2 acknowledges the phonological perception in L2 perception due to the L2 learning effects, which makes L2 perception different from naïve non-native perception. However, due to the gestural, phonetic and phonological components and the great difference between naïve and L2 learners, together with the L2 learning and developmental effects, it is still very difficult to predict the perceptual patterns in L2 perception, especially at the different stages of L2 learning.

The Speech Learning Model (SLM)

Flege's Speech Learning Model (SLM) (Flege 1995; Flege, Schirru, & MacKay, 2003) is primarily concerned with the ultimate attainment of L2 pronunciation in both production and perception. The SLM claims that L1 and L2 phonetic subsystems exist within a single phonological space in experienced L2 learners. L1 and L2 segments can be related along a continuum from identical through similar to new, defined empirically in terms of acoustic similarity or perceived cross-language similarity. A new phonetic category can be established for an L2 sound that differs phonetically from the closest L1 sounds. However, the category formation for an L2 sound may be blocked by the mechanism of equivalence classification. When this occurs, a single phonetic category will be used to process perceptually linked L1 and L2 sounds. Moreover, the SLM proposes that the mechanisms and processes used in learning the L1 sound system remain intact over the life span, and can be applied to L2 learning. The L2 development is constrained by four main factors: perceived cross-language similarity, age of arrival, L1 use, and the storage of L1 and L2 categories in a common phonological space. The SLM proposes that there is a direct relationship between perception and production, because the production will be inaccurate without accurate perceptual targets to guide the sensor motor learning of L2 sound. The model is also explicit about the bi-directional nature of the interactions between the native and non-native languages in an individual.

The predictions of the SLM have been attested using both consonants (Flege, 1987; MacKay *et al.*, 2001) and vowels (Flege, MacKay, & Meador, 1999; Flege & MacKay, 2004). However, Guion *et al.* (2000) find some evidence against the predictions with relatively inexperienced L2 learners. The SLM explicitly claims that it aims at predicting the perception of the experienced L2 learners, which seems to dismiss the counter-evidence found in Guion *et al.* (2000). However, according to the SLM, whether a new phonetic category will be established depends on the phonetic similarity of the L2 sound to the closest L1 sound, which happens at the initial state (or first contract with the sound) of L2 learning. However, this contradicts the SLM claims for its applicability only to the experienced L2 learners. Moreover, with L2 learners' increase exposure to L2, their inter-language system will consist of L1, L2 and U (Major, 2000) and their perception will involve more than the matching of L2 sounds with the L1 sound. In this sense, the PAM-L2 does a better job than the SLM. What is more, as Flege (1995: 264-265) points out, it is difficult to measure the perceived phonetic difference between the L1 and L2 sounds. Thus, it is difficult to predict how an L2 sound will be processed, whether treated as a new phonetic category, or treated as similar or identical to the native phonetic category. More complication will occur, if more exposure or experience improves the L2 perception. All of these issues show that, even though the SLM acknowledges the learning effects in perception, it still needs to formulate the explicit mechanisms about how to incorporate the L2 experience or exposure when predicting the perceptual patterns.

The Native Language Magnet Model (NLM)

The Native Language Magnet Model (Kuhl, 1991; Kuhl, 2000; Kuhl & Iverson, 1995) aims at explaining the development of speech perception from infancy to adulthood. The NLM assumes a general auditory-acoustic mechanism. According to the NLM, exposure to the native language alters the perceived distance between category exemplars in the acoustic space and these alterations lead to long-term changes in speech perception patterns. Phonetic prototypes are formed through exposure to the distribution of sounds in the native language. Sounds close to the prototypes are perceptually drawn to them and, therefore, the perceived distance between the prototype and other members of the category shrinks. This change in perceptual space can help listeners perceive the non-native sounds by heightening the relative salience of essential acoustic cues if these cues are used for native between-category discrimination. However, this change in perceptual space may hinder the discrimination by attenuating cues that signal within-category variation in the native language, but are important for making non-native between-category variation. The NLM assumes that the perception of one's primary language is completely different from that required by other languages. Kuhl (2000) proposes that "no speaker of any language perceives acoustic reality; in each case, perception is altered in service of language". Due to the existence of an L1 language-specific perceptual filter, the NLM predicts that learning an L2 is difficult because later learning is constrained by the initial mental mappings that have shaped neural structure.

The NLM-Expanded Version (NLM-e), Kuhl *et al.* 2007) formulates the five principles guiding the NLM-e: (1) distributional patterns and infant-direct speech are agents of change; (2) language exposure produces neural commitment that affects future learning; (3) social interaction influences early language learning at the phonetic level; (4) the perception-production link is forged developmentally; and (5) early speech perception predicts language growth. The NLM-e proposes that the development of speech perception undergoes four phases: (1) Phase 1: Early in life infants discriminate all phonetic units in the world's languages; (2) Phase 2: infants' sensitivity to the distributional patterns and exaggerated cues of infant-directed speech cause phonetic learning; (3) Phase 3: enhanced speech perception abilities improve three independent skills that propel infants towards word acquisition: the detection of phonotactic patterns, the detection of transitional probabilities between segments and syllables, and the association between sound patterns and objects; and (4) Phase 4: analysis of incoming language has produced relatively stable neural representations – new utterances do not cause shifts in the distributional properties coded neurally, and future learning is affected by native language patterns.

The NLM-e provides an encompassing model of the multiple of factors influencing the infants' early phonetic learning and also offers specific hypotheses that are amenable to empirical investigation (Kuhl *et al.* 2007: 16). The NLM-e suggests that non-native phonetic performance reveals uncommitted neural circuitry, whereas native language phonetic performance is indicative of neural commitment to the native language. However, both the NLM and the NLM-e should explicitly address the particular mechanisms of L2 learning, such as how the separation of perceptual mappings for two languages is achieved and how it is influenced by different levels of L2 proficiency.

The L2 version of the Linguistic Perception Model (L2LP)

The model of Linguistic Perception (LP), (Escudero & Boersma, 2004; Escudero, 2006) aims at describing, explaining, and predicting the knowledge underlying speech perception and the acquisition of this knowledge in learning a first or a second language. The LP model is entrenched in the theoretical framework of Functional Phonology (Boersma, 1998), which claims that cognitive linguistic knowledge underlies speech perception. In the LP model, the language-specific knowledge underlying speech perception consists of (1) a linguistic and grammatical processor, i.e., a perception grammar, which maps (i.e. categorizes) the variable and continuous acoustic signal; and (2) perceptual representations of perceptual input. According to the LP model, the acoustic signal is linguistically analyzed bottom-up without top-down application of lexical knowledge. With respect to the workings of the perception grammar, the LP model proposes that an optimal listener will construct the perceptual categories that are most likely to have been intended by the speaker and thus will pay attention to the acoustic cues that are most reliable in the environment when perceiving sound segments.

With respect to the L2 learning, Escudero & Boersma (2004) and Escudero (2005) propose an L2 version of the linguistic perception model (L2LP). The L2LP makes a clear distinction between the perceptual mapping and sound representations (or category formation). There are three possible learning scenarios, new, similar and subset.

For the initial state in L2 perception, the L2LP proposes that the learners automatically create a "copy" of the L1 perception grammar and the L1 lexical representations. Thus the L2LP assumes that L2 perception is handled by a separate perceptual system. Take the similar scenario for example. The two L2 phonemes in this situation will be perceptually mapped to two L1 phonemes.

With respect to the L2 development, it is proposed that L2 learners have access to the same learning mechanisms, performed by the Gradual Learning Algorithm (GLA); (Boersma & Hayes, 2001), which were available for L1 learning – namely, auditory-guided category formation and lexicon-guided boundary shifting for phonological categories. The results of the initial state might result in the mismatch between the copied L1 perception and the near-optimal L2 perception. In this situation, the learner will not be able to correctly categorize all L2 tokens. Thus the learner's GLA, which in this situation acts as an error-driven constraint re-ranking mechanism triggered by the mismatches between the output of perception and the lexicon, will change their perception grammar by small steps in order to decrease the probability of semantic mismatches. Finally, an optimal L2 perception will be attained when such mismatches no longer occur.

With respect to the ultimate attainment in the L2, it is proposed that native or native-like perception in a learner's two languages is possible because L1 and L2 are two separate systems (i.e., two separate sets of perceptual categories and two perception grammars). Both L2 development and L1 stability are predicted provided that the two languages are each used on a regular basis.

The L2LP model provides a detailed developmental model for L2 perception with respect to the initial, developmental, and end states, with the most comprehensive description, explanation, and prediction of L2 sound perception. As compared with other perception models, this is a great strength. The L2LP treats the L2 perception grammar as a separate system, which copies all L1 perception grammar and lexical representation at the initial state, and can become optimal gradually with the help of GLA. Even though the L2LP proposes the separation of two separate systems for L1 and L2, the L2LP supports the cognitive interplay of the L1 and L2 language systems during acquisition. However, the L2LP is only a working model and so far has only been tested in limited studies on vowels (see Escudero, 2005).

The Automatic Selective Perception (ASP)

Strange's Automatic Selective Perception (ASP) model (Strange, 2006; Strange, 2007; Strange & Shafter, 2008) focuses on the development of L1 and L2 phonetic perception. In this model, infants are language-general perceiver, born (or shortly after birth) with the ability to discriminate phonetically-relevant acoustic properties of speech sound. However, by the end of the first year of life, infants start to become language-specific perceivers: their perceptual abilities have been reorganized so that they begin to reflect the phonological structure of the native language input. Meanwhile, they have learned to selectively attend to those phonetic differences that are phonologically relevant in the native language, and to ignore many of the acoustic-phonetic differences not present or not used to distinguish phonological contrasts in native language. Over the next several years of life, children's selective perceptual processes are further modified such that the weighting of multiple acoustic parameters comes to resemble the adult patterns of the native language. More reliable acoustic parameters are given more weight, while phonologically irrelevant variations are given less weight. This allows the children to cope with the inherent variability in the phonetic realization of phonological segments which occurs within and across speakers, phonetic/phonotactic contexts, and speaking rates/styles. In adults, native-language phonetic perception is robust and automatic. The ability to extract the phonetic message form the acoustic signal, even in non-optimal situations (unfamiliar talkers, competing noise, distracting tasks requiring the listeners' attention) requires few cognitive resources on the part of the native listener.

With respect to L2 phonetic perception, the ASP model assumes the language-specific phonetic perception. Language-specific patterns of performance are not due to differences in the basic auditory capabilities of adult speakers of different languages. Rather, they reflect highly over-learned and efficient patterns of selection and integration of acoustic-phonetic information by which phonetic sequences are reorganized. In adult listeners, these language-specific patterns of categorization have become automatic (requiring few cognitive resources) and highly robust even in difficult listening perceptions. Strange (2006) refers to these automatic, language-specific patterns of perception as Selective Perceptual Routines (SPRs).

Beginning L2 learners initially come to the L2 listening task using their automatic L1 SPRs, which, in some cases, are not attuned to the most appropriate acoustic information for L2 phonetic segments (i.e. , L1 interference). This results in perceptual difficulties on some non-native contrasts; when tested with stimulus materials and perception tasks that tap these selective perception processes, they show significant perceptual deficits, as compared to native listeners. However, because auditory sensory capabilities remain intact, perception of non-native contrasts can and usually does improve with experience with the L2 phonological structures. Selective perceptual processes are re-educated with L2 experience such that many late L2 learners come to be able to perceptually differentiate even difficult contrasts under optimal listening conditions. That is, L2 SPRs can be acquired in adulthood. However, due to the influence of the L1, L2 SPRs may be based on different (non-optimal) weightings of acoustic parameters than those used by native listeners, even after years of immersion experience. Under difficulty listening conditions which challenge the perceptual capacities of the listeners, L2 speakers' performance deteriorates more rapidly than native speakers' performance. It appears, then, that L2 SPRs may differ from those of native listeners and may never be as fully automated as L1 SPRs.

The ASP model shares some similarities with the PAM-L2, such as the intact sensory capabilities（also similar to the SLM model）, the attentional effects (the effects of stimulus and task conditions), and the re-education of L2 learners. However, the ASP has yet to address how L1 SPRs will be employed in different scenarios, because it is expected that the L1 SPRs will be used differently when the L1 categories are similar to or different from the L2 categories. Actually, this is a common problem confronting all perception models.

Summary

The L2 perception models reviewed above can be categorized into two groups: (1) models that are primarily concerned with predicting relative difficulty in the perception (and production) of non-native contrasts by naïve listeners and later L2 listeners, such as OPM, SLM, NLM, PAM and PAM-L2; and (2) the models that consider the particular processing mechanism in L2 phonetic perception, such as NLM-e, L2LP and ASP. The models in the first category do not consider in detail the online process involved in recovering the phonetic message from acoustic signals, and how those processes may differ for perception of L1 vs. L2 phonetic sequences, or for inexperienced vs. experienced L2 learners (Stranger & Shafter, 2008: 173). Even though in PAM-L2, Best and Tyler (2007) introduce the notion of attentional focus at either a phonetic or a phonological level of analysis, they do not address under what conditions these differences in attentional focus are invoked. By contrast, the models in the second category explicitly and in detail address the above issues. The L2LP is different from NLM-e and ASP in assuming that the L2 perception may be optimal at the end state. On the other hand, L2LP and ASP provide more detailed and explicit description and prediction for the L2 perception patterns than NLM-e. In spite of that, both the ASP and the L2LP are still working models, which have yet to be tested in more perceptual studies. Even so, it is expected that the L2LP and ASP may turn out to be able to provide more explanatory power for the L2 perception.

Most of the models require the assessment of the similarity between the native sounds and the non-native sounds. Thus, how to assess the perceived phonetic similarity has become an important issue, and has been discussed in some studies. With respect to how to assess the perceived phonetic similarity, different approaches have been adopted, including acoustic comparisons cross languages (Strange *et al.*, 2004), direct assessment of similarity by explicitly asking listeners to place non-native segments into native categories (Guion, 2003; Iverson *et al.*, 2003; Strange *et al.*, 2004), listener transcriptions of non-native sounds using native spelling and other descriptions (Best, McRoberts, Goodell, 2001), judgments of non-native phonemes' fit into native categories (Guion *et al.*, 2000; Iverson *et al.*, 2003), and multidimensional scaling (MDS) (Bent, 2005). However, there are still no widely accepted criteria in evaluating phonetic similarity to date.

Application of the L2 perception models to prosody perception

All the models reviewed above are primarily concerned with the perception of segments. Most studies in L2 perception are also focused on the perception of segments, even though recent years have witness the increase of the studies on prosody perception (Chen, 2005; Grabe *et al.*, 2003). However, most of these studies on prosody perception are conducted without reference to any existing perception model. Grabe *et al.* (2003) find that the perception of similarities and differences among intonation contours calls upon universal auditory mechanisms whose output is molded by experience with one's native language. Even though these findings might be accommodated by some perception models, such as L2LP and ASP, little attempt has been made.

There are multiple factors that might lead to the predominance of studies on segment perception over prosody perception: (1) the evident production errors in some L2 segments lead researchers to investigate whether L2 learners have difficulty in perception as well; (2) the perception of prosody consists of a variety of prosodic phenomena, such as tones and intonation, stress/prominence, and rhythm; and the complexity of prosody results in the limited attempts at the prosodic perception; (3) due to the difference in prosodic structure in different languages, it is difficult to determine the units of analysis and assess the perceived similarity when examining L2 prosodic perception; and (4) prosody itself is not fully and clearly understood in the field.

However, prosodic perception is an integral and indispensable part of the L2 learning. So, one question to ask is whether we can investigate the prosodic perception within the frameworks of the present L2 perception models. Before we start to consider the applicability of the present perception models to prosodic perception, another question has to be addressed, that is, whether the prosodic phenomena, such as tones and intonation, are processed in the same manner as segments. To answer this question requires the examination of two modes of processing in perception, namely, the auditory mode vs. linguistic/phonetic mode.

The auditory processing refers to the situation in which listeners do not refer to their native linguistic system and merely use psychoacoustic (i.e., language-independent) abilities when perceiving the non-native categories. It is expected that if only auditory mode is used, the perception of non-native categories should be very similar, regardless of their native languages. The linguistic/phonetic processing refers to the situation in which listeners interpret non-native categories with reference to their native linguistic systems. With respect to the processing modes, the perception of the non-native categories varies, depending on the similarity between the native and non-native linguistic systems. However, in most cases, L2 perception often involves both auditory and linguistic processings. Infants are innate language-general perceivers, because they mainly depend on the auditory processing. However, their L1 linguistic experience will adjust their perceptual patterns, rendering them more attuned to the native categories, and then linguistic processing gradually takes over. However, as most of the perception models argue that learners' auditory processing capability remains intact through their life span, the auditory processing will also be employed even in adulthood. The particular mode in perceiving non-native categories is dependent on the similarity between L1 and L2 categories, and the degree that the linguistic processing interacts with the auditory processing. Meanwhile, as Strange (1995) and Strange & Shafter (1998) point out, the stimulus materials and task exert influence on the perception modes, because different experimental paradigms reflect different modes of online processing of speech input. Therefore, the best way to examine the perception modes is to design studies that reflect real-world stimulus and task constraints (including those of the language classrooms and the L2 work environment) while maintaining experimental control and vigor (Strange & Shafter, 1998: 185).

With respect to prosodic perception, the matter becomes even more complicated in that different prosodic categories may be perceived in different manners, depending on the similarity between L1 and L2 prosodic systems. On the one hand, prosodic categories vary in different languages. On the other hand, even though the same prosodic categories are used in both L1 and L2, their acoustic cues might be different. And even if the same cues are employed, there might be some individual difference in contexts, and the cues may have different weightings. Moreover, some prosodic categories may span more than one syllable, i.e., intonation. Since the prosodic categories have a much higher variability than segments, the prosodic perception is sure to pose more challenge to listeners, especially to L2 learners. Thus, to answer the question posed above, the prosodic phenomena may or may not be processed in the same manner as segments, depending on whether L1 and L2 share the similar phenomena. It can be expected that stress/prominence might be perceived in similar manners as segments, whereas tone and intonation perception may or may not. Meanwhile, it can be predicted that the perception of tones and intonation might pose more difficulty to L2 learners. Due to the complexity of prosodic perception, both auditory and linguistic processing might be employed and it is also likely that in some cases auditory processing may be more dominant than linguistic processing, because of the difficulty of associating the L2 categories to the L1 ones.

Very limited studies have been conducted on the prosodic perception and most of the extant studies are not conducted within the framework of any model mentioned above. Bent (2005), "the production and perception of non-native prosody", is the only study (to my knowledge) that examines the perception of L2 Mandarin prosody within the framework of the PAM. Bent (2005) examines the perception and production of Mandarin tones by naïve adult American English speakers. Bent used a non-word syllable /ra/ with different tones as the building blocks of the test stimuli. The perceptual results indicate that naïve American listeners process Mandarin tones mainly in the auditory mode in that their linguistic experience does not hinder their ability to discriminate non-native contrasts, though linguistic (phonological) perception also existed when tonal contrasts were in different tonal frames. Some attempts were made to assess the correspondence of the Mandarin pitch contour to native English intonation patterns and the perceptual results were explained within the framework of the PAM (Best, 1995). The English listeners' perception (sensitivity in Bent) to one tone pair, the level-rising tone pair, differed significantly across different tonal frames, namely, listeners' perception is much better in the level-falling tone frame than in the falling-rising tone frame. Bent argues that the level-rising tone pair in the level-falling tone frames may correspond to the uncategorized versus categorized assimilation patterns (UC type) in the PAM in which discrimination is expected to be very good, while the level-rising tone pair in the falling-rising tone frames correspond to the same English intonation pattern: both had a rising patter or L*+H, namely a single-category assimilation (SC type) in the PAM in which discrimination is expected to be poor. Figure 1, cited form Bent (2005: 101) illustrates the above points.

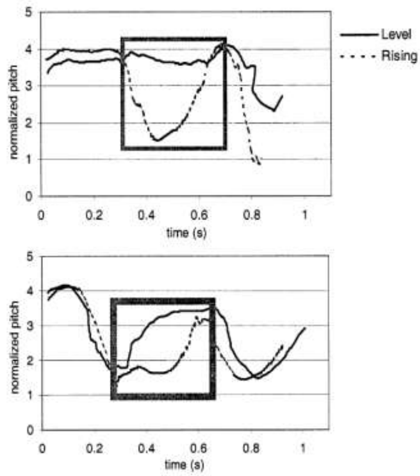

Figure 1(cited from Bent [2005:101]): The level-rising tone pair shown in two different tone frames. In the top graph, the tone pair is shown in the level-falling tone frame; in the bottom graph, the tone pair is shown in the falling-rising tone frame. The tone pair in the top graph may correspond to a categorized vs. uncategorized pattern, while the pair in the bottom graph may correspond to a single category assimilation pattern.

Bent's attempt shows that the PAM can be applied to the perception of tones by naïve listeners. Due to the lack of lexical tones in English, Bent associates Mandarin tone contours with English intonation pitch contours. It seems that the association of tones with intonation is a big difference from that in segmental perception, in which non-native sounds are usually assimilated to other native sounds. However, even in segmental perception, the non-native sounds might be heard as non-speech; thus it is justified in doing so, because intonation is the closest prosodic category that English listeners can associate with Mandarin tones. Bent's study also highlights the importance of phonetic context in perception, especially prosodic perception. Due to the coarticulation of tones in Mandarin Chinese, tones may have different pitch contours depending on the context.

Bent's study also shows the strengths of the PAM in explaining and predicting L2 perception. Due to the clearly-spelt-out possible scenarios in the PAM, the predictions of the PAM are easy to be testified (or falsified) in research.

As mentioned in the summary in Section 2.7, there are two types of L2 perception models: one type to predict the L2 perceptual patterns, and the other type to investigate the particular processing mechanisms in L2 perceptual development. Therefore, which model to choose in research depends on the research goal and the research design. For instance, if the research is to explain the L2 perceptual difference among learners of different proficiency level, the perception models in the second group, such as L2LP, NLM-e or ASP, might be better candidates to choose from. However, if the research is to predict what will happen for a particular group of learners, the models in the first group might be adequate enough. However, our discussion above has shown that the L2LP, NLM-e an ASP can not only explain and predict the perceptual patterns, their predictions also take into consideration the learning effects in the process of L2 development. Even though the predictions in the models of the second type may not be so clear-cut as predicted by the PAM or the PAM-L2, these models can provide a more comprehensive and accurate picture of the L2 learning and processing. Therefore, the L2LP, NLM-e and ASP may turn out to be more adequate models in L2 perception in the long run, even though they are still working models now.

The above discussion suggests that the L2 perception models, though proposed primarily for segmental perception, may also be applied to the prosodic perception. However, when examining the prosodic perception within these models, great caution should be made in several aspects: (1) the stimuli and task should be appropriate; (2) context factor must be taken into consideration; and (3) more complicated means of similarity assessment need to be adopted. Multidimensional scaling might be a better means. Meanwhile, before examining the L2 prosodic perception, a good understanding of the L1 and L2 prosodic systems is a must, without which no predictions can be made. Even though different research goals may justify the use of different perception models, the easy-to-predict models may not be the best choice, because many complicated factors are involved in L2 speech (including prosodic) perception.

Conclusion

We have reviewed the perception models in SLA. It is expected that some latest models, such as L2LP or ASP, will prove to have more explanatory power in L2 speech perception in the long run. However, due to the different working frameworks, purposes and assumptions, it is difficult to collapse these models into a universally accepted model. Meanwhile, even though most of these models are developed with the purpose of examining the processing of segments in L2, it is pointed out that these models may also be applied to the prosodic perception. The prerequisite for applying these models to prosodic perception is to gain an adequate understanding of both L1 and L2 prosodic systems. Moreover, great caution should be taken when applying these models to prosodic perception, in order to avoid many confounding factors when drawing any conclusion.

Acknowledgements

This work is supported by the Social Science Fund Project of Liaoning province (Project number: L18BYY004), and Dr. Start-up Fund Project of Liaoning province (Project number: JH1/102).

References

Bent, T. (2005). *Perception and Production of Non-native Prosodic Categories*. Unpublished Ph. D dissertation. Northwestern University.

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

Best, C.T., & Tyler, M.D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In: O.-S. Bohn& M. J. Munro (eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, Amsterdam / Philadelphia: John Benjamin, pp. 13-34.

Best, C.T., McRoberts, G., Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, *109*, 775-794.

Boersma, P. (1998). *Functional phonology*. Doctoral Dissertation, University of Amsterdam. The Hague: Holland Academic Graphics.

Boersma, P., & Hayes, B. (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry, 32*, 45-86.

Caramazza, A., Yeni-Komshian, G., Zurif, E., & Carbone, E. (1973). The acquisition of a new phonological contrast: the case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America 5*, 421–428.

Chen, A. (2005). *Universal and language-specific perception of paralinguistic intonational meaning*, Radboud University Nijmegen.

Escudero, P. (2005). *Linguistic perception and second language acquisition: explaining the attainment of optimal phonological categorization*. Doctoral dissertation. University of Utrecht.

Escudero, P. (2006). Second Language Phonology: The Role of Perception. In M. Pennington (ed.) *Phonology in context.* Macmillan, UK.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*, 551-585.

Flege, J.E. & Eefting, W. (1987). Cross-language switching in stop consonant production and perception by Dutch speakers of English. *Speech Communication*, *6*, 185-202.

Flege, J.E. & MacKay, I.R.A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, *26*, 1-34.

Flege, J.E. (1987). The production of 'new' and 'similar' phones in a foreign language: evidencefor the effect of equivalence classification. *Journal of Phonetics 15*, 47-65.

Flege, J.E. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America*, *93*, 1589-1608.

Flege, J.E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience* (pp. 233-277). Timonium, MD: York Press.

Flege, J.E., MacKay, I.R.A. & Meador, D. (1999). Native Italian speakers' production and perception of English vowels. *Journal of the Acoustic Society of America*, *106*, 2973-2987.

Flege, J.E., Schirru, C., & MacKay, I.R.A. (2003). Interaction between the native and second language. *Speech Communication* 40, 467-491.

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

Grabe, E., Rosner, B.S., Garcia-Albea, J.E., & Zhou, X. (2003). Perception of English intonation by English, Spanish, and Chinese listeners. *Language and Speech*, *46*(4), 375-401.

Guion, S.G. (2003). The vowel systems of Quicha-Spanish bilinguals: An investigation into the age of acquisition effects on the mutual influence of the first and second languages. *Phonetica*, *60*, 98-128.

Guion, S.G., Flege, J.E., Akahane-Yamada, R., & Pruitt, J.C. (2000). An investigation of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America*, *107*, 2711-2724.

Iverson, P., Kuhl, P.K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Ketermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*, B47-B57.

Kuhl, P. K., Conboy, B.T., Coffey-Corina, S., & Padden, D., *et al.* (2007). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of The Royal Society* B, *363,* 979-1000.

Kuhl, P.K. (1991). Human adults and human infants show a 'perceptual magnetic effect' for the prototypes of speech categories, monkeys do not. Perception & *Psychophysics*, *50*, 93-107.

Kuhl, P.K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences USA*, *97*, 11850–11857.

Kuhl, P.K., & Iverson, P. (1995). Linguistic Experience and the perceptual magnet effect. In: W. Strange (Ed.), *Speech Perception and Linguistic Experience* (pp. 121-154). Timonium, MD: York Press.

Leather, J. (1999). Second-language speech research: an introduction". In: J. Leather (ed.), *Phonological Issues in Language Learning*, 1-58. Oxford: Basil Blackwell

Llisterri, J. (1995). Relationships between speech production and speech perception in a second language. In: K. Elenius & P. Branderud (eds.), *Proceedings of the 13th International Congress ofPhonetic Sciences,* Vol. 4, 92-99. Stockholm: KTH / Stockholm University.

MacKay, I.R.A., Flege, J.E., Piska, T., & Schirru, C. (2001). Category restructuring during second-language speech acquisition. *Journal of the Acoustical Society of America*, *110*, 516-528.

Major, R.C. (2001). *Foreign Accent*. Mahwah, NJ: Lawrence Erlbaum Associates.

Major, R.C. (2002). The phonology of the L2 user. In: V. Cook (ed.) *Portraits of the L2 User*, 65-92.Clevedon: Multilingual Matters.

Neufeld, G.G. (1988). Phonological asymmetry in second language learning and performance. *Language Learning*, *38*, 531-559.

Polivanov, E.D. (1931). La perception des sons d'une langue étrangère. *Travaux du Cercle linguistique de Prague*, *4*, 79-96.

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

Strange, W. (1995). Cross-language studies of speech perception: A historical review. In: W. Strange (Ed.), *Speech Perception and Linguistic Experience* (pp. 3-45). Timonium, MD: York Press.

Strange, W. (2006). Second-language speech perception: The modification of automatic selective perceptual routines. *Journal of the Acoustical Society of America*, *120*, 3137.

Strange, W. (2007). Selective perception, perceptual modes, and automaticity in first- and second-language speech processing (A) *Journal of the Acoustical Society of America*, *122*(5), 2970-2970.

Strange, W., & Shafter, V.L. (2008). Speech perception in second language learners: The re-education of selective perception. In: Edwards, J.G.H. & Zampini, M.L. (Eds), *Phonology and Second Language Acquisition* (pp.153-191).Amsterdam/ Philadelphia: John Benjamins Publishing Company.

Strange, W., Bohn, O.S., Trent, S.A., & Nishi, K. (2004). Acoustic and perceptual similarity of north German and American English vowels. *Journal of the Acoustical Society of America*, *115*, 1791-1807.

Trubetzkov, N.S. (1969). *Principles of Phonology*. Berkeley, CA: University of California Press.

Williams, L. (1977). The perception of consonant voicing by Spanish English bilinguals. *Perception& Psychophysics*, *21*, 289-297.