The ‘Parity and Disparity’ approach to speech perception in second language learning: A dynamical systems perspective*

Naoshi Hosomura
University of Toronto

In the studies of speech perception in second language (L2) learning, researchers seldom claim that speech motor control plays a role in the perception of speech. On the other hand, Walker (2000) claims that the greater perceptual difficulty certain non-native speakers experience in perceiving native speakers’ speech is rooted in the limited motor control skills these non-native speakers have. If so, rigorous training of speech motor skills will enable these learners to better parse speech signals produced by native speakers. Converging evidence from the empirical studies of the motor theory of speech perception and recent studies of cross-language speech perception provides theoretical grounds to such a ‘motor skill approach’ to perceptual learning. In this paper, a new approach, called the Parity and Disparity (P&D) theory, is presented as a synthesis of these theoretical backgrounds, using general concepts of Dynamical Systems Theory (DST) as applied to speech motor control.

1. Introduction

In 1959, Charles Percy Snow expressed his deep concerns over the disintegration of knowledge in modern society in his influential lecture titled “The Two Cultures” (Snow, 1959). As both a literary man and a scientist, Snow criticized the widening gap and lack of communication between the practitioners of science and humanities. Snow’s criticisms on the loss of integration of knowledge are even more applicable to the state of today’s intellectual world, where increasing specialization of knowledge is leading to dichotomy not only between science and humanities, but even within narrow subdisciplines of sciences, and more importantly, between the industry and the academia (Tachibana, 2000). Second language (L2) acquisition research is not an exception to this inherent problem of a highly specialized academic society. While speech is probably the most complex motor behavior humans can perform, theorists in L2 acquisition seldom make reference to motor control theory, or the fundamental laws of physics that govern them. In addition, there is a breakdown of communication between academic research in this area and English language

* The author is grateful to Laura Colantoni and Pascal van Lieshout for useful comments and guidance in the preparation of this paper.

Toronto Working Papers in Linguistics 25: 103–113
Copyright © 2005 Naoshi Hosomura
teaching industry, as evident from the fact that the literature of the former does not reflect the practices of the latter.

Learning to perceive a foreign language is a challenge, and thus auditory perception has been the focus of much research in the field of L2 acquisition studies. In the past two decades of cross-language speech perception research, much attention has been paid to the perception of phonemic segments (excellent reviews can be found in Strange, 1995). While research on the perception of segments has certainly elucidated the nature of speech perception, there is an elusive aspect of speech perception that cannot be captured from the studies of segmental perception. That is, while native speakers of some languages seem to experience much greater difficulty in the perception of certain L2 than the native speakers of other languages, it is difficult to observe such global cross-language trends in perceptual ability from studies of segmental perception. For example, at segmental levels, Dutch speakers’ perception of English vowels can be discussed in the same way as Japanese speakers’ perception of English vowels, without regard to the difference in the ability to catch English between native speakers of the two languages. In the case of the native Japanese speaker’s perception of English, much attention has been paid to their perceptual (in)ability to discriminate between /l/ and /r/. But the nature of the difficulty Japanese speakers experience in perceiving English is, I argue, rooted in a more fundamental problem than the perception of non-native phonemic contrasts. I propose that the fundamental problem lies in a difference in speech motor skills between native English and Japanese speakers.

Walker (2005) argues that for adult ESL learners whose native languages are outside of the Indo-European language group, pronunciation expertise is the most important element contributing to listening comprehension ability of English. According to Walker, the primary reason adult Japanese-speaking ESL learners experience considerable difficulty in perceiving English is that they cannot imagine how native speakers move their mouths, because they cannot move their articulators adroitly enough to produce native-like English (Walker, 2005). While learners typically approach the problem by spending long hours at listening training, Walker claims that these learners should devote their energy to training their pronunciation skills. He argues that developing the ability to spontaneously produce English in a native-like manner will, in turn, enable the learners to be skilled at ‘feeling’ the way native speakers of English employ their musculatures during vocalization (Walker, 2005).

Motor involvement in auditory perception is not frequently addressed in the context of L2 speech learning. Although Walker’s approach has long been practiced in Japan, and anecdotal comments made by his students unanimously include improved perceptual ability, empirical evidence in support of such claims are extremely limited. Yet, Walker’s claims are compatible with the claims of the Revised Motor Theory of Speech Perception, namely the claim that speech perception makes use of the speech motor system, and that perception is made in reference to potential vocal tract action (Liberman and Mattingly, 1985).

The goal of the theoretical work presented in this paper is to provide a synthesis of theoretical backgrounds for a ‘motor skill approach’ to speech perception, from a Dynamical Systems Theory (DST) perspective (for a comprehensive review of DST and its application in speech, see van Lieshout, 2004). Citing evidence from recent cross-language studies of speech perception, I will provide a motor theorist view of cross-language speech perception. A sketch of the Parity and Disparity (P&D) theory is presented, which relates speech motor skills to speech perception in DST terms, in a broad cross-linguistic context.
appropriate, general concepts of coordination dynamics are used. I conclude by discussing some aspects of Walker’s approach, as well as possible future research directions.

1. A motor theorist view of cross-language speech perception

The Revised Motor Theory of Speech Perception (Liberman and Mattingly, 1985) claims that the objects of speech perception are the speaker’s intended phonetic gestures, and that the motor system is involved in the perception of speech. Speech gestures are represented in the brain as motor control structures, and the same entities that are used in speech perception also command movements of the articulators in speech production. There is a body of behavioral and neurophysiological evidence in favor of the motor theory’s claim of linkages between mechanisms that support motor performance and those that support perception, although the ubiquity of such links disfavors motor theory’s claim of a unique module to perceive speech gestures (e.g., Kerzel and Bekkering, 2000; Fowler and Galantucci, 2005).

It has been commonly argued that perceptual mastery must precede accurate production of new L2 phones and phonological contrasts (e.g., Flege, 2003). Therefore, standard pedagogy consists of training of aural discrimination of target phonemes by means of minimal pair contrast drills (e.g., Yamada, 1995). However, recent experimental evidence favors Walker’s claim that speech motor skill training is a more effective method to improve the ability to perceive L2. Reed and Warsi (2005) investigated whether improved ability to produce English /l/ and /r/ accurately would enable adult Japanese-speaking ESL learners to perceive these English liquids correctly. While control group subjects listened to an audiocassette and repeated words and sentences, similar to typical perceptual learning conditions, the experimental group was not provided with external auditory modeling. Instead, diagrams and instruction explaining articulatory movements were combined with explicit feedback from the investigators which informed subjects when their pronunciation approximated the target. The training produced better perception and production of /l/ and /r/ in the experimental group, suggesting that speech production which converges with the target sound will facilitate speech perception, in support of Walker’s claim.

There is yet another body of experimental evidence for the motor involvement in speech perception. While for native listeners intelligibility of the native talkers is greater than the intelligibility of any of the non-native talkers, for non-native listeners, non-native talkers from the same native language background are often more intelligible than the native talkers (e.g., Bent and Bradlow, 2003; van Wijngaarden et al., 2002; Sereno et al., 2004). In other words, not all foreign speakers of a certain L2 are equally easily understood, and the difficulty depends upon both the language of the speaker and the listener (Sebastian-Galles, 2005). From a motor theoretical point of view, these findings suggest that speech intelligibility is crucially dependent upon the similarity in speech motor skill between the listener and the talker.

2. Explaining cross-language speech perception: a sketch of the P&D theory

The P&D Theory has been developed to provide theoretical accounts to the global trends of cross-language speech perception, based upon the assumption that there is a close link between production and perception mechanisms. In physics, parity refers to the mirror-image symmetry of this world; that is, the laws of physics are neither right-handed nor left-handed.
For example, situations such as the law of gravity always acting stronger to the left direction than to the right direction, or heat always flowing to the left direction and left side being always hotter than the right side, will not occur in nature (Tachibana, 2000). The conservation of parity is considered to be one of the fundamental laws of nature (although parity violations do exist, see Feynman, 1963). In the context of speech perception, ‘parity’ means “a relation of sufficient equivalence in phonological messages sent and received” (Fowler and Galiantucci, 2005; p 634). I refer to the failure to achieve this ‘parity’ requirement as ‘disparity’. Here I assume that the objects of speech perception are, minimally, gestures realized by coordinative structures, by which synergies or functional grouping of muscles are connoted (Fowler and Rosenblum, 1989; Kelso et al., 1983). I argue that, in order for the listener to accurately perceive the language forms that the talker produces, there must be ‘parity’ in the motor control skills between the listener and the talker. In this view, the listener’s ability to parse the speaker’s intended phonetic gestures, including the ability to decode the effect of coarticulation, will crucially depend upon the degrees of (dis)similarity in the intrinsic motor coordination tendencies between the listener and the speaker in performing the same speech motor task. In other words, ‘disparity’ in the coordination dynamics between the listener and the talker in performing a given speech motor task will limit the listener’s ability to recover the talker’s intended phonetic gestures.

‘Motor skill’ may not necessarily be behaviorally observable. An obvious counter-example to the above claims would be that there are individuals whose pronunciation is far from native-like but have good listening comprehension skills. However, I claim that there is motor involvement in the perception of speech even for such individuals. Although they may not be able to implement native-like motor coordination at the articulatory level, they may have acquired the appropriate abstract gestural configuration or score, which is independent from the actual vocal tract configuration (e.g. Browman and Goldstein, 1992). In other words, it may be the case that although they do share ‘parity’ with the talker in the intended phonetic gestures and thus are able to accurately parse the talker’s gestures, they have problems in actually implementing the gestures due to the competitions between the task demands and intrinsic motor coordination tendencies. It was shown in the coordination of finger movements that even when subjects tried to generate a variety of different phase relationships48 in their relative movement of right and left index fingers, the phase relationships converged to a few stable patterns (Zanone and Kelso, 1992). I argue that similar competitive processes occur for speech motor control in the context of second language learning. Learning is difficult when intrinsically stable movement patterns do not cooperate with new task demands, and in such cases learners have difficulty in implementing vocal tract gestures in a native-like manner. The complications mentioned here are applicable to discussions presented in the following sections.

It is also important to define the scope of the theory. The P&D theory concerns only with the listener’s ability to parse phonetic gestures produced by the talker, without regard to the process of associating the gestures with meaning. For example, the theory predicts that native English speakers can perceive non-sense words produced by a native English speaker better than native Japanese speakers can. Here, meaning of the gestures are not relevant.

---

48 The relative phase relationship of 0 degrees, for example, means that both fingers go up and down at the same time, and the relative phase relationship of 180 degrees means that the two fingers are moving up and down in the opposite directions.
Language imposes an arbitrary coding on the relationship between speech gestures and perception, so the same gesture could mean different things in different languages. But the scope of the P&D theory is on the perceptual ability to parse phonetic gestures, and it does not matter what the gestures encode. Below, I will introduce the essential components of the P&D theory.

Disparity Index (DI) is a measure of the magnitude of dissimilarity in the coordination dynamics that exist between the speaker and the listener in performing given speech motor tasks. The coordination dynamics express the organism’s preferred and stable coordination tendencies that exist at any moment (Zanone and Kelso, 1994). Disparity in the coordination dynamics refer to the discrepancy between the two communicative parties in their motor strategies to control the degrees of freedom in the vocal tract system. Such motor control strategies can be modeled in terms of attractor states in which specific interactions between the component articulators form relatively stable patterns (van Lieshout, 2004). The magnitude of the (dis)similarity in the coordination dynamics between the speaker and the listener is defined along the continuum of parity and disparity. DI can be measured by obtaining behavioral information of the motor performance during the execution of specified motor tasks, typically with reference to native speakers’ motor performance on the same tasks.

For example let us set the reference to be native English motor performance in a specified speech motor task. I set the object of measurement as the ‘disparity’ in L2 speakers’ performance from native English motor performance. The magnitude of DI could be measured as, for example, 1 for a Dutch speaker, 2 for a German speaker, and 6 for a Japanese speaker (based upon Walker, 2000).

Of course, ‘native performance’ may not be in a singular dimension. Since coordination dynamics are an emergent product of internal preferences given certain biomechanical, physical, and physiological constraints, task demands, and environmental embedding, coordination dynamics is to an extent unique to each speaker and listener, even within the same language environment. This was shown for specific kinematic parameters as the variability in obstruent gestural specification in the study by Alfonso and van Lieshout (1997). While there is certainly such intra-subject variability in many measures of performance even among native speakers, I claim that there are “common coordination patterns” in the motor performance of the native speakers of a given language.

The “common coordination patterns” distinguish between skilled (in this context, native) and less skilled (non-native) performers. There is often an underlying coordination pattern that is remarkably consistent within different skill groups and that needs to be assembled before an individual can progress to the next level of skill (Temprado et al., 1997; Bennett, 2003). In the context of this paper, the ‘different skill groups’ mean native speakers of different languages. There are examples from speech data showing a consistent relative phase relationship for the coupling between gestures in specific speech tasks among native speakers of the same language (van Lieshout et al., 1999). I claim that such “common coordination patterns” are language-specific, and for this reason speech motor skill levels of speaking English, for example, is different for the native speakers of different languages. Therefore, native Dutch speakers, with low DI, meaning there is large similarity in the speech motor skills to produce native-like Dutch and English, generally sound far more native-like in speaking English compared to native Japanese speakers, whose speech motor skills are much more dissimilar to those of native English speakers, characterized by high DI.
Since the P&D theory is based upon the motor theory’s assumption that the same entities used in speech perception also command movements of the articulators in speech production, ‘disparity’ in the motor domain will be mirrored in the perceptual domain. Tuller and Kelso (1990) showed that in saying /ip/ repeatedly at increasing speaking rates, the relative phase of articulator movements changed markedly at critical values (i.e. /ip/ became /pi/), and these transition points corresponded precisely to perceptual changes in syllable affiliation of the consonant. Listeners were able to accurately perceive changes in gestural information even though such transition was not clear in acoustically. By taking such a close fit between perception and production into account, I propose that DI is not only a measure of dissimilarity in speech motor skills, also a measure of perceptual difficulty.

In a motor theoretical interpretation of speech signals, the portions of speech signals from which the listener cannot recover the talker’s intended phonetic gestures are processed as non-speech signals, and thus can become a source of perceptual difficulty. The question is, when can the speaker’s intended phonetic gestures not be parsed as phonetic gestures by the listener? It has been shown that, for example, when listeners hear lingual consonants, there is often enhanced muscle activity in the tongue (Fadiga et al. 2002). Simply put, gestures that cannot trigger such a response in a listener will not be perceived as phonetic gestures by the listener. More specifically, when the speaker’s intended phonetic gestures involve coordination patterns that cannot be readily organized by the listener’s existing motor skills, then such gestures or the combinations of them will be parsed as something unusual or unknown to listener, and thus will not be perceived as speech gestures. Coarticulation is also relevant. The capacity to coarticulate must have coevolved with that of decoding the effects of coarticulation in the acoustic signal, and therefore these capacities are both grounded in a common mechanism (Fowler and Galantucci, 2005). Thus, when the talker’s coarticulatory patterns are disparate from those of the listener, then the listener will have difficulty in decoding the effects of coarticulation and thus experience difficulty in perceiving the talker’s speech. The more unusual or unknown the phonetic gestures of the speaker are to the listener, the greater the component which listeners cannot accurately perceive. Thus, when the speaker’s coordination of phonetic gestures are more disparate from what the listener produces, i.e. the higher the DI, speech perception becomes more difficult.

This concept of DI captures precisely what I alluded to in the introduction as an “elusive aspect” of cross-language speech perception. I hypothesize that the primary reason why native Japanese speakers have disproportionate difficulty in parsing English speakers’ speech is due to high DI. This is why I identified the fundamental problem of L2 acquisition to be at the level of speech motor skills. Native Japanese speakers’ DI is very large in the task space of speaking English, and this large DI affects both speech production and perception. This might explain why native Japanese speakers typically find normal speed utterance of English to be “too quickly spoken to catch” (Walker, 2005).

Another important point about the notion of DI is that it is largely independent of what language (adult) talkers produce. Sebastian-Galles (2005) notes that “for Spanish listeners, Italian and Greek natives are easier to understand than German and Japanese natives, when speaking in English. But this is not the case for Dutch listeners (p. 555)”. Assuming the same levels of lexical and syntactic knowledge in the speaker and the listener, this trend will also apply for other languages. That is, for Spanish listeners, Italian natives will be more intelligible than Japanese natives when speaking any other foreign languages. The key point to note here is the low DI between Italian and Spanish speakers, regardless of the languages.
these talkers produce. This is because whatever language typical late learners of foreign languages try to speak, their motor control patterns converge to the attractor states that they formed while learning their native language. In other words, the attractors for speaking their native languages compete with the coordination patterns those speakers try to generate in speaking a L2. Another example is the polyglot who learned foreign languages after puberty. If that polyglot’s native language is Japanese, then he or she will sound as if speaking Japanese even though he or she is actually speaking a foreign language. The DI will be near zero for native Japanese listeners who hear that polyglot’s speech, and thus native Japanese listeners will be able to catch the polyglot’s phonetic gestures accurately even though they may not know the meaning of the gestures.

The last component of the P&D theory relates to evolution. The magnitudes of DI are more or less proportionate to the diachronic separation between the learner’s native language and the target L2 (cf. Walker, 2000). In other words, DI is roughly a function of time, as speech ‘task spaces’ naturally drift for a variety of reasons (e.g. imperfect imitation and creation, much as mutation and recombination of the DNA, see de Boer, 2001). It should also be noted that speaker migration and language contact will decrease the magnitudes of DI. For example, it has been shown that the intonational convergence of Buenos Aires Spanish and Italian is likely to be the consequence of speaker migration (Colantoni and Gurlekian, 2004). In P&D terms, native speakers of Buenos Aires Spanish will have lower DI, in terms of intonation, in performing specified speech tasks of speaking Italian, compared the speakers of other Spanish dialects. It is expected that native speakers of Buenos Aires Spanish will be able to speak Italian with native-like Italian intonation, as their intrinsically stable motor control tendencies will effectively cooperate with the new task demands, whereas native speakers of other Spanish dialects may have more difficulty due to the competition between the existing attractor states and the new task demands.

3. A ‘parity’ fostering approach to speech perception in L2 learning

In this section, I illustrate how Walker approaches the problem of speech perception by his ‘speech motor skill approach’, from a DST perspective. Walker devised a comprehensive English pronunciation workout called The Jingles, aimed at developing native-like speech motor skills (Walker, 2000). To use terms of the P&D approach, Walker identified which synergies and coordination patterns must be practiced in order to establish behavioral attractors that would allow native-like gestural coordination patterns in producing English, and designed the system so that it can effectively facilitate the formation of appropriate attractor states. For example, the first sentence of the system is: “From early on, three thoroughly surly thirsty thugs thought they could take baths throughout the night.” Any non-native speaker of English will notice that the motor performance of pronouncing this sentence clearly distinguishes native and non-native speakers of English, as the attractors needed to produce the sequence of gestures in the sentence with accuracy and stability are likely to be found only in the motor control system of a native English speaker.

What sets The Jingles apart from any other kinds of pronunciation drilling is the use of ‘training mode’ speech, which is characterized by the production of extremely effortful and forceful gestures (for a demonstration of the ‘training mode’, refer to the audio CD in Walker, 2000). The use of such effortful movements seems to fit very well with current theoretical notions about noise in force production and the principles of coordination
It is known that with more force output, noise in the motor system increases (Slifkin and Newell, 2000). From a DST point of view, noise is a mechanism for change. That is, by increasing noise, one can make instabilities in the motor system, allowing it to probe for new solutions. This mechanism is probably how The Jingles can allow a native speaker of Japanese to find the appropriate attractor states for English pronunciation. Application of extreme force would induce noise in the motor system, which potentially destabilizes existing solutions and allows the system to explore new stabilities for the L2 pronunciation. Although the mechanism described here is still speculative, this is well in line with the principles of DST and the concept of coordination dynamics.

In the training, students first practice hyper-speech, and then they gradually shift toward practicing hypo-speech, in Lindblom’s H&H paradigm (1990). Experimental evidence indicates that smaller movements (i.e. hypo-speech) could induce instabilities in gestural or articulatory coupling (van Lieshout, 2004). Since hypo-speech is inherently more unstable than hyper-speech, it is typically very difficult for L2 speakers to effectively undershoot gesture targets in hypo-speech without losing “sufficient discriminability”. The Jingles allows students to develop native-like speech motor skills, enabling them to pronounce English in a native-like manner along the full range of the hyper- and hypo-speech. This in turn allows them to accurately perceive the hypo-speech of native English speakers. Importantly, this rigorous motor skill training is not limited to oral motor skills, but the training also works on the coupling dynamics of all associated musculature, including laryngeal and respiratory muscles. By following the training program outlined above, students can develop native-like attractor landscape in the motor domain, and their perceptual skills improve accordingly.

The Jingles has another major practical application in addition to its utility for motor skill training in ESL teaching. The analysis of the motor performance of the speech tasks defined by The Jingles would allow the quantification of DI, as the rigorous task demands of the Jingles sentences can reveal spontaneous coordination patterns of the speakers. I argue that the Jingles sentences could be used to create a reference database for coordination dynamics as would be typical for native English speakers, and the differences found for ESL learners. For example, one can gain insights into the differences in the intrinsic oral motor coordination tendencies between native Japanese and Dutch speakers, by asking them to produce the sentence “From early on, three thoroughly surly thirsty thugs thought they could baths throughout the night” and comparing their performance to that of native English speakers. Thus, The Jingles can serve as a powerful research tool, allowing researchers to probe the coordination dynamics of the subjects.

There is yet another benefit of the Jingles training. Walker (2000) claims that as students approximate the ‘native equivalent level’ in English pronunciation, then their pronunciation of related languages to English, such as German, also becomes significantly more intelligible compared to monolingual native Japanese speakers. In P&D terms, this is because the newly formed attractors allow them to pronounce German at native English speakers’ level (DI ≈ 2) as opposed to native monolingual Japanese speakers’ level (DI ≈ 5). This is an interesting claim, but definitely in need of experimental verification.
4. Conclusion

The aim of theoretical work presented above was to provide a unifying theory to the theoretical backgrounds which previously existed independently of each other. I have provided a synthesis of “The Two Cultures” – The Jingles from the ESL industry, and the motor perception theory, DST (coordination dynamics, task dynamics), and studies in cross-language speech perception from academia. When seen from the perspectives of the P&D approach, Walker’s approach to speech production and perception in L2 learning seems to fit very well with current theoretical notions about motor involvement in speech perception and general principles of dynamical systems theory, in particular coordination dynamics. Walker’s work deserves further attention and efforts, as it offers new tools to studying the relation of speech production to speech perception. Such a ‘motor skill approach’ may lead to improved pedagogical methodology in language education. In addition, The Jingles model could provide researchers with a powerful research tool. The system can provide a way to quantify speech motor skills, using the paradigm of the P&D approach.

The P&D approach unifies a variety of relevant theoretical backgrounds related to cross-language speech perception. To use David Deutsch’s words49, the ‘Fabric of Cross-Language Speech Perception and Production’ is woven from four strands: (1) The Jingles (Walker); (2) DST (Kelso, van Lieshout); (3) gestural theory of speech perception (Liberman, Fowler); and (4) the theory of evolution (Darwin, Dawkins), with the resulting ‘fabric’ being the P&D theory. The P&D paradigm and its practical counterpart The Jingles can thus provide a unique opportunity for motor control theorists, phoneticians, and L2 acquisition researchers to work together to explore alternative research designs. Although the theoretical work is still in its developmental stage and needs empirical evidence for its claims, I hope the P&D approach will contribute to the unity of knowledge in the study of cross-language speech perception.

References


49 David Deutsch, a prodigy of quantum computation, identifies the four strands of ‘the Fabric of Reality’ to be: the quantum physics of multiverse, Popperian epistemology, the Darwin-Dawkins theory of evolution, and a strengthened version of Turing’s theory of universal computation.


