The perception (and production) of English word-initial consonants by native speakers of Cantonese

Christina Chan Pik Ha
The University of Hong Kong

Abstract

With the view to develop their professionalism, second language (L2) teachers and educators are expanding their research and teaching diversity. However, phonology has remained an area of general neglect both in terms of classroom practice and research. A closer look into current phonology programs reveals that the majority of L2 teachers approach their teaching from a restricted perspective that focuses mainly on the place and manner of sound production. This ignores other levels of speech processing, thus contributing to learners’ achievements being limited. This study attempts to explore a higher level of speech processing, the perceptual level, and its relationship with the production of the same sound. Local Cantonese subjects, aged between 21 and 29 and of similar English experience and proficiency, were assigned to two groups. The subjects in Group A were those who consistently mispronounced the English word-initial consonants /v, ð, z, r/ as [w, f, s, w] respectively while those in Group B consistently produced the same sounds correctly. These subjects were involved individually in a forced-choice phonemic perception test of the contrast-pairs. Results indicate a significantly poorer perceptual performance of Group A as compared to Group B which support the perceptual based hypothesis of speech accuracy. A discussion of the possible implications of these findings for L2 speech pedagogy and future research directions conclude this paper.

Current practice in second language phonology

In order to develop their professionalism, second language (L2) teachers and educators have been expanding their research and teaching diversity. However, in spite of the increased demand for L2 phonology training, especially from those who have achieved proficient L2 grammar, phonology has been generally neglected in L2 research and teaching (Leather, 1983; Morley, 1991; Pennington, 1998). It was not until the 1980s that the situation began to change.

Pennington (1998) observes that the neglect of L2 phonology teaching is partly due to the long-standing theoretical (but controversial) notion of the Critical Period Hypothesis (Lenneberg, 1967) that may lead many L2 teachers
to believe that teaching pronunciation\(^1\) is useless. The second reason for the neglect relates to \textit{practical} issues. Due to the scarce empirical and methodological directions available, many L2 teachers find phonology a difficult area to handle and one where significant success in classroom practice seems less warranted. Even for those who intend to include pronunciation in their curricula, there is little detailed guidance available on what the syllabus should include or how best to approach the teaching (Leather, 1983).

The Critical Period Hypothesis, however, may not provide a reasonable explanation for the difficulty learners encounter. Long (1990), for example, points out that even if plasticity loss is a fact, it is not a sudden effect that causes the same sudden loss in L2 learning. It would also be difficult to accept that the specific effect of such a “barrier” applies solely to language learning (and especially phonology learning) but not to other kinds of learning. First language (L1) interference or negative transfer would thus appear to be a more plausible reason for L2 learners’ accent.

The acceptance of this view has significant theoretical and practical implications. First, it implies that learners’ pronunciation problems may not be resistant to change and second that greater success may be achieved when learners are provided with the appropriate teaching methodology to minimise interference from the L1. However, lack of research support and the ongoing neglect of phonology teaching in classroom practice are now generating a vicious circle which is hindering further empirical development and classroom success in this area.

From a practical perspective, a detailed evaluation of the present pedagogical materials and teaching approaches may also help explain the current deficiencies in the teaching of phonology. Little attempt appears to have been made in current materials to analyze learners’ level of breakdown in speech processing and the study of L1 influence on L2 phonology learning has also been largely neglected.

Many researchers (e.g. Ferguson & Macken, 1980; Hewlett, 1990) have noted that speech production is a process (rather than a product) comprising different phases --- \textit{perceptual}, \textit{programming}, \textit{processing} and \textit{execution} levels. However, most teaching materials of L2 phonology attend mainly (if not only) to the \textit{place} (i.e. where the tongue and lips make contact) and \textit{manner} (i.e. how the airstream from the lungs and vocal tract is shaped) of articulation. Such

\(^1\) The terms “pronunciation”, “articulation” and “speech production” in this study are synonymous.
focus on the *execution phase* of speech production represents a very restricted analysis of the learners’ difficulties. Failure to consider the problem in relation to other levels and working only on the physiological muscular sequence may not guarantee learners’ success. The need to explore a higher level of speech processing, specifically the level of perception, is justified and thus forms part of the basis for this study.

Another problem with the current methodology is that even when perception exercises are involved, the influence of the learners’ L1 phonological system (i.e. L1 transfer) is often ignored. This results in a standard minimal pair approach to perceptual training within the confines of the L2. In constructing the representation of an L2 sound, many researchers (e.g. Best and Strange, 1992; Flege, 1991; Hacin-Bhatt, 1994) believe that learners tend to perceive the L2 sounds by drawing similarities between these unfamiliar sounds and L1 sound categories. Therefore, without including the L1 sound system in designing the L2 curriculum, L2 teachers will continue to encounter much difficulty in understanding learners’ level of breakdown in correct speech production. More importantly, learners will not receive the corresponding feedback for attending to the specific characteristic aspects of L2 sounds.

### The importance of perception to production

In addition to the theoretical concerns for perceptual training in classroom practice, much of the current empirical evidence justifies the corresponding need to study the relationship between perception and production. An increasing number of positive findings have supported this perceptual based hypothesis of speech accuracy.

Spanish has been one of the most extensively studied languages in the literature of L2 perception and production. Flege and Eefting (1987), for example, examined the difference in the perceptual performance of two groups of subjects. One consisted of adult English monolinguals. The other group comprised native Spanish speakers, aged between 9 and 10 with about 4 years of English experience, who often produced the English /d/ and /t/ with a Spanish accent. The subjects in each group participated in a perceptual decision task in which they made forced-choice decisions on whether the presented auditory stimuli were English /da/ or /ta/ in the corresponding voice-
The perception (and production) of English word-initial consonants

onset time\(^2\) continuum. Results indicate that, similar to their production performance, the native Spanish speakers showed a significantly poorer perceptual performance as compared to the English controls. The researchers conclude firstly that the perception and production of English /d/ and /t/ by the Spanish speakers were parallel. Secondly, the influence of the L1 might affect the subjects in their construction of the perceptual representation of L2 speech sounds.

In another study that looked at the effectiveness of perceptual training on speech accuracy, Schneiderman et al. (1988) explored the perception and production of adult learners of French before and after a training program on pronunciation. The subjects were of different L1 backgrounds, including English, Chinese, Tamil, Hindi, Turkish, Spanish and German. These subjects, aged between 18 to 60 and at beginner and low-intermediate levels of French, were further divided into trained and untrained groups. Each group participated in the discrimination and production tests before and after a training period. The discrimination test comprised the same-different judgment on two levels - (i) pairs of French words which were either identical or which differed by only one segment and (ii) sentences of the same or different intonation and rhythmic patterns of French. The production test investigated their imitation ability of French words and sentences. Two important findings were obtained. First, the subjects’ discrimination and production scores correlated significantly with one another both pre- and post-training. This suggests that perception and production are always parallel at various stages of L2 learning. Secondly, the trained groups were noted to have made significant gains in both perception and production on all measures except for the discrimination of rhythmic patterns, while the untrained groups did not show evidence of such gain. The researchers conclude that the production performance of a second language, at least on the segmental level, is positively associated with the perceptual performance and that an improved discrimination ability would result in a more native-like production. These findings provide more direct support for both the learning mechanism and for methodological directions for L2 phonology teaching. In particular, the need for perceptual training is further justified.

Motivated by these positive findings, studies have been extended to explore the generality of the perceptual based hypothesis of speech accuracy in some Asian languages. For instance, Bradlow et al. (1997) observe that perceptual training has enhanced the intelligibility of Japanese learners’

\(^2\) Crystal (1991) defines “voice-onset time” as the time spot at which the vocal cords start to vibrate. In a voiced plosive, the vocal cords vibrate throughout while in a voiceless plosive, there is a delay before voicing starts. The delay is also language-specific.
productions of the English /r-l/ contrast. In this study, Japanese speakers of university level were involved in a perceptual training program of English /r-l/ minimal pairs. Results indicate that all subjects showed a significant improvement in the perception of the same contrast after training. Such perceptual learning was also evidenced in novel items spoken by other native speakers. Their post-training production of the same /r-l/ minimal pairs was more accurately identified by native English listener judges. Bradlow et al. conclude that the perceptual knowledge gained in the perceptual training of /r-l/ was transferred to subjects’ production domain.

Despite the cumulative positive findings, some researchers including, Flege and his co-researchers (1996, 1997) question the validity of the results because of the loose control of confounding variables like the subjects’ starting age of L2 learning, length of L2 experience and intensity of native speaker contact. Ingram and Park (1997) thus adopted a carefully controlled design in their investigation of the perception of non-native vowels in Japanese and Korean learners of English, who are observed to produce English /a/ and /æ/ without differentiating between them. Subjects were grouped according to their English experience. The inexperienced group, aged in their twenties, had been resident in Australia for under a year while the experienced group, aged in their thirties, had more than 5 years residence in Australia. The researchers employed a modified forced-choice perception test. On listening to a tape of stimuli, the subjects were to circle one of the 5 English words in the form of /h_d/ where the blank was embedded with the vowels /i:/, /æ/, /æ:/, /a:/ which are also contrastive in their languages. Ingram and Park account for this by explaining that due to the non-contrastiveness of [e] and [æ] in their native languages, Japanese and Korean subjects showed difficulty in tuning in to the difference. They further suggest that the English /e/ and /æ/ might be stored in the same sound category which form the same basis for their production. This further supports the mirror notion of perception and production.

Although scarce in number, these empirical results provide encouraging support for the continuous exploration of the impact of L2 speech perception on production.
Motivation for the present study

Despite the limited success of current L2 phonology programs, L2 phonology teaching should be valued equally with other aspects of L2 teaching. As Leather (1983) and Morley (1991) point out, with the rise of communicative language teaching and the learner-centered approach, there is in fact a growing need for fluent communicative speech patterns (i.e. pronunciation) from different learners. Supported by the positive evidence in the literature, the perceptual performance and its relationship to production in L2 phonology continues to be worthwhile direction for further research. Findings from such research could have important implications for our understanding of the learning mechanisms employed in L1 and L2 learning and most importantly, suggest the WHAT and HOW for future L2 programs. In particular, the emphasis of perceptual training using bilingual minimal pair discrimination exercises in relation to the learners’ native language influence would be better justified. The present study thus extends the interests of these perceptive ideas to the Cantonese population which has not been studied extensively. Specifically the research set out to address the following questions:

1. Do Cantonese speakers of English with speech errors in the specific English word-initial consonants /v, θ, ð, z, r/ have poorer phonemic perception of the same sounds, presented as the contrast-pairs of /v-w/, /θ-f/, /ð-d/, /z-s/, /r-w/, as compared to speakers with correct production?
2. Is there a positive relationship between perception and production in general?
3. What are the implications of the findings for L2 phonology teaching?

Methodology

Minimal pair contrasts

The choice of consonants was based on the differences between the Cantonese and English sound systems. As /v, θ, ð, z, r/ are not present in the Cantonese sound system, Cantonese speakers might not be able to transfer the production skills positively from their L1 into speaking English. It is a commonly observed phenomenon from daily interaction that these are the English phones frequently mispronounced by many Cantonese speakers. It was hoped that selecting data in word-initial position would provide some insight into determining the influence of any L1 transfer effect. Controlling the present
investigation to the word-initial position was also necessary because perceptual performance might vary with syllable position (Mochizuki, 1981).

Subjects

The present study employs a quasi-experimental design. Two groups of subjects, each comprising 30 native Cantonese speakers of English, participated in the study. Group A consisted of subjects who consistently mispronounced all English words with word-initial consonants /v, ð, z, r/ across three non-consecutive trials in a production test with the following error patterns:

- /v/ → [w]/ # ______
- /θ/ → [f]/ # ______
- /ð/ → [d]/ # ______
- /z/ → [s]/ # ______
- /r/ → [w]/ # ______

The second group, Group B, was made up of Cantonese speakers of English who consistently pronounced the same words correctly in the same production test.

The control of their production error pattern was also necessary as the perceptual performance (including the accuracy and error pattern) was thought to mirror participants’ production pattern. Thus controlling the error pattern enabled a direct comparison of the relationship between the two domains. To strengthen the internal validity of the study, the two groups were also matched for their age and length of time they had been learning English. In addition, other confounding variables like their English educational level, hearing and oro-motor function were also controlled. Though not explicitly mentioned in the previous studies, the last two were potential complicating factors to the results. A hearing deficit would affect the basic auditory sensations, let alone the interpretation and identification of speech sounds, while problems with the oral articulatory movement would inherently contribute to a discrepancy between perception and production. Therefore, an interview before the perception and production tests was used to ensure that all subjects:

- are local Hong Kong Cantonese, between 21 to 29 years of age, who have not studied abroad and have had no intensive contact with English speakers;

---

3 The words in the perceptual and production tests were the same.
• speak Cantonese and English only without the influence of other languages or dialects since childhood;
• started to learn English as a second language in English-medium schools from about the age of 6 (i.e. at primary 1);
• have attained a form 5 academic level with Grade D or E in English (syllabus B) in the Hong Kong Certificate Education of Examination (HKCEE);
• have been working with not much spoken English use in their present work, thus their English exposure is of about 10 years (from primary 1 to form 5);
• have been diagnosed by a qualified speech therapist as having adequate oro-motor function for speech production;
• have no apparent hearing deficit as evidenced by the pure-tone hearing screening test at octave frequencies between 0.5 and 4kHz at 20dBHL (ANSI, 1970);
• have reported no previous history of speech and language delay in L1 development.

Speaker

A phonetically trained Hong Kong female Cantonese speaker, in her mid-twenties, who received her primary and secondary schooling in one of the best-known girls’ English-medium schools in Hong Kong, was invited as the speaker. She completed her Bachelor degree in Speech and Hearing Sciences at the University of Hong Kong and has just received her Master’s degree in Audiology from the same university. The speaker recorded 2 tokens of each item in the phonemic perception test in a soundproof audiological testing room. Items were elicited in their citation form on the understanding that they would be used in a listening test for non-native speakers. The speaker was judged by two British teachers of English as having a standard British accent. She was also considered to be an educated near-native speaker, representative of the model that the average Hong Kong learner hears most frequently through the media and in teaching environments and is also the model which learners aspire to achieve themselves.

Stimuli

A hundred English monosyllabic words, divided into two parts of 50 words, were used in the phonemic perception test. These comprised 3 examples each of the 5 tested contrast-pairs (see appendix) with thrice occurrence (3x5x2x3)
plus 10 distracter items. The recordings were made using a Sony professional microphone connected to a Sony MZ-R55 Mini-disc Walkman on a Sony professional mini-disc at the intensity of about 65 dBSPL. The auditory stimuli were subsequently digitized at the sampling rate of 44.1kHz with 16-bit resolution.4

The choice of the stimuli was based on three criteria. First, they are monosyllabic. Second, each of them should have a minimally paired substitute which represents the production error pattern. Finally, they should be known and familiar to the subjects so as to avoid them choosing only words they know (Barton, 1980).

To reduce the auditory interference of the previous item, each testing item was separated by a 5-second interval while the two tokens of each item were paused by 1.5 seconds (Ingram and Park, 1997). The likelihood of chance level response was also controlled by providing three choices for each testing item - one target word, one representing the subjects’ error pattern and one distracter which rhymed with the target. To control any response by a particular serial position, targets were randomized with the occurrence at every three possible serial positions. For example, the target “zinc” occurred at the

first position in item 24: zinc sink link
second position in item 60: link zinc sink
third position in item 38: sink link zinc

In an attempt to motivate the subjects to continue with the test, there were 10 relatively easy items inserted in between the targets at a ten-item interval.

Procedure

Pre-test recruitment procedure

Subjects were each seen twice with about a two-week interval between the two sessions. In the first session, they were interviewed to obtain information about their language background, English experience and use, oro-motor and hearing function. They were also asked to tick off words they knew in a randomized list which included more words than those used in the perception and production tests. Only those who knew all the words in the testing targets

4 These sampling rate and resolution degree guarantee a high quality stimulus image.
were recruited in the second stage of the study which included the tests of perception and production. The whole procedure in the second stage lasted about 45 minutes per individual subject.

The perception test

The second session started with the perception test which was carried out in a quiet room. The subjects listened to the stimulus tape binaurally at a comfortable level (about 72 dB peak syllable intensity as suggested by Bohn and Flege, 1990) via Sony high quality headphones connected to the Sony MZ-R55 Mini-disc Walkman. On listening to each stimulus item, the subjects were asked to make a forced-choice response on the response sheet by circling one of the three alternatives written in conventional English orthography. They were also instructed on the tape and from the instructions written on the response sheet to make the best guess in case of uncertainty. Each item was presented twice with intra-stimulus duration of 1.5 seconds before going to the next item. In an attempt to minimize the subjects’ fatigue, there was a 15-minute break between parts I and II of the test. To become familiar with the testing format, each subject practised with 5 unanalyzed items before the real test began.

The production test

In the production test, the subjects were asked to read aloud 45 English words comprising 3 examples of each of the 5 target phones in thrice random non-consecutive order. The inclusion of each example in thrice random order was a control measure to guarantee the consistency of their production performance. The productions were also recorded simultaneously on the Sony high-quality mini-discs using a Sony professional microphone connected to the Sony MZ-R55 Mini-disc Walkman in a quiet room for subsequent inter-rater and intra-rater reliability checks. Only those subjects with the production patterns of: /v/ \rightarrow [w]; /θ/ \rightarrow [f]; /θ/ \rightarrow [d]; /z/ \rightarrow [s]; /r/ \rightarrow [w] were assigned to Group A, while those with correct production to Group B. To avoid any bias or cueing to the items on the perception test, the subjects finished the perception test before the production test.
Data analysis and results

Comparison of perceptual performance

To compare the perceptual performance of the subjects, one mark was given to each correctly identified item and zero to each wrongly identified one on the hundred-item phonemic perception test. Since there were 10 distracter items in the test which were discarded in the analysis, the total possible mark was 90.

To determine if the observed difference between the two groups is one that could likely result from sampling error, a one-way analysis of variance (ANOVA) was conducted. Table 1 below summarizes the mean perception scores, standard deviations and the F-ratio of the group effect for the whole test and for the 5 contrast-pairs across the two groups. The results indicate a significant difference between the groups on the total perceptual scores, calculated for all test items \( F (1, 48) = 125.68, p < .05 \) as well as a significant differences \( (p < .05) \) between the two groups for all the contrast-pairs.

Table 1: Group mean scores for perceptual performance for the whole test and the target contrast-pairs

<table>
<thead>
<tr>
<th></th>
<th>Group A ( (n = 30) )</th>
<th>Group B ( (n = 30) )</th>
<th>F-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole test</td>
<td>54.23 (7.19)</td>
<td>78.30 (5.52)</td>
<td>125.68*</td>
</tr>
<tr>
<td>/v-w/</td>
<td>9.53 (3.16)</td>
<td>16.10 (1.90)</td>
<td>55.18*</td>
</tr>
<tr>
<td>/θ-ʃ/</td>
<td>10.87 (2.26)</td>
<td>14.10 (2.51)</td>
<td>18.90*</td>
</tr>
<tr>
<td>/θ-ð/</td>
<td>11.80 (2.25)</td>
<td>16.50 (2.03)</td>
<td>42.27*</td>
</tr>
<tr>
<td>/z-s/</td>
<td>9.80 (2.63)</td>
<td>14.63 (2.57)</td>
<td>29.98*</td>
</tr>
<tr>
<td>/r-w/</td>
<td>12.23 (2.75)</td>
<td>17.00 (1.43)</td>
<td>47.67*</td>
</tr>
</tbody>
</table>

* = \( p < .05 \), df for group = 1

The results show that the perceptual performance of the subjects who were not able to successfully produce the identified sounds (Group A) is statistically different from that of the subjects who were successful (Group B). Since the latter consistently outperformed the former in discriminating the contrast across all pairs, as evidenced by their higher mean perception scores, it would appear that subjects with production errors on /v, θ, ð, z, r/ have poorer perceptual performance on the corresponding contrast-pairs than those able to produce the sounds.
Relationship between perception and production

Though the mean perception scores between Groups A and B were found to be statistically different, it was still necessary to determine whether the subjects’ perceptual accuracy was associated with and varied according to their groups (and hence their production performance). Since the subjects’ perception scores were recorded on an interval scale while the grouping was dichotomous, the Means analysis was employed which allows the relationship between perception score and grouping to be explored by breaking down the perception score by the group type (Bryman and Cramer, 1997). Table 2 illustrates the strength of the relationship which is determined by the eta-squared\(^5\) value.

Table 2: The F-ratio and eta-squared of the total perceptual scores and scores for each contrast-pair between Groups A and B

<table>
<thead>
<tr>
<th>Contrast-pair</th>
<th>F-ratio</th>
<th>Eta(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total score</td>
<td>125.68*</td>
<td>.79</td>
</tr>
<tr>
<td>/v-w/</td>
<td>55.18*</td>
<td>.62</td>
</tr>
<tr>
<td>/θ-f/</td>
<td>18.90*</td>
<td>.32</td>
</tr>
<tr>
<td>/ð-d/</td>
<td>42.27*</td>
<td>.55</td>
</tr>
<tr>
<td>/z-s/</td>
<td>29.98*</td>
<td>.47</td>
</tr>
<tr>
<td>/r-w/</td>
<td>47.67*</td>
<td>.55</td>
</tr>
</tbody>
</table>

* = \( p < .05 \)

As Table 1 shows, the mean perceptual scores for Group B were consistently and significantly higher than those of Group A as a whole and across all the 5 contrast-pairs. This suggests that a higher perceptual accuracy tends to be associated with Group B and a lower accuracy with Group A. Since the grouping was assigned in terms of the production accuracy of the tested phonemes (unsuccessful producers being assigned to Group A and successful producers to Group B), the implication is that perception performance relates positively to production performance. The strength of such a relationship, as indicated by the magnitude of the eta-squared in Table 2, is the strongest for /v-w/, followed by /θ-d/ and /r-w/, /z-s/ and /ð-f/.

\(^5\) Bryman & Cramer (1997) define that eta-squared is a tool for exploring the strength of the relationship between two variables. Specifically, it refers to the extent of variation in the dependent variable (i.e. perception score) being accounted for by the independent variable (i.e. group type).
Further analysis

Table 3 below shows the confusion matrices for the two groups of subjects across the 5 contrast-pairs. This was drawn up by working out the accuracy percentage across all contrast-pairs between the two groups, calculated as the total observed accurate identification in each contrast derived, divided by the total possible score, x 100%; the total possible score being 540 (18 marks for each contrast-pair x 30 subjects in each group).

Table 3: Confusion matrices derived from the accuracy percentage across all contrast-pairs between the two groups

<table>
<thead>
<tr>
<th>RESPONSE PATTERN</th>
<th>GROUP A</th>
<th>GROUP B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>v-w</td>
<td>ð-f</td>
</tr>
<tr>
<td>v-w</td>
<td>53</td>
<td>60</td>
</tr>
<tr>
<td>ð-f</td>
<td>89</td>
<td>78</td>
</tr>
</tbody>
</table>

This analysis reveals that there was perceptual confusion within the contrast-pairs of /v-w/, /ð-f/, /ð-d/, /z-s/ and /r-w/ by Group A as their perceptions were of approximately chance-level accuracy, ranging from the lowest of 53% accuracy for /v-w/ to the highest of 68% for /r-w/. Group B, on the other hand, demonstrated a relatively consistent above-chance level accuracy, ranging from the lowest of 78% accuracy for /ð-f/ to the highest of 94% accuracy for /r-w/.

An investigation into the subjects’ error pattern in the perception test reveals that all the perceptual errors for the target items /v, ð, z, t/ were the ones corresponding to their mispronounced versions /w, f, d, s, w/, that is, “vent” was perceived as “went”. This kind of misidentification between /v-w/, /ð-f/, /ð-d/, /z-s/ and /r-w/ for Group A, further suggests that each of the English contrast-pairs may exist as a single phonemic category instead of two different contrastive ones. In other words, the subjects may categorize the English /v, ð, z, t/ phonetically in the Cantonese classes of /w, f, d, s, w/ respectively, instead of building up new phonemic categories for them. Consequently, a chance-level perceptual response of Group A may result. This analysis further supports the notion that there is a positive relationship between perception and production.
**Discussion**

Two important findings arise from the present study. First, the subjects with better perception tend to have better production and vice versa. This finding supports the Motor Theory of speech perception and production (Liberman and Mattingly, 1985) which suggests that in speech learning, the acoustic consequence of a speech sound is transformed into an articulatory-based code which contributes to the speakers’ internal sound representation. Such internal representation is then used to devise a motor plan for one’s own production template. In accounting for such parallel performance, Bradlow et al. (1997) suggest that there might be a common mental representation that determines both speech perception and production.

The second finding relates to the level of breakdown in L2 speech learning. As evidenced by numerous cross-linguistic findings of L1 phonological acquisition, speech accuracy may not be related to corresponding perceptual difficulties but largely to incorrect realization of rules between input and output lexicons (e.g. Brettetal et al., 1987). However, the nature of L2 speech errors represents a different picture. Results of the present study provide strong evidence as does supportive evidence in the literature that the breakdown in L2 speech learning may well be at the level of internal perceptual representation. In particular, this finding supports Flege’s (1991, 1992) model of L2 speech learning that L2 learners tend to perceive L2 sounds categorically within the sound classes of their L1. Cantonese speakers of English, who have acquired the complete Cantonese sound system, may perceive speech sounds in the L2 according to Cantonese sound characteristics which would account for why subjects in the present study displayed difficulties in discriminating the L2 sounds /v, ð, z, r/ which are unfamiliar to them. However, due to these sounds being acoustically similar to the Cantonese /w, f, d, s, w/ respectively, subjects might have equated these with their corresponding Cantonese classes. Such phonological transfer may represent a universal phenomenon of L2 learning. However, the specific sound substitution (i.e. error pattern) will not be the same cross-linguistically due to the difference in typological phonological distance between the L1 and L2.

In the teaching of L2 pronunciation a common approach is for teachers to focus on the place and manner of articulation (Leather, 1983; Morley, 1991; Pennington, 1998). The present findings suggest that, in addition to these

---

6 According to distinctive feature specification (Chomsky & Halle, 1968), the values of /w, f, d, s/ in English and Cantonese are very similar.
aspects of pronunciation, more focus on sound perceptual training would be a worthwhile direction for future L2 phonology programs. Since perception and production appear to go hand in hand, the very first step in any of these programs should be to analyze the learners’ perception of problematic L2 sounds in relation to the phonology of their mother tongue. This would represent a more holistic approach to decision-making. To help learners assess and adapt their schemas for correct perceptual representation, real words in the form of differentiation or discrimination tasks, as opposed to same-different judgments, would be preferable.

As the learners’ L1 would also influence their L2 speech perception, comparative minimal pair exercises would be a second feasible approach in teaching. This would involve minimal pair perceptual training requiring learners to compare and discriminate the targets and their substitutes. This kind of training would go beyond the current minimal pair exercises which are usually used within the L2 sound system. As acknowledged by Bowen and Marks (1992) and Pennington (1998), this approach would encourage learners to break their L1 habits in both perception and production, and construct and reform actively their perceptual categories for L2 speech sounds. In other words, this would help L2 learners shift their phonemic analysis to phonetic analysis and enable them to incorporate the minute differences they have assumed into their representation constructions for L2 sounds.

Phonology has hitherto remained a much neglected area of teaching and research. It is time for it to be given the due consideration it deserves and for it keep pace with the ever expanding research in other areas of applied linguistics. Encouraged by recent L1 and L2 studies of the higher levels of speech processing, the present study has hopefully brought L2 phonology to the foreground as a starting point for further work in the area. Most importantly, this study aimed at providing a basis for more discussions into redefining the philosophy underlying current pedagogy and for developing greater professionalism towards L2 phonology teaching.

Note: This paper is based on the author’s dissertation for M.A. (Applied Linguistics), the University of Hong Kong, 1999.
References


<table>
<thead>
<tr>
<th>Phoneme</th>
<th>Targets &amp; phonetic transcription</th>
<th>Errors &amp; phonetic transcription</th>
</tr>
</thead>
<tbody>
<tr>
<td>/v/</td>
<td>vent [ventʰ]</td>
<td>went [wentʰ]</td>
</tr>
<tr>
<td></td>
<td>vest [vestʰ]</td>
<td>west [westʰ]</td>
</tr>
<tr>
<td></td>
<td>V [viː]</td>
<td>we [wiː]</td>
</tr>
<tr>
<td>/ð/</td>
<td>thin [θɪn]</td>
<td>fin [fɪn]</td>
</tr>
<tr>
<td></td>
<td>thirst [θɜːstʰ]</td>
<td>first [fɜːstʰ]</td>
</tr>
<tr>
<td></td>
<td>three [θriː]</td>
<td>free [friː]</td>
</tr>
<tr>
<td>/θ/</td>
<td>though [ðəθ]</td>
<td>dough [dəθ]</td>
</tr>
<tr>
<td></td>
<td>they [ðet]</td>
<td>day [det]</td>
</tr>
<tr>
<td></td>
<td>there [ðeər]</td>
<td>dare [dər]</td>
</tr>
<tr>
<td>/z/</td>
<td>zeal [ziːl]</td>
<td>seal [sɪːl]</td>
</tr>
<tr>
<td></td>
<td>zinc [zɪŋkʰ]</td>
<td>sink [sɪŋkʰ]</td>
</tr>
<tr>
<td></td>
<td>zip [ziːpʰ]</td>
<td>sip [sɪpʰ]</td>
</tr>
<tr>
<td>/r/</td>
<td>rare [reər]</td>
<td>where [weər]</td>
</tr>
<tr>
<td></td>
<td>right [raɪtʰ]</td>
<td>white [wɜːtʰ]</td>
</tr>
<tr>
<td></td>
<td>rail [reɪl]</td>
<td>whale [weɪl]</td>
</tr>
</tbody>
</table>