Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/-/d/ contrast

James Emil Flege and Chipin Wang

Department of Biocommunication, University of Alabama at Birmingham, University Station, Birmingham, AL 35294, U.S.A.

Received 6th October 1988, and in revised form 14th July 1989

A contrast between /t/ and /d/ exists in the initial but not the final position of Chinese words. This study examined identification of English word-final /t/ and /d/ tokens from which closure voicing and release burst cues had been removed. The performance of three Chinese groups was compared before, during, and after feedback training. The Cantonese subjects were expected to perform best because their L1 permits unreleased /p, t, k/ in word-final position, and the Mandarin subjects were expected to perform most poorly because their L1 permits no word-final obstruents. An intermediate level of performance was expected from the Shanghainese subjects, whose L1 permits /ʔ/ word-finally. As predicted, the Cantonese subjects were significantly more sensitive to the English /t/-/d/ contrast than the Mandarin subjects, with the Shanghainese subjects showing an intermediate level of performance. The subjects in all three groups showed a significant increase in sensitivity as a result of the training. The increase in correct identifications averaged 25%. The results were interpreted to mean that native-language phonotactic constraints influence how syllables are processed. More specifically, it is concluded that the Cantonese subjects focused greater attention on the end of the consonant-vowel-consonant stimuli than the Mandarin subjects, which better enabled them to use remaining acoustic cues to the /t/-/d/ contrast such as F1 offset frequency.

1. Introduction

At least some pronunciation errors in the speech of non-natives derive from underlying perceptual problems. The purpose of the present study was to provide additional information concerning how the perception of native and non-native speakers may differ. The hypothesis tested was that the number of obstruents permitted in the final position of words in the native language (henceforth L1) may influence the extent to which listeners allocate attention to the end of syllables, and that such learned syllable processing strategies transfer into the perception of stop consonants in a second language (L2). This led to the prediction that native speakers of Cantonese, Shanghainese and Mandarin would differ in ability to perceive the contrast between phonologically voiced and voiceless stops in the final position of English words.
1.1. L2 perceptual problems

Weinreich (1953) observed that sounds found in an L2 are typically "identified" with sounds found in the L1, causing non-natives to produce L2 sounds as if they were L1 sounds. Perceived distortions or substitutions may result if the acoustic differences between L1 and L2 sounds so identified are sufficiently great. Echoing Trubetzkoy's (1939) observation that the phonology of L1 acts like a "sieve" through which L2 sounds are processed, Borden, Gerber & Milsark (1983, p. 500) suggested that L2 learners are "constrained" to categorize L2 sounds according to the "phoneme contrasts" of their L1. This implies that difficulty will arise in the perceptual differentiation of two L2 sounds if they are not judged to be realizations of different categories in L1, and that the difficulty might be reversible because the locus of the problem is identified as occurring at a relatively late stage of processing (see, e.g., Werker and Logan, 1985).

L2 research has not yet established definitively which acoustic, articulatory and possibly phonological properties lead to the interlingual identification of acoustically different sounds in L1 and L2, nor what is the basic unit of interlingual identification. Several possibilities exist: the syllable or demisyllable, the abstract morphophoneme of generative phonology, the less abstract systematic phoneme of structural linguistics, the major allophone, positional allophones, context-sensitive allophones (either universal or language-specific), or even individual phones. It is likely that several of these units are important, depending on the task and level of processing. A recent finding by Flege (1989a) suggested that the basic unit of interlingual identification may be the positional allophone, that is, a class of phones occurring in a specific syllable position as realized in a specific language. Therefore, in the present article, the L1 and L2 phonic elements that are identified with one another will be referred to as "phones".

Wode (1980) and later Flege (1988a) hypothesized that interlingual identification need not occur between phones in L2 and phones in L1 if they differ sufficiently in articulatory-acoustic terms. This view has been supported by evidence from speech production and perception experiments. It appears that certain L2 learners, especially young ones, may establish phonetic categories for phones found in the L2 but not the L1 (e.g., Underbakke, Polka, Gottfried & Strange, 1988; Flege & Eefting, 1988). This might be expected to reduce perceptual difficulties if it reduces the likelihood that phonetically different L1 and L2 phones that are not contrastive in the L1 phonetic system will be judged, incorrectly, to be realizations of a single category.

1.2. The critical period hypothesis

Since the seminal work of Lenneberg (1967), a critical period has often been hypothesized to exist for both first and second language acquisition. The pronunciation of foreign languages has received special attention, since it seems to conform best to the critical period notion. (For reviews of this complex issue, see Oyama, 1979; Flege, 1987; Snow, 1987.) Studies using interval scaling have amply confirmed that the speech of many L2 learners is marked by a detectable foreign accent. Phonetic studies have demonstrated that non-native speakers differ from native speakers in the production and perception of specific sounds and phonetic contrasts in ways that are predictable from a comparison of how phonemes are implemented.
Chinese perception of word-final /t/-/d/ in English

phonetically in L1 and L2. For example, Flege & Eefting (1987) showed that Dutch speakers of English produced English /t/ with voice onset time (VOT) values that were too short for English, and that the extent to which the Dutch subjects undershot the English phonetic norm was correlated with their degree of perceived foreign accent.

Foreign accent studies have provided evidence that the age at which an L2 is learned will influence how well an L2 is pronounced (Asher & Garcia, 1969; Fathman, 1975; Tahta, Wood & Lowenthal, 1981; Oyama, 1982; Flege, 1988b). More narrowly focused studies have shown that age of learning affects segmental articulation. For example, Flege, (1989b; see also Flege & Eefting, 1988) showed that native speakers of Spanish who learned English by the age of 5–6 years, but not those who learned English as adults, matched native speakers’ VOT values for /t/.

Differences like those obtained in observational studies of naturalistic L2 learning may also be observed in laboratory studies. Cochrane (1977) had more success training Japanese children than Japanese adults to distinguish /r/ from /l/, a contrast found in English but not in Japanese.

Such age of learning effects raise the issue of whether phonetic perception remains malleable after the L1 phonetic system has been established. One possibility is that L2 speech perception problems exist because it is necessary for innate sensory-based discrimination abilities to be exploited at a relatively early age in order for them to be maintained beyond early childhood (Aslin & Pisoni, 1981; Werker & Tees, 1984; Werker, 1989). If so, the human speech learning ability might be bounded by a specific period of sensitivity to environmental stimulation in a manner analogous to the song-learning abilities of certain avian species (Marler, 1970). On the other hand, at least some of the perceptual difficulties seen in L2 learning may arise because L2 phones are processed using reversible attentional strategies which are optimal for L1 phones but non-optimal (or inappropriate) for L2 phones. This can be illustrated by considering the perception by native speakers of Japanese of English /r/ and /l/.

Japanese learners’ difficulty perceiving the English /r/-/l/ contrast is well documented (e.g., Mochizuki, 1981; Shimuzu & Dantsuji, 1983). In a classic study, Miyawaki, Strange, Verbrugge, Liberman, Jenkins & Fujimura (1975) showed that Japanese subjects could not identify reliably the /r/ and /l/ endpoints of a synthetic speech continuum in which third-formant frequency (F3) was varied. However, the Japanese subjects closely resembled native English speakers in discriminating the F3 frequency portion of the stimuli, which sounded like chirps when presented in isolation. The Japanese subjects’ speech perception problem seemed to derive from difficulty in processing relevant acoustic dimensions phonetically, not from purely auditory limitations (see, e.g., Werker & Logan, 1985; Burnham, 1986; Flege, 1988a; Best, McRoberts & Sithole, 1988; Werker, 1989).

If an irreversible sensory-based deficit in an L2 arises during or as a result of maturation, one would not expect much better performance for certain individuals than others. Nor would one expect difficulty with English /r/ and /l/ to be overcome simply as the result of phonetic input during naturalistic acquisition. However, the extent of the /r/-/l/ problem for Japanese learners varies widely between individuals (Underbakke et al., 1988). Japanese speakers with a great deal of conversational English experience, but not those without such experience, may resemble native speakers in perceiving the English /r/-/l/ contrast categorically (MacKain, Best & Strange, 1981).
1.3. Training studies

Another expectation one would have if L2 speech perception problems were the result of a sensory-based deficit (or arose because a critical period had been passed) is that problems perceiving the /r/-/l/ contrast should not be reversible as the result of training. However, laboratory, clinical, and classroom training studies have provided evidence of improvements in the perception of /r/-/l/ for subjects whose L1 does not possess such a contrast (Gillette, 1980; Borden et al., 1983; Strange & Dittmann, 1984; Logan, Pisoni & Lively, 1989).

These improvements were by no means dramatic. Laboratory training studies have also resulted in improvements for a number of other non-native contrasts. A study by Jamieson and Morosan (1986) yielded a small (11%) but significant increase in the rate of correct identifications of word-initial English /θ/-/ð/ by French subjects. Other studies have shown varying amounts of improvement for word-initial stop consonants (Lisker, 1970; Kalikow & Swets, 1972; Strange & Jenkins, 1978; Pisoni, Aslin, Perey & Hennessy, Pisoni & Carrell, 1982).

Most training studies to date have focused on what might be called unfamiliar non-native contrasts since the two phones being trained contrasted phonetically in the L2 whereas they did not contrast phonetically in any syllable position or phonetic context in the listeners' L1. Most previous research has also focused on contrasts in the initial position of syllables (henceforth, words). The present study differed from previous studies in two important ways: it focused on a phonetic contrast in word-final position; and the phonetic contrast being trained was a familiar non-native contrast in the sense that it existed in syllable-initial position in the subjects' L1.

1.4. Chinese–English differences

The present study examined the identification of word-final English /d/s and /t/s by Chinese subjects. The L1s of the subjects examined possess a contrast between phonologically voiced and voiceless stops in word-initial but not word-final position (Cheng, 1973; Howie, 1976; Li & Thompson, 1981; Shinn, 1985; Heyer, 1986). The contrast between phonologically voiced and voiceless stops in Chinese is implemented as a contrast between voiceless unaspirated stops with short-lag VOT values and voiceless aspirated stops with long-lag VOTs (Lisker & Abramson, 1964; Clumeck, Barton, Macken & Huntington, 1981).

As predicted by contrastive analysis (e.g., Lado, 1957), Chinese learners of English are often unable to produce a perceptually effective contrast between phonologically voiced and voiceless stops in the final position of English words. Transcriptional data from several studies have shown that Chinese speakers of English devoice /b, d, g/, delete both voiced and voiceless stops, and/or add schwa-like vowels following word-final stops (Tarone, 1980; Eckman, 1981; Anderson, 1983, 1987; Flege & Davidian, 1985; Heyer 1986; Weinberger, 1987; Flege, McCutcheon & Smith, 1987). Detailed analyses of the production of /p/ and /b/ in word-final position have shown that Chinese subjects produce a much smaller (albeit significant) duration difference between vowels preceding /p/ vs. /b/, are significantly less likely than native speakers to actively enlarge the oral cavity to sustain voicing in /b/, and do not produce /p/ with a greater force of labial constriction than /b/ (Flege et al., 1987; Flege, 1988c).
The presence of word-initial contrasts between /p, t, k/ and /b, d, g/ in the L1 should enable L2 learners to produce the same contrasts word-finally if speech production were based on freely commutable phonemes (see Flege & Port, 1981). The speech production studies just cited, however, show that a word-initial contrast does not guarantee the ability to produce a contrast in the word-final position. They suggest that speech production skills must to be learned on an allophone-by-allophone basis (see also Brière, 1966). Perhaps the production problems noted in the previous studies have derived from an attempt by Chinese learners of English to produce word-final stops as if they were stops in the word-initial position, that is, with little or no modification.

A recent study by Flege (1989a) led to a similar conclusion concerning speech perception. Native English subjects were able to identify the final stops in words like beat and bead at high rates in a forced-choice test, even when closure voicing and release burst cues had been removed by editing. This result, taken together with the results of previous studies, suggested that spectral and temporal properties of the preceding vowel and the formant transitions leading into the final consonant were sufficient to cue the /t/-/d/ distinction for native speakers (Halle, Hughes & Radley, 1957; Malécot, 1958; Wang, 1959; Winitz Scheib & Reeds, 1972; Ohde & Sharf, 1977; Raphael, 1981; Revoile, Picket, Holden & Talkin, 1982; Flege & Hillenbrand, 1987; Wolf, 1978; Hillenbrand, Ingrisano, Smith & Flege, 1984).

Chinese subjects did not differ from native speakers in identifying stops containing all available acoustic cues to the English /t/-/d/ contrast. However, their rate of correct identifications was far lower for stops from which the closure voicing and release bursts had been removed (64% vs. the near-perfect rate for native speakers). Since word-final /b, d, g/ are frequently devoiced in conversational English, and both voiced and voiceless stops are often produced without audible release bursts, the Chinese subjects' difficulty with the edited /t/s and /d/s might be indicative of difficulty perceiving word-final stop voicing contrasts in normal conversational speech.

This finding was interpreted to mean that the Chinese speakers of English attempted to use acoustic cues that distinguish /t/ from /d/ in the initial position of Chinese words to identify the word-final English stops. This strategy worked well for stops with release bursts but not for stops without release bursts.

1.5. Aims of the study

The purpose of this study was to replicate and extend the Flege (1989a) study. The first aim was to determine if Chinese subjects' sensitivity to the English /t/-/d/ contrast could be improved by a small amount of feedback training. To this end, a subset of the edited stimuli that caused difficulty for Chinese subjects in the Flege (1989a) study were again used. The stimuli were tokens of beat, bead, bet and bed from which closure voicing and release burst cues had been removed. The same training technique was used (presenting multiple natural tokens for identification with feedback; see Jamieson & Morosan, 1986).1

1 Identification with feedback was used rather than discrimination training because the aim of training was to enable the Chinese to identify cues to the /t/ and /d/ categories rather than the boundary between those categories (Repp, 1984).
The second aim was to test a hypothesis concerning the influence of L1 syllable structure constraints on perception of the word-final /t/-/d/ contrast in English. Psycholinguistic research has shown that listeners tend to focus more attention on segments occurring early than later in a processing unit (e.g., Marslen-Wilson & Walsh, 1978; but cf. Slowiaczek, Nusbaum & Pisoni, 1987). The extent to which this occurs, and the extent to which listeners pay attention to the end of syllables, may vary according to the syllable structure constraints of the L1.

Languages differ according to the number of segments in their phonetic inventory and according to restrictions on where segments may occur. This leads to large differences in syllables inventory size. For example, Maddieson (1984) reported that Hawaiian has 162 syllable types whereas Thai has an inventory of more than 2000 syllable types. It seems reasonable to think that differences in the number and kind of syllables permitted in languages will lead to measurable differences in how native speakers of those language will process syllables in real time.

The body of L2 research which now exists (see Flege, 1988a) strongly suggests that if syllable processing strategies are learned during L1 acquisition then they will be used in perceiving phones in an L2. Results obtained by Walley (1987, 1988) support the assumption that syllable processing strategies may evolve during the course of L1 acquisition. Results obtained by Cutler, Mehler, Norris & Segui (1986) are consistent with the view that processing strategies developed for the segmentation of syllables in the L1 may transfer to the processing of syllables in a foreign language, even when those strategies are not optimal.

The potential effect of differing syllable processing strategies on the identification of L2 phonetic segments has not been examined in L2 research. Flege (1989a) noted a trend for native speakers of Shanghainese to identify tokens of English /t/ and /d/ (from which voicing and release burst cues were removed) better than native speakers of Mandarin, and for the Shanghainese subjects to benefit more from feedback training than the Mandarin subjects. Flege hypothesized that the number of obstruents in word-final position in the L1 determines how much attention listeners will allocate to the rapid spectral changes which accompany the constriction of final consonants. If so, the Shanghainese subjects may have allocated more attention to the end of syllables than the Mandarin subjects because no obstruents are permitted in the final position of words in Mandarin whereas a glottal stop is permitted at the end of words in Shanghainese. The difference would have been motivated by the need of the Shanghainese speakers to differentiate syllables with relatively fast vs. slow decreases in energy (that is, syllables ending in a glottal stop vs. open syllables or syllables ending in a resonant consonant).

Results obtained by Walley (1987, 1988) are consistent with the view that syllable processing strategies may be learned. Walley hypothesized that the priority given to word-initial over word-final consonants in word recognition may show a developmental course. Children may listen to longer portions of a word before recognition occurs to ensure correct recognition of familiar words and to permit the acquisition of new ones because their lexicon is small. In support of this, Walley (1987) found that in contexts where meaning was not semantically constrained, 4- and 5-year-olds did not detect more mispronunciations in initial than final position. In semantically constraining contexts they did so, thereby resembling adults. Walley (1988) found that adults but not children gave lower ratings to word-initial than word-final consonants that had been degraded by noise.

Both Mandarin and Shanghainese have word-final /m/ and /ŋ/ (Cheng, 1973; Li & Thompson, 1981). The presence of these final consonants may not cause listeners to focus attention on the rapid spectral changes which accompany constriction since nasal consonants can be identified on the basis of the nasal murmur during constriction (Repp & Svastukula, 1988).
If the syllable processing hypothesis is correct, then native speakers of Shanghai­
nese should be somewhat better able to identify edited tokens of word-final English
/t/ and /d/ than native speakers of Mandarin. Since the difference between
Mandarin and Shanghai­inese in terms of the number of obstruents permitted in
word-final position is small, Mandarin subjects were also compared to native
speakers of Cantonese, a language which permits /m, n, g, p, t, k/ in word-final
position (e.g., Heyer, 1986). By hypothesis, the Cantonese subjects should
outperform the Mandarin subjects to an even greater extent than the Shanghai­inese
subjects. They might also benefit to a greater extent from the feedback training by
virtue of having learned to focus attention on the end of syllables, where acoustic
cues to the edited /t/-/d/ are to be found.

2. Methods

The stimuli were derived from an audio recording of an adult native English
speaker’s reading of beat, bead, bet and bed in the carrier phrase Now I will say——. As described previously (Flege, 1989a), the first four tokens of each word
were low-pass filtered at 4 kHz and digitized at 10 kHz with 12-bit resolution, then
amplitude normalized. The words selected for the speech training experiment had
all been spoken clearly with prevoicing in the initial /b/s. The final stop in each
word had an audible release burst, and the /d/s all had closure voicing. The release
bursts and closure voicing were removed by editing the 16 waveforms (two
vowels x two final stops x four replicate tokens of each word). It is important to
note that cues to /t/-/d/ contrast remained after the editing. For example, vowels
were longer in the bead than beat tokens (335 ms vs. 198 ms) and longer in the bed
than bet tokens (327 ms vs. 201 ms).

A total of 27 Chinese subjects in three L1 groups completed a language
background questionnaire and passed a pure tone hearing screening test before
participating. Characteristics of the nine paid subjects in each group are shown in
Table I. Individuals who began learning English before the age of 12 years or had
participated in previous related research were excluded from the study. Native
speakers of Mandarin were excluded if they spoke Shanghai­ese or Cantonese;
Shanghai­ese subjects were excluded if they spoke Cantonese.

The stimuli were presented in three approximately 6 min blocks, each containing
10 separate randomizations. The first block was designated the “pre-training” block
because it occurred before feedback training. In the second block, designated “FB”,
the subjects were informed of the correct answer immediately after pushing a “t” or
“d” button. No feedback was provided in the third block, which was designated the
“post-training” block. The stimuli were presented binaurally over headphones at
74 dB(A) peak syllable intensity. The experiment was self-paced (a stimulus was
presented 1.0 s after a response was received for the preceding trial). The subjects
were told to guess if uncertain.

4 One might argue that because formant transitions in syllables ending in nasal consonants provide
place of articulation cues as potent as the nasal murmur (Repp & Svastukula, 1988), Mandarin subjects
may have learned to focus attention on the end of syllables with /ŋ/. There is no difference in the number
of syllable-final consonants in Mandarin and Shanghai­ese if one counts the clitic /r/ in Mandarin words.
The glottal stop of Shanghai­ese occurs in short, checked syllables which have a distinct tone from long
syllables, so Shanghai­ese speakers may not need to focus on the end of syllables to distinguish syllables
ending with and without a glottal stop.
J. E. Flege and C. Wang

TABLE I. Characteristics of the subjects in three Chinese groups. "INS" is the number of years of formal English-language instruction; "AOA" is the age of arrival in the U.S., in years; "YOR" is the number of years of residence in the U.S.; "PER" is the subjects self estimates of their percentage of daily use of English; and "Origin" is the place where they grew up; standard deviations are in parentheses.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Age</th>
<th>Sex</th>
<th>INS</th>
<th>AOA</th>
<th>YOR</th>
<th>PER</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mandarin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>25</td>
<td>F</td>
<td>6</td>
<td>17</td>
<td>1.0</td>
<td>60</td>
<td>Beijing</td>
</tr>
<tr>
<td>2</td>
<td>28</td>
<td>M</td>
<td>7</td>
<td>27</td>
<td>3.0</td>
<td>40</td>
<td>Taipei</td>
</tr>
<tr>
<td>3</td>
<td>29</td>
<td>F</td>
<td>7</td>
<td>28</td>
<td>4.9</td>
<td>10</td>
<td>Taipei</td>
</tr>
<tr>
<td>4</td>
<td>27</td>
<td>M</td>
<td>7</td>
<td>26</td>
<td>1.9</td>
<td>10</td>
<td>Kaoshing</td>
</tr>
<tr>
<td>5</td>
<td>27</td>
<td>F</td>
<td>7</td>
<td>26</td>
<td>1.3</td>
<td>20</td>
<td>Taichug</td>
</tr>
<tr>
<td>6</td>
<td>27</td>
<td>M</td>
<td>7</td>
<td>26</td>
<td>2.0</td>
<td>20</td>
<td>MiaoLi</td>
</tr>
<tr>
<td>7</td>
<td>27</td>
<td>M</td>
<td>3</td>
<td>25</td>
<td>1.0</td>
<td>50</td>
<td>Anshan</td>
</tr>
<tr>
<td>8</td>
<td>27</td>
<td>M</td>
<td>15</td>
<td>25</td>
<td>1.6</td>
<td>60</td>
<td>Taipei</td>
</tr>
<tr>
<td>9</td>
<td>27</td>
<td>M</td>
<td>6</td>
<td>26</td>
<td>1.4</td>
<td>50</td>
<td>Beijing</td>
</tr>
<tr>
<td>M</td>
<td>27</td>
<td>6M/3F</td>
<td>7</td>
<td>25</td>
<td>2.0</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td></td>
<td>27.2</td>
<td>(1.1)</td>
<td>7.2</td>
<td>(3.3)</td>
<td>25.1</td>
<td>(3.2)</td>
<td>1.3</td>
</tr>
<tr>
<td>Shanghainese</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>23</td>
<td>F</td>
<td>5</td>
<td>22</td>
<td>8.2</td>
<td>20</td>
<td>Shanghai</td>
</tr>
<tr>
<td>2</td>
<td>27</td>
<td>M</td>
<td>2</td>
<td>24</td>
<td>1.0</td>
<td>30</td>
<td>Shanghai</td>
</tr>
<tr>
<td>3</td>
<td>39</td>
<td>M</td>
<td>8</td>
<td>34</td>
<td>0.8</td>
<td>50</td>
<td>Shanghai</td>
</tr>
<tr>
<td>4</td>
<td>33</td>
<td>M</td>
<td>3</td>
<td>31</td>
<td>1.0</td>
<td>90</td>
<td>Shanghai</td>
</tr>
<tr>
<td>5</td>
<td>26</td>
<td>F</td>
<td>3</td>
<td>25</td>
<td>1.0</td>
<td>30</td>
<td>Shanghai</td>
</tr>
<tr>
<td>6</td>
<td>27</td>
<td>F</td>
<td>15</td>
<td>25</td>
<td>0.7</td>
<td>20</td>
<td>Shanghai</td>
</tr>
<tr>
<td>7</td>
<td>28</td>
<td>F</td>
<td>3</td>
<td>27</td>
<td>2.4</td>
<td>30</td>
<td>Shanghai</td>
</tr>
<tr>
<td>8</td>
<td>30</td>
<td>M</td>
<td>12</td>
<td>28</td>
<td>2.0</td>
<td>40</td>
<td>Shanghai</td>
</tr>
<tr>
<td>9</td>
<td>30</td>
<td>M</td>
<td>5</td>
<td>29</td>
<td>2.0</td>
<td>30</td>
<td>Shanghai</td>
</tr>
<tr>
<td>M</td>
<td>29</td>
<td>5M/4F</td>
<td>6</td>
<td>27</td>
<td>2.1</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29.2</td>
<td>(4.6)</td>
<td>6.2</td>
<td>(4.5)</td>
<td>27.2</td>
<td>(3.7)</td>
<td>2.4</td>
</tr>
<tr>
<td>Cantonese</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>24</td>
<td>F</td>
<td>13</td>
<td>16</td>
<td>6.8</td>
<td>60</td>
<td>Malaysia</td>
</tr>
<tr>
<td>2</td>
<td>27</td>
<td>F</td>
<td>13</td>
<td>26</td>
<td>1.6</td>
<td>30</td>
<td>Hong Kong</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>M</td>
<td>10</td>
<td>24</td>
<td>6.9</td>
<td>100</td>
<td>Hong Kong</td>
</tr>
<tr>
<td>4</td>
<td>22</td>
<td>M</td>
<td>14</td>
<td>20</td>
<td>2.6</td>
<td>30</td>
<td>Malaysia</td>
</tr>
<tr>
<td>5</td>
<td>29</td>
<td>M</td>
<td>10</td>
<td>23</td>
<td>6.6</td>
<td>60</td>
<td>Malaysia</td>
</tr>
<tr>
<td>6</td>
<td>25</td>
<td>F</td>
<td>9</td>
<td>24</td>
<td>1.4</td>
<td>80</td>
<td>Canton</td>
</tr>
<tr>
<td>7</td>
<td>22</td>
<td>M</td>
<td>10</td>
<td>19</td>
<td>1.9</td>
<td>80</td>
<td>Hu Bei</td>
</tr>
<tr>
<td>8</td>
<td>19</td>
<td>F</td>
<td>11</td>
<td>18</td>
<td>1.4</td>
<td>70</td>
<td>Malaysia</td>
</tr>
<tr>
<td>9</td>
<td>30</td>
<td>F</td>
<td>5</td>
<td>27</td>
<td>3.6</td>
<td>40</td>
<td>Canton</td>
</tr>
<tr>
<td>M</td>
<td>25</td>
<td>4M/5F</td>
<td>11</td>
<td>22</td>
<td>3.6</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25.3</td>
<td>(2.7)</td>
<td>10.6</td>
<td>(3.7)</td>
<td>22.0</td>
<td>(2.4)</td>
<td>3.6</td>
</tr>
</tbody>
</table>

The rate at which each subject correctly identified the final stop in bead, beat, bet and bed was calculated. Sensitivity to the English /t/-/d/ contrast was evaluated by calculating A', an unbiased measure of sensitivity (Grier, 1971) which ranges from 0.0 to 1.0 (perfect sensitivity). This nonparametric index of sensitivity was used because of the small number of responses by each subject for the four words, and
Chinese perception of word-final /t/-/d/ in English

because of the requirement of normal distribution for \(d'\) analyses. The A' scores were calculated separately for the stops in *beat–bead* and in *bit–bid*. Hits were defined as instances in which a /\(t\)/ token was labelled "\(t\)" and false alarms were defined as those instances in which /\(d\)/ tokens were incorrectly labelled "\(t\)". The A' scores were submitted to a Group (Mandarin, Shanghainese, Cantonese) × Preceding Vowel (/\(i\)/, /\(e\)/) × Block (pre-training, FB, post-training) ANOVA with repeated measures on the last two factors.

3. Results

3.1. Identification of final stops

The mean percentage of correct identifications for subjects in the three L1 groups is presented in Table II for the pre-training, FB, and post-training blocks and in Fig. 1(a). As predicted, the correct identification rate was higher for the native speakers of Cantonese than for the Shanghainese subjects who, in turn, had a higher correct identification rate than the native Mandarin subjects (88%, 79%, 69%). Performance improved as the result of training. Averaged across the three groups, the rates were higher in the post-training than in the FB block, and higher there than in the pre-training block (89%, 83%, 64%). The difference between the pre- and post-training blocks was somewhat greater for the Cantonese and Shanghainese subjects (26% and 28%) than for the Mandarin subjects (19%).

The effect of training also varied as a function of word. The difference between the pre- and post-training blocks was 34% for *bead* and 32% for *bet*, but only 18% for *beat* and 14% for *bed*. The percent correct identification rate was about the same for /\(t\)/ and /\(d\)/ (78%, 80%), but stops were identified correctly somewhat more often following /\(e\)/ than /\(i\)/ (81% vs. 76%).

A' scores are shown in Fig. 1 as a function of group and block. (The A' scores in Fig. 1 have been averaged across the two vowel contexts because the Vowel factor did not interact with either Group or Block). Sensitivity to the English /\(t\)/-/\(d\)/ contrast differed across the three subject groups, leading to a significant Group main effect \([F(2, 24) = 4.51, p < 0.05]\. The A' scores were higher for the Cantonese

<table>
<thead>
<tr>
<th>Subject group</th>
<th>Mandarin</th>
<th>Shanghainese</th>
<th>Cantonese</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Word</td>
<td>Pre</td>
<td>FB</td>
</tr>
<tr>
<td>Beat</td>
<td>61</td>
<td>68</td>
<td>75</td>
</tr>
<tr>
<td>Bead</td>
<td>52</td>
<td>68</td>
<td>73</td>
</tr>
<tr>
<td>Bet</td>
<td>52</td>
<td>72</td>
<td>81</td>
</tr>
<tr>
<td>Bed</td>
<td>72</td>
<td>76</td>
<td>84</td>
</tr>
<tr>
<td>M</td>
<td>59</td>
<td>71</td>
<td>78</td>
</tr>
</tbody>
</table>

TABLE II. The mean percentage of correct identifications of word-final stops in English words by three Chinese groups before, during, and after training ("pre", "FB", "post"); standard deviations are in parentheses.
The mean percentage of correct identifications of edited word-final tokens of English /t/ and /d/ by native speakers of Mandarin, Shanghainese, and Cantonese before (□), during (■), and after (□) feedback training. Mean perceptual sensitivity to the English /t/-/d/ contrast by subjects in the three Chinese groups as measured by A' scores based on the proportion of /t/ responses given to final stops in two minimal pairs (beat-bead, and bet-bed); the brackets enclose one standard error.

Subjects than for the Shanghainese subjects who, in turn, had higher A' scores the Mandarin subjects (0.898, 0.840, 0.726). Post-hoc tests (Newman–Keuls, alpha = 0.05) revealed that the Cantonese subjects' A' scores were significantly higher than the Mandarin subjects'. No other between-group difference reached significance.

Sensitivity to the /t/-/d/ contrast increased systematically across the three blocks in a similar fashion for all three groups, leading to a highly significant main effect of Block \(F(2, 48) = 15.9, p < 0.05\). The post-training A' scores were higher than those in the FB block which, in turn, were higher than those in the pre-training block (0.912, 0.878, 0.676). Post-hoc tests revealed that the A' scores were significantly higher in the post-training and FB blocks than in the pre-training block \(p < 0.05\). The Group × Block interaction was non-significant \(F(4, 48) = 0.32, p > 0.10\).
The last finding of interest was that subjects performed better on one of the minimal pairs examined than on the other. The main effect of Preceding Vowel was significant \[ F(1, 24) = 7.99, p < 0.05 \] because the \( A' \) scores were higher for \textit{bet–bed} than for \textit{beat–bead} (0.853 vs. 0.789).

The differences between the three groups supported the hypothesis that the larger the number of final obstruents permitted in the L1, the better non-natives' identification of a word-final voicing contrast will be (if such a contrast does not exist in the L1). An alternative explanation is that the between-group differences derived from differences in L2 experience. This explanation was explored in the section to follow.

### 3.2. Factors affecting between-group differences

The Cantonese subjects performed better than subjects in the other two groups. Questionnaire data (Table I) indicated that they had studied English longer at school and had lived longer in the U.S. than either the Mandarin or the Shanghainese subjects. Moreover, the Cantonese subjects reported speaking English more often on a daily basis, and had arrived in the U.S. at a somewhat earlier age than the Mandarin or Shanghainese subjects. All of these factors point in the same direction: the Cantonese subjects had more, and perhaps better, L2 phonetic input than the other subjects. To determine whether significant differences existed between the three groups, the questionnaire data were submitted to one-way ANOVAs. Significant effects of Group were followed up with Newman-Keuls \textit{post-hoc} tests (alpha = 0.05).

The results support the belief that the difference between the Cantonese and Mandarin subjects was due to L1 syllable structure differences rather than to differences in L2 experience. There was not a significant difference between groups in chronological age or in the length of residence in the U.S. The tests did yield a number of significant between-group differences, however. There was a significant difference in the number of years of formal English-language instruction \[ F(2, 24) = 3.62, p > 0.05 \]. The Cantonese subjects had received significantly more instruction than the Shanghainese subjects. There was a significant difference in age of arrival in the U.S. \[ F(2, 24) = 4.92, p > 0.05 \]. The Cantonese subjects had arrived earlier in the U.S. than the Shanghainese subjects. The effect of Group on self-estimated percent daily use of English was also significant \[ F(2, 24) = 3.65; p < 0.05 \], but the \textit{post-hoc} test failed to reveal any significant pairwise differences. In no instance did the \textit{post-hoc} tests reveal a significant difference in L2 experience between the two groups which differed significantly in sensitivity to the /t/-/d/ contrast, viz. the Cantonese and Mandarin subjects.

To further explore the possibility that prior L2 experience may have led to the observed between group difference, questionnaire variables were submitted to a series of forward step-wise multiple regression analyses. The five predictor variables examined were: native language, years of formal English instruction, age of arrival in the U.S., length of residence in the U.S., and the percentage of daily use of English.\(^5\) The criterion variables were the \( A' \) scores obtained for the 27 Chinese

\(^5\) Sex was excluded because the author's prior experience has shown that gender has little influence on L2 production or perception. Chronological age was not examined because it was redundantly specified in terms of the age of arrival and length of residence variables.
subjects. Six regression analyses were performed, one each for beat–bead and bet–bed before, during, and after training.

The multiple regression models were unable to account for a significant amount of variance in A’ scores for beat–bead or bet–bed in the pre-training block (7% and 8% respectively) \([F(1, 25) = 2.85; F(2, 24) = 2.10, p > 1.0]\). They were, however, able to account for a significant amount of variance for both beat–bead and bet–bed in the FB block (56% and 42% respectively) \([F(4, 22) = 9.28; F(2, 24) = 10.5, p < 0.05]\). Smaller, but still significant amounts of variance were accounted for in the post-training blocks (33% and 27%, respectively) \([F(2, 24) = 7.43; F(2, 24) = 5.75, p < 0.05]\). As expected, the effect of Group was significant in the analyses examining the FB and post-training blocks, both for beat–bead and bet–bed \((p < 0.05)\). The most important finding was that three factors which related directly to amount of L2 experience (years of instruction in English, length of residence in the U.S., percentage of daily use of English) were not identified as significant predictors of the A’ scores. A fourth experiential factor, age of arrival in the U.S., was a significant predictor in the FB and post-training blocks for beat–bead, but not for bet–bed \((p < 0.05)\). Recall that the Cantonese subjects arrived at a significantly earlier average age in the U.S. than the Shanghainese subjects but not the Mandarin subjects. Thus the significant difference in A’ scores between the Cantonese and Mandarin subjects cannot be explained simply as the result of the confound of age of arrival and L1 background.

4. Discussion

Flege (1989a) found that Chinese speakers of English closely resembled native speakers of English in identifying unedited tokens of /t/ and /d/ in the final position of English words, whereas they performed much more poorly than English native speakers when attempting to identify stops from which voicing and release burst cues had been removed. The native languages (L1s) of the Chinese subjects tested all possess a contrast between /t/ and /d/ in word-initial but not in word-final position, which suggested that the Chinese subjects’ relatively poor identification of burstless /t/s and /d/s derived from their attempt to use word-initial cues for word-final stops.

The present study replicated the Flege (1989a) study in showing that burstless /t/s and /d/s were difficult for Chinese subjects to identify correctly, and in showing that their sensitivity to the /t/-/d/ contrast could be increased significantly by a small amount of feedback training. The native speakers of Mandarin, Shanghainese, and Cantonese in the present study were all able to identify edited /t/s and /d/s better after than before training. The rate of correct identifications increased by an average of 25%, which is larger than the approximately 10% increase obtained by Flege (1989a) using the same stimuli and training technique. The Mandarin subjects in the present study identified stops in the pre-training block at a slightly lower rate than Mandarin subjects examined in the earlier study (59% vs. 64%) but at a higher rate after training (78% vs. 69%).

It is uncertain why a larger training effect was obtained here than in the Flege (1989a) study. One possibility is that more subjects in the present study than in the earlier study had a special aptitude for perceiving non-native phonetic contrasts (see Underbakke et al., 1988). The training effect obtained here was also larger than
those obtained in studies which have applied similar training techniques to other phonetic contrasts (e.g., Jamieson & Morosan, 1986; Logan et al. 1989). Perhaps the effect of training was greater here than in other studies because the /t/-/d/ contrast was a familiar non-native contrast to the subjects in the sense that it existed in word-initial position in their L1s. Another possibility is that voicing contrasts are relatively easy to learn—and train—because they are psychoacoustically “robust” (see, e.g., Burnham, 1986).

It appears that, as a result of the training, the Chinese subjects became better able to make use of acoustic cues that remained in the stimuli after the closure voicing and release burst cues were removed, such as preceding vowel duration and quality. One finding suggested that they also may have begun to use F1 offset frequency as a cue to the word-final stop voicing distinction. The Chinese subjects showed significantly greater sensitivity to stops following /e/ than /i/. A number of studies have shown that F1 frequency is lower at the end of formant transitions leading into voiced than voiceless English stops in word-final position, and that native speakers of English may use this acoustic difference as an acoustic cue to the voiced-voiceless distinction (Wolf, 1978; Revoile et al., 1982; Hillenbrand et al., 1984; van Summers, 1987, 1988). Hillenbrand et al. (1984) provided acoustic evidence that the F1 difference may be greater following mid than high vowels owing to differences in F1 frequencies in the preceding vowels. For example, the F1 offset frequency differences between words ending in /g/ and /k/ were greater following /e/ than /i/ (181 Hz vs. 15 Hz).6

It is uncertain exactly how the Chinese subjects’ perception changed as the result of the feedback training administered in the present study. One possibility is that they began to allocate more attention to the end of the CVC stimuli. Psycholinguistic research has shown that consonants at the beginning of syllables are processed differently, and may receive more attention, than consonants at the end of syllables (e.g., Marslen-Wilson & Walsh, 1978; Slowiaczek et al., 1987; Walley, 1987, 1988). Before training, the Chinese subjects may have focused less attention on the end of the CVC stimuli than native speakers of English, who must learn to pay more attention to the end of syllables because many words in their L1 are distinguished crucially by rapid spectral changes which accompany constriction for post-vocalic consonants. Chinese subjects, on the other hand, may allocate relatively little attention to syllable endings because relatively few Chinese words are so distinguished.

A second important finding of the study was that the native speakers of Cantonese were significantly more sensitive to the English word-final /t/-/d/ contrast than the native speakers of Mandarin. They also showed a non-significant tendency to outperform native speakers of Shanghainese. Although subjects in the present study were not matched according to L2 experience, an examination of questionnaire data suggested that the differences between the Cantonese and Mandarin subjects were not due to differences in L2 experience. There were no significant differences between them in years of English-language instruction, length of residence in the U.S., or percentage daily use of English. Of the variables examined in multiple

---

6 van Summers (1987) found that F1 frequencies were also higher in the steady-state portion of vowels preceding voiceless than voiced stops. In a later perceptual study using synthetic stimuli, van Summers (1988) showed that F1 steady-state frequency had an effect on voicing judgments that was independent of the effect of F1 frequency offset differences.
regression analyses, only age of arrival in the U.S. proved to be a significant predictor of variance in A' scores. The Cantonese and Mandarin subjects did not arrive at significantly different ages in the U.S.

A more likely explanation for the difference between the Cantonese and Mandarin subjects (and the non-significant difference between the Cantonese and Shanghainese subjects) was the hypothesized syllable processing differences mentioned above. Cantonese permits unreleased /p, t, k/ in word-final position and Shanghainese has a glottal stop. Mandarin, on the other hand, permits no obstruents to occur in word-final position. The results support the hypothesis that the greater the number of word-final obstruents in the L1, the easier it will be for non-natives to perceive a voicing contrast in the final position of L2 words (if no such contrast exists in their L1). By hypothesis, the Cantonese subjects outperformed the Mandarin subjects because they attended more closely to the spectral changes that occurred near the constriction of the final /t/s and /d/s. Although there is no voiced-voiceless contrast in the final position of words in Cantonese, Cantonese speakers may need to focus attention on the end of syllables in order to distinguish the unreleased tokens of /p, t, k/ that occur at the end of Cantonese words.

The present research leaves unanswered a number of important questions concerning the perception of non-native phonetic contrasts. It is uncertain whether the improved identification of word-final tokens of /t/ and /d/ that can be obtained rather quickly through feedback training will persist, and if it will improve subjects' recognition of English words during conversation. One wonders if the feedback training would have been less successful had the subjects not been familiar with a /t/-/d/ contrast in the word-initial position of their L1. It would be interesting to compare the rate and the extent to which Chinese subjects and speakers of a language like Finnish (which has only a marginal contrast between voiced and voiceless stops in any position) would learn the /t/-/d/ contrast.

Although there was no evidence that the between-group differences in sensitivity to the /t/-/d/ obtained in the present study could be accounted for by differences in prior L2 experience, it is nevertheless possible that some other, more fine-grained, measures of English-language experience might have led to a different conclusion. Six of the Cantonese subjects came from places (Malaysia, Hong Kong) where English is a major cultural force, so they may have received a great deal of passive exposure to English prior to their arrival in the U.S. Since it is difficult to gauge amount and quality of L2 experience, it would be useful to replicate the results obtained here with Chinese subjects who had no previous English-language experience. This would make it possible to answer another important question: Do phonotactic constraints in the L1 influence the effectiveness of feedback training?

Studies of language development have shown that children's prior knowledge may affect to some extent the effectiveness of training (Johnston, 1988). The same might be true for speech learning. The present study did not yield a significant Group × Block interaction. However, the Cantonese subjects may nevertheless have benefited more from training than the Mandarin subjects as the result of paying greater attention to the ends of the stimuli. The interaction may have been obscured by the ceiling effect which arose because the Cantonese subjects' sensitivity to the /t/-/d/ contrast was nearly perfect after the training.

In summary, the results obtained here suggest that L1 phonotactic constraints may affect how well listeners learn a familiar non-native contrast. Subjects whose L1
Chinese perception of word-final /t/-/d/ in English

(Cantonese) has word-final obstruents perceived the (burstless) word-final English /t/-/d/ contrast better than subjects whose L1 (Mandarin) permits no obstruents in word-final position. The subjects in all three Chinese groups examined benefited from training, which suggests that strategies of attentional allocation learned during L1 acquisition can be modified by appropriate sensory stimulation. Additional research is needed to determine whether the training effects obtained in short-term laboratory training experiments persist and generalize to the recognition of words in conversational speech.

The research was supported by NIH grant NS20963. The authors thank Elizabeth James for help preparing the stimuli, Laurie Skelton for help with data analysis, and O-S. Bohn and two anonymous reviewers for comments on a previous draft of the article.

References


Chinese perception of word-final /t/-/d/ in English


