

Language Experience in Second Language Speech Learning

In honor of James Emil Flege

Edited by

Ocke-Schwen Bohn

Aarhus University

Murray J. Munro

Simon Fraser University

© 2007 – John Benjamins B.V.

John Benjamins Publishing Company

Amsterdam/Philadelphia

Temporal remnants from Mandarin in nonnative English speech¹

Yue Wang and Dawn Behne

Second language (L2) production can take on the form of an interlanguage, a relatively stable system bearing the nature of both the native language (L1) and L2. Within such a system, acoustic-phonetic components of a syllable are known to bear interlanguage characteristics, but how do these interlanguage components interact within the syllable? The present study investigates temporal patterns of L1 and L2 in interaction within a syllable. Audio recordings were made of English stop-vowel syllables produced by native speakers of Mandarin who were fluent in English (ChE). Native English productions (AmE) of these syllables, and native productions of comparable Mandarin (ChM) stop-vowel syllables were also recorded. Temporal measures included stop closure duration, voice-onset time (VOT), vowel duration, and syllable duration. Results show that the internal timing of syllable components in ChE productions often deviates from ChE in the direction of AmE, with the closure duration, VOT, and vowel duration being intermediate to ChM and AmE. These temporal deviations of individual components were also compensated by temporal adjustments of other components in the syllable, maintaining a balanced distribution between the consonant and vowel. These findings are discussed in the context of previous research on interlanguage behavior and the gradual process of acquiring a target language.

Introduction

The pronunciation of a second language (L2) is generally believed to be influenced by the phonetic system of one's native language (L1) (Best, 1995; Flege, 1995a). Despite this interference, L2 learners often do not directly substitute a sound from their L1 phonemic inventory for an L2 sound. Instead, L2 productions may take on a systematic form which is intermediate to the L1 and target L2. Such an interlanguage (Selinker, 1972; see also Corder, 1967, and Nemser, 1971) is characterized as a relatively stable system and bears the nature of both the native and the target language (e.g., Flege, 1980; Flege & Port, 1981, Flege & Hillenbrand, 1984; Flege & Eefting, 1987a). Flege

(1980) has hypothesized that if L2 pronunciation is the output of an interlanguage, the L2 sounds produced by learners would be phonetically intermediate to similar sounds produced in the L1 and L2 by native speakers of those languages, and that L2 sounds (including mispronunciations) would be produced consistently in the same phonetic context by language learners. His research (e.g., Flege, 1980; Flege & Port, 1981) and that of others (e.g., Caramazza, Yeni-Komshian, Zurif, & Carbone, 1973; Gass, 1984; Fourakis & Iverson, 1985) has investigated the phonetic categories of interlanguage systems. Here, we extend this topic to syllable-internal timing and address how acoustic-phonetic components of an interlanguage interact within the syllable.

Voice-onset time (VOT) has been a highly effective means of testing the interlanguage hypothesis. The stop categories of many languages divide the VOT continuum into three general VOT ranges, corresponding to the three voicing features of {voiced}, {voiceless unaspirated}, and {voiceless aspirated} (e.g., Lisker & Abramson, 1964; Keating, 1984). VOT has thus been used to examine cross-language differences of stops and the interactions of L1 and L2 stops, with different languages having different oppositions among stop consonants along the VOT continuum (Ladefoged, 1971).

Research examining L2 VOT productions support the interlanguage hypothesis. For example, VOT produced by L2 learners is often found to be intermediate in value to that in their L1 and L2. Examining the VOT production of the English stop consonants by Canadian French-English bilinguals, Caramazza et al. (1973) found that bilingual speakers showed a shift in their production of the English VOT toward the target English distribution range, but that they did not totally align themselves with the native English norms. As a result, their VOT values for voiceless stops were intermediate to the French and English norms. This provides support for the notion of an interlanguage system, characterized as a relatively stable system which bears the nature of both one's L1 and L2 (e.g., Flege, 1980).

Moreover, unique interlanguage patterns have been identified. Gass (1984) focused on the VOT of initial English /b/ and /p/ produced by non-native speakers of English. She reported that non-native speakers tended to overcompensate for the VOT differences between native and target language by exaggerating a target language attribute, showing a tendency to overshoot the target language norm, and resulting in a pull away from the native language. In contrast, examining the VOT and closure duration of word-initial voiceless stops in Arabic-accented English, Fourakis and Iverson (1985) found that Arabic-accented English fell not between the norms established by native English (comparatively shorter closure, longer VOT) and Arabic (comparatively longer closure and shorter VOT), but rather constituted an articulation type characterizable as more "Arabic" than Arabic itself (namely, with even longer closure and shorter VOT than is characteristic of Arabic). In other words, the interlanguage stops pulled away from, rather than toward the target language. These unique kinds of patterns contribute to the diversity of the interlanguage phenomena.

Furthermore, VOT depends not only on the voicing feature of a stop, but is also affected by its acoustic-phonetic context (Lisker & Abramson, 1967b; Klatt, 1975; Port

& Rotunno, 1979). Among these is speaking rate (e.g., Miller, Green & Reeves 1986). Magloire & Green (1999) showed that the effects of speaking rate on VOT can differ between languages such as English and Spanish, and that L2 learners have difficulty incorporating the language specific VOT adjustments to accommodate speaking rate in their productions. Previous research has also investigated the relationship between VOT and the duration of a following vowel. Although earlier studies suggest that VOT and vowel duration are independently controlled (Lisker & Abramson, 1967b; Port & Rotunno, 1979; Weismer, 1980), recent research showed a trade-off between VOT and vowel duration, but with vowel duration not fully compensating VOT (Allen & Miller, 1999).

Other research has shown compensatory effects with VOT from within a syllable. Weismer (1980) argued that the duration of VOT in English and an adjacent closure interval are not independent. Flege and Port (1981) found VOT to be shorter in Arabic accented English than in both English and Arabic norms, while the closure intervals were found to be lengthened relative to both English and Arabic closures. With this, Flege and Port argued for a compensatory relation between VOT and the closure duration of voiceless stops. Fourakis and Iverson (1985) suggested that the timing for VOT and closure duration can be learned, though not be authentically grasped, by foreign speakers of English.

That VOT covaries with other acoustic-phonetic components within the syllable and that these differ across languages suggests that the syllable-internal timing of an L2 would have to be learned. An examination of the timing of L2 stop productions, including both VOT and its neighboring components, will extend findings on the interlanguage phenomenon to the syllable level.

The present study examined the temporal patterns of English stop-vowel syllables produced by Mandarin speakers of English. Mandarin differs from English in that it has no [+voice] stops in its phonemic inventory. Although in both English and Mandarin, the [-voice] stops can be phonetically implemented as {voiceless aspirated} and {voiceless unaspirated}, the VOT for Mandarin aspirated and unaspirated stops tends to be longer than for English (Dow, 1972; Svantesson, 1987). Moreover, the English {voiceless unaspirated} stops only occur in consonant clusters after /s/, whilst the Mandarin ones appear syllable initially.

The different VOT patterns of Mandarin and American English make Chinese learners of English a good testing ground for how acoustic-phonetic components of an interlanguage temporally interact within a syllable.

Method

An experiment was carried out to obtain three temporal measures in a stop-vowel syllable (stop closure duration, VOT, and vowel duration) for three types of speech productions (1) English syllables by native speakers of American English (AmE), (2)

English syllables by native speakers of Mandarin Chinese with English as an L2 (ChE), and (3) Mandarin syllables by native speakers of Mandarin Chinese (ChM).

Participants

Sixteen adults between 20–30 years old with no history of speech or hearing impairment participated in the experiment. The participants included eight native speakers of American English, and eight native speakers of Mandarin Chinese. The American speakers were students on a one-year exchange program at the University of Oslo, who had been in Norway for eleven months at the time of the recording. They reported to have North or Mid-western American as their principal dialect. The Chinese speakers were also students at the University of Oslo and had been in Norway for two to three years at the time of the study. All the Chinese speakers started formal L2 class instruction (ca. five hours/week for six to eight years) in English at around twelve years old. None of them had ever lived in an English-speaking country, but based on their TOEFL (Test of English as a Foreign Language) scores, all were considered competent users of the language. According to self-evaluation, they spoke English with slight or moderate accents. All speakers claimed to use Mandarin as their principal Chinese dialect.

Materials

Sets of English and Mandarin monosyllabic real words were selected, and are shown in Table 1. The words included all combinations of voicing and place of articulation for the prevocalic stops found in the respective languages. The prevocalic stops were followed by the vowel /aɪ/ which has approximately the same quality in both languages. In order to match English intonation, all Mandarin words had Tone 4 (high-low) which corresponds to a “high-low” intonation pattern in English (Dow, 1972).

Table 1. Words used to elicit stops in American English and Mandarin Chinese

	Labial	Alveolar	Velar
<i>English</i>			
[Voiceless aspirated]	pie /p ^h /	tie /t ^h /	Kai /k ^h /
[Voiceless unaspirated]	spy /p/	sty /t/	sky /k/
[Voiced]	buy /b/	die /d/	guy /g/
<i>Chinese</i>			
[Voiceless aspirated]	pai /p ^h /	tai /t ^h /	kai /k ^h /
[Voiceless unaspirated]	bai /p/	dai /t/	gai /k/

The target words were each used in the carrier phrases (1) "Let me say ____ again." in English, and (2) "Qing ni gei ____ zhu yin." (Please transcribe ____ to the phonetic alphabet.) in Chinese. The English and Mandarin carrier sentences were chosen to have similar syllabic structures, with the vowels (/eɪ/) preceding the target words being phonetically similar in the two languages. Five repetitions of each test sentence were used. Materials were randomized for each language.

Procedure

Audio-recordings of the participants were made in a sound attenuated room at the Department of Linguistics and Phonetics, University of Oslo, using a SONY TCD-10 PRO tape recorder and a SONY ECM MS-5 microphone, which was placed ca. 20cm from the speakers' mouth.

American English participants were recorded producing the English sentences. Chinese participants were recorded producing both the English and the Mandarin sentences, with the materials for the two languages presented in a counterbalanced order. Speakers were encouraged to read at a normal speaking rate. In total, the American speakers provided 360 productions (8 speakers x 9 target words x 5 repetitions) and the Chinese speakers produced 360 productions in English (8 speakers x 9 target words x 5 repetitions) and 240 productions in Mandarin (8 speakers x 6 target words x 5 repetitions).

Measurements

Recordings were acoustically analyzed using the Computerized Speech Research Environment software. Three measurements were made for each target word: closure duration, VOT, and vowel duration. In cases where stops occurred in a /s/-stop cluster, the duration of /s/ was also measured.

Closure duration was measured from the offset of the preceding vowel, indicated by a sharp decrease in energy in the region of F1 and F2, to the beginning of the noise burst signaling the release of the stop. VOT lag was measured from the beginning of the release to the onset of voicing indicated by the sudden onset of vertical striations in the second and higher formants (Klatt, 1975). VOT lead was measured as the negative interval between the release burst and the beginning of the low frequency voicing-bar in the stop closure. Vowel duration was measured from the onset of voicing to the offset of energy in the region of F1 and F2.

Results

Results were analyzed based on production type (AmE, ChE, ChM), voicing category (voiceless aspirated, voiceless unaspirated, voiced), and place of articulation (labial, alveolar, velar) with the timing components (VOT, closure duration, vowel duration) as dependent variables. For each timing component, two types of comparisons were made. First, voicing categories were compared to determine the voicing effects of stops in the productions by AmE, ChE, and ChM. Thus for each production type, a 2-factor analysis of variance (ANOVA) was calculated with voicing category and place of articulation as factors. Second, the three production types were directly compared to determine the extent to which the voicing effects of stops for ChE differ from AmE or ChM. This was done for each voicing category using a 2-factor ANOVA with production type and place of articulation as factors. In addition, data were analyzed at the syllable level, combining the results of VOT, closure duration and vowel duration, to determine the syllable internal timing relations in the productions by ChE compared to the AmE and ChM patterns.

Voice Onset Time

Table 2 lists mean VOT values as a function of voicing category and place of articulation produced by AmE, ChE, and ChM.

Table 2. Mean VOT (ms) as a function of voicing category and production type (AmE: American speakers of English; ChE: Chinese speakers of English; ChM: Chinese speakers of Mandarin), with standard deviations in parentheses.

Speakers	Voiceless aspirated				Voiceless unaspirated				Voiced*			
	p ^h	t ^h	k ^h	Mean	p	t	k	Mean	p	t	k	Mean
AmE	82 (24)	95 (20)	94 (10)	90 (20)	12 (5)	18 (4)	23 (5)	17 (7)	7 (3)	14 (28)	23 (9)	15 (10)
ChE	76 (21)	78 (18)	93 (20)	82 (21)	13 (7)	15 (7)	28 (14)	19 (12)	15 (14)	14 (7)	28 (12)	19 (20)
ChM	82 (20)	92 (21)	99 (22)	91 (22)	15 (8)	14 (5)	34 (34)	22 (23)	-	-	-	-

* The lead VOT values are not included here.

As expected, the main effect of voicing category on VOT revealed a reliable difference for all three types of production (AmE: $F(2, 357)=1401.2, p<.0001$, ChE: $F(2, 357)=728.1, p<.0001$, and ChM: $F(1, 238)=592.7, p<.0001$). As is illustrated in Table 2, in all three production types, the VOT values are substantially longer (Tukey post hoc analyses) for aspirated than for either unaspirated or voiced stops.² For AmE, the mean VOT values are also reliably longer for voiceless unaspirated than for voiced stops,

$F(1,238)=7.2, p<.008^3$. In contrast, the ChE productions showed no significant difference between voiceless unaspirated and voiced stops, $F(1,238)=1.6, n.s.$ The main effect of place of articulation also revealed a reliable difference for AmE ($F(2,357)=29.3, p<.0001$), ChE ($F(2,357)=23.7, p<.0001$), and ChM ($F(2,237)=6.1, p<.002$). For AmE the mean VOT values increase progressively as the place of articulation moves from the front to the back in the vocal tract. However, for both ChE and ChM, VOTs are substantially shorter for labial and alveolar than for velar stops. These findings are consistent with previously observed general patterns of effects on VOT for American English and Mandarin Chinese (e.g., Lisker & Abramson, 1964; Dow, 1972; Klatt, 1975; Volaitis & Miller, 1992).

The results were further analyzed by directly comparing the three production types (AmE, ChE, and ChM) within each voicing category. For voiceless aspirated stops, VOT values were significantly different among the three production types, $F(2, 357)=6.6, p<.002$, and as is shown in Table 2, VOTs were shorter for ChE than for both AmE and ChM. No reliable difference was observed among the production types for voiceless unaspirated stops, $F(2, 357)=2.3, n.s.$ However, a significant interaction between place of articulation and production type was observed, $F(4,355)=2.5, p<.044$. Post hoc analyses revealed that, for AmE, the mean VOT values increase progressively from /p/ to /t/ to /k/, whereas for both ChM and ChE, VOTs are similar for /p/ and /t/, but increase sharply for /k/ (Table 2). For voiced stops, the VOTs for ChE were significantly longer than those for AmE, $F(1, 238)=7.8, p<.006$. An interaction between production type and place of articulation was also observed, $F(2, 237)=3.6, p<.03$, with ChE having comparable VOT values for /p,t/ and a strikingly longer VOT for /k/, while AmE had a progressive increase with the backness of the place of articulation, similar to the patterns in the voiceless unaspirated condition. The only lead VOT values for AmE were produced by one speaker for the target word "buy" in the fourth and fifth repetitions (mean: -27ms). One Chinese speaker produced all five repetitions of the English "buy" (mean: -45ms) and one repetition of "die" (-45ms) with prevoicing.

In summary, in the Chinese speakers' English productions VOTs are generally consistent with those of the English norms, with voiceless aspirated stops having long-lag VOT values, and both voiceless unaspirated and voiced stops having short-lag VOTs. However, VOTs in the AmE and ChE productions do differ. For voiceless aspirated stops, whereas the VOTs in native English and native Mandarin productions are approximately the same, the Chinese speakers' productions of English tend to deviate from both the native English and the Mandarin productions, with reliably shorter VOT values than those in English and Mandarin. For voiceless unaspirated and voiced stops, the Chinese speakers' English productions resembled the VOT patterns of Mandarin rather than those of English, with comparable VOT values for /p,t/ and an abrupt increase for /k/, rather than a progressive increase with the backness of the place of articulation. Moreover, while the American speakers showed a reliable difference between the VOT values for English voiceless unaspirated and voiced stops, no such difference was observed for the Chinese speakers of English.

Closure duration

Mean stop closure durations produced by AmE, ChE, and ChM are shown in Table 3.

Table 3. Mean closure durations (ms) as a function of voicing category, and production type (AmE: American speakers of English; ChE: Chinese speakers of English; ChM: Chinese speakers of Mandarin), with standard deviations in parentheses.

Group	Voiceless aspirated				Voiceless unaspirated				Voiced			
	p ^h	t ^h	k ^h	Mean	p	t	k	Mean	p	t	k	Mean
AmE	99 (28)	90 (41)	66 (16)	85 (32)	80 (17)	45 (14)	47 (17)	57 (23)	116 (29)	101 (31)	102 (59)	107 (42)
ChE	105 (43)	114 (93)	83 (25)	101 (62)	81 (15)	74 (30)	57 (10)	71 (22)	114 (29)	120 (51)	99 (33)	111 (39)
ChM	82 (17)	77 (17)	83 (54)	81 (33)	133 (70)	107 (34)	88 (36)	108 (51)	-	-	-	-

A main effect of voicing category on closure duration was observed for all three types of production (AmE: $F(2, 357)=60, p<.0001$, ChE: $F(2, 357)=27.2, p<.0001$, and ChM: $F(1, 238)=13.5, p<.0001$). Post hoc analyses revealed that for AmE the mean closure duration for voiceless unaspirated stops is shorter than that for voiceless aspirated stops, which are in turn both shorter than for voiced stops. For ChE, the closure duration for voiceless unaspirated stops is significantly shorter than for both voiceless aspirated stops and voiced stops, but the duration of the latter two does not differ. The short closure durations for voiceless unaspirated stops are likely due to having been produced in consonant clusters, where the duration of cluster components is generally known to be shorter than for singletons (e.g., Klatt, 1973). In contrast, for ChM in which the voiceless unaspirated stops occur in syllable initial position, the mean closure duration is significantly longer for voiceless unaspirated than for aspirated stops. A main effect of place of articulation was observed for AmE ($F(2, 357)=18.8, p<.0001$), ChE ($F(2, 357)=9.8, p<.0001$), and ChM ($F(2, 238)=4.6, p<.001$). For both AmE and ChM, the closure duration is longer for labial than for alveolar and velar stops (AmE: 98ms, 79ms, 72ms, respectively; ChM: 108ms, 92ms, 86ms), while for ChE, it is significantly longer for labial and alveolar than for velar (100ms, 103ms, 80ms) stops.

A comparison among production types was carried out within each voicing category. For voiceless aspirated stops a significant difference was observed, $F(2, 357)=3.3, p<.04$, with ChE (101ms) having longer closure duration than both AmE (85ms) and ChM (81ms). For voiceless unaspirated stops the closure duration was significantly different among the three production types, $F(2, 357)=52.8, p<.0001$. Post hoc analyses showed that the closure duration is shorter for ChE (71ms) than that for AmE (57ms) in the consonant cluster context, which are in turn both shorter than that for ChM

(108ms) in the syllable initial position. A significant interaction between production type and place of articulation was also observed, $F(2,237)=3.9, p<.004$: For AmE, the mean closure duration is the longest for /p/, but decreases sharply for both /t/ and /k/, whereas for ChE, it decreases sharply from /p/ and /t/ to /k/ (Table 3). For voiced stops, closure duration was not reliably different when produced by AmE (107ms) and ChE (111ms), $F(1,238)=0.4, n.s.$

To summarize, for voiceless aspirated stops, the closure duration produced by ChE is longer than by either AmE or ChM, which have approximately the same values. When producing the English voiceless unaspirated stops, the ChE productions tend to resemble the target English pattern, showing a shortening effect in consonant clusters, although the values are intermediate to the native English and Mandarin productions. However, whereas the closure duration is substantially longer for voiced than for voiceless aspirated stops for AmE, no such difference was observed for ChE between these two voicing categories.

Vowel duration

Table 4 shows the mean durations for the English /aɪ/ produced by the American and Chinese speakers, and the Mandarin /aɪ/ produced by the Chinese speakers.

Table 4. Mean vowel durations (ms) as a function of preceding voicing category, and production type (AmE: American speakers of English; ChE: Chinese speakers of English; ChM: Chinese speakers of Mandarin), with standard deviations in parentheses.

Group	Voiceless aspirated				Voiceless unaspirated				Voiced			
	p ^h	t ^h	k ^h	Mean	p	t	k	Mean	p	t	k	Mean
AmE	311 (43)	318 (66)	315 (57)	315 (55)	320 (54)	326 (51)	323 (54)	323 (53)	338 (57)	356 (62)	359 (65)	351 (62)
ChE	263 (53)	269 (50)	254 (63)	262 (55)	282 (55)	289 (68)	280 (47)	284 (57)	301 (47)	298 (50)	290 (54)	296 (50)
ChM	172 (25)	173 (23)	171 (27)	172 (25)	210 (42)	197 (32)	201 (30)	202 (35)	-	-	-	-

A reliable difference in overall mean vowel duration was observed among all three types of production, $F(2, 932)=423.5, p<.0001$. Post hoc analyses showed that the duration of the English /aɪ/ produced by ChE (281ms) tends to be intermediate to the vowel durations produced by AmE and ChM, being significantly shorter than AmE, and longer than ChM.

The vowel durations produced by AmE, ChE, and ChM were compared based on the voicing of the preceding stop. For AmE the mean vowel duration was longer following a voiced stop (351ms) than following either a voiceless unaspirated stop

(323ms) or an aspirated one (315ms), while the latter two do not differ significantly, $F(2, 357)=12.9, p<.0001$. This pattern occurs, above and beyond any temporal compensation (e.g., Klatt 1976) that may be going on in the voiceless unaspirated condition to accommodate the preceding /s/. However, for ChE, the vowel duration does not differ significantly following a voiced stop (296ms) or voiceless unaspirated stop (284ms), both being longer than when following a voiceless aspirated stop (262ms), $F(2, 357)=10.7, p<.0001$. This pattern is similar to ChM, for which vowel duration is longer following an unaspirated stop (202ms) than an aspirated stop (172ms), $F(1, 238)=55.9, p<.0001$. The place of articulation of the preceding stops had no significant effect on vowel duration for AmE ($F(2, 357)=0.3, n.s.$), ChE ($F(2, 357)=2.5, n.s.$) or ChM ($F(2, 357)=0.2, n.s.$).

Overall, the vowel duration for ChE is intermediate to those for AmE and ChM. Vowels produced by AmE are longer following voiced stops than following either voiceless unaspirated or aspirated stops. For ChE, vowel durations are longer following voiced and voiceless unaspirated stops than following voiceless aspirated stops, similar to the ChM pattern of a shorter vowel duration following voiceless unaspirated stops than aspirated stops.

Syllable-internal timing

Figure 1 cumulatively illustrates the mean closure duration, VOT, and vowel duration with initial stops differing in voicing, as well as the duration of /s/ preceding voiceless unaspirated stops, produced by AmE, ChE and ChM. Figure 2 presents the corresponding relative values within a syllable, calculated by dividing each measure by the duration of the whole syllable it was extracted from. These relative values provide the timing distribution of each component within a syllable.

The overall syllable duration for ChE (448ms) is intermediate to that for AmE (504ms) and ChM (338ms), being significantly shorter than AmE and longer than ChM, $F(2, 930)=280.1, p<.0001$. For both AmE and ChE, a significant difference in syllable duration was observed as a function of voicing category (AmE: $F(2, 357)=27.6, p<.0001$, ChE: $F(2, 357)=12.2, p<.0001$); in both cases the syllable containing a voiceless unaspirated stop in a consonant cluster is, not surprisingly (e.g., Klatt, 1973), significantly longer than those containing either voiceless aspirated (490ms) or voiced (473ms) stops. For ChM, no difference in syllable duration was observed as a function of voicing category, $F(1, 238)=0.3, n.s.$

The relative durations presented in Figure 2 provide a timing distribution of closure duration, VOT and vowel duration as components of the syllable, by offsetting differences in overall syllable length across conditions.

Overall, for both the native English and native Mandarin productions, a tendency is observed for the timing components to compensate for each other to maintain a relatively balanced consonant/vowel distribution (e.g., Klatt 1976), as well as relatively balanced overall syllable durations across voicing categories.

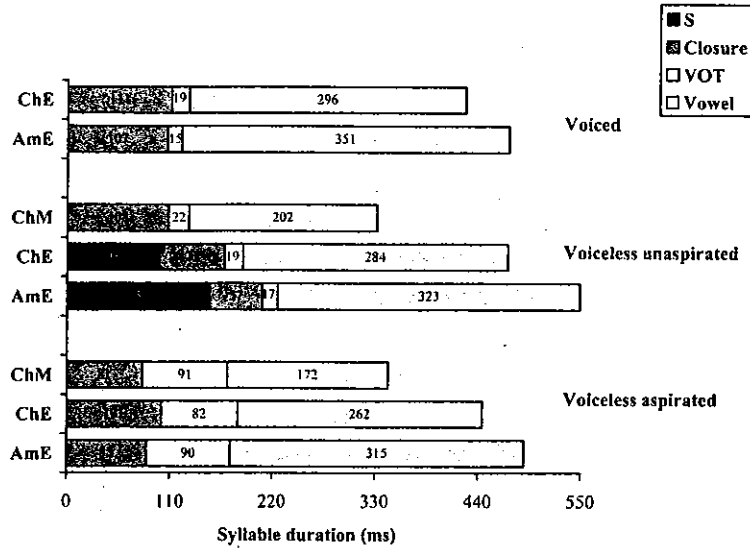


Figure 1. Mean durations of syllable, /s/, stop closure, VOT, and vowel produced by American speakers of English (AmE), Chinese speakers of English (ChE), and Chinese speakers of Mandarin (ChM). Voiced stops with lead VOT values are not included

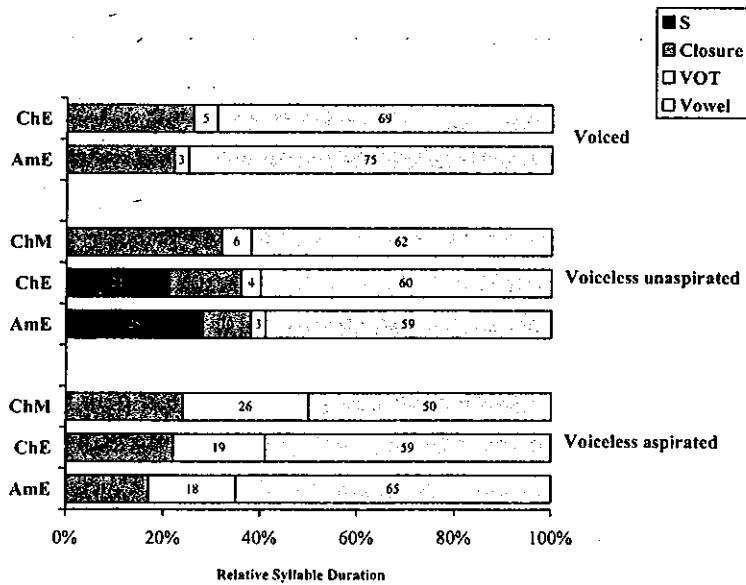


Figure 2. Mean relative durations of stop closure, /s/, VOT, and vowel produced by American speakers of English (AmE), Chinese speakers of English (ChE), and Chinese speakers of Mandarin (ChM). Voiced stops with lead VOT values are not included

In general, ChE tends to follow the patterns of AmE, maintaining a balanced consonant/vowel distribution as well as overall syllable duration across voicing categories. However, the internal timing components of ChE sometimes deviate from the native English productions, with the corresponding values being typically intermediate to AmE and ChM. Detailed analyses of the ChE productions of the syllable internal timing patterns, as compared to AmE and ChM, were conducted for each voicing category.

For the syllables containing voiceless aspirated stops, the syllable duration is significantly longer for AmE (490ms) than for ChE (445ms), which is in turn longer than that for ChM, $F(2,351)=108.2$, $p<.0001$. The results presented earlier revealed no difference between AmE and ChM for either VOT or closure duration, while the vowel duration is significantly shorter for Mandarin than for English. As such, the difference in overall syllable duration between the two languages is likely attributable to the vowel duration difference. As Figure 2 illustrates, vowel duration in AmE maintains a longer portion (65%) of the syllable than the consonant (closure duration + VOT: 35%), whereas in Mandarin, the portion of the syllable distributed between the consonant (50%) and vowel (50%) are more comparable. For ChE, interestingly, the VOT and closure durations revealed inverse patterns compared to AmE and ChM. As shown in the previous sections, while both AmE and ChM have reliably longer VOT and shorter closure duration than ChE, ChE has a reliably shorter VOT and longer closure duration than AmE and ChM. Thus the overall consonant duration for ChE (183ms) tends to be comparable to AmE (175ms) and ChM (172ms). However, the vowel duration for ChE (262ms) is significantly shorter than that for AmE (315ms), but significantly longer than ChM (172ms), $F(2, 357)=261.1$, $p<.0001$. Overall, the distribution of the consonant (41%) and vowel (59%) for ChE is still intermediate to that for AmE (35%:65%) and ChM (50%:50%).

The syllable for voiceless unaspirated stops contains a consonant cluster (/s/+stop) for AmE and ChE, whereas for ChM, the voiceless unaspirated stops occur syllable initially. However, the syllable durations for both AmE (549ms) and ChE (471) are substantially longer than for ChM (341ms), $F(2, 357)=215.7$, $p<.0001$, not only in the consonant cluster condition, but also when /s/ was excluded, $F(2, 357)=36.2$, $p<.0001$. Post hoc analysis also showed that the syllable duration for ChE is significantly shorter than that for AmE. As shown in the previous sections, whereas their VOT values are comparable, the closure duration for ChE is significantly longer than that for AmE. Conversely, the preceding /s/ duration is significantly shorter for ChE (98ms) than for AmE (152ms), $F(1, 178)=208.5$, $p<.0001$, and ChE (284ms) also has a reliably shorter vowel duration than AmE (323ms), $F(1, 238)=23.1$, $p<.0001$. In the timing distribution, it is interesting to note that, although the internal distributions are different for AmE (/s/: 28%, closure: 10%, VOT: 3%, Vowel: 59%) and ChE (21%, 15%, 4%, 60%, respectively), with ChE having a shorter /s/ and longer closure duration than AmE, the relative consonant and vowel proportion for ChE (40%:60%) approximates that for AmE (41%:59%)

For English voiced stops, a reliable difference was observed between AmE and ChE, $F(1,238)=17.8$, $p<.0001$, with the mean syllable duration for AmE (472ms) being longer than that for ChE (423ms). While no reliable difference occurred between AmE and ChE for closure duration, the Chinese speakers did produce a longer VOT (cf. the statistical results in the previous sections) and shorter mean vowel duration than the American speakers, $F(1, 238)=42.8$, $p<.0001$. Cumulatively, the consonant takes a larger portion of the syllable for ChE (31%) than for AmE (25%). Furthermore, since the voiced stops for AmE and ChE appear syllable initially, and are phonetically realized as voiceless unaspirated stops, which resembles the patterns for ChM voiceless unaspirated stop, a comparison was made among these production types. The results show that the closure duration for the three types of production does not differ, $F(2, 357)=0.7$, n.s. The VOT values for voiceless unaspirated stops in ChM do not differ from the voiced stops in ChE, but are longer than those of AmE voiced stops, $F(2, 357)=3.4$, $p<.035$. These results indicate that the production of the ChE voiced stops resembles that of the voiceless unaspirated stops in their native language Mandarin. However, since the vowel duration of ChE voiced stops is intermediate to AmE voiced stops and ChM voiceless unaspirated stops, $F(2, 357)=244.1$, $p<.0001$, the overall proportion of the consonant in the syllable (31%) is still intermediate to AmE (25%) and ChM (38%).

Taken as a whole, the syllable duration of AmE is longer than that of ChM. As compared to ChM, the AmE vowel duration maintains a larger portion of the syllable than the consonant. Overall, the syllable duration and its internal components produced by ChE are typically intermediate in value to AmE and ChM. More importantly, the ChE speakers tend to approximate the syllable timing patterns of the native English speakers, with necessary temporal compensations to maintain a balanced timing distribution; although internal to the syllable, their adjustments are sometimes still inconsistent with the native patterns. In the voiceless aspirated condition, the shorter (than AmE) VOT for ChE is compensated by a longer closure for a consonant duration that is more comparable to that of the AmE. Similarly, in the voiceless unaspirated condition, ChE approximated the native AmE patterns with the shortening of closure duration to accommodate the preceding /s/ in consonant clusters, although the value is still longer compared to that of AmE. However, the longer (than AmE) closure for ChE is partially compensated by a shorter (than AmE) /s/ duration, resulting in an overall consonant distribution comparable to that of AmE. For syllables containing voiced stops, the timing distribution does not differ much for ChE and AmE. With relatively longer VOT and shorter vowel duration, the consonant for ChE takes a larger portion in the syllable than that for AmE.

Discussion

This investigation assumed that in acquiring English stops, particularly their timing characteristics, Mandarin speakers would be able to establish an interlanguage system

which bears intermediate characteristics of both Mandarin and English, and that their productions would approximate the patterns of the English model, but not necessarily with full accuracy.

The results confirm this prediction. The timing patterns of the stops of English and Mandarin contain language-specific characteristics. Although the Chinese speakers' productions of English usually tended in the direction of patterns exhibited by the native American speakers, they often failed to authentically realize all the phonetic details of the stops in the English norm. However, phonetic interference from Mandarin to English does not mean that the Chinese speakers would simply substitute similar sounds from Mandarin for those in English. Rather, their productions of English closure duration, VOT, and vowel duration are intermediate to the native English and Mandarin norms, with the temporal deviations often compensated by temporal adjustments of other components in the syllable, suggesting the establishment of a structured interlanguage system.

The results for voiceless aspirated stops showed that the mean VOT produced by native speakers of American English (AmE) are similar to those produced by native Mandarin speakers (ChM). However, when producing English, the Chinese speakers deviated from both English and Mandarin by producing shorter VOTs than either of the two languages concerned. Closer inspection of each stop, /p^h, t^h, k^h/, showed that the VOT for ChE is shorter than that for AmE and ChM, suggesting that the production pattern is very systematic. Furthermore, like the patterns in the English norm, VOT produced by ChE increased systematically as the place of articulation moves from the front to the back in the vocal tract. This phenomenon is rather unique, since it is different from the common finding that VOT produced by an L2 learner is often intermediate to the L1 and L2 norms (e.g., Caramazza et al, 1973; Flege & Hillenbrand, 1984; Flege & Eefting, 1987a). As described previously, Gass (1984) showed that non-native speaker productions display a tendency for VOT values to exaggerate target language norms, resulting in a departure from the native language. Fourakis and Iverson (1985), on the other hand, revealed a reverse pattern of a direction away from the target language, resulting from overshooting a native language characteristic. The present data exhibit another unique interlanguage phenomenon, with a deviation from both the native and the target language, which have the same VOT value.

This pattern repeats in the production of ChE closure duration. AmE and ChM have similar closure durations for voiceless aspirated stops. If the Chinese speakers directly transferred this temporal characteristic from Mandarin, they would be expected to produce a near authentic American English closure duration. However, the Mandarin speakers produced a longer closure duration than either AmE or ChM, indicating that the Chinese speakers' attempt to modify their productions of voiceless aspirated stops in English. It is interesting to note that the sum of the VOT and stop closure intervals for ChE tends to approximate that for AmE, suggesting that the Chinese speakers may decrease VOT in English to compensate for their relatively long closure duration.

This is consistent with Flege and Port's (1981) finding of a compensatory relation between VOT and the closure duration of voiceless stops, in that VOT was shorter in Arabic accented English than in both English and Arabic, while closure intervals were found to lengthen relative to both English and Arabic closures. The present results also support Weismer's (1980) argument of a constant "voiceless interval" (VLI), which can be learned by L2 speakers (Fourakis & Iverson, 1985). The Chinese speakers seemed to be able to implement the VLI as a unit of timing in English, although internal to this interval, they produced both VOT and closure duration markedly different from both the English and Mandarin norms.

Further evidence of the Chinese speakers' formation of an interlanguage system is their production of the temporal patterns associated with voiceless unaspirated stops in English. For example, the closure duration of voiceless unaspirated stops by ChE followed the pattern of the English norm with a shorter duration in a consonant cluster context (after /s/), and did so for all places of articulation. The values they produced were, nevertheless, intermediate to those in English and Mandarin. On the other hand, their production of /s/ duration was shorter than that of the native English productions.

As described previously, Weismer (1980) claimed that the voiceless interval (VLI) is constant in English for word initial consonants regardless of place and manner of articulation. In the case of an /s/-stop cluster, it is argued that VLI will still be constant (Fourakis & Iverson, 1985), with the VLI based on the sum of the fricative, stop closure, and VOT duration. Fourakis and Iverson (1985) found that English VLI in word-initial clusters can also be learned by foreign speakers of English. The Chinese speakers' production of a longer closure duration compensated by a shorter /s/ duration than the English norm indicates an approximation of a constant VLI, suggesting the relevance of syllable-internal timing in an interlanguage system.

In addition to timing of the stops, the Chinese speakers' articulation of the English vowel supports the hypothesis from another angle; their vowel production patterns systematically follow the English norm in the context of stops across voicing categories and places of articulation. The results reveal that the Chinese speakers' production of the vowel duration of English /aɪ/ is intermediate to the English and Mandarin norms. This pattern was observed in each context examined; that is, in the contexts of prevo-calic stops across three places of articulation and three voicing categories. The Chinese speakers' productions of the English /aɪ/ agrees with Flege's (1980) description of an interlanguage system; on the one hand, it is temporally intermediate to the English and Mandarin /aɪ/, while on the other hand, it is produced with consistency and stability in the phonetic contexts concerned. Moreover, the relative timing of voicing correlates of the English stops and the following vowel is comparable to that of the American speakers, providing further evidence of the Chinese speakers' systematic approximation to the English norm.

Overall, the present results reveal systematic interlanguage behavior of the Mandarin speakers of English, both at syllable level and from temporal relationships among components within the syllable. Comparing the consonant and vowel distribu-

tion within the syllable, Chinese speakers' productions of English consonant-to-vowel timing are intermediate between the balanced consonant-to-vowel proportions of the Mandarin speakers and the relatively longer vowel durations of the American English speakers. As described above, the internal timing components of ChE often deviate from AmE, with the closure duration, VOT, and vowel duration being intermediate to AmE and ChM. However, ChE temporal deviations were often compensated by temporal adjustment of other components in the syllable, approximating a more balanced native-like distribution between consonant and vowel.

Interlanguage behavior has been explained in terms of phonemic and phonetic categories. For example, when articulating an L2 stop which has no counterpart in their L1 phonemic inventory, learners can usually produce the VOT value that lies in the L2 category (Flege & Port, 1981; Port & Mitleb, 1983). However, when producing an L2 stop which has an L1 phonemic counterpart, but is phonetically implemented in a different category from that in their L1, learners often fail to establish the phonetic category for the L2 stop. Consequently, their productions of VOT remain in the L1 category (Flege & Port, 1981; Flege & Hillenbrand, 1984). On this basis, Flege and Port (1981) argue that the most important interference from one's L1 to L2 during the process of L2 acquisition occurs at the level of phonetic implementation rather than at the phonemic level (Flege & Port, 1981).

In addition to the phonemic and the phonetic implementation levels, there is a third level of phonetic motor realization, providing language-specific details (Flege, 1991a). In this account, learners may be able to produce the VOT in an L2 phonetic category, but their productions may not yet be an authentic L2 stop as a consequence of the application of different phonetic realization rules. Little research has examined the case in which L1 and L2 speech sounds lie in the same phonetic category, but differ in phonetic (language-specific) details. In terms of motor realization, is it possible that L2 learners are unable to realize the L2 specific rules, resulting in the inaccurate production of the L2 sound within the category they have already formed (in L1)? The present study has provided evidence of L1-L2 interaction at the phonemic, phonetic, and motor levels.

When producing the English voiced stops which have no counterparts in the Mandarin phonetic inventory, the Chinese speakers were expected to form an English VOT category for voiced stops. Results of the present study are consistent with this hypothesis in two ways. First, in most cases, the phonemically voiced English initial stops are phonetically implemented as {voiceless unaspirated} with short-lag VOT values by the native speakers. Although voiced stops do not occur in Mandarin at the phonological level, they are not "new" at the phonetic level for the Mandarin speakers. The Chinese speakers thus do not need to establish a new category for phonemically new English phones, and they succeeded in phonetically realizing the English voiced stops as {voiceless unaspirated}, producing VOT values in that category. Second, occasionally, the English voiced stops are phonetically implemented as {voiced} with lead VOT values. In the present study, only a few cases of prevoicing were realized, and

they occurred among productions by both American and Chinese speakers of English. That the Chinese speakers can produce English voiced stops with prevoicing provides convincing evidence of category formation. In addition to VOT, the Chinese speakers' successful production of another temporal correlate, closure duration for English voiced stops, provided further evidence that the Chinese speakers are able to produce phonemically "new" phones in English.

At the phonetic level, the Chinese speakers' productions of VOT for English significantly differed between aspirated and unaspirated stops, just as the American speakers' productions did. In terms of phonetic implementation, the Chinese speakers authentically realized the English voiceless aspirated and voiceless unaspirated stops as the stops in the same phonetic categories as corresponding stops in Mandarin, resulting in the correct implementation of the stops. However, the Chinese speakers did not accomplish all of the temporal and/or phonetic details specific to English. According to the supposition concerning the level of phonetic realization, an L2 learner may be able to produce the VOT in an L2 phonetic category, but his/her production might not yet authentically use the phonetic realization rules that are specific to the L2 (Flege, 1991a). On this basis, even for English and Mandarin stops using the same phonetic category, the Chinese speakers may produce unauthentic English VOT and closure durations by using different phonetic realization rules to output a single phonetic category.

The present results are consistent with this hypothesis. Although usually approaching the English norms, the Chinese speakers did not manage to produce VOT values for English stops authentically. For example, although both English and Mandarin voiceless aspirated stops lie in the same phonetic category, the Chinese speakers' productions revealed that other realization rules than those specific to either English or Mandarin VOT were used to implement the English stops. Likewise, the inverse pattern of the closure duration and VOT for English voiceless aspirated stops produced by the Chinese speakers as compared to the English norms suggests that the temporal details specific to English stops are something that the Chinese speakers partially implemented, but had not fully achieved.

Conclusions

The present study shows that American English and Mandarin Chinese have similar VOT values for voiceless aspirated and unaspirated stops, but each language manifests language-specific temporal patterns within the syllable. The Chinese speakers successfully implemented the English voiced stops which have no counterparts in the Mandarin phonemic inventory as having either short-lag (in most cases) or lead (occasionally) VOT values. Although they were able to implement the English stops in the correct phonetic categories, they often failed to accomplish the detailed temporal characteristics that are English-specific. However, their temporal implementations were systematic and indicative of an interlanguage.

Overall, the results revealed that the Mandarin speakers' productions of English stops and their temporal characteristics were phonetically intermediate to the English and Mandarin norms, again reflecting the existence of an interlanguage system. On the one hand, they follow the pattern exhibited by the native American speakers; on the other hand, they are not fully independent of interference from Mandarin.

As is evident from this study, as well as much previous research, the concept "interlanguage" continues to account for and shed light on L2 acquisition phenomena. Along the path towards a second language, temporal remnants of the native language and learned patterns of the second language are evident in the learner's interlanguage. Speakers of a foreign language appear to modify pre-existing phonetic patterns, including components of a syllable and their relative timing. By modifying these patterns, they make slow progress in acquiring the acoustic-phonetic norms of a target language, leading to the appropriate final destination of the target second language.

Notes

1. A portion of this research was conducted as part of a M.A. thesis at the Norwegian University of Science and Technology by the first author under the direction of the second author (Wang, 1995). Portions of this study were reported at the 147th Meeting of the Acoustical Society of America, New York.
2. Tukey tests were used for all post hoc analyses reported in the present study.
3. The mean differences (below 5ms) might not be perceptually salient. Note that all the findings in the present study are acoustic differences; whether these differences are below what would be perceptible will be examined separately.