

CROSS-LANGUAGE SWITCHING IN STOP CONSONANT PERCEPTION AND PRODUCTION BY DUTCH SPEAKERS OF ENGLISH

James Emil FLEGE

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

Wieke EEFTING

Institute of Phonetics, University of Utrecht, Utrecht, The Netherlands

Received 28 April 1986

Revised 28 October 1986

Abstract. Voiceless /p,t,k/ are implemented as aspirated stops in English, but as unaspirated stops in Dutch. We examined identification of a voice onset time (VOT) continuum ranging from /da/ to /ta/ in two language "sets" designed to induce native Dutch subjects to perceive the stimuli as if they were listening to Dutch or English. The effect of language set was highly significant, but the boundary shift was very small (2.1 ms longer in English than in Dutch) for three groups of subjects differing widely in English language proficiency. It nevertheless showed the subjects were aware of acoustic differences distinguishing Dutch and English /t/ and that the procedures were effective in creating differing language sets. Nearly every subject *produced* a longer mean VOT in English than Dutch /t/. The magnitude of the production shift was significantly greater for proficient than non-proficient subjects. Proficient Dutch speakers of English produced Dutch /t/ with shorter VOT values than non-proficient subjects, suggesting they formed a new category for English /t/. We speculate that the language set effect was small because subjects used their English /t/ category to identify stops in both sets. This was probably due to the fact that the synthetic stimuli, which were modeled on the English /t/ – /d/ contrast, differed substantially from their Dutch /t/ category.

Zusammenfassung. Stimmlose /p,t,k/ werden im Englischen als aspirierte, im Niederländischen hingegen als nicht-aspirierte Explosivlaute klassifiziert. Untersucht wurde die Identifikation eines VOT-Kontinuums beim Übergang von /da/ nach /ta/ in zwei unterschiedlichen Sprach-"Umgebungen", die niederländischen Versuchspersonen den Eindruck vermitteln sollten, als hörten sie die Stimuli in englischem oder niederländischem Sprachmaterial. Die Umgebung zeigte hochsignifikante Auswirkungen, doch konnte bei drei Versuchspersonengruppen unterschiedlicher Beherrschung des Englischen nur eine geringe Verschiebung der Beurteilungsgrenze beobachtet werden (VOT 2,1 ms länger im Englischen). Immerhin wurde deutlich, dass die Versuchspersonen die akustischen Unterschiede zwischen dem niederländischen und englischen /t/ bemerkten und dass die Prozeduren bei der Bildung unterschiedlicher sprachlicher Umgebungen wirksam waren. Fast jede Versuchsperson *produzierte* eine längere durchschnittliche VOT für ein englisches /t/. Dabei war der Betrag der Verschiebung in der Produktion der Fortgeschrittenen deutlich grösser. Diese Gruppe artikulierte ein niederländisches /t/ mit kürzeren VOT-Werten als die übrigen Versuchspersonen, was nahelegt, dass sie für das englische /t/ eine neue Kategorie bildete. Wir vermuten, dass die Wirkung der Sprachumgebung gering war, da alle Versuchspersonen ihre englische /t/-Kategorie zur Identifikation der Explosivlaute in beiden Umgebungen verwendeten. Dies hing wahrscheinlich mit dem Umstand zusammen, dass die synthetischen Stimuli, die aus dem /t/-/d/-Kontrast des Englischen abgeleitet waren, sich deutlich von ihrer niederländischen /t/-Kategorie absetzten.

Résumé. Les consonnes sourdes /p,t,k/ sont classées comme consonnes aspirées en anglais, mais comme non-aspirées en néerlandais. Nous avons examiné l'identification d'un continuum de délai d'établissement du voisement (VOT) allant de /da/ vers /ta/ pour deux "ensembles" linguistiques conçus pour induire des sujets néerlandophones à percevoir les stimuli comme s'ils écoutaient soit du néerlandais, soit de l'anglais. L'effet de l'ensemble linguistique s'est montré très significatif mais le déplacement de la frontière a été très faible (2,1 ms. de plus en anglais) pour trois groupes de sujets différant largement par leur maîtrise linguistique en anglais. Il en est ressorti néanmoins que les sujets étaient conscients des différences acoustiques distinguant le /t/ en néerlandais du /t/ en anglais et que les procédures étaient efficaces quant à la

création d'ensembles linguistiques différents. Quasiment chaque sujet *a produit* un VOT moyen plus long pour un /t/ en anglais qu'en néerlandais. La taille du déplacement à la production était significativement plus importante chez les sujets avancés que chez les sujets moins avancés. Les sujets plus avancés dans leur maîtrise de l'anglais ont produit un /t/ néerlandais avec des valeurs de VOT plus brèves que les sujets moins avancés ce qui suggère qu'ils ont créé une nouvelle catégorie pour le /t/ anglais. Nous avançons l'hypothèse que l'effet d'ensemble linguistique a été faible parce que les sujets ont utilisé leur /t/ anglais pour identifier les occlusives dans les deux ensembles. Ceci a probablement été dû au fait que les stimuli, synthétisés sur le modèle du contraste /t-/d/ en anglais, différaient considérablement de leur /t/ en néerlandais.

Keywords. Phonetic categories, speech perception, speech production, second language learning, second-language proficiency.

1. Introduction

The authentic pronunciation of phones in a foreign or second language (L2) may require the establishment of new phonetic categories. For example, native speakers of Spanish must establish new vowel categories for English because the English vowel system contains more vowels (15) than the five-vowel system of Spanish (/e/, /i/, /o/, /u/ and /a/ (J. Harris, 1989)). Perhaps the greatest single cause of foreign accent (J.E. Flege, 1984) is the “new wine in old bottles” phenomenon: production of L2 phones via the nearest possible L1 category. The success or failure of native Spanish L2 learners¹ in producing English vowels such as /æ/ depends on their ability to recognize the separate category status of a class of phones (i.e., the allophones of /æ/). Once this happens, they must determine which lexical items in L2 contain /æ/, establish a central phonetic representation ({æ}) that specifies “prototypical” properties of the newly established category, and develop sensorimotor realization rules for outputting the new category rapidly and reliably in spontaneous speech.

It remains uncertain in many instances, however, whether L2 learners do—or should—establish new phonetic categories. Take, for example, the /i/ found in English and Danish. The Danish vowel is “higher” and “fronter” than its English

counterpart, leading to a significant cross-language acoustic difference (S.F. Disner, 1986). The observed difference in production—which is apt to be detectable auditorily (J.E. Flege, 1984)—leads to the inference that the phonetic categories established for /i/ by native speakers of Danish and English differ. The question arises whether native speakers of Danish will establish a new category for /i/ when learning English L2 as readily as Spanish L2 learners will establish a new category for English /æ/. Perhaps the difference between the L1 and L2 phones must, collectively, exceed some “phonetic distance” criterion before the processes leading to establishment of a new category are triggered.²

These questions have been explored in some detail by studies examining the production and perception of stop consonants in word-initial (pre-vocalic) position. Non-native adults with relatively little English language experience produce the long-lag /p,t,k/ of English with short-lag voice onset time (VOT) values if /p,t,k/ are realized as voiceless unaspirated stops in their L1 (i.e., as short-lag stops). Those with relatively more L2 experience typically produce English /p,t,k/ with VOT values *intermediate* to the long-lag phonetic norm of English and the short-lag VOT norm for corresponding stops in their L1 (A. Caramazza et al., 1973; J.E. Flege, 1980; 1987b; J.E. Flege et

¹ We use the term “L2 learner”, rather than the more common term “bilingual”, because we view the former term as being somewhat more neutral with respect to a number of important empirical and theoretical issues. It should not be regarded as an implicit claim concerning the frequency of use of L1 versus L2, a statement concerning how the L1 and L2 were learned, nor an indication of the relative proficiency in L1 versus L2.

² For example, Danes might go on saying English words with an unmodified “Danish” /i/ no matter how much experience they have with “English” /i/ tokens spoken by native speakers of English. Or, they might distinguish /i/ when speaking their two languages by realizing a single phonetic category ({i}) with slightly different realization rules (P.A. Keating, 1984). This is apparently what monolinguals do in order to produce vowels differently in differing social contexts (W. Labov, 1981).

al., 1987a; 1984; 1981; R.C. Major, 1987; R.F. Port et al., 1983).

The seeming inability of non-natives to produce English /p,t,k/ with authentic VOT values extends to adults who learned English L2 as children, as well as to children (J.E. Flege et al., 1987a; L. Williams, 1979).³ This minimizes the likelihood that differences between native and non-native adults is simply due to their having passed a "critical period" for speech learning. Instead, Flege (1981, 1984, 1987) hypothesized that adult L2 learners fail to make effective use of the sensory information associated with certain phones in L2 as the result of a mechanism termed "equivalence classification". If short-lag and long-lag realizations of /p,t,k/ in L1 and L2, respectively, are judged to be realizations of the same categories, L2 learners will not establish new phonetic categories for the long-lag L2 stops, even though the acoustic differences distinguishing the corresponding L1 and L2 stops may be detected auditorily.

This hypothesis led to the strong claim that adult L2 learners will never match native English speakers in producing /p,t,k/ (J.E. Flege et al., 1984). The production studies cited above are consistent with this claim, but are insufficient to prove that differences between native and non-native speakers arose from failure on the part of the L2 learners to establish new phonetic categories for English /p,t,k/. The subjects in previous studies might, for example have established new categories, but not yet have learned to realize those categories effectively.⁴ Another possibility

is that the L2 learners *did* establish new categories, and realized them just as they were represented centrally, but that their *new categories* differed from those of English native speakers (perhaps due to L2 phonetic input from other non-native speakers).

If adult learners of English L2 develop new phonetic categories for English /p,t,k/, it should be evident in their perception of stops. Previous research has indicated that groups of subjects who learned English after about the age of 5–6 years show phoneme boundaries at VOT values that are often *intermediate* in value to the norms established for the L1 and L2 (A. Caramazza et al., 1973; J.E. Flege et al., 1987a; L. Williams, 1977; 1979).⁵ In some instances, individual subjects have category boundaries at values that are genuinely "intermediate" to the L1 and L2 norms, suggesting the phonetic merger (or assimilation) of the properties of corresponding stops in L1 and L2. In other instances, however, the individual subjects seem to have had boundaries at locations close to the L1 norm or the L2 norm.

One difficulty in interpreting the results of previous identification experiments is that it is not always clear *how* the subjects processed the stimuli. Differing results might be obtained depending on whether the subjects employed the phonetic criteria used to identify stops in L1, or to identify L2 stops.

Three previous identification experiments addressed this issue by attempting to create different "language sets".⁶ The principal method used for doing so in two studies which employed synthetic consonant–vowel stimuli differing in VOT (A. Caramazza et al., 1973; L. Williams, 1977) was to interact with the subjects in L1 or L2. Neither

³ There is some evidence that "early childhood bilinguals" (i.e., individuals who learn two "first" languages concurrently in early childhood (B. McLaughlin, 1978)) may produce both the long-lag realizations of /p,t,k/ in English, and the short-lag realizations of /p,t,k/ in their other language, authentically (M. Mack, 1983; L. Williams, 1977). If English is clearly the *second* language learned, however, exposure to English in early childhood may *not* lead to an authentic pronunciation of English /p,t,k/. Flege and Eefting (1987a) found that two groups of native Spanish adults who began learning English by the age of 5–6 years produced English /p,t,k/ with significantly shorter VOT values than monolingual English adults. The same was true for native Spanish children who began learning English at about the same age.

⁴ Zlatin and Koenigsknecht (1973) found that the voiced-voiceless boundaries of English adults and children did not

differ significantly in an identification experiment using several VOT continua, but that adult subjects produced /p,t,k/ with significantly longer VOT values than children. This implies that, in L1 acquisition, the development of realization rules may lag behind the establishment of the phonetic categories used in identification.

⁵ We use the term VOT "norm" for a language to refer to the mean value observed for groups of (hopefully) monolingual subjects, as opposed to more subjective notions of how a phone "ought" to sound (P. Linell, 1982).

⁶ Another, more recent, study (J. Werker et al., 1984) manipulated psychological set in a discrimination experiment. Changes in instruction did not influence performance.

study showed a shift in the location of the mean voiced-voiceless category boundary as a function of language set.

This suggested that the subjects did *not* employ different phonetic criteria for identifying the acoustically different stops of L1 and L2.⁷ It is possible, of course, that these studies yielded negative findings because the procedures used for creating the different language sets did not succeed in inducing subjects to process the stimuli *as if* they were listening to *two different languages*.

A later study by Elman et al. (1977), which used more elaborate procedures, did show a significant effect of language set. The language used to interact with the subjects was again varied. In addition, the stimuli were appended to L1 (Spanish) or L2 (English) carrier phrases, and L1 or L2 “filler” words were interspersed among the stimuli in the two language sets. The subjects identified five natural syllables as /ba/ or /pa/ (one stimulus produced by a “bilingual” talker with lead VOT, one with long-lag VOT, and three with short-lag VOT values). As expected, the short-lag tokens were usually identified as /pa/ by monolingual Spanish subjects, but as /ba/ by monolingual English subjects. “Strong”, but not “moderate/weak English-Spanish bilingual” subjects identified significantly more syllables as /pa/ (the Spanish mode of labeling) in the Spanish than English set.

The authors concluded that the language set manipulation led some, but not all, subjects to vary the placement of “category boundaries”. The significant correlation between “degree of fluency” in Spanish and English, and the magnitude of the shift in the percentage of /pa/ judgments given in response to short-lag tokens, suggested that only those subjects who were relatively proficient in English L2 identified stops differently in the two language sets. This, in turn, suggested that new phonetic categories for English /p,t,k/ are established as L2 learners gain proficiency in English.

⁷ Caramazza et al (1973) examined native French speakers of English who were selected on the basis of self-evaluated ability to speak English, and according to reading speed. Data was gathered in two sessions separated by 2–3 weeks. Williams (1977) examined “bilingual” subjects who spoke both Spanish and English with little evidence of a foreign accent.

There are several reasons for caution, however, in accepting the conclusion that category boundaries showed a shift across language sets in the Elman et al. (1977) study. The increase in /pa/ responses in the Spanish compared to English set was small. Spanish monolinguals identified about 81% more short-lag tokens as /pa/ than English monolinguals. In comparison, even the “strong” bilingual subjects showed only a 17% increase in /pa/ judgments across sets. This finding cannot be directly interpreted as a shift in the location of “phoneme boundaries” because only three natural tokens, not the members of a VOT *continuum*, were identified.⁸

The shift that was observed may have derived from application of a phonological rule, rather than from the use of differing phonetic criteria for identifying stops in the two language sets. The carrier phrase appended by Elman et al. to stimuli in the Spanish set (“Escriba la palabra...”) may have triggered a Spanish phonological rule that causes /b,d,g/, but *not* /p,t,k/, to be realized as fricatives in word-initial position (J.E. Flege et al. 1984; J. Harris, 1969).⁹ This is because it created a context in which a word-initial stop such as /b/—but not /p/—would be fricativized in the *production* of Spanish. The native Spanish subjects may therefore have judged more syllables as /pa/ in the Spanish than English set because the CV stimuli were initiated by a *stop*.

There is also reason for caution in accepting the conclusion that the magnitude of the language set effect depended on subjects’ proficiency in L2. Subjects’ L2 proficiency was not assessed objectively in the Elman et al. (1977) study.¹⁰ The aim

⁸ Some tentative inferences concerning a possible shift in category boundaries might have been drawn, for the three short-lag stimuli had slightly different VOT values. Unfortunately, the percentage increase in /pa/ judgments was not reported separately for the three stimuli. The authors indicate that pilot testing revealed the lack of a language set effect for synthetic stimuli.

⁹ Macken (1980) hypothesized that children acquiring Spanish as an L1 may learn to identify word-initial obstruents as voiceless (i.e., /p,t,k/) not on the basis of VOT, but on the basis of whether or not they hear a *stop* rather than a fricative.

¹⁰ Degree of “fluency” in Spanish and English was estimated by two judges using an 8-point rating scale. One judge spoke both Spanish and English; the other judge was apparently monolingual in one of those languages.

of the present study was therefore twofold: to determine if subjects identify stops differently in their two languages as a function of language set; and, if so, whether the magnitude of the shift depends on objectively-determined differences in L2 proficiency.

The study examined the identification of the members of a continuum ranging from /d/ to /t/ by 50 native speakers of Dutch who began learning English at the age of 12 years. In Dutch, /p,t,k/ are implemented as voiceless unaspirated stops (L. Lisker et al., 1964; A. Slis et al., 1969). Pilot work, as well as previous experiments examining the identification of stops by native speakers of Spanish and French (A.S. Abramson et al., 1973; J.E. Flege et al., 1986; L. Williams, 1977), suggested that Dutch subjects would show category boundaries at shorter VOT values than native English subjects. If the Dutch subjects identified the stimuli differently as a function of language set, they should therefore show category boundaries at longer VOT values in an English than a Dutch set.

To assess the possible effect of L2 proficiency on the magnitude of cross-language shifts in identification, it was necessary to develop procedures with which to objectively assess L2 proficiency. The "extreme group" approach was used to test the hypothesis regarding L2 proficiency level. Subjects were assigned to "high-", "mid-" and "low- proficiency" subgroups according to objective criteria. The lack of a Language Set \times L2 Proficiency level interaction would indicate that the magnitude of a cross-language shift in perception does not vary as a function of L2 proficiency.

2. Method

2.1. Subjects

We attempted to select a relatively homogeneous group of native Dutch adults differing principally according to overall ability to speak English. A total of 50 subjects between the ages of 20–35 years were chosen to participate. Half were male, half were female, and all had normal hearing according to self-report. Forty subjects were students in the Department of English at the Univer-

sity of Utrecht. Since most of them appeared to speak English quite well, 10 other subjects were recruited from a technical college in Delft. Like the 40 "English majors", the "Engineering majors" first began to study English at the age of 12 years. Unlike the English majors, however, they stopped studying English after the six years required in high school.

Responses to a language-background questionnaire confirmed our belief that the Engineering majors had much less need, opportunity, or desire to speak English than the English majors. The strong foreign accents evident among most of the Engineering majors suggested they had received much less exposure to English stops than the English majors. Since L2 learning may influence L1 production (J.E. Flege, 1987b; J.E. Flege et al. 1987a), the Engineering majors were expected to provide a better estimate of Dutch phonetic norms than the English majors.¹¹

Data was also obtained from five adult native speakers of British English (4 males, 1 female) to provide a tentative estimate of English phonetic norms. It should be born in mind, however, that the native English subjects were living in The Netherlands at the time of the study, and could all speak Dutch to some extent. Although English was their principal language at the time of the study, learning Dutch may have influenced their production and perception of /t/ in English (J.E. Flege, 1987b). The speech of the native English subjects was also used to provide a benchmark in the procedure developed to assess degree of global foreign accent in the English spoken by the 50 Dutch subjects (see below).

2.2. Procedure

The experiment was divided into two approximately 20 min parts, referred to here as "language sets". All speech material and recorded instructions were in English in the "English set". The

¹¹ It would have been difficult to establish a matched group of Dutch subjects who were truly monolingual. English is spoken well and readily by most college-educated Dutch subjects in the 20–35 year age range because English is often regarded as a prerequisite to career advancement in many fields. Moreover, English can be heard routinely on Dutch radio and TV.

speech material and translation-equivalent instructions were, of course, in Dutch in the Dutch set. The order of participation in the two language sets was counterbalanced across subjects. The English instructions were recorded by a female native speaker of English, the Dutch instructions by a female native speaker of Dutch. Recorded instructions were used to reduce the possibility of experimenter bias, as well as possible psycho-social or sociolinguistic effects. The experimenter (WE) was a native speaker of Dutch who speaks English with a moderate degree of foreign accent, in the first author's opinion. The experimenter interacted minimally with the subjects. The interactions which, of necessity, did occur were usually in English or Dutch, as appropriate.

The subjects first read some speech material from written lists. As shown in Appendix I, the speech material used in the Dutch and English sets was carefully balanced. In the first part of each set, subjects read 5 short sentences, 3 of which were later used for evaluating foreign accent. This was followed by the reading of 72 sentences (to be analyzed in future studies), then by 30 sentences of the form "The word is *CVC*." One of the words produced in the last set was used in the VOT analysis, to be described below. The recordings were made on a portable cassette tape recorder (Sony Model TCD6C) with an electret condenser microphone (Sony Model ECM-939LT).

The subjects then identified the 16 members of a VOT continuum ranging from /da/ to /ta/. The members of this continuum, which was synthesized using a version of the Klatt (1981) software synthesizer, resembled careful, isolated tokens of "da" and "ta" produced by a native speaker of English (JEF). The stimuli have been described in detail elsewhere (J.E. Flege et al. 1986). Briefly, VOT was increased in 10 ms steps from -60 to +90 ms. All 16 stimuli contained the same low-intensity 10 ms release burst. The relatively long (ca. 100 ms) duration of the formant transitions into the "vowel" portion of the stimuli were expected to yield category boundaries for the native English subjects at somewhat longer VOT values than those obtained in previous studies using other stimulus continua. Since the /a/ of English and Dutch differ little in quality, it is likely that

only the initial stop would seem "non-Dutch" to the Dutch subjects, especially those stimuli beginning with long-lag VOT values of 40-90 ms.

The stimuli were presented binaurally over headphones (TDH-49) at 76 dB SPL(A) using a portable cassette tape recorder (Sony Model TCD6C). The interstimulus interval was fixed at 2.5 s, with a 5 s pause between 10 different randomizations of the stimuli. The subjects were told to circle "da" or "ta" on the answer sheet, and to guess in case of uncertainty.

The instructions were written in such a way as to suggest that two different continua would be presented in the two language sets. The instructions included the following:

... Speech sounds may be produced differently in Dutch and English ... (which has) ... an important practical significance. When we program computers to produce speech, we must synthesize speech sounds differently for Dutch and English listeners. Our goal is to improve methods for synthesizing speech. We have asked you to participate because you are bilingual. People like you should have a good understanding of the pronunciation of both Dutch and English ...

Remarks made by some subjects indicated they had indeed inferred (incorrectly) that physically different stimuli were used in the Dutch and English sets.

We did not append a Dutch or English carrier phrase to the CV stimuli presented in the two sets because we felt that a carrier phrase might, after several repetitions, be ignored. Instead, the subjects were required to *actively process* Dutch and English speech in the two sets. The sentences presented in Appendix II were interspersed between blocks of stimuli in the Dutch and English sets. The subjects were informed they would hear the same questions in the Dutch and English sets, and that their written responses to the questions would be compared to verify they could "understand both languages". The advantage of this approach over the use of a carrier phrase was that it could be objectively determined whether the sentences were actually processed. A check verified that, in fact, the subjects gave the same answers to the questions asked in Dutch and English.

2.3. Assessing L2 proficiency

The design called for an objective evaluation of the English language proficiency of the Dutch subjects. At present, no suitable instrument exist for evaluating L2 proficiency as it relates to phonetic perception and segmental articulation. Therefore, three separate factors thought to be relevant to how the subjects might identify stops in L1 and L2 were used to evaluate L2 proficiency.

The most likely index of L2 proficiency seemed to be a rating of overall degree of foreign accent (J.E. Flege, 1987a). (The "fluency" ratings obtained by Elman et al. in 1977 may, in fact, have represented an impressionistic evaluation of global foreign accent). Previous studies have established that degree of foreign accent can be estimated validly and reliably (E. Brennan et al., 1981; 1975; R.C. Major, 1987; E. Ryan et al., 1975). A foreign accent score was obtained for all 50 Dutch subjects based on their production of the sentences "I can read this for you", "The good shoe fits Sue", and "The red book was good". Eight native speakers of British English rated the 3 sentences produced by 50 Dutch subjects and 5 native speakers of British English. British rather than American listeners were chosen because British English is taught in Dutch schools.

The 165 foreign accent sentences (50 Dutch + 5 English subjects \times 3 sentences) were low-pass filtered at 8 kHz, digitized at 20 kHz with 12-bit amplitude resolution, edited, normalized for overall rms intensity, and stored on disk for later on-line presentation. Three repetitions of each were presented binaurally (TDH-49) to the judges at a peak syllable level of 76 dB SPL(A) after D/A conversion and low-pass filtering. The experiment was self-paced: each sentence was presented 1.0 s after a response was received for the preceding sentence.

The judges were told to estimate degree of foreign accent by moving the lever on a response box between scale endpoints marked "No Foreign Accent" and "Strong Foreign Accent". The lever was attached to a linear potentiometer which, in turn, was connected to an A/D converter that returned values ranging from 1–256. The listeners were told that some talkers were native speakers

of English, but not what proportion (11%). The three sentences were presented in separate blocks, the order of which was counterbalanced across listeners. A foreign accent score was calculated by averaging the 72 rating scale judgments (8 listeners \times 3 sentences \times 3 repetitions) obtained for each talker.

The second index of L2 proficiency was the Dutch subjects' own estimate of their ability to pronounce English. Previous research suggests that L2 learners' self-reports concerning how much they use an L2 may be valid (K. MacKain et al., 1981), and that L2 learners are aware, at least to some extent, of their own foreign accent (Neufeld, 1979). A 10 point self-evaluation score was created. It was based on written responses to 4 questions presented aurally between blocks of the identification experiment in the English set (3,7,12,18 in Appendix II).¹²

The third index of L2 proficiency was a specific measure of segmental articulation. Previous research suggests a relationship exists between how authentically specific L2 phones are produced, and overall degree of foreign accent (E. Gathbonton, 1975; R.C. Major, 1987; E. Ryan et al., 1975). However, this does not rule out the possibility that an L2 learner might have a good overall accent in L2 but fail to produce some subset of sounds—such as prevocalic stops—authentically. It seemed reasonable to think that the extent to which Dutch speakers approximate the long-lag VOT norm of English /p,t,k/ in speech production is related to how they perceive English (and, possibly, Dutch) stops. Therefore, the mean VOT value produced by the Dutch subjects in prevocalic English /t/ (in the word-initial position of "tot", in the carrier "the word is _____") was determined. Three tokens of /t/ were measured to the nearest 0.5 mm (2.0 ms) from spectrograms (Kay Model 7800). The VOT interval was defined as extending from the beginning of the release burst, to the onset of periodicity in the "vowel" portion.

¹² Responses to the four questions were converted to a 10 point scale by subtracting 2 from the total number of points accorded each subjects based on their responses.

2.4. Analyses

A category boundary was calculated by linear interpolation of the 50% cross-over in the identification functions obtained for all 50 native Dutch subjects in the English and Dutch sets. However, not all of the boundary scores were used in the mixed-design Language Set \times L2 Proficiency Level ANOVAs used to test the hypothesis that the magnitude of perceptual switching increases as a function of L2 proficiency. We reasoned that if the original population of 50 subjects were simply split into halves based on any of the three criteria, there might be overlap in terms of actual L2 proficiency. This would tend to obscure the existence of a significant interaction, if indeed one existed, so a conservative approach to data analysis was adopted. The category boundary scores for subjects in three subgroups of 10 subjects each, designated that "high-", "mid-", and "low-proficiency" groups, were used in analyses. Three separate ANOVAs were performed, each based on scores for 30 subjects selected on the basis of one of the three proficiency criteria.

Most of the 10 Engineering majors were assigned to the "low-proficiency" subgroups determined by rankings for the three proficiency criteria. It should be noted, however, that the same subsets of subjects were not always assigned to the three subgroups; and that the "Engineering major" subgroup is not synonymous with the "low-proficiency" subgroups. Since we had no a priori way of knowing which of the three criteria was the most valid index of L2 proficiency, it was decided that a significant two-way interaction in *any one* of the three analyses would be interpreted as support for the hypothesis that cross-language switching increases along with L2 proficiency. Similarly, a significant effect of Language Set in any of the three ANOVAs would be regarded as support for the hypothesis that the listeners identified stops differently in the two language sets.

The criteria used to assess the L2 proficiency were themselves of interest. They are therefore described, and the inter-relationships among them explored using correlational techniques, following presentation of the speech perception data.

3. Results

3.1. Identification of Stops

Figure 1 plots the mean category boundaries obtained in the English and Dutch sets for subjects assigned to three extreme groups differing in English language proficiency on the basis of global foreign accent scores. The category boundaries occurred at slightly longer values in the English than Dutch set for all three proficiency subgroups. The mean difference averaged 1.2 ms for the 10 subjects with the best accent in English (i.e., the high-proficiency subjects), 3.5 ms for the mid-proficiency subjects, and 3.9 ms for subjects with the strongest accent (i.e., the low-proficiency subjects).

Highly similar results were also obtained when subjects were assigned to high-, mid-, and low-proficiency subgroups according to the self-evaluation scores, and according to English VOT. For subjects selected on the basis of the self-evaluation scores, the difference in category boundaries averaged 3.6 ms for the high-proficiency subjects, 1.4 ms for the mid-proficiency subjects, and 3.4 ms for subjects assigned to the low-proficiency subgroup. For subjects selected on the basis of mean VOT, the difference averaged 2.9 ms for the high-proficiency subjects, 1.6 ms for the mid-proficiency subjects, and 4.6 ms for the low-proficiency subjects.

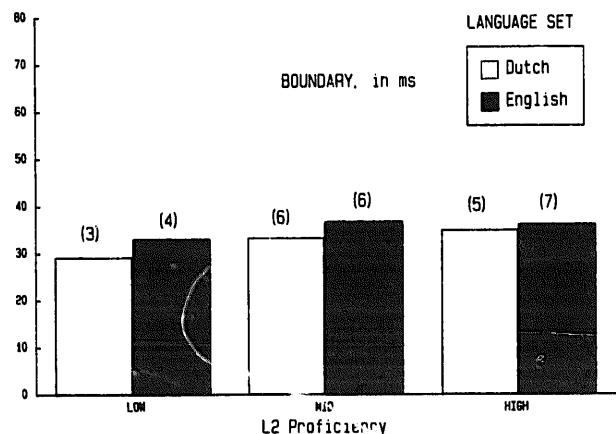


Fig. 1. The mean /d/-/t/ phoneme boundaries, in ms, observed in the Dutch and English sets for Dutch subjects assigned to high-, mid- and low-proficiency groups (10 each) according to the foreign accent scores (see text).

Table 1

The mean /d/-/t/ boundaries obtained in Dutch and English sets for 5 native speakers of English, 50 native speakers of Dutch, and two subsets of the native Dutch population (students majoring in English or Engineering), in ms. The "difference" score is the mean of individual subject between-set differences; inter-subject standard deviations are in parentheses

		Language Set		
		Dutch	English	Difference
Native English	(<i>n</i> = 5)	43.9(4.7)	47.1(16.4)	3.2(12.5)
Native Dutch	(<i>n</i> = 50)	33.5(5.7)	35.6(5.4)	2.1(3.8)
English majors	(<i>n</i> = 40)	34.5(5.8)	36.1(5.7)	1.6(3.8)
Engineering majors	(<i>n</i> = 10)	29.6(2.7)	33.7(3.9)	4.1(3.4)

The main effect of Language Set was highly significant in the analysis examining subjects on the basis of foreign accent scores [$F(1,27) = 19.5$], self-evaluation scores [$F(1,27) = 19.8$], and mean English VOT [$F(1,27) = 23.0$, $p < 0.001$]. Neither the effect of L2 Proficiency Level, nor the interaction of Language Set \times L2 Proficiency Level, was significant in any analysis.

The mean /d/-/t/ category boundaries obtained for all 50 Dutch subjects in the English and Dutch sets, and for the 5 native English subjects, are presented in Table 1. The mean category boundary for the 50 Dutch subjects averaged 2.1 ms longer in the English than Dutch set. A two-tailed *t*-test indicated that the order in which the Dutch subjects participated in the two language sets did not significantly affect the location of category boundaries.

Surprisingly, the difference in boundaries between language sets was smaller for the 40 English majors (1.6 ms) than the 10 Engineering majors (4.1 ms). The native English subjects also showed a larger effect of Language Set (3.2 ms) than the English majors. It was our impression that the English majors spoke English better than the En-

gineering majors, and that they also spoke their L2 (English) better than the native English subjects spoke their L2 (Dutch). These findings therefore appear to be inconsistent with the hypothesis that perceptual switching increases along with L2 proficiency.

3.2. L2 proficiency measures

Table 2 presents the mean values of the three criteria used to evaluate L2 proficiency. As expected, the 50 native Dutch subjects were generally accorded lower foreign accent scores than the 5 native speakers of English. The native English subjects' scores averaged 239, which was very near the "No Foreign Accent" end of the scale. The native Dutch subjects' scores averaged 159. Only two Dutch subjects received a foreign accent score that was higher than the *lowest* score received by any native English speaker (i.e., 230).

In agreement with our subjective impressions (see above), the English majors received a higher mean score (178) than the Engineering majors (86). In fact, only one of the 40 English majors received a score that was lower than the *highest*

Table 2

The mean foreign accent scores, self-evaluation scores, and VOT values (in ms) in /t/ spoken in English and Dutch words by native speakers of English, 50 native speakers of Dutch, and two subgroups of the native Dutch population (students majoring in English and Engineering); inter-subject deviations are in parentheses

		Accent	Self-eval.	English VOT	Dutch VOT
Native English	(<i>n</i> = 5)	239(9)	-	90.4(23.6)	51.2(23.5)
Native Dutch	(<i>n</i> = 50)	159(51)	5.0(1.9)	60.3(31.8)	22.8(7.2)
English majors	(<i>n</i> = 40)	178(36)	5.4(1.8)	68.6(29.9)	23.3(7.6)
Engineering majors	(<i>n</i> = 10)	86(29)	3.3(1.5)	26.9(7.8)	20.9(5.3)

score received by an Engineering major (i.e., 130). The self-evaluation scores of the English majors were also higher (an average of 5.4) than the Engineering majors' (3.3 on a 10 point scale). The self-evaluation scores and foreign accent scores were positively correlated [$r(49) = 0.557$, $p < 0.01$]. This suggests that the 50 Dutch subjects' evaluation of their accent in English was, at least to some extent, valid.

As expected from previous research (J.E. Flege, 1987a; J.E. Flege et al., 1981), the Dutch subjects produced /t/ with a shorter mean VOT value in English (60 ms) than the native speakers of English (90 ms). Also as expected, the 40 English majors produced English /t/ with a longer—and therefore more English-like—mean VOT value (67 ms) than the 10 Engineering majors (27 ms). Figure 2 plots the mean VOT values in English /t/ observed for all 50 Dutch subjects as a function of the foreign accent scores. A significant positive correlation existed between these two indices of English-language proficiency [$r(49) = 0.341$, $p < 0.01$]. Close inspection of Fig. 2 reveals that the significance of the correlation was due largely to the fact that a small subset of subjects with low foreign accent scores (i.e., "strong" accents) produced English /t/ with short-lag VOT values. Most of these subjects were Engineering majors. They seem to have produced English words with a "Dutch" /t/.

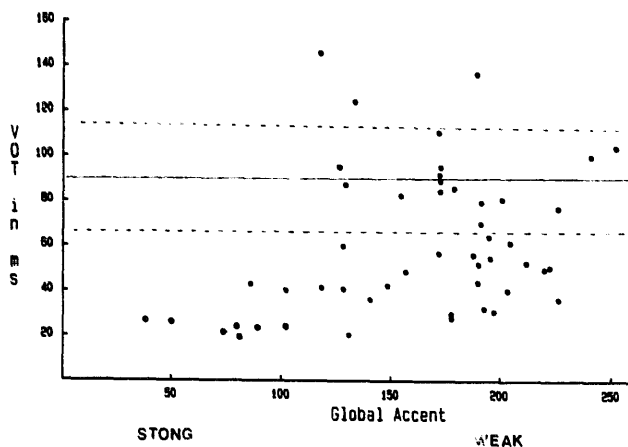


Fig. 2. The mean VOT in ms in the /t/ produced in English words by 50 native Dutch subjects as a function of the foreign accent scores. The mean VOT measured in stops produced by native English subjects is represented by a solid horizontal line; the dashed lines enclose +/- one standard deviations from the mean native English VOT.

Figure 2 also reveals that some of the native Dutch subjects produced English /t/ with English-like VOT values. The solid line in Fig. 2 represents the mean VOT value produced in English /t/ by the native speakers of English. The dashed lines enclose +/- one standard deviation around the native English mean. While it is true that most (64%) Dutch subjects' mean VOT was more than one standard deviation below the native English mean, some subjects' (30%) mean values were less than +/- one standard deviation from the English mean. A few (6%) of the Dutch subjects actually produced English /t/ with mean VOT values that exceeded the native English mean by more than one standard deviation.

Table 2 also presents the mean VOT values observed for Dutch /t/. Almost all (48/50) of the native Dutch subjects, and all 5 native English subjects, produced /t/ with longer VOT values in English than Dutch. As expected, the 5 native English subjects produced Dutch /t/ with a substantially longer mean VOT (51 ms) than the 50 native speakers of Dutch (23 ms). The 40 English majors produced Dutch /t/ with somewhat longer—and therefore English-like—VOT values than the Engineering majors. However, the difference between the English majors and the Engineering majors was non-significant [$F(1,48) = 0.84$].

Figure 3 plots the mean VOT of /t/ in English and Dutch words spoken by the Dutch subjects

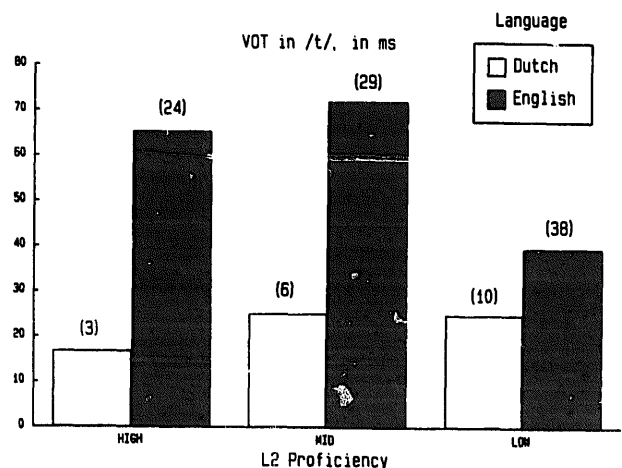


Fig. 3. The mean VOT in ms in English /t/ produced by native Dutch subjects assigned to high-, and low-proficiency groups (10 each) according to the foreign accent scores.

who were assigned to the high-, mid-, and low-proficiency subgroups according to the foreign accent scores. The subjects in all three proficiency subgroups produced /t/ with longer VOT values in English than Dutch. However, the VOT difference between English and Dutch /t/ was greater for subjects in the high-proficiency subgroup (48 ms) and the mid-proficiency subgroup (47 ms) than for subjects assigned to the low-proficiency subgroup (15 ms).

The differing magnitude of the VOT difference between Dutch versus English /t/ for subjects differing in L2 proficiency led to a significant interaction of Language Set \times L2 Proficiency Level [$F(2,27) = 4.67, p = 0.02$]. The simple main effect of Language Set was significant for the high-proficiency [$F(1,9) = 42.1$] and mid-proficiency subjects [$F(1,9) = 23.2, p < 0.01$], but not for the low-proficiency subjects. The simple main effect of Proficiency Level was significant for the VOT in Dutch /t/ [$F(2,27) = 4.34, p = 0.02$]: it narrowly missed reaching significance for English /t/ [$F(2,27) = 3.06, p = 0.06$]. Post-hoc tests for Dutch /t/ revealed that the high-proficiency subjects produced significantly *shorter* VOT values (17 ms) than subjects in the mid- (25 ms) or low-proficiency (25 ms) subgroups ($p < 0.05$).¹³

T-tests (two-tailed) revealed there was not a significant difference in VOT for Dutch /t/ between subjects participating in the Dutch and English sets first. However, subjects who participated in the Dutch set first produced *English* /t/ with a significantly longer mean VOT value (70 ms) than those who participated in the English set first (51 ms, $t(48) = 2.19, p < 0.05$).

Further tests were performed to determine whether the order in which subjects participated in the two language sets may have differentially affected subjects differing in L2 proficiency. The 25 subjects who participated in the same order (of language set) were subdivided into two extreme groups of 10 subjects each based on the foreign accent scores. Two separate ANOVAs revealed that there was not a significantly different

effect of the order of participation on subjects differing in L2 proficiency.

A multiple correlation analysis was performed to determine what relationship, if any, existed between the category boundary values obtained for the 50 Dutch subjects and the three criteria used to assess the Dutch subjects' L2 proficiency. The foreign accent scores, self-evaluation scores, and English VOT did not account for a significant portion of the variance in the category boundaries obtained in the English set. Nor was there a significant simple correlation between any single criteria and the category boundaries observed in the English set.

A multiple correlation analysis was also performed to examine the relationship between the L2 proficiency variables and the VOT values produced by the Dutch subjects in English /t/. The foreign accent and self-evaluation scores together accounted for a significant 34% of the variance in VOT values /t/ ($p < 0.01$). The foreign accent scores and self-evaluation scores were positively correlated with one another [$r(49) = 0.557, p < 0.01$], suggesting they may have reflected a single dimension. As mentioned earlier, there was a significant correlation between English VOT and foreign accent scores. However, there was not a significant simple correlation between English VOT and the self-evaluation scores.

Additional tests indicated there was not a significant correlation between the VOT in English /t/ and the category boundary values obtained in the Dutch set. There was not a significant correlation between the magnitude of the VOT difference produced between /t/ in English and Dutch and the three L2 proficiency variables. Finally, there was not a significant correlation between category boundaries obtained in the Dutch and English sets.

4. Discussion

Native speakers of Dutch identified a voice onset time (VOT) continuum ranging from /da/ to /ta/ in two language "sets" intended to induce them to process stimuli *as if* they were listening to Dutch or English. The language set manipulation exerted a highly significant effect on the loca-

¹³ Only 8 of the 10 subjects assigned to the low-proficiency group according to the foreign accent scores were Engineering majors. The mean VOT values in Fig. 3 therefore differ somewhat from the values in Table 2.

tion of category boundaries. Boundaries were longer in the English than Dutch set, which is consistent with the fact that /t/ is realized with much longer VOT values in English than Dutch (L. Lisker et al., 1964). This suggests that the Dutch subjects, who were able to speak English, were tacitly aware of acoustic differences distinguishing the long-lag /t/ of English from the short-lag /t/ of Dutch.

The present study may have showed a significant effect of language set, unlike two previous studies (A. Caramazza et al., 1973; L. Williams, 1977), because the procedures used to create differing language sets were more elaborate, and thus probably more effective (see Methods section). In addition, the synthetic consonant-vowel stimuli used here were probably more natural-sounding than the stimuli used in the previous studies.

The absolute magnitude of the language set effect on category boundaries was small in absolute terms: 2.1 ms. The largest effect of language set obtained for any Dutch subject was only 11 ms. Interestingly, the subject showing this relatively large effect was one of two Dutch subjects whose global foreign accent score fell within the range of values observed for native speakers of English (That is, this was a subject who could probably "pass" as a native speaker of English in some circumstances.)

The effect of language set was undoubtedly smaller than the difference in boundaries that would be obtained for Dutch and English monolinguals identifying the same continuum. We saw a 17 ms difference in the boundaries obtained for Dutch Engineering students in the Dutch set, and native English speakers in the English set.¹⁴ An earlier study (J.E. Flege et al., 1986) using the same stimuli showed a 14 ms difference between monolingual English and Spanish subjects.

¹⁴ These two mean values may actually underestimate the difference that would be obtained from groups of Dutch and English monolinguals. The Dutch L2 experience of the British subjects could have led to a reduction in values of their English boundary; and the English L2 experience of the engineering students might have led to an increase in the VOT values obtained in the Dutch set.

An earlier finding by Elman et al. (1977) suggested that not all individuals who learn a second language (L2) will show the same effect of language set. More specifically, it suggested the hypothesis that Dutch subjects proficient in English would show a larger difference in the location of category boundaries in a Dutch versus English set than relatively non-proficient Dutch subjects. This was tested by examining 50 Dutch subjects representing a wide range of proficiency levels in English. Some subjects spoke English frequently, and with little accent; others spoke English only infrequently, and with a strong foreign accent. Three objective criteria were used to assign subjects from this population to subgroups differing in L2 proficiency.

The most important finding of this study was that the magnitude of the boundary shift between language sets did *not* vary as a function of L2 proficiency. The same result was obtained when subjects were assigned to L2 proficiency subgroups according to global foreign accent scores, their own self-evaluation concerning ability to pronounce English, and the mean VOT values produced in English /t/.

The analysis of perception in two language sets was followed by an examination of the criteria used to assess L2 proficiency. The foreign accent scores were correlated with the self-evaluation scores, and with the mean VOT produced in English /t/ by the native Dutch subjects. The first finding indicates that L2 learners' judgments of their own foreign accent (G. Neufeld, 1979) is valid, at least to some extent. The second finding is in close agreement with Major's (1987) report of a significant correlation between native Portuguese subjects' English VOT and global foreign accent scores. Other studies have also shown significant correlations between specific aspects of L2 segmental articulation and degree of accent (E. Gatbonton, 1975; R.C. Major, 1987; E. Ryan et al., 1975). It is likely that divergences from L2 phonetic norms for many segmental and suprasegmental dimensions contribute to foreign accent (J.E. Flege, 1984; 1987a). This raises the question of whether approximation to the phonetic norm for *all phones* in L2 will lead to equal changes in degree of perceived foreign accent (J.E. Flege,

1984).¹⁵

The cross-language difference in VOT seen in speech production was an order of magnitude larger than the category boundary shift between /d/ and /t/ observed in the English and Dutch sets. The Dutch subjects' /t/ had a mean VOT value that was 38 ms longer in English than Dutch words. Conversely, the native English subjects' /t/ had a mean VOT that was 38 ms *shorter* in Dutch than English. Despite their ability to alter VOT when switching from Dutch to English, the 50 native Dutch subjects, taken as a whole, produced English /t/ with a shorter mean VOT (60 ms) than the native speakers of English (90 ms). This agrees with previous studies of L2 speech production (A. Caramazza et al., 1973; J.E. Flege et al., 1987a; 1984; 1981; R.C. Major, 1987; R.F. Port et al., 1983) showing that L2 learners usually fail to match native speakers in producing /p,t,k/ if these stop phonemes are implemented as voiceless unaspirated stops in the learners' native language (L1).

On the surface, at least, this seems to support the claim (J.E. Flege, 1981; 1984; 1987a,b) that adult L2 learners may approximate L2 phonetic norms, but never match native speakers in producing L2 phones which differ acoustically from corresponding stops in L1. An inspection of individual subject data, however, showed that while most of the Dutch subjects did produce English /t/ with substantially shorter (and therefore Dutch-like) VOT values than the native English speakers, some of them produced English /t/ with VOT values closely resembling the native English mean. Does this mean that they had acquired an truly English-like production of /p,t,k/ ?

¹⁵ Another question raised by the relationship between foreign accent and VOT is whether L2 learners' approximation to the long-lag VOT norm of English will affect intelligibility. One study (D.E. Metz et al., 1985) showed the VOT difference between English /t/ versus /d/ produced by deaf speakers of English accounted for a significant 54% of the variability in intelligibility scores. The first author's impression is that a Dutch talker's production of English /t/ with short-lag VOT values will usually be perceived as /t/ by native English speakers. However, non-authentic VOT values might increase processing time so that, for example, it would take English subjects longer to identify Dutch-produced /t/ tokens than native-produced /t/ tokens in a speeded classification task.

We speculate that the English stop production observed here may not have been truly representative of how the Dutch subjects typically produced English /p,t,k/. VOT was measured in key words produced in the carrier phrase "The word is _____". This undoubtedly focused subjects' attention on the minimal contrast between "tot" and "dot" (J.E. Flege, 1984; J.E. Flege et al., 1984). The subjects who produced English /t/ with the most English-like VOT values were students majoring in English. They are likely to have received explicit instruction pertaining to the VOT difference distinguishing English versus Dutch stops. It is therefore possible they exaggerated VOT¹⁶ in producing English /t/. This hypothesis was supported by the observation that several Dutch subjects "overshot" the English VOT norm for /t/. It is further supported by the finding that subjects who participated in the Dutch set first produced English /t/ with significantly longer VOT values than subjects who first participated in the English set. It would be useful to examine production of English stops by a comparable population of subjects using a speech sample likely to result in "unmonitored" performance (J.E. Flege et al., 1984).

The production of Dutch /t/ was also examined. This analysis has lead us to an important inference that, at first, seems divergent from the perception results. We found that the Dutch subjects with the best accents in English produced Dutch /t/ with significantly *shorter* VOT values than subjects with stronger accents in English. This suggests that learning English as an L2 influenced how the proficient Dutch speakers of English produced stops in their *native language*.

This is not the first time an effect of learning an L2 on production of the native language has been observed. Flege and Eefting (1987a) obtained a similar result for native Spanish speakers of English. The subjects in two adult groups and one child group realized the short-lag /p,t,k/ of

¹⁶ It is unlikely, of course, that talkers directly manipulate the acoustic VOT dimension in stops. VOT variation in /p,t,k/ is one consequence of changes in the laryngeal devoicing gesture used for voiceless stops or, more importantly, of changes in the relative timing of glottal and supraglottal gestures.

Spanish with significantly shorter VOT values than groups of age-matched native Spanish subjects who had not learned English. Flege and Eefting (1987a) interpreted this "L2 effect on L1 production" as evidence for the establishment of new phonetic categories. This interpretation was supported by the finding (J.E. Flege et al., 1987b) that, whereas the Spanish speakers of English imitated a VOT continuum ranging from -60 ms to +90 ms with stops in the three modal VOT ranges (lead, short-lag, long-lag), Spanish and English *monolinguals* used only two modal VOT ranges in imitating the same continuum.

Dutch learners of English might be expected to shorten the VOT in Dutch /t/ if they established a long-lag phonetic category for English /t/. Keating (1984) hypothesized that the voiceless unaspirated realizations of Spanish /p,t,k/ have somewhat longer VOT values than the voiceless aspirated realizations of English /b,d,g/ because of "polarization", a universal phonetic factor which serves to enhance the acoustic differences between phonetic categories. Whereas the short-lag stops used to implement Spanish voiceless /p,t,k/ "polarize" away from the prevoiced stops used to implement voiced /b,d,g/ in Spanish, the short-lag realizations of voiced /b,d,g/ in English polarize away from the *long-lag* stops used in English to implement voiceless /p,t,k/. If the proficient Dutch subjects added a phonetic category for English /t/, their Dutch /t/ would be expected to polarize from it, that is, be produced with somewhat shorter VOT values. This conclusion must, of course, remain tentative in the absence of data for Dutch monolinguals.

Flege (1987b) observed significant L2 effects on L1 production that went in the opposite direction. Proficient native French speakers of English produced French /t/ with longer—and therefore English-like—VOT values than French monolinguals. Conversely, native English speakers of French produced English /t/ with shorter—French-like—VOT values than English monolinguals. This suggested a *merger* (or assimilation) of the phonetic properties of French and English /t/, rather than an enhancement (polarization) of the differences between them.

The direction of the L2 effect on L1 production observed by Flege (1987b) may have been due to

the chronological *age* at which subjects in that study began learning their L2. Whereas they began learning their L2 (French or English) as young adults, the subjects examined by Flege and Eefting (1987a) began learning their L2 (Spanish) by the age of 5–6 years. Flege and Eefting (1987b) hypothesized that young children establish new phonetic categories for L2 phones more readily than older children and adults because their phonetic systems are still in the process of development.

The Dutch subjects examined in the present study were intermediate in age to the subjects examined by Flege (1987b) and Flege and Eefting (1987a). They first began to study English in school at the age of 12 years. One might speculate that individuals who first begin L2 learning by the age of 12 years will, like younger children, establish a new phonetic category for English /t/. However, this conclusion is premature for two reasons. We have no information concerning exactly when the Dutch subjects were first passively exposed to English.¹⁷ Also, the subjects examined here constituted a special population. The Dutch subjects who showed an apparent shortening of VOT in Dutch /t/ were the subjects with the greatest aptitude for pronunciation selected from a group which, as a whole, probably possessed relatively great aptitude for L2 learning.

If, as suggested by the production data, the proficient Dutch speakers of English developed a phonetic category for English /t/, we need to ask why the effect of establishing a category was not manifested by substantial differences in boundaries between language sets. That the Dutch subjects' category boundaries averaged 34 ms strongly suggests they did not identify stimuli based on the presence/absence of a detectable temporal order discontinuity between F_1 onset and the onset of F_2 and higher formants. This value does not diverge greatly from the mean value obtained for a group of monolingual native English subjects (43 ms) in a previous study (J.E. Flege et al., 1986).

¹⁷ It should be noted that the Dutch L2 learners examined in this study probably heard their L2 (English) routinely on radio and TV. The same could not be said for the subjects examined by Flege (1987b) and Flege and Eefting (1987a).

We speculate that the Dutch subjects did establish a new phonetic category for English /t/, and made use of it when identifying stimuli in both the Dutch and English sets. The stimuli were modeled on the *English* phonetic contrast between /da/–/ta/; and the continuum contained stops with long-lag VOT values, which do not occur on the phonetic surface of Dutch. Since the subjects were required to divide a continuum which contained both prevoiced and long-lag stops into two categories, the task may have been facilitated by the use of a long-lag category for identifying /t/ tokens.

This hypothesis is consistent with research examining the voiced-voiceless contrast in other languages which, like Dutch, implement /p,t,k/ as voiceless unaspirated stops. Williams' (1977) post-hoc acoustic analysis of Spanish stops suggested to her that native speakers of Spanish give greater weight to prevoicing as a cue to voicedness than native English speakers; and give greater weight to the presence of an audible release burst, and to the absence of low-frequency energy immediately following the burst, as cues to voicelessness.¹⁸ Serniclaes (1976) found that, for stops without prevoicing, French listeners showed a rapid increase in /t/ judgments as VOT was increased in small increments beyond 0 ms VOT. However, the effect of increasing VOT on percent voiceless judgments reached asymptote at a much shorter value (ca. 8 ms) than it did in the present study (about 40 ms).

This leads to the hypothesis that the VOT dimension is used in a qualitatively, rather than just quantitatively, different manner in Dutch and English. That is, the relative weight, as well as the absolute parameter values for VOT, might be specified differently in the phonetic category prototypes established for /t/ by native speakers of Dutch and English.

If so, the Dutch subjects may not have invoked the Dutch prototype for /t/ when identifying stimuli in this study because the stimuli—especially those with long-lag VOT values—differed

too much from the Dutch prototype. This should be examined in future research by testing subjects with two different VOT continua, one derived from natural tokens of Dutch /t/, the other based on naturally produced English /t/ tokens. It would also be useful to vary the range of VOT values in the stimulus continua.

Lieberman and Mattingly (1985) proposed that a distal object of perception in stops is the timing of laryngeal and supralaryngeal gestures (or intended gestures). This suggests that individuals who succeed in producing /t/ with substantial differences in L1 versus L2 also perceive the difference between them. The results obtained here for L2 learners agree closely with Caramazza et al. (1973) in showing a disparity between production and perception: the cross-language shift in production was far greater than the one observed in speech perception.¹⁹ One possible interpretation of the VOT difference between Dutch and English /t/ produced by the Dutch subjects is that they used different realization rules to output a single phonetic category. Such a mechanism is needed to explain how, for example, monolinguals produce vowels differently as a function of social context (W. Labov, 1981). However, the great magnitude of the cross-language switch in VOT production suggests a different explanation, namely that the Dutch subjects implemented /t/ using different phonetic categories when reading Dutch and English word lists.

4.1. Conclusions

We found that Dutch speakers of English produced a substantial VOT difference between the /t/ in Dutch and English, but that even those who were highly proficient in English showed only a small (albeit significant) shift in the location of phoneme boundaries when identifying stops in a

¹⁸ Williams further speculated that the rate at which energy increases at syllable onset may be a relatively more important as an acoustic cue to the voicing feature in prevocalic stops for native speakers of Spanish than English.

¹⁹ Perhaps the VOT values measured in speech production and the category boundaries obtained in forced-choice identification should not be compared directly. Elsendoorn (1981, 1984) found that Dutch high school students showed relatively little effect of vowel duration in identifying word-final English stops. However, they showed a native-like effect when asked to *adjust* vowel durations to "prototypical" values in the context of voiced versus voiceless word-final stops.

Dutch versus English set. The significant effect of the language set manipulation demonstrated the procedures used to create differing language sets were effective. The finding that highly proficient Dutch subjects produced Dutch /t/ with shorter VOT values than relatively non-proficient Dutch subjects was interpreted as evidence they had established a phonetic category for English /t/. We speculated that the Dutch subjects made use of their English phonetic category for /t/, but not their Dutch /t/ category, in judging the synthetic stimuli because the stimuli differed greatly from the subjects' prototypes for Dutch /t/.

Appendix 1

Materials used to elicit speech production in the English and Dutch sets

English set

- E1. a. I can read this for you. (2 × each)
 b. A big man sat here
 c. The good shoe fits Sue
 d. The red book was good.
 e. A brown cow ate hay.
- E2. Seep through a CVC (18 ×)
 Sheep through a CVC (18 ×)
 Sip through a CVC (18 ×)
 Ship through a CVC (18 ×)
 CVC = heat, hoot, hit, hat, hot, hut.
- E3. The word is CVC (30 ×)
 CVC = bought, dot, cot, pot, tot, got.

Dutch set

- D1. a. Ik kan niets voor je doen (2 × each)
 b. Het boek is erg goed.
 c. Het rood is erg mooi.
 d. De tand is niet rot.
 e. Een bruin kalf at hooi.
- D2. Tiep van een CVC (18 ×)
 Siep van een CVC (18 ×)
 Tip van een CVC (18 ×)
 Sip van een CVC (18 ×)
 CVC = hiet, hoed, het, hut, hit, had.

- D3. Het woord is CVC (30 ×)
 CVC = bod, dot, kot, pot, tot, mot.

Appendix 2

Questions interspersed between the 10 blocks of stimuli in the identification experiment in the English set (Dutch translation equivalents were used in the Dutch set); an asterisk marks questions used in forming the self-evaluation scores.

1. What is your age?
2. What is your name?
- *3. How well do you speak English? (1–3 pts).
4. Where were you born?
5. How many brothers do you have?
6. How many languages do you speak fluently?
- *7. Evaluate your own accent in English using a 1 to 5 rating scale where: 5 stand for native-like pronunciation, 1 for a strong Dutch accent, and numbers in between for an intermediate degree of accent.
9. Where was your mother born?
10. Where was your father born?
11. Have you ever been to England?
- *12. Do native speakers of English ever compliment you for your pronunciation of English? (if yes, 2 pts.)
13. Have you ever lived in an English-speaking country?
14. Do you often see films in English?
15. Do you ever speak English at home?
16. Do you speak English with your close friends?
17. Do you read books in English?
- *18. Have you ever been mistaken for a native speaker of English? (if yes, 2 pts.)
19. When did you first begin to study English?
20. How many years did you study English before you came to the University?
21. Do you ever dream in English?
22. Do you speak Dutch better than English?
23. Can your mother speak English?
24. Can your father speak English?
25. Can you speak and understand English over the telephone?
26. Can you understand English at a noisy

party?

27. Can you speak English when you are tired?

Acknowledgments

This study was funded by NIH grant NS20963 to the University of Alabama at Birmingham. The authors would like to thank Bruno Repp and Roy Major for commenting on an earlier draft of this article.

References

- A.S. Abramson and L. Liskers (1973), "Voice-timing perception in Spanish word-initial stops", *J. Phonetics*, Vol. 1, pp. 1-8.
- E. Brennan and J. Brennan (1981), "Accent scaling and language attitudes: Reactions to Mexican-American English speech", *Lang. Speech*, Vol. 24, pp. 207-221.
- E. Brennan, E. Ryan and W. Dawson (1975), "Scaling of apparent accentedness by magnitude estimation and sensory modality matching", *J. Psycholing. Res.*, Vol. 4, pp. 27-36.
- A. Caramazza, G. Yeni-Komshian, E. Zurif, E. and E. Carbone (1973), "The acquisition of a new phonological contrast: The case of stop consonants", *J. Acoust. Soc. Am.*, Vol. 54, pp. 421-428.
- S. Disner (1986), "On describing vowel quality", in: J. Okala and J. Jaeger, eds., *Experimental Phonology* (Academic Press, New York) pp. 69-79.
- J. Elman, R. Diehl and S. Buchwald (1977), "Perceptual switching in bilinguals", *J. Acoust. Soc. Am.*, Vol. 62, pp. 971-974.
- B.A. Elsendoorn (1981), "The influence of vowel duration in English CVC-words on the word-final perception of the fortis-lenis contrast by Dutch speakers of English", *PRIPU* (Institute of Phonetics, Utrecht), vol. 6, pp. 3-13.
- B.A. Elsendoorn (1984), "Production and perception of English vowel duration by Dutch speakers of English", in: M. van den Broecke and A. Cohen, eds., *Proc. Tenth Internat. Congress of Phonetic Sci.* (Foris, Dordrecht), pp. 673-676.
- J.E. Flege (1980), "Phonetic approximation in second language acquisition", *Lang. Learning*, Vol. 30, pp. 117-134.
- J.E. Flege (1981), "The phonological basis of foreign accent: A hypothesis", *TESOL Quarterly*, Vol. 15, pp. 443-455.
- J.E. Flege (1984), "The detection of French accent by American listeners", *J. Acoust. Soc. of Am.*, Vol. 76, pp. 692-707.
- J.E. Flege (1987a), "The production and perception of foreign languages," in: H. Winitz, ed., *Human Communication and Its Disorders*, Vol. III. (Ablex, Norwood, N.J.), in press.
- J.E. Flege (1987b), "The production of 'nəw' and 'similar' phones in a foreign language: Evidence for the effect of equivalence classification", *J. Phonetics*, Vol. 15, pp. 47-65.
- J.E. Flege and R.D. Davidian (1984), "Transfer and developmental processes in adult foreign language speech production", *Appl. Psycholing.*, Vol. 5, pp. 323-347.
- J.E. Flege and W. Eefting (1986), "Linguistic and developmental effects on the perception and production of stop consonants", *Phonetica*, Vol. 43, pp. 155-171.
- J.E. Flege and W. Eefting (1987a), "The production and perception of English stops by Spanish speakers of English", *J. Phonetics*, Vol. 15, pp. 67-83.
- J.E. Flege and W. Eefting (1987b), "Imitation of a VOT continuum by Spanish speakers of English bilinguals: Evidence for the emergence of a new phonetic category", *J. Acoust. Soc. Amer.*, to appear.
- J.E. Flege and J. Hillenbrand (1984), "Limits on pronunciation accuracy in adult foreign language speech production", *J. Acoust. Soc. Amer.*, Vol. 76, pp. 708-721.
- J.E. Flege and R.F. Port (1981), "Cross-language phonetic interference: Arabic to English", *Lang. Speech*, Vol. 24, pp. 125-146.
- E. Gatbonton (1975), *Systematic Variations in Second Language Speech: A Sociolinguistic Study*, Unpublished Ph.D. thesis, McGill University.
- J. Harris (1969), *Spanish Phonology*, (MIT Press, Cambridge, MA).
- P.A. Keating (1984), "Phonetic and phonological representation of stop consonant voicing", *Language*, Vol. 60, pp. 286-319.
- D. Klatt (1981), "Software for a cascade/parallel formant synthesizer", *J. Acoust. Soc. Am.*, Vol. 67, pp. 971-995.
- W. Labov (1981), "Resolving the neogrammarian controversy", *Language*, Vol. 57, pp. 267-309.
- A.M. Liberman and I. Mattingly (1985), "The motor theory of speech perception revised", *Cognition*, Vol. 21, pp. 1-36.
- P. Linell (1982), "The concept of phonological form and the activities of speech production and speech perception", *J. Phonetics*, Vol. 10, pp. 37-72.
- L. Lisker and A.S. Abramson (1964), "A cross-language study of voicing in initial stops: Acoustical measurements", *Word*, Vol. 20, pp. 384-422.
- M. Mack (1983), *Psycholinguistic consequences of early bilingualism: A comparative study of performance of English monolinguals and French-English bilinguals in Phonetic, Syntactic, and Semantic Experiments*, Unpublished Ph.D. thesis, Brown University, Providence, RI.
- K. MacKain, C. Best and W. Strange (1981), "Categorical perception of English /r/ and /l/ by Japanese bilinguals", *Appl. Psycholing.*, Vol. 2, pp. 369-390.
- M.A. Macken (1980), "Aspects of the acquisition of stop systems: A cross-linguistic perspective", in: G. Yeni-Komshian, J. Kavanagh and C. Ferguson, eds. *Child Phonology*, Vol. 1, Production (Academic Press, New York), pp. 143-168.
- R.C. Major (1987), "English voiceless stop production by speakers of Brazilian Portuguese", *J. Phonetics*, Vol. 15, to appear.

- B. McLaughlin (1978), *Second-language acquisition in childhood* (Erlbaum, Hillsdale, N.J.).
- D.E. Metz, V.J. Samar, N. Schiavetti, R.W. Sitler and R.L. Whitehead (1985), "Acoustic dimensions of hearing-impaired speakers' intelligibility", *J. Speech Hear. Res.*, Vol. 28, pp. 345-355.
- G. Neufeld (1979), "Towards a theory of language learning ability", *Lang. Learn.*, Vol. 29, pp. 227-241.
- R.F. Port and F. Mitleb (1983), "Segmental features and implementation in acquisition of English by Arabic speakers", *J. Phonetics*, Vol. 11, pp. 219-231.
- E. Ryan, M. Carranza and R. Moffie (1975), "Reactions toward varying degrees of accentedness in the speech of Spanish-English bilinguals", *Lang. Speech*, Vol. 20, pp. 267-273.
- S.F. Disner (1986), "On describing vowel quality," in: J. Ohala and J. Jaeger, eds., *Experimental Phonology*, (Academic Press, New York), pp. 69-80.
- W. Serniclaes, (1976), "Prevoisement et délai d'établissement du voisement: Deux indices indépendants pour la perception des occlusives," *Rapport d'Activités de l'Institut de Phonetique, Bruxelles*, Vol. 10, pp. 83-103.
- A. Slis and A. Cohen (1969), "On the complex regulating the voiceless-voiced distinction", *Lang. Speech*, Vol. 12, pp. 80-102, 137-155.
- K.S. Stevens (1983), "Design features of speech sound systems", in: P.F. MacNeilage, ed., *The Production of Speech* (Springer, New York) pp. 247-263.
- L. Williams (1977), "The voicing contrast in Spanish", *J. Phonetics*, Vol. 5, pp. 169-184.
- L. Williams (1977), "The perception of stop consonant voicing by Spanish-English bilinguals", *Percep. Psych.*, Vol. 21, pp. 289-297.
- L. Williams (1979), "The modification of speech perception and production in second-language learning," *Percep. Psych.*, Vol. 26, pp. 95-104.
- J. Werker, and R. Tees (1984), "Phonemic and phonetic factors in adult cross-language speech perception," *J. Acoust. Soc. Amer.*, Vol. 75, pp. 1858-1865.
- M. Zlatin, and R. Koenigsnecht (1973), "Development of the voicing contrast: A comparison of voice onset time in stop perception and production," *J. Speech Hear. Res.*, Vol. 19, pp. 93-111.