Effects of age of second-language learning on the production of English consonants

James Emil Flege a,*, Murray J. Munro a,1, Ian R.A. MacKay b

a Department of Biocommunication, VH 503, University of Alabama at Birmingham, Birmingham, AL 45294-0019, USA
b Department of Linguistics, University of Ottawa, Ottawa, Canada

Received 3 December 1993; revised 8 August 1994

Abstract

This study examined the production of English consonants by native speakers of Italian. The 240 adult native Italian speakers of English who participated had begun learning English when they emigrated to Canada between the ages of 2 and 23 years. Word-initial, word-medial and word-final tokens of English stops and fricatives were assessed through forced-choice judgments made by native English-speaking listeners, and acoustically. The native Italian subjects' ages of learning (AOL) English exerted a systematic effect on their production of English consonants even though they had lived in Canada for an average of 32 years, and reported speaking English more than Italian. In all but two instances, one or more native Italian subgroup defined on the basis of AOL differed significantly from subjects in a native English (NE) control group. The AOL of the first native Italian subgroup to differ from the NE subjects varied across consonant and syllable position. The results are discussed in terms of hypotheses proposed in the literature concerning the basis of segmental errors in L2 speech production.

Zusammenfassung


* Corresponding author.
1 Now at Department of Linguistics, Simon Fraser University.

0167-6393/95/$09.50 © 1995 Elsevier Science B.V. All rights reserved
SSDI 0167-6393(94)00044-1

Résumé


Dans tous les cas sauf deux, au moins un des sous-groupes constitué sur la base de l’AOL a obtenu des résultats significativement différents de ceux du groupe de contrôle (NE) formé de sujets de langue maternelle anglaise. Pour ce qui concerne l’identification des consonnes et des frontières syllabiques, l’AOL à partir duquel on note une différence entre le premier sous-groupe de sujets de langue maternelle italienne et les sujets du groupe de contrôle (NE) varie. Les résultats sont discutés par rapport aux hypothèses proposées dans la littérature concernant les erreurs segmentales liées à la production d’une seconde langue.

Keywords: Speech production; Second-language; Consonants; Bilingualism; Phonetic interference; Foreign accent; Critical period; Italian

1. Introduction

Research examining overall degree of foreign accent in the production of a second language (L2) has revealed strong effects of age. If L2 learning commences by the age of roughly 7 years, the L2 is spoken without a detectable accent by at least some individuals who have used the L2 as their primary language for many years. Beyond that age of learning (AOL) the second language, the degree of perceived foreign accent increases with AOL up to early adulthood. Few if any individuals with AOLs greater than 15 years manage to speak their L2 without a detectable foreign accent (e.g. (Oyama, 1976; Suter, 1976; Flege and Fletcher, 1992; Flege et al., 1994b)).

The foreign accent findings just cited suggest that a critical period exists for the learning of speech (e.g. (Lenneberg, 1967; Scovel, 1969, 1988)). However, such a theory does not explain what causes foreign accents (see (Flege, 1987; Long, 1990)). For example, it is uncertain if the AOL effects on degree of foreign accent are due primarily to neurofunctional reorganization which affects the storage of new phonetic information in long-term memory, cognitive changes which affect processing, or psychological and/or sociolingustic factors (e.g. (Penfield and Roberts, 1959; Neville et al., 1992; Paradis, 1995)). Nor is it known whether observed errors in production have a motoric basis, or derive indirectly from an inability to discern perceptually the phonetically relevant properties of L2 vowels, consonants and prosodic dimensions. At a more descriptive level, we do not yet know whether the same age effects are evident for the full range of phonetic and phonological dimensions along which any L1–L2 pair might differ.

Many previous studies have examined the production of vowels and consonants in an L2, but
these studies have been limited in a number of respects. For example, in some studies (e.g. (Koutsoudas and Koutsoudas, 1983)), the individuals whose production errors are reported are not described, and it is not made clear how production errors were observed nor in what situational context(s) or style(s) of speech the observed errors occurred. A great many other, more descriptively adequate studies have focused on individuals in early stages of L2 learning who were unlikely to have reached their ultimate attainment in L2 pronunciation (e.g. (Hammarberg, 1988; Major, 1992)). Given the slow course of speech learning in the native language (L1), errors made by the subjects examined in these studies might reflect insufficient phonetic input or learning in progress, rather than a permanent inability to learn to accurately pronounce L2 vowels and consonants. For example, many studies have shown that native Japanese speakers typically have difficulty producing and perceiving English /ʃ/ and /l/. Two recent studies showed as expected that most Japanese adults who had lived in the United States (US) for less than 3 years produced /ʃ/ and /l/ tokens that were identified incorrectly and judged to be foreign-accented. However, the majority of Japanese subjects who had lived in the US for more than 12 years were found to produce the two English liquids accurately (Flege et al., 1995b; Flege and Takagi, 1995).

Other studies examining highly experienced speakers of an L2 have provided evidence of correct production (or near approximation) of L2 vowels and consonants, as well as evidence of production errors that persisted after many years of speaking the L2 (e.g. (Flege, 1987b; Busà, 1992; Peng, 1993; Munro, 1993; Munro et al., 1995)). This points to the need for examining a wide range of L2 vowels and consonants. It has been hypothesized, for example, that L2 segmental articulation will eventually be more accurate in instances where L2 vowels and consonants are treated as “new”, but not when L2 vowels and consonants are treated as distorted realizations of L1 categories (e.g. (Flege, 1988, 1992a,b)). Although several studies have compared the performance of small groups of “early” versus “late” learners of an L2 (e.g. (Flege, 1991b; Hazan and Boulakia, 1993)), no previous study has examined the segmental production accuracy of non-native subjects sampling a wide range of AOLs. One question of interest is whether, like degree of global foreign accent, segmental production accuracy decreases steadily beyond early childhood, or whether it manifests a sharp decline at some particular AOL.

The non-native subjects examined in the present study were the 240 native Italian (NI) speakers whose overall pronunciation of English was assessed by Flege et al. (1994b). These subjects were first massively exposed to English when they emigrated to Canada between the age of 2 and 23 years. All of them had lived in a predominantly English-speaking environment for at least 15 years at the time of the study (M = 32 years), and nearly all of them reported using English more often than Italian. It therefore seemed likely that these NI subjects would have reached their ultimate level of English pronunciation. The English consonants that were examined either differed phonetically from consonants found in the Italian inventory, or else did not have an obvious counterpart in Italian. The aims of this study were to assess the relation between AOL and the NI subjects’ accuracy in producing English consonants, and to determine if NI subjects could produce some or all of the English consonants accurately.

Consonant production accuracy was assessed using the auditory judgments of native-English speaking listeners, and through acoustic measurements. It would have been ideal to sample the entire inventory of English consonant allophones. However, given the large number of subjects examined, the study was confined to tokens of /p t k l d g/ produced in word-initial position, word-medial tokens of /k g/, and word-final tokens of /p t k b d g/. The consonants examined occurred in a set of 14 consonant-vowel-consonant (CVC) English words elicited using a delayed repetition technique. The study yielded a number of AOL effects on the NI subjects’ production of English consonants. Hypotheses regarding the production of specific consonants will be introduced as each experiment is presented. In Section 7, the results
are discussed in terms of a number of general hypotheses found in the literature concerning the basis of L2 segmental production errors.

2. General method

2.1. Talkers

A total of 240 native speakers of Italian (110 males, 130 females) and 24 native speakers of English (10 males, 14 females) participated. The native Italian (NI) subjects were first massively exposed to English when they emigrated to Canada from Italy between the ages of 2 and 23 years. All of the native English (NE) subjects were born in Canada and so will be designated as having an age of arrival in Canada of 0 years. As summarized in Table 1, the NI subjects were slightly older on average than were the NE subjects (44 versus 39 years). On average, the NI subjects arrived in Canada at the age of 13 years and lived there for 32 years (range = 15 to 44 years). They reported using English more than Italian on a daily basis (69% versus 28%), but said they pronounced Italian better than English (5.5 versus 5.2 on a 7-point scale ranging from "poor" to "good").

The population averages just reported are somewhat misleading because of confounds with the NI subjects' ages upon arrival in Canada. For example, the NI subjects' estimates of how frequently they used English and Italian varied as a function of their age upon arrival in Canada. Those who arrived in early childhood reported using English far more than Italian. The gap narrowed as age of arrival increased, but on the average even those NI subjects who arrived in Canada as adults reported using English somewhat more than Italian. The NI subjects' estimates of how well they pronounced their two languages also varied as a function of age of arrival. Most NI subjects who arrived in Canada before the age of 12 years reported pronouncing English better than Italian, whereas most who

### Table 1
Characteristics of the 110 male and 130 female native Italian subjects who participated

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronological age</td>
<td>44.1</td>
<td>6.3</td>
<td>24.7</td>
<td>56.5</td>
</tr>
<tr>
<td>Age of arrival in Canada</td>
<td>12.5</td>
<td>5.9</td>
<td>1.9</td>
<td>23.2</td>
</tr>
<tr>
<td>Length of residence in Canada</td>
<td>31.5</td>
<td>6.0</td>
<td>14.6</td>
<td>44.3</td>
</tr>
<tr>
<td>% Use of English</td>
<td>69%</td>
<td>21%</td>
<td>5%</td>
<td>100%</td>
</tr>
<tr>
<td>% Use of Italian</td>
<td>28%</td>
<td>21%</td>
<td>0%</td>
<td>95%</td>
</tr>
<tr>
<td>Ability to pronounce English a</td>
<td>5.2</td>
<td>1.5</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Ability to pronounce Italian a</td>
<td>5.5</td>
<td>1.6</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Time spent in Italy</td>
<td>0.5</td>
<td>0.8</td>
<td>0</td>
<td>10.0</td>
</tr>
</tbody>
</table>

a Self-estimates on a scale ranging from "poor" (1) to "good" (7).

### Table 2
Characteristics of the ten subgroups of 24 native Italian (NI) subjects each who participated

<table>
<thead>
<tr>
<th>AOL(%) &amp; Arrival age</th>
<th>First exposure c</th>
<th>Acquisition d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>Range</td>
</tr>
<tr>
<td>3 (92%) 3.1</td>
<td>1.9– 4.1</td>
<td>4.3</td>
</tr>
<tr>
<td>5 (87%) 5.2</td>
<td>4.2– 6.4</td>
<td>5.4</td>
</tr>
<tr>
<td>7 (82%) 7.5</td>
<td>6.5– 8.6</td>
<td>7.5</td>
</tr>
<tr>
<td>9 (77%) 9.6</td>
<td>8.7–10.6</td>
<td>9.6</td>
</tr>
<tr>
<td>11 (74%) 11.6</td>
<td>10.6–12.6</td>
<td>11.6</td>
</tr>
<tr>
<td>13 (69%) 13.6</td>
<td>12.7–14.8</td>
<td>13.6</td>
</tr>
<tr>
<td>15 (65%) 15.8</td>
<td>15.0–16.8</td>
<td>15.9</td>
</tr>
<tr>
<td>17 (62%) 17.5</td>
<td>16.8–18.5</td>
<td>17.5</td>
</tr>
<tr>
<td>19 (57%) 19.2</td>
<td>17.8–20.2</td>
<td>20.3</td>
</tr>
<tr>
<td>21 (56%) 21.5</td>
<td>20.2–23.2</td>
<td>21.5</td>
</tr>
</tbody>
</table>

a AOL(%): the NI subjects' age of learning English as indexed by their age upon arrival in Canada, and the percentage of their lives that had been spent in Canada.
b Arrival age: the NI subjects' age of emigration to Canada, in years.
c First exposure: the NI subjects' self-reported age of exposure to English in Canada, in years.
d Acquisition: the self-reported age at which NI subjects said they could first speak English "comfortably", in years.
arrived after the age of 12 years said the opposite (see (Flege et al., 1994b)). There was a modest correlation between the NI subjects' age upon arrival and their length of residence in Canada ($r = -0.437$; $df = 238$, $< 0.01$). Weak but significant correlations were observed between the NI subjects' self-reported frequency of use of English and ability to pronounce English ($r = 0.348$), as well as between use of Italian and Italian pronunciation ability ($r = 0.468$).

The NI subjects were asked to identify which of their two languages they spoke best. Ninety percent of the subjects who arrived in Canada before the age of 12 years indicated that English was the better of their two languages, but only 26% of those who arrived in Canada later in life gave that response. The NI subjects were also asked which of their two languages they would be more reluctant to lose through injury or illness. Despite self-reported differences in ability to speak English, roughly the same percentages of NI subjects who arrived in Canada before versus after the age of 12 years reported being less willing to lose English than Italian (86% versus 79%). This suggests that English was very important to all of the NI subjects, not just the early arrivals.

The 240 NI subjects were assigned to subgroups based on age of arrival in Canada, which will serve as an index of their age of learning (AOL) English. As shown in Table 2, the mean arrival ages of subjects in ten NI subgroups increased in roughly 2-year increments. Table 2 also presents the mean age at which the NI subjects estimated having first been exposed massively to English, which occurred on average 0.2 years after their arrival in Canada. The subjects estimated that, on average, they could speak English "comfortably" 1.6 years after their arrival in Canada. The subjects estimated that, on average, they could speak English "comfortably" 1.6 years after their arrival in Canada. (The self-reported time needed was apparently greater than average for subjects who arrived in the pre-school years, late adolescence or early adulthood.) The age ranges of subjects assigned to the various subgroups varied somewhat but, for convenience, the NI subgroups will be designated by their average arrival age, rounded down to the nearest whole number. For example, the subgroup consisting of subjects who arrived in Canada at an average age of 5.2 years (range: 4.2 to 6.4 years) will be referred to as the subgroup with an AOL of "5" years.

2.2. Speech materials

The subjects were recorded one at a time in a quiet room at a Catholic church in Ottawa by one of us (MJM) who does not speak Italian. The experimenter verified that each subject could speak English well enough to respond to a detailed language background questionnaire (LBQ), and did not have an obvious hearing problem. The subjects' productions of English words were recorded immediately after the LBQ was administered using a head mounted microphone (Shure Model SM10A) and a portable cassette tape recorder (Sony Model TC-D5ProII). Of the 25 words recorded, just 14 were examined in the present study: pick, peak, tack, tag, tacking, tagging, cap, cab, read, raid, they, then, thought, thief.

The test words were elicited twice each using orthographic and aural cues. The list of words to be spoken appeared on a written list placed in front of the subjects, who said each word in a carrier phrase (Now I say ...) after hearing the word played out over a loudspeaker in a different carrier phrase (. is the next word). This elicitation procedure ensured that difficulty in reading would not masquerade as foreign accent. Although the words to be repeated were modeled by a native English speaker on the tape, the delay between the model and the subjects' repetition of it, as well as the presence of intervening speech material, was expected to prevent direct (i.e., phonetically unmediated) imitations from sensory memory.

2.3. Stimulus preparation

The second available token of each word was low-pass filtered at 10.0 kHz, then digitized at 22.05 kHz using a PC. The test words were edited from their original carrier phrase using a Kay Computerized Speech Lab (CSL), then normalized for peak amplitude. The stimuli were ramped on and off using special software to remove transients arising from editing (zero amplitude to full amplitude over the initial 5 ms; full amplitude to
zero amplitude over the final 20 ms). This process yielded a total of 3,696 digitized waveforms.

2.4. Perceptual evaluation

Young adults who were monolingual native speakers of English auditorily evaluated consonants spoken by the NE and NI subjects in three experiments. The listeners all reported hearing normally, and passed a pure-tone hearing screening prior to participating. The digitized words were presented binaurally over headphones in a random order to listeners, who were tested one at a time in a sound booth. The interval between stimuli and the listeners' responses was a constant 1.0 s. Given the large number of stimuli, the stimuli could be presented just one time each. However, each block of stimuli began with 10 or 20 practice trials to permit listeners to become accustomed to the stimuli and task.

2.5. Analyses

The dependent variables in certain experiments were the percentage of tokens that were identified in terms of the intended phonetic categories (e.g., /t/ called /t/, not /d/), and A* scores derived from the percent correct scores. In other instances, the dependent variable was the percentage of tokens judged to have been produced “correctly”. Dependent variables were submitted to analyses of variance (ANOVAs) in which Group (11 levels) served as a between-subjects factor, and any other variable was treated as a repeated (i.e., within-subjects) factor. When the Group effect was significant, Williams' test (Williams, 1971) was used to determine which of the NI groups, if any, differed significantly from the NE comparison group. An alpha level of 0.01 was used in testing the significance of main effects and interactions, and for all post-hoc comparisons.

3. Experiment 1. Word-initial /I/, /a/, /0/

The purpose of this experiment was to assess word-initial tokens of English /I a fl/ spoken by the NI and NE subjects described earlier. Flege (1988, 1992a,b) hypothesized that even adult learners can produce L2 consonants accurately provided that the L2 consonants are judged to be distinct phonetically from any consonant in the L1 inventory, and provided that they receive sufficient native-speaker input. From a phonetic standpoint, the three English consonants examined here differ from any consonant in the Italian inventory. Italian has five fricatives (labiodental /f v/, apical /s z/, alveo-palatal /f/), but does not possess the interdental fricatives /θ/ . Italian /r/ is an apico-postalveolar trill, rather than the kind of dorsal velar approximant /l/ found in English (Vagges et al., 1978; Magno Caldognetto et al., 1979; Ferrero and Magno Caldognetto, 1986). Despite this, no specific predictions could be generated from the hypothesis mentioned above because no previous study has directly assessed the perceived relation between English /l δ θ/ and consonants in the Italian inventory. What little is known suggests that English /δ/ and /θ/ will not be treated as “new” by NI learners of English. Morosan and Jamieson (1989) reported that speakers of another language which lacks /θ/, Canadian French, may have difficulty perceptually distinguishing English /δ/ from /θ/, and English /δ/ from /d/. If the same held true for NI learners of English, one might expect them to produce English /θ/ and /θ/ as do certain French Canadians, that is, to realize these fricative as stops.

3.1. Method

One token each of words beginning in /l/ (read, raid), /θ/ (they, then) and /θ/ (thought, thief) spoken by each subject was digitized. The word-final consonant and the second one-half of the vowel of each token was removed digitally to prevent variations in these portions of the words from influencing listeners' judgments. Edited tokens of the six test words were presented in separate, counterbalanced blocks to native speakers of Canadian English (8 males and 2 females with a mean age of 27 years) residing in Birmingham, Alabama. Only one intended consonant (known beforehand to the listeners) was evaluated in each block.
The listeners were offered the following response alternatives with which to label CVs edited from words beginning in /θ/ and /ð/:

1. "correct th", 2. "distorted th", 3. "s", 4. "t", 5. "f", or 6. "d". The response alternatives provided for CVs edited from words beginning in /r/ were 1. "correct r", 2. "distorted r", 3. "w", 4. "l", 5. "th", and 6. "tr". These response alternatives were based on the first author's evaluation of the corpus of CVs. In his opinion, they captured the full range of English consonants that might be heard by native English-speaking listeners without phonetic training. Although the first author never heard NI subjects say /l/ for /l/, "l" was included among the response alternative for words beginning in /l/ because such substitutions are common among other L1 groups (e.g., native Japanese). The listeners were required to respond to each stimulus, and were told to guess if uncertain. The number of times each response alternative was used was tabulated for the words beginning in /r/, /θ/ and /ð/. The dependent variable examined in statistical analyses was the number of "correct" productions of each consonant, out of a maximum possible of 20.

3.2. Results and discussion

The results obtained here indicate that NI subjects who begin learning English after a certain age will persist in producing /r/, /θ/ and /ð/ inaccurately. Fig. 1 shows the mean percentage of /r/, /θ/ and /ð/ tokens produced by the 11 groups of subjects that were judged to have been produced "correctly" by the English-speaking listeners. In this figure and those to follow, the NE subjects are designated as having an arrival age of 0 years; and the NI subjects' age of arrival in Canada is used as an index of their age of learning (AOL) English. Each mean value is based on a total of 480 forced-choice judgments (24 talkers × 10 listeners × 2 tokens per talker).

Inspection of Fig. 1 reveals that the percentage of /r/ tokens judged to have been produced correctly decreased as AOL increased. Most /r/ tokens that were not judged to have been produced correctly were judged to have been produced in a "distorted" manner rather than being heard as /w/, /l/, /tr/ or /θ/. The ANOVA examining the percentage of "correct r" judgments yielded a significant effect of Group ($F(10,253) = 18.9, p < 0.01$). A Williams post-hoc test revealed that the NI subjects with AOLs of 11 to 21 years, but not those with AOLs of 3 to 9 years, produced accurate /r/ tokens significantly less often than did the NE subjects ($p < 0.01$). Significant effects of Group were also obtained in the ANOVAs examining the percentage of /θ/ tokens ($F(10,253) = 20.2$) and the percentage of /ð/ tokens ($F(10,253) = 28.6$) that were produced "correctly" ($p < 0.01$). Post-hoc tests revealed that just the NI subjects with AOLs of 9 to 21 years differed significantly from the NE subjects in producing /θ/, and just those with AOLs of 9 to 21 years produced /θ/ correctly significantly less often than did the NE subjects ($p < 0.01$). Incorrect realizations of /θ/ were usually heard as /d/, whereas incorrect realizations of /θ/ were usually heard as /t/. That is,
when the English fricatives were replaced by a stop, the intended fricatives’ voicing feature was preserved.

It appears that highly experienced Italian speakers of English seldom if ever produce English /θ/ and /ð/ as a more posterior fricative, /s/ or /z/. Such a substitution pattern is common for speakers of other languages, such as Japanese (Weinberger, 1990). Japanese, like Italian, possesses apical stops and fricatives, but not interdental fricatives. The substitution of stops rather than fricatives for English /θ/ by speakers of certain L1s has suggested a number of possible explanations as to how L2 consonants are mapped onto consonants in the L1. It may be that certain features in L2 consonants are more prominent, or receive higher weightings, than do other features owing to the overall structure of feature contrasts in the L2, or to differences in the frequency with which various features are used for contrasting words in the L1 (e.g. (Ritchie, 1968; Weinberger, 1990; Hancin-Bhatt, 1993)). Although the perceptual features used in identifying Italian consonants have been examined (e.g. (Magno Caldognetto et al., 1988)), we know of no research that provides a direct explanation for why our NI subjects apparently attended to the non-stridency feature of English /θ/ rather than to their continuancy feature in choosing the closest Italian consonants, which are apparently /t/ and /d/. It is worth noting, however, that speakers of European French apparently substitute both fricatives and stops for English /θ/ δ/. The relative frequency of the variants used may vary according to English-language proficiency and speaking style (Wenk, 1979).

As noted earlier, the NI subjects’ age of arrival in Canada was confounded with their length of residence in Canada. However, the effects shown in Fig. 1 are unlikely to have been due to variations in amount of L2 experience. Modest correlations between the subjects’ ages of arrival and the percent correct scores for the three consonant were obtained when variations in length of residence (LOR) in Canada were partialled out (range: \( r = -0.536 \) to -0.617, \( p < 0.01 \)). However, correlations between the percent correct scores and LOR were all non-significant when variations in age of arrival were partialled out (\( p > 0.10 \)).

It is uncertain how best to account for variation in ability to pronounce English /θ/ δ/ among the NI subjects with AOLSs greater than about 10 years. One possibility is that the AOL effects arose from the loss of motoric ability to produce new consonants once the L1 phonetic inventory had been established. Another possibility, mentioned earlier, is that some NI subjects failed to note perceptually the phonetic differences which distinguish /θ/, /ð/ and /θ/ from the most similar consonants in Italian. Still another possibility is that attitudes towards the L2 or motivation to pronounce it without accent may play a role. Research reported by Segalowitz and Gatbonton (1977) indicated that French Canadian (Quebecois) subjects’ production of English /ð/ θ/ were related to strength of “ethnic identification”. The more nationalistic were the subjects’ political attitudes, the less English-like were their productions of English /ð/ θ/.

4. Experiment 2. Word-final /t/ and /d/

Here we examined the NI and NE subjects’ productions of word-final /t/ and /d/ tokens. The aim of this experiment was to test the hypothesis (e.g. (Flege, 1988)) that word-final stops in an L2 can eventually be produced accurately, even by adult learners, if the L1 lacks word-final stops. This hypothesis was predicated on the assumption that interlingual identification of L1 and L2 consonants occurs at a phonetic (i.e., allophonic) level rather than at a phonemic level. In terms of the present study, the assumption being made was that Italian learners of English will treat tokens of /t/ and /d/ in the final position of English words as “new” rather than relating them perceptually to /t/ and /d/ tokens found in the word-initial or word-medial positions of Italian words.

Previous L2 production research has shown that non-native speakers whose L1 lacks word-final stops are more likely to produce /t/ accurately in the final position of English words than /d/ (e.g. (Flege and Davidian, 1984)). Like chil-
Children learning English as their L1, non-native adults frequently devoice English word-final /d/ stops, perhaps because voiced stops are more "marked" (physiologically complex) in word-final position than are voiceless stops (Eckman and Iverson, 1993). Native English-speaking listeners in a study by Flege et al. (1992) identified, as /t/ or /d/, word-final stops in English words spoken by groups of native Mandarin and Spanish who had begun learning English as adults and had lived in the US for averages of 5.5 and 9.0 years, respectively. Neither Mandarin nor Spanish has a contrast between /t/ and /d/ in word-final position. Contrary to hypothesis, many non-native subjects did not produce a perceptually effective contrast between /t/ and /d/, and subgroups of subjects who were experienced in English did not succeed better in doing so than did subgroups who were relatively less experienced.

The findings of Flege et al. (1992) may mean that adult learners of an L2 are unable to learn to produce word-final consonants accurately, or that the assumption concerning the nature of interlingual identification is incorrect. Another possibility is that the native Mandarin and Spanish subjects examined had not yet received sufficient native-speaker phonetic input. This last possibility was tested here. Like Mandarin and Spanish, Italian does not possess a contrast between /b d g/ and /p t k/ in word-final position (e.g., Ferrero et al., 1979). One question of interest was whether the NI subjects who began learning English as adults would be judged to have produced both word-final /t/ and /d/ tokens accurately. If they did not, it was of interest to know how early in life L2 learning must commence in order to ensure the correct production of both word-final English stops.

4.1. Method

A single token of four words (read, raid, beat, bait) spoken by each of the NI and NE subjects were digitized, then presented in random order one time each via headphones to listeners. The listeners (4 males, 6 females) were monolingual native speakers of American English with a mean age of 27 years (range: 24 to 33 years). The digitized words were presented in four approximately 12-min blocks, the order of which was counterbalanced across listeners. Each block contained all four words spoken by one-fourth of the subjects, that is, by six NE and 60 NI subjects.

The listeners were told to push one of four buttons depending on what they heard. The buttons were marked "exaggerated t", "t", "d" and "exaggerated d". The subjects were told that a stop should be identified as "exaggerated" if it was produced in a way that was "more clear or emphatic" than is typical for English. Two response alternatives were offered for both stop consonants so that we could test the possibility that NI subjects may have re-syllabified the English words. For example, if they treated the final stop of raid as a syllable-initial rather than syllable-final stop, one might expect to hear a paragogic vowel, or perhaps a very strong voiced release burst. However, preliminary analyses did not reveal any systematic differences between groups in the use of the label "exaggerated", and so we have collapsed responses over the two alternatives for /t/ and /d/. The percentage of times the /t/ and /d/ tokens spoken by each subject were identified correctly was calculated. Each score was based on 20 forced-choice identifications (10 listeners × 2 tokens per stop).

There are two disadvantages to examining percent correct identification scores such as those just described. One is that such scores are often at ceiling for native speakers. Another is that listeners may be biased towards one of the two response alternatives. We therefore calculated unbiased estimates of the listeners' sensitivity to the /t/-/d/ contrast produced by each subject. In calculating the A’ scores, correct identifications of /d/ tokens as /d/ were treated as hits and identifications of /t/ tokens as /d/ were treated as false alarms (see Snodgrass et al., 1985, p. 452).

4.2. Results and discussion

The mean percent correct scores obtained for the word-final tokens of /t/ and /d/ are shown in Fig. 2. The A’ scores obtained for the /t/-/d/ tokens for the 11 groups of subjects are presented.
Fig. 2. Mean rate at which word-final /t/ and /d/ tokens spoken by 11 groups of subjects were identified correctly; see the caption of Fig. 1.

in Table 3. The /t/ tokens produced by the subjects in every group, including the latest-arriving NI subjects, were identified correctly in nearly every instance. However, the /d/ tokens spoken by NI subjects with AOLs of 15 to 21 years were sometimes misidentified as /t/. The ANOVA examining the A' scores yielded a significant effect of Group ($F(10,253) = 5.67, p < 0.01$). A post-hoc test revealed that the NI subjects with AOLs of 15 to 21 years, but not those with AOLs of 3 to 13 years, differed significantly from the NE subjects ($p < 0.01$). This finding suggests that, beyond a certain AOL, at least some non-natives cannot produce English /d/ correctly even if they have spoken English for many years. It leads one to question the possibility that the Mandarin and Spanish subjects examined by Flege et al. (1992) produced word-final English stops incorrectly because they had not yet received sufficient native-speaker input.

Still, we think it is noteworthy that the NI subjects seldom if ever omitted final stops or added a paragogic vowel. Both phenomena have been reported for native speakers of a variety of L1s (see references in (Flege et al., 1992)). Fonda (1984) suggested that adding vowels after a word final consonant is typical for native Italian learners of English. Also, the Italian “late learners” examined in the present study seem to have been much more successful in producing word-final English stops than were the Mandarin and Spanish late learners examined previously (Flege et al., 1992). The /t/ and /d/ tokens spoken by the 48 NI subjects with AOLs greater than 18 years were identified correctly in 96% of instances on average, whereas the percent correct identification rates obtained for the less experienced Mandarin and Spanish subjects averaged just 62% and 73% correct, respectively.

Given how long they had spoken English, the

### Table 3

<table>
<thead>
<tr>
<th>AOL</th>
<th>/b/ versus /p/</th>
<th>/d/ versus /t/</th>
<th>/g/ versus /k/</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PC b A'(SD)</td>
<td>PC b A'(SD)</td>
<td>PC b A'(SD)</td>
</tr>
<tr>
<td>0</td>
<td>95 0.972 (0.034)</td>
<td>100 0.999 (0.004)</td>
<td>97 0.982 (0.047)</td>
</tr>
<tr>
<td>3</td>
<td>93 0.960 (0.070)</td>
<td>99 0.995 (0.014)</td>
<td>97 0.985 (0.020)</td>
</tr>
<tr>
<td>5</td>
<td>97 0.984 (0.021)</td>
<td>99 0.997 (0.013)</td>
<td>98 0.988 (0.015)</td>
</tr>
<tr>
<td>7</td>
<td>92 0.947 (0.103)</td>
<td>99 0.996 (0.011)</td>
<td>96 0.978 (0.028)</td>
</tr>
<tr>
<td>9</td>
<td>94 0.967 (0.062)</td>
<td>99 0.997 (0.006)</td>
<td>96 0.976 (0.034)</td>
</tr>
<tr>
<td>11</td>
<td>86 0.921 (0.106)</td>
<td>98 0.991 (0.019)</td>
<td>94 0.967 (0.064)</td>
</tr>
<tr>
<td>13</td>
<td>88 0.925 (0.113)</td>
<td>98 0.991 (0.011)</td>
<td>91 0.932 (0.165)</td>
</tr>
<tr>
<td>15</td>
<td>93 0.963 (0.051)</td>
<td>94 0.971 (0.036)</td>
<td>90 0.946 (0.061)</td>
</tr>
<tr>
<td>17</td>
<td>86 0.904 (0.163)</td>
<td>95 0.975 (0.036)</td>
<td>88 0.902 (0.179)</td>
</tr>
<tr>
<td>19</td>
<td>80 0.862 (0.173)</td>
<td>96 0.979 (0.025)</td>
<td>89 0.928 (0.116)</td>
</tr>
<tr>
<td>21</td>
<td>83 0.880 (0.148)</td>
<td>95 0.973 (0.036)</td>
<td>84 0.904 (0.107)</td>
</tr>
</tbody>
</table>

---

*a* AOL: the subjects age of learning English (see Table 2 for more exact information).

*b* PC: mean percentage of correct identifications of word-final stops.
fact that some of the NI subjects devoiced /d/ tokens is of interest. None of the NE subjects did so, nor did NI subjects who began learning English prior to the age of 15 years. One possibility is that, having passed a critical period for speech learning (Scovel, 1969, 1988), some of the NI subjects were incapable of the motoric learning needed to produce English /d/ correctly in word-final position. For example, they may have been unable to learn to actively expand the volume of the supraglottal cavity during /d/ constriction in order to sustain closure voicing (Flege et al., 1987).

Alternatively, the production difficulty observed for certain late learners may have a perceptual origin. The NI late learners who devoiced /d/ may have failed to discern the acoustic phonetic features that distinguish /t/ from /d/ in the final position of English words. The features used to contrast /t/ from /d/ in the medial position of Italian words appear to differ from those used to contrast /t/ from /d/ in the final position of English words. For example, the presence versus absence of closure voicing, and the duration of stop closure intervals, appear to be more important for word-medial Italian stop voicing contrasts than for word-final English contrasts, whereas vowel duration may be more important for word-final English than word-medial Italian contrasts (see (Vagges et al., 1978; Magno Caldognetto et al., 1979, Farnetani and Kori, 1991)). If certain NI learners of English were to note only those features important to stop voicing contrasts in Italian, one would not expect them to develop accurate perceptual representations for word-final English /t/ and /d/ tokens. This is because word-final English stops are often unreleased in conversational English (thereby removing closure duration cues), and final /d/s are often produced with silence during part or all of their closure interval (see (Flege et al., 1992) for discussion). And, if NI subjects’ perceptual representations for word-final English stops were not nativelike, one might expect their productions of those stops to diverge from NE speakers’ (assuming, of course, that their motoric output conformed to their perceptual representations).

To help determine why certain NI late learners’ /d/s were misidentified, we examined three acoustic dimensions relevant to the perception by native English-speaking listeners of the voicing feature in word-final stops. This acoustic analysis focused on the final /d/ in *read* and *raid* tokens spoken by subjects in three groups. One group of 20 individuals, designated the “poor producers” of /d/, consisted of 20 NI subjects whose /d/s were often misidentified. All of these subjects had AOLs greater than 15.0 years. Another group of 20 individuals, designated the “poor producers,” consisted of NI subjects who were matched in AOL to the poor producers, but whose /d/s were always identified correctly. The final group of 20 subjects consisted of randomly chosen subjects from the NE control group. The duration of the “vowel” in each word was measured from the point of rapid spectral change, indicating release

<table>
<thead>
<tr>
<th></th>
<th>Poor producers</th>
<th>Good producers</th>
<th>Native English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of learning a</td>
<td>18.4</td>
<td>18.4</td>
<td>—</td>
</tr>
<tr>
<td>Correct identifications b</td>
<td>70%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Closure voicing</td>
<td>44 (29)</td>
<td>73 (20)</td>
<td>51 (22)</td>
</tr>
<tr>
<td>Vowel duration</td>
<td>201 (36)</td>
<td>223 (38)</td>
<td>250 (48)</td>
</tr>
<tr>
<td>Stop closure duration</td>
<td>124 (33)</td>
<td>98 (35)</td>
<td>65 (16)</td>
</tr>
</tbody>
</table>

a AOL: age of learning English as an L2, in years.
b Correct identifications: Rate at which native English-speaking listeners were able to identify correctly the /d/ tokens spoken by talkers in the three groups.
of the initial liquid, to the point of complete constriction of the final stop, signalled by a decrease in waveform amplitude and simplification of the waveform shape. The duration of closure voicing was measured from the constriction of /d/ to the cessation of low-amplitude periodicity. The cessation of periodicity (i.e., glottal pulsing) often, but not always, occurred before the release of /d/ closure, which was signalled by a release burst. The durations of /d/ closure intervals were measured from the constriction of /d/ to the beginning of the release bursts.

The acoustic measurements are summarized in Table 4. The effect of Group on the duration on stop closure voicing was significant \((F(2,57)=7.83, p < 0.001)\). A post-hoc test revealed that the "good" producers sustained closure voicing significantly longer than did the "poor" producers and the NE subjects \((p < 0.01)\). The poor producers did not differ significantly from the NE subjects \((p > 0.10)\), which seems to rule out the most likely candidate for a motorically based explanation of their /d/ production errors.

The effect of Group on vowel duration was also highly significant \((F(2,57)=7.46, p < 0.001)\). Post-hoc tests revealed that the NE subjects made vowels significantly longer than did the poor producers \((p < 0.01)\), but their vowels did not differ significantly from those of the good producers \((p > 0.10)\). Finally, the effect of Group on the duration of stop closure intervals was significant \((F(2,57)=19.9, p < 0.001)\). Interestingly, the poor producers were found to sustain /d/ closures significantly longer than the good producers, who sustained /d/ closures longer than did the NE subjects \((p < 0.01)\).

The prolongation of /d/ closures by the poor producers may have led to the cessation of closure voicing prior to stop release. In all but five (12%) of their /d/ tokens, voicing died out before the end of the stop closure interval, resulting in a voiceless release burst. A voiceless release is a powerful cue to voicelessness (i.e., /t/) for native English-speaking listeners. It is uncertain why the poor producers seem to have prolonged /d/ closure intervals. However, as will be shown in Section 6, it may have been part of a more general phenomenon in Italian-accented English.

5. Experiment 3. Word-final /p b/ and /k g/

This experiment extended Experiment 2 to two additional places of articulation by examining the NI and NE subjects' productions of word-final tokens of /p b/ and /k g/.

5.1. Method

One token of the words cap, cab, tack and tag spoken by each subject was digitized and presented via headphones to listeners in a forced-choice identification experiment. The listeners were monolingual native speakers of American English (5 males, 5 females) with a mean age of 30 years (range: 25 to 36 years). Four of these listeners had participated previously in Experiment 2. The digitized words were presented in six counterbalanced blocks, each of which lasted about 7 min. In three blocks, the cap and cab tokens spoken by different one-thirds of the NE and NI talkers were presented. In the remaining three blocks, the tack-tag tokens were presented. The listeners were told to push one of two buttons, depending on what they heard. The response alternatives offered for the cap and cab tokens were "p" and "b"; the alternatives for tack and tag were "k" and "g". The listeners were required to respond to each word, and were told to guess if unsure. As in the last experiment, we calculated the percentage of times the stops spoken by each subject were identified correctly, and calculated A' scores for the /p/-/b/ and /k/-/g/ tokens produced by each of the 264 subjects.

5.2. Results and discussion

Fig. 3 shows the mean rates at which word-final tokens of /p/, /b/, /k/ and /g/ spoken by the 11 groups of subjects were identified correctly. (The mean percent correct scores and A' scores are presented in Table 3.) The bilabial and velar stops spoken by the NE subjects were identified correctly in 95% and 97% of instances, respectively. The rates for NI talkers who began learning English in adolescence or early adulthood were somewhat lower. The A' scores com-
Fig. 3. Mean rate at which word-final tokens of /p/ and /b/ (top) and /k/ and /g/ (bottom) spoken by the 11 groups of subjects were identified correctly; see the caption of Fig. 1.

puted for /p/-/b/ and /k/-/g/ contrasts were submitted to separate one-way ANOVAs, which both yielded significant effects of Group /p/-/b/: $F(10,253) = 3.34$; /k/-/g/: $F(10,253) = 2.91$; $p < 0.01$). Post-hoc tests revealed that the NI subjects with AOLs of 19 to 21 years, but not those with AOLs of 3 to 17 years, differed significantly from the NE subjects in producing /p/-/b/ (p < 0.01). Just the NI subjects with AOLs of 17 to 21 years were found to differ significantly from the NE subjects in producing /k/-/g/ contrasts (p < 0.01). These results corroborate the Experiment 2 results in showing that some NI late learners had difficulty producing a contrast between voiced and voiceless stops in the final position of English words.

6. Experiment 4. Acoustic measurements

The aim of this experiment was to examine acoustically word-initial, word-medial and word-final English consonants spoken by the NE and NI subjects. The acoustic variables measured here were the voice onset time (VOT) values of /p t k/ tokens in six words (peak, pick, tack, tag, cap, cab); and the duration of vowels, stop closure intervals, and closure voicing in medial and final stops found in four words (tack, tag, tacking, tagging).

Our interest in VOT derived from the results obtained in previous research. First, it has been established that English monolinguals are acutely aware of VOT variations in word-initial /p t k/ tokens. They can reliably choose which members of a synthetic VOT continuum have the most appropriate VOT values for English (Miller and Volaitis, 1989), and can detect small divergences from the VOT norm of English in stops spoken by non-natives (Flege and Hammond, 1982; Bohn and Flege, 1993; Flege and Munro, 1995). Second, many previous studies have examined the production of English /p t k/ by speakers of LIIs in which /p t k/ are realized with short-lag rather than long-lag VOT values. For example, Flege (1991b) found that most native Spanish speakers who began learning English in adulthood produced English /t/ with shorter (Spanishlike) VOT values than did NE speakers. However, most "early learners" who began learning English by the age of 6 years were found to produce English stops with accurate VOT values.

The stops /p t k/ are produced with shorter VOT values in the pre-stressed word-initial position of Italian than English words (Ferrero and Magno Caldognetto, 1986). Not surprisingly, NI subjects require a shorter VOT interval to hear word-initial stops as phonologically voiceless than do NE subjects (Magno Caldognetto et al., 1979; Lisker and Abramson, 1964, 1970). An assumption underlying the present experiment was that although all NI learners of English are likely to detect cross-language phonetic differences auditorily, those who learned English relatively late in life will tend not to treat English /p t k/ as "new" (non-Italian) consonants. That is, they will tend not to treat auditorily detected differences between English /p t k/ and Italian /p t k/, such as VOT differences, as being phonetically relevant. If so, then the data presented here afforded an opportunity to test the hypothesis (e.g. (Flege, 1991b)) that Italian speakers of English
who begin learning English as adults will produce English /p t k/ with VOT values showing the effect of cross-language interference, as has previously been shown for native Spanish learners of English (e.g., (Flege, 1991b)).

Adult learners of English often produce English /p t k/ with “compromise” VOT values, that is, with average VOT values that are intermediate to the mean values observed for English monolinguals and monolingual speakers of their own L1 (e.g. (Caramazza et al., 1973; Williams, 1979; Flege, 1987b)). This led to the expectation that the NI subjects who began learning English as adults would differ from the NE subjects to a greater extent in producing English /p t/ than in producing /k/. The mean value reported for Italian /k/ by Vagges et al. (1978) was 50 ms. This differed less from values reported for English /k/ than the values reported for Italian /p/ and /t/ (12 and 17 ms, respectively) differed from those typical for English /p/ and /t/ (see also (Magno Caldognetto et al., 1979; Ferrero and Magno Caldognetto, 1986; Farnetani, 1989)).

Acoustic measurements of phonetic intervals in tack-tag and tacking-tagging provided a way to assess the relative accuracy of word-medial and word-final stops. Previous work has shown that speakers of certain L1s fail to distinguish English /b d g/ and /p t k/ by means of closure voicing (Flege et al., 1992). However, there was reason to think that such a finding might not be obtained for Italian learners of English. It appears that /b d g/ and /p t k/ are contrasted more reliably by the presence versus absence of closure voicing in the intervocalic position of Italian than English words (see (Farnetani, 1989)). Earlier, we raised the possibility that Italians might relate word-final English stops to word-medial, or even word-initial Italian stops. If this happened, and if Italian production patterns carry over into the production of word-final English stops, we might expect the NI subjects to produce larger, not smaller, closure voicing contrasts between /k/-/g/ than do NE speakers. This was supported earlier by an analysis of closure voicing in word-final /d/ tokens by a subset of NI subjects. (Despite having begun to learn English after the age of 15 years, these subjects managed to produced /d/ accurately.) The earlier analysis of word-final /d/ tokens suggested that some NI speakers of English may sustain stop closure intervals longer than NE speakers. The analyses of /k/ and /g/ closures afforded an opportunity to learn if this is a general phenomenon, or one restricted to word-final /d/.

We expected all of the NE and NI subjects to make /æ/ shorter in one-syllable (tag, tack) than in two-syllable words (tacking, tagging). However, a report by Vayra et al. (1984) suggested the possibility that the NI subjects would show less difference between one and two-syllable words than the NE subjects. Other published data suggested that the NI subjects might produce a somewhat smaller difference in the duration of vowels preceding /k/ versus /g/, but larger stop closure duration differences between these stops than would the NE subjects (Magno Caldognetto et al., 1979; Vagges et al., 1978; Ferrero and Magno Caldognetto, 1986; Farnetani, 1989).

6.1. Method

A single research assistant made all acoustic measurements using displays created by a Kay CSL. The acoustic measures were based primarily on time domain waveforms, with secondary reference to digital spectrograms. VOT was measured to the nearest millisecond from the beginning of the release burst in tokens of /p/, /t/ and /k/, to the first positive peak in the periodic portion (“vowel”) of the waveforms. A test–retest analysis suggested that VOT measurement error was small. Fifty randomly selected tokens were re-measured after two weeks had elapsed. The two sets of VOT measurements diverged by only 1.8 ms on average (range: 0 to 12 ms). In statistical analyses, the VOT values obtained for the two tokens of each stop spoken by each subject were averaged.

Vowel durations were measured from the first positive peak in portions of the waveform corresponding to /æ/ to the last positive peak in the vowel portion. The end of the /æ/ intervals was defined on the basis of changes in waveform
shape and/or intensity, and by a cessation of energy in the region of F2 and higher formants. The closure intervals of /g k/ tokens were measured from the end of the vowel, as just defined, to the beginning of the release bursts (which were almost always evident). The duration of closure voicing was measured from the end of the vowel to the last positive peak (i.e., glottal pulse) in the closure interval. As expected, voicing was often observed to continue for one or two glottal pulses into the /k/ closure intervals, and for longer intervals into the /g/ closure intervals. Closure voicing measures could not be obtained in 19 instances (1.8% of tokens). The duration of stop closure intervals could not be measured in 29 (2.7%) tokens. In these instances, the missing values were replaced by means of the subgroup to which a subject with missing data belonged.

6.2. Results and discussion

6.2.1. VOT

As shown in Fig. 4, the subjects in all ten NI subgroups showed the expected effect of place of articulation on VOT: longer VOT for /k/ than for /t/, and longer VOT values for /t/ than for /p/. More importantly, VOT decreased (became less Englishlike) as AOL increased. The VOT values obtained for all 11 groups were submitted to a Group × Place of Articulation ANOVA, with repeated measures on the Place factor. A significant two-way interaction was obtained ($F(20,506) = 4.18, p < 0.01$) because, as expected, the NE and NI subjects differed more for /p/ and /t/ than they differed for /k/. The simple effect of Group was significant for all three stops ($p < 0.01$), yet post-hoc tests revealed that no NI group differed significantly from the NE subjects for /k/. On the other hand, NI subjects with AOLS of 17 to 21 years produced /p/ with significantly shorter VOT values than did the NE subjects; and NI subjects with AOLS of 11 to 21 years produced /t/ with significantly shorter VOT values than did the NE subjects ($p < 0.01$).

The mean VOT values obtained for the NI subjects with an AOL of 21 years were of special interest. These were the NI subjects who had begun learning English latest in life of all the NI subjects examined here. These subjects produced all three English voiceless stops with average VOT values that were almost exactly intermediate to the values obtained for our NE subjects (/p/-57, /t/-78, /k/-77 ms) and the values reported by Vaggis et al. (1978) for Italian /p/-12, /t/-17, /k/-50 ms. The NI subjects with an AOL of 21 years produced English stops with the following average VOT values: /p/-35, /t/-45, and /k/-65 ms.

As shown in Table 2, the NI subjects with an AOL of 21 years had spend just over half (56%) of their lives in Canada. Their VOT values may therefore have reflected the overall distribution of VOT values heard in both English and Italian realizations of /p t k/. (This assumes, of course, that these subjects did not have separate representations for Italian and English voiceless stops.) Following this line of reasoning, the mean VOT values shown in Fig. 4 may have decreased (become less Englishlike) as AOL increased because the proportion of all stops heard over the course of a lifetime that were English stops decreased.

We believe, however, that a more likely explanation for the observed AOL effect on VOT is that the likelihood of new phonetic categories being established for L2 vowels and consonants diminishes with increasing AOL, as discussed by Flege (1995). On this view, L2 stop consonants can be produced accurately only if L2 learners
note all relevant phonetic differences between corresponding stops in the L1 and L2, such as (but not limited to) the VOT differences that distinguish realizations of /p t k/ in Italian and English. Flege claimed that when cross-language phonetic differences are noted, a bilingual may establish new phonetic categories for L2 stops. These hypothesized new categories will exist in long-term memory alongside the categories established previously for L1 stops.

To assess the hypothesis concerning category formation, a series of histograms were prepared. The bin size used for plotting VOT values was 4 ms. Recall that the mean VOT values obtained for each individual were based on just two observations obtained in a delayed repetition task. The histograms showed the number of NI subjects in two AOL ranges who produced English stops with various VOT values. For the purposes of this discussion, the 80 NI subjects with AOLs of 2 and 9 years will be designated “early” learners, and the 80 NI subjects with AOLs of 16 to 23 years will be designated “late” learners.

Inspection of Fig. 5 reveals that both the early and late learners produced English stops with a wide range of VOT values. Far more late learners than early learners produced English stops with Italian-like short-lag VOT values. More importantly, some NI late learners produced English /p/ and /t/ with English-like long-lag VOT values, whereas others produced these stops with Italian-like short-lag VOT values. Such a pattern was not evident for the early learners, whose distributions were more nearly unimodal. The data for the NI late learners agree with the results obtained in a recent study examining the production of English /p/ by native Spanish late learners of English (Schmidt and Flege, 1994). One possible interpretation of our results for NI subjects is that whereas most early learners had established new phonetic categories for English /p t/, just a subset of the NI late learners – those who produced /p/ and /t/ with long-lag VOT values – may have established new categories for these English stops.

Indirect support for the view that some but not all highly experienced bilinguals may establish additional new phonetic categories for L2 stops was provided by two recent speech perception studies. These studies examined the effect of speaking rate on stop perception by English monolinguals and Spanish/English bilinguals (Flege and Schmidt, 1994a, 1994b). The stimuli consisted of two voice onset time (VOT) continua in which VOT ranged from values appropriate for short-lag realizations of English /b/, to values greatly exceeding those typical for English /p/. Speaking rate was simulated by varying the overall duration of stimuli in the two continua. An important assumption underlying these studies was that goodness ratings of long-lag stops would not vary systematically as a function of stimulus duration (i.e., perceived speaking rate) unless listeners had a phonetic category representation for long-lag stops. The English monolingual subjects were found to give high goodness ratings to stimuli with a wider range of VOT values in the slow- than fast-rate VOT continuum. Spanish monolinguals showed much less effect of the speaking rate manipulation than did the English monolinguals. Most importantly, na-
tive Spanish early learners were found to resemble the English monolinguals to a greater extent than did native Spanish late learners.

The VOT results presented here, as well as the results of auditory analyses in the first three experiments, suggest that AOL had a strong effect on the NI subjects' production of English consonants. Hazan and Boulakia (1993) recently claimed, however, that strength of bilingualism and language dominance, not AOL, were the primary determinants of the VOT values produced in word-initial English stops by French/English bilinguals. The subjects' strength of bilingualism was estimated by evaluating degree of perceived foreign accent in three English sentences. (The ratings were obtained from ten listeners, who used a 5-point scale.) Language dominance was a binary variable based on responses to several questionnaire items (country of birth, language use, length of residence in a predominantly English-speaking country, and self-reported "primary" language).

To evaluate the claim of Hazan and Boulakia (1993), we regressed three values obtained from questionnaires completed by our NI subjects onto VOT values measured produced by these subjects in English /p t k/. Our estimates of the NI subjects' "strength of bilingualism" were the average foreign accent ratings obtained for their productions of five English sentences. (Each estimate was based on a total of 150 continuous-scale ratings; see (Flege et al., 1994b).) We treated dominance as a scale rather than as a binary variable. Following Hazan and Boulakia (1993), the NI subjects received one point if they indicated that English, not Italian, was their "better language"; and one point if they said that Italian was the language they would sacrifice if they had to lose one of their two languages through injury or illness. Matched questions were asked regarding the frequency of use of English and Italian in four settings (home, work/school, social events and overall). Subjects received one point whenever they indicated using English more than Italian. Finally, one point was added if the NI subjects had lived for more than one-half of their life in Ottawa. (Country of birth was not used because all 240 subjects were born in Italy.) The "dominance" scores obtained in this way ranged from 0 to 7, averaging 4.85 (SD = 1.75).

The regression models developed for /p/ and /t/ reached significance at the 0.01 level ($F(3,236) = 18.7$ and 21.3, respectively), but they accounted for little variance in VOT values (18% and 20%, respectively). It is nevertheless of interest to note that, in both instances, AOL was identified as a significant predictor of VOT ($p < 0.001$) but not strength of bilingualism or language dominance ($p > 0.10$). It is uncertain why the present results diverged from those obtained by Hazan and Boulakia (1993). The divergence may be related to the fact that our NI subjects were first exposed to English at a wider range of ages (2 to 23 years), and had spoken their L2 longer ($M = 32$ years) than had the French/English bilinguals examined by Hazan and Boulakia (1993).

As is typical for immigrant populations in North America, the NI subjects who arrived in Canada as children were more likely to be "English dominant" and to pronounce English well than were the NI subjects who arrived in adolescence or early adulthood. The correlation between AOL and the dominance scores was $r = -0.559$ ($df = 238$, $p < 0.01$); and the correlation between AOL and the foreign accent ratings was $r = 0.611$ ($p < 0.01$). Most importantly, the correlation between AOL and VOT was $r = 0.215$ when the other two variables were partialled out. Although very small, this correlation was nonetheless larger than were the correlations obtained between VOT and strength of bilingualism and dominance ($r = 0.060$ and 0.062, respectively) when effects of the other two variables were partialled out. Taken together, these results suggest that AOL was a more important determinant of the NI subjects' production of English stops than was language dominance or strength of bilingualism, at least as defined by Hazan and Boulakia (1993).

### 6.2.2. Vowel duration

The mean durations of the vowel /æ/ in four words (tack, tag, tacking, tagging) spoken by the subjects in each group are shown in Fig. 6(a). As expected, vowels were longer in the one-syllable...
than two-syllable words ($M = 206$ versus $248$ ms), and longer before /g/ than /k/ ($M = 197$ versus $157$ ms). It also appeared that overall vowel duration decreased as AOL increased. The vowel durations were submitted to a Group $\times$ Stop $\times$ Number of Syllables ANOVA, which yielded a significant Group main effect ($F(10,239) = 3.24$, $p < 0.01$). A post-hoc test revealed that just the NI subjects with an AOL of 21 years produced shorter vowel durations than did the NE subjects ($p < 0.01$).

The significant three-way interaction that was obtained ($F(10,239) = 2.93$, $p < 0.01$) appears to have been due to differences that existed between the NE subjects and NI subgroup consisting of subjects with an AOL of 21 years. To explore the interaction, we computed for each subject the extent to which vowels preceding /g/ exceeded in duration the vowels preceding /k/ in the one- and the two-syllable words. These two scores were submitted to separate one-way ANOVAs. The Group factor was found to be significant for the one-syllable words ($F(10,253) = 4.01$, $p < 0.01$), but missed reaching significance for the two-syllable words ($F(10,253) = 2.15$, $p = 0.022$). A post-hoc test revealed that, for tack--tag, just the NI subjects with an AOL of 21 years produced a significantly smaller vowel duration difference than did the NE subjects ($p < 0.01$). We next computed the extent to which vowels in one-syllable words (tack--tag) exceeded in duration vowels in two-syllable words (tacking--tagging). These scores were submitted to separate one-way ANOVAs. The Group factor was significant for tag versus tagging ($F(10,239) = 3.54$, $p < 0.01$), but not for tack versus tacking ($F(10,253) = 0.83$, $p > 0.10$). A post-hoc test revealed that just the NI subjects with an AOL of 21 years differed from the NE subjects in terms of the tag versus tagging vowel difference ($p < 0.01$). These subjects, but not NI subjects who learned English earlier in life, may have maintained Italian word-level timing characteristics (see (Vayra et al., 1984)) in their production of English.

6.2.3. Stop closure duration

The durations of /k g/ stop closure intervals in four words (tag, tack, tagging, tacking) are shown in Fig. 6(b). As expected, stop closures were shorter overall in two- than in one-syllable words ($M = 70$ versus $107$ ms), and shorter for /g/ than /k/ (76 versus 102 ms). The stop closure duration values were submitted to a Group $\times$ Stop $\times$ Number of Syllables ANOVA which yielded an $S \times NS$ interaction ($F(1,253) = 59.1$, $p < 0.01$). The interaction apparently arose because the /g/ versus /k/ differences were larger in one-syllable ($M = 125$ versus 90 ms) than in two-syllable words ($M = 80$ versus 61 ms). The lack of a significant three-way interaction ($F(10,253) = 0.91$, $p > 0.10$) suggests that the subjects in all 11 groups showed much the same effect of stop voicing and number of syllables on the durations of stop closure intervals. More importantly, a significant Group main effect was
obtained ($F(10,253) = 5.55, p < 0.01$). A post-hoc test revealed that NI subjects with AOLs of 11 to 21 years produced stops with significantly longer closure intervals than did the NE subjects ($p < 0.01$). This effect is opposite in direction to the vowel duration effect reported earlier, and is therefore unlikely to have reflected differences in overall speaking rate.

The between-group difference in stop closure duration might be attributed to the use made of closure duration in Italian. Italian possesses contrasts between phonemically long (i.e., geminated) and phonemically short consonants. Although the duration of vowels preceding long versus short medial consonants varies, Italian listeners base their identifications of medial consonants primarily on the duration of the consonants themselves (Bertinetto and Vivaldi, 1978). We speculate that the temporal patterns observed for the NI subjects who began learning English in late adolescence or early adulthood reflect an Italianlike temporal organization at the syllable or word level. If so, then one might conclude that certain cross-language differences in speech timing (or rhythm) can be successfully negotiated only if L2 learning commences by the age of about 11 years.

6.2.4. Stop closure voicing

The measures of closure voicing duration in /k/ and /g/ tokens were submitted to a Group X Stop X Number of Syllables ANOVA. A significant main effect of Stop was obtained ($F(1,253) = 1170.8, p < 0.01$) because closure voicing was sustained longer in /g/ than /k/ ($M = 56$ versus 14 ms). A significant $S \times NS$ interaction was obtained ($F(1,253) = 19.2, p < 0.01$) because the /g/ versus /k/ difference was smaller in two-syllable than one-syllable words. However, neither the main effect of Group ($F(10,234) = 1.85, p = 0.054$) nor any interaction involving the Group factor reached significance. This finding agrees with the data presented earlier for word-final /d/ tokens. It suggests that native Italian learners of English are able to produce English /b d g/ with as much, or perhaps even more, closure voicing than NE speakers (see also (Flege et al., 1992)). It may be that the relatively great importance (weight) accorded closure voicing as a cue to the voicing feature in Italian stops was maintained by the NI bilinguals in their processing of English stops.

6.2.5. Individual differences

Despite the many strong AOL effects noted here, it should be emphasized that a fairly substantial amount of inter-subject variability was evident. Some but not all NI subjects who learned English in late adolescence or early adulthood were found to produce English consonants accurately. To estimate the number of NI subjects in each subgroup who were successful in producing word-initial tokens of /d/ (Experiment 1) and the voicing contrast between /p b/, /t d/ and /k g/ in word-final position (Experiments 2 and 3), we determined how many of the 24 subjects in each NI subgroup received scores that fell within $\pm 2$ standard deviations (SDs) of the mean value observed for the 24 NE subjects. Percent ‘correct’ scores were used in the analysis of /d/. A’s scores were used in analyses involving productions of word-final /p–/b/, /t–/d/ and /k–/g/ contrasts.

The number of NI subjects in each subgroup (maximum = 24) who met this criterion is shown in Table 5. Of the subjects with AOLs of 3 to 7 years, the accuracy criterion just described was met in 196 instances (91%) for word-initial consonants, and in 198 (92%) instances for word-final consonants. For NI subjects with AOLs of 9 to 13 years, this held true in 126 instances (58%) for initial consonants and 168 instances (78%) for final consonants. For NI subjects with AOLs of 15–21 years, the accuracy criterion was met in 58 (20%) of instances for word-initial consonants and 181 (63%) of instances for word-final consonants. 2

Although the measures of word-initial and word-final consonants were not equivalent, the

---

2 It is at least possible that some word-final stops produced by subjects were foreign-accented despite being intelligible (Flege et al., 1995b). However, it is likely that the values reported here for word-initial consonants reflect tokens that were both intelligible and free of foreign accent.
results suggest that the NI subjects' production of final consonants were more accurate than their production of word-initial English consonants. If so, this might be taken as indirect evidence in support of the claim (Flege, 1995) that interlingual identification occurs at a position-sensitive allophonic level rather than at a phonemic level. That is, the bilinguals may have been more successful in producing word-final English stops because there are no word-final Italian stops, and so no L1 phonetic structures which existed to influence or constrain their learning of English stops in the word-final position.

7. General discussion

This study examined the production of word-initial, word-medial and word-final English consonants by ten subgroups of native Italian (NI) subjects who were all highly experienced in English. The NI subjects differed primarily according to the chronological age at which they had emigrated to Canada, and thus according to their age of learning (AOL) English as a second language. These subjects had spoken English for 32 years on average, and most said they spoke English more than Italian. Thus, the NI subjects enrolled in the present study were likely to have reached their ultimate attainment in English pronunciation.

The NI subjects' production of English consonants was evaluated auditorily by English-speaking listeners in three experiments, and acoustically in a fourth experiment. In two instances, none of the ten Italian subgroups differing in AOL were found to differ significantly from a native English (NE) control group. This held true for word-final tokens of /t/, and for the VOT values produced in word-initial tokens of /k/. The finding for /k/ was likely to have been due to the fact that the VOT difference between Italian and English realizations of /k/ are too small to yield measurable effects of cross-language phonetic interference.

In most instances, however, strong effects of AOL were noted for the ten subgroups of NI subjects. As expected, we generally found that "younger is better" as far as pronouncing English consonants is concerned. The study revealed that some of the NI subgroups produced word-initial tokens of /l/ /r/ /b/ /t/ /d/ correctly less often than did NE subjects, and some NI subjects produced less effective perceptual contrasts between /b/-/p/, /t/-/d/ and /k/-/g/ in the final position of English words. As assessed perceptually, significant differences from the NE subjects were evident in NI subjects who had begun learning English as early as the age of 9 years. Acoustically measured VOT differences between native and nonnative subjects emerged as early as 11 years. Also, NI subjects who had begun learning English as early as the age of 11 years were found to differ significantly from the NE subjects in terms of measures of segmental timing such as stop closure duration.

Given the NI subjects' vast experience with the English language, these findings might seem to disconfirm the hypothesis that late learners will eventually learn to produce in a completely nativelike fashion L2 consonants that either are not found in the native language inventory, or else differ phonetically from a counterpart in the L1 (see (Flege 1988, 1991a, 1992a, 1992b) for review and discussions). However, the results for word-
final stops are more supportive than they might seem at first. The NI subjects had little difficulty producing word-final tokens of /p t k/ even though previous studies have shown that speakers of an L1 without final stops may fail to do so (see (Flege et al., 1992; Flege, 1995)). Difficulty producing final stops was restricted to /b d g/, and many NI subjects who had begun learning English relatively late in life managed to produce these stops accurately.

The results obtained for word-final /t/ and /d/ were of special interest. Almost without exception, /t/ was produced accurately. Just a subset of NI subjects with AOLS greater than 15 years produced /d/ inaccurately (their /d/s were heard as /t/). One possible explanation for their devoicing of word-final /d/ tokens is that they had acquired awareness of the phonological distinction between English /t/ and /d/ but had not mastered the language-specific cues used in English to contrast phonologically voiced from voiceless stops. Another possibility is that, having passed a critical period, they were unable to learn the gestures needed to produce voiced word-final stops effectively.

Acoustic analyses suggested another explanation, however. The NI subjects who devoiced word-final /d/s managed to sustain closure voicing as long as did the NE speakers, and the duration of their preceding vowels were, for the most part, appropriate. It may be that syllable timing characteristics of Italian influenced the NI subjects' production of word-final /d/s. Their prolongation of /d/ closures may have resulted in the production of voiceless release bursts, which gave rise to the perception of devoicing by native English-speaking listeners. If this explanation is correct, it points to prosodic interference as the source of what at first appears to have been an error in segmental articulation.

Not all errors can have such an explanation, of course. For example, the cause of persistent errors in the NI subjects' production of word-initial English consonants remains uncertain. A wide range of hypotheses have been advanced in the literature for such errors (see (Flege, 1988, 1992a, 1992b)). These explanations include, but are not limited to, the following:

1. **Inadequate phonetic input.** With increasing age, the phonetic input received by L2 learners may become less adequate in terms of quantity or quality (e.g., hearing others speak the L2 with a foreign accent).

2. **Psychosocial factors.** With increasing age, psychosocial factors may increasingly inhibit L2 learners from wanting to sound like a native speaker; or, there may be an increasing desire to maintain allegiance to the L1-speaking community by producing the L2 with an L1 accent.

3. **Motivation.** With increasing age, motivation for learning to produce L2 phones (especially “difficult” ones) may decline if articulatory errors do not impede communication.

4. **Habit formation.** L2 learners, especially older ones, are often forced into using the L2 to communicate almost immediately. Given the challenge of producing many unfamiliar words in meaningful sentences, L1 phones may be substituted for L2 phones not found in the L1, especially if such substitutions (e.g., saying think as “tink”) do not lead to misperceptions. Once a “shortcut” articulatory pattern has been adopted, it might become habitual, or “fossilized”.

5. **Motoric difficulty.** Children require a number of years to learn to produce the phones of their L1 according to its language-specific phonetic norms. Some phones are more complex articulatorily than others, and are acquired relatively late by most children. Beyond a certain age, L2 learners may have difficulty at a motoric level in modifying pre-established patterns of articulation, or in learning new patterns of speech articulation. The difficulty might be especially evident for articulatorily complex phones.

6. **Incorrect perception.** During L1 acquisition, speech perception becomes attuned to the phonetic elements of the L1. L2 learners may fail to perceive accurately the phonetic details of L2 phones and contrasts owing to the assimilation of L2 phones by L1 categories, or to an inability to attend to acoustic properties (features) that are unimportant in the L1. If deprived of accurate perceptual “targets” to
guide sensorimotor learning of L2 phones, production of the L2 phones may necessarily be inaccurate.

7. **Phonetic system effects.** Bilinguals may have difficulty keeping two phonetic/phonological systems separate. If so, the expected effect of “system” pressure, which works to optimize the distinction between elements in a single system, would be expected to cause phones in both the L1 and L2 to diverge from monolingual norms.

We can probably rule out the first three hypotheses as explanations for why our NI subjects pronounced certain English consonants incorrectly. As stated earlier, the NI subjects examined here were likely to have reached their ultimate level of attainment in English pronunciation. Flege et al. (1995a) submitted questionnaire data obtained from the 240 NI subjects to a principle components analysis, then regressed the derived underlying factors onto production scores obtained for word-initial consonants. AOL accounted for 43% of the variance in the consonant production scores. Language use factors and motivation factors accounted for just 5% and 4% of variance, respectively. No other factors, such as a language loyalty factor, were identified as significant predictors.

An explanation for persistent L2 consonant production errors can probably be found among the remaining hypotheses listed above, either singly or in some combination. The easiest of the remaining hypotheses to reject (if indeed it is incorrect) will be the one that posits a perceptual cause for L2 articulation errors (#6). The type of perceptual difficulty we envisage is categorical rather than continuous in nature (see e.g. (Flege, 1991a, 1992a, 1992b, 1995)). It is unlikely that native and non-native speakers will differ in ability to discriminate L2 phones auditorily in a same/different (AX) task. However, L2 learners’ prior linguistic experience may nonetheless cause them to ignore certain properties of L2 phones that are phonetically relevant (e.g. (Yamada and Tohkura, 1992)). We hypothesize that an L2 phone must be perceived in a fully nativelike fashion if it to be produced in a fully nativelike fashion. (This assumes, of course, that the L2 differs acoustically from the nearest L1 phone.) On this view, one might attribute an NI subject’s production of English /t/ with Italian-like VOT values to a lack of recognition perceptually that long-lag realizations of English /t/ are distinct phonetically from short-lag realizations of Italian /t/ (see (Flege, 1991b)). To take another example, NI learners of English might realize /d/ as [d] if they fail to note the array of relevant phonetic differences between realizations of English /d/ and the closest Italian consonant (which is presumably /d/).

It is generally agreed that adult learners of an L2 are more analytic in their processing of the new language being acquired than are children learning their L1, and that adult learners give relatively greater emphasis to top-down as opposed to phonetically oriented bottom-up processing strategies. Also, adults are known to be language-specific perceivers of speech (e.g. (Flege and Hillenbrand, 1987; Best et al., 1988; Best and Stange, 1992)). Thus, we suspect that L2 research in the future will eventually show that a substantial proportion of L2 production errors have a perceptual basis. A perceptual error might arise, for example, from failure to attend to properties or features that are phonetically relevant in the L2 but not the L1 (Weinreich, 1957). For example, certain NI subjects examined in the present study may have devoiced word-final tokens of /bdg/ because they gave less weight than do NE subjects to preceding vowel duration, and relatively more weight to closure voicing cues, as the result maintaining the cue weighting pattern which typifies word-medial stop voicing distinctions in Italian.

To take another example, we know of no cross-language or L2 perception research that relates directly to the production of English by NI subjects. However, research with native French subjects is consistent with a perceptual explanation for why some NI subjects err in producing English /θ/ and /ð/. French, like Italian, lacks /θ/ and /ð/. Native speakers of Canadian French who learn English, like the NI subjects examined here, tend to substitute /d t/ for English /θ ð/. Morosan and Jamieson (1989) found that native Canadian French subjects had diffi-
difficulty in differentially identifying multiple natural tokens of English /ð/ and /θ/, and also tokens of /ð/ and /d/. The question of interest, then, is whether our NI subjects, had they been tested, would have shown difficulty in recognizing the phonetic distinction that exists between English /ð/ and either English /d/ or Italian /d/.

How might such a question be addressed? To be successful, an L2 perceptual test must be tailored to the kinds of errors actually observed in speech production (see Locke, 1980a, 1980b for discussion). Also, an L2 perceptual test should focus on the issue of whether or not an equivalence class of L2 phones (e.g., word-initial singleton /t/s) are judged to be the “same” as an equivalence class of phones in the L1. The critical issue is not whether L1 and L2 phones can be discriminated auditorily in an AX test, but whether L2 learners recognize that phonetic distinctions exist between certain L1 and L2 phones. The categorical oddity discrimination task developed by Flege et al. (1994a) is ideally suited for this purpose. It discourages within-category discrimination, and promotes the grouping of phones into phonetically relevant equivalence classes. Among other things, Flege et al. (1994) found that native speakers of Spanish, who often substitute /a/ for /æ/ in English, were unable to discriminate Spanish /a/ and English /æ/ tokens, although they readily discriminated triads containing vowels likely to map onto distinct L1 categories. For the perceptual hypothesis (#6 above) to receive support, one would need to show that subjects who produced an L2 phone inaccurately are unable to discriminate it categorically from phones used as a substitute for it. For example, one would need to show that the NI subjects who substituted /d/ for English /ð/ would perform poorly on triads containing /d/ and /ð/ tokens, whereas AOL-matched subjects who produced /ð/ correctly would perform well on such triads.

Perceptual barriers to accurate L2 segmental production, if they exist, should be relatively simple to document if the appropriate test is applied to the right phones. In the event that L2 learners’ consonant production errors can not be related to specific perceptual failures, it would be necessary to choose from among the remaining alternatives. One might reject the “motoric difficulty” hypothesis (#5 above), for example, by showing that the NI subjects who produced /ð/ as /d/ in the present experiment could produce /ð/ correctly when some other elicitation technique was used (e.g., an imitation task, or when asked to produce words “emphatically”). However, providing positive support for the hypothesis that production errors arise because of the loss of motoric ability, or the hypotheses concerning incorrect habit formation and negative psychosocial factors, may prove more difficult.

In summary, the present study showed that even highly experienced native Italian (NI) speakers of English may persist in producing certain English consonants inaccurately after having spoken English for several decades. The absence of an L2 consonant from the L1 inventory does not ensure ultimate mastery of L2 consonants as we had supposed earlier (e.g., Flege, 1988). Effects of age of learning (AOL) the L2 were noted for nearly every consonant examined. The AOL at which the NI subjects’ productions first diverged measurably from the NE subjects’ varied across the consonant allophones that were examined. It was not possible, on the basis of the data presented here, to choose from among a number of alternate hypotheses concerning the cause of AOL effects on L2 consonant production.

Acknowledgements

This research was supported by NIH grant DC00257. The authors are very grateful to Mrs. Jennie Prosperine and to the parishioners of St.
Anthony’s parish in Ottawa, Ontario, for their kind participation in our study. Thanks are also extended to Veena Reddy for testing listeners, to Jane Lesnick for making acoustic measurements, and to Naoyuki Takagi, Anna Schmidt and two anonymous reviewers for comments on a previous version of this article.

References


J. Flege, M. McCutcheon and S. Smith (1987), “The develop-


D. Williams (1971), “A test for differences between treatment means when several dose levels are compared with a zero dose control”, Biometrics, Vol. 27, pp. 103–117.
