Talker and listener effects on degree of perceived foreign accent

James Emil Flege and Kathryn L. Fletcher
Department of Biocommunication, University of Alabama at Birmingham, Birmingham, Alabama 35294-0019

(Received 11 September 1990; accepted for publication 8 August 1991)

Four experiments were carried out to examine listener- and talker-related factors that may influence degree of perceived foreign accent. In each, native English listeners rated English sentences for degree of accent. It was found that degree of accent is influenced by range effects. The larger the proportion of native (or near-native) speakers included in a set of sentences being evaluated, the more strongly accented listeners judged sentences spoken by non-native speakers to be. Foreign accent ratings were not stable. Listeners judged a set of non-native-produced sentences to be more strongly accented after, as compared to before, they became familiar with those sentences. One talker-related effect noted in the study was the finding that adults' pronunciation of an L2 may improve over time. Late L2 learners who had lived in the United States for an average of 14.3 years received significantly higher scores than late learners who had resided in the United States for 0.7 years. Another talker-related effect pertained to the age of L2 learning (AOL). Native Spanish subjects with an AOL of five to six years were not found to have an accent (i.e., to receive significantly lower scores than native English speakers), whereas native Chinese subjects with an average AOL of 7.6 years did have a measurable accent. The paper concludes with the presentation of several hypotheses concerning the relationship between AOL and degree of foreign accent.

PACS numbers: 43.70.Kv, 43.71.Hw

INTRODUCTION

Flege (1991a) recently examined voice onset time (VOT) in stop consonants produced by native speakers of Spanish who had learned English as a second language (henceforth, L2). Those who had learned English as adults—called late learners—produced phonologically voiceless stops with significantly longer VOT in English than in Spanish words. Their English stops nevertheless had significantly shorter VOT values than the stops produced by native speakers of English. A different pattern of results was observed for early learners who began learning English by the age of 5–6 years. They produced Spanish /t/ with appropriate short-lag VOT values and English /t/ with appropriate long-lag VOT values. In so doing, the early learners fully differentiated corresponding Spanish and English stop consonants in terms of VOT. One might infer from this that early L2 learners are able to fully separate the phonetic systems of their two languages, but such a conclusion would diverge from the view that “mixing” of the LI and L2 is inevitable because a bilingual's two language systems are both constantly engaged, at least to some extent (Grosjean, 1983, 1989). A mutual influence of the LI and L2 systems on one another has been observed for processing in the semantic domain (e.g., Lambert and Rawlings, 1969; Obler and Albert, 1978; Magiste, 1979; Mack, 1986), in the syntactic domain (e.g., Blair and Harris, 1981; Mack; 1986), and in the phonological domain (e.g., Altenberg and Cairns, 1983; Cutler et al., 1989). Moreover, a mutual influence of the LI and L2 phonetic systems has also been observed, at least for late learners (e.g., Flege, 1987a; Major, 1991).

The VOT finding obtained in the Flege (1991a) study does not mean, necessarily, that early L2 learners can speak both of their languages without a perceptible accent, or even that their production of phonetic segments in the LI and L2 is completely authentic. The Flege (1991a) study focused on a single phonetic segment (viz., /t/), the production of which was elicited in a formal list-reading experiment. Two languages may differ phonetically in terms of many different aspects of segmental and suprasegmental articulation of which many—perhaps all—may contribute to the perception of foreign accent (Flege, 1984, 1988a,b, 1991d). The early L2 learners who managed to differentiate English /t/ from the /t/ of Spanish in terms of VOT might have failed to produce the English /t/ with an alveolar, as opposed to a Spanish-like dental place of articulation. They may have failed to fully differentiate other “similar” English sounds from their Spanish counterparts, or to produce “new” English sounds (or suprasegmental dimensions) that have no direct counterpart in Spanish with complete authenticity. The primary purpose of the present study was therefore to determine if an L2 can be pronounced without a perceptible accent if it is learned early enough in life.

Age of L2 learning (AOL) is a variable that has figured prominently in the literature on second language acquisition. AOL refers to the chronological age at which an individual first begins receiving massive input from native speakers of an L2 in a naturalistic context and begins, slowly, to learn to speak and understand the L2. Research examining morphosyntactic aspects of language acquisition has shown that AOL affects ultimate attainment in both production and comprehension. Long (1990) suggested that native-like control of L2 morphosyntax is not possible if L2 learning begins after the age of 15 yrs. This view has received empirical support from a number of studies (e.g., Coppieters, 1987; Johnson and Newport, 1989). However, it is not yet cer-
tain—in part because of the problem of how to define morphosyntactic “norms” for the L2—at what AOL a less than fully nativelike control of L2 morphosyntax first becomes apparent. However, fine-grained measures have shown the existence of certain specific differences between native speakers and individuals who began learning the L2 well before puberty, perhaps as early as the age of 6 years (John-

Other studies have examined the acquisition of American Sign Language (ASL), which may be learned as a first language beyond the age at which oral languages are normally acquired by hearing children (Newport, 1984, 1990; Newport and Suppala, 1991; Mayberry and Fischer, 1989; Mayberry and Eichen, 1991). These studies have tended to focus on subjects who have used ASL for many years, and thus have reached their ultimate attainment in ASL. Their results mirror those obtained in studies of L2 syntax, including evidence of specific deficits in individuals who learned ASL long before puberty.

In experiment 1 of the present study we had native English listeners rate English sentences that had been spoken by native Spanish “early” and “late learners” for degree of perceived foreign accent. The sentences were spoken by subjects whose production of Spanish and English /t/ was examined previously by Flege (1991a). The native English listeners also evaluated sentences spoken by other native speakers of English. If the early L2 learners received significantly lower scores than the native English speakers, it would indicate the presence of a perceptible accent and so help define the age of L2 learning (AOL) at which non-native control of the phonology of an L2 first becomes evident.

It is well known that most adults who learn an L2 will speak it with an accent, but no previous study has attempted to identify the AOL at which foreign accents first emerge. A number of investigators have linked the emergence of foreign accents with the end of a “critical period” near the onset of puberty (e.g., Lenneberg, 1967; Scoval, 1988; Patkowski, 1989). Of course, the age at which puberty occurs varies. The age of 12 or 13 years is most often mentioned as the AOL, beyond which accent-free pronunciation of an L2 is no longer possible. Data presented by Patkowski (1980, 1989) suggested 15 yrs as the AOL demarcating nativelike and foreign-accented pronunciation of an L2.

In agreement with the syntactic studies mentioned earlier, other research has suggested that foreign accents might emerge long before puberty. Like adults, children may make errors in producing unfamiliar foreign language speech sounds (Lambert and MacNamara, 1969; Politzer and Weiss, 1969; Locke, 1969; Cochrane, 1980). Children may also have difficulty in correctly imitating foreign sounds (Olson and Samuels, 1973; Snow and Hoefnagel-Höhle, 1982; Ekstrand, 1982; Lowenthal and Bull, 1984; but cf. Tahta et al., 1981b). Not surprisingly, children may manifest a foreign accent when speaking an L2, at least initially. Seliger et al. (1975) surveyed 394 adults who had learned English or Hebrew as an L2. Of those who began learning their L2 by the age of 10 years, 8% reported speaking it with an accent. Flege (1988b) had native English listeners rate English sentences spoken by native and non-native speakers for degree of foreign accent. Even though they spoke English fluently, Chinese adults who began learning English at an average age of 7.6 years were found to have a perceptible accent. Tahta et al. (1981b) found that 68% of subjects who began learning English between the ages of 7–12 years had some degree of accent. Of the subjects examined by Patkowski (1989), 54% of those who began learning English between the ages of 5–15 years appear to have had perceptible accents.

Taken together, the results of the studies just cited provide counterevidence to the widespread belief that children typically pronounce foreign languages without an accent. Most subjects examined had lived for many years in a predominantly L2-speaking environment, the minimum period being 2 years. The Chinese early learners examined by Flege (1988b), for example, had lived in the United States for an average of 12 years, so their foreign accents were probably not due to a lack of L2 experience. These results thus raise the issue of whether it is possible for any bilingual to fully separate the sound systems of the L1 and L2.

Many investigators accept that a critical period exists for the learning of pronunciation, either that of an L2 (Scoval, 1969) or a dialect of the L1 (see Payne, 1980). Long (1990) suggested recently that a sensitive period for speech learning occurs at about the age of 6 years, not at puberty, as proposed by other investigators (e.g., Lenneberg, 1967; Scoval, 1988; and Patkowski, 1989). If so, then accent-free L2 pronunciation might be possible if the L2 is learned in early childhood. In support of this, Thompson (1984) found that two Russian adults who began learning English at the age of 4 years apparently did not have perceptible accents. The study by Tahta et al. (1981a) included ten subjects who began learning English at the age of 6 years. None of these individuals was judged to speak with a foreign accent. As mentioned earlier, Seliger et al. (1975) reported evidence that some individuals who learn an L2 in childhood may speak it with an accent. The authors suggested, however, that these individuals may have had accents because they had not received sufficient L2 input.

A study by Asher and Garcia (1969), on the other hand, provided counterevidence to the hypothesis that early L2 learners may evade accent detection. Of the 30 native English children examined, 23 were correctly judged to be native speakers of English. Not one of 71 native Spanish (Cuban) children who were examined were so identified, including individuals who had arrived in the United States between the ages of 1–6 years. One might argue, of course, that the native English listeners’ identification of seven native English children as non-native indicated the presence of a response bias that could have resulted in the incorrect classification of some unaccented Cuban children as accented. A four-point scale was used to classify each talker (“native speaker,” “near native speaker,” “slight foreign accent,” “definite foreign accent”). Perhaps some non-native speakers in the Asher and Garcia (1969) study would not have been found to differ from the native English speakers had a more fine-grained scaling procedure been used. Indeed, our previous research has suggested that listeners can resolve far more than four degrees of foreign accent (Flege and Eefting,
Experiment 1 of the present study made use of a fine-grained technique for assessing degree of accent in English sentences spoken by native Spanish speakers who differed according to age of L2 learning (AOL). In addition to comparing early versus late L2 learners, we compared late learners who differed according to length of residence (LOR) in the United States. We expected most if not all of the late learners to speak with an accent, but it was uncertain from previous research if those who had lived in the United States for many years would have a lesser degree of accent than newly arrived late learners.

Additional experiments were undertaken to provide insight into factors that may influence non-native talkers' degree of foreign accent and listeners' assessment of their foreign accent. More specifically, the purpose of experiment 2 was to help pinpoint the AOL at which a foreign accent first emerges and to test the hypothesis that foreign accent judgments are subject to range effects. Experiment 3 examined inter- and intrasubject reliability of foreign accent judgments. In experiment 4, listeners assessed degree of foreign accent in sentences produced by talkers in three groups: native speakers of English, native speakers of Spanish who began learning English by the age of 5–6 years, and native speakers of Chinese who began learning English at an average age of 7.6 years.

I. EXPERIMENT 1

The primary purpose of experiment 1 was to determine if native Spanish subjects who learned English as children can produce English sentences without a foreign accent. Native English listeners rated sentences spoken by "early L2 learners" who started learning English at the age of five or 5–6 years. The listeners also rated sentences spoken by two groups of native Spanish "late learners" who started learning English as adults. The sentences were read from a list, and thus may have represented the non-natives' optimal pronunciation of English. If sentences produced by the early learners received significantly lower ratings than sentences produced by native English speakers, it would support a "constant dual activation" hypothesis (Grosjean, 1985, 1989). Such a finding would imply that cross-language phonetic interference may be evident no matter how early an L2 is learned.

Experiment 1 also assessed the role of amount of L2 experience on adults' pronunciation of an L2. Selinker (1972) observed that adults' pronunciation of an L2 may "fossilize," which suggests that further improvement in the pronunciation of an L2 does not occur beyond a certain point (see also Scovel, 1988). Flege (1988b) found that the degree of accent of experienced and inexperienced Chinese late learners did not differ significantly, suggesting that additional L2 experience may not lead to a measurable improvement in L2 pronunciation. In experiment 1, a group of late learners consisting of individuals who had lived in the United States for less than 1 year was compared to a group of late learners who had lived in the United States for more than 7 years. The two groups differing in length of residence (LOR) in the United States will be referred to as the "inexperienced" and "experienced" late learners. The LOR difference between the two Spanish groups examined here was substantially larger than the LOR difference between the two Chinese groups examined by Flege (1988b).

A. Methods

1. Speech materials

The talkers were seated in a sound booth when their speech was recorded using portable equipment (Marantz Model PMD420). They read a list of sentences that included five tokens of each of The good shoe fits Sue, I can read this for you, and The red book was good. These sentences will be referred to here as the "Sue," "Read," and "Book" sentences. We made use of sentences rather than paragraph-length materials as in some previous experiments (e.g., Oyama, 1976) because they more nearly approximate the short utterances typical of conversational speech, and because sentences fit our testing format (see below) better than paragraphs.

The three English sentences used here were employed in two previous foreign accent experiments (Flege and Eefting, 1987a; Flege, 1988b). They were not designed specifically to contain sounds and sound sequences that would be especially difficult for native speakers of Spanish. (Had this been our goal, we would have chosen sentences with many interdental fricatives, lax vowels, word-final singleton obstruents, and consonant clusters.) Still, the sentences examined did contain a number of vowels (/æ/, /I/, /u/, /a/, /e/, /3/), consonants (/t/, /d/, final /z/), and a consonant cluster (/ts/) that are known to be mispronounced by Spanish learners of English (Ornstein, 1974; Brennan and Brennan, 1981; Dowd, 1984; Hammond, 1986; MacDonald, 1989). The second (or, in a few instances third) token of each sentence was bandpass filtered (60–8000 Hz) and digitized at 20 kHz with 12-bit amplitude resolution. The waveforms were normalized for peak intensity and stored on disk for later on-line presentation to listeners.

2. Talkers

The English sentences were spoken by five groups of paid participants, each consisting of five males and five females. One group consisted of native speakers of English affiliated with the University of Alabama at Birmingham (mean age = 26 years; s.d. = 4). Some of the native Spanish subjects in this study (see below) had learned English primarily in Birmingham. Others had learned English in Texas. Acoustic analyses have shown that early learners closely match the speech patterns of the L2 native speakers to whom they have been exposed in childhood (Flege and Eefting, 1987b; Flege, 1991b,e). If the Spanish early learners we recorded in Texas had learned a Texas dialect of English perfectly, and if the native English-speaking listeners from Birmingham who later rated their sentences misinterpreted a dialect difference as foreign accent (see Giles, 1972), one might falsely conclude that the early learners had a foreign accent. To guard against this, we recorded a second native English group in Austin, Texas. Most of these individuals (mean age = 24 years; s.d. = 7) had been born and raised in Texas; all were students at the University of Texas. The sub-
jects in both native English groups were monolinguals who reported no history of speech or language learning difficulty.

The three native Spanish groups consisted of individuals who had all learned English as an L2. None of them spoke a third language. All of the bilinguals produced materials in both English and Spanish in counterbalanced order (see Flege, 1991a). The native English speakers produced only the English speech materials. Data was elicited from the native Spanish speakers in the appropriate language by the same bilingual research assistants who elicited data, in English, from the native English speakers.

Characteristics of the three native Spanish groups are summarized in Table I. Subjects in the "early learner" group (mean age = 23 years) were first exposed to English as young children. They reported being unable to speak English prior to attending an elementary school in Texas at the age of five or six years. Since early learners exposed to Spanish-accented English as young children may not pronounce English authentically (Flege and Eefting, 1987b), an important criterion for inclusion in the early L2 group was exposure in childhood to native-produced English. All of the early learners attended elementary schools in which English was the sole language of instruction. They all reported having been taught by native English-speaking teachers in the first three primary grades and/or having a majority of native English classmates in those grades. All of the early learners were recorded in Texas.

We considered age of L2 learning (AOL) to be the most important variable affecting degree of accent (see below). AOL and age of arrival in the United States (AOA) were the same for the late learners we examined. (This assumes, of course, that they began learning English immediately upon arriving in the United States.) There was a difference between AOL and AOA for some of the early learners, how-

### TABLE I. Characteristics of talkers in three native Spanish groups whose English sentences were rated for degree of foreign accent by native English listeners in experiment 1. Standard deviations are in parentheses.

<table>
<thead>
<tr>
<th>Sub.</th>
<th>Age</th>
<th>Sex</th>
<th>POB</th>
<th>EDU</th>
<th>AOA</th>
<th>AOL</th>
<th>LOR</th>
<th>USE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early L2 learners</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>23</td>
<td>F</td>
<td>Edinburgh, TX</td>
<td>12</td>
<td>0</td>
<td>5-6</td>
<td>23</td>
<td>90</td>
</tr>
<tr>
<td>2</td>
<td>26</td>
<td>M</td>
<td>Mexico</td>
<td>12</td>
<td>6</td>
<td>5-6</td>
<td>20</td>
<td>70</td>
</tr>
<tr>
<td>3</td>
<td>19</td>
<td>M</td>
<td>Guadalajara</td>
<td>13</td>
<td>5</td>
<td>5-6</td>
<td>14</td>
<td>80</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>F</td>
<td>Farr, TX</td>
<td>13</td>
<td>0</td>
<td>5-6</td>
<td>20</td>
<td>80</td>
</tr>
<tr>
<td>5</td>
<td>23</td>
<td>F</td>
<td>Dallas, TX</td>
<td>12</td>
<td>0</td>
<td>5-6</td>
<td>23</td>
<td>90</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>F</td>
<td>Mexico</td>
<td>11</td>
<td>3</td>
<td>5-6</td>
<td>15</td>
<td>80</td>
</tr>
<tr>
<td>7</td>
<td>23</td>
<td>M</td>
<td>Mexico</td>
<td>13</td>
<td>1</td>
<td>5-6</td>
<td>22</td>
<td>75</td>
</tr>
<tr>
<td>8</td>
<td>21</td>
<td>F</td>
<td>Mexico</td>
<td>13</td>
<td>0</td>
<td>5-6</td>
<td>21</td>
<td>90</td>
</tr>
<tr>
<td>9</td>
<td>24</td>
<td>M</td>
<td>Laredo, TX</td>
<td>16</td>
<td>0</td>
<td>5-6</td>
<td>24</td>
<td>80</td>
</tr>
<tr>
<td>10</td>
<td>26</td>
<td>M</td>
<td>Taft, TX</td>
<td>13</td>
<td>0</td>
<td>5-6</td>
<td>26</td>
<td>85</td>
</tr>
<tr>
<td>M</td>
<td>23</td>
<td></td>
<td>(3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1)</td>
<td></td>
<td></td>
<td>(2)</td>
<td></td>
<td>(0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(4)</td>
</tr>
<tr>
<td>Experienced late L2 learners</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>28</td>
<td>M</td>
<td>Mexico</td>
<td>0</td>
<td>18</td>
<td>18</td>
<td>10</td>
<td>70</td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>M</td>
<td>Mexico</td>
<td>6</td>
<td>13</td>
<td>13</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>37</td>
<td>F</td>
<td>Mexico</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>19</td>
<td>F</td>
<td>Mexico</td>
<td>7</td>
<td>11</td>
<td>11</td>
<td>8</td>
<td>70</td>
</tr>
<tr>
<td>5</td>
<td>41</td>
<td>M</td>
<td>Chile</td>
<td>6</td>
<td>24</td>
<td>24</td>
<td>17</td>
<td>73</td>
</tr>
<tr>
<td>6</td>
<td>38</td>
<td>F</td>
<td>Equador</td>
<td>12</td>
<td>14</td>
<td>14</td>
<td>24</td>
<td>90</td>
</tr>
<tr>
<td>7</td>
<td>53</td>
<td>F</td>
<td>Panama</td>
<td>0</td>
<td>35</td>
<td>35</td>
<td>18</td>
<td>95</td>
</tr>
<tr>
<td>8</td>
<td>28</td>
<td>M</td>
<td>Nicaragua</td>
<td>7</td>
<td>18</td>
<td>18</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>9</td>
<td>33</td>
<td>M</td>
<td>Guatemala</td>
<td>5</td>
<td>26</td>
<td>26</td>
<td>7</td>
<td>75</td>
</tr>
<tr>
<td>10</td>
<td>40</td>
<td>F</td>
<td>Chile</td>
<td>4</td>
<td>26</td>
<td>26</td>
<td>14</td>
<td>35</td>
</tr>
<tr>
<td>M</td>
<td>34</td>
<td></td>
<td>(4)</td>
<td></td>
<td></td>
<td>(8)</td>
<td></td>
<td>(8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(4)</td>
<td></td>
<td></td>
<td>(8)</td>
<td></td>
<td>(7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1)</td>
<td></td>
<td></td>
<td>(2)</td>
<td></td>
<td>(0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(4)</td>
<td></td>
<td></td>
<td>(8)</td>
<td></td>
<td>(7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(19)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inexperienced late L2 learners</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>25</td>
<td>F</td>
<td>Colombia</td>
<td>10</td>
<td>24</td>
<td>24</td>
<td>0.7</td>
<td>77</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>M</td>
<td>Colombia</td>
<td>7</td>
<td>30</td>
<td>30</td>
<td>0.3</td>
<td>67</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>M</td>
<td>Equador</td>
<td>6</td>
<td>30</td>
<td>30</td>
<td>0.8</td>
<td>70</td>
</tr>
<tr>
<td>4</td>
<td>23</td>
<td>M</td>
<td>Puerto Rico</td>
<td>7</td>
<td>22</td>
<td>22</td>
<td>0.8</td>
<td>47</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>M</td>
<td>Puerto Rico</td>
<td>6</td>
<td>19</td>
<td>19</td>
<td>0.8</td>
<td>45</td>
</tr>
<tr>
<td>6</td>
<td>31</td>
<td>M</td>
<td>Venezuela</td>
<td>5</td>
<td>31</td>
<td>31</td>
<td>0.5</td>
<td>50</td>
</tr>
<tr>
<td>7</td>
<td>18</td>
<td>F</td>
<td>Chile</td>
<td>6</td>
<td>17</td>
<td>17</td>
<td>1.0</td>
<td>67</td>
</tr>
<tr>
<td>8</td>
<td>35</td>
<td>F</td>
<td>Mexico</td>
<td>10</td>
<td>35</td>
<td>35</td>
<td>0.2</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>25</td>
<td>F</td>
<td>Venezuela</td>
<td>5</td>
<td>24</td>
<td>24</td>
<td>0.8</td>
<td>16</td>
</tr>
<tr>
<td>10</td>
<td>26</td>
<td>M</td>
<td>Mexico</td>
<td>2</td>
<td>25</td>
<td>25</td>
<td>1.0</td>
<td>87</td>
</tr>
<tr>
<td>M</td>
<td>26</td>
<td></td>
<td>(5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(2)</td>
<td></td>
<td></td>
<td>(6)</td>
<td></td>
<td>(0.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(27)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a Age, chronological age at testing, in years.
b POB, place of birth.
c EDU, formal education in English, in years.
d AOA, age of arrival in the United States, in years.
e AOL, age of L2 learning, in years.
f LOR, length of residence in the United States, in years.
g USE, self-estimated percentage daily use of English.
ever. Four early learners were born in Mexico, where Spanish is the predominant language, and six were born in United States towns located near the Mexican border. The early learners born in the United States are listed in Table I as having an AOA of "O." Since they reported not being able to speak English when they went to school, their AOL is listed as 5–6 years. More detailed information about the early learners' language history is unavailable. It is at least possible that the early learners who were born in the United States or arrived there prior to school age were passively exposed to English—in conversations heard on the street and on television and radio—prior to the age of 5–6 years. Tees and Werker (1984) found that adults exposed to Hindi in the first two years of life were better able to discriminate Hindi sounds than individuals who had not received such early exposure. Early exposure to English, if it did occur, might have improved the early learners' pronunciation of English.

Two other native Spanish groups consisted of individuals who began learning English as adults. The two late learner groups differed primarily according to length of residence (LOR) in the United States. The "experienced" late learners had lived in the United States for an average of 14.3 years at the time of testing; those in the "inexperienced" group had lived in the United States only 0.7 years, on the average. LOR has often been used in previous studies to provide an estimate of amount of L2 experience. We can be sure that the experienced late learners had received considerably more L2 input than the inexperienced ones. However, the exact relationship between LOR and amount of native speaker input is uncertain. As alluded to above, one might live in a predominantly L2-speaking environment without interacting with L2 native speakers. Even for those L2 learners who interact routinely with native speakers of the L2, the relationship between LOR and amount of native-speaker input may not be linear. One might suppose, for example, that the better one speaks the L2 the more the L2 is likely to be used on a daily basis.

The two late learner groups differed little according to several variables represented on a language-background questionnaire. Compared to the inexperienced late learners, the experienced late learners reported using English somewhat more often on a daily basis (75% vs 68%). They had received about one year more of formal instruction in English than the inexperienced late learners, and had arrived in the United States at a somewhat earlier age (20 vs 26 years). Four of the experienced late learners were recorded in Austin, six in Birmingham. All of the inexperienced late learners were recorded in Birmingham.

3. Listeners

The English sentences were rated for degree of foreign accent by ten monolingual native speakers of American English (eight female, two male). These listeners, who were recruited through newspaper advertisements in Birmingham, Alabama, had a mean age of 28 years (s.d. = 6). None of them were judged to speak English with a marked regional accent. They were drawn from the same community as many of the native Spanish speakers, and are thus likely to have spoken a variety of English approximating the "target" L2 of the native Spanish speakers. Familiarity with foreign accents was not a criterion in selecting the listeners, but they were all required to pass a pure-tone hearing screening (500–4000 Hz at 20-dB HL). The listeners in this experiment, as well as all those in subsequent experiments, were paid to participate.

4. Procedures

The technique used here to assess degree of foreign accent was probably better suited to listeners' ability to resolve differences in degree of accent than the techniques employed in some previous studies (e.g., Asher and Garcia, 1969; Oyama, 1976; Tahta et al., 1981a). The listeners positioned the lever on a response box after hearing each sentence to indicate degree of perceived accent. The lever was connected to a linear potentiometer, which was connected to an 8-bit A/D converter. Depending on the lever's position, a value ranging from 1–256 was returned. The number of gradations of foreign accent that listeners can reliably resolve is unknown. Although it is unlikely to be as many as 256, it is surely greater than the three to nine degrees of resolution implied by the three- to nine-point rating scales used in previous studies. We felt that it was prudent to err in the direction of using too fine a scale than in using too coarse a scale. This is because we were testing the hypothesis that early learners do not have a perceptible accent. If a negative finding were obtained with a seven-point scale, for example, it might be due to the use of too coarse a scale.

The English sentences were presented binaurally over headphones at a comfortable level of about 75-dB SPL(A) peak syllable intensity. The listeners were told that they would hear sentences spoken by an unspecified proportion of native and non-native talkers. The listeners were instructed to indicate the degree of foreign accent in each sentence by positioning the response box lever at some place along its 7-cm range. The scale was defined by the labels "no foreign accent," at one end; "medium foreign accent," at the middle of the scale; and "strong foreign accent," at the other end. The better a sentence was judged to have been pronounced, the higher the rating it received (up to a maximum of 256). The listeners were not told the background of the non-native speakers. They were told to use the whole scale, and to guess if uncertain.

Each sentence was presented 1.0 s after a response was received for the preceding sentence. The three sentences were presented in separate, counterbalanced blocks, each lasting about 14 min. Within a block, each sentence was presented three times, yielding a total of 450 judgments per listener (five groups X ten talkers X three sentences X three repetitions). The first presentation of each sentence in each block was not analyzed. This was to ensure that listeners were familiar with the entire range of possible accents before giving responses that were actually analyzed. An average was computed from the remaining two ratings given to each sentence by each listener. An average value for each sentence spoken by the 50 talkers was then computed based on the ten listeners' mean values. This resulted in 150 foreign accent scores (five groups X ten talkers X three sentences), each
based on a total of 20 ratings (ten listeners × two judgments). The mean foreign accent scores were submitted to a (5) group × (3) sentence mixed-design ANOVA.

B. Results

1. Effects of age of L2 learning and length of residence in the United States

The mean foreign accent scores obtained for the two native English groups and the three native Spanish groups are shown in Fig. 1. The early learners' scores closely resembled the scores accorded sentences spoken by the native speakers of English from Birmingham and Austin (242 vs 246 and 250). As expected, the experienced and inexperienced late learners' scores were considerably lower (91 and 44, respectively) than those of the native English speakers, which resulted in a significant main effect of group [F(4,45) = 116.5, p < 0.05]. Newman–Keuls post-hoc tests revealed that the early L2 learners did not differ significantly from either native English group (p < 0.05). Both groups of late learners, on the other hand, had significantly lower scores than both native English groups (p < 0.05). The experienced late L2 learners nevertheless had significantly higher (i.e., more authentic, nativelike) scores than the inexperienced late learners.

Fig. 1 shows much the same pattern of between-group differences for the Sue, Read, and Book sentences. Despite this, the ANOVA yielded a significant group × sentence interaction [F(8,90) = 3.24, p < 0.05]. The simple main effect of group was significant for all three sentences (p < 0.05). Both groups of late learners, on the other hand, had significantly lower scores than both native English groups (p < 0.05). The experienced L2 learners nevertheless had significantly higher (i.e., more authentic, nativelike) scores than the inexperienced late learners.

2. Factors affecting degree of foreign accent

The results presented thus far are consistent with the view that the age of L2 learning (AOL) is an important determinant of degree of accent. Previous research has indicated that factors other than AOL may also influence degree of accent, although perhaps to a lesser extent. Correlation techniques were used to determine the relationship between degree of foreign accent and variables from the language background questionnaire administered to the 30 Spanish subjects. We chose to examine age of arrival in the United States rather than age of L2 learning because there was more variation in age of arrival than in age of L2 learning (always 5–6 years) for the early L2 learners. Also, the age of arrival in the United States of the 30 native Spanish subjects was highly correlated with their age of L2 learning (r = 0.989). Table II presents the simple correlations between degree of accent and the questionnaire variables. Degree of accent was not found to be correlated with percentage daily use of English, gender, and chronological age. However, degree of accent was correlated significantly, at a per-experiment error rate of 0.05, with years of formal English-language instruction, age of arrival in the United States, and length of residence in the United States. These correlations indicate that the more formal education in English the Spanish subjects received, the earlier they arrived in the United States,

![Fig. 1. Mean foreign accent (FA) scores given by native English listeners to three English sentences that had been spoken by talkers who were native speakers of English or learners of English as an L2 (ten per group). Talkers in the two native English groups were from Birmingham, AL or Austin, TX. Talkers in the "early L2" groups learned English as children, whereas those in two "late L2" groups learned English as adults. Each mean is based on 200 observations (ten talkers × ten listeners × two replicate judgments); the brackets enclose +/− one standard error.](image-url)
and the longer they had lived in the United States, the more native-like was their pronunciation of English. The significant correlation between degree of accent and age of arrival in the United States was hardly surprising. (As noted above, age of arrival was correlated with the age of L2 learning.) Age of arrival was also correlated with length of residence in the United States. As noted earlier, experienced late learners received higher scores than inexperienced late learners. The number of years of English-language instruction was correlated with degree of accent almost as strongly as age of arrival in the United States. It was also correlated with length of residence in the United States.

A forward stepwise multiple regression analysis was carried out to explore the relationship between degree of accent and the six questionnaire variables. A one-factor model with age of arrival (AOA) accounted for a significant 79.8% of the variance in the foreign accent scores ($p < 0.05$). A model with both AOA and number of years of English-language instruction (EDU) accounted for 85.0% of the variance. None of the four remaining factors were identified as significant predictors of degree of accent. We further explored the possibility that amount of L2 phonetic input might be related to degree of accent by deriving a variable called "exposure" (EXPOS). The values for EXPOS were calculated by multiplying length of residence in the United States by self-estimated daily use of English. EXPOS replaced both the length of residence and percentage daily use of English variables in a second multiple regression. The second analysis yielded a two-factor model with AOA and EDU that accounted for slightly more variance (86.7%) than the earlier two-factor model. The exposure variable was not identified as a significant predictor of degree of accent, however ($p < 0.05$).

### C. Discussion

Experiment 1 showed that native speakers of Spanish who began learning English at the age of 5–6 years produced three English sentences without a perceptible foreign accent. Half of the subjects had been speaking Spanish with a fellow native speaker of Spanish only moments before recording the English speech materials (see Flege, 1991a). If concurrent activation of the L1 phonetic system were likely to influence early learners' pronunciation of an L2 (Grosjean, 1989), then we would have expected to see an effect on these subjects' pronunciation of the English sentences (if not that of all ten early learners).

The present results do not rule out the concurrent activation hypothesis, however, for we can think of several reasons why a Spanish accent was not observed. First, the early learners may have been able to suppress an effect of the L1 phonetic system on their pronunciation of L2 because they were merely asked to read sentences from a list. That is, the L1 and L2 systems may have been activated concurrently, but such activation may be evident only in less "guarded" and/or more spontaneous speech. Second, the L1 phonetic system may have influenced the early learners' production of the English sentences, but it was not auditorily perceptible to our listeners. We cannot rule out the possibility that differences between early learners and native speakers of English would have been evident in a fine-grained acoustic or physiological study. Third, although the L1 system did not perceptually influence the early learners' pronunciation of their L2, an influence of the L2 on the L1 might have been evident had we examined their pronunciation of Spanish. The data presented here do not allow us to assess these hypotheses. Clearly, more work is needed to test the constant dual activation hypothesis of Grosjean (1989).

The apparent absence of foreign accent in the early learners' production of the English sentences was probably not due to the method we used to assess degree of foreign accent. Flege (1988b) used the same method and sentences in a study examining native Chinese speakers who began learning English at an average age of 7.6 years. These early learners received significantly lower ratings than did one of the two groups of native English speakers examined in the present study (viz., the Birmingham group). Thus, when the results of experiment 1 are taken together with the results of the Flege (1988b) study, one might conclude that a foreign accent first becomes perceptible at an age of L2 learning (AOL) of roughly 5 to 8 years.

This conclusion should be regarded with caution, however, for several reasons. First, we examined the production of only three sentences, all read from a list. The results might not generalize to longer passages of read speech or to conversational speech. Second, age of L2 learning (AOL) was

### TABLE II. Simple correlations between degree of perceived foreign accent (ACC) and six variables derived from a questionnaire administered to the 30 native Spanish subjects who participated as talkers in experiment 1 (see the text). An asterisk indicates significance at a per-experiment error rate of 0.05 (that is, $p < 0.0023$).

<table>
<thead>
<tr>
<th></th>
<th>ACCa</th>
<th>Ageb</th>
<th>Sex</th>
<th>EDUb</th>
<th>AOAc</th>
<th>LORd</th>
<th>USEe</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACCa</td>
<td>1.000</td>
<td>-0.432</td>
<td>0.087</td>
<td>0.827*</td>
<td>-0.893*</td>
<td>0.736*</td>
<td>0.431</td>
</tr>
<tr>
<td>Ageb</td>
<td>...</td>
<td>1.000</td>
<td>0.090</td>
<td>-0.425</td>
<td>0.595*</td>
<td>0.116</td>
<td>-0.083</td>
</tr>
<tr>
<td>Sex</td>
<td>...</td>
<td>...</td>
<td>1.000</td>
<td>0.143</td>
<td>-0.103</td>
<td>0.191</td>
<td>-0.016</td>
</tr>
<tr>
<td>EDUb</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>1.000</td>
<td>-0.761*</td>
<td>0.582*</td>
<td>0.254</td>
</tr>
<tr>
<td>AOAc</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>1.000</td>
<td>-0.728*</td>
<td>-0.481</td>
<td>0.524</td>
</tr>
<tr>
<td>LORd</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

* ACC, degree of perceived foreign accent.  
* Age, chronological age, in years.  
* AOA, age of arrival in the United States, in years.  
* EDU, formal education in English, in years.  
* LOR, length of residence in the United States, in years.  
* USE, self-estimated percentage daily use of English.
confounded with native language. If the sentences we examined contained sounds and/or suprasegmental dimensions that are more difficult for Chinese than Spanish learners of English—something we cannot quantify at this time—it could explain why the Chinese subjects in the Flege (1988b) study, but not the Spanish subjects examined in the present experiment, had perceptible accents. Finally, the AOL variable used to distinguish the Chinese and Spanish early learners was inexact. As noted earlier, it is possible that some of the Spanish early learners were exposed to English prior to their nominal AOL, that is, before the age of 5–6 years when they enrolled in an English-speaking school. If such early (possibly only passive) exposure occurred, it may have contributed to the early learners' accent-free pronunciation of English. It will be necessary in future studies examining the relationship between AOL and degree of accent to probe early childhood language experience in greater detail. This may require eliciting questionnaire data from older relatives and friends of the subjects being tested.

We can probably rule out the use of different groups of native English listeners as an explanation for why the Chinese early L2 learners in the Flege (1988b) study, but not the Spanish early L2 learners who were examined in the present experiment, had a perceptible accent. We obtained the same results in the present study when mean foreign accent scores were calculated in two ways (i.e., based on the responses given by each of ten listeners to all of the talkers in a group, and based on an average of scores given by all ten listeners to each talker). Also, the native English listeners who participated in the Flege (1988b) study and in experiment 1 performed the accent scaling task in much the same way. The two sets of ratings of sentences spoken by the native English speakers from Birmingham differed very little (means of 248.2 vs 246.3; standard deviations of 13.7 vs 16.4). Also, ratings the two listener groups gave were strongly correlated with one another (r = 0.925, df = 28, p < 0.001), even though the native produced sentences they rated represented only a small portion of the scale (i.e., scores ranging from 164 to 256).

A second important finding of experiment 1 was the evidence it provided that late learners' pronunciation of English may improve as a function of length of residence (LOR) in the United States. Late learners who had lived in the United States for an average of 14.3 years received higher ratings than those who had lived there for only 0.7 yr. Although adults' pronunciation of an L2 may fossilize (Selinker, 1972), it appears that some improvement is possible. Flege (1988b), on the other hand, did not observe a significant difference in the degree of accent of talkers in two Chinese late learner groups who had lived in the United States for 1.1 and 5.1 years on the average. A difference between the two Chinese groups may not have been evident because the LOR difference between them was not as large as that between the two Spanish groups examined here. If so, one might infer that late learners' pronunciation of English improves so slowly that an improvement will be evident after 14 years, but not after just 5 years. This interpretation is counterintuitive, however, for most improvement in L2 pronunciation seems to take place in the first year of L2 learning (see (see Baci, 1956, p. 217). We suspect that the Spanish but not the Chinese subjects showed an LOR effect because the inexperienced Spanish subjects had lived for a shorter period of time in the United States than the inexperienced Chinese subjects (0.7 vs 1.1 years). The inexperienced Chinese subjects may therefore have come closer to their ultimate level of L2 pronunciation than the inexperienced Spanish subjects.

Finally, multiple regression analyses provided some insight into factors that may affect degree of foreign accent. As expected, age of arrival in the United States—which was correlated with the age of L2 learning—was identified as the most important predictor of degree of accent. In agreement with Purcell and Suter (1980; see also Cochrane, 1980), percentage daily use of English was not a significant predictor. The number of years of formal education in English was identified as a significant predictor, although the amount of variance it accounted for was small (viz., 5%). In our study, the more education in English the non-native speakers had received, the better they pronounced English. Suter (1976) noted the opposite relation, perhaps because he examined spontaneous conversations (see Oyama, 1976; Dowd, 1984). The finding that gender was not a significant predictor of foreign accent agrees with results obtained by Purcell and Suter (1980) but not Asher and Garcia (1969) or Thompson (1984).

The simple correlation between length of residence in the United States (LOR) and degree of accent was significant. LOR was not identified as a significant predictor of degree of foreign accent, however. This agrees with some previous studies (Oyama, 1976; Tahta et al., 1981a; Thompson, 1984) but not others (Asher and Garcia, 1969; Purcell and Suter, 1980; see also Gras, 1983). From this and other studies (Seliger et al., 1975; Fathman, 1975; Oyama, 1976; Purcell and Suter, 1980; Patkowski, 1980; Tahta et al., 1981a; Thompson, 1984; Flege and Eefting, 1987a; Flege, 1988b) one might conclude that LOR is a less important determinant of degree of accent than age of arrival in an L2-speaking environment (or age of L2 learning, see above). While this may be true, such a conclusion cannot be drawn validly from the present experiment because an insufficiently wide range of values for the length or residence (LOR) and the age of L2 learning (AOL) variables were available. The design of experiment 1 called for the comparison of extreme groups differing in AOL and LOR. As for AOL, most subjects had begun learning English either before the age of 7 or after the age of 20 years. As for LOR, most subjects had lived in the United States for less than one year or more than ten years.

A consideration of previous studies suggest two conclusions regarding the relative importance of age of L2 learning (AOL) and length of residence (LOR) on degree of accent. It appears that in those studies that have not yielded a significant LOR effect the range of LOR values was too narrow, only individuals who began learning the L2 as adults were included, or both. The strength of the relationship between AOL and degree of perceived accent also seems to depend on the range of AOL values sampled in a study (e.g., Seliger et al., 1975; Oyama, 1976; Tahta et al., 1981a; Thompson, 1984).
AOL has usually been identified as an important predictor of degree of foreign accent. However, in a study by Purcell and Suter (1980) in which AOL was not a significant predictor, most subjects had apparently learned English after about the age of 12 years. A significant relationship between degree of accent and AOL may be evident only when the population being sampled includes child learners in addition to adult or adolescent learners.

II. EXPERIMENT 2

The primary purpose of experiment 2 was to test the hypothesis that range effects will influence listeners' judgments of foreign accent. Range effects are known to influence judgments of many aspects of speech. This includes the identification of consonants, which are often thought to be perceived in a categorical—and absolute—fashion. Identification judgments may be affected by the relative frequency with which the stimuli making up a continuum are presented, or by changes in the range of stimuli in the continuum (Lisker, 1970; Ades, 1977; Brady and Darwin, 1978; Rosen, 1979). For example, Keating et al. (1981) found that the location of Polish subjects’ /da/-/ta/ phoneme boundaries differed as a function of the range of VOT values used in synthesizing perceptual continua. The subjects required longer VOT values before giving predominantly /t/ judgments for a continuum containing few prevoiced tokens than they did for continua that included a larger number of prevoiced stimuli.

One might suppose that qualitative judgements of foreign accent would be more easily influenced by range effects than the discrete categorization of consonants. Long (1990) mentioned two factors that may influence whether a particular talker is judged to be native or non-native. One is the listener's "tolerance" for variations in pronunciation. For example, listeners from a large metropolitan area who have been exposed to many varieties of speech, including dialects of their L1 and various kinds of foreign accent, may be more reluctant to classify a talker as non-native than listeners without such an experience of linguistic diversity. A second factor mentioned by Long (1990) is the proportion of native speakers included in the speech samples being rated. It does not appear that the effect of either factor has been tested empirically.

In experiment 1 we reached the tentative conclusion that foreign accents may first emerge at an age of L2 learning of from five to eight years. This was based on the observation that native Spanish subjects who began learning English by the age of 5-6 years (experiment 1) did not have a perceptible accent, whereas Chinese subjects who began learning English at an average age of 7.5 years (Flege, 1988b) did have an accent. This conclusion would be strengthened if it were shown that range effects influence foreign accent judgments. The listeners in both experiments just mentioned were instructed to use the whole scale. Of the 50 talkers examined in experiment 1, 20 (40%) were native speakers of English, whereas only 10 (21%) of the 47 subjects examined by Flege (1988b) were native speakers of English. Since there was a larger proportion of native English speakers in experiment 1, the listeners who participated in that experiment may have been more apt to classify early learners as non-native than were listeners in the Flege (1988b) experiment. That is, one might hypothesize that the larger the proportion of native and near-native talkers represented in a sample, the more likely listeners will be to detect—or report—a slight trace of foreign accent in the speech of fluent non-native speakers.

On the other hand, if range effects do influence foreign accent judgments, the conclusion that a sufficiently long period of residence in an L2-speaking environment may lead to an improvement in L2 pronunciation would be weakened. Suppose that the larger the proportion of native speakers in a sample, the more strongly accented non-native speakers will seem to be. If such a putative range effect was greater for strongly accented non-native speakers than for non-natives with a relatively good pronunciation of English, then the relatively larger proportion of native speakers included in experiment 1 than in the Flege (1988b) study might have contributed to a significant effect of LOR in the former but not the latter study.

In the present experiment we directly compared the ratings given to English sentences spoken by the Spanish speakers from experiment 1 and the Chinese speakers from the Flege (1988b) study. No native English speakers were included. If range effects influence foreign accent judgments, the non-native subjects should receive higher ratings here than in the previous experiments. More importantly, the increase in ratings should be greater for the Spanish than Chinese subjects because they were originally presented with twice as many native English talkers. We were also interested in learning if the early Spanish learners would receive significantly higher scores than the early Chinese learners; and if the length of residence (LOR) effect noted in experiment 1 would again be obtained.

Experiment 2 provided a way to indirectly assess a potential explanation offered earlier for why the Chinese early learners in the Flege (1988b) study may have had an accent but not the Spanish early learners in experiment 1. Direct phonetic interference from L1 to L2 is usually most obvious in the speech of individuals who are just beginning to learn an L2. If it were shown that inexperienced Chinese late learners pronounced an English sentence as well as inexperienced Spanish late learners, it would suggest that the sentence was equally difficult to pronounce for native speakers of Spanish and Chinese. This would argue against the interpretation that a difference between the two experiments was due to a confound of L1 background.

A. Method

Experiment 2 had two parts, both of which assessed degree of foreign accent in the sentence I can read this for you. The sentences examined in part 1 were spoken by 27 native speakers of Spanish drawn from experiment 1 of the present study and 37 native speakers of Chinese drawn from the Flege (1988b) study. Because only 64 waveforms could be included in a single block due to a software limitation, three of the original ten Spanish early learners from experiment 1 were excluded. In part 2, all of the Spanish subjects from experiment 1 were included, and three of the original 37
Chinese subjects (all inexperienced late L2 learners from mainland China) were excluded. No native English speakers were included in either part 1 or part 2.

The 64 sentences used in part 1 were randomly presented to three native English listeners, all neuroscience graduate students in their mid-twenties. The 64 sentences used in part 2 were presented to 11 native English speakers (three males, eight females) with a mean age of 27 years (s.d. = 5; range = 20–36), all of whom were affiliated with the University of Alabama at Birmingham. All of the neuroscience students had been exposed to foreign-accented speech, but several listeners in part 2 indicated that they had little familiarity with foreign-accented English.

The listeners all passed a pure-tone hearing screening in the range of 500–4000 Hz (20 dB HL) before participating. They were told to use the whole scale, which ranged from "no foreign accent" (256) to "medium foreign accent" (128) to "strong foreign accent" (1). The mean rating given by each listener to each sentence was calculated based on the final three presentations. A mean based on the scores obtained for all of the listeners was then calculated for the 64 tokens of the Read sentence.

B. Results and discussion

1. Part 1

The mean rating given to the Read sentences in part 1 (viz., 122) came close to the midpoint of the scale, which was 128. The lowest foreign accent score obtained (for one of the inexperienced Chinese late learners) was 20; the highest score obtained (for one of the Spanish early learners) was 249. It thus appears that the listeners heeded the instruction to use the whole scale.

Table III presents the mean foreign accent scores obtained in part 1 for the three native Spanish groups and the four native Chinese groups. The foreign accent scores obtained in experiment 1 (for the native Spanish groups) and in the 1988 study by Flege (for the Chinese groups) are presented next to the mean values obtained in the present experiment. As predicted by the hypothesis that range effects influence foreign accent judgments, the late L2 learners’ sentences received higher scores here in experiment 2, where no native speakers of English were included, than in the earlier experiments, where native English talkers were included. This is not in itself a convincing proof of the existence of range effects, however, for the listeners were in fact instructed to use the whole scale. More convincing was the fact that the difference in foreign accent scores between the earlier experiments and the present experiment was greater for Spanish late learners (65 vs 116) than for the Chinese late learners (70 vs 80).

To test the significance of the between-experiment differences, we identified 14 sentences spoken by native Chinese speakers that had received the same, or very nearly the same, ratings in the earlier experiment (Flege, 1988b) as did 14 sentences spoken by native Spanish subjects in experiment 1. The average difference in ratings for the 14 pairs of sentences was just 3.6 (range = 0–14). Eight pairs, designated the “strongly accented” sentences, had scores ranging from 21 to 95; six pairs, designated the “weakly accented” sentences, had scores ranging from 186 to 254. A total of 28 scores (14 from experiment 2, 14 from earlier experiments), were submitted to an ANOVA in which strength of accent (strong vs weak) and language (Spanish vs Chinese) served as between-subject factors, and experiment (experiment 2 vs earlier experiments) served as a repeated measure.

The three-way interaction was nonsignificant [F(1,12) = 1.69, p = 0.218]. An experiment × language ANOVA yielded a significant interaction [F(1,13) = 9.99, p < 0.01] because of a difference in scores obtained for sentences produced by the Chinese and Spanish talkers in experiment 2 (163 vs 126), but not when these talkers’ sentences were rated in the earlier experiments (132 for both groups). Paired t tests revealed that the Spanish subjects’ scores were significantly higher in experiment 2 than in the earlier experiment [t = 3.026, p < 0.01], whereas ratings given to sentences spoken by the Chinese talkers in experiment 2 did not differ significantly from scores obtained in the earlier experiment [t = 1.12, p > 0.10].

As discussed in the Introduction, a demonstration that range effects influence foreign accent judgments would tend to undermine the conclusion that increased L2 experience leads to improved pronunciation of an L2. It was of interest, therefore, to determine if the same length of residence (LOR) effects obtained in the earlier experiments would be obtained in experiment 2, where no native English speakers were included. The 64 mean foreign accent scores were submitted to a one-way ANOVA, which yielded a significant effect of group [F(6,57) = 21.7, p < 0.01]. Newman–Keuls post-hoc tests indicated that, as before, there was a significant difference between the experienced and inexperienced Spanish late learners (152 vs 80), whereas the experienced Chinese late learners did not differ from either group of inexperienced Chinese late learners (82 vs 74 and 73) (p > 0.05). Although a range effect may have contributed to the LOR effect noted earlier for the Spanish late learners, it does not seem to have been an artifactual finding. Adults’ pronunci-
tion of an L2 apparently can improve.

We raised the issue of whether the Chinese but not Spanish early learners had an accent in the earlier experiments because the English sentences examined were somewhat more difficult for native speakers of Chinese than Spanish. It is not possible to evaluate this hypothesis at present due to the lack of equally detailed descriptions of the phones used contrastively in Chinese, Spanish, and English. Data from the present experiment, however, are not consistent with the “differential difficulty” hypothesis. The post-hoc tests revealed that the foreign accent ratings for the inexperienced Spanish and Chinese subjects did not differ significantly. Had the Read sentence been more difficult for Chinese than Spanish learners of English, we would have expected a difference between the inexperienced Chinese and Spanish subjects. The conclusion that English is no more difficult for Chinese than Spanish speakers to pronounce can, of course, be made with confidence only in regards to the Read sentence. (However, the conclusion may extend to the other two sentences examined. With but a few exceptions, the same pattern of differences between native Spanish groups were evident in experiment 1 for the Sue, Read, and Book sentences.)

Based on the results obtained in experiment 1 and by Flege (1988b) it appears that a foreign accent may be evident if one begins learning an L2 at the age of 7-8 years, but not at the age of 5-6 years. One might surmise that if the Chinese and Spanish early learners were compared directly, the Chinese early learners would receive significantly lower ratings than the Spanish early learners. A trend in the expected direction was noted (208 vs 235) in part 1, but post-hoc tests revealed it to be nonsignificant (p < 0.05). Although this might appear to undermine the conclusion that foreign accents first emerge between AOLs of about 5-8 years, it should be recalled that three of the original ten Spanish early learners had to be eliminated from part 1 of this experiment (see above).

2. Part 2

Much the same results were obtained in part 2 as in part 1. The mean ratings given by 11 listeners to the Read sentence was 119 (slightly below the scale midpoint of 128). Scores for individual talkers ranged from 16 to 247. The mean scores obtained for the seven groups of non-native speakers in part 2 are presented in Table IV. As expected, the early learners received much higher scores than the late learners. The between-group differences resulted in a significant effect of group in a one-way ANOVA \(F(6,57) = 27.6, p < 0.001\). Post-hoc tests revealed that the difference between the Spanish and Chinese early learners (234 vs 197) was not significant at the 0.05 level, so the absence of a significant difference in part 1 probably not due to the elimination of three Spanish early learners. The possibility that the lack of an effect was due to the absence of native English speakers in the sample will be examined in experiment 4.

Part 2 replicated, with a new and larger group of listeners, the length or residence (LOR) effects obtained in part 1. The scores for the experienced and inexperienced Spanish late learners (126 vs 75) differed significantly, whereas the experienced Chinese late learners did not differ significantly from either group of inexperienced Chinese late learners (85 vs 61 and 65) \(p < 0.05\). The Read sentence was apparently no more difficult for Spanish than Chinese learners of English, as indicated by the lack of a significant difference between the two inexperienced Chinese late learner groups and the inexperienced Spanish late learners group (61 and 65 vs 75) \(p > 0.05\).

The analysis described earlier for part 1 was again performed to test for the existence of range effects. Fourteen pairs of sentences spoken by Chinese and Spanish talkers were selected. The sentences in each pair received much the same ratings in experiment 1 and the Flege (1988b) study. The original scores for these sentences, were submitted along with the mean scores obtained from the 11 native English listeners in part 2, to a mixed-design ANOVA. The three-way interaction was nonsignificant \(F(1,12) = 0.42, p = 0.526\), just as it was in part 1. An experiment \(x\) language ANOVA yielded a significant interaction \(F(1,13) = 51.9, p < 0.001\) because scores for the Chinese and Spanish talkers differed in part 2 (150 vs 118), but not in the original experiments (132 in both instances). Paired \(t\) tests revealed that scores were significantly higher in part 2 than in experiment 1 for the Spanish subjects \(t = 3.27, p = 0.006\). Ratings accorded sentences spoken by the Chinese subjects did not increase as much as ratings for the Spanish subjects' sentences; in fact, they decreased \(t = 3.27, p < 0.01\). The greater increase in ratings for the Spanish than the Chinese talkers was predicted, but it is uncertain why scores obtained for the Chinese talkers decreased in part 2 as compared to the Flege (1988b) experiment.

III. EXPERIMENT 3

Experiment 2 showed that foreign accent judgments are subject to range effects. Experiment 3 examined the reliability of foreign accent judgments. Listeners rated tokens of a sentence that had been spoken by seven native speakers of Chinese and seven native speakers of Spanish. They rated the same sentences twice, separated by an approximately 5-min

---

**TABLE IV.** Mean foreign accent scores given by 11 native English listeners in experiment 2, part 2, to an English sentence (I can read this for you) spoken by talkers in three native Spanish and four Chinese groups (see the text). \(N\) indicates the number of talkers in each group who were rated for degree of accent.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early learners</td>
<td>10</td>
<td>234</td>
<td>13</td>
</tr>
<tr>
<td>Experienced late learners</td>
<td>10</td>
<td>126</td>
<td>74</td>
</tr>
<tr>
<td>Inexperienced late learners</td>
<td>10</td>
<td>75</td>
<td>27</td>
</tr>
<tr>
<td>Chinese</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early learners</td>
<td>7</td>
<td>197</td>
<td>26</td>
</tr>
<tr>
<td>Experienced late learners</td>
<td>10</td>
<td>85</td>
<td>48</td>
</tr>
<tr>
<td>Inexperienced late learner a</td>
<td>10</td>
<td>61</td>
<td>24</td>
</tr>
<tr>
<td>Inexperienced late learner b</td>
<td>7</td>
<td>65</td>
<td>16</td>
</tr>
</tbody>
</table>

* Native speakers of Taiwanese.
* Native speakers of Mandarin.
interval. The interval was filled by an activity that was designed to focus the listeners’ attention on the acoustic phonetic characteristics of the sentences. In so doing, the activity may have served to make the listeners more aware of the non-native speakers’ divergences from the phonetic norms of English. We wanted to learn if the foreign accent ratings were stable, or if they would change from the first to the second accent rating session.

Experiment 3 also provided an opportunity to further test the conclusion about range effects drawn from the results of experiment 2. The pairs of sentences spoken by the seven Chinese and seven Spanish subjects were matched for degree of foreign accent based on the foreign accent scores obtained in part 1 of experiment 2. In that experiment, roughly 25% of the non-native speakers examined spoke with little or no accent. All 14 sentences selected for the present experiment were among the remaining sentences that had strong accents. Since the listeners in the present experiment (like those in previous experiments) were told to use the whole scale, we expected to obtain higher ratings for the 14 sentences than were obtained in experiment 2.

We expected the same increase in ratings for the Chinese and Spanish talkers’ sentences, a prediction that differs from the one made in experiment 2. Recall that, in experiment 2, the Chinese and Spanish talkers were paired on the basis of scores obtained in earlier experiments in which differing proportions of native English speakers were included along with the non-native speakers. Sentences spoken by Chinese and Spanish talkers that were examined here, on the other hand, were paired on the basis of scores obtained in a single experiment (viz., experiment 2) with no native speakers. Therefore, the range effects noted in experiment 2 were not expected in experiment 3.

A. Method

Degree of foreign accent was assessed in the sentence I can read this for you as spoken by seven Chinese and seven Spanish native speakers who were matched closely in terms of the mean ratings obtained in part 1 of experiment 2. The 14 sentences received scores ranging from 31–71, which fell near the “strongly accented” end of the scale. The sentences were rated for degree of accent by nine listeners (three males, six females), all young adults who were monolingual native speakers of American English. The listeners were all affiliated with the University of Alabama at Birmingham, and passed a pure-tone hearing screening. None of them had previous experience rating sentences for degree of foreign accent.

The listeners participated in three sessions. They rated the 14 sentences (randomly presented four times each) for degree of foreign accent in session 1 and session 3 using the procedures described earlier. As in earlier experiments, mean foreign accent scores were computed for each sentence based on the final three judgments given by each listener. A mean rating for each of the 14 sentences was computed based on the means obtained from the nine listeners. The 28 resulting mean foreign accent scores obtained at session 1 and session 3 were then submitted to a (2) language × (2) session ANOVA.

The listeners participated in an identification experiment during session 2 (i.e., in the interval between the two rating sessions). The same 14 sentences were randomly presented eight times each. On each trial, the listeners had to decide if a sentence had been spoken by a native speaker of Chinese or Spanish. Feedback concerning the correct response was given immediately after each trial. The percentage of correct identifications in each of the eight separate randomizations was calculated for each listener. This was done in order to determine if the feedback training significantly improved the listeners’ ability to identify the talkers’ L1. We were uncertain if the feedback training would be effective. Regardless of whether it was or was not, we thought it would serve to familiarize the listeners with the acoustic phonetic characteristics of the 14 sentences, which might be expected to change the listeners’ perception of foreign accent. The feedback training in session 2 could lower the listeners’ ratings, if it caused them to note more divergences from the phonetic norms of English; or, it could raise their ratings, if it caused the listeners to become more tolerant of the deviations they noted.

B. Results and discussion

Figure 2 presents the data obtained in session 2. It shows the mean percentage of correct identifications of the non-native talkers’ L1 that was obtained in response to the eight random presentations of the 14 sentences. The listeners’ ability to identify the talkers’ L1 background changed little as the result of the feedback training. L1 background was identified correctly in just 66% of instances, on the average. A one-way ANOVA indicated that the correct identification rate did not change significantly over the eight randomizations [F(7,56) = 0.50, n.s.].

The 14 sentences received mean ratings that ranged from 31 to 71 in part 1 of experiment 2. In the present experiment, where we presented no sentences with little or no accent, the same 14 sentences received a mean rating of 127 (which comes very close to the midpoint of the scale). The ratings given during session 3 were lower than those given at

![FIG. 2. The mean percentage of times that native English listeners correctly identified the native language of strongly accented Chinese and Spanish late L2 learners (seven each) on the basis of the acoustic properties of a single English sentence. Eight randomizations of the 14 sentences were presented with feedback. The brackets enclose + / - one standard deviation.](image-url)
becomes better acquainted with them over the course of session 2 without becoming more tolerant.

If so, listeners might give higher ratings to sentences spoken by a non-native talker in a single session after becoming familiar with his/her speech than before such familiarization has taken place. It would be useful to examine the effect of more substantial increases in familiarity (i.e., days or weeks of exposure) than were available in the present study. We speculate that a U-shaped function would be obtained: scores at first decreasing, as listeners became more aware of divergences from English phonetic norms, then increasing as the listeners established correspondence "rules." The correspondence rules would be expected to increase intelligibility by relating regular forms of mispronunciation to standard English pronunciations (Wingstedt and Schulman, 1987).

The lack of a significant language X session interaction \[ F(1,124) = 0.68 \] suggested that the decrease in scores from session 1 to session 3 was no greater for sentences spoken by the Spanish subjects than by the Chinese subjects. This strengthens the conclusion drawn from experiment 2 concerning range effects. In that experiment, scores for native Spanish speakers increased more than those for native Chinese speakers, apparently because the original scores for the Spanish speakers were obtained in an experiment that included a larger proportion of native English speakers.

We were also interested in assessing inter-listener reliability. Oyama (1976) reported that two listeners' ratings of paragraphs spoken by 60 Italian and 10 native speakers of English were strongly correlated \( r = 0.860 \). Patkowski (1980) had two judges rate non-native speakers for syntactic control of English and pronunciation of English using nine-point scales. The two judges were highly experienced teachers of English as a second language. They were given only a "few minutes" of training on how to rate 30-s excerpts of speech for degree of accent, whereas they received training over a two week period on how to rate L2 syntactic proficiency. Despite this disparity, the interjudge reliability was somewhat higher for the ratings of pronunciation than syntax \( r = 0.840 \) vs \( r = 0.790 \). From this, Patkowski concluded that degree of accent is "more easily perceived and judged" than control of L2 syntax.

The results obtained here clearly demonstrate that listeners are sensitive to small variations in degree of accent. First, we found that intra-rater reliability was high. The correlation coefficients for ratings assigned to the 14 sentences by the nine listeners during sessions 1 and 3 averaged 0.795 (s.d. = 0.123; range = 0.536-0.957). We also assessed inter-rater agreement among the nine listeners in experiment 3. Cronbach's reliability coefficient was slightly higher at session 1 than session 3 \( (0.920 \text{ vs } 0.907) \). We computed the 36 simple correlations between all possible pairings of the nine listeners. The correlations at session 1 averaged 0.596 \( (r = 0.210 \text{ to } 0.873) \). Of the 36 correlation coefficients, 23 were significant at the 0.05 level. The coefficients averaged 0.536 at session 3 \( (r = -0.016 \text{ to } 0.921) \). Nineteen of the correlations were significant at the 0.05 level. An ANOVA revealed that the correlation coefficients were significantly smaller at session 3 than at session 1 \( F(1,135) = 4.87, p = 0.034 \).

The average inter-rater correlations obtained here were somewhat lower than those reported for a pair of expert judges in studies by Oyama (1976) and Patkowski (1980). It is important to remember, however, that the 14 sentences were drawn from a very narrow range of the continuum of possible foreign accents. We think it likely that they would have occupied just one of the four discrete categories used in the Asher and Garcia (1969) experiment (viz., the "definitely accented" category). Recall, too, that our listeners had no experience with speech research (most were graduate students in the biological sciences) and were given no training or instructions whatsoever on how to rate degree of foreign accent. Despite this, several pairs of our listeners had somewhat higher inter-rater coefficients than the pair of expert judges who participated in the Patkowski (1980) study.

IV. EXPERIMENT 4

Chinese and Spanish early learners were again compared. Although the native Spanish speakers received higher scores than the native Chinese speakers in experiment 2, the difference was nonsignificant. Recall that sentences spoken by the native speakers were not presented along with sentences produced by native English speakers in that experiment. The final experiment tested the hypothesis that a significant difference between the two early learner groups would be obtained if their sentences were presented in a single experiment along with sentences spoken by native speakers of English.

A. Method

Degree of foreign accent was examined in The good shoe fits Sue, I can read this for you, and The red book was good. These English sentences were rated for degree of foreign accent by nine monolingual native speakers of American English (four males, five females) with a mean age of 28 years (s.d. = 5). All listeners were affiliated with the University of Alabama at Birmingham, and all passed a pure-tone hearing screening. The listeners rated sentences spoken by ten monolingual native speakers of English from Birmingham and ten native Spanish early L2 learners (all drawn from experiment 1). Also included were sentences spoken by seven Chinese early learners (drawn from the Flege, 1988b, study).

After being bandpass filtered (80-8000 Hz), the digitized Sue, Read, and Book sentences were presented over headphones at a comfortable level in separate, counterba-
lanced blocks. The listeners were told that they would hear sentences spoken by 27 individuals, and that they were to rate each sentence for degree of foreign accent. They were told to position the lever at some point along the continuum marked “no foreign accent” at the top, “medium foreign accent” at the middle, and “strong foreign accent” at the bottom.

As before, the listeners were told nothing about the listeners they were evaluating; and they were told to use the whole scale. The sentences in each block were randomly presented four times each, and an average foreign accent score was based on the listeners’ final three responses to each sentence. Mean scores based on the mean ratings of the nine listeners were then calculated and submitted to a (3) group × (3) sentence ANOVA.

B. Results and discussion

Telling listeners to use the whole scale implied that at least some talkers would have a strong accent. After the experiment, several listeners commented that, despite the instructions, few if any talkers seemed to speak with a strong accent. In the experiments presented thus far, the overall mean ratings obtained for all talkers in an experiment have fallen near the midpoint of the scale (i.e., a score of about 128). In the present experiment, the mean rating for the 27 talkers was somewhat higher (viz., 165). It does appear, however, that the listeners tried to use the whole scale. The highest possible value of 256, which corresponded to the “no foreign accent” label, was used for only about 10% of the 2187 responses given (3 sentences × 27 talkers × 9 listeners × 3 repetitions), despite the fact that all of the sentences evaluated had little or no accent.

The ANOVA indicated that the small differences between the Sue, Read, and Book sentences (161, 163, 171) were nonsignificant [F(2,48) = 1.31, p = 0.280]. The sentence × group interaction was also nonsignificant [F(4,48) = 0.65, p = 0.629]. The group factor was highly significant, however [F(2,28 = 12.66, p < 0.001]. Newman–Keuls post-hoc tests indicated that the native English speakers and the Spanish early learners received significantly higher ratings than the Chinese early learners (187 and 187 vs 102) (p < 0.01). The average ratings obtained for the native English speakers ranged from 132 to 220. Only one of the ten Spanish early learners’ scores fell below the English range, but five of the seven Chinese early learners’ scores did so. Of the two Chinese early learners whose scores fell into the native English range, one was an 18-year-old male who had arrived in the United States at the age of 3 years. The other was an 18-year-old female who had arrived in the United States at the age of 5 years. These talkers received scores of 149 and 155, respectively. The Chinese talker who received the lowest score (viz., 20) was a 20-year-old male who reported having arrived in the United States at the age of 10 years.

The results obtained here support the hypothesis that a foreign accent emerges at an age of L2 learning (AOL) between about 5 and 8 years. Recall that the native Spanish subjects started learning English when they began attending school at the age of 5–6 years, whereas the Chinese listeners first arrived in the United States—and presumably first began learning English—at an average age of 7.6 years. This conclusion must be considered tentative, however, for several reasons. We examined very few subjects spanning the AOL at which foreign accents are hypothesized to emerge. Moreover, there is some uncertainty about the actual chronological age at which our subjects were first massively exposed to native-produced English and actually began learning English. Although the native Spanish subjects all reported being unable to speak English before going to school, some of them were born, or were living in the United States before the age of 5–6 years. Of these, some may have been passively exposed to English prior to the age of 5–6 years, and this early exposure may have contributed to their accent-free pronunciation of English. Also, the two early L2 groups spoke different native languages (although we presented evidence earlier that the English sentences examined may have been no more difficult for speakers of Chinese than Spanish). Clearly, additional research with subjects from a single L1 background will be needed to accurately pinpoint the AOL at which foreign accents emerge.

In experiment 3 we found that inter-listener reliability was high for sentences spoken with a strong foreign accent (i.e., with low scores). A high degree of inter-listener reliability was also evident in sentences spoken with little or no accent. Cronbach’s alpha was computed for the nine listeners’ ratings of the Sue, Read, and Book sentences. The coefficients were 0.952, 0.915, and 0.939, respectively. Simple correlations were computed for all 36 possible inter-listener correlations for each sentence. For the Sue, Read, and Book sentences the average correlations were 0.710 (range = 0.434 to 0.860), 0.560 (range = 0.156 to 0.814), and 0.647 (range = 0.241 to 0.848), respectively.

V. GENERAL DISCUSSION

This study yielded findings that shed light on listeners’ evaluation of degree of accent and on talker-related factors that lead to variations in degree of perceived foreign accent. We will discuss the listener-related findings before turning to the talker-related findings, which concern differences between groups of non-native talkers. In the final section we discuss some general issues relating to L2 speech learning.

A. Listener-related factors

Experiment 2 provided evidence that listeners’ judgments of degree of foreign accent is not absolute, but is influenced by the range of talkers included in a sample. Scores obtained for native speakers of Chinese and Spanish in experiment 2, where no native English talkers were included, were compared to the scores obtained from the same non-native talkers in two earlier experiments. There were twice as many native speakers included in the earlier experiment with Spanish subjects (experiment 1) than in the earlier experiment with Chinese subjects (Flege, 1988b). As predicted, the non-natives’ scores were higher in experiment 2 than in the earlier experiments, owing to the absence of native English speakers. Also, the Spanish talkers’ scores increased significantly more than the Chinese talkers’.
The results of experiment 3 strengthened the conclusion that the differential change in scores for Spanish and Chinese talkers in experiment 2 was due to the proportion of native English talkers included in the earlier experiments. In experiment 3, sentences spoken by seven Spanish speakers were paired with seven sentences spoken by Chinese speakers on the basis of foreign accent scores obtained in experiment 2. All of the non-native speakers had strong accents. Since a single range of foreign accents was represented in experiment 2, we expected any change in scores between experiment 3 and experiment 2 to be equal for the Chinese and Spanish speakers. Overall, the mean score for the 14 sentences was higher in experiment 3 than in experiment 2. (This was expected because only talkers with strong accents were included in experiment 3.) As predicted, however, the increase was much the same for the Spanish and Chinese talkers.

The present study provided evidence that listeners can make very fine-grained and reliable judgments of degree of accent. Inter-listener reliability was good. We would, of course, expect strong correlations to be obtained if two groups of listeners were asked to assess a set of sentences including strongly accented sentences as well as unaccented sentences spoken by native or near-native speakers of English. The findings go beyond that, however. We found that interlistener correlations were also quite high for a homogeneous set of sentences with little accent as well as for a homogeneous set of sentences with strong accents. Native English listeners in two experiments (experiment 1, Flege, 1988b) rated sentences spoken by a group of native English talkers. The two sets of ratings were strongly correlated \( r = 0.925 \) even though all of the sentences received ratings in the range of 164 to 256. We also assessed inter-listener agreement for the nine listeners who participated in experiment 4. Most of the talkers examined in that experiment spoke English with little or no accent. Cronbach's alpha averaged 0.935 for the three sentences examined. The average inter-listener correlations ranged from 0.560 to 0.710.

In experiment 3, listeners rated 14 strongly accented sentences on two occasions. The Cronbach reliability coefficients averaged 0.914. The average inter-listener correlation for all possible pairings of the 11 listeners (0.566) was lower than that obtained for the two expert judges who participated in a study by Patkowski (1980). This was hardly surprising since the listeners in experiment 3 had no prior experience with speech research and had received no prior training. Perhaps just as importantly, they heard only a single sentence (rather than paragraph-length material) and they were asked to rate speech samples that occupied a very narrow range of foreign accents. It is therefore noteworthy that a few pairs of our nonexpert listeners had inter-subject correlations as high as Patkowski's two expert listeners.

Intra-listener reliability in making foreign accent judgments was also shown to be quite high. Correlations were computed for the two sets of ratings given to 14 sentences in experiment 4 by nine listeners. The coefficients ranged from \( r = 0.536 \) to \( r = 0.957 \), averaging 0.795.

The scaling technique used here was sufficient to show small differences between groups of talkers. It is not certain, however, if it is the best way to evaluate degree of accent, for the nature of the foreign accent scale is unknown. Ryan et al. (1977) compared the foreign accent ratings given by listeners to excerpts of Spanish-accented English using the method of equal appearing intervals (EAI) and direct magnitude estimation (DME). The EAI scores correlated significantly with the DME scores \( (r = 0.880) \). The strength of the EAI-DME correlation increased somewhat when scores were converted to logarithms \( (r = 0.970) \) because a curvilinear relationship existed between the EAI and the DME scales. This suggests that, just as for speech intelligibility (Schiavetti et al., 1981), foreign accents may represent a prothetic continuum, like loudness, rather than a metathetic continuum, like pitch (Stevens, 1974).

Unequal differences between stimuli representing different degrees of accent might be due to variations in listeners' resolution of accent differences at various locations along the foreign accent scale (Stevens, 1974). If so, equal appearing intervals, which have been employed in many previous studies, may be inappropriate for scaling degree of foreign accent. For nonlinear, prothetic continua, direct magnitude estimation is the preferred scaling technique. This raises the issue of whether the scaling technique used here, which involved having listeners position a lever at some point along a 7-cm range, is more nearly equivalent to EAI or DME scaling.

B. Talker-related factors

Two important conclusions were drawn from the present study concerning talker-related factors that may influence degree of foreign accent. The first conclusion is that, among individuals who learn an L2 as adults, those who are highly experienced in the L2 will pronounce it better than those who are relatively inexperienced. In experiment 1, it was shown that a group of Spanish late learners who had lived in the United States for 14.3 years received significantly higher ratings than a group of late learners who had lived in the United States for only 0.7 years. Flege (1988b), on the other hand, did not observe a significant difference between groups of Chinese late learners who had lived in the United States for 1.1 and 5.1 years on the average.

It is possible that a significant length of residence (LOR) effect was observed here, but not in the Flege (1988b) study, because the two Spanish groups differed more according to the LOR variable than the two Chinese groups. However, the difference between the two studies may have more to do with how well the relatively inexperienced late learners spoke English. We speculate that the experienced Chinese subjects may have come closer to achieving their ultimate level of proficiency in English pronunciation than the inexperienced Spanish subjects. It will be important in future longitudinal research to plot the rate of change in L2 pronunciation by adult learners in terms of their ultimate attainment. We hypothesize that most improvements in pronunciation occur during the first year of L2 learning. If so, this would distinguish speech learning
from various aspects of language learning, such as lexical learning, where one might expect continued progress over many years.

Another important conclusion drawn from the study was that a foreign accent may first emerge when L2 learning commences between about the ages of five and eight years. In experiment 1 it was shown that a group of native Spanish early learners who began learning English at the age of 5–6 years did not have a measurable accent. An earlier study using the same speech materials and procedures (Flege, 1988b) showed that a group of Chinese early learners with a mean age of L2 learning (AOL) of 7.6 years did have an accent. The two early L2 groups were directly compared in experiment 2. Although the Spanish early learners had somewhat better accents (i.e., received higher ratings) than the Chinese early learners, the difference did not reach significance.

The nonsignificance of the difference between the Spanish and Chinese early learners seems to have been due to the absence of native English talkers in the sample being evaluated. Listeners have native speech patterns stored in long-term memory (Flege, 1984). Nevertheless, their ability to make fine-grained judgments of degree of accent, which may be based on a determination of the degree of divergence from stored native-speaker norms, appears to be facilitated by the availability of native-produced sentences. When sentences spoken by the Spanish and Chinese early learners were presented together in experiment 4, the Spanish early learners were found to have significantly higher scores than the Chinese early learners; also, the Chinese, but not the Spanish early learners, differed significantly from the native English speakers.

The conclusion that a foreign accent first emerges at an age of L2 learning (AOL) of between 5 and 8 years supports a hypothesis advanced recently by Long (1990). It appears to differ from the view that accents are associated with the passing of a neuropsychologically triggered “critical period” ending at puberty, that is, at a chronological age of about 12 or 13 years (e.g., Lenneberg, 1967; Scovel, 1988) or even later (Patkowski, 1989). Several points need to be made about the apparent divergence of views. Our conclusion concerning the AOL at which foreign accents emerge must be regarded as tentative for reasons given earlier. The two early L2 groups compared in the present study differed in L1 background as well as in AOL (but see above). Also, some members of the Spanish early L2 group may have been passively exposed to English prior to their nominal AOL of 5–6 years. The divergence of views may be more apparent than real if one considers the proportion of individuals at various AOLS who speak the L2 without a perceptible accent (e.g., Seliger et al., 1975) rather than average degree of accent. One might hypothesize that a perceptible foreign accent is highly unusual for individuals who begin learning an L2 before reaching the age of 6 years, is almost always evident for post-pubescent learners, and is evident in an increasing proportion of individuals who begin L2 learning between the ages of about 6 to 12 years.

Although it has often been linked to puberty, we can think of no satisfying explanation as to why the effects of a critical period should first be evident at the age of 12–13 years. Of course, one might argue that a critical period for speech learning is centered at puberty but that the “beginning of the end” of a period of heightened sensitivity to speech stimulation occurs earlier in life. Many studies have shown a relationship between AOL and degree of foreign accent (e.g., Oyama, 1976; Tahta et al., 1981a; Thompson, 1984), but the exact relationship between AOL and degree of accent remains uncertain. This is because no study we know of has examined degree of accent with sufficient resolution in the speech of talkers from a single L1 background who represent a sufficiently wide range of AOLS.

As a first approximation to a function relating AOL to degree of accent, we have displayed the data obtained in experiment 1 for 30 Spanish subjects in the top panel of Fig. 3. (Recall that 20 of these subjects were late learners and the remaining 10 were early learners.) We do not at present have sufficient data to fill out the function. The bottom panel of Fig. 3 is an idealized function fitted to hypothetical data. Each data point is meant to represent multiple subjects in successive AOL “bins.” The purpose of the bottom panel is to graphically depict four hypotheses that might serve to guide future research.

The first hypothesis illustrated by the hypothetical data in Fig. 3, which was suggested by the results of the present study, is that foreign accents first become perceptible at AOLS of between about 5 and 8 years. The emergence of accent would be defined by the first, downward deflection in a function fitted to the data.

The second hypothesis is that there is a linear increase in degree of foreign accent after the AOL at which accents first emerge. This hypothesis is derived from data presented by Oyama (1976) and Tahta et al. (1981a). As suggested by Oyama (1979), a behavior as complex as speech is unlikely to be described by a rectangular function that would be needed if accents increased precipitously at a certain AOL. The “linear increase” hypothesis seems to stand in contrast with the results of studies of certain avian species, where song learning ability may be characterized by fairly well-defined periods of sensitivity to environmental stimulation (Nottebohm, 1969, 1989; Marler and Peters, 1989; Gould and Marler, 1989). It also seems to stand in contrast with Lenneberg’s view (1967, p. 142) that a critical period for speech ends at the “age (sic) of puberty,” and with Patkowski’s (1989) conclusion that subjects learning an L2 before and after the age of 15 years differ markedly. We speculate that foreign accents become increasingly strong beyond puberty.

The third hypothesis illustrated by the hypothetical data is that a plateau in pronunciation is reached at some time after the teens. It is consistent with the finding of many investigators (e.g., Patkowski, 1989) that relatively little change in degree of accent is evident for subjects with AOLS greater than about 18 years.

One other hypothesis is illustrated in Fig. 3. It is that inter-subject variability increases steadily from the AOL at which foreign accents first emerge until the “plateau” just mentioned has been reached. The variability hypothesis is represented by the widening of the error bars, which stands for standard deviations associated with the means for the
FIG. 3. (top) Relation between the mean foreign accent scores accorded English sentences spoken by the 30 native Spanish subjects in experiment 1 as a function of their age of L2 learning; (bottom) an idealized function in which hypothetical foreign accent scores are related to the age of L2 learning (AOL) of non-native speakers of English (see the text for a discussion).

The variability hypothesis is based on the belief that, as one ages, an ever wider range of factors may influence one’s pronunciation of an L2. The catalog of potential factors is large, including differences in personality (Singleton, 1981; Guirao et al., 1972; Taylor et al., 1969; Schumann, 1975, 1978; Krashen, 1985), degree of motivation to learn the L2 well and speak it authentically (Van Els and De Bot, 1987), and the quantity and quality of L2 phonetic input.

The variability hypothesis is consistent with several previous findings. It agrees with a finding by Baci (1956), who showed increasing variation in “use-of-L2” indices with increasing AOL, and with Patkowski’s (1989) observation that the range of foreign accents is greater for individuals who began learning L2 after, as opposed to before, the age of 15 years. It is consistent also with the results obtained in a study of L2 syntax learning by Johnson and Newport (1989). These authors provided evidence that inter-subject variability is substantially greater for subjects with AOLS of 17-39 years than 3-15 years. Finally, the variability hypothesis is consistent conceptually with an observation made by Lenneberg (1967). He observed that speech learning by young children appears to be “automatic” whereas that of adults is “labored,” and thus presumably subject to higher-level cognitive strategies.

It is important to note that an adequate definition of the relationship between AOL and degree of accent would not constitute an explanation for why foreign accents occur. Many investigators would interpret an increase in accents—abrupt or otherwise—as evidence for the existence of a critical period (see Penfield and Roberts, 1959; Lenneberg, 1967; Scovel, 1969, 1988; Fathman, 1975; Lamendella, 1977; Singleton, 1981; Snow and Hoehnle-Hohle, 1982; van Els and de Bot, 1987). However, positing that speech learning is subject to maturational constraints will not explain why early learners usually pronounce an L2 better than late learners (see Flege, 1987b). Oyama posed three questions in 1976 that we believe are as timely now as then. She asked “What kinds of phonology learning are subject to age restraints and what kinds are not?”, “How are various aspects of phonology learning related to more general cognitive processes?”, and “What is the role of motor coordination [in the learning of L2 phonology]?”

Work carried out within the framework of Flege’s Speech Learning Model (SLM) has attempted to address such questions (e.g., Flege, 1988, 1991b,c,d,e; Bohn and Flege, 1990). According to the SLM, a gradual increase in degree of foreign accent might arise because of differences in the state of development of the phonetic system at the time L2 learning commences. It is hypothesized that learners of all ages remain able to establish additional phonetic categories when they encounter certain “new” L2 sounds that do not have a counterpart in the L1 (also see Best et al., 1988; Best and McRoberts, 1989). However, the ability to do so for “similar” sounds that have an L1 counterpart may decrease after about the age of 5-6 years. Beyond that age, L2 learners may become more likely to equate similar L2 sounds with corresponding L1 sounds and, as a result, become increasingly likely to ignore auditorily accessible acoustic differences that distinguish corresponding L1 and L2 sounds. If so, foreign accents might increase with AOL because a longer time is needed for L2 learners to recognize that certain L2 sounds without an L1 counterpart are “new,” because an increasingly larger proportion of sounds in the L2 inventory are equated with L1 sounds, or both.

These hypotheses are consistent with observations made by Scott (1978) concerning sensitive periods. Scott noted that, in general, organizational processes are modified most easily at the time they are proceeding most rapidly. It seems reasonable to think that phonetic systems are changing more rapidly at the age of 5 years than at the age of 15 years. Scott (1978) also noted that behavioral development is cumulative in the sense that as more behaviors are added, they are “integrated into specialized systems and subsystems” which may interfere with the acquisition of subsequent new behaviors; and that changes in behavior become progressively more difficult as organizational processes become more stable.

A full understanding of why accents arise will have to take into consideration factors other than the age of L2 learning (AOL). As noted by many investigators (e.g., Oyama, 1979; Snow, 1979, 1987; Flege, 1987b), a variety of cognitive, social, and psychological factors often covary with AOL and have been confounded with AOL in many foreign accent studies (including the present study). For example, Baci (1956) reported that AOL accounted for
45% of variance in use-of-L2 scores for Hebrew L2 learners in Israel. Generally speaking, the older an individual was at the time of immigration to Israel, the less Hebrew was used on a daily basis. Length of residence (LOR) in Israel also accounted for 45% of variance: the greater the LOR (up to about two years) the greater the use of Hebrew. Individuals who arrived in Israel as young children probably had a great deal more Hebrew input in a given period of time than individuals who arrived as adults.10

In summary, this study has led to four important conclusions. First, foreign accent judgments, like many other aspects of speech perception, are subject to range effects. Second, they are not stable, but may shift over as short a time period as five minutes. Third, adult L2 learners' pronunciation of English may improve measurably over time. And finally, the age of L2 learning (AOL) at which foreign accents first become perceptible occurs long before puberty, probably between the ages of about five to eight years. Much additional research is needed, not only to clarify the perceptual mechanisms that underlie the perception of accent, but also to more fully assess the relative effect on degree of accent of variations in AOL and amount of L2 input. A "large N" study, including subjects who first began learning their L2 from ages ranging from about 2 to 25 years is needed to pinpoint the AOL at which accents first emerge, and to accurately describe the relationship between AOL and degree of accent. Given what Bornstein (1987, p. 3) called the "interactional" nature of development, one might suppose that factors other than AOL, especially those relating to the quantity and quality of L2 input, will affect L2 pronunciation. An optimal model of L2 speech learning will need to account for all relevant factors, including input and social-psychological factors, and their interactions.

ACKNOWLEDGMENTS

This study was funded by NIH Grant No. DC002547. The authors thank A. Levitt, M. Mack, M. Munro, C. Snow, A. Walley, and an anonymous reviewer for comments on earlier drafts of the paper. We also thank C. Mena and L. Skelton for help with data collection and analysis.

1 The only difference between the talker- and listener-based ANOVAs was that a few more significant differences between sentences were noted in the listener-based ANOVA. For the experienced late learners, the Book sentence received higher ratings than the Read sentence, which received higher ratings than the Sue sentence (p < 0.05). For the native speakers of English from Birmingham, the Book and Read sentences received higher ratings than the Sue sentence.3

2 The basis for the few between-sentence differences that were noted is uncertain. For the experienced late learners, the Sue sentences received lower (i.e., more accented) scores than the other sentences. One might hypothesize that the Sue sentences contained more "difficult" sounds (e.g., sounds that differed to a greater extent from sounds in Spanish) than the other sentences. One problem with this interpretation, however, is that the same effect was not evident for the inexperienced late learners.

3 The Chinese subjects were not asked when they were interviewed at what age—in years and months—they first arrived in the United States. In what now seems to be an important oversight, they were not asked detailed questions about their first year in the United States. No elementary school in Birmingham, as far as we know, was conducting classes in Chinese a decade ago. If the Chinese subjects were enrolled in school immediately, then their age of arrival (AOA) in the United States would be equivalent to their age of L2 learning (AOL).

4 The confounding of L2 input and AOL can be illustrated by considering the study of Patkowski (1980). Groups of L2 learners differing in AOL (9 vs 27 years) were compared. Although the two groups were nicely matched in terms of LOR (20 vs 19 years in the United States) the "presubjective" group were estimated to have received 44% more L2 input ("exposure") than the late learners. AOL was strongly correlated with L2 syntactic performance after the effects of the exposure differences were partialed out statistically. The exposure variable was, nevertheless, significantly correlated with the dependent variable under investigation.