The Use of Global Foreign Accent Rating in Studies of L2 Acquisition

A Report Prepared for the
Language Research Centre
University of Calgary

by

Karen Jesney
Department of Linguistics
1. Introduction
Foreign accent, as an identifying aspect of nonnative speech, has been widely discussed in both the theoretical and pedagogical literature. Still, the elements that contribute to the perception of foreign accent, and, indeed, the objective characteristics of it, remain ill defined. With this in mind, numerous studies have sought to quantify degrees of foreign accentedness by asking groups of listeners to rate speech samples for degrees of foreign accent using a Likert scale. These ratings are then correlated with a range of linguistic and cognitive variables and generalizations drawn. The precise methodology utilized in making such foreign accentedness ratings remains unstandardized, however, and, as a consequence, numerous differences in both specific approaches and findings have been noted across studies. This paper will first summarize the range of approaches represented in the literature. It will then consider listener-based variables, such as degree of familiarity with accented speech, that can influence the judgments made by listeners and the reliability of these. Finally, the findings of these studies insofar as the correlation of accentedness ratings with talker-based factors (such as age of L2 acquisition) and stimulus factors (such as rate of speech) are concerned, will be considered and conclusions drawn.

2. The Methodology of Accentedness Ratings

2.1 Rating Scales

Ratings of global foreign accent are most commonly obtained by asking native speakers to judge speech samples using a Likert scale, the ends of which are given contrasting labels such as ‘no foreign accent’ and ‘very heavy foreign accent.’ The number of gradients on such scales has varied widely, ranging from three to ten, with the most common approach in more recent studies being the adoption of a nine-point scale (e.g., Munro & Derwing 2001, Riney, Takada & Ota 2000, Yeni-Komshian, Flege & Liu 2000).

<table>
<thead>
<tr>
<th>Scale</th>
<th>Example studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three-point scale</td>
<td>Tahta, Wood &amp; Loewenthal (1981)</td>
</tr>
<tr>
<td>Four-point scale</td>
<td>Asher &amp; García (1969); Flege, Frieda &amp; Nozawa (1997); Munro (1995)</td>
</tr>
<tr>
<td>Five-point scale</td>
<td>Olson &amp; Samuels (1973); Oyama (1976); Snow &amp; Hoefnagel-Höhle (1977); Major (1987); Anderson-Hsieh &amp; Koehler (1988); Piper &amp; Cansin (1988); Patkowski (1990); Thompson (1991); Bongaerts, Planken &amp; Schils (1995); Elliott (1995); Bongaerts, van Summeren, Planken &amp; Schils (1997)</td>
</tr>
<tr>
<td>Six-point scale</td>
<td>Suter (1976); Moyer (1999)</td>
</tr>
<tr>
<td>Seven-point scale</td>
<td>Brennan &amp; Brennan (1981b); Anderson-Hsieh, Johnson &amp; Koehler (1992); González-Bueno (1997); Magen (1998); Southwood &amp; Flege (1999)</td>
</tr>
<tr>
<td>Nine-point scale</td>
<td>Flege &amp; Munro (1994); Munro &amp; Derwing (1994); Munro &amp; Derwing (1995a); Munro &amp; Derwing (1995b); Derwing &amp; Munro (1997); Munro &amp; Derwing (1998); Riney &amp; Flege (1998); Munro &amp; Derwing (1999); Munro, Derwing &amp; Flege (1999); Riney &amp; Takagi (1999); Flege, Yeni-Komshian &amp; Liu (1999); Guion, Flege &amp; Loftin (2000);</td>
</tr>
</tbody>
</table>
As evidenced in (1), while Likert scales dominate in such studies, other approaches have also been adopted. Prime among these is the use of a sliding scale (e.g., Major 1986, Flege & Fletcher 1992) where raters adjust a lever or cursor along a continuum upon which only the endpoints are marked. The position selected by the rater is then interpreted as a numerical score by a computer. This approach has the benefit of allowing much finer distinctions, up to 256 in many such studies, to be recorded than is the case when a simple Likert scale is used. At the same time, however, raters using a sliding scale approach are unaware of the individual gradients, raising questions about the reliability of these fine distinctions.

Direct magnitude estimation has also been adopted in some studies, including those of Brennan, Ryan & Dawson (1975) and Brennan & Brennan (1981a). In this case, raters accord a numerical score to the first token presented and then judge subsequent tokens on its basis, multiplying by two for a token deemed twice as accented as the first, etc. Under this approach, individual raters effectively devise their own scales, with the relationship between the individual numbers, rather than the raw scores themselves, being the variable of interest.

Given this range of approaches, it is worth noting that at least one study has considered the amenability of foreign accent to scalar and direct magnitude rating. Specifically, Southwood & Flege (1999) compared the ratings given to a data set by judges using direct magnitude estimation and a 7-point Likert scale. It was found that results obtained using the two methods mirrored each other, a finding that was interpreted as an indication that interval scales can be effectively employed in rating degrees of foreign accentedness. At the same time, Southwood & Flege (1999) found that direct magnitude estimation enabled greater degrees of distinction among those tokens at the top end of the scale than did the seven-point Likert scale. The researchers therefore suggest that ceiling effects may have been at play and that use of a nine- or eleven-point scale may be more effective than use of a seven-point scale. Based on these results, as well as the strong trends that have emerged across a range of studies employing Likert scales (see section 3), it seems reasonable to continue the use of such an approach in assessing degrees of global foreign accent.

### 2.2 Tokens

Just as the types of scales used in accentedness studies have varied considerably, so too have the types of stimuli upon which accentedness judgments have been based. As suggested by the comparison in (2), however, single sentences are have been the most common type of stimuli used in such studies, an approach which has the benefit of controlling the range of non-
phonological variables which might contribute to judgments of foreign accent (e.g., Flege, Munro & MacKay 1995, Jilka 2000, Munro & Derwing 1995a,b).

<table>
<thead>
<tr>
<th>Stimuli Length</th>
<th>Example studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>30ms</td>
<td>Flege (1984)</td>
</tr>
<tr>
<td>Syllable</td>
<td>Flege (1984)</td>
</tr>
<tr>
<td>Word</td>
<td>Olson &amp; Samuels (1973); Snow &amp; Hoefnagel-Höhle (1977); Major (1987); Flege &amp; Munro (1994); Bongaerts, Planken &amp; Schils (1995); González-Bueno (1997); Moyer (1999)</td>
</tr>
<tr>
<td>Phrase</td>
<td>Flege (1984); Major (1986); Major (1987); Derwing &amp; Munro (1997); Magen (1998)</td>
</tr>
<tr>
<td>Sentence</td>
<td>McDermott (1986); Thompson (1991); Flege &amp; Fletcher (1992); Bongaerts, Planken &amp; Schils (1995); Flege, Munro &amp; MacKay (1995); Munro (1995); Munro &amp; Derwing (1995a); Munro &amp; Derwing (1995b); Bongaerts, van Summeren, Planken &amp; Schils (1997); Flege, Frieda &amp; Nozawa (1997); Munro &amp; Derwing (1998); Riney &amp; Flege (1998); Flege, Yeni-Komshian &amp; Liu (1999); Moyer (1999); Munro &amp; Derwing (1999); Riney &amp; Takagi (1999); Southwood &amp; Flege (1999); Guion, Flege &amp; Loftin (2000); Jilka (2000); Yeni-Komshian, Flege &amp; Liu (2000); Munro &amp; Derwing (2001); Piske, MacKay &amp; Flege (2001); MacKay &amp; Flege (to appear)</td>
</tr>
<tr>
<td>2 Sentences</td>
<td>Brennan, Ryan &amp; Dawson (1975)</td>
</tr>
<tr>
<td>4 Sentences</td>
<td>Asher &amp; Garcia (1969)</td>
</tr>
<tr>
<td>Clip – 10 seconds</td>
<td>Munro, Derwing &amp; Flege (1999)</td>
</tr>
<tr>
<td>Clip – 45 seconds</td>
<td>Oyama (1976); Derwing &amp; Rossiter (2003)</td>
</tr>
<tr>
<td>Clip – 2 minutes</td>
<td>Suter (1976)</td>
</tr>
</tbody>
</table>

Despite the prevalence of sentence-length stimuli in such studies, foreign accentedness judgments can be effectively made based on tokens shorter than a sentence. Flege (1984), for example, found that even 30 milliseconds of speech can be sufficient for raters to make a reasonably accurate decision as to whether a token was produced by a native speaker or a nonnative speaker. This was demonstrated by extracting the first 30 milliseconds of English /t/ segments produced by native speakers of French and native speakers of English. Using a forced-choice comparison task, judges were, on average, able to identify those tokens produced by the French speakers with 69% accuracy. On the other hand, in the same study Flege (1984) found that judges’ accuracy increased to 89% when full phrases were presented to them. It would thus seem that while very short tokens can be accurately judged, longer excerpts of speech are more informative for raters.
This being said, other studies have made use of much longer sections of text, either paragraphs of prose (e.g., Anderson-Hsieh, Johnson & Koehler 1992, Moyer 1999), or excerpts of narratives / descriptions ranging in length from 10 seconds (Munro, Derwing & Flege 1999) to two minutes (Suter 1976). Here, as with the shorter stimuli, good levels of interrater reliability are typically obtained, though, as McDermott (1986) points out, morphosyntactic or lexical errors can influence accentedness judgments when extemporaneous speech is used. It is therefore sometimes difficult to determine to what extent phonological variables are determining ratings of these longer tokens.

Also of relevance in considering the stimuli used in making global foreign accent ratings must be the ‘naturalness’ of the speech samples employed. The techniques used in eliciting speech samples range from direct repetition based on native speaker models (e.g., Olson & Samuels 1973, Snow & Hoefnagel-Höhle 1977), to delayed repetition with or without written support (e.g., Piske, MacKay & Flege 2001, Yeni-Komshian, Flege & Liu 2000), to reading (e.g., Major 1987, Munro & Derwing 2001), to extemporaneous speech produced in response to a prompt (e.g., Elliott 1995, Thompson 1991). This range of approaches raises questions about the extent to which data obtained under these different conditions can be effectively compared.

With this in mind, Oyama (1976) found that, generally, the extemporaneous speech of Italian-English bilinguals was deemed to be more accented than was the read speech, though scores for the two types of stimuli were highly correlated. Munro & Derwing (1994) hypothesized that such effects might be due to the ability of speakers in extemporaneous speech, but not in read speech, to select words and expressions with which they are comfortable and to avoid those that they find difficult. This hypothesis was directly tested by having speakers first recount a free narrative and then, a few days later, read an orthographically transcribed version of the same text. Under these conditions, no significant differences between the accentedness ratings given to the two types of stimuli were. It would seem, then, that whether extemporaneous or read speech is used is not crucial, provided that method of elicitation is controlled and comparisons of studies’ findings bear such factors in mind.

Still other studies have used data mechanically manipulated for speech rate (e.g., Munro & Derwing 1998), the correction of perceived segmental errors (e.g., Magen 1998), alternations in factors such as VOT (e.g., González-Bueno 1997), or even the removal of all segmental cues (e.g., Jilka 2000). While all of these factors likely do not affect the naturalness of the speech stream to the same extent, there appears to be at least some indication that naturalistic and synthesized data may not be judged in precisely the same manner. Thus, Flege & Munro (1994) found that when judges rated naturalistic exemplars of the word ‘taco’ they tended to focus attention on the beginning of the word, and particularly the VOT of the initial /t/ segment, whereas when judges rated synthesized exemplars of the same word they tended to focus attention more evenly throughout the word in making their judgments. It seems, therefore, that conclusions based on ratings of synthesized stimuli should be treated with some caution and, further, that direct comparisons of findings from global accentedness studies more generally should be conducted with care.

2.3 Speakers
By far the majority of global accent studies have examined the second language pronunciation of learners of English. Speakers in these studies have come from a range of L1 backgrounds, including Spanish, Russian and Korean. A much smaller number of studies have examined the global foreign accent of English speakers learning other languages, or, in case of Guion, Flege & Loftin (2000), the global foreign accent in Spanish of native speakers of Quichua. A summary of the language combinations considered in such studies is given in (3) and (4).

(3) **Learners of English**

<table>
<thead>
<tr>
<th>L1 background</th>
<th>Example studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brazilian Portuguese</td>
<td>Major (1986); Major (1987)</td>
</tr>
<tr>
<td>Chinese</td>
<td>Anderson-Hsieh &amp; Koehler (1988); Flege &amp; Fletcher (1992)</td>
</tr>
<tr>
<td>French</td>
<td>Flege (1984)</td>
</tr>
<tr>
<td>Japanese</td>
<td>Riney &amp; Flege (1998); Riney &amp; Takagi (1999); Riney, Takada &amp; Ota (2000)</td>
</tr>
<tr>
<td>Korean</td>
<td>Flege, Yeni-Komshian &amp; Liu (1999); Yeni-Komshian, Flege &amp; Liu (2000)</td>
</tr>
<tr>
<td>Russian</td>
<td>Thompson (1991)</td>
</tr>
<tr>
<td>Various</td>
<td>Suter (1976); Purcell &amp; Suter (1980); Tahta &amp; Loewenthal (1981); Piper &amp; Cansin (1988); Anderson-Hsieh, Johnson &amp; Koehler (1992); Derwing &amp; Munro (1997); Munro &amp; Derwing (2001); Derwing &amp; Rossiter (2003)</td>
</tr>
</tbody>
</table>

(4) **Learners of other languages**

<table>
<thead>
<tr>
<th>L2 being learned</th>
<th>L1 background</th>
<th>Example studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dutch</td>
<td>English</td>
<td>Snow &amp; Hoefnagel-Höhle (1977)</td>
</tr>
<tr>
<td>German</td>
<td>English</td>
<td>Olson &amp; Samuels (1973); Elliott (1995); Moyer (1999); Jilka (2000)</td>
</tr>
<tr>
<td>Spanish</td>
<td>English</td>
<td>González-Bueno (1997)</td>
</tr>
<tr>
<td></td>
<td>Quichua</td>
<td>Guion, Flege &amp; Loftin (2000)</td>
</tr>
</tbody>
</table>

Numbers of nonnative speakers from whom stimuli have been collected for such studies have ranged widely, from as few as one (González-Bueno 1997) or three (Anderson-Hsieh & Koehler 1988) to 240 (Yeni-Komshian, Flege & Liu 2000). Depending upon the purpose of the study, speakers with varying degrees of experience with the second language have been selected, or a relatively more homogeneous group has been used. The specific findings associated with these different types of approaches are discussed in section 3.

Control native speakers are generally included in such studies as well, with their main purpose being the establishment of native-speaker norms. Furthermore, such groups may serve to identify those raters who are unable to recognize native speech (and therefore consistently rate the speech of control speakers as accented). Typically, such raters are considered to be unreliable and their judgments are discarded before any statistical analysis is conducted.
This being said, it is worth noting that the relative proportion of native-speaker stimuli as compared to nonnative-speaker stimuli may have impacts upon the ratings given to nonnative speakers. Specifically, Flege & Fletcher (1992) suggest, based on a comparison of scores accorded to nonnative speakers in groups comprising 40% and 20% native-speaker controls, that range effects may play an important role in the specific raw scores given to nonnative speakers. Thus, where native speakers are relatively more numerous, the scores accorded to nonnative speakers may be depressed. If fewer (or no) native speakers are included among the stimuli, nonnative speakers’ scores may be relatively higher than might otherwise be the case. This, of course, further complicates the direct comparison of specific scores from different studies difficult at best, though trends may still be compared.

2.4 Listeners

Just as the number of speakers included in global accent studies has ranged considerably, so too has the number of individuals who have served as judges in rating stimuli. At the lowest end of the continuum, Snow & Hoefnagel-Höhle (1977) used only a single judge to rate all of the stimuli (though that judge rated all stimuli twice); at the other end of the spectrum, Anderson-Hsieh & Koehler (1988) made use of 224 judges, though not all of them rated all of the stimuli. Most studies, of course, fall between the two extremes (see section 5 for individual study details).

Beyond numbers, of course, selection of individual judges for the rating task is also an important consideration, though relatively little research has directly considered this issue. This being said, in the vast majority of cases, native speakers have been used in rating the accentedness of tokens, though in some cases near-native speakers have also been used (e.g., Elliott 1995; Olson & Samuels 1973). Given the finding that very few nonnative speakers ever achieve levels of accentedness within native speaker ranges (see section 3.1), however, it would seem prudent to avoid the use of nonnative speakers as judges wherever possible, and, instead, to confine accentedness judgments to native speakers.

Assuming that native speakers are used as raters, then, of particular interest must be those factors inherent to different judges that might influence the way in which they assess individual tokens. McDermott (1986) examined such listener factors in detail and concluded that while overall ratings of speakers may not always be significantly affected, individual raters differ in the degree of importance that they accord to different factors in making judgments. Specifically, she concluded that while some listeners make use of the same set of criteria in rating all speakers, other listeners may place greater or lesser importance on different criteria depending upon the speaker that they are rating. Such variation in the application of criteria may particularly cloud the ratings of speakers occupying the middle range of accentedness. Munro & Derwing (1995b) found similar variation in the specific criteria used by different raters in making judgments, noting that the degree to which comprehensibility and accentedness ratings were correlated differed considerably between individual judges.

McDermott (1986) further suggests that the backgrounds of individual listeners may affect the specific factors upon which they choose to focus in rating tokens. At the same time, however, she found that the listeners’ backgrounds were not associated with any significant differences in
the overall accentedness scores given. Brennan & Brennan (1981a) also considered this issue, comparing the ratings given to Spanish-accented speech by high school judges with and without Mexican ancestry. Again, no significant differences in overall accentedness ratings were noted, though the two groups did differ in the extent to which these ratings were correlated with feelings of solidarity. On the other hand, Flege, Frieda & Nozawa (1997) compared accentedness ratings given by native speakers of Canadian and Alabama English to Italian-accented speech and found that, overall, ratings given by native speakers of Canadian English were more accurate (i.e., more strongly correlated with factors such as age of L2 learning) than were those given by native speakers of Alabama English. The researchers suggest that this might be due to the Canadian group’s presumed greater familiarity with a range of foreign accents. It also seems plausible, however, that the fact that the Italian speakers were learning Canadian English may have also influenced the judgments of individual raters from the two groups, a supposition supported by the finding of Bongaerts, van Summeren, Planken & Schils (1997) that the dialect spoken by speakers may affect the way in which their accents are rated.

At least one study has indicated that training in linguistics and familiarity with other languages may lead to greater degrees of interrater reliability. Thompson (1991) compared the accentedness ratings given by two groups of eight judges, an experienced group (whose members spoke at least one foreign language fluently, had taken linguistics courses and had frequent contact with nonnative speakers) and an inexperienced group (whose members spoke no foreign languages fluently, had not taken linguistics courses and had only infrequent contact with nonnative speakers). The inexperienced group was found to generally evaluate tokens as more accented than did the experienced group. As well, a lesser degree of interrater reliability was obtained within the inexperienced group than within the experienced group. Flege and Fletcher (1992), for their part, compared the accentedness ratings assigned by a group of three neuroscience graduate students (considered to be ‘experienced’) and a group of eleven students who had less familiarity with foreign-accented speech. In this case, however, the ratings accorded by the two groups of listeners were comparable, suggesting that, contrary to the findings of Thompson (1991), inexperienced judges may not necessarily be less reliable than are experienced judges. Good levels of agreement between ratings made by experienced and inexperienced judges were also obtained by Bongaerts, van Summeren, Planken & Schils (1997) and Anderson-Hsieh & Koehler (1988).

When it comes to the training of listeners specifically for the task of rating global foreign accent, no study, to my knowledge, has compared the judgments made by trained and untrained raters. It is therefore unclear the extent to which task-specific training may affect the accuracy, reliability, or raw scores obtained in this type of study. This being said, there is considerable variation reported in the literature as to the type of training received by raters for this type of task. Among the most extensive training must be that reported by Suter (1976). The fourteen judges selected for this study all participated in a two-hour training session two days before the actual experiment, and also had the opportunity to listen to a ‘sampler’ tape of the first thirty seconds of each stimuli, aimed at familiarizing them with the range of accentedness, before completing any actual ratings (Brennan, Ryan & Dawson 1975 had judges listen to the full data set twice, only making judgments the second time). Furthermore, all judges selected for the study by Suter (1976) had completed an introductory linguistics course with an ‘A’ grade, a criteria instituted with the goal of ensuring that judges would be able to distinguish between the
phonological factors influencing accent and other grammatical characteristics of nonnative speech, a distinction thought to be particularly important given that extemporaneous speech was used as stimuli.

On the other hand, some studies (e.g., McDermott 1986) explicitly provide no training to judges, with the goal of capturing the perceptions of naïve native speakers and avoiding possible biases that training sessions might encourage. More commonly, however, judges complete a small number of practice trials using data not included among the test stimuli before beginning the actual experiment (e.g., Major 1986; Munro, Derwing & Flege 1999). As well, when the same stimuli are presented in multiple blocks, frequently the judgments given in the first block are considered ‘practice’ (though the raters are not always made aware of this) and are not included in statistical analyses (e.g., Flege, Frieda & Nozawa 1997; Guion, Flege & Loftin 2000; Riney & Takagi 1999).

With this in mind, it is worth noting that there is some experimental indication that, in fact, familiarization with the data set may lead to harsher judgments. Specifically, Flege & Fletcher (1992) found that accentedness judgments of a single data set were consistently harsher following a brief task designed to focus attention on specific aspects of the speakers’ foreign accents. These effects may hold true even when the data presented in two blocks are different. Thus, Munro & Derwing (1994) found that judges consistently rated the data set they heard second more harshly than the first data set, despite the fact that the two sets were not identical. It would therefore seem that randomization of stimuli presentation across judges is of considerable importance in order to prevent ordering effects and that more research on the effects of task-specific training in this regard is needed.

3. The Findings of Accentedness Studies

3.1 Speaker Background Factors

Turning from the methodology of global accent studies, a number of findings have emerged over the years with respect to those characteristics of speakers that are associated with either greater or lesser degrees of perceived foreign accent. Prime among these must the age at which an individual speaker was first exposed to / began learning his or her second language. Most such studies have indicated that earlier exposure is associated with lesser degrees of foreign accent. This is often noted in terms of a largely linear increase in degree of perceived accentedness as age of learning increases (e.g., Asher & García 1969; Flege & Fletcher 1992; Flege, Munro & MacKay 1995; Flege, Yeni-Komshian & Liu 1999; Oyama 1976; Piper & Cansin 1988; Piske, MacKay & Flege 2001; Moyer 1999; Tahta, Wood & Loewenthal 1981; Thompson 1991; Yeni-Komshian, Flege & Liu 2000), though Patkowski (1990) found some evidence of discontinuous progression such that the speech of those who began learning English between the ages of 5 and 15, as a group, was clearly less accented than that of those with an age of learning greater than 15. In at least two studies, judges have demonstrated the ability to distinguish, at statistically significant levels, the speech of people who began learning English at age 3 from that of native speakers of English (Flege, Munro & MacKay 1995; Flege, Yeni-Komshian & Liu 1999). Generally, then, an earlier age of learning has been associated with lesser degrees of foreign
accent (though not necessarily the attainment of native-speaker norms) with considerable consistency in the literature.

Still, some studies have found contrary results with respect to the influence of age of learning on degree of global foreign accent. Snow & Hoefnagel-Höhle (1977), in particular, found that older Anglophone learners of Dutch had better global pronunciation in their second language than did younger learners, at the early stages of acquisition. By the end of the year-long study, however, the younger subjects had surpassed the older subjects on the pronunciation of some sounds. Olson & Samuels (1973), for their part, found that junior high and college age subjects attained better pronunciation of German words after ten sessions of oral drills than did elementary-age subjects. In this case, however, students did not receive any kind of communicative language training, raising questions about the ability of these findings to inform our understanding of naturalistic language acquisition.

Further to this, it is worth noting that while in many global accent studies nonnative speakers are consistently distinguished from native speakers (see above), some studies have identified a limited number of nonnative speakers whose global accent scores fall within native-speaker norms. For example, in a study of 24 motivated graduate students in German, Moyer (1999) found one learner who was rated within native-speaker range. Bongaerts, van Summeren, Planken & Schils (1997) identified three highly-proficient native speakers of Dutch whose English pronunciation on all tokens was within native-speaker range. Still, such speakers are by far the minority among participants in the global accent studies reported in the literature, and are typically only found in studies which have expressly sought out highly successful L2 learners.

Of course, age of L2 learning is typically correlated with numerous other factors, including length of residence in a second language environment and amount of L1 use. Despite these correlations, however, these factors, in addition to age of L2 learning, have at times been found to exert independent influence. Thus, increased length of residence in an L2-dominant environment has been demonstrated in some studies to be significantly independently correlated with improved global foreign accent scores (e.g., Asher & García 1969; Flege & Fletcher 1992), although some studies (e.g., Oyama 1976) have found no such effect. First language use, for its part, has also been found to be significantly negatively correlated with global foreign accent in a number of studies. For example, Guion, Flege & Loftin (2000) found that native speakers of Quichua who continued to use their L1 more frequently were judged to have stronger foreign accents when speaking Spanish than were those who used Quichua on a less regular basis, even when age of learning was controlled for. Pronunciation in Quichua, for its part, was not significantly affected. Similarly, Yeni-Komshian, Flege & Liu (2000) found that accentedness scores among Korean-English bilinguals were significantly correlated with L1 use, such that those who used Korean more frequently typically had stronger foreign accents in English. In this case, however, the same principle held true for L1 pronunciation, with those who used Korean less frequently being more likely to have stronger perceived (English) foreign accents when speaking Korean. For most subjects, degree of accentedness in the two languages was inversely related.

Formal instruction, for its part, has only occasionally been reported to be associated with lesser degrees of foreign accent. Among such studies, Derwing & Rossiter (2003) found that the global
foreign accent scores of adult ESL students who received global pronunciation training (focusing on prosodic factors) improved between the beginning and end of an academic term whereas the scores of the control group and those who received training focusing on segmental factors did not. Similarly, Moyer (1999) found that lesser degrees of perceived foreign accent in German were correlated with having received a combination of segmental and suprasegmental pronunciation training, but not with having received segmentally-focused pronunciation training alone. Flege, Yeni-Komshian & Liu (1999), for their part, found that formal instruction was more strongly associated with rule-based aspects of morphosyntax than with perceived global foreign accent.

A number of other factors related to the speakers’ backgrounds, including gender (Asher & García; Flege, Munro & MacKay 1995; Thompson 1991), mimicry ability (Purcell & Suter 1980; Suter 1976; Thompson 1991), professional motivation (Moyer 1999), immersion in an L2 environment (Elliott 1995; Purcell & Suter 1980; Riney & Flege 1998), attitude toward the second language (Elliott 1995) and field independence (Elliott 1995) have also been found to be significantly correlated with global accentedness scores in at least some studies. These results, however, have not been consistently replicated, and require further examination before any firm conclusions can be drawn.

As it stands there is reasonable support for the influence of at least age of L2 learning, length of residence, L1 use, and, possibly, the role of global pronunciation training on global foreign accent. Researchers considering such studies should therefore seek to control for these factors where relevant and possible.

3.2 Phonetic Factors

We turn now to the specific characteristics of the tokens produced by speakers have also been associated, in various studies, with degrees of perceived foreign accent. One of the broadest such considerations must be that of Magen (1998) who edited English phrases produced by native speakers of Spanish to correct elements thought to be associated with foreign accent. In this context, adjustments to syllable structure (such as removing epenthetic segments), consonant manner of articulation and word stress were found to produce the most substantial effects in decreasing degree of perceived foreign accent. Adjustments to voice onset time, on the other hand, had little effect. Many of these individual aspects of the stimuli have been considered in other studies as well (often with similar results), as will be reviewed in the remainder of this section.

González-Bueno (1997), for her part, considered the role of stop voicing by manipulating the voice onset time of the initial segment [k] in the Spanish word ‘casa’ [kasa] spoken by a native speaker of English. In judging the foreign accentedness of the single word, the raters identified those instances where the VOT of the [k] was between 15 and 35 milliseconds as most native-like, suggesting that in controlled contexts VOT can influence degrees of perceived accentedness. Using natural stimuli collected in a longitudinal study of the English pronunciation of Japanese learners, on the other hand, Riney & Takagi (1999) only found limited support for a correlation between the VOT of stop segments and global foreign accent ratings. Over the forty-two months of their study, only three learners showed significant improvements in
global foreign accent and the English VOT of only one of these speakers showed movement toward English native-speaker norms. The overall role of voice onset time in contributing to global foreign accent in naturalistic speech, then, was not strongly supported.

The accuracy of liquid pronunciation has also been considered within this context, though, again, the relationship between segmental accuracy and global foreign accent is not been firmly established. Riney, Takada & Ota (2000), for example, found that substitution of the Japanese flap for the liquids /l/ and /r/ in L2 English speech was significantly correlated with global foreign accent (such that lesser degrees of such substitution were associated with lesser degrees of perceived foreign accent) only at the beginning of the longitudinal study; no such association was evident at the end. Riney & Flege (1998) point out, based on the same data, that while liquid identifiability and global foreign accent followed similar paths for some speakers, this relationship did not hold true for others.

A few studies have considered the contribution of vowel quality to the perception of foreign accent in speech. Munro, Derwing & Flege (1999) found, for their part, that the pronunciation of the diphthong /aj/ alone was sufficient to allow experienced raters to distinguish between native speakers of Canadian and Alabama English dialects with a high degree of accuracy. Major (1987) considered the pronunciation of English /æ/ and /ɛ/ by native speakers of Brazilian Portuguese and found that as global foreign accent scores and accuracy of English /æ/ pronunciation improved, /ɛ/ pronunciation actually became less accurate. It would seem, then, that when it comes to the phonetic implementation of individual segments – whether they be stops, liquids or vowels – the relationship to global foreign accent is not as straightforward as one might assume, a finding also reflected in the lack of influence of purely segmental pronunciation training on global foreign accent (see section 3.1).

Moving beyond the individual segment, however, prosodic accuracy seems to be more closely linked to global foreign accent scores. Thus, the influence of epenthesis on perceived foreign accent identified by Magen (1998) is also reflected in Major (1986). Specifically, Major (1986) found that, among native speakers of Brazilian Portuguese learning English, higher rates of epenthesis were significantly correlated with stronger global foreign accent. The use of epenthetic [i] (as opposed to schwa) was particularly indicative of stronger foreign accent.

Other prosodic aspects of speech have also been demonstrated to be correlated with global foreign accent. Munro (1995), for example, found that even when all segmental information was removed from the speech stream judges were able to distinguish between English passages spoken by native speakers of English and native speakers of Mandarin. Jilka (2000) found similar results, with the accuracy of sentence-level intonation being significantly correlated with the degree of perceived foreign accent in the German speech of native speakers of English. Anderson-Hsieh, Johnson & Koehler (1992) echo the importance of prosodic factors in influencing perceived foreign accent, identifying them as more important than both segmental and syllable structure factors in their study of English learners from a range of L1 backgrounds.

The influence of speech rate has also been considered by some researchers. MacKay & Flege (to appear), for example, found that among late Italian-English bilinguals shorter sentences were perceived to be less foreign accented. Munro & Derwing (1998), for their part, found that the
English speech of native speakers of Mandarin was deemed to be more accented when slowed and that at least some speakers’ accents were found to be less strong when their speech was speeded. Further to this, Munro & Derwing (2001) suggest that the natural speaking rate of nonnative speakers is typically somewhat slower than optimal, and found speech that was slightly speeded to be correlated with improved foreign accent. On the other hand, the authors note that too great of an increase in rate of speech may, in fact, impede comprehensibility.

Finally, comprehensibility and intelligibility have also been found to be correlated with global foreign accent scores. Munro & Derwing (1995a, 1999), for example, found a moderate correlation between the three factors (comprehensibility, intelligibility and global foreign accent) in the English speech of native speakers of Mandarin. Munro & Derwing (1995b), however, point out that the degree of correlation between comprehensibility and global accentedness varied considerably between individual raters (see also section 2.4). Thus, while accentedness scores may be correlated with the comprehensibility of speech, highly accented speech can still be intelligible and comprehensible to native speakers, dependent upon the context and the specific characteristics of the given utterance (see also Derwing & Munro 1997).

4. Summary

Studies employing global foreign accent ratings have been used by a number of researchers who have considered the correlations of these ratings to a wide range of variables. As such, these studies have demonstrated that numerous factors, including dialect, range of variation in the stimuli, length of tokens, and individual, idiosyncratic characteristics of raters, can influence the way in which judgments are made. At the same time, speaker characteristics, such as age of learning, L1 use and length or residence, as well as token characteristics such as prosodic accuracy have been demonstrated to have significant effects on ratings accorded to individual speakers. In the following section, the specific findings of some studies which have employed global foreign accent ratings are summarized.

5. Study Details


This study compared the comprehensibility and perceived accentedness of English passages read by three native speakers of Chinese and one control native speaker of English. It was found that both the accentedness ratings assigned by naïve judges and the passages’ comprehensibility were correlated with the TSE (Test of Spoken English) pronunciation scores of the speakers. The effect of speaking (reading) rate was also assessed, and it was found that, for all speakers, comprehensibility of the ‘fast’ passage was significantly less than that of the same passage read at ‘regular’ or ‘slow’ rates. Few effects of speaking rate on accentedness were noted, though in one case a slow passage was deemed significantly less accented than the corresponding ‘regular’ or ‘fast’ versions. There was also some indication that prosodic rather than segmental factors might be more important in determining accentedness rating (particularly at faster rates of speech). Judges who held more positive attitudes toward foreigners and foreign accents gave
slightly (but significantly) better accentedness ratings to the fastest and most accented passages. In addition, it was noted that speakers with more pronounced accents were perceived as speaking more rapidly, even when this was not, objectively, the case.

- **Judges:** 224 native speakers of English acted as judges in the study. Each judge only rated a limited number of passages.
- **Scale:** A five-point accentedness scale, ranging from ‘heavy foreign accent; very difficult to understand’ to ‘no foreign accent; very easy to understand’ was used (the accentedness scale was based on the TSE descriptors). In addition, judges rated the speaking rate on a five-point scale of ‘too slow’ to ‘too fast.’ In both cases, each point on the scale was labeled with a descriptor. Comprehensibility was assessed by having raters answer multiple-choice comprehension questions based on the passages.
- **Training:** No specific training was provided, though background information about each judge was collected.
- **Stimuli:** Ratings were based on six read passages of 310 to 475 syllables. Speakers recorded the passages at three rates: regular, slow and fast. The native speaker recorded all six passages, while the three nonnative speakers each recorded two passages.


Segmental, prosodic, and syllable structure errors were compared with global foreign accent ratings for 60 non-native speakers of English from a range of L1 backgrounds and proficiency levels. It was found that, overall, errors in all three domains were significantly correlated with global accent / intelligibility, but that prosodic factors accounted for the greatest amount of variance in accentedness. It is suggested that the relative importance of segmental versus syllable structure errors in relation to prosodic factors may vary based on speakers’ first languages.

- **Judges:** Three experienced ESL teachers who had served as SPEAKtest raters in the past acted as judges in the study.
- **Scale:** A seven-point scale ranging from ‘heavily accented and unintelligible’ to ‘near-native’ was employed. The individual descriptors were largely similar to those used on the SPEAKtest, although more distinctions were made.
- **Training:** Judges were experienced SPEAKtest raters.
- **Stimuli:** Paragraph-length passages read by 60 male non-native speakers of English served as the stimuli for the study. A short excerpt of the same passage read by a native speaker of English was played between each experimental stimulus to provide judges with a native-speaker reference point.

This study compared degrees of perceived foreign accent with the age of arrival in the US of 71 Cuban immigrant children. Thirty control native speakers of American English were also included in the study. All 71 Cuban children were identified as non-native speakers of English by the judges, and age of arrival was a significant predictor of individuals’ degree of perceived foreign accent. Thus, 68% of children with an AOA of between 1 and 6 achieved near-native pronunciation scores whereas only 7% of children with an AOA above 13 did. Length of residence was also a significant predictor of accentedness scores, with 51% of subjects who had lived in the US for five to eight years achieving near-native pronunciation scores and only 15% of those who had lived in the US for four years or less achieving similar scores. Length of residence, however, was a less important variable than age of arrival. Gender also accounted for significant variance, with females achieving better accentedness scores than did males, especially among those who had lived for less time in the US.

- **Judges:** 19 American high school students acted as judges in the experiment.
- **Scale:** A four-point scale was employed in the study, with individual points labeled ‘native speaker,’ ‘near native speaker,’ ‘slight foreign accent’ and ‘definite foreign accent.’
- **Training:** Examples of each point on the scale were given to raters before the experiment began.
- **Stimuli:** Each speaker read, rehearsed and then recorded a series of four sentences specifically selected to contain segments and sequences of segments thought to be difficult for native speakers of Spanish. Raters accorded a single score for the full series of four sentences from each speaker.


This study compared the foreign accentedness scores of three groups of subjects: a control group of 5 native speakers of English, a group of 10 native speakers of Dutch deemed to be highly successful learners of English, and a group of 12 native speakers of Dutch with a wide range of levels of accent in English. All of the nonnative speakers had begun learning English at or around the age of 12. Four tasks were included in the study: an extemporaneous narrative, a short text (read), 10 English sentences (read), and 25 English words (read). On all tasks the scores of the third group were significantly lower than those of groups one and two. The scores of group two (highly proficient nonnative speakers) were not significantly worse than those of the native control group on any task, and were actually higher than those of the control group on the sentence reading task. The results are taken to indicate that there are nonnative speakers who can achieve fully native-like accents. Still, it is worth noting that in this case the native speaker control subjects were not consistently identified as such by the judges. It is suggested, therefore, that the fact that the control subjects spoke a somewhat different regional dialect of English than did the judges and that the highly proficient nonnative speakers had received explicit training in Received Pronunciation may have influenced the raters’ judgments.

- **Judges:** 4 native speakers of British English (from a different region than the control subjects) acted as raters in the study. All were considered to be ‘naïve.’
• **Scale:** A five-point scale, ranging from ‘very strong foreign accent: definitely non-native’ to ‘no foreign accent at all: definitely native’ was employed. Descriptors were also included for the midpoints.

• **Training:** Judges completed two practice samples for each task before beginning to rate the experimental stimuli.

• **Stimuli:** Stimuli from each task were presented to the judges separately in pseudo-randomized order. In the case of the narratives, excerpts 16 to 20 seconds in length were extracted to control for vocabulary and syntax.


This study compared perceived foreign accent in the English speech of 10 native speakers of (nonregional) British English, 11 native speakers of Dutch identified as being highly successful L2 learners of English, and 20 native speakers of Dutch at a range of L2 proficiency levels. All non-native speakers had begun learning English at age 12 or later. Unlike in the study reported by Bongaerts et al. (1995), which found low scores for NSs and an inability of judges to distinguish them from the highly proficient group, in this study NSs more were highly rated, with mean scores significantly above those of the other two subject groups. Nonetheless, five L2 speakers scored within two standard deviations of the native speaker mean and were thus considered to be within native speaker ranges. For three of these speakers, this was true on all of the stimuli considered. It is thus suggested that L2 learners who begin learning their second language after puberty are sometimes able to achieve accentedness scores similar to those of native speakers. No significant differences were noted between the strictness of the scores given by the experienced and inexperienced judges in the study.

• **Judges:** 13 native speakers of English deemed to speak a nonregional variety of British English acted as judges in the study. Six of these judges were either EFL teachers of phoneticians and were thus considered ‘experienced;’ the other seven judges had no formal training in languages or linguistics and were considered ‘inexperienced.’

• **Scale:** A five-point scale, ranging from ‘very strong foreign accent: definitely nonnative’ to ‘no foreign accent at all: definitely native,’ was employed. Midpoints were also labeled.

• **Training:** Ten practice items were included before the experimental ratings began.

• **Stimuli:** Each subject read six sentences three times each. These sentences were specifically selected to include a wide range of both more difficult and less difficult English phones.


In this study assessments of global accent made by naïve raters were compared with ratings of solidarity and status made by these same raters. All speech samples in the study came from nine Mexican-American speakers of English whose speech as accented to various degrees. The accentedness ratings given by the judges were found to be highly correlated with ‘Accentedness Index’ scores independently assigned by the researchers. Global accentedness ratings were also found to be correlated with status scores, with those whose speech was deemed more accented
also being thought to be of lower status. No significant differences between the accentedness ratings made by Mexican-American and Anglo-American raters were found, though solidarity scores were significantly correlated with perceived accentedness among the Mexican-American raters.

- **Judges:** 80 high school students served as raters in the study. 43 of these indicated that their families were of Mexican origin, while 37 indicated no such ties.
- **Scale:** Direct magnitude estimation was used in assessing global accent in this study. Status and solidarity scores were attained through the use of eight seven-point scales, the endpoints of which were given labels such as ‘kind/cruel,’ ‘successful/unsuccessful.’
- **Training:** All judges completed a brief training session before completing the experimental judgments. For accentedness, this included rating the length of lines using direct magnitude estimation, listening to the full tape once before giving any judgments, and an opportunity to ask questions. Judges completed a practice rating of one non-test speaker and had an opportunity to ask questions before completing status / solidarity ratings.
- **Stimuli:** A single short paragraph read by nine speakers with differing degrees of foreign accent was used as the stimulus in the study.


This study compared global foreign accent scores of the English speech of native speakers of (Mexican) Spanish with the occurrence of 18 specific phonological variables thought to be associated with Spanish accent in English. Indeed, 11 of the individual variables were found to be highly significantly correlated with global accent ratings. This was true both of the ratings completed by trained linguists and of the ratings completed (in an earlier study) by naïve native speakers of English. Data from 32 native speakers of Spanish and 4 control native speakers of English were included in the study.

- **Judges:** 3 doctoral candidates in linguistics.
- **Scale:** A seven-point scale ranging from ‘Spanish accent’ to ‘no Spanish accent’ was employed. Individual variables were assessed on a presence / non-presence basis.
- **Training:** The judges were all trained in phonetics and were familiar with both the characteristics of Spanish accent and the local English dialect.
- **Stimuli:** 1.5 minute selections from a read (five-minute long) text were used as stimuli. The specific selection chosen for rating contained instances of all of the specific phonological variables of interest in the study.


In this study judges rated the degree of foreign accent found in the English L2 speech of 8 Spanish-English bilinguals using two techniques: magnitude estimation (where listeners assigned proportional number scores based on degree of perceived accent) and sensory modality matching (where listeners squeezed a hand dynamometer more or less tightly based on degree of perceived
accent). Results using both approaches were found to be correlated with the number of phonological errors independently tabulated by the researchers.

- **Judges**: 72 speakers of standard American English served as raters in the study.
- **Scale**: No formal scale was employed. Rather, for the direct magnitude estimation task, listeners assigned a number they deemed appropriate to the first example and then a proportionate number to following examples based on the amount of foreign accent (e.g., if the second example was 3 times as accented as the first example, they would give a number 3 times as great as that given to the first example – regardless of what the raw numbers were).
- **Training**: Listeners heard the whole tape once before giving any formal ratings. In addition, all listeners completed a full set of 8 ratings of separate sentences before beginning the actual test stimuli.
- **Stimuli**: Ratings were based on a read passage two sentences in length. All 8 speakers read the same passage.


In this study global accent ratings were compared with the intelligibility (measured based on listeners’ transcriptions) and comprehensibility (rated by listeners based on their perceptions of the speech’s intelligibility) of nonnative speech. Data from 48 intermediate-level learners of English were included in the study; speakers were divided into 4 groups of 12 based on first language background. Native speakers of Cantonese, Polish, Japanese and Spanish were included. It was found that while all three of the factors under consideration (accent, intelligibility and comprehensibility) were interrelated, they were not strictly equivalent. Furthermore, it was found that comprehensibility and intelligibility could remain relatively high even when speech was judged to be fairly heavily accented.

- **Judges**: 26 native speakers of Canadian English acted as judges in the study. Ten of these indicated some degree of familiarity with a language other than English.
- **Scale**: A nine-point scale ranging from ‘no accent’ to ‘extremely strong accent’ was employed in measuring global foreign accent. As well, a nine-point scale ranging from ‘extremely easy to understand’ to ‘extremely difficult or impossible to understand’ was used to measure comprehensibility.
- **Training**: All judges answered questions regarding their familiarity with other languages and specific foreign accents.
- **Stimuli**: Full phrases ranging from 4.5 to 10.5 seconds were taken from the beginning of the speakers’ retellings of a story based on cartoon images.


Global accent, fluency and comprehensibility were compared for 48 intermediate ESL students (from a variety of L1 backgrounds) who underwent 12 weeks (20 hours) of pronunciation training. Three experimental conditions were devised: Segmental (where training focused on segmental phonology), Global (where training focused on prosodic factors), and Control (a group
that received no specific pronunciation training). In the end, the ratings for comprehensibility and fluency of the Global group improved between times 1 and 2; no significant improvements were found for members of the other groups. Based on an analysis of specific errors, it was suggested that the extensive focus on segmental factors of the Segmental group might have detracted from their ability to focus on more global fluency and prosodic aspects of speech.

- **Judges**: Six ESL teachers with extensive background in pronunciation training served as judges in the study.
- **Scale**: *Nine*-point scales were employed. The comprehensibility scale ranged from ‘very easy to understand’ to ‘impossible to understand,’ the accentedness scale ranged from ‘no accent’ to ‘very strong accent’ and the fluency scale ranged from ‘very fluent’ to ‘extremely dysfluent’ (fluency was taken to refer to speech rate / hesitations).
- **Training**: Judges were experienced ESL teachers. Three practice stimuli were rated at the beginning of each session.
- **Stimuli**: Stimuli consisted of 45-second excerpts from narratives recounted by students based on a series of cartoon pictures. The same pictures were used at both time 1 (beginning of training) and at time 2 (11 weeks later). Excerpts were randomized during rating sessions.


This study compared the Spanish pronunciation of 66 native speakers of English (taking undergraduate intermediate Spanish) with a number of variables. No native speaker control group was included. Three tasks focused on the pronunciation of specific segments, while in the final task global pronunciation was rated over a narrative text. In this case, it was found that attitude toward Spanish pronunciation accounted for the greatest degree of variance in global accentedness scores, followed by field independence (linked to better pronunciation) and right-hemisphere specialization. Attitude, itself, was also significantly correlated with years of Spanish instruction and travel to Spanish-speaking countries, both of which were independently associated with better pronunciation scores. In addition, global accent scores were also highly correlated with scores on the first three tasks where pronunciation of individual segments within words and phrases was assessed.

- **Judges**: Three judges rated the stimuli in this experiment. One was a native speaker of Spanish while the other two were graduate students specializing in Spanish (including the author of the study).
- **Scale**: A *five*-point scale ranging from ‘very poor accent’ to ‘native’ was used in rating the narrative texts.
- **Training**: All judges had significant academic training in phonology and Spanish.
- **Stimuli**: Global accent was rated based on a 90 second narrative based on a picture. This was preceded by word reading and repetition tasks, and by a sentence repetition task, each of which were used only to assess the accuracy of the pronunciation of individual segments.

Five experiments were conducted assessing the ability of native speakers of English to recognize French (non-native) accent in English speech. Judges did not rate degrees of foreign accent, but, instead, in different tasks, either identified tokens as native or non-native or directly compared two stimuli and select the one they believed to be non-native. It was found that, as a group, both experienced and inexperienced judges were able to correctly identify non-native stimuli in a significant proportion of cases. When phrases were used, non-native tokens were accurately identified 89% of the time. When only syllables, /tu/ or /ti/, were used, accuracy rates for non-native tokens were 95% in the forced choice task and 77% in independent choice task. Using hybrid syllables (with one segment produced by a NS of French and one produced by a NS of English), the presence of a French segment prompted an assessment in a forced-choice task of ‘non-native’ in 67% of cases. Finally, when only the initial 30ms of a /t/ segment were presented, 69% of non-native tokens were accurately identified in a forced-choice task. In this final case, however, significant variation existed among the accuracy rates achieved by individual judges. This study was taken by the researcher to indicate that non-native speech can be correctly identified based on as little as 30ms.

- **Judges**: Nine to twelve native speakers of English with different degrees of phonetics and linguistics training acted as judges in the different experiments.
- **Scale**: A Likert scale was not used in this experiment. Rather, judges simply identified tokens as native or non-native in free or forced-choice tasks.
- **Training**: No indication of specific training is given. Judges in experiments 1, 2, 4 and 5 all had some degree of linguistics or phonetics training. Judges in experiment 3 were considered to be naïve.
- **Stimuli**: In the first experiment two phrases were used, in experiments two and three single syllables were used, in experiment 4 hybrid syllables created by splicing segments were used, and in experiment 5 the first 30ms of a /t/ segment were used.


Four experiments measuring degree of perceived foreign accent in English were conducted involving native speakers of Spanish and Chinese (Mandarin and Taiwanese). In the first experiment, ratings accorded to 50 speakers – 10 each of native speakers of Alabama English, Texas English, early (age 5-6) Spanish learners of English, late experienced (LOR 14.3 yrs) Spanish learners of English, and late inexperienced (LOR 0.7 yrs) Spanish learners – were compared. Early L2 learners’ scores did not differ significantly from the scores accorded to either group of native English speakers. Length of residence was a significant predictor of accent among L2 learners, as were years of formal English instruction and age of arrival in the US. In the second experiment, native speaker data was excluded and data from 30 NSs of Spanish (same as experiment 1) and 37 NSs of Chinese were instead compared. LOR effects similar to those found in experiment 1 were again noted, and it is suggested that ratings given to speakers are independent of L1 background. It is further suggested that the size of the NS control group (particularly the relatively large number of NSs used in experiment 1 as compared to the smaller number involved in the original experiment involving the NSs of Chinese) may
significantly affect the scores accorded to talkers. In the experiment 3, 7 NSs of Spanish and 7 NSs of Chinese, all of whom were considered to have highly accented speech in experiment 2, were compared. Here, sentences were rated twice, with a brief exercise designed to focus raters’ attention on phonetic factors taking place in between. Ratings according on the second pass were significantly lower than those accorded on the first; it is suggested, therefore, that through the training listeners became more aware of divergences from NS norms without becoming more tolerant. Good degrees of interrater reliability were attained, suggesting that fine-grained judgments can be made even when range of accentedness is limited. Finally, in experiment 4 sentences spoken by 10 NSs of English, 7 early NSs of Chinese and 10 early NSs of Spanish were compared. In this case NSs of English and Spanish were rated as significantly better than NSs of Chinese, a finding attributed to the later mean age of learning of the Chinese subjects. Again, good interrater reliability was obtained.

- **Judges:** In the first experiment, 10 monolingual NSs of American English acted as judges. In the second experiment, two groups of English NSs were employed, 3 experienced neuroscience graduate students and 11 inexperienced raters (both groups attained similar results). In experiments 3 and 4 separate groups of 9 monolingual native speakers of American English acted as judges.
- **Scale:** In all four experiments, a sliding scale, interpreted as giving scores between 1 and 256, was employed. Only the ends and midpoint of the scale were labeled (‘no foreign accent,’ ‘medium foreign accent’ and ‘strong foreign accent’).
- **Training:** No specific accent-rating training was provided in any of the experiments, though in each case judges were urged to use the full scale.
- **Stimuli:** Three separate sentences were read by the speakers and a selection of these was used in the experiments. In experiments 1 and 4, all three sentences were used; in experiments 2 and 3, only one sentence was used.

---


This study was a reanalysis and reevaluation of data collected in an earlier study (Flege, Munro and MacKay 1995). Data (three English sentences) from 60 subjects (20 native speakers of English, 20 native speakers of Italian who used Italian frequently [HiUse] and 20 native speakers of Italian who used Italian infrequently [LoUse]) were submitted to global accent ratings. Two groups of judges rated the sentences – 12 native speakers of Canadian English and 12 native speakers of Alabama English. It was found that overall all of the native speakers of Italian had detectable foreign accents, though these were more pronounced among the HiUse group than among the LoUse group, despite the fact that all came to Canada as children. It is therefore claimed that it is amount of L1 use (activation) that is the critical factor in determining the amount of L1 phonetic interference in the L2, and not age of acquisition. Also, it was found that the native speaker of Canadian English judges were slightly, but significantly, better at assessing foreign accent than were the American judges (a claim based on the fact that the Canadian ratings were more sensitive to factors such as AOA and HiUse vs. LoUse).

- **Judges:** 12 native speakers of Canadian English and 12 native speakers of Alabama English acted as judges. The judgments of the Canadian English speakers were held to be more
accurate; the authors suggest that this may be due to that group’s greater familiarity with a range of accents and dialects.

- **Scale**: A *four*-point scale was employed, with individual buttons to be pushed for each point. The points were labeled: ‘definitely Italian,’ ‘probably Italian,’ ‘probably English’ and ‘definitely English.’

- **Training**: Judges were told they would hear sentences, some of which would be spoken by native speakers of Italian, and that they were to judge the background of the talkers using the 4-point scale. All passed a hearing test.

- **Stimuli**: Recordings of 3 *sentences* by 60 speakers were randomized in each block with a pause of 1.5 seconds between each sentence. Each listener completed 4 blocks (i.e., judged each sentence 4 times) but only the judgments from the final three blocks were analyzed.

---


This study compared the global accent ratings obtained and reported in Piske et al. (2001) with various measures of bilingual dominance (self-ratings, translation efficiency, etc.) obtained from the same 72 Italian-English bilinguals. It was found that age of arrival in Canada, length of residence and L1 (Italian) use were all associated with language dominance, but that language dominance also had independent effects. With respect to global accent then, all of these factors were important. New to this study, though, it was found that sentences produced by balanced bilinguals and Italian-dominant bilinguals, but not English-dominant bilinguals, received significantly lower ratings (i.e., were more foreign-accented) than were sentences produced by the native English controls.

- **Judges**: 9 monolingual native speakers of Canadian English acted as judges in the experiment.

- **Scale**: A *nine*-point scale was employed, ranging from ‘very strong foreign accent’ to ‘no foreign accent.’

- **Training**: See Piske et al. (2001)

- **Stimuli**: Three English *sentences* produced by the native English controls and the 72 Italian-English bilinguals were used as stimuli in the experiment. Each version of each sentence was presented three times to each of the judges. See Piske et al. (2001)

---


This study examined the role of various phonetic factors in the perception of the single word ‘taco’ as relatively English- or Spanish-accented. Stimuli produced by native speakers of Spanish and English were used in experiment 3, synthetic stimuli in experiment 4, and stimuli produced by native speakers of Spanish (who were either early or late learners of English) were used in experiment 5. It was found that listeners made use of phonetic factors across the full word in making judgments. This was particularly true in the case of the synthetic stimuli. The VOT of the initial /t/ was given a relatively heavier weight by listeners judging natural stimuli, though other phonetic factors also accounted for important variance. It is argued that the word is a meaningful unit within the L2 phonological system.
• **Judges:** In experiment 3, 3 trained speech scientists familiar with Spanish-accented English acted as judges. In experiments 4 and 5, 30 monolingual native speakers of English acted as judges (15 of whom used the rating scale and 15 of whom engaged in a forced-choice task).

• **Scale:** In experiment 3, a *sliding* scale (interpreted as 1 to 256) was used. In experiments 4 and 5, a *nine*-point scale (ranging from ‘least English-like’ to ‘most English-like’) was used.

• **Training:** Judges in experiment 3 were trained speech scientists. Judges in experiments 4 and 5 did not receive any specific training.

• **Stimuli:** In experiment 3, 42 examples of the word ‘taco’ produced by native speakers of English and Spanish were used. In experiment 4, 63 exemplars of the same word, synthetically produced, were used (each word was presented 8 times). In the final experiment, tokens produced by 14 native speakers of Spanish were used.


In this study global accent ratings were compared to the age at which native speakers of Italian began to learn English. 240 native speakers of Italian and 24 native speakers of Canadian English participated in the study. It was found that ‘age of learning’ was the most significant variable in accounting for degree of global foreign accent. All judges ranked the native speakers of English, as a group, above any of the L2 speaker groups. The most discerning judges were even able to distinguish the foreign accent of subjects with a mean AOL of 3.1 years (based on scores at least 2 standard deviations from the control subjects’ norms). The least discerning judges identified foreign accents only in subjects with a mean AOL of 15 years or greater. Other factors (including extent of English use, age of arrival and gender) also played a role, though AOL was the most important.

• **Judges:** 10 native speakers of Canadian English acted as judges for the study.

• **Scale:** A *sliding scale* was used, with judges positioning a lever along a continuum. The endpoints were labeled ‘no foreign accent’ and ‘native speaker of Italian – strongest foreign accent,’ while the midpoint was labeled ‘medium foreign accent.’

• **Training:** Judges were told that they would hear speakers, the majority (but not all) of whom were native speakers of Italian. They were further told to reserve the lowest point on the scale for the one speaker with the strongest foreign accent and to use the top of the scale only for native speakers of English. The first rating for each exemplar was not included in the statistical analysis (raters were told that it would serve as ‘practice’).

• **Stimuli:** Five *sentences*, elicited through a delayed repetition technique, served as stimuli in the study. Each exemplar was presented to the judges four times.


This study compared degree of global foreign accent in English and scores on a morphosyntactic grammaticality judgment test with age of arrival in the United States of 240 native speakers of Korean. Data from 24 native-English-speaking control subjects were also included in the study. It was found that AOA was the most significant factor in determining variance in degree of foreign accent. This relationship was essentially linear in nature. The utterances of speakers in
each AOA group as a whole (including the group with an average AOA of 3 years) were deemed significantly more accented than were those of native English speakers. Language use patterns were also significantly independently correlated with accentedness scores (with more English-language use being associated with a lesser degree of foreign accent). Similarly, higher numbers of morphosyntactic judgment errors were correlated with higher AOA. However, here it was suggested that ‘rule-based’ and ‘lexically-based’ errors might be subject to different factors, with lexically-based errors being more heavily influenced by language use patterns (similar to global accent) and rule-based errors being more heavily influenced by formal education in English.

- **Judges**: 10 native speakers of English acted as judges of global foreign accent.
- **Scale**: A nine-point scale of global foreign accent was used. It ranged from ‘very strong foreign accent’ to ‘no accent.’
- **Training**: Judges completed 22 practice sentences that spanned the range of accentedness before beginning to rate the experimental stimuli. As well, they were instructed to use the full scale in making their judgments.
- **Stimuli**: Five sentences elicited using a delayed repetition technique (with written support) were used as stimuli in the experiment. The same five sentences were used for each of the 264 subjects.


In this study accentedness ratings were used to assess the effect of VOT and aspiration on the perceived foreignness of second language Spanish speech (Spanish as spoken by native speakers of English). A single item ‘casa’ [kasa] was recorded from an L2 learner of Spanish and the VOT and aspiration modified in 5ms increments. It was found that [k] was perceived as most native-like when the VOT was between 15 and 35ms. Aspiration had no significant effect.

- **Judges**: 18 monolingual native speakers of Spanish who reported normal hearing served as judges in the study.
- **Scale**: A seven-point scale was employed, ranging from ‘native’ to ‘foreign.’ As well, a five-point confidence scale was used, ranging from ‘sure’ to ‘not sure.’ Listeners recorded their responses on an answer sheet.
- **Training**: Judges were given a brief overview of the procedure and encouraged to use the full scale. Information regarding the judges’ educational backgrounds and familiarity with foreign accents in Spanish was also collected.
- **Stimuli**: 28 variations of the single-word ‘casa,’ modified for VOT and aspiration, were rated in the study. Each variation was repeated 3 times for a total of 72 items which were randomly ordered.


This study examined the degree of foreign accent in the Quichua (L1) and Spanish (L2) pronunciation of 30 bilingual speakers who used the L2 to different extents in the bilingual community of Otavalo, Ecuador. Five Spanish monolinguals and five near-monolingual Quichua speakers were also included. Twelve of the thirty bilinguals received native-like ratings in Spanish. Others had varying degrees of Quichua accent in Spanish; degree of accent in the L2
was correlated with the extent to which speakers used the L1 (i.e., those who used Quichua more typically had a stronger Quichua accent when speaking Spanish). All bilinguals had begun acquiring Spanish between the ages of 5 and 8. Accent in Quichua was not affected by the extent to which speakers used Quichua in day-to-day life (all attained scores similar to the near-monolinguals); in other words, the L1 system did not appear to be influenced by extent of L2 use.

• **Judges:** 5 native speakers of Spanish rated the Spanish sentences and 5 near-monolingual speakers of Quichua rated the Quichua sentences.

• **Scale:** A nine-point scale was used, ranging from ‘most accented’ to ‘without accent.’ The scale was presented as individual buttons on a computer screen.

• **Training:** Judges were told to use all nine buttons and were informed of the type of accent being rated (i.e., Spanish for those rating Quichua and Quichua for those rating Spanish). Only the last three ratings for each sentence were included in the statistical evaluation.

• **Stimuli:** Stimuli were repeated sentence-length utterances matched for syllable-count and usage of high-frequency vocabulary. Each of the recorded sentences was repeated 4 times (35x4=140) in blocks such that all examples of sentence 1 were presented, then all of sentence 2, etc. (there were five sentences in Spanish and five in Quichua). Recordings were randomized within blocks.


Global accent rating was used in the first two tasks of this study to assess the degree to which prosody (sentence- and phrase-level intonation) contributed to degree of perceived foreign-accentedness in the German speech of native speakers of English. In the first task, original and resynthesized versions of sentence-length utterances were directly compared by judges. Resynthesized versions were specifically altered to more closely match the intonation patterns of native German speakers. Judges simply indicated which version displayed a greater degree of foreign accent. In the second task, judges used the scale to rate degree of foreign-accentedness of the randomized original and resynthesized sentences. Intonation was found in these two experiments to make a significant contribution to degree of perceived foreign accent.

• **Judges:** 57 native speakers of German served as judges in the study

• **Scale:** In task two, 10 a ten-point visual analogue scale was used. Judges moved the bar on the scale to the point they judged appropriate (there were no intermediary marks on the scale).

• **Training:** The test was self-paced and conducted over the worldwide web. Judges were encouraged to use the entire scale; there was potential for different interpretations by different judges as to the relative degrees of foreign accent.

• **Stimuli:** German *sentences* in both original and resynthesized forms spoken by native speakers of English acted as stimuli in the experiment.

Global accent rating was used in the third experiment reported in this paper to assess the degree to which sentence duration (speeded vs. unspeeded conditions) was associated with perceived foreign accent. It was found that for early bilinguals, shorter sentences were perceived to be more foreign accented, whereas for late bilinguals shorter sentences were perceived to be less foreign accented. In all cases, native English speakers’ speech was held to be the least accented, followed by that of early (Italian-English) bilinguals, and, finally, that of late (Italian-English) bilinguals.

- **Judges:** 15 native speakers of English served as judges in the study. All passed a pure tone hearing test.
- **Scale:** A nine-point scale was employed in the study.
- **Training:** No definition of foreign accent was provided; judges were told to use their best guess if unsure and to use the entire scale.
- **Stimuli:** Sentences in unspeeded and (mechanically) speeded conditions spoken by 11 native speakers of English, 24 early Italian-English bilinguals and 25 Italian-English bilinguals were used as stimuli in the experiment.


Global accent rating was used in this study to assess the extent to which specific characteristics of non-native speech (Spanish speakers learning English) affected perceived degree of foreign accent. Two native Spanish speakers produced a number of target phrases (originals); these were then edited to remove / adjust specific non-native characteristics (e.g., removal of epenthetic schwas, etc.). Adjustments to aspects of syllable structure, consonant manner of articulation and stress were found to have the greatest positive results on the perceived degree of foreign accent. Adjustments to voicing (VOT) had little effect. Notably, even where only a single adjustment affecting 35-200ms was made, judges were in many cases able to accurately rate the adjusted version as ‘less foreign accented’ than the original.

- **Judges:** 10 monolingual native speakers of English acted as judges in the experiment.
- **Scale:** A seven-point scale was used with a separate button corresponding to each point. The ends of the scale were labeled ‘closer to native English’ and ‘less close to native English.’
- **Training:** Judges were told that they would hear a series of English phrases produced by one non-native speaker (the data from each of the speakers was treated separately). Judges were instructed to push the relevant button after hearing the entire phrase and were given some initial practice stimuli.
- **Stimuli:** 32 original phrases plus three edited versions of each were used as stimuli in the experiment. These phrases were presented orthographically on a monitor as they were being heard. This was done in order to increase intelligibility; however, there were some indications that ‘spelling pronunciations’ may have influenced the judgments of the raters in some cases (e.g., seeing the full vowel in the word may have encouraged raters to assume that pronunciations without vowel reduction were more accurate).

This study used global accent rating to assess the degree of foreign accent in the English speech of 53 native speakers of Brazilian Portuguese (and 7 control native speakers of English). All judges were clearly able to distinguish the control subjects from the Brazilian speakers of English. Degree of accent among the speakers of Brazilian Portuguese was then compared with the extent of [i] and schwa epenthesis. It was found that epenthesis was significantly correlated with higher degrees of foreign accent and that [i] epenthesis was especially indicative of higher accent (as well as lower proficiency based on amount of time spent studying English and less formal speaking style – text reading vs. word list).

- **Judges:** 10 native speakers of American English acted as judges in the study.
- **Scale:** A *sliding* scale was used with one end being labeled ‘no foreign accent at all’ and the other ‘very heavy foreign accent.’ Judges slid a handle along the scale to select the appropriate level. The computer recorded ratings between 1 and 256, but no numbers were apparent on the scale and the judges were unaware of the numbers being recorded. Only the endpoints and midpoint were marked.
- **Training:** Judges were told they would be hearing Americans and Brazilians and should estimate the degree of foreign accent, if any, using the full range of the scale. A practice session involving 8 speakers was conducted immediately before the actual experiment.
- **Stimuli:** Three *phrases* (3 to 5 seconds long) from each speaker formed the stimuli. All examples of each phrase were judged within a single block, with each phrase being included in two blocks. A pause of 1 second followed each response and a pause of 5 seconds followed each block.


This study compared the degree of nativeness of the English phonemes /ɛ/ and /æ/ as produced by native speakers of Brazilian Portuguese to these same speakers’ global foreign accent. Fifty-three native speakers of Brazilian Portuguese and 7 control native speakers of American English were included in the study. Overall, the accuracy of rating (control subjects over L2 speakers) was much greater for global accent (measured over sentences) than for accent specific to the individual segments (measured over single words). Thus, L1 speakers were rated better on global accent than on individual segments whereas L2 speakers were rated better on individual segments than on global accent. Interestingly, it was found that overall accuracy of /ɛ/ pronunciation decreased as global foreign accent and /æ/ pronunciation improved. It is suggested that this may be the result of an interaction between the original /ɛ/ phoneme and the newly acquired /æ/ phoneme of the learners.

- **Judges:** 10 native speakers of American English rated the global foreign accent of the sentences and a different 10 native speakers of American English rated the degree of foreign accent specific to the individual segments.
- **Scale:** In rating the sentences, a *sliding* scale with the endpoints labeled as ‘no foreign accent at all’ and ‘very heavy foreign accent’ was used. The computer recorded the results as numbers between 1 and 256, though the judges were unaware of the specific numbers at the
time. In rating the degree of foreign accent specific to the individual words, judges used a five-point scale ranging from ‘very heavy foreign accent’ to ‘no foreign accent at all.’

• **Training**: Judges were told that they would hear phrases / words produced by Americans and Brazilians, though they were not told the number of speakers in each group.

• **Stimuli**: In the first task, three short phrases that had been read by each speaker were employed. Each phrase was repeated twice, for a total of 360 tokens. In the second task, recordings of the single words ‘bet’ /bɛt/ and ‘sat’ /sæt/ were employed to rate the degree of foreign accent associated with the pronunciation of the vowels.


This study argues that foreign accent is multidimensional in nature. Especially for untrained judges rating extemporaneous stimuli, a wide range of factors (including those which are not phonological in nature) may influence judgments. Data from 9 native speakers of Puerto Rican Spanish (speaking English) were considered in the study, which primarily focused on variance in judges’ rankings. The results suggested that different judges may focus to a greater or lesser extent on different elements of accent (e.g., one may place greater importance on rhythm while another may place greater importance on grammatical errors), and, further, that, for some raters at least, the factors which are given most importance vary based on the particular speaker being assessed. These variations in focus of importance most clouded the rankings of middle-proficiency speakers (i.e., those who, in comparison to others, were neither particularly strong nor particularly weak). As well, sociocultural factors, particularly familiarity with foreign languages and accents, were found to influence the degree of emphasis placed on different factors by the individual judges (though these factors did not significantly affect the overall accentedness scores given to tokens).

• **Judges**: 114 speakers of English from different sociocultural backgrounds served as judges in the study.

• **Scale**: A ten-point scale for the rating of global accent was employed, as well as 11 similar scales related to individual factors thought to contribute to perceived foreign accent (e.g., rhythm, hesitations, comfortableness, etc.). Other methods of determining individual speakers’ rankings (e.g., direct comparisons between two speakers) were also employed.

• **Training**: Judges were not trained in order to gauge the types of rankings made by naïve native speakers. All raters completed a background questionnaire.

• **Stimuli**: Sentences and series of sentences repeated by the speakers and based on models obtained from natural conversation served as stimuli in the experiment. Speakers were able to alter the specific lexical and grammatical structure of the sentences (i.e., they were not rote repetitions).


This study assessed the German L2 proficiency of 24 graduate students in German, all of whom were first exposed to their second language at age 11 or later and who were assumed to be highly motivated. Four native speakers of German served as control subjects in the study. Overall, NNSs as a group were clearly distinguished from NSs by their scores; only one L2 speaker
scored firmly within the native-speaker range. Three of the four raters were able to consistently distinguish NNSs and NSs; one rater, however, was considerably less consistent and accurate in making such distinctions. Age of first exposure played a significant role in predicting the accentedness scores of NNSs, with a largely linear relationship obtaining between the two variables. It is suggested, however, that, given confounding factors, this age effect is not as marked as has been claimed in previous studies. Professional motivation and having received a combination of suprasegmental and segmental pronunciation training also accounted for significant variation in scores. Length of immersion did not have a significant effect, nor did task type. It is suggested that more than simply phonological factors entered into raters’ judgments, with morphosyntactic and lexical factors also playing an important role.

- **Judges**: Four native speakers of German served as raters in the study, though only two raters assessed each token.
- **Scale**: Judges classified subjects as either native or nonnative speakers and then indicated their degree of certainty using a three-point scale. The results were combined to form a six-point continuum of ratings interpreted as ranging from ‘definitely native’ to ‘definitely nonnative.’
- **Training**: No information about judges’ training (if any) is provided in the paper.
- **Stimuli**: Speakers provided four types of stimuli. These included reading a word list, reading a list of sentences and reading a paragraph, as well as telling a free narrative in response to a written prompt. Certain words, etc. read by the subjects were specifically selected based on the presence of segments thought to be difficult for native speakers of English.


This study used global accent rating to assess the degree to which low-pass filtered sentences and narrative passages retained characteristics of foreign accent. It was found that even when all of the segmental information was removed, the listeners were still able to accurately distinguish English utterances spoken by native speakers of Mandarin from those spoken by the control group of native English speakers. It is suggested that rate of speech, intonation contours and lack of native-like reduction may all have been important factors in informing the judgments.

- **Judges**: Twenty native speakers of Canadian English served as judges in the experiment. Ten of the listeners assessed the sentence stimuli and ten of the listeners assessed the narrative passages.
- **Scale**: A four-point scale was employed, with the individual points labeled: ‘definitely spoken with a foreign accent,’ ‘probably foreign-accented,’ ‘probably spoken by a native speaker of English,’ and ‘definitely spoken by a native speaker of English.’ The ratings of the sentences were recorded by pressing buttons on the computer keyboard, while the ratings of the narrative passages were recorded on an answer sheet.
- **Training**: Before beginning the rating session, listeners were familiarized with Mandarin-accented English by listening to 20-second unfiltered passages from 2 NNSs of English and 1 NS of English. Judges were told that they would hear only the “musical” portion of the speech of native English and native Chinese speakers during the test and should rate the likelihood that the individual passages were spoken by native speakers of English. Judges were encouraged to use the full four-point scale.
• **Stimuli**: Four low-pass filtered read sentences (2 declaratives, 1 y/n question, 1 wh-question) and 20 second-long low-pass filtered passages from ‘free’ narratives served as stimuli in the experiment. Data from 10 native speakers of Mandarin and 10 native speakers of English were included in the test. Each stimulus was presented twice and the order was randomized. In the sentence test, all of the first sentences were presented in a single block, then all of the second sentences, etc.; as well, judges were given a written version of the individual sentences (this was not done during the rating of the narrative passages).


This study compared ratings of global accent among 10 native speakers of Mandarin (advanced English speakers) and 2 control native speakers of English. Ratings of each speaker were obtained for two conditions: extemporaneous (freely-produced) recountings of a story and read versions of transcripts of these same stories. No significant differences in ratings were found between the two types of text (note that familiarity of vocabulary and grammatical structures is controlled here). However, it was found that whichever version judges heard second (either extemporaneous or read) tended to be rated more harshly. The native speakers of Mandarin showed no differences in speaking rate between the two conditions (unlike the native speakers of English who were significantly faster in the reading condition).

• **Judges**: 44 native speakers of Canadian English in undergraduate linguistics or ESL teaching methodology courses served as judges.

• **Scale**: A *nine*-point scale was used, ranging from ‘unaccented’ to ‘heavily accented.’

• **Training**: Judges completed 3 warm-up samples (not included among the experimental stimuli) immediately before beginning the rating session. Judges were not told that they would be hearing extemporaneous and read versions of the same text.

• **Stimuli**: Clips of approximately 20 seconds were taken from the beginning of each speaker’s extemporaneous storytelling and reading of their own transcript. Hesitations, etc. were retained in the transcripts and speakers were told to read the text exactly as it was written.


In this study global accent ratings were compared with the accuracy of listeners’ transcriptions and with their judgments of the relative comprehensibility of speech. Sentences produced by ten advanced ESL students who were native speakers of Mandarin and two native speakers of English were elicited and judged. It was found that while intelligibility, comprehensibility and accent were correlated, generally these correlations were only moderate. Some highly intelligible utterances were still identified as relatively highly accented. It was further suggested that segmental factors played a greater role in determining accent than in determining intelligibility.

• **Judges**: 18 native speakers of English acted as judges in the study. All had a basic knowledge of articulatory phonetics.
Scale: A nine-point scale was used. In rating comprehensibility, it ranged from ‘extremely easy to understand’ to ‘impossible to understand.’ In rating global foreign accent, it ranged from ‘no foreign accent’ to ‘very strong foreign accent.’

Training: Two practice stimuli for comprehensibility rating and transcription were included before the experimental stimuli. The same two stimuli acted as practice before the global accent rating session.

Stimuli: Clauses or phrases were excerpted from a story told by the speakers on the basis of a series of cartoons. Three excerpts from the first 30 seconds of the story were selected for each speaker; these could not be standardized for length.


This study measured the response latencies of listeners in judging the truth values of simple English propositions uttered by 10 native speakers of English and 10 native speakers of Mandarin. These were compared to the ratings of accentedness and comprehensibility assigned to the same sentences by the judges. Overall, the utterances of native speakers of English were accurately verified significantly more often and were judged significantly more quickly. In addition, they were deemed significantly less accented and significantly more comprehensible. There was a significant correlation between accentedness and comprehensibility, but the size of the correlation varied considerably between listeners; it is suggested that some listeners placed much more emphasis on comprehensibility when rating global accent than did others. Comprehensibility but not accentedness scores had a significant impact on response times. Judges’ exposure to nonnative speakers did not have a significant impact on any variable.

Judges: 20 native speakers of English acted as judges in the study.

Scale: Nine-point scales were used in accessing both accentedness and comprehensibility. The endpoints of the scales were labeled ‘no foreign accent at all’ / ‘very strong foreign accent’ and ‘not difficult to understand at all’ / ‘very difficult to understand,’ respectively.

Training: Three passes through the data occurred. On the first pass, judges simply indicated whether a proposition was true or false and transcribed (in standard orthography) the sentences. On the second and third passes, accentedness and comprehensibility were judged; raters were therefore reminded that they had already heard the full data set and were instructed to reserve the top and bottom of the scales for the most extreme examples.

Stimuli: 40 sentences, half of which were true and half of which were false, acted as stimuli in the experiment. All speakers produced all of the statements, but each rater heard each statement only once from one speaker. Furthermore, only 2 sentences from a single speaker occurred in any given data set and no two data sets were identical.


This study assessed the effect of speaking rate on the relative accent and comprehensibility ratings of native and non-native English speech. 10 native speakers of Mandarin (advanced ESL students) and 10 native speakers of Canadian English were included in the experiment. Overall,
NNSs naturally produced slower readings of the text than did NSs and slowed their readings less, when asked to do so, than did NSs. The readings of NSs were rated the same for accent in the normal and slowed conditions. The readings of NNSs were deemed more accented in the slowed condition. Comprehensibility also suffered in the slowed condition. In the second experiment, the normal-speed sentences were adjusted using speech compression software, resulting in the speeding and slowing of tokens to different mean rates. In this case, at least some NNSs received better ratings when their speech was speeded. While the optimal rate for NNSs was slower than for NSs, further slowing did not help and could even hurt comprehensibility and accentedness.

• **Judges:** 20 native speakers of English served as judges in the study (a different group was used in each of the two experiments).

• **Scale:** A nine-point scale was employed. In the first experiment sentences were rated for both accent (‘no accent’ to ‘very strong accent’) and comprehensibility (‘very easy to understand’ to ‘very difficult to understand’). In the second experiment rate of speech was directly rated, with the extremes of the scale being labeled ‘too slow’ and ‘too fast.’

• **Training:** In the first experiment, judges rated three practice tokens not included in the actual experiment before commencing the experiment. In the second experiment, they rated 12 practice tokens taken from tokens actually used in the experiment and selected to represent the full range of speaking rates. In this case, they were told that some of the tokens had been manipulated by the computer and in both cases they were encouraged to use the full scale.

• **Stimuli:** The first sentence from a read text was used as the stimulus for the experiment. In the first experiment the speakers’ slow and normal versions of the sentence were used. In the second experiment, the normal version was manipulated to match the English mean, the Mandarin speakers’ mean and a further slowed version. The unmanipulated versions were also included in this experiment.


This paper is a reprise of Munro & Derwing (1995a).


Global accent and comprehensibility ratings were compared with the speech rate of non-native speakers of English in this paper. In the first study reported, the productions of 48 adult intermediate ESL students from a range of L1 backgrounds and 4 native speakers of Canadian English were recorded and rated for accent and comprehensibility. It was found that speech rate (based on syllables per second in the read sentences) accounted for a small but significant amount of variation in global accent and comprehensibility scores. In the second study, sentences read by 10 native speakers of Mandarin (advanced ESL speakers) and 7 native speakers of Canadian English were digitally compressed 10% and expanded 10%. These, along with the originals, were rated and again it was found that speech rate accounted for a small but
significant amount of variation in global accent and comprehensibility. It is suggested that usually ESL students speak at a rate somewhat slower than optimal, but that too much increase in rate would also be detrimental.

- **Judges:** 48 monolingual native speakers of Canadian English acted as judges in the first experiment. In the second experiment, 27 native speakers of Canadian English (a few of whom had fluency in a second language) acted as judges.

- **Scale:** In both experiments *nine*-point scales were used. The extremes of the accent scale were labeled ‘no accent’ and ‘very strong accent,’ while the extremes of the comprehensibility scale were labeled ‘very easy to understand’ and ‘impossible to understand.’

- **Training:** In the first experiment judges were told they would hear native speakers of several languages including English. Three warm-up items were completed at the beginning of the experiment and the raters were told to use the full scale. In the second experiment judges were not told that they would hear the same stimuli at different rates of speech.

- **Stimuli:** Read sentences (4 from each subject) were used in both experiments.


This study used global accent ratings to gain insight into the degree to which adult migrants between dialect areas can acquire aspects of a second dialect. Speakers from three groups of ten were compared: Canadians in Canada, Alabamans in Alabama, and Canadians who immigrated to Alabama as adults. It was found that, indeed, Canadian immigrants to Alabama had accents intermediary between those of Canadians in Canada and Alabamans in Alabama (though there was considerable variance between the individual members of the Canadians in Alabama group). Phoneticians who later analyzed only the /aj/ diphthongs in the same data also found similar results.

- **Judges:** 22 native speakers of Canadian English and 27 native speakers of Alabama English acted as judges in the experiments.

- **Scale:** A *nine*-point scale was used. In the first experiment this ranged from ‘very Canadian’ to ‘very American’ and in the second experiment this ranged from ‘definitely from Alabama’ to ‘definitely not from Alabama.’ Results using the two versions of the scale were comparable. Results were recorded on paper by the judges.

- **Training:** Judges were told that they would hear speech from three groups of speakers, which were identified, though no specific information about the individual speakers was provided during the experiment itself. Judges did 4 practice judgments before the experiment began.

- **Stimuli:** 10-second clips from extemporaneous speech by members of the three groups were used as stimuli in the experiment. All clips came from the subjects’ retelling of a story based on a series of pictures and were taken from the beginning of the story to control for content.

This study measured degree of foreign accent in German words and phrases produced by 60 native speakers of English before and after a 3-week series of oral drills (no actual meaning-based language instruction was provided). Speakers from three age groups – elementary (age 9.5-10.5), junior high (age 14-15) and college (18-26) – were included in the study. No significant effect of age group or gender was found in the pretest. After training, there was again no significant effect of gender but the accent of both junior high and college groups was deemed significantly better than that of the elementary group.

- **Judges**: One native speaker of German and one native speaker of English who was doing graduate studies in German acted as judges in the study.
- **Scale**: A five-point scale was used in assessing accent. Points of the scale were labeled: ‘native-like pronunciation,’ ‘good,’ ‘average,’ ‘poor but an attempt,’ ‘no attempt.’
- **Training**: Judges were not made aware of the purpose of the experiment. No information was given about any other training provided.
- **Stimuli**: Short words and phrases repeated after a native speaker model were collected in both the pretest and posttest. These words and phrases served as the stimuli in the experiment.


This study examined the perceived foreign accent of 60 Italian-born male immigrants to the US, all of whom had lived in the United States for at least five years, and 10 control native speakers of English. Two types of speech were assessed: a read paragraph and an oral recounting of a personal narrative (spontaneous speech). An extremely strong age of arrival effect was noted, particularly in the read paragraph task; virtually no length of residence effect was found. The overall age of arrival effect was quite linear, with the youngest arrivals performing within the range of the native speaker controls. Generally, the stories were deemed to be more accented than were the read paragraphs, though there was a strong correlation between scores on the two tasks.

- **Judges**: Two American-born linguistics graduate students acted as judges in the study.
- **Scale**: A five-point accentedness scale, ranging from ‘no foreign accent’ to ‘heavy foreign accent,’ was employed.
- **Training**: Six practice samples were scored before the experimental tokens were presented. Both judges were trained in linguistics.
- **Stimuli**: Two types of stimuli were used in the study: a short, read paragraph, and a free personal narrative, the last 45 seconds of which were included in the study. The two types of stimuli were rated by the judges in separate sessions.

This paper reconsiders the accentedness findings of an earlier study by the same author, within the framework of a discussion of the Critical Period Hypothesis. Data from 67 second language learners of English were included in the study (no control subjects are described). In this case, it was found that the distribution of accentedness scores among those speakers who began learning English after the age of 15 (n=34) followed a normal bell-shaped curve. The scores of those who began to acquire English between the ages of five and fifteen (n=33), on the other hand, were strongly skewed toward the top (native-like) end of the scale. These results are described as similar to those obtained by the speakers on syntactic measures in the earlier study, and are held to provide support for the Critical Period Hypothesis.

- **Judges**: The judges are not described in this paper.
- **Scale**: A five-point scale, with intermediate ‘+’ values, was employed in the study.
- **Training**: Training of the judges is not described in the paper.
- **Stimuli**: The nature of the stimuli employed is not described in the paper.


This study assessed degree of global foreign accent in 29 advanced ESL students from a variety of L1 backgrounds (and 3 control native speakers of English). Only one factor – age of arrival in Canada - was found to be significantly correlated with degree of foreign accent. Specifically, students who arrived before the age of 6 were found to have significantly less accented speech than were those who arrived between the ages of 12 and 17. Length of time resident in Canada and extemporaneous vs. read speech were not significant factors in relation to the degree of perceived foreign accent.

- **Judges**: 13 native speakers of English training to be ESL teachers served as judges in the experiment.
- **Scale**: A five-point scale ranging from ‘no foreign accent’ to ‘very heavy foreign accent’ was employed.
- **Training**: Judges did not receive specific training, though all were enrolled in an advanced applied linguistics course at the time.
- **Stimuli**: 20-second segments from two different tasks served as stimuli. In the first case, a clip from a read text was used and in the second case a clip from an extemporaneous personal narrative was used. No significant differences were found between the ratings accorded to the two types of stimuli.


This study measured degree of global foreign accent in English among 72 native speakers of Italian and compared these ratings to age of learning (early vs. late bilinguals) and amount of L1 use (high- vs. low-use). Data from 18 native speakers of English were also included in the study. It was found that both age of learning and amount of L1 use were significant independent predictors of degree of foreign accent, with late bilinguals and those who continued to use Italian
to a greater extent producing more accented English. Gender and self-reported L1 proficiency were not significant independent predictors. All groups of bilinguals were deemed to produce more accented English than was the English native speaker control group.

- **Judges:** Nine native speakers of Canadian English acted as judges in the study.
- **Scale:** A nine-point scale of accentedness was employed, ranging from ‘very strong foreign accent’ to ‘no foreign accent.’
- **Training:** Judges were told that they would hear sentences produced by native speakers of Canadian English and Italian immigrants to Canada. In addition, judges were able to correct their ratings one at a time.
- **Stimuli:** Three sentences elicited using a delayed repetition technique with written support were used as stimuli in the experiment. Versions of each sentence were presented to judges in separate blocks and each individual token was rated three times by each judge.


This paper reanalyzes the data presented in Suter (1976), considering the correlations between individual variables. Under this analysis, only five of the original factors emerged as significant determiners of global foreign accent: L1, aptitude for oral mimicry, years spent in an English-speaking country, months spent residing with a native speaker of English (years in an English-speaking country and months riding with a NS of English were correlated with one another), and concern for pronunciation. No other factor emerged as significant. It is worth noting that L1 was not included in any of the ‘factors’ and was instead entered into the regression independently (despite correlation with a number of other variables).

- **Judges:** See Suter (1976)
- **Scale:** See Suter (1976)
- **Training:** See Suter (1976)
- **Stimuli:** See Suter (1976)


This study assessed the global accent of 11 Japanese learners of English at two points, 42 months apart (5 native speakers of English were also included). All native speakers of Japanese either improved their accents or remained constant between the two tests. The greatest factor determining whether or not a speaker’s accent score would improve was the amount of time spent in English-speaking countries, as, for example, on a student exchange. Overall, only three speakers showed improvement in global foreign accent ratings between time one and time two. Global accent scores were also compared to accuracy of liquid pronunciation in initial singletons and clusters. For some speakers, global accent and liquid identifiability followed similar paths, though this was not true for all speakers.

- **Judges:** Five native speakers of English acted as global accent raters in the first portion of the experiment.
- **Scale:** A nine-point scale, ranging from ‘strong foreign accent’ to ‘no foreign accent,’ was used.
• **Training**: No specific training was provided to judges, though only the final three sets of ratings (out of 4) were included in the statistical analyses. Judges were urged to use the full scale in making their ratings.

• **Stimuli**: Five read sentences, collected from each speaker at two points 42 months apart, were used as stimuli in the experiment. A read word list was also collected from each speaker and used in assessing the accuracy of the learners’ liquid pronunciation.


This study compared voice onset time with degree of global foreign accent measured at two points in time 42 months apart among 11 native speakers of Japanese (see Riney & Flege 1998). Japanese was assumed to have an intermediate lag VOT, while English was held to have a long lag VOT. Few speakers showed improvement in VOT (i.e., movement toward English native speaker norms) over the 42 months. However, the study did provide general support for a correlation between global foreign accent and voice onset time. Only one Japanese learner of English showed improvement in VOT between T1 and T2; this was one of the same few speakers who were found to have improved in global foreign accent between the two data collection sessions. It is suggested that difficulty in forming new phonetic categories corresponding to the English long-lag voiceless stops can be accounted for based on the Speech Learning Model, with the similarity between intermediate-lag and long-lag stops prompting L2 learners to classify English /p/, /t/ and /k/ as equivalent to Japanese /p/, /t/ and /k/.

• **Judges**: Five native speakers of English acted as global accent raters in the experiment (see Riney & Flege 1998).

• **Scale**: A nine-point scale, ranging from ‘strong foreign accent’ to ‘no foreign accent,’ was used (see Riney & Flege 1998).

• **Training**: No specific training was provided to judges, though only the final three sets of ratings (out of 4) were included in the statistical analyses. Judges were urged to use the full scale in making their ratings (see Riney & Flege 1998).

• **Stimuli**: Five read sentences, collected from each speaker at two points 42 months apart, were used as stimuli for global accent rating. Measurement of stop VOT was conducted by the researchers based on six words from the word list read by speakers at time one and time two.


This paper is a reprise of portions of the data reported in Riney & Flege (1998). The relationship between global accent and the percentage substitution of the Japanese flap for the English liquids was assessed. The subjects were 11 native speakers of Japanese learning English and 3 control native speakers of American English. The substitution of Japanese flaps was confirmed / identified by native speakers of Japanese trained in linguistics. It was found that, indeed, a higher percentage of flap substitution was correlated with a greater degree of foreign accent; this was more important and apparent at time 1 (when the subjects were in first year of college) than at time 2 (when the subjects were in fourth year of college). Over the four years, improvement in overall degree of foreign accent was noted in only three of the subjects. It was found that flaps
were more frequently substituted for English /l/ than for English /r/, in onset singletons than in onset clusters, and in spontaneous speech than in read prompts.

- **Judges**: See Riney & Flege (1998)
- **Scale**: See Riney & Flege (1998)
- **Training**: See Riney & Flege (1998)
- **Stimuli**: See Riney & Flege (1998)


Two experiments examining the Dutch pronunciation of native speakers of English who began learning Dutch at different ages were conducted. In the first experiment 136 learners of Dutch repeated words selected for difficulty of pronunciation following recorded native speakers models. The youngest groups of speakers (with ages of arrival of 5 or 6) achieved the worst pronunciation scores of any of the groups while the oldest groups of speakers (with ages of arrival in their late teens or later) achieved the best pronunciation scores. In the second experiment, 47 learners of Dutch were recorded three times over the course of a year in direct repetition and ‘spontaneous production’ tasks. Significant age effects were found only at the time of the first recording where greater age was associated with better accentedness scores. By the time of the third recording, the younger subjects had surpassed the older speakers in their pronunciation of some, but not all, sounds.

- **Judges**: One judge scored each token in both of the studies.
- **Scale**: A five-point scale was employed in both experiments, ranging from ‘uninterpretable as target sound’ to ‘indistinguishable from a native speaker’s pronunciation’ (intervening points were also labeled).
- **Training**: Three judges were employed in the first experiment (to rate different sections of the data) and they worked together on some tokens to establish ratings and to ensure that they were all interpreting the criteria in the same way. In the second experiment the judge scored each item twice.
- **Stimuli**: In the first experiment each speaker repeated Dutch words selected for their difficulty 20 times each. In the second experiment direct repetition was also employed; in addition, subjects named items from pictures. Ratings of words focused specifically on the segments thought to be difficult for native speakers of English learning Dutch.


This study compared the adequacy of two methods in rating the degree of foreign accent evident in the English speech of 90 native speakers of Italian and 6 native speakers of English. Interval scaling (a 7-point Likert scale) and direct magnitude estimation were used. Results in the two cases were comparable and were taken to indicate that interval scaling can be effectively employed in quantifying foreign accentedness. Some response biases were noted – these were taken to be the result of differing internal standards of accentedness among listeners. As well, some ceiling effects with the 7-point scale were noted, suggesting that a 9- or 11-point scale might be more effective in capturing the full range of variation.
• **Judges:** 20 monolingual native speakers of English acted as judges in the study.

• **Scale:** A *seven*-point scale ranging from ‘least accented’ to ‘most accented’ was used. In addition, direct magnitude estimation was employed, with raters assigning scores to stimuli based on their relationship to a model token. Raters in this condition chose any number to be assigned to the model token (see also Brannen, Ryan & Dawson 1975).

• **Training:** Raters were given a short training session, including practice ratings, before the experiment commenced. Interval scaling and direct magnitude estimation were employed in separate sessions.

• **Stimuli:** Two separate *sentences* elicited from 90 native speakers of Italian and 6 native speakers of English acted as stimuli for the study. Only half of the tokens were assessed by each rater.


This study measured the degree of perceived foreign accent in the English speech of 61 native speakers of Arabic, Japanese, Persian and Thai. Two control native speakers of English were also included; their scores were consistently distinguished from those of the NNSs. The most significant predictor of foreign accent was L1, with NSs of Persian and Arabic receiving significantly better scores than did NSs of Japanese and Thai. Also significant were: concern for pronunciation, amount of conversation with native speakers of English, age of learning, length of residence in English-speaking countries, mimicry ability, intensive English training, formal classroom instruction (negative correlation), integrative orientation (negative correlation), age of arrival in an English-speaking country, percentage of conversation at home conducted in English, and residence with native speakers of English. No significant effects were found for formal pronunciation training, percentage of native English-speaking teachers, gender or extroversion (among others). No correlations between individual variables were examined (but see Purcell & Suter 1980).

• **Judges:** 14 native speakers of English, all of whom had received an ‘A’ grade in introductory linguistics and none of whom had had contact with speakers of the four L1s considered here, served as judges.

• **Scale:** A *six*-point scale ranging from ‘best’ to ‘worst’ was used in the study.

• **Training:** Two days before the experiment began, judges participated in a two-hour training session. In addition, before beginning to rate the samples on the day of the experiment, judges first heard a ‘sampler’ tape which included the first 30 seconds of each speaker’s narrative to provide a sense of the range. Judges were told to focus on pronunciation rather than on grammatical or other factors (it was thought that having completed an introductory linguistics course would facilitate this).

• **Stimuli:** Speakers each gave a free oral description of a holiday or celebration in their native country. The first two minutes of each recording served as the experimental stimuli.

This study measured degree of foreign accent among L2 learners of English who began learning the language at different ages upon immigration to the UK. A total of 109 subjects were included in the study, with 10 per age of learning from 6 through 14 and another 19 having an age of learning of 15 or greater. Speakers came from a wide range of L1 backgrounds and all had been living in the UK for at least two years at the time of the study. Age of learning proved to be the biggest factor in accounting for variance in degree of perceived foreign accent; those who began learning English at or before age six were ‘invariably accent-free’ while the speech of those with an AOL of 13 or greater was invariably accented. Use of English at home also accounted for important variance, being correlated with lesser degrees of accent. Female gender was also associated with less accented speech.

- **Judges:** Three native speakers of English acted as judges in the study.
- **Scale:** A three-point scale was used in measuring global foreign accent. The points were labeled ‘no foreign accent,’ ‘detectable but slight accent’ and ‘marked accent.’
- **Training:** Training of the judges (if any) is not described.
- **Stimuli:** Each speaker read the same paragraph of English prose that served as the stimuli in the experiment.


This study compared degree of perceived foreign accent in the speech of 36 Russian immigrants to the US as rated by experienced and inexperienced raters with a range of independent variables. Ten native speakers of American English were included as control subjects in the experiment. Age of arrival in the US was the most significant factor, accounting for 66% of the variance in accentedness scores; gender (with females receiving better scores), mimicry ability and global speaking proficiency (independently assessed) also accounted for significant, though lesser, variance. Attitudinal and motivational variables did not account for any variance. Only two subjects, both of whom came to the US at age 4, received any perfect accentedness scores. Generally, sentence reading was deemed to be more accented than was spontaneous speech. Inexperienced raters generally perceived greater accentedness than did experienced raters, and reliability coefficients were higher for experienced than for inexperienced raters.

- **Judges:** Two groups of 8 judges, all of whom were native speakers of English, were included in the study. The inexperienced judges had little knowledge of foreign languages or contact with non-native speakers and had not taken linguistics courses. The experienced judges spoke at least one foreign language fluently, had taken linguistic courses and had frequent contact with non-native speakers of English.
- **Scale:** A five-point scale, ranging from ‘no foreign accent’ to ‘heavy foreign accent,’ was employed.
- **Training:** Judges were told that the speech samples were produced by Russians and Americans. As well, they rated three practice samples before beginning the experiment.
- **Stimuli:** Speakers completed three tasks: reading a list of 20 sentences seeded with difficult sounds, reading a 160-word paragraph, and recounting a personal narrative.
This study compared age of arrival in the United States of 240 Korean-English bilinguals with their global accent in L1 and L2 as rated by naïve native speakers of the two languages. It was found that age of arrival was significantly correlated with proficiency in both languages. Specifically, the English pronunciation of the bilinguals with AOA of 1-5 was better than that of those with AOA 6-9, which was better than that of those with AOA 10-13, which, in turn, was better than that of those with AOA 14-23. All of the bilingual groups had English accentedness scores that differed significantly from those of monolingual English speakers. At the same time, Korean pronunciation was worse for those who had earlier AOA in the United States. Only those who arrived in the USA after the age of 12 had Korean accent scores comparable to those of Korean monolinguals. For most speakers, L1 and L2 pronunciation scores were inversely related. In addition to age of arrival, Korean and English use were found to be significant independent factors in determining proficiency in the two languages. Finally, speakers’ self-rated proficiency in both L1 and L2 was strongly correlated with the judges’ global accent ratings.

• **Judges:** 10 monolingual English speakers judged the pronunciation of the English sentences, and 10 near-monolingual Korean speakers judged the pronunciation of the Korean sentences.

• **Scale:** A nine-point scale was employed in each case. The endpoints of the English scale were labeled ‘no accent’ and ‘very strong accent,’ while the endpoints of the Korean scale were labeled ‘very good pronunciation’ and ‘very poor pronunciation’ (the differences in scale labeling were due to translation factors).

• **Training:** Raters were explicitly instructed not to base their judgments on regional dialects. As well, all were trained in practice trials and told to use the full scale in making their judgments.

• **Stimuli:** Five sentences in each language were elicited using a delayed repetition technique with written support. All sentences were controlled for length, and each stimulus was presented three times to the raters. Each rater judged all five sentences in their native language, with one sentence being rated each day over the course of five days.

6. References


Bongaerts, Theo, Chantal van Summeren, Brigitte Planken and Erik Schils. 1997. Age and


Major, Roy C. 1987. Phonological Similarity, Markedness, and Rate of L2 Acquisition. *Studies