Which Features of Spanish Learners’ Pronunciation Most Impact Listener Evaluations?

Kara McBride

Hispania, Volume 98, Number 1, March 2015, pp. 14-30 (Article)

Published by Johns Hopkins University Press
DOI: 10.1353/hpn.2015.0001

For additional information about this article
http://muse.jhu.edu/journals/hpn/summary/v098/98.1.mcbride.html
Which Features of Spanish Learners’ Pronunciation Most Impact Listener Evaluations?

Kara McBride
Saint Louis University

Abstract: This study explores which features of Spanish as a foreign language (SFL) pronunciation most impact raters’ evaluations. Native Spanish speakers (NSSs) from regions with different pronunciation norms were polled: 147 evaluators from northern Mexico and 99 evaluators from central Argentina. These evaluations were contrasted with ratings from non-native SFL instructors. Evaluators rated speech samples in terms of comprehensibility and pleasantness, and identified the specific features of the learners’ pronunciation that they found either acceptable or unsatisfactory. For all groups, negative personal attributes were associated with less accurate pronunciation. Significant quantitative differences were found in groups’ ratings, with the non-native raters falling between the two groups of NSSs. The different groups of NSS raters were largely in agreement in terms of the specific pronunciation features that they identified as important, however. Compared to non-native raters, the NSSs focused more on suprasegmentals and global voice qualities. Phonemes that appear to have the greatest impact on listeners’ evaluations are identified, and teaching implications are discussed.

Keywords: affective response/reacción afectiva, comprehensibility/inteligibilidad, nativeness/nativo versus no nativo, pleasantness/agrado, pronunciation/pronunciación

Introduction

Clear pronunciation is required for listeners to understand the words that a second language (L2), speaker produces. However, it is not comprehensibility alone that contributes to how speech is perceived. In both first language (L1) and L2 speech, pronunciation can strongly impact the impressions that listeners form of the speaker (Bresnahan et al. 2002; Derwing and Munro 2010; Lippi-Green 1997; Llurda 2000). These impressions can affect the quality of interactions between speakers and listeners (Cunningham-Andersson 1997; Golombek and Jordon 2005; Lippi-Green 1997).

Choices about which pronunciation features one displays are decisions about identity expression (Golombek and Jordon 2005). In order to make well-informed choices in this arena, L2 speakers need information about how different aspects of their speech may be perceived by speakers of the L2. This kind of information can also help teachers make informed decisions about pronunciation instruction.

Features of SFL Speech

Table 1 summarizes the features that Bergen (1974) proposed as the most essential pronunciation features for Spanish as a foreign language (SFL) learners to master for a clear and nonirritating accent—what he calls a “learner’s dialect” (480). Arteaga (2000) explicitly borrowed Bergen’s concept of a learner’s dialect but advocated for a slightly different set of features (343) (see Table 1). Neither of these authors, however, offered empirical evidence for their claims that
these were the most important features of SFL pronunciation. Research that has investigated the question of which aspects of SFL pronunciation most merit instruction time has approached the issue largely from the angle of learnability (Elliott 1997; Lord 2005). In choosing instructional priorities, it is also important to establish which aspects of L2 pronunciation have the greatest impact on successful communication (Jenkins 2002). Schairer’s (1992) study related measurements of specific pronunciation features to native Spanish speakers’ global evaluations, but again an explanation for the basis for selection of the particular features examined was not offered.

Table 1. SFL pronunciation features highlighted in publications

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Segmentals</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vowels</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Trilled /r/</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Stops /b, d, g/ (fricative where required)</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Tap /ɾ/</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Not aspirating /p, t, k/</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Lateral /l/ (not velarized [l])</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Nasal assimilation</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Diphthongs</td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pronunciation of the letter ‘v’ ([b] or [β] and not [v])</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Voicing of /s/</td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dental or interdental /d/</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pronunciation of the letter ‘ñ’</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pronunciation of the letter ‘g’</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A more fricative /x/</td>
<td></td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not pronouncing ‘h’ (except in ‘ch’)</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Suprasegmentals</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linking</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Rhythm</td>
<td></td>
<td></td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>Stress</td>
<td></td>
<td></td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>Intonation</td>
<td></td>
<td></td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>Rate of speech</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Foreignness**

A layman’s description of poor L2 pronunciation would be to say that the speaker has a strong foreign accent, but this way of characterizing L2 pronunciation is problematic. To classify an accent as not native-like requires comparing the learner’s pronunciation to a specific pronunciation norm (Hayes-Harb and Watzinger-Tharp 2012). There are, however, multiple pronunciation norms in Spanish (Schwegler and Kempff 2007). The acceptable learner’s dialect that
Bergen (1974) and Arteaga (2000) sought was neither tied to one particular Spanish norm nor, for that matter, meant to sound exactly native-like.

Another problem with using foreign accentedness as a means of evaluating L2 speech is that it implies that the L2 speaker’s goal can and should be to produce the L2 in a way that is indistinguishable from the oral production of a native speaker. Ample research suggests that most adult L2 learners cannot reach such a goal in the area of pronunciation (Flege, Munro, and MacKay 1995; Levis 2005; Moyer 1999, 2004). Furthermore, because accent is intimately tied to identity and the presentation of self (Golombek and Jordon 2005), many L2 learners may not want to attain a native-like accent.

In Hayes-Harb and Watzinger-Tharp’s (2012) study, there was no significant correlation between foreign accentedness and comprehensibility. Foreign accentedness, then, may be at times a counterproductive means by which to measure L2 speakers’ pronunciation. This is not to say, however, that it is not possible to formulate meaningful evaluations of L2 pronunciation in ways that can usefully inform L2 instruction. What is important in communication is first whether listeners can understand the spoken message, and second, what kind of effect the speaker’s pronunciation has on listeners’ perception of the speaker.

Comprehensibility

Not all L1 phonetic transfer causes difficulties in L2 comprehensibility. The severity of the impact that a non-native-like pronunciation feature has on listeners’ comprehension varies depending on the phonemes in question (Cunningham-Andersson 1997; Jenkins 2002; Munro and Derwing 2006). Furthermore, the key to achieving highly comprehensible L2 speech may not be exclusively or even primarily in specific phonemes, but may be highly associated with more global features such as stress, rhythm, and intonation (Derwing, Munro, and Wiebe 1998).

The studies thus far cited in this discussion of comprehensibility were done with learners of English as a second language (ESL). One study has investigated the question of which SFL pronunciation features correlate the highest with differences in listener evaluations. Schairer (1992) had native speakers of Spanish rate speech samples of SFL learners. She also analyzed each speech sample in terms of the nine features indicated in Table 1. Using regression analysis, Schairer (1992) concluded that vowels accounted for 90% of the variance in the native speakers’ evaluations (315). It is possible, however, that other SFL pronunciation features that were not included in Schairer’s analysis were also significant in shaping native speakers’ evaluations. If so, their absence in Schairer’s study may have led to an overrepresentation of the impact of vowels. Because of this possibility, the present study included a number of open-ended questions.

The quantitative measure of comprehensibility in the present study used Derwing and Munro’s (2009) definition of the term: “The listener’s perception of how easy or difficult it is to understand a given speech sample” (478). Using evaluators’ impressions to measure this variable is in keeping with how the comprehensibility of L2 speakers’ speech is judged in most situations. As Trofimovich and Isaacs (2012) point out, interlocutors in everyday communicative encounters and L2 teachers and examiners respond to the L2 speech that they hear impressionistically, and not, generally, by using objective measures of the kind that machine analysis could render. Thus, eliciting evaluators’ perceptions corresponds to what is meant by this term in teaching and communicative contexts.

Affective Responses

Investigations into listeners’ attitudes toward accents generally ask evaluators to indicate their responses using semantic differential scales describing the characteristics being investigated, such as the evaluators’ sense of solidarity with or the prestige of the speaker (Garrett, Coupland, and Williams 2003). The design of such studies must try to isolate the impact of pronunciation
from other factors. By having subjects read prepared texts, researchers attempt to eliminate
differences in grammar and word choice from one speech sample to another (Garrett, Coupland,
and Williams 2003). The matched-guise technique, in which the same speaker speaks with
different accents and is presented to evaluators as different people, is frequently employed to
minimize issues of voice quality. It is not always possible for the same speaker to accurately
represent different accents, however. The alternative is to use a large number of speech samples.

Cunningham-Andersson’s (1997) study used the matched-guise paradigm to study atti-
tudes about ESL pronunciation. She found that different sounds had different levels of impact
on how friendly and intelligent raters considered each speaker to be. The same speakers were
judged differently, even by the same rater, depending on which type of pronunciation error was
heard. It is, therefore, important to look not only at overall impressions but specific features of
L2 pronunciation.

In the present study, a seven-point Likert scale from ‘irritating’ to ‘pleasant’ was included.
This was used to gauge listeners’ overall sense of the acceptability of the SFL learners’ pronuncia-
tion and calculate its relation to other scores. In order to further explore the range of affective
responses that the evaluators had to SFL speech and the bases upon which they were founded,
an open-ended item, in which the raters were asked to describe the speaker, was included.

Allowing participants to choose their own descriptors avoids some problems associated with
closed-ended items presented to diverse populations (Al-Hindawe 1996). One consideration
in the present study was the possibility of nonequivalence between translations: a given word
may be the best available translation but have different connotations that could lead subjects
to interpret the item in a way not meant by the researcher. Also possible was that some words
could have different connotations for the two main groups of native Spanish speakers (NSSs) that
participated in this study. In such a case, a translated item might be understood as intended by
one NSS group but not by the other. By first seeing what the evaluators wrote and then consulting
with native speakers from the evaluators’ region, it was hoped that a more accurate interpretation
of the data could be made. Another danger in relying on closed-ended items in attitude studies
is that participants may vary the associations that they make with a given adjective, depending
on whether they apply it to one group or another (e.g., males or females) (Al-Hindawe 1996).

Affective responses to pronunciation can be significantly affected by listeners’ attitudes
toward the L2 speakers’ ethnic group (Gatbonton, Trofimovich, and Magid 2005; Llurda 2000)
and whether the listeners consider the speakers to be their status peers (Cargile and Giles 1998).
For these reasons and others that will be explained below, it is important to consider differences
between groups of evaluators.

Types of Evaluators

The more familiar a rater is with a particular L2 accent, the more comprehensible he or she
will find the speech of speakers with that same L1 (Bradlow and Bent 2008). Familiarity can affect
other aspects of a listener’s reception of an accent as well. DeShields and Kara (2011) found that
NSS evaluators from Mexico were more likely to find speech samples convincing if they were
familiar with the L2 accent. Evaluators who lived in a region where residents had frequent contact
with native English speakers rated native Spanish speakers and (US) SFL speakers as equally
convincing in Spanish, whereas evaluators who lived in a region of Mexico where contact with
native English speakers was rare found NSSs significantly more convincing than SFL speakers.

In Cunningham-Andersson’s (1997) study, raters with previous second language teaching
experience were generally the most lenient judges of mispronunciation severity. On the other
hand, Flege and Fletcher (1992) found that trained raters can become more critical of foreign
accents on materials that they are more familiar with. Although trained raters can detect more
in a speech sample than untrained raters, and this can obviously benefit research, Munro (2008)
argued in favor of using untrained raters as well, precisely because training alters perception.
For these reasons, the present study used both linguistically trained English speakers and untrained NSS evaluators.

Another characteristic of evaluators that may influence their perception is which variety of the target language they speak (Hayes-Harb and Watzinger-Tharp 2012). Arteaga’s (2000) attempt to establish an SFL learner’s dialect was intended to be “unmarked,” but the wide variety of pronunciation norms in the Spanish-speaking world throws into question the possibility of such neutrality (343). It could be that NSSs with very different pronunciation norms would respond differently to the same L2 speech. The present study investigated this possibility by including evaluators from northern Mexico and central Argentina.

Finally, there is evidence to suggest that native speakers and non-native speakers judge L2 learners’ pronunciation differently (Fayer and Krasinski 1987; Kang 2012; Rossiter 2009). Generally speaking, non-native speakers focus more on form errors, whereas native speakers tend to be more influenced by issues of comprehensibility. Thus, while this study was primarily designed to investigate native Spanish speakers’ views of SFL pronunciation, evaluations were also collected from non-native SFL teachers for the purpose of comparison.

**The Study**

This study sought to answer the following questions:

1) Which characteristics of SFL speech do listeners find to be the greatest barriers to comprehensibility?

2) Which features of SFL speech do evaluators classify as irritating, and which as pleasant?

3) Are the answers to the above the same for NSSs from northern Mexico and from central Argentina?

4) Are there any consistent differences between the evaluations of naive NSSs as compared to those of NSSs with extensive linguistic training and an advanced knowledge of the speakers’ L1, English?

5) Do non-native teachers of SFL focus on the same features that the NSSs do in their evaluations?

6) For each group of evaluators, are the obstacles to comprehensibility that they identify the same features that are categorized as irritating, and are these the features that are identified as signaling foreignness?

**Materials**

**Recordings**

The SFL speakers were 60 first-semester university Spanish students in their thirteenth week of the semester. They read two verses of the poem “Versos sencillos” by José Martí (see Appendix). This poem appeared in their textbook (Terrell, Andrade, Egasse, and Muñoz 2006) in the chapter they were studying at the time (163). Students were asked to study the poem before class. In class, students listened to a professional recording of the poem read by a native, Latin American Spanish speaker. Next, students, still with the text in front of them, heard the poem read line by line and were recorded repeating each line during the pauses. This was done in order to minimize mispronunciations caused by a lack of familiarity with Spanish orthography. Originally over 200 recordings were made. These were recorded on HP desktop computers using the software Sony Virtuoso/Soloist 9.0. Sixty recordings were chosen semirandomly, selecting 30 men and 30 women, evenly distributed across classes. A few originally selected recordings were rejected for having background noise or low volume. These were replaced by other recordings.
representing similar demographics. The recordings that the evaluators listened to were edited so that only the students’ voices were heard, as if reading the poem straight through, with natural, brief pauses between lines. They were about 50 seconds long each.

**Questionnaire**

All evaluators were given the same questionnaire, translated into their native language, to fill out for each recording they listened to. At the top of the questionnaire was the text of the poem. Evaluators were encouraged to mark the text to help them remember what they heard, as they were instructed to listen to each recording only once. Next they were asked to answer the questions, “How easy is it to understand what the speaker is saying?” and “How pleasant is the speaker’s accent?” by marking on a scale of 1 (easy) to 7 (difficult) and from 1 (irritating) to 7 (pleasant), respectively. To further capture evaluators’ impressions of the speakers, the last item of the questionnaire was “Give two adjectives that could describe this person.”

Six other open-ended items were included to explore the question of which pronunciation features were most influential. These were, in their original order:

1) List aspects of the speaker’s accent that are good or unproblematic.
2) What is irritating about the speaker’s way of speaking?
3) What is pleasant about the speaker’s way of speaking?
4) If the speaker was difficult to understand, what made it difficult?
5) What about the speaker’s accent makes him/her sound foreign?
6) What aspect(s) of the speaker’s pronunciation need(s) improvement?

**Evaluators**

The five groups of judges were:

- 40 Mexican college students studying the teaching of ESL (ME). These judges spoke English at an advanced level and had all taken at least one class in phonology.
- 107 students from the same Mexican university (MN). These judges are considered naive listeners because they had a wide range of other majors and did not possess an advanced level of English, having been recruited from beginning and low-intermediate ESL classes.
- 50 Argentine college students majoring in English (AE). These students also had received courses in phonology and spoke English at an advanced level.
- 49 students from the same Argentine university but who were not language majors (AN). These naive listeners were recruited from a variety of undergraduate science and humanities classes and self-reported having low proficiency in English.²
- Three native English speakers who were SFL instructors at the SFL students’ university (NES). All three had taken classes in Spanish phonology.

The Mexican raters lived in a city near the US border, and the Argentine raters lived in the central province of Córdoba. There are a number of substantial differences in the dominant dialects of those areas.² Besides differences in pronunciation, it was thought that the Mexican and Argentine evaluators might display different attitudes toward a US accent in SFL, given that the two groups have very different levels of frequency of contact with native speakers of American English. Whereas the Mexican participants lived in a region that is frequented by English speakers from the United States and where many of its residents often visit or reside temporarily in the United States, the Argentine participants on the whole would have had far less contact with people from the United States.
Procedure

Using headphones, the NSSs listened to the recordings on mp3 players. 24 playlists of five recordings each were created so that each recording appeared twice. A recording’s position (early, middle, or late) was alternated between lists. Attention was also given to alternating whether a speaker was followed by a speaker of the same sex or not. NSS evaluators typically had one hour for the procedure. They rated an average of four recordings each. This resulted in 595 evaluations from the Mexican group and 395 from Argentina. The NES group listened to the recordings on a computer, using headphones. NES evaluators each rated all 60 recordings (180 evaluations total). The playlist had been counterbalanced for gender and the speakers’ SFL instructors. The NES evaluators did not have information about the speakers’ nor their instructors’ identities.

Results

Likert-scale Data

Table 2 shows the average ratings by group, on a scale of 1 to 7. The Mexican raters were the most generous, on both scales, followed next by the non-native raters and last by the Argentines. A one-way analysis of variance (ANOVA) test confirmed that scores differed significantly by the rater’s nationality, both for comprehensibility, $F(2, 1160) = 6.02, p = .003$, and for pleasantness ratings, $F(2, 1161) = 7.02, p = .001$. Among the NSSs, the English majors rated the speakers higher than did the other evaluators: $F(1, 977) = 4.3, p = .038$ for comprehensibility and $F(1, 978) = 7.68, p = .006$ for pleasantness. Comprehensibility correlated significantly to pleasantness ratings: $r = .66$, $p$ (1-tailed) $= .01$ for NSSs and $r = .51$, $p$ (1-tailed) $= .01$ for NESs.

Table 2. Average ratings by group

<table>
<thead>
<tr>
<th></th>
<th>Mexico</th>
<th></th>
<th></th>
<th>Argentina</th>
<th></th>
<th></th>
<th>United States</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Comprehensibility</td>
<td>Pleasantness</td>
<td>Comprehensibility</td>
<td>Pleasantness</td>
<td>Comprehensibility</td>
<td>Pleasantness</td>
<td>Comprehensibility</td>
<td>Pleasantness</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>5.09</td>
<td>4.49</td>
<td>4.74</td>
<td>4.02</td>
<td>4.98</td>
<td>4.14</td>
<td>4.98</td>
<td>4.14</td>
<td></td>
</tr>
<tr>
<td>English majors</td>
<td>5.37</td>
<td>4.77</td>
<td>4.86</td>
<td>4.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other majors</td>
<td>4.98</td>
<td>4.39</td>
<td>4.62</td>
<td>3.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Qualitative Data

Analysis of the open-ended responses involved iterative stages of reading, code development, and coding of the data (Miles and Huberman 1994). Responses that were judged to fall into the same category were compared, and frequently the original sound files were listened to again to guide interpretation. The coding of these answers required an understanding of the connotations of many expressions, and NSSs from the two areas where the NSS evaluators were recruited were consulted.

Positive Aspects

In order to determine which characteristics of the speakers’ pronunciation featured most prominently in the responses of each group of evaluators, percentages were calculated. Thus, if we look at Table 3, which presents each group’s 5 most frequent answers to the item “List...
Table 3. Most frequently named positive aspects of speech samples

<table>
<thead>
<tr>
<th>Mexico (100% response rate)</th>
<th>Argentina (70% response rate)</th>
<th>United States (100% response rate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ME</td>
<td>MN</td>
<td>AE</td>
</tr>
<tr>
<td>Comprehensibility (22%)</td>
<td>Fluency (18%)</td>
<td>’r’ (47%)</td>
</tr>
<tr>
<td>Intonation (13%)</td>
<td>Fluency (12%)</td>
<td>Pausing (11%)</td>
</tr>
<tr>
<td>Fluency (12%)</td>
<td>Intonation (8%)</td>
<td>’r’ (4%)</td>
</tr>
<tr>
<td>Pausing (11%)</td>
<td>’r’ (8%)</td>
<td>’r’ (5%)</td>
</tr>
<tr>
<td>’r’ (5%)</td>
<td></td>
<td>Word stress (2%)</td>
</tr>
</tbody>
</table>

Table 4. Most frequently named pleasant aspects of speech samples

<table>
<thead>
<tr>
<th>Mexico (87% response)</th>
<th>Argentina (55% response)</th>
<th>United States (75% response)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ME</td>
<td>MN</td>
<td>AE</td>
</tr>
<tr>
<td>Voice (16%)</td>
<td>Voice (24%)</td>
<td>Fluency (14%)</td>
</tr>
<tr>
<td>Effort (15%)</td>
<td>Comprehensibility (14%)</td>
<td>Comprehensibility (13%)</td>
</tr>
<tr>
<td>Pausing (8%)</td>
<td>N/A (7%)</td>
<td>Voice (11%)</td>
</tr>
<tr>
<td>Intonation (7%)</td>
<td>Fluency (7%)</td>
<td>Accuracy (8%)</td>
</tr>
<tr>
<td>Foreignness (7%)</td>
<td>Calm (5%)</td>
<td>Intonation (8%)</td>
</tr>
</tbody>
</table>
was created to capture answers such as “Es una persona que aunque le es dificil en algunas ocaciones, continua intendentandolo.””

The category ‘N/A’ was applied only when respondents wrote in answers expressing that little to nothing was pleasant about the speech sample, but this sentiment was probably also shared by most of the raters who did not answer the question. If so, roughly half of the evaluators from Argentina found nothing pleasant about the SFL learners’ speech. In contrast, 7% of ME’s answers to this question were that they enjoyed the non-native accent of the speaker. Other frequently remarked upon features, for all groups, were the kindness or self-assurance that the evaluator sensed from the speaker’s voice.

Negative Aspects

Next we turn to the problematic features that were identified by evaluators. Table 5 shows the top responses to the question about comprehensibility. Again, ‘N/A’ was assigned to written responses and not items left blank. If, however, blank answers also indicated that the item was not applicable, then in over half of the cases, the NSSs felt that there was nothing about the learner’s speech that impeded comprehension.

A typical ‘omissions’ answer was, “se come letras.” The category ‘transfer’ was used when the evaluator stated that L1 English sounds had substituted the target Spanish sounds. Answers counted as ‘too fast’ tended to state also that the speaker had run his or her words together. Specific mentions of phonemes besides /l/ and /ɾ/ were few: 6 for /l/, 5 for /d/, and 3 for /ɾʃ/.

Table 5. Obstacles to comprehensibility

<table>
<thead>
<tr>
<th>Mexico (52% response)</th>
<th>Argentina (51% response)</th>
<th>United States (85% response)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ME</td>
<td>MN</td>
<td>AE</td>
</tr>
<tr>
<td>N/A (16%)</td>
<td>N/A (22%)</td>
<td>N/A (8%)</td>
</tr>
<tr>
<td>‘r’ (7%)</td>
<td>‘r’ (10%)</td>
<td>Fluency (6%)</td>
</tr>
<tr>
<td>Fluency (5%)</td>
<td>Fluency (7%)</td>
<td>Intonation (4%)</td>
</tr>
<tr>
<td>Transfer (1%)</td>
<td>Too fast (3%)</td>
<td>Vowels (3%)</td>
</tr>
<tr>
<td>Intonation (1%)</td>
<td>Intonation (2%)</td>
<td>‘r’ (2%)</td>
</tr>
</tbody>
</table>

Table 6 shows the top answers to the question, “What is irritating about the speaker’s way of speaking?” ‘Inaccuracies’ was used to categorize comments such as “confused pronunciation.” ‘Attitude’ was assigned to answers such as “parece arrogante” or comments about the speaker sounding lazy, tired, or sad.

In answer to “What about the speaker’s accent makes him/her sound foreign?” many evaluators wrote that it was the learner’s accent. This response, along with comments about transfer, although quite frequent, were not considered illuminating for the present purposes and so are not among the categorized responses that are presented in Table 7. Answers such as “Porque batalla para pronunciar algunos sonidos” were put in the ‘difficulty’ category. Because several evaluators named more than one feature, we can see that in Table 7, percentages reported for some groups of evaluators are greater than 100% (see note 6).
Table 6. Irritating aspects of speech samples

<table>
<thead>
<tr>
<th>Mexico (84% response)</th>
<th>Argentina (67% response)</th>
<th>United States (83% response)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ME</td>
<td>MN</td>
<td>Fluency</td>
</tr>
<tr>
<td>N/A 32%</td>
<td>N/A 32%</td>
<td>N/A 12%</td>
</tr>
<tr>
<td>'r' 9%</td>
<td>Fluency 11%</td>
<td>Fluency 13%</td>
</tr>
<tr>
<td>Inaccuracies 8%</td>
<td>Fluency 14%</td>
<td>'r' 14%</td>
</tr>
<tr>
<td>Fluency 6%</td>
<td>'r' 9%</td>
<td>Fluency 9%</td>
</tr>
<tr>
<td>Attitude 4%</td>
<td>Attitude 3%</td>
<td>Inaccuracies 9%</td>
</tr>
</tbody>
</table>

Table 7. Factors that contribute to foreignness

<table>
<thead>
<tr>
<th>Mexico (88% response)</th>
<th>Argentina (84% response)</th>
<th>United States (100% response)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ME</td>
<td>MN</td>
<td>'r' (61%)</td>
</tr>
<tr>
<td>'r' (36%)</td>
<td>'r' (33%)</td>
<td>'r' (61%)</td>
</tr>
<tr>
<td>Vowels (10%)</td>
<td>Fluency (12%)</td>
<td>Fluency (13%)</td>
</tr>
<tr>
<td>Fluency (8%)</td>
<td>Intonation (6%)</td>
<td>Fluency (16%)</td>
</tr>
<tr>
<td>Intonation (4%)</td>
<td>Word stress (4%)</td>
<td>Vowels (11%)</td>
</tr>
<tr>
<td>/tl/ (3%); /dl/ (3%)</td>
<td>/tl/ (11%); /s/ (3%)</td>
<td>/tl/ (10%)</td>
</tr>
</tbody>
</table>

Here, as elsewhere, when vowels were mentioned, the most common comment was to say that the 'o' sounded like the diphthong /ou/. The other frequently mentioned vowel mispronunciations were 1) pronouncing the letter 'e' (/e/) as a diphthong or as /i/; 2) converting the /a/ to a schwa; and 3) that pronouncing the wrong vowel sometimes resulted in incorrect gender assignment to words. Although remarks about the pronunciation of the grapheme 'r' are grouped together in Table 7, it was often clear whether the trill /r/ or the flap /ɾ/ was meant. For all groups of evaluators, there were more (roughly twice as many) mentions of the flap /ɾ/. Many raters said that the worst pronunciation of the letter 'r' was when it appeared in word-final position. Other frequent answers for this item were linking, word stress, and, among the Mexican evaluators, /x/, saying that the pronunciation was too strong, which was interpreted to mean too fricative.

Table 8 reports the most common answers to “What aspect(s) of the speaker’s pronunciation need(s) improvement?” The consonant clusters mentioned by the AE group were letter combinations of either 'I' or 'r' with one other consonant. The sound /l/ appeared frequently in the NSSs' responses and the sound [v] among the NESs, although these sounds do not appear in any other top-five lists.

The final question on the survey asked the raters to describe how they imagined the speaker. Participants were told that anything that came to mind, including both personality...
and physical characteristics, was valid. Although the wording of the question requested “two adjectives,” evaluators usually answered in full phrases. The average comprehensibility and pleasantness scores for all recipients of each of these coded categories of descriptions were then calculated. Table 9 shows the categories that received the five highest and five lowest average comprehensibility scores according to each nationality group, and Table 10 shows those data for average pleasantness scores.

**Discussion**

Responses from the four groups of NSS evaluators were more similar to evaluations from other NSSs than to those from NESs. The NES evaluators focused largely on vowels and...
transferred L1 sounds, whereas the NSSs commented on a somewhat different list of segmentals and, more frequently, on suprasegmental features and comprehensibility. Patterns of differences among the NSS responses can be discerned, however, according to the NSS evaluators’ level of previous exposure to SFL speakers and to regional differences in their own NSS pronunciations.

NSSs

The NSSs associated comprehensibility with pleasantness. When evaluators found the speech to be clear, this in itself was cited among the speaker’s merits. The qualities that raters attributed to the speakers corroborate these findings. Speakers judged to be highly comprehensible were imagined to be attractive, nice people, whereas low comprehensibility scores were associated with negative characteristics. Some pronunciation inaccuracies, even when not disrupting comprehensibility, were described as irritating, especially by Argentine raters, the group least familiar with the speakers’ L2 accent.

Tables 9 and 10 show that in terms of characteristics that NSSs associated with better and worse pronunciations, there was substantial agreement between the Argentine and Mexican participants (with the curious exceptions of weight and athleticism categorizations). Quantitatively (Table 2), however, there was a significant difference between the Mexicans and Argentines in terms of their overall level of acceptance of the learners’ pronunciation. This agrees with previous research that suggests that familiarity with an accent leads to a greater acceptance (DeShields and Kara 2011) and comprehension (Bradlow and Bent 2008). One implication of this is that L2 speakers who intend to associate with native speakers unaccustomed to their L2 accent may have stronger reasons to work on their L2 pronunciation. The English majors among the NSSs were significantly more generous in their ratings than their non-expert counterparts. Again, this may be a result of greater familiarity. It may also be that many English students chose their field because of an interest in the English-speaking world.

In the current study, suprasegmentals and other global features of the speech samples received a great deal of attention. The category of fluency appears in all but one of the NSSs’ 24 top-five lists. Many of the SFL speakers read the text quite fluently, and this was commented upon favorably by the NSS evaluators. Although the nature of this study’s production task eliminated many of the fluency challenges associated with unrehearsed speech, there were

<table>
<thead>
<tr>
<th>Mexico</th>
<th>Argentina</th>
<th>United States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-assured</td>
<td>Self-assured 5.0</td>
<td>Intelligent 5.3</td>
</tr>
<tr>
<td>Nice 4.8</td>
<td>Interested; Hard-working; Thin; Nice 4.8</td>
<td>Serious 5.0</td>
</tr>
<tr>
<td>Attractive; Young 4.7</td>
<td>Attractive 4.6</td>
<td>Self-assured 4.9</td>
</tr>
<tr>
<td>Calm 4.6</td>
<td>Intelligent 4.2</td>
<td>Nice; Happy 4.7</td>
</tr>
<tr>
<td>Serious; Overweight 4.6</td>
<td>Short 4.1</td>
<td>Hard-working; Quiet 4.6</td>
</tr>
</tbody>
</table>

Table 10. Descriptions with highest and lowest average pleasantness scores
occasions in some recordings where the SFL speakers stumbled over a word or spoke haltingly. These disfluencies, although minor compared to how students at that level might otherwise speak when extemporizing, appear to have had a major impact on the NSSs’ perceptions; fluency is mentioned in all NSS groups’ top-five lists regarding obstacles to comprehensibility, irritating features, markers of foreignness, and aspects needing improvement.

Intonation was the second most commonly mentioned suprasegmental. Raters were also influenced by the related characteristics of voice quality and the attitude that they perceived in speakers. These latter speech characteristics are issues that are not normally mentioned in L2 pedagogical materials, but they have been discussed with respect to some groups of ESL students (Derwing 2008). Self-assurance, calmness, and sincere effort were responded to positively, while perceived boredom and arrogance provoked condemnation.

While suprasegmentals and other global characteristics dominated NSSs’ remarks about comprehensibility and pleasantness, segmental features received more of their attention when they listed foreign-sounding features. Pronunciations of the letter ‘r’ (especially tap /ɾ/) were the most important to the NSSs, appearing on every top-five list except in Table 4, about pleasant features, where it is absent. Although vowels were named by the NSSs as features marking foreignness and needing improvement, vowels were not identified as important obstacles to comprehensibility. This contrasts sharply with Schairer (1992), in which vowels were found to be overwhelmingly the most important feature. One possible explanation for this is that other factors, such as fluency, intonation, and the tap /ɾ/—some of the most frequently remarked upon features in the present study but which were not taken into account in Schairer—may correlate highly with the mispronunciation of vowels. If so, vowel pronunciation would have served as a proxy factor in Schairer’s regression analysis, picking up the variation that would have been attributed to other factors had they been included in the analysis. This points to a disadvantage in preselecting features when investigating speech perception.

According to the NSSs of this study, the segmentals that followed the pronunciation of the letter ‘r’ in importance were /d/ and /t/. The numerous mentions of /d/ are likely connected to the obligatory but not phonemic (that is, not representing minimal pairs) fricative-stop contrasts that also apply to /b/ and /ɡ/, and, similarly, mentions of /t/ speak to the importance of repressing an English-like aspiration of /p, t, k/. The data contained mentions of /b, g, p, k/ as well, but these were considerably less frequent. These four phonemes were somewhat less numerous in the text that was read aloud. Another likely explanation for the greater focus on /d/ and /t/ is that, in addition to the Spanish-English contrasts just mentioned, these phonemes have a different point of articulation in Spanish (in which they are dental) and English (alveolar) (Schwegler and Kempff 2007: 143, 206). In fact, some NSS remarks on the SFL pronunciation of /d/ specified that the learners’ /d/ sounded like the alveolar flap /ɾ/ of ‘r.’

The other segmental that was frequently noted was the /l/, which in Spanish is never realized as the ‘dark l’ found in syllable-final position in English (Schwegler and Kempff 2007: 266). Transfer of this L2 pronunciation can in turn affect surrounding sounds. The surprising number of times that evaluators identified an ‘m’ as mispronounced appears to have been, in all cases, due to the influence of a mispronounced ‘l’ in the word alma.

Differences as to which features were named across nationalities were few. One striking difference, however, was with ‘y’ and ‘ll,’ which Argentines pronounce differently from the majority of the Spanish-speaking world. Argentines not only labeled this as a signal of foreignness but also as a feature needing improvement. Similarly, although not to be found within top-five lists, some Mexicans complained that many learners’ /x/ was too strongly fricative. This phoneme is typically not strongly fricative in that region of Mexico (Lipski 2008: 86). Finally, mentions of learners’ ‘s’ were frequent in the AN raters’ list of foreign-sounding features. The few raters who explained said that the sound was held for too long.
NESSs

Like the NSSs, the NES evaluators attributed positive characteristics to the speakers with clearer accents. NESSs remarked on rhythm less and segmentals more than their NSS counterparts. Top among the segmentals were vowels, in contrast with the NSSs. This is likely a byproduct of the NESs’ training: having been instructed that vowel pronunciation is essential to good Spanish pronunciation (e.g., Schwegler and Kempff 2007), they assumed that it heavily impacted comprehensibility. This assumption may also have come from their experiences with non-native speakers of English. Since mispronounced vowels can seriously reduce comprehensibility in English (Jenkins 2002), the NESs may have assumed the same was true in SFL. Two other sounds that the NESs often mentioned were the [v] pronunciation of the grapheme ‘v’ and the error of pronouncing the letter ‘h.’ These are features the NSSs mentioned rarely. Since many NSSs find it difficult to distinguish between [v] and [b] when they learn English (Avery and Ehrlich 1992: 149), it may be that few NSSs can detect this error in Spanish.

Since NES raters shared the SFL learners’ L1 and were SFL instructors at the learners’ university, they were very familiar with the pronunciation of this group and had little difficulty understanding the recordings. Therefore the features that they listed as obstacles to comprehensibility were largely conjectures on their part. The evidence suggests that the NESs judged the L2 speech against a list of features of Spanish phonology that they, as advanced speakers and teachers of SFL, were aware of. This helps to explain why, in keeping with previous research, non-native speakers in this study focused more on details of form and less on issues of comprehensibility.

As for the questionnaire item about personal attributes, differences between evaluator groups appear to have more to do with peer status than with native status. The NESs had reason to imagine the L2 speakers as potential students of theirs, while the NSSs, being fellow college students, may have imagined the L2 speakers as potential classmates and friends. NSSs made numerous comments about the speakers’ level of attractiveness, while NESs made no such observations. NESs focused instead on intelligence, effort, and seriousness. They also, like many teachers might, detected some boredom in the best performing students but found their boredom irritating.

Limitations and Future Research

Findings would have been more potentially generalizable if there were a similar number of NES raters as NSSs. It is possible that the three NES raters displayed a high degree of agreement in their evaluations as a result of working on the same teaching team. SFL teachers from other institutions might focus on a somewhat different set of features. It would be interesting to investigate the ways in which the beliefs of non-native L2 teachers, across institutions, vary in terms of what they value most in L2 pronunciation.

The effect of the experimental task should also be further explored. While studies of accent perception typically rely on read texts, it would be important to see how results vary with a change in the type of text or with spontaneous speech. Using spontaneous speech would likely result in speech samples that varied much more in terms of fluency, and this, along with the speakers’ word choice and differing grammaticality, could impact evaluators’ perception of pronunciation.

Teaching Implications

The most fundamental implication of the results of this study is that L2 pronunciation matters, not only in terms of comprehensibility, but because it has a serious impact on native and non-native speakers’ initial impression of the speaker. The findings summarized in Tables 9 and
10 demonstrate that negative personality characteristics are frequently attributed to L2 speakers with inaccurate or unpleasant pronunciation. Sharing this information with L2 learners may help them to understand the importance of improving their pronunciation.

SFL pronunciation instruction has tended to focus on segmentals (Arteaga 2000). This study, however, indicates that—as has been found in ESL studies (e.g., Derwing, Munro, and Wiebe 1998)—suprasegmentals and global speech characteristics heavily impact listeners’ perceptions. The high frequency with which fluency was mentioned by NSSs—despite the fact that the speech samples were of rehearsed speech—suggests that fluency plays an important role in shaping native speakers’ perception of SFL speech. Fluency promotion activities such as the ones described in Derwing, Munro, and Wiebe (1998) and Rossiter et al. (2010) might be adapted to target SFL fluency. Because some of these activities involve formulaic expressions and mimicry, they may strike some teachers as incompatible with communicative teaching practices, but Iruela (2007) and Rossiter et al. (2010) argue convincingly that this is not the case.

Also striking in the present study was the extent to which evaluators reacted to the attitudes that they perceived in the L2 speakers’ voices. This factor is rarely mentioned in pronunciation teaching materials but appears to merit some attention. If not reminded of the importance of the overall quality of their speech, students may focus too exclusively on very fine details, which can be damaging to fluency (Derwing, Munro, and Wiebe 1998; Iruela 2007).

Virtually every contrastive sound difference between Spanish and English appeared at least once in the data. This means that in order to remove all markers of a non-native pronunciation, all such sounds would eventually need to be attended to. L2 learners are not, however, and cannot become monolingual native speakers of their target language. Reasonable and useful learning goals will be those that most promote successful communication via the L2. The present study has helped to identify which features of SFL pronunciation can be most important in achieving such goals.

ACKNOWLEDGEMENTS

This study was supported by Saint Louis University’s Summer Research and Mellon Faculty Development Grants. Many thanks to research assistants Mariela López and Haley Jacobson, Nolvia Cortez in Mexico, and Lidia Soler in Argentina for their assistance in conducting the study and ideas on how to proceed. I am also indebted to all of the study’s participants, the cooperating university instructors, and several consultants from the three regions involved.

NOTES

1 ‘SFL’ is used here to mean specifically native English speakers from the United States.
2 The actor had been professionally trained to speak in a dialect-neutral style (Scott Tinetti, Senior Director of Development, World Languages, McGraw-Hill, email message to author, September 6, 2013).
3 Whereas the Mexican university required all undergraduate students to take English, no such requirement existed at the Argentine university. This made recruiting large numbers of AN students from beginning ESL classes impossible.
4 Among these are 1) while the Mexican participants would be expected to produce most syllable final /-s/ with a full sibilant pronunciation ([s]), the Argentines frequently elide or aspirate ([h]) it; 2) the graphemes ‘y’ and ‘ll’ represent the sound [ʝ] for the Mexican population but [ʃ], [ʃʃ], [ʒ] or [dʒ] for most Argentines; 3) the phoneme /x/ in Mexican Spanish tends to be less strongly fricative than in other Spanish dialects, including the Argentine dialects; 4) vowel nasalization and vowel reduction are common processes in Mexico but not in Argentina; and 5) there exist dramatic differences in the typical intonation patterns of these two groups (Schwegler and Kempff 2007).
5 When a new evaluator listened to an mp3 player, the next recording up would play, further randomizing the order in which recordings were heard.
Percentages were calculated by dividing the total number of one kind of answer by the total number of evaluation sheets from that group, whether or not an answer for that particular question was written.

Comments are reproduced as originally written, including errors.

Argentine English instructors consulted confirmed that these are taught explicitly in English classes there.

WORKS CITED


APPENDIX

The selection of Rubén Darío’s “Versos sencillos” that the learners read:

Yo soy un hombre sincero
De donde crece la palma,
Y antes de morirme quiero
Echar mis versos del alma.

Y vengo de todas partes
Y hacia todas partes voy:
Arte soy entre las artes.
En los montes, monte soy.

Oigo un suspiro, a través
De las tierras y la mar,
Y no es un suspiro, —es
Que mi hijo va a despertar.

Con los pobres de la tierra
Quiero yo mi suerte echar
El arroyo de la sierra
Me complace más que el mar.