Native Speakers' Perceptions of Nonnative Speakers: Related to Phonetic Errors and Spoken Grammatical Errors.

A study investigated the perceptions of native English-speakers concerning the spoken grammatical and phonetic (accent) errors of non-native speakers. Speech samples were collected from three non-native speakers of English of varied linguistic backgrounds (German, Spanish, and Arabic) and one speaker of North American English. Each of the four speakers recorded two speech samples, one containing a grammatical error only and one containing a phonetic error only. A group of 124 Florida college freshmen then heard the speech samples and completed a semantic differential scale questionnaire for each. Analysis revealed significant differences in reactions across languages and error types. Phonetic error was rated lower than grammatical error, despite the fact that grammatical errors sometimes obscured meaning while phonetic errors did not. Spanish and German accents were rated about the same, but the Arabic accent was rated lower than the others. It is suggested that American students be taught greater tolerance of international students' speech. The appendix includes one table and two graphs. Contains 38 references. (MSE)
Native Speakers' Perceptions of Nonnative Speakers:
Related to Phonetic Errors
and Spoken Grammatical Errors

Ruth Johnson
Southern Illinois University
Department of Linguistics
Mailcode 4517
Carbondale IL 62901

Frederick L. Jenks
The Florida State University
CIES
918 West Park Avenue
Tallahassee FL 32306

Acknowledgements
The authors thank Drs. Paul Angelis, Usha Lakshmanan,
and Kim Hughes Wilhelm for their comments on earlier
versions of this article.
Ruth Johnson is assistant professor of applied linguistics/TESL in the Department of Linguistics, Southern Illinois University at Carbondale. Frederick L. Jenks is professor of TESL in the Multilingual/Multicultural program, and director of the Center for Intensive English Studies, Florida State University, Tallahassee.
Abstract

The relationship between the perception of native speakers towards nonnative speakers and the spoken grammatical and phonetic errors (accent) of nonnative speakers was investigated. Speech samples were collected from three nonnative speakers of English--German, Spanish, and Arabic--and an American English control. Each speaker recorded two speech samples--one that contained grammatical error only and one that contained phonetic error only. One hundred and twenty-four subjects were randomly selected from among college freshmen. Using the matched guise technique, each respondent heard the eight speech samples and completed a semantic differential scale questionnaire for each. Statistical analyses revealed significant differences between the reactions across languages and between error type.
Native Speakers' Perceptions of Nonnative Speakers:
Related to Phonetic Errors
and Spoken Grammatical Errors

The stream of speech may be perceived by the interlocutor as a unified entity; in other words, s/he may not usually distinguish one component from another. The applied linguist, however, views the stream of speech as comprised of several components--segmentals, suprasegmentals (prosody), subsegmentals (voice quality settings), grammar. Much is known about how native speakers react to nonnative speech as a unified entity. Research (detailed below) has investigated how native speakers react to accented speech as compared to native-like speech and how native speakers react to various degrees of accented speech of a given language background. These studies have looked at the speech signal as a unified entity. Delamere (1986) conducted research in which the nonnative speech sample was broken into two components: one part was speech with an accent; the other part, speech with an accent and grammatical error. In this case, the grammatical error was identical for each nonnative speech sample, regardless of language background. What is needed is
research that investigates native speaker reaction to various components of nonnative speech. This study investigated the reactions that native speakers of English have towards nonnative speakers of English when the latter's speech contains only phonetic errors ("accent") or only grammatical errors. The three native language backgrounds that were chosen for this study are German, Spanish, and Arabic.

Research has showed that native speakers are always able to determine whether or not a particular person or group of people speak with an accent and, subsequently, make judgments about those people based on their accent (Lambert, Hodgson, Gardner, & Fillenbaum, 1960; Anisfeld, Bogo, & Lambert, 1962; Lambert, Frankel, & Tucker, 1966; Arthur, Farrar, & Bradford, 1974; Ryan & Carranza, 1975; Ryan, Carranza, & Moffie, 1977; Brennan & Brennan, 1981). Furthermore, accents are not all equal; that is, they are not valued equally: native speakers like the sounds of some accents, dislike the sounds of other accents, and like and dislike the people who speak with these accents, respectively (Delamere, 1986). In addition, people with a particular accent are labeled as a homogeneous
group by those outside their speech community while within their own community their accent is used to distinguish one from another (Trudgill, 1974; Riches & Foddy, 1989).

Classification by persons untrained in linguistics is accomplished based upon a portion of language—the speech signal. The speech signal itself has remained difficult to research. Speech, in actuality, is a continuous signal produced by a continuous series of modifications in the vocal tract and larynx. Portions of speech include the segmentals, suprasegmentals, structures of syllables, and voice qualities. For the nonnative speaker, besides errors in these portions of the speech signal, grammatical errors peculiar to the influence of his/her native language are often present. All of these, phonological and grammatical errors, contribute to the reactions of native speakers (Delamere, 1986).

For a native speaker listening to the speech produced by a nonnative speaker, it has been established that there is what can be called a "threshold of comprehensibility." At the level below this threshold, speech is incomprehensible or extremely
difficult to comprehend (Anderson-Hsieh, Johnson, & Koehler, 1992; Johnson, 1990). The speech signal is affected in many ways. It may contain many segmental errors, on both the phonetic and phonemic levels (Flege, 1981; Beebe, 1987), or it may deviate far from native-speaker stress, intonation, and rhythm, so that, for example, the speech will have a pronounced sing-song intonation or syllable-timed rhythm (Dauer, 1983; Bolinger, 1986; Tarone 1987b). Also, it may contain numerous syllable structure errors, usually in consonant deletion and vowel insertion (Broselow, 1987; Sato, 1987; Tarone, 1987a), and/or it may contain many grammatical errors (Delamere, 1986).

The aggregate of errors is such that the native speaker cannot stretch his/her speech synthesizer (Clark & Clark, 1977) far enough to accommodate the accent, and s/he simply makes no further attempt to listen because s/he cannot comprehend. In other words, in dealing with speech that is non-comprehensible, the native speaker must spend too much time figuring out, for example, what the phonological representation of the speech is so that as a result, the message does not get attended to (Prideaux, 1985). The native speaker
will express irritation at having to listen to such speech (Ludwig, 1982). Accent is not simply noise; rather, the native speaker begins with the presumption that the speech signal contains meaning (Abbott, 1986). When listener-meaning and speaker-meaning do not match (due to the speech containing too many errors), communication breaks down.

However, when the threshold of comprehensibility has been reached, the issue of accent takes on another dimension. In this case, the native listener's inability to comprehend is not involved: his/her unwillingness to listen is. This is an issue of attitude. The native speaker may find accented speech irritating to attend to, not because s/he cannot understand the speech signal due to the presence of too many errors, but rather because s/he does not "like" the sound. The native speaker's listening itself affects the message and its reception (Ghiglione & Beauvois, 1983). Interlocutors view their listeners as active participants in the dialogue (Nisbet, Zanella, & Miller, 1984; Harrigan, 1985; Marcon, 1985). Thus, a nonnative speaker may have expectations for his/her native speaker while the native speaker may be
disinterested for one reason or another. One reason native speakers may be disinterested is that they are affected by their attitudes and/or their perceptions of people who are "different" from them (Gass & Varonis, 1985). People bring prior expectations to a communication exchange, and these expectations affect the communication itself.

In many settings, a particular language or dialect or accent will be "marked"; that is, its use will place the speaker in a lower societal category than that of the listener. In earlier studies addressing this issue, English-speakers in Canada were rated more favorably than French-speakers by both French-speaking Canadians and English-speaking Canadians (Lambert, et al., 1960). This was shown to be true when the raters were as young as twelve years old (Lambert, et al., 1966). Persons with a Yiddish accent were not judged favorably by gentiles on any criterion and were judged favorably by Jews on only some criteria (Anisfeld, et al., 1962). English-speaking college students in California rated Chicano-English unfavorably in all cases (Arthur, et al., 1974), and the heavier the accent, the more unfavorable the rating (Ryan, et al., 1974).
1977). Even when native speakers cannot identify language backgrounds (Palmer, cited in Eisenstein, 1983), there was a significant difference in how the speakers were rated.

More recently, in a comparison of various "foreign accents," it was shown that grammatical errors in accented speech both hinder and enhance the nonnative speaker's speech, depending upon the language background of the speaker (Delamere, 1986). For example, for persons who speak English with an Arabic or Farsi (Iran) accent, the reaction of a native speaker is more positive if the nonnative speaker produces speech with fewer grammatical errors. For persons who speak with a French or Malay accent, the reaction of a native speaker is more positive if the nonnative speaker produces speech with more grammatical errors. For persons who speak English with a Spanish accent, the native-listener's reaction is negative with or without grammatical errors. There is a link between the accented language used by a speaker and the native listener's perception of the speaker. Reaction to accent may be more predominant than reaction to grammatical error (Delamere, 1986).
Research is needed that investigates the question of how native speakers perceive people who speak with various "accents" (due to first language background), comparing them to each other. Of interest also is an investigation into selected components of the speech signal to determine which components a native speaker reacts to more negatively.

The present study investigates: (1) the perceptions that native speakers have of nonnative speakers who are of various language backgrounds and (2) the relationship between native listeners' perceptions of nonnative speakers and speech produced by those speakers which, on the one hand, is only accented (contains phonetic error but no grammatical error) and, on the other hand, contains only grammatical error (that is, is "accent-free"). It is predicted that accented speech will be viewed less favorably than will speech with grammatical error, and that the language background of the nonnative speaker will influence how favorably or unfavorably the native speaker views nonnative speech.
Method

The researchers employed the matched guise technique (Lambert, et al., 1966) in which three nonnative male speakers of English and one native male speaker of English each produced two speech samples (for a total of eight speech samples) that were then listened to by the subjects. Personality traits of the speakers were rated by the subjects, using the semantic differential rating scale (Osgood, Suci, & Tannenbaum, 1957).

Subjects

Subjects for this study were drawn from the freshmen classes of three institutions of higher learning in Tallahassee, Florida; namely, The Florida State University, Florida A & M University, and Tallahassee Community College. A random group sampling from all freshmen English classes was made, and all students within each of the randomly selected classes were used as subjects, with a sample size of 124.\(^1\) The particular three institutions of higher learning were chosen because each represents a different population within this region of the United States. The Florida State University is largely Anglo-American, Florida A &
M University is largely African-American, and Tallahassee Community College is a balanced mixture of race. Therefore, the sample would be a fair representation of the adult student population.

**Instrumentation**

In this study, the matched guise technique and the semantic differential rating scale were used.

Each guise (there was a total of three) was a nonnative speaker of English who could speak English without a detectable accent and with good conversational grammar (See "Speech Samples" section for a detailed definition of "detectable accent" and "good conversational grammar."). For one speech sample, the guise produced speech that was accented but contained no grammatical errors. For the matched speech sample, the guise produced speech that contained grammatical errors (typical of the type made by persons of his/her language background learning English) but with no accent. This technique coincides with that used in numerous studies that have measured perception and attitude.

The semantic differential rating scale involved subjects listening to speech samples and evaluating the
speakers' personality characteristics on a bipolar scale. This scale tends to reveal a respondent's reactions to a speaker at the connotative level. (See Appendix A for a copy of the questionnaire.)

The matched guise technique and the semantic differential rating scale combined are particularly useful in exposing a respondent's private feelings and stereotyped attitudes towards a group whose language, dialect, or accent is different (Delamere, 1986). The combination is also useful because it uniquely taps perceptions. It is irrelevant whether or not a respondent can accurately identify a language, dialect, or accent. What is important is that the perceptions of the speaker are revealed in his/her responses to the guises. In addition, the semantic differential rating scale and the matched guise technique are useful because, in combination, they are valid when investigating stereotypes of the group to which a respondent belongs. That is, group biases in evaluating others are revealed. Lambert, et al. (1966) reports good reliability in the sense that when different samples of subjects, drawn from a particular
subpopulation, are used, the same profile of traits for that group appears.

Speech samples

Criteria for the speech samples were set using characteristics of the first-language backgrounds being investigated (German, Spanish, and Arabic) that were unique to each in respect to phonetic error and grammatical error.

Audiotape recordings of many different speakers of each language background were listened to by the primary researcher. From these were selected those that (1) scored no higher than "novice-high" on the ACTFL scale and (2) scored no higher than 1.5 (from a composite score of three raters) on section 2 of the SPEAK test, the read-aloud passage, and between 150 and 180 on overall comprehensibility on the SPEAK test. Ratings on these two scales were set at these levels to ensure that the speech (1) would be minimally comprehensible (Anderson, et al., 1992) but (2) would contain a "heavy" accent and many grammatical errors.

Contents of the ACTFL-scaled interviews were used to obtain grammatical errors for the "typical" ESL learner from each of the three first-language
Thus, although the number of errors per speech sample was equal across first-language backgrounds, the type varied according to what each "typical" ESL learner spoke as an interlanguage at that level. Surely, this could be salient in the native speaker's reaction to the speech samples, but it is necessary to establish that it is the grammatical error peculiar to a language background as compared to another that is reacted to. It is of little use to have the speakers utter the same grammatical errors when, in actuality, the errors are not the same nor even equal in the "real world" environment.

Similarly, samples of speech of Section 2 of the SPEAK test (the one-minute, read-aloud passage) were used to determine what a "heavy" accent of an ESL learner for each of the three first-language backgrounds would be.

From these, "composites" were made for each language background and for each error type; that is, from numerous speech samples of a given language background was constructed a passage that contained only grammatical error and one that would be read with an "accent" and no grammatical error. The passages
that were prepared to be read with phonetic error were passages from section 2 of the SPEAK test. The passages that were prepared to be read with grammatical error were passages containing information about the speaker: education, family, home town, hobbies. No attempt was made to make these composite speech samples "equal," either in type of grammatical error or in type of pronunciation error of segmentals resulting in accent. Rather, the "typical" ESL speaker of German, Spanish, and Arabic vis-a-vis segmental error and grammatical error was presented to the subjects. Thus, for example, the Arabic speech sample containing grammatical error contains more errors related to tense/time than does the German speech sample containing grammatical error. For example, the German speech sample containing grammatical error has many "she's" in it with reference to inanimate objects: "Near the bridge is a small town. And she became big because of large water roads and railroads." On the other hand, the Arabic speech sample contained errors such as this: "I am after finish English, I'm come back to my country" in reference to a future event.
For all the speech samples, the suprasegmental aspects of the nonnative speech—rhythm, intonation, stress—were controlled; i.e., these were native-like. Likewise, the subsegmental aspects of the native and nonnative speech—voice quality settings—were controlled. Each of the speakers used in this study had a more lenis pronunciation and spoke with an "open" jaw, "retroflex" articulation, "nasal" voice, "creaky" (low pitch range) voice, and "lowered" larynx (Esling & Wong, 1983).

Once the accented speech sample and the grammar-error speech sample for each first-language background were determined, the speech samples were audiotaped in English by three male international bilingual speakers: one a native Spanish speaker, one a native Arabic speaker, and one a native German speaker. Each read one passage that contained only grammatical error "typical" of the ESL learner of his/her first-language background and a second passage with no grammatical error but with a "heavy" accent; that is, an accent matching that of a person of that particular language background who scores about 180 on the SPEAK test for comprehensibility.
The three speech samples that contained grammatical error but no accent were judged to be free of accent by the researchers and two others trained in linguistics. Likewise, the three speech samples that contained accent but no grammatical error were judged to be free of grammatical error by the researchers and two others trained in linguistics.

What was of interest in this study was the comparison of the reactions of native speakers to two characteristics of nonnative speech—ungrammatical utterances and accent (phonetic error)—across three first-language backgrounds.

Administration of the test

Subjects listened to an audiotape containing random combinations of the guises, including a native-English-speaker guise that served as a control. (The two speech samples from the guise contained neither grammatical error nor phonetic error. The speech sample control used for grammatical error contained biographical information similar to that in the speech samples containing grammatical error for the nonnative speech samples. Likewise, the speech sample control used for phonetic error was a passage read by the
native speaker obtained from section 2 of the SPEAK test).

Subjects were told that they would be hearing voices of students in Tallahassee, that they were to make a judgment on the probability of success of the student in his university studies by completing a 20-item questionnaire (the semantic differential rating scale, explained above), and that they might begin rating a speech sample whenever they had a clear idea of their judgment. Subjects were not told they would be hearing each voice twice, but rather that they would be hearing eight speakers, each one talking for about one minute. The presence or absence of accent or grammatical error was not mentioned.

Results

The data were analyzed using a 2 x 4 ANOVA (repeated measures) analysis to compare evaluations of language backgrounds and evaluations of error types (see Table 1).

The first step was a comparison of means for main effects; namely, (1) language background and (2) error type and for interaction between the main effects. Both effects and the interaction were statistically
significant. Language background effect had an F value of 39.12 (DF=3; p<0.001); error type effect had an F value of 33.78 (DF=1; p<0.001); and the interaction of language background effect with error type effect had an F value of 5.54 (DF=3; p<0.001).

The next step was a Scheffe post-hoc comparison of means for language background effect. The group mean for English (the control) was statistically different from all other language backgrounds as was the group mean for Arabic, with the English control being rated the highest and Arabic being rated the lowest. The group means for German and Spanish, while statistically different from those of English and Arabic, were not statistically different from each other.

The final step, a comparison of group means for simple effects, showed that, for each language background, with the exception of the English control, the difference between the group means for grammatical error and for phonetic error (accent) within each language background was statistically different (see Figure 1), with the grammatical error guise rated more highly than the phonetic error guise. A comparison of the group means for grammatical error and phonetic
error of the English control had an F value of 0.00 (DF=1; p<0.962); that of German had an F value of 26.78 (DF=1; p<0.001); that of Spanish had an F value of 13.83 (DF=1; p<0.001); and that of Arabic had an F value of 17.60 (DF=1; p<0.001).

Within the grammatical error type, comparisons of the group means of the language backgrounds were ranked in this order: the English control, German, Spanish, and Arabic. The group mean of the English control was statistically different from Spanish and Arabic but not from German. The group means of German and Spanish were not statistically different from each other while that of Arabic was statistically different from those of every other language background (see Figure 2). The F value of these comparisons of means was 14.51 (DF=3; p<0.001).

Within the phonetic error type, comparisons of the group means of the language backgrounds were ranked in this order: the English control, German, Spanish, and Arabic. The group mean of the English control was statistically different from those of every other language background. The group means of German and Spanish were not statistically different from each
other while that of Arabic was statistically different from those of every other language background (see Figure 2). The F value of these comparisons of means was 30.35 (DF=3; p<0.001).

Discussion

The ratings given the American English control guise in comparison to all of the others was expected. The fact that the subjects rated the group of speech samples with phonetic error lower than they did the group of speech samples with grammatical error was predicted, as was the fact that for a given language background, speech with phonetic error was rated lower than that with grammatical error. From an applied linguist's point-of-view, this is interesting: The speech samples containing phonetic error were comprehensible. All that was asked of the listener was to speech that was accented but comprehensible; in other words, to "tolerate" an accent. In the case of the speech samples containing grammatical error, however, some of them obscured meaning more so than others.

What this study establishes once again is that American speakers of English rate an accent negatively.
Also, the reaction of American speakers of English to two components of speech—phonetic error and grammatical error—and their comparison to each other has been shown.

However, regrading ratings of accents, what is different in the results obtained in this study, as compared to those of Delamere (1986), is that German and Spanish accents were not rated differently while an Arabic accent was rated differently, and lower, than both German and Spanish. The rating of the Spanish accent, in particular, went contrary to expectations. The Spanish accent was expected to have been rated lowest, and certainly was expected to have been rated differently from the German accent.

The explanation may lay in the fact that German and Spanish are from the same language branch (Indo-European) while Arabic is from another (Semitic).

Or, as noted by Zadidi (1967), people’s reactions to those of another nationality could be swayed by changes in their government’s stance towards the foreign country and/or by changes in general public opinion towards people from the foreign country. Subjects in this study could have rated the Arabic
accent due to influence of the (then) relatively recent political events, the Gulf War and the bombing of the New York City World Trade Center. Similarly, subjects could have rated the Spanish accent higher than the Arabic one and equal to the German one because an Hispanic accent has become "acceptable" in north Florida, where a sizeable percentage of the population is of Hispanic origin.

These questions have implications for further research. Already, data for Phase II of this project have been collected, this time in southern Illinois, with a view to comparing reactions across regions within the U.S. and to establish a profile of American-English speakers towards accent and grammatical error in nonnative speech. Another issue that needs consideration is that of gender: Speech samples of female voices can be used to determine if NS reactions are the same for each language background when gender changes. Also, speech samples from more language backgrounds, such as those of Asians, could be investigated.

Perhaps a more important implication of this study concerns the training of American students. As the
population of international students increases on American college campuses and, in particular, as the number of international students who are hired as teaching assistants increases in numbers, the task becomes not only of one of reducing the phonetic and/or grammatical error of the international students but of increasing the awareness of and acceptance of international students by American students.

Thus, training can be selective: reducing a nonnative speaker's accent where the accent is causing negative reactions, reducing a nonnative speaker's grammatical errors where they are causing negative reactions, and reducing the native speaker's negative reaction toward nonnative speech when the latter is comprehensible.
References


Anderson-Hsieh, J., Johnson, R., & Koehler, K.


Stanford University Press.


Appendix A

**Questionnaire**

<table>
<thead>
<tr>
<th>Name</th>
<th>Sex</th>
<th>F</th>
<th>M</th>
<th>Native Language</th>
<th>Other lgs.</th>
</tr>
</thead>
</table>

In this experiment you will rate each voice on the characteristics listed below. Listen to each voice and begin to score as soon as you feel you can judge what type of person each speaker is. Give a rating for each trait. Do this in the same way as you do when listening to someone's voice for the first time on the telephone or radio. Rate in this way: ___: X:___

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Very Little</th>
<th>Very Much</th>
</tr>
</thead>
<tbody>
<tr>
<td>good looks</td>
<td></td>
<td>very much</td>
</tr>
<tr>
<td>leadership</td>
<td></td>
<td>very much</td>
</tr>
<tr>
<td>sense of humor</td>
<td></td>
<td>very much</td>
</tr>
<tr>
<td>intelligence</td>
<td></td>
<td>very much</td>
</tr>
<tr>
<td>honesty</td>
<td></td>
<td>very much</td>
</tr>
<tr>
<td>self-confidence</td>
<td></td>
<td>very much</td>
</tr>
<tr>
<td>height</td>
<td></td>
<td>very tall</td>
</tr>
<tr>
<td>friendliness</td>
<td></td>
<td>very much</td>
</tr>
<tr>
<td>generosity</td>
<td></td>
<td>very much</td>
</tr>
<tr>
<td>entertainingness</td>
<td></td>
<td>very much</td>
</tr>
<tr>
<td>nervousness</td>
<td></td>
<td>very much</td>
</tr>
<tr>
<td>good-heartedness</td>
<td></td>
<td>very much</td>
</tr>
<tr>
<td>reliability</td>
<td></td>
<td>very much</td>
</tr>
<tr>
<td>ambition</td>
<td></td>
<td>very much</td>
</tr>
<tr>
<td>stability</td>
<td></td>
<td>very much</td>
</tr>
<tr>
<td>good character</td>
<td></td>
<td>very much</td>
</tr>
<tr>
<td>likability</td>
<td></td>
<td>very much</td>
</tr>
<tr>
<td>wealth</td>
<td></td>
<td>very much</td>
</tr>
<tr>
<td>cleanliness</td>
<td></td>
<td>very much</td>
</tr>
<tr>
<td>prestige</td>
<td></td>
<td>very much</td>
</tr>
</tbody>
</table>
Table 1

**Simple Effects and Marginal Group Means**

<table>
<thead>
<tr>
<th>Error Type</th>
<th>Language Background</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E</td>
</tr>
<tr>
<td>Gram</td>
<td>73.35</td>
</tr>
<tr>
<td>Phon</td>
<td>73.41</td>
</tr>
<tr>
<td>Marg</td>
<td>73.38</td>
</tr>
</tbody>
</table>

**Key**
- Gram = grammatical error
- Phon = phonetic error
- Marg = marginal means
- E = English
- G = German
- S = Spanish
- A = Arabic
- M = marginal means
Figure 1. Language background and error type means for semantic differential rating

Key
E = English
G = German
S = Spanish
A = Arabic
Figure 2. Error type and language background means for semantic differential rating

Key
G = grammatical error
P = phonetic error
1. There were instances in which a respondent had not placed an "X" on a particular space for a particular item. These were considered missing data; consequently, scores on all 20 items on that questionnaire were excluded from the data set. Hence, varying sample sizes resulted for each dependent variable (error type) for each language background (categorical variable).