Language Learning 42:4, December 1992, pp. 529–555

The Relationship Between Native Speaker Judgments of Nonnative Pronunciation and Deviance in Segmentals, Prosody, and Syllable Structure

Janet Anderson-Hsieh
Iowa State University

Ruth Johnson
Florida State University

Kenneth Koehler
Iowa State University

This study investigated the relationship between experienced SPEAK Test raters' judgments of nonnative pronunciation and actual deviance in segmentals, prosody, and syllable structure. Sixty reading passage speech samples from SPEAK Test tapes of speakers from 11 language groups were rated impressionistically on pronunciation and later analyzed for deviance in segmentals, prosody, and syllable structure. The deviance found in each area of pronunciation was then correlated with the pronunciation ratings using Pearson correlations and multiple regression. An analysis of the 60 speakers showed that whereas deviance in segmentals,

We acknowledge the contributions made to this study by Phil Iverson, Lucy Hanson, Sandy Humphrey, Cindy Myers, and Janet Searls. We also acknowledge the careful review given to our paper by the two anonymous referees. Their comments were very helpful to us in improving and revising the paper.

Requests for reprints may be sent to the first author at Department of English, Iowa State University, 203 Ross Hall, Ames, Iowa 50011.

prosody, and syllable structure all showed a significant influence on the pronunciation ratings, the prosodic variable proved to have the strongest effect. When separate analyses were done on two language subgroups within the sample, prosody was always found to be significantly related to the global ratings, whereas this was not always true for the other variables investigated.

Although nonnative pronunciation is characterized by deviance in several areas, when native speaker judges evaluate nonnative pronunciation, they most often judge it in terms of the speaker's overall intelligibility, the irritability of the accent, or its acceptability (Johansson, 1978; Ludwig, 1982). In such judgments of pronunciation, the evaluators base the pronunciation score they assign on their overall impression, without attempting to count the various types of errors that occur.

An example of such judgments can be found in the pronunciation subtest of the SPEAK Test, which is used widely for evaluating the speaking proficiency of International Teaching Assistants at universities throughout the United States.[1] The criteria used for pronunciation in the SPEAK Test are based on considerations of intelligibility and acceptability, and the raters are instructed to judge errors mainly as they affect intelligibility. For example, a rater would select the lowest point on the scale when frequent errors are present that cause the speaker to be unintelligible.[2]

A question concerning the judgments of pronunciation from speech samples used in the SPEAK Test is whether native speaker judges react equally to deviance in all major areas of pronunciation or whether each area carries a different weight in influencing the scores that are assigned. The major areas of pronunciation are segmentals, prosody (supraseg-mentals), syllable structure, and voice quality. Deviance in segmentals involves errors in consonants and vowels, such as the substitution of one sound for another or the modification of a sound (Dickerson, 1975; Hecht & Mulford, 1982; Altenberg & Vago, 1983; Beebe, 1984; Flege & Hillenbrand, 1984; Nathan, Anderson, & Budsayamongkon, 1987). Errors in prosody involve deviations in patterns of stress and intonation as well as in timing, phrasing, and rhythm (Adams, 1979; Taylor, 1981; Grover, Jamieson, & Dobrovolsky, 1987; Flege & Bohn, 1989; Juffs, 1990; Shen, 1990). Syllable structure errors involve the addition of a segment or syllable, the deletion of a segment or syllable, or the reordering of segments in syllables, the most common types of errors being consonant deletion and vowel insertion (Tarone, 1980; Anderson, 1983; Broselow, 1983, 1984; Sato, 1984; Karimi, 1987). Voice quality refers to characteristics of pronunciation that affect entire utterances (Abercrombie, 1967; Laver, 1980). Voice quality deviance in nonnative speech can involve inappropriate posturing of the articulators, such as the tight-jawed posture and dentalized tongue body setting often found in Chinese ESL learners (Esling & Wong, 1983). However, generally, voice quality settings in nonnative speech have not been as well characterized in the second-language literature as have been other kinds of deviance.

The question of the relative importance of these different areas of nonnative pronunciation for intelligibility has been discussed, although the relative importance of prosody versus segmentals has been of particular interest. It has been argued that whereas both of these aspects of pronunciation are important in intelligibility, the most critical area is prosody (Hatch, 1983; Gilbert, 1984; Pennington & Richards, 1986; McNerney & Mendelsohn, 1987; Wong, 1987, Dickerson, 1989; Acton, Gilbert, & Wong, 1991). One of the major arguments presented is that prosody provides the framework for utterances and directs the listener's attention to information the speaker regards as important. However, it is necessary to note that this view is apparently based on informal observations rather than on empirical evidence from research studies. The question of whether prosody in fact carries more weight than do the other areas of pronunciation in native speakers' judgments is an empirical question, the answer to which can be found only through research studies that investigate native speakers' reactions to nonnative speech.

Yet a survey of the literature indicates that although numerous studies have investigated native speaker reactions to nonnative speech and several reviews of the literature have been written (Ludwig, 1982; Eisenstein, 1983; Ryan, 1983), most of the studies have not investigated the effects of different aspects of pronunciation on native speakers' reactions. Instead, they have investigated the relationship between other, often external variables and native speaker reactions. Examples of some of the questions investigated are the relationship between factors in the speakers' backgrounds and judgments of pronunciation accuracy (Suter, 1976; Purcell & Suter, 1980; Thompson, 1991), the comprehensibility of different types of accents (Smith & Bisazza, 1982), and the relationship between stereotypical reactions and the type of accent (Lambert, Hodson, Gardner, & Fillenbaum, 1960; Anisfeld, Bogo, & Lambert, 1962; Bradford, Farrar, & Bradford, 1974; Brennan, Ryan, & Dawson, 1975; Ryan & Carranza, 1975; Ryan, Carranza, & Moffie, 1977; Brennan & Brennan, 1981). Other studies investigating native speaker reactions to nonnative speech have investigated the effects of the subsystems of language (e.g., grammar and phonology) on native speakers' judgments (Politzer, 1978; Varonis & Gass, 1982), the effects of familiarity on comprehensibility (Gass & Varonis, 1984), and the effects of speaking rate on native speakers' comprehension of nonnative speech (Anderson-Hsieh & Koehler, 1988).

Whereas the studies cited above address important problems, they do not offer much information about the extent to which each area of pronunciation contributes to the native speakers' reactions. Although the Anderson-Hsieh and Koehler (1988) study, cited above, suggested that prosody may be more critical for comprehension than are segmentals, this was true only at faster rates of speech.

Only three studies have investigated the relative effects of prosody and segmentals on native speakers' reactions to pronunciation at normal rates of speech, and the results of these studies are not strongly conclusive. James (1976) reported a study in which he attempted to determine the relative effects of articulation and prosody on native speakers' perceptions of accentedness in the speech of English Canadian students learning French. Speech samples and oscillograms of the students were examined and three types of pronunciation were identified: good intonation with poor articulation, poor intonation and good articulation, and poor intonation and poor articulation. The averaged ratings of the native French listeners indicated that good intonation and poor articulation was more acceptable than was poor intonation with good articulation, supporting the primacy of prosody over articulation. However, it is important to note that statistical tests were not reported so that it is not known whether the differences found were significant.

Johansson (1978) also investigated the relative effects of prosody and segmentals on native speakers' judgments of pronunciation. In his study, British English judges reacted to passages of extended text and individual words and sentences read by Swedish ESL learners. He reasoned that a lower score on the extended text than on the word list and sentences for the same speakers would indicate less tolerance for prosodic errors, because he felt that rhythm and intonation are more apparent and important at the discourse level than they are at the level of isolated words and sentences. He found that the judges consistently rated the extended speech samples more severely than they did the word lists and sentences, suggesting that prosody weighed more heavily in the native speakers' judgments than did segmentals. Further evidence for less tolerance for prosodic deviance was found in a comparison between a speaker with good prosody and poor segmentals and another speaker with poor prosody and good segmentals, the ratings being higher for the former. These observations are consistent with the results of James' (1976) study (reported above) investigating the global judgments of learners of French as a second language. However, the results of this study cannot be considered strongly conclusive either, because statistical tests of significance were not reported.

Fayer and Krazinski (1987) investigated the reactions of native speakers of English and native speakers of Puerto Rican Spanish to the oral communication of Puerto Rican ESL learners. They were asked to make judgments of intelligibility, irritability, and annoyance in reaction to several aspects of the ESL learners' speech, including grammar, intonation, pronunciation (apparently articulation), hesitations, and lexical errors. It was found that whereas the Puerto Rican judges rated the speech samples more severely than did the native judges, both groups rated form more severely than they did intelligibility, and they found deviant pronunciation and hesitations to be more distracting than they did intonation. The lower importance given to intonation by the raters conflicts with the results of the studies by James (1976) and Johansson (1978) reported above. However, it is not clear from the article how the authors distinguished between pronunciation, which is normally thought to include intonation, and how the linguistically naïve judges in the study were instructed to make such a distinction.

More research is needed that investigates the relationship between native speaker judgments of pronunciation and deviance in segmentals and prosody. Whereas more evidence has been presented to support the primacy of prosody over segmentals in impressionistic ratings of nonnative pronunciation, the results of the studies supporting this were not strongly conclusive. It is also important to note that none of the studies cited above attempted to determine the relative contribution of syllable structure errors to impressionistic judgments of pronunciation and whether they play a role as important as do prosody and segmentals. Because numerous studies in interlanguage phonology have shown that many groups of ESL learners have difficulty with the syllable structure of English, (Tarone, 1980; Eckman, 1981, 1991; Anderson, 1983; Broselow, 1983, 1984; Sato, 1984; Hodne, 1985; Karimi, 1987; Weinberger, 1987), it should of be of interest to determine the effect such errors have on native speaker reactions to nonnative pronunciation.

The purpose of the present study is to investigate the relationship between deviance in three areas of pronunciation—prosody, segmentals, and syllable structure—and impressionistic judgments of pronunciation in samples of speech taken from the SPEAK Test. Voice quality is not considered in the study because of the problems inherent in reliably describing and quantifying voice quality features of nonnative speech. In this study, segmentals, prosody, and syllable structure are all predicted to be significantly correlated with the pronunciation ratings, although the strongest correlation is predicted for prosody.

METHOD

The method involved selecting nonnative speech samples from SPEAK Test tapes, rating the speech samples impressionistically on pronunciation, analyzing the samples for phonological errors, and analyzing the relationship between the errors and the pronunciation ratings through statistical procedures. Each of these steps will be more fully described below.

SPEECH SAMPLES

The speech samples that were used in the study were taken from the pool of SPEAK Test tapes available at Iowa State University. A total of 60 tapes were selected. To ensure that pronunciation—and not grammar or fluency—was being evaluated, a decision was made to use only the oral reading passage from the SPEAK test, and to control for content, only one form of the test, Form A, was used (see Appendix A for the text of the reading passage).

Only male speakers were selected. Also important in the selection of tapes was establishing a balance between high and low proficiency speakers, using the SPEAK comprehensibility score as the means of selection. Half of the speakers in the sample had comprehensibility scores above 200 and the other half had comprehensibility scores below 200, the scores ranging from 140 to 300. It was felt that a wide range of proficiencies would ensure greater variability in error frequencies, and this was needed to establish whether any relationship existed between the global pronunciation judgments and deviance in the areas of pronunciation investigated. The sample contained speakers representing a variety of language backgrounds. The linguistically heterogeneous sample used in the study was consistent with the way in which SPEAK Test tapes are normally rated. Speakers from many language backgrounds are rated at the same time instead of being rated in separate groups according to native language. The language groups represented in the sample were Arabic, Armenian, Chinese, Farsi, German, Greek, Indonesian, Korean, Serbo-Croatian, Spanish, and several language groups from the Indian subcontinent: Assamese, Hindi, Kannada, Malayam, Punjabi, and Tamil. Because the tapes were drawn from a pool of tapes containing very few speakers in some language groups, the sample does not contain an equal number of speakers in each language group.

It is also important to note that the language groups were not balanced with regard to language proficiency. For example, whereas in the Indian group, there was a greater proportion of speakers who had scored above 200 on comprehensibility than had scored below 200, the opposite was true for the Chinese and Korean groups. Table 1 presents the mean SPEAK comprehensibility scores and the number of speakers for all of the language groups represented in the sample. This reflects the realities of SPEAK Test ratings at Iowa State University, in the sense that the pattern of distribution of languages and language proficiencies in the sample is typical of nonnative graduate students at Iowa State University who take the SPEAK Test. However, the lack of balance in the number of speakers and the

Table 1
Mean SPEAK Comprehensibility Scores for Each Language Group in the Sample

Language Group	*n*	Mean SPEAK Comprehensibility Scores
Arabic	3	286.67
Armenian	1	230.00
Assamese	1	150.00
Chinese	21	173.33
Farsi	2	280.00
German	1	290.00
Greek	2	260.00
Hindi	3	260.00
Indonesian	1	220.00
Kannada	5	284.00
Korean	12	172.50
Malayam	2	250.00
Punjabi	1	220.00
Serbo-Croatian	1	270.00
Spanish	3	236.67
Tamil	1	240.00
Total	60	

speaking proficiencies within various language groups limits the generalizability of the study. Separate analyses for some language subgroups were therefore performed to examine the consistency of the results across language groups, although the analysis was limited by the relatively small numbers of speakers in the groups analyzed.

All 60 nonnative speech samples were randomly dubbed onto the same tape for the global ratings. A short segment of the same passage read by the nonnatives was read by a native speaker of American English and recorded, and then dubbed onto the tape in between the nonnative speech samples. These native speaker segments, which were not evaluated, served as a native speaker reference, their purpose being to reduce the influence that one nonnative speech sample might have on the next one being rated.

RATING OF PRONUNCIATION

The tape was rated by three experienced ESL teachers, all of whom had served as SPEAK Test raters, in a single sitting in approximately two hours. All raters heard the same tape.

Similar to the SPEAK Test pronunciation criteria, both intelligibility and acceptability were rated using the same scale. The lowest point on the scale represented heavily accented speech that was unintelligible. The midpoint represented accented but intelligible speech, and the highest point on the scale represented near-native speech. The rating scale departed from the SPEAK Test scale in that seven points, instead of four, were used. These were 0, 0.5, 1.0, 1.5, 2.0, 1.5, and 3.0. The scale was changed because a tendency had been observed on the SPEAK Test to avoid the extreme points on the four-point scale that is used, effectively reducing the scale to two points. For the purposes of the study, it was felt that finer distinctions in pronunciation were needed. The raters were encouraged to use the extreme points of the scale and the decimal points in between the whole numbers whenever appropriate.

After the samples had been rated, interrater reliability coefficients were computed between all possible pairs of raters for the three raters who evaluated the tapes. These correlations were found to be strong, ranging from 0.80 to 0.89. Having established that the interrater reliability was acceptable, the three scores given by the three raters for each speaker were averaged, and the mean score was used as the pronunciation variable. The group mean pronunciation score was 1.55, whereas the group mean score for the pronunciation subtest of the SPEAK Test for the same individuals, but rated by different raters, was 2.02. This difference in mean score was expected because of the changes made in the scale and the rating procedure. It is important to note, however, that the Pearson correlation between the two sets of scores was 0.83, indicating that although the absolute values of the scores were different between the two sets of ratings, the ranking of speakers according to their scores remained fairly constant. The mean pronunciation ratings and the native languages of all of the speakers investigated are presented in Appendix B.

ANALYSIS OF PRONUNCIATION DEVIANCE

Several steps were taken to determine pronunciation deviance in the speech samples: establishing native speaker norms for the reading passage, transcribing the speech samples phonetically, and finally, analyzing them for phonological errors. Each of these steps will be described more fully below.

Native Speaker Norms of Pronunciation. Speech samples of three male native speakers of American English reading the same passage read by the nonnatives were recorded on tape and used as a native speaker reference group. Any simplifications or changes from citation pronunciation identified in these native speech samples were not counted as errors when the speech samples of the nonnatives were analyzed. For example, the deletion of *d* in *and* was not counted as an error because such simplifications were common in the three native speech samples. Also, sources on connected speech phenomena (Wolfram & Johnson, 1982; Prator & Robinett, 1985) were consulted, and any additional native simplifications noted in these sources were not counted as errors either.

Transcribing the Speech Samples. The transcribers were two individuals whose phonetic transcription skills had been judged to be good by the first author, who teaches a course in phonetics and phonology in the graduate TESL program at Iowa State University. The basis for this judgment was the performance of the raters on phonetic transcription tasks while they were students in the course. Because of the well-known difficulty in achieving consistency among transcribers, only

Table 2
Examples of Consonant, Vowel, and Syllable Structure Errors from the Nonnative Speech Samples

Error Type	Word in Which Error Occurred	Error	Phonetic Transcription
Segmental			
Consonant			
Phonemic	winters	w—>v	[vɪntɚz]
Subphonemic	unnecessary	r—>ɾ	[ənɛsəsɛɾi]
Vowel			
Phonemic	put	ʊ—>u	[p^{h}ut]
Subphonemic	trapped	æ—>æ̂	[t^{h}ræ̂pt]
Syllable Structure			
Epenthesis: Consonant	tight	ø—>s	[t^{h}a^{ɪ}ts]
Epenthesis: Vowel	avoided	ø—>ə	[əvɔ^{ɪ}dədə]
Deletion: Consonant	cold	l—>ø	[ko^{ʊ}d]
Deletion: Vowel-Syllable	activity	ɪ—>ø	[ækt^{h}ɪvt^{h}i]
Metathesis	clothing	lo—>ol	[k^{h}ɔlðɪŋ]

those features that it was felt could be transcribed reliably were included in the phonetic transcription, which can be described as "moderately narrow". Any differences between the transcriptions of the two transcribers was resolved by the first author.

Analysis of Errors. The major error categories were errors in sound segments, syllable structure, and prosody. The segmental category consisted of both phonemic errors (sound substitutions) and subphonemic errors (modifications) in consonants and vowels. The syllable structure category consisted of vowel epenthesis (insertion) errors, consonant epenthesis errors, vowel/syllable deletion errors, consonant deletion errors, and metathesis (order reversal) errors. The errors in the phonetic transcripts were identified and counted by one of the authors, and checked for accuracy by another. Areas of dis-

Prosodic Criteria

1. Stress	The correct syllables are stressed in words and appropriate words are accented in tone groups.
2. Rhythm	Stressed syllables are sufficiently prominent and occur at fairly regular intervals. Unstressed syllables and function words are sufficiently reduced.
3. Intonation	Intonation contours are appropriate and pitch range is sufficiently wide.
4. Phrasing	Phrasing and pausing are appropriate, with pauses occurring at syntactic boundaries.
5. Overall Prosody	Overall, prosody is native-like.

Rating Scale

Least Native-like — 0 — 1 — 2 — 3 — Most Native-like

Figure 1. Criteria and scale for rating prosody

agreement were discussed and resolved. The error rates were calculated by dividing the number of errors in each category by the number of total possible occurrences of error. For example, to calculate the syllable structure error rate, the number of syllable structure errors was divided by the total number of syllables (199) in the reading passage. Table 2 presents examples of all of the error types investigated.

Prosody was rated impressionistically because of the difficulties of quantifying most kinds of prosodic deviance in terms of discrete errors. Adams (1979) had used impressionistic ratings in a study on nonnative rhythm, and had shown that they were reliable. The criteria used for the ratings were adapted from Adams (1979), Prator & Robinett (1985), and Wong (1987). Stress, rhythm, phrasing, intonation, and overall prosody were all rated after listening to each speech sample once. "Overall prosody" measured the raters' overall impression of stress, rhythm, phrasing, and intonation. A four-point scale, with midpoints between the whole numbers was used.

The speech samples were rated on prosody by the same three individuals who transcribed the tapes. After discussing the criteria and rating some speech samples for practice, the raters assigned scores independently except when the three scores were not within one point of each other on the rating scale. In such cases, the raters discussed their judgments with each other, and then rated the speech sample again. The interrater reliability was very high, the Pearson correlation coefficients between all possible pairs of raters being above 0.9. Also, because the correlations among all the prosodic subscores were also all above 0.9, only the overall prosodic score was used in the analysis. The decision to use only the overall prosodic score was supported by a subsequent factor analysis (see discussion under *Results*). The prosody scores for the three raters were averaged, and the mean score served as the prosodic variable. The criteria used for evaluating prosody are presented in Figure 1.

QUANTIFICATION AND ANALYSIS OF THE DATA

The dependent variable in the study was the pronunciation score. This was expressed positively in terms of degree of correctness on a four-point scale. The three independent variables selected for the analysis were the phonological variables: the segmental error rate, the prosodic score, and the syllable structure error rate. Similar to the pronunciation rating, the prosodic score was expressed positively in terms of degrees of correctness on a 4-point scale, and the segmental and syllable structure error rates were expressed negatively, in terms of percentages of error.

The selection of the independent variables for the analysis was supported by a factor analysis of the larger set of subscores based on the error categories presented in Table 2 and the prosodic features presented in Figure 1. For example, the prosodic subscores all had nearly the same loading on one factor, whereas the other variables showed only a small loading on that factor. As indicated above, this supported the decision to use only the overall prosodic score in the analysis. The two additional independent variables—the segmental and syllable structure error rates—which corresponded to other combinations of subscores were similarly supported by other factors. Also, regression models were fit to all possible subsets of the independent variables, and the three independent variables selected for the analysis—the prosodic score and the segmental and syllable structure error rates—were selected as the subset that maximized the adjusted R^2 value and minimized the Mallows Cp statistic (Rawlings, 1988, pp. 180–186).

RESULTS

The mean values, standard deviations, and minimum and maximum values for all the major variables investigated are presented in Table 3. The mean values for the prosodic score and the segmental and syllable structure error rates all indicate a substantial rate of error, and the minimum and maximum values indicate a sufficient range of values for each variable. It is important to note again that the mean values for prosody are based on ratings using a 4-point scale, whereas the mean values for segmentals and syllable structure are based on percentages of errors.

Table 4 presents the correlations among all the variables for the whole group (N=60). The correlations between the pronunciation score and the other variables are moderate to very strong and significant at the .0001 level. There is a strong positive correlation between the pronunciation score and the prosodic score (r=0.90). As expected, the pronunciation score had negative correlations with the syllable structure error rate (r=–0.77) and the segmental error rate (r=–0.67), which indi-

Table 3
Mean Scores, Standard Deviations, and Minimum and Maximum Values for Major Variables Investigated (N=60)

Variable	Mean	Standard Deviation	Minimum Value	Maximum Value
Pronunciation Ratings	1.55	0.74	0.33	3.00
Prosody Score	1.95	0.48	0.87	2.90
Segmental Error Rate	5.86	2.24	1.27	12.16
Syllable Structure Error Rate	6.44	5.92	0.00	23.12

cates that lower pronunciation ratings are associated with higher error rates for syllable structure and segmentals. It is also important to note that the correlations among all the independent variables are significant at least at the .01 level. The strongest correlation found among the independent variables was that between the prosody and syllable structure variables (r=–0.69). The results indicate that whereas all of the phonological variables are significantly correlated with the pronunciation ratings, the strongest relationship found was that between the prosodic variable and the pronunciation ratings, as predicted.

Because the independent variables were all significantly correlated with each other, a multiple regression analysis was performed regressing the pronunciation score on the three independent variables. Because the prosodic score is not reported on the same scale as were the segmental and syllable structure error rates, standardized regression analysis was used to facilitate comparisons among the estimated regression parameters. Standardized regression analysis is performed by regressing the standardized values of the pronunciation score on the standardized values of the three independent variables

Table 4
Pearson Correlation Coefficients for Pronunciation Rating, Prosody, and Segmental and Syllable Structure Error Rates (N=60)

Variable	Pronunciation Rating	Prosody	Segmental Errors	Syllable Structure Errors
Pronunciation	1.00[a]	0.90	–0.67	–0.74
Rating	0.0[b]	0.0001	0.0001	0.0001
Prosody Score		1.00	–0.55	–0.69
		0.0	–0.0001	0.0001
Segmental			1.00	0.32
Error Rate			0.0	0.01
Syllable Structure				1.00
Error Rate				0.0

The direction of the correlations is dependent upon the scale used for measuring the variables.
[a]Estimated correlation coefficient
[b]*P*-value for test of hypothesis of zero correlation

(Weisberg, 1980, p. 168). Each regression coefficient is a measure of partial association showing the average amount of change in the standardized pronunciation score for a one standard deviation increase in the corresponding independent variable. The standardized regression analysis revealed an R^2 of 0.89, indicating that 89% of the variation in the pronunciation ratings can be associated with changes in the independent variables. The corresponding F-ratio is 153.08 with (3, 56) degrees of freedom ($p<.0001$). Estimates of the standardized regression coefficients are presented in Figure 2. Each of the listed coefficients is significantly different from zero at the .0001 level, and because the variables in the regression were standardized, the intercept is zero. These results indicate that the prosody appears to have a greater influence on the pronunciation rating than do either the segmental or syllable structure error rates.

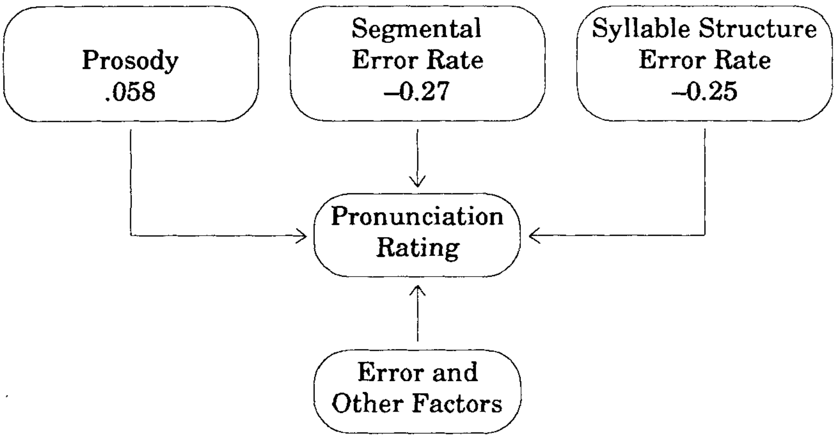

*Figure 2. Standardized multiple regression coefficients for prosod, segmental error rate, and syllable structure error rate (*N*=60)*

The question still remains, however, whether the same relative influence of prosody versus the other variables found for the whole group would remain constant across the language subgroups in the sample. However, because of the small number of speakers in many of the language groups, it was felt that a statistical analysis on each of the groups was not warranted. Thus, to further analyze the data, some of the individual language groups were combined to form major language groups consisting of speakers of more than one language from the same geographical areas. Although these procedures do not allow strong conclusions to be drawn, they at least provide some indication about whether any major patterns found in the analysis of the whole group remain unchanged when separate analyses are done on language subgroups in the sample.

The language subgroups analyzed were the East Asian Group, consisting of native speakers of Chinese and Korean and the Indian Subcontinent Group consisting of native speakers of Assamese, Hindi, Kannada, Malayam, Punjabi, and Tamil.[3] The number of speakers in the European and Middle Eastern

Table 5
Pearson Correlations Between the Pronunciation Ratings and the Phonological Variables for Two Language Subgroups in the Sample

Language Group	n	Prosody	Segmentals	Syllable Structure
East	33	0.72[a]	–0.57	–0.68
Asian		.0001[b]	.0006	.0001
Indian	13	0.89	–0.82	–0.25
Subcontinent		.0001	.0006	.4010

[a]Estimated correlation coefficient
[b]*P*-value for test of hypothesis of zero correlation

groups was too small to warrant an analysis of these groups. The one Indonesian speaker in the sample was also excluded from the analysis.

The Pearson correlations between the pronunciation ratings and the three independent variables for the two subgroups analyzed are presented in Table 5. In examining the correlations *within* each subgroup relative to each other, the correlations between the pronunciation ratings and the prosodic score were highly significant (p>.0001) and stronger than were the correlations between the pronunciation ratings and the other phonological variables in each group. Also, although the pronunciation ratings have negative correlations with both the segmental and syllable structure error rates for both subgroups as expected, the correlations between the pronunciation rating and the syllable structure error rate is not significant for the Indian subcontinent group. Similar patterns in partial association were observed in the estimated coefficients for separate regressions of the standardized pronunciation scores on the standardized independent variables for each group. Whereas the coefficient for the prosody score was significant for both groups (p>.02) the segmental and syllable structure error rates were not always significant.

When making comparisons *across* the two groups, the strength of the coefficients varied (see Table 5). The correlations between the pronunciation score and the prosodic score and the pronunciation score and the segmental error rate were significantly lower for the East Asian group (p>.01) than for the Indian subgroup. On the other hand, the correlation between the pronunciation score and the syllable structure error rate was significantly higher for the East Asian group than it was for the Indian subcontinent group (p>.02).

These results suggest, although they do not show conclusively, that whereas the strength of association between the pronunciation score and the phonological variables shows variation *across* groups, *within* each group, the prosodic score is highly significantly associated with the pronunciation score, and relatively speaking, this association is at least as high as is the association between the pronunciation score and any of the other independent variables, which are not always significant. However, as noted earlier, this analysis of language subgroups must be interpreted with caution because of the linguistic heterogeneity of the subgroups and the relatively small number of speakers and lack of language proficiency balance in the subgroups.

CONCLUSIONS AND IMPLICATIONS FOR FURTHER RESEARCH

This study has shown that the phonological variable that is most strongly associated with the pronunciation scores is the prosodic variable, regardless of language subgroup. These results on prosody agree with the results of the studies by James (1976) and Johansson (1978) cited above. On the other hand, the results suggest, although they do not show conclusively, that the effects of the segmental and syllable structure error rates on pronunciation scores *relative* to prosody may be more dependent on native language.

However, it is important to note, once again, the limitations of the study due to the numbers of speakers and the language proficiency balance in the language groups. More research is needed which investigates the same questions posed in this study using larger samples of speakers with a good balance of language proficiencies in each group. Such studies are needed to show more clearly the extent to which the relative influence exerted by the different areas of pronunciation on native speaker judgments of nonnative pronunciation may vary depending on the speakers' native languages.

In addition, research is needed that further investigates the specific ways in which prosody affects pronunciation judgments because, at this point in research, prosody appears to exert the strongest influence. An experimental approach using a speech synthesizer would allow for strict control and manipulation of prosodic variables, such as syllable duration and pitch range and direction. Using such an approach, one could investigate whether certain aspects of prosody are more critical than are other aspects in native speakers' judgments of pronunciation and the extent to which any differences in these areas of prosody can predict differences in pronunciation judgments.

In addition, the effects of voice quality features, such as tongue body settings and nasalization on global judgments of pronunciation need to be considered in future research, although work is first needed in reliably describing and quantifying the voice quality features of nonnative speech.

It is hoped that more research investigating the relationship between native speaker reactions to nonnative speech and phonological variables will follow. An understanding of the phonological factors that weigh the most heavily in native speaker reactions to nonnative speech should be helpful in establishing valid priorities for teaching pronunciation to second-language learners. The study reported here has made a small contribution toward that goal.

NOTES

[1]The SPEAK Test actually uses retired forms of the Test of Spoken English (TSE)—a test developed by the Educational Testing Service—that measures a speaker's capacity to communicate comprehensibly. The TSE consists of subtests on overall comprehensibility, pronunciation, grammar, and fluency. The test is administered individually through the use of a tape recorder, and is evaluated by two trained evaluators who respond to the recorded speech using a different four-point scale for each of the aspects of speech being evaluated. A third person evaluates the speech if the scores from the two raters do not fall within a prescribed range. The test has been shown to have good reliability and construct validity (Clark & Swinton, 1979, 1980).

[2]Acceptability is defined as the extent to which language deviates from native speaker norms (Ludwig, 1982). Although acceptability and intelligibility are often measured separately using different scales (Fayer & Krazinski, 1987), the TSE/SPEAK pronunciation subtest uses descriptors of both intelligibility and acceptability on the same rating scale, and as noted in the text, acceptability is rated primarily according to the extent to which errors, or deviations from native speaker norms, affect intelligibility.

[3]This grouping of languages according to East Asian and Indian Subcontinent regions is not consistent with the traditional classification of languages in which Korean is classified as Altaic and Chinese as Sino-Tibetan and in which the Indian languages are classified according to whether they are Indic or Dravidian. Nevertheless, many of the patterns of error among the speakers of different languages in the same subgroups were similar enough to justify this grouping. For example, syllable structure errors and r/l confusion were common among both Korean and Chinese speakers, whereas similar nonnative patterns of rhythm and intonation and errors in aspiration were common among the Indian Subcontinent group.

REFERENCES

Abercrombie, D. (1967). *Elements of general phonetics*. Chicago, Illinois: Aldine.

Acton, W. R., Gilbert, J., & Wong, R. (1991, March). *Integrating pronunciation into the second language curriculum*. Paper presented at the 25th Annual International TESOL Convention, New York, New York.

Adams, C. (1979). *English rhythm and the foreign language learner*. The Hague, Holland: Mouton.

Altenberg, E. P., & Vago, R. M. (1983). Theoretical implications of an error analysis of second language phonology production. *Language Learning*, *33*, 427–447.

Anderson, J. (1983). Syllable simplification in the speech of second language learners. *Interlanguage Studies Bulletin*, *3*, 4–36.

Anderson-Hsieh, J., & Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning, 38*, 561–613.

Anisfeld, M., Bogo, N., & Lambert, W. (1962). Evaluational reactions to accented English speech. *Journal of Abnormal and Social Psychology, 65*, 223–231.

Beebe, L. (1984). Myths about interlanguage phonology. In S. Eliasson (Ed.), *Theoretical issues in contrastive phonology* (pp. 51–62). Heidelberg: Groos.

Bradford, A., Farrar, D., & Bradford, B. (1974). Evaluation reactions of college students to dialect differences in the English of Mexican Americans. *Language and Speech, 17*, 255–270.

Brennan, E. M., & Brennan, J. S. (1981). Accent scaling and language attitudes: Reactions to Mexican American Speech. *Language and Speech, 24*, 207–221.

Brennan, E. M., Ryan, E. B., & Dawson, W. E. (1975). Scaling of apparent accentedness by magnitude estimation and sensory modality matching. *Journal of Psycholinguistic Research, 4*, 27–36.

Broselow, E. (1983). Non-obvious transfer: On predicting epenthesis errors. In S. M. Gass & L. Selinker (Eds.), *Language transfer in language learning* (pp. 269–280). Rowley, Massachusetts: Newbury House Publishers.

Broselow, E. (1984). An investigation of transfer in second language phonology. *International Review of Applied Linguistics in Language Teaching, 22*, 253–269.

Clark, J., & Swinton, S. (1979). *An exploration of speaking proficiency measures in the TOEFL context* (TOEFL Research Report No. 4). Princeton, New Jersey: Educational Testing Service.

Clark, J., & Swinton, S. (1980). *The Test of Spoken English as a measure of communicative ability in English-medium instructional settings* (TOEFL Research Report No. 7). Princeton, New Jersey: Educational Testing Service.

Dickerson, L. J. (1975). The learner's interlanguage as a system of variable rules. *TESOL Quarterly, 9*, 401–407.

Dickerson, W. (1989). *Stress in the stream of speech: The rhythm of spoken English: Teacher's manual*. Urbana, Illinois: University of Illinois Press.

Eckman, F. R. (1981). On the naturalness of interlanguage phonological rules. *Language Learning, 31*, 195–216.

Eckman, F. R. (1991). The structural conformity hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition, 13*, 23–41.

Eisenstein, M. (1983). Native reactions to nonnative speech: A review of empirical research. *Studies in Second Language Acquisition, 5*, 160–176.

Esling, J., & Wong, R. (1983). Voice quality settings and the teaching of pronunciation. *TESOL Quarterly, 17*, 89–94.

Fayer, J., & Krazinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning, 37*, 313–326.

Flege, J. E., & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustical Society of America, 76*, 708–721.

Flege, J. E., & Bohn, O. (1989). An instrumental study of vowel reduction and stress placement in Spanish-accented English. *Studies in Second Language Acquisition, 11*, 35–62.

Gass, S. M., & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning, 34*, 65–89.

Gilbert, J. (1984). Pronunciation and listening comprehension. In H. J. Morley (Ed.), *Current perspectives on pronunciation* (pp. 33–39). Washington, DC: TESOL.

Grover, C., Jamieson, D. G., & Dobrovolsky, M. B. (1987). Intonation in English, French, and German: Perception and production. *Language and Speech, 30*, 277–295.

Hatch, E. M. (1983). *Psycholinguistics: A second language perspective*. Rowley, Massachusetts: Newbury House Publishers.

Hecht, B. F., & Mulford, R. (1982). The acquisition of a second language phonology: Interaction of transfer and developmental factors. *Applied Psycholinguistics, 3*, 313–328.

Hodne, B. (1985). Yet another look at interlanguage phonology: The modification of English syllable structure by native speakers of Polish. *Language Learning, 35*, 405–422.

James, E. (1976). The acquisition of prosodic features using a speech visualizer. *International Review of Applied Linguistics in Language Teaching, 14*, 227–243.

Johansson, S. (1978). *Studies of error gravity: Native reactions to errors produced by Swedish learners of English*. Göteborg, Sweden: Acta Universitatis Gothoburgensis.

Juffs, A. (1990). Tone, syllable structure and interlanguage phonology: Chinese learners' stress errors. *International Review of Applied Linguistics in Language Teaching, 28*, 99–117.

Karimi, S. (1987). Farsi speakers and the initial consonant cluster in English. In G. Ioup & S. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 305–318). Cambridge, Massachusetts: Newbury House Publishers.

Lambert, W., Hodson, R., Gardner, R., & Fillenbaum (1960). Evaluational reactions to spoken language. *Journal of Abnormal and Social Psychology, 60*, 44–51.

Laver, J. (1980). *The phonetic description of voice quality*. Cambridge: Cambridge University Press.

Ludwig, J. (1982). Native speaker judgments of second language learners' efforts at communication: A review. *The Modern Language Journal, 66*, 274–283.

McNerney, M., & Mendelsohn, D. (1987). Putting suprasegmentals in their place. *TESL Talk*, *17*, 132–140.

Nathan, G., Anderson, W., & Budsayamongkon, B. (1987). On the acquisition of aspiration. In G. Ioup & S. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 204–212). Cambridge, Massachusetts: Newbury House Publishers.

Pennington, M., & Richards, J. (1986). Pronunciation revisited. *TESOL Quarterly*, *20*, 207–225.

Politzer, R. (1978). Errors of English speakers as perceived by German natives. *The Modern Language Journal*, 62, 253–261.

Prator, C. H. & Robinett, B. W. (1985). *Manual of American English pronunciation* (4th ed.). New York, New York: Holt, Rinehart and Winston.

Purcell, E., & Suter, R. (1980). Predictors of pronunciation accuracy: A reexamination. *Language Learning*, *30*, 271–287.

Rawlings, J. (1988). *Applied regression analysis: A research tool*. Pacific Grove, California: Wadsworth & Brooks/Cole.

Ryan, E. B. (1983). Social psychological mechanisms underlying native speaker evaluations of nonnative speech. *Studies in Second Language Acquisition*, *5*, 148–159.

Ryan, E. B., & Carranza, M. A. (1975). Evaluative reactions of adolescents toward speakers of standard English and Mexican-American accented English. *Journal of Personality and Social Psychology*, *31*, 855–863.

Ryan, E. B., Carranza, M. A., & Moffie, R. (1977). Reactions toward varying degrees of accentedness in the speech of Spanish-English bilinguals. *Language and Speech*, *20*, 267–273.

Sato, C. (1984). Phonological processes in second language acquisition: Another look at interlanguage syllable structure. *Language Learning*, *34*(4), 43–47.

Shen, X. (1990). Ability of learning the prosody of an intonational language by speakers of a tonal language: Chinese speakers learning French prosody. *International Review of Applied Linguistics in Language Teaching*, *28*, 119–134.

Smith, L., & Bisazza, J. (1982). The comprehensibility of three varieties of English for college students in seven countries. *Language Learning*, *32*, 259–269.

Suter, R. (1976). Predictors of pronunciation accuracy in second language learning. *Language Learning*, *26*, 233–253.

Tarone, E. (1980). Some influences on the syllable structure of interlanguage phonology. *International Review of Applied Linguistics in Language Teaching*, *18*, 139–152.

Taylor, D. (1981). Nonnative speakers and the rhythm of English. *International Review of Applied Linguistics in Language Teaching*, *19*, 221–226.

Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning*, *41*, 177–204.

Varonis, E. M., & Gass, S. M. (1982). The comprehensibility of nonnative speech. *Studies in Second Language Acquisition*, *4*, 114–136.

Weinberger, S. (1987). The influence of linguistic context on syllable simplification. In G. Ioup & S. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 401–417). Rowley, Massachusetts: Newbury House Publishers.

Weisberg, S. (1980). *Applied linear regression*. New York: John Wiley & Sons.

Wolfram, W., & Johnson, R. (1982). *Phonological analysis: Focus on American English*. Washington, DC: Center for Applied Linguistics.

Wong, R. (1987). *Teaching pronunciation: Focus on rhythm*. Englewood Cliffs, New Jersey: Prentice-Hall.

APPENDIX A: READING PASSAGE FROM FORM A OF THE SPEAK TEST

During cold winters, people must be extra careful to prevent excessive exposure to cold and serious loss of body heat. Layers of relatively light, loose clothing give better protection than one thick, heavy item. Between each layer, there is a film of trapped air, which when heated by the body, acts as excellent insulation. Tight clothing should be avoided, because it does not leave room for the trapped air. When people exercise or work hard, layered clothing becomes particularly important. As they move about, they may get overheated. If a person becomes too warm, layers of clothing can be removed during the active time, and put back on when the exercise is stopped. By wearing layers of clothing during activity, a person can avoid an unnecessary chill.

APPENDIX B

Table B–1
Native Languages and Global Pronunciation Scores of Speakers

Speaker	Score	Native Language	Speaker	Score	Native Language
1.	2.00	Indonesian	31.	1.17	Chinese
2.	2.33	Chinese	32.	2.83	Farsi
3.	1.17	Korean	33.	1.33	Chinese
4.	1.00	Punjabi	34.	1.00	Assamese
5.	2.83	Arabic	35.	3.00	Hindi
6.	2.83	Greek	36.	0.67	Chinese
7.	0.83	Chinese	37.	0.83	Korean
8.	1.67	Korean	38.	1.50	Spanish
9.	2.00	Chinese	39.	3.00	Arabic
10.	2.00	Kannada	40.	1.27	Korean
11.	1.67	Arabic	41.	0.67	Chinese
12.	1.17	Chinese	42.	1.50	Malayam
13.	1.00	Korean	43.	1.83	Chinese
14.	3.00	Farsi	44.	0.67	Chinese
15.	0.67	Chinese	45.	2.00	Hindi
16.	2.33	Kannada	46.	1.33	Chinese
17.	2.17	Kannada	47.	1.50	Hindi
18.	0.83	Korean	48.	0.50	Chinese
19.	1.50	Armenian	49.	1.17	Spanish
20.	1.17	Korean	50.	1.33	Chinese
21.	2.00	Spanish	51.	1.50	Korean
22.	1.17	Chinese	52.	2.50	German
23.	1.33	Korean	53.	2.17	Greek
24.	2.50	Serbo-Croatian	54.	0.67	Chinese
25.	2.33	Kannada	55.	0.50	Chinese
26.	0.33	Chinese	56.	0.83	Korean
27.	2.00	Korean	57.	2.33	Malayam
28.	0.67	Chinese	58.	2.33	Tamil
29.	1.17	Korean	59.	0.50	Chinese
30.	1.00	Chinese	60.	1.67	Kannada