SSLA, **20**, 1–16. Printed in the United States of America.

ACCENT, INTELLIGIBILITY, AND COMPREHENSIBILITY

Evidence from Four L1s

Tracey M. Derwing
University of Alberta

Murray J. Munro
Simon Fraser University

This study was designed to extend previous research on the relationships among intelligibility, perceived comprehensibility, and accentedness. Accent and comprehensibility ratings and transcriptions of accented speech from Cantonese, Japanese, Polish, and Spanish intermediate ESL students were obtained from 26 native English listeners. The listeners were also asked to identify the first language backgrounds of the same talkers and to provide information on their familiarity with the four accents used in this study. When the results of this study were compared with the Munro and Derwing (1995, *Language Learning*, *45*, 73–97) study of learners of high proficiency, speaker proficiency level did not appear to affect the quasi-independent relationships among intelligibility, perceived comprehensibility, and accentedness; however, the relative contributions of grammatical and phonemic errors and goodness of prosody differed somewhat. Ability to identify the speakers' first languages was influenced by familiarity.

We thank Ron Smyth for his assistance with the data analysis, Bruce Derwing for his comments on an earlier version of this paper, and Liz Karra and Wendy McFeely for their assistance with data collection. We also thank the anonymous reviewers for their helpful comments. We are very grateful to the staff and students at Grant MacEwan Community College, Edmonton, Alberta, for their contribution to this study. This research was partially funded by a President's Research Grant from Simon Fraser University and a SARAC grant from the Faculty of Education, University of Alberta.

Address correspondence to Tracey M. Derwing, Department of Educational Psychology, 6th Floor, Education North, University of Alberta, Edmonton, Alberta, T6G 2G5, Canada; e-mail: tracey.derwing@ualberta.ca.

© 1997 Cambridge University Press 0272-2631/97 $7.50 + .10

https://doi.org/10.1017/S0272263197001010 Published online by Cambridge University Press

Second language (L2) speakers vary considerably in terms of the extent to which native listeners judge their speech to be accented. In a number of studies listeners have been asked to assess degree of accent using rating scales (e.g., Flege, Munro, & MacKay, 1995). However, it is recognized that a crucial aspect of L2 speech is intelligibility. Nelson (1982) defined intelligibility as "the apprehension of the message in the sense intended by the speaker" (p. 63). Munro and Derwing (1995a, 1995b), who offered a very similar definition (the extent to which the native speaker understands the intended message), attempted to measure this dimension by asking native listeners to transcribe in standard orthography utterances produced by L2 speakers. This notion of intelligibility corresponds fairly closely to that of Smith (1992), which emphasizes word and utterance recognition. Despite the prevalent view that accent and intelligibility are inextricably bound, Munro and Derwing's (1995a, 1995b) findings indicated that a strong accent does not necessarily preclude fully intelligible speech (see also Flege, Takagi, & Mann, 1995; Munro, Flege, & MacKay, 1996). Given that accent and intelligibility are related but partially independent dimensions, it seems reasonable to explore the nature of the relationship between the two. A third factor that is closely related to accent and intelligibility is comprehensibility, or native speakers' *perception* of intelligibility. Comprehensibility, as defined here, refers to judgments on a rating scale of how difficult or easy an utterance is to understand. Note that the term *comprehensibility* has been used in many different ways; Gass and Varonis (1984), for instance, used it in the sense of intelligibility as defined here. Smith (1992) regards it as being at a higher level of understanding than intelligibility. Samar and Metz (1988), in a study of the intelligibility of deaf speech, observed that transcription or "write down assessment" provided a "more accurate alternative for speech intelligibility assessment" (p. 307) than rating scale judgments. In this study we will examine the interrelationship of accent, intelligibility, and perceived comprehensibility for ESL speakers from four first language backgrounds: Cantonese, Japanese, Polish, and Spanish.

Although it has been claimed that puberty is a "turning point" for pronunciation (Patkowski, 1990), the evidence now suggests that this is not the case. Long (1990) argued that even individuals who began to learn their L2 at the age of 6 may exhibit accented speech. Recent experimental work involving large subject samples suggests that some individuals who learned their L2 in very early childhood can have detectable accents as adults (Flege, Munro, & MacKay, 1995). The inevitability of a foreign accent for most L2 learners has not discouraged activity in the area of pronunciation pedagogy; indeed, there appears to be a revival of interest in the teaching of pronunciation that had diminished somewhat with the demise of audiolingualism (Morley, 1991). Although the value of a "comfortable" intelligibility over perfect accuracy has been stressed by researchers for decades (Abercrombie, 1949; Morley, 1994), in recent years accent reduction programs that promise to eliminate foreign accented speech (e.g., Catran, 1985) have become very common. The renewed interest in accent may also have been stimulated by legal cases in which accent has been cited as a deciding factor in certain employment disputes (Sato, 1991).

https://doi.org/10.1017/S0272263197001010 Published online by Cambridge University Press

Factors Influencing Accent

Purcell and Suter (1980) carried out one of the most extensive studies on factors that influence L2 pronunciation in adults. On the basis of an earlier study by Suter (1976), Purcell and Suter identified four significant predictors of accented speech: first language, aptitude for oral mimicry, length of time in the L2 environment, and strength of concern for pronunciation accuracy. They concluded that a language instructor can have very little impact on students' accuracy, because the strength of concern for accuracy is the only significant predictor over which the instructor can have any influence. Thompson (1991), in a study of Russian speakers of English, found that the factors that best predicted native speakers' perception of accent were age of arrival, gender, self-ratings of oral mimicry skills, and overall proficiency in the L2. Flege, Munro, and MacKay (1995) found that, in addition to age of arrival, the factors of length of residence, the speaker's gender, and relative use of the first and second languages affected degree of perceived accent.

Factors Affecting Comprehension of Accented Speech

Varonis and Gass (1982) undertook a study to determine which factors affect listening comprehension in native speakers of English exposed to L2 accents. They concluded that grammar and pronunciation interact to influence overall intelligibility. In a later study, Gass and Varonis (1984) examined the effects of familiarity on NS comprehension of accented speech. Transcription tasks indicated that familiarity with topic, NNS speech, a particular accent, and a particular speaker all had an effect on intelligibility. Wingstedt and Schulman (1984) also found that familiarity with a particular accent facilitates comprehension.

In an innovative qualitative study that compared lectures originally produced by a Mandarin speaker and a native English speaker, both of which were delivered by a native English speaker to university student listeners, Tyler (1992) found that a number of discourse factors affected the overall comprehension of the lecture content, apart from pronunciation. Although the speaker in her study spoke with no L2 accent, lexical discourse markers, issues of lexical specificity, and syntactic relationships all affected the listeners' ability to follow the L2 speaker's discourse. The current study does not directly address discourse factors, but clearly they can play a significant role in overall intelligibility.

In a comparison of NNS errors in prosody, segments, and syllable structure with accent comprehensibility ratings, Anderson-Hsieh, Johnson, and Koehler (1992) found that prosody was rated as affecting accentedness and perceived comprehensibility to a greater extent than the other factors. Unfortunately, these authors did not actually measure intelligibility. Tajima, Port, and Dalby (1994) manipulated temporal variables in NS and NNS speech samples such that the English productions of a speaker of Taiwanese and Mandarin resembled NS patterns. Similarly, a native speaker's utterances were altered to reflect the temporal patterns of the Chinese NNS. The adjustment to the NNS's speech had a beneficial effect on intelligibility,

https://doi.org/10.1017/S0272263197001010 Published online by Cambridge University Press

whereas the change to the NS's utterances resulted in a reduction in overall intelligibility.

Scalar judgments of accent have been used in a number of studies (Flege, Munro, & MacKay, 1995; Major, 1987). Munro and Derwing (1995a) compared transcription scores (intelligibility) as well as accent and perceived comprehensibility ratings with phonetic, phonemic, and grammatical errors, and goodness of intonation ratings in Mandarin speakers' English productions. They found that for the majority of their listeners, accent ratings were significantly correlated with all the error types and goodness of intonation ratings. When comprehensibility ratings were compared with the linguistic measures, goodness of intonation was correlated for the majority of listeners, but phonetic errors dropped in significance. Finally, when intelligibility scores were compared with the measures, phonetic errors were not correlated for any of the listeners, whereas phonemic and grammatical errors and goodness of intonation ratings were correlated for only a minority of listeners. Interestingly, although accentedness, perceived comprehensibility, and intelligibility were interrelated, it was clear that they were considered separate dimensions. In other words, a strong foreign accent does not necessarily interfere with intelligibility, although NSs may require extra processing time to understand NNS speech, which may lead to lower perceived comprehensibility ratings (Munro & Derwing, 1995b).

The present study was designed to explore further the relationships among intelligibility, perceived comprehensibility, and accentedness (with a focus on grammatical errors, phonemic errors, nativeness of prosody, and speaking rate) that were revealed in Munro and Derwing (1995a, 1995b). We extend the previous work by examining listeners' reactions to speech from a variety of language backgrounds. Familiarity with speakers from particular language backgrounds was also investigated, as was native-speaking listeners' ability to recognize the four accents used in the study. We also wanted to determine whether the relationships observed in the speech of high-proficiency speakers would be evident in the speech of intermediate NNSs. To address these issues, we asked NSs to evaluate and transcribe the speech of Cantonese, Japanese, Polish, and Spanish intermediate learners of English.

METHODOLOGY

Participants

The 48 speakers were ESL students, enrolled at a community college academic English preparation program in Edmonton, Alberta (Canada). They were of intermediate to high intermediate proficiency according to the placement test used by the college; all were well educated in their own countries, and most planned to enter English language postsecondary education programs. The 12 students who had taken a TOEFL test achieved scores ranging from 400 to 565, with a mean of 479. (Only one student exceeded 540; the minimum TOEFL score used in Munro and Derwing [1995a] was 550.) The students who had taken a TOEFL test represented the upper range of proficiency in this sample. The speakers ranged in age from 19 to 64 years; the mean age was 29.1 years. There were 12 speakers from each of four language

https://doi.org/10.1017/S0272263197001010 Published online by Cambridge University Press

backgrounds: Cantonese, Japanese, Polish, and Spanish. It was not possible to balance the groups equally for gender (18 males and 30 females).

The 26 native English listeners, 6 males and 20 females, were enrolled in undergraduate university courses. Their mean age was 24.7 years with a range of 18–48 years. They were all born and raised in Canada. Most had studied another language, but 16 of the 26 reported that they could not speak any language fluently other than English. Ten participants reported having a good knowledge of at least one other language, usually French.

Procedure

The NNSs were presented with a series of cartoons that depicted an amusing story about two men who go on a hunting trip but who end up taking photographs of deer instead of shooting them. Before describing the story in their own words, the speakers were asked to examine the pictures and were given an opportunity to ask one of the researchers for clarification about the drawings or for any vocabulary they might have needed (several people asked for the words *deer* and *rifle*). Once the speakers were comfortable with the story, they narrated it onto audiotape. The recordings were made in a language lab using a Sony Console LLC-9000 system. So that the speech samples could be randomly presented to listeners for ratings, they were digitally recorded on a Macintosh computer (22 kHz with 16-bit resolution). Rather than arbitrarily controlling duration or number of words, we opted to select full phrases (taken from the beginning of each narrative) to keep the utterances as natural as possible. The samples varied in length from 6 to 19 words, with a mean of 12.4 words. Each sample was between 4.5 and 10.5 seconds long, with a mean of 7.0 seconds. (A subsequent test revealed that length had no effect on the results.) Two audiotapes were then made, each with a different random ordering of utterances.

The listeners were randomly assigned to one of two conditions, which differed only in terms of task order and one open-ended question. In the first condition, the participants immediately transcribed (in standard orthography) each speech sample. They then made a scalar judgment of accentedness (1 = no accent, 9 = extremely strong accent). Once the listeners had heard all 48 stimuli, they were given a break of about 5 minutes. They were then asked to listen to the samples again (in a different order) and make judgments of comprehensibility (1 = extremely easy to understand, 9 = extremely difficult or impossible to understand) after hearing each one. At the same time they were asked to circle which accent they thought the speaker had—Cantonese, Japanese, Polish, or Spanish. Finally, the listeners assigned to condition 1 were asked to comment on their perceptions of what makes some people more difficult to understand than others. The participants assigned to the second condition heard a different randomization of language samples; furthermore, they were asked to make comprehensibility judgments on the same round as the transcriptions. On the second round, they were asked to judge accentedness and identify individual accents. The participants also answered an open-ended question, this time regarding their perceptions of which aspects of an accent are most noticeable.

https://doi.org/10.1017/S0272263197001010 Published online by Cambridge University Press

All listeners completed a language background questionnaire at the outset of the experiment in which they indicated their degree of familiarity with languages other than English. They also indicated on a scale of 1–5 (where 1 = very limited exchanges and 5 = everyday/extended conversations) the amount of contact they had had with people who speak with Cantonese, Japanese, Polish, and Spanish accents specifically and with people who speak with any other foreign accent generally.

The recorded utterances were examined for the following accent features: grammatical errors, phonemic errors, nativeness of prosody, and speaking rate. Grammatical errors included the following: inappropriate verb tense, incorrect use of plurals, incorrect choice of prepositions, and incorrect pronoun assignment. The number of errors per utterance was determined. Phonemic errors were also counted (the deletion or insertion of a segment or the substitution of a segment that was clearly interpretable as an English phoneme other than the correct one). The global prosodic goodness of each utterance was evaluated on a 9-point scale. One of the difficulties associated with assigning ratings of this type is that judges may find it difficult to assess prosody independently of segmental factors. For this reason we used a low-pass filtering technique that removed most of the segmental information from the signal, while leaving rhythmic and intonational features largely intact (Munro, 1995; Van Els & De Bot, 1987). Utterances produced by female speakers were digitally filtered at 350 Hz using Signalyze software on a Power Macintosh computer; those produced by males were filtered at 250 Hz. The experimenters listened to the filtered samples twice. On the second pass they rated the prosody of each sample (1 = perfectly nativelike, 9 = extremely accented), while referring to a transcript of the utterance being evaluated. The scores were averaged to give a prosodic goodness rating. Speaking rate was measured in syllables per second.

RESULTS

Accent and Perceived Comprehensibility Judgments

The mean rating data for the two participant groups were compared to determine whether there was an effect of order of tasks. Because independent *t*-tests revealed that there were no significant differences between accent ratings, $t(1, 24) = 0.34$, $p > .05$, two-tailed, or between perceived comprehensibility ratings, $t(1, 24) = 0.57$, $p > .05$, two-tailed, the groups can be considered equivalent.

Reliability coefficients (standardized alpha) of 0.95 and 0.94 were calculated for the accent and perceived comprehensibility ratings, respectively. When these values were converted to Pearson correlations (as recommended in Hatch & Lazaraton, 1991), they both had a value of $r = 0.74$, $p < .01$. It appears then that, just as in the Munro and Derwing (1995a) study, there was a notable degree of interrater agreement in both sets of ratings, especially given that the listeners were not provided with training in the rating procedures. Hatch and Lazaraton (1991, p. 534) observed that under rating conditions such as these a Pearson correlation value in this range is an expected result.

https://doi.org/10.1017/S0272263197001010 Published online by Cambridge University Press

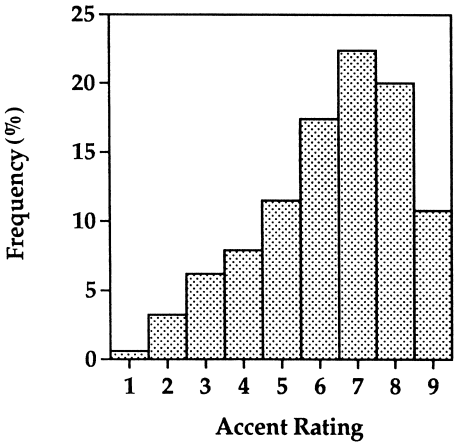


Figure 1. The distribution of accent ratings assigned to 48 speakers by 26 listeners using a 9-point scale, where 1 = no accent and 9 = extremely strong accent.

Each of the 26 listeners assigned 48 accent ratings and 48 perceived comprehensibility ratings (i.e., one pair of ratings for each L2 speaker). For each listener a Pearson correlation coefficient for the 48 pairs was computed. Although most participants showed a significant correlation ($p < .05$), the relationship in general was quite weak: The mean of the correlations was 0.45, and the *R*-squared values ranged from 0.01 to 0.52. As can be seen in Figure 1, the accent ratings were distributed across the scale, but the majority of ratings fell between 6 and 8. The comprehensibility ratings, by comparison, were less harsh. Figure 2 indicates that the majority of the utterances were assigned a rating of 5 or less and that the modal response was 3.

Transcription and Intelligibility Scores

Each of the listeners' 1,248 orthographic transcriptions was coded for exact word matches against a transcription of the recorded utterances. (All the recorded utterances had originally been transcribed by one of the experimenters and checked for accuracy by a research assistant. Any discrepancies were checked again by both parties.) For each listener transcription, the percentage of words exactly matching the original transcription was calculated. An overall transcription score was also determined for each of the 48 utterances by taking the mean of the 26 listeners' exact word match scores for the utterance. The mean transcription scores by utterance ranged from 55% to 99%. The mean individual listener scores ranged from 68% to 92%.

Errors were also categorized as trivial or nontrivial. Regularizations such as correcting minor grammatical errors (e.g., writing *two men* instead of *two mans*) and omitting a repetition of a word (writing *the car* instead of *the, the car*) were considered

https://doi.org/10.1017/S0272263197001010 Published online by Cambridge University Press

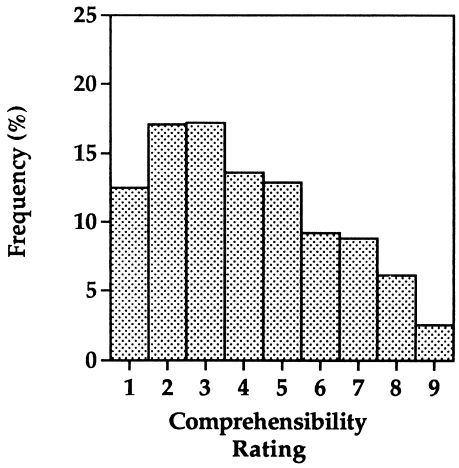


Figure 2. The distribution of comprehensibility ratings assigned to 48 speakers by 26 listeners using a 9-point scale, where 1 = extremely easy to understand and 9 = extremely difficult or impossible to understand.

to be trivial. Other word omissions and substitutions of words that did not appear in the stimuli were defined as nontrivial errors. The nontrivial errors were totalled across listeners for each utterance and a mean error score for each utterance was then calculated. These scores ranged from 0.0 to 6.31 per utterance. Trivial errors, although coded, were ignored in defining the intelligibility of an utterance. The number of nontrivial errors determined the intelligibility score. This procedure was undertaken in order to compare the results with the more rigourous exact word match technique, which slightly penalizes listeners for transcription errors that are unrelated to intelligibility.

Mean transcription scores across the stimuli were calculated by L1: Cantonese, 85%; Japanese, 84%; Spanish, 82%; and Polish, 81%. The listeners' mean nontrivial errors were also calculated for each L1: Spanish, 1.61; Polish, 1.27; Cantonese, 1.13; and Japanese, 1.06. When the intelligibility measure was used, ignoring trivial errors, the mean scores were as follows: Cantonese, 91%; Japanese, 89%; Polish, 89%; and Spanish, 89%. As can be seen in Figure 3, the distribution of transcription scores is highly skewed; the majority of transcriptions were at least 80% correct. A full 327 transcriptions, or 26%, were transcribed perfectly. When trivial errors were excluded from the analysis of intelligibility, fully 60% of the transcriptions were completely accurate, as can be seen in Figure 4.

We used Pearson correlations to determine whether number of words or utterance duration (measured in seconds) may have had an effect on intelligibility. Neither was significantly correlated with the mean intelligibility scores. We also conducted a one-way ANOVA with L1 as the between-subjects variable to see

https://doi.org/10.1017/S0272263197001010 Published online by Cambridge University Press

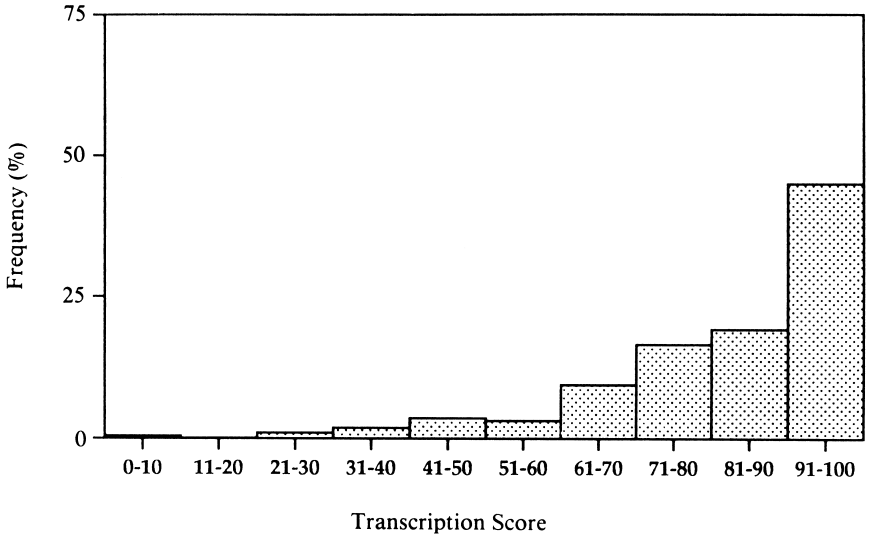


Figure 3. The distribution of transcription scores (based on exact word-matches) for the full stimulus set.

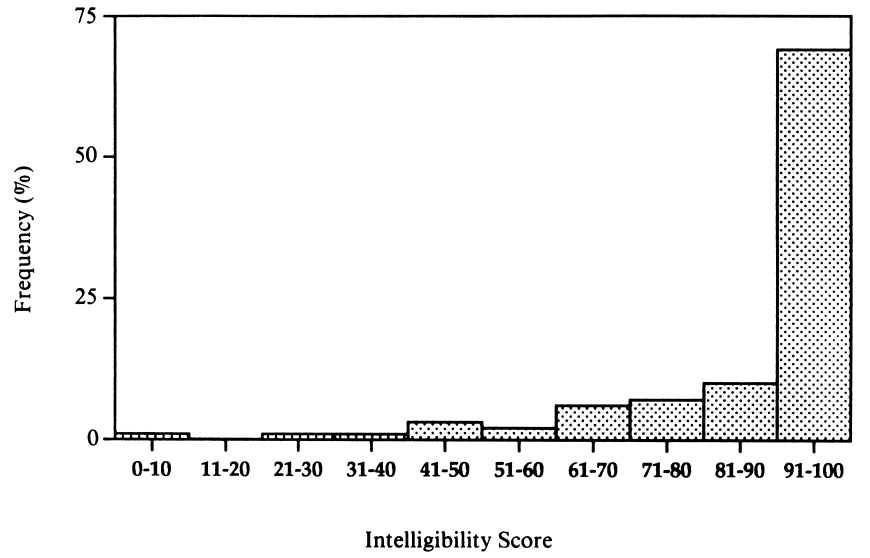


Figure 4. The distribution of intelligibility ratings (based on the nontrivial errors only) for the full stimulus set.

https://doi.org/10.1017/S0272263197001010 Published online by Cambridge University Press

Table 1. Number of listeners' significant correlations of accent features with task scores ($N = 26$)

Accent Feature	Accent Rating	Comprehensibility Rating	Intelligibility Score
Grammar score	13 (50%)	14 (54%)	4 (15%)
Phonemic score	4 (15%)	4 (15%)	2 (8%)
Prosodic score	7 (27%)	9 (35%)	2 (8%)
Speaking rate	6 (23%)	10 (38%)	2 (8%)

whether any of the four language groups' utterances differed from the others in length and duration. There were no significant differences between L1 groups for either number of words, $F(3, 44) = 1.70$, $p = .18$, or duration, $F(3, 44) = 1.58$, $p = .21$. Another one-way ANOVA was performed to determine whether the gender of the speakers affected the transcription scores. No significant differences based on gender were identified, $F(1, 46) = 0.97$, $p = .33$.

Cross-Task Comparisons

Pearson correlations between accent ratings and transcription scores were carried out for each listener. There were 20 significant correlations out of a possible 26, $p < .05$. The mean of the significant correlations, computed using a Fisher Z transformation and then converted back to Pearson r (Hatch & Lazaraton, 1991) was –0.46, indicating, as expected, that transcription scores tended to increase as the degree of accent decreased. When comparisons were made of the perceived comprehensibility ratings and transcription scores, all 26 correlations were significant at the $p < .05$ level. The mean of the Pearson r values (again computed using a Z transformation) was 0.51.

Parallel comparisons were made between accent ratings and the intelligibility scores for each listener. Very small differences were found between the exact word transcription and intelligibility scores overall. Pearson correlations between accent ratings and intelligibility indicated that there were 21 significant correlations, $p < .05$. The mean of the significant correlations transformed with a Fisher Z score was 0.41. When comparisons were made of the perceived comprehensibility ratings and intelligibility scores, 24 of the correlations were significant at the $p < .05$ level. The mean of the significant Pearson r values (Z-transformed) was 0.54.

Comparison of Accent Features with Task Scores

Correlations were computed between accent features and the accent judgments, comprehensibility judgments, and intelligibility scores for each individual. The number of listeners who had significant correlations is shown in Table 1.

Listeners' responses to the open-ended questions regarding which factors they considered to be important when judging accentedness and comprehensibility were quite varied. We classified the responses into eight categories, as can be seen in Table 2.

https://doi.org/10.1017/S0272263197001010 Published online by Cambridge University Press

Table 2. NSs' perceptions of factors contributing to accent and comprehensibility

Response Category	Causes of Accent ($n = 13$)	Factors Affecting Comprehensibility ($n = 13$)
Segmentals	92%	0%
Grammar	46%	38%
Rate	15%	38%
Prosodic features	23%	23%
Fluency	8%	15%
Enunciation (mumbling)	38%	46%
Speaking volume	0%	8%
Vocabulary	8%	0%

Language Identification Task

The listeners' ability to recognize the four accents under study was measured by means of a forced-choice task in which the four language names were supplied. The overall mean number of correct identifications across all listeners was 51.5%. The participants found Cantonese the easiest language to identify (62.5%) followed by Spanish (52.9%), Polish (49.6%), and Japanese (41%). Figure 5 indicates the incorrect choices made as well as the accurate identifications. The speakers of the two Asian languages were most often mistaken for each other, as were the speakers of the two Indo-European languages. A multiple regression was carried out to determine whether language, familiarity with a given language, or both predicted successful identification of samples, $F(2,\ 101) = 2.67,\ p = .07$. Although the F was marginal, when the two factors were examined, language was not a significant predictor but familiarity was, $p = .04$. In other words, familiarity with a language predicted successful identification; no one language was intrinsically more difficult or easier to identify. A significant Pearson correlation was found between language recognition scores and intelligibility scores, $r = 0.41, p < .05$, indicating that recognition of a first language speaker's background was moderately associated with better understanding of the speaker's productions.

DISCUSSION

The relationship among accent ratings, perceived comprehensibility ratings, and intelligibility scores, whether measured by exact word match or nontrivial errors for speakers of Cantonese, Japanese, Polish, and Spanish, closely replicate Munro and Derwing's (1995a) findings for Mandarin. In each instance, accent ratings are harsher than perceived comprehensibility ratings, which in turn are harsher than actual intelligibility scores. Results of the correlational analyses indicate that the three dimensions are clearly related but are not equivalent. The results provide additional evidence in support of the claim that, although some features of accent may be highly salient, they do not necessarily interfere with intelligibility. A clear implication of this finding is the need to disassociate accent ratings and intelligibility

https://doi.org/10.1017/S0272263197001010 Published online by Cambridge University Press

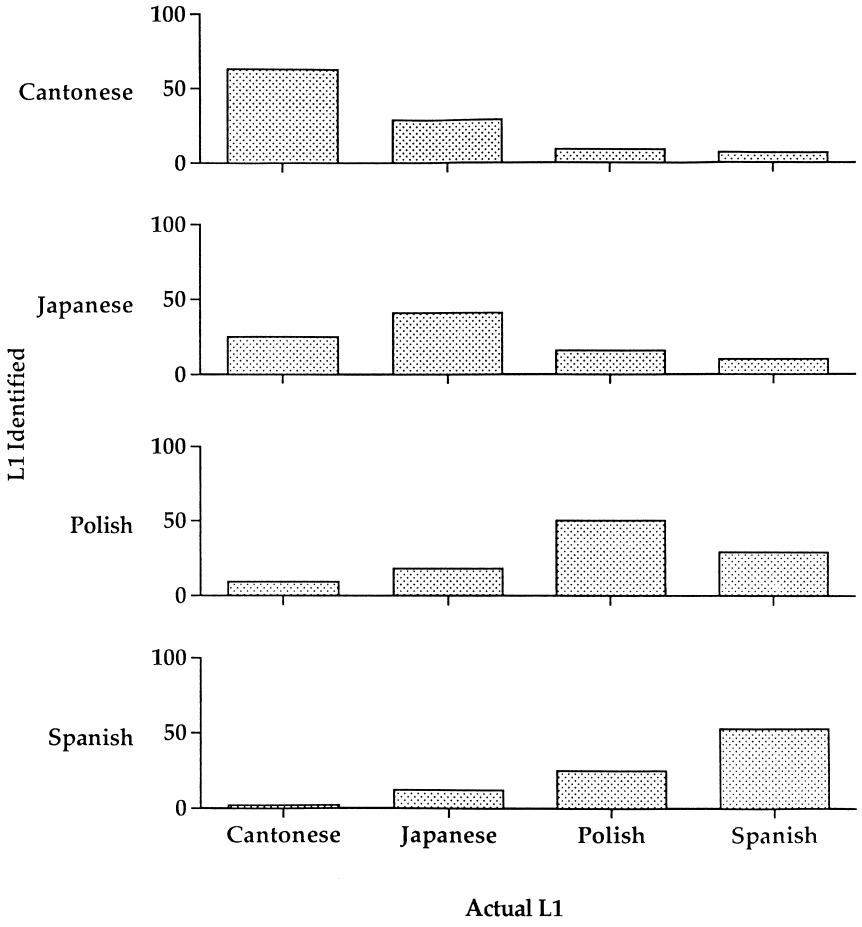


Figure 5. L1 confusions on the language identification task.

in language assessment instruments, which often confound the two dimensions. The discrepancy between perceived comprehensibility ratings and transcription success suggests that some accented but fully intelligible utterances may require additional effort or processing time (cf. Munro & Derwing, 1995b), which leads listeners to rate them as difficult to understand.

The individual contributions of four accent features—grammatical and phonemic errors, prosody, and speaking rate—were assessed in relation to accent and comprehensibility judgments and intelligibility scores. Subphonemic errors were not recorded because of their lack of effect on intelligibility in the Mandarin study (Munro & Derwing, 1995a). As in the Mandarin study, each of the accent features was significantly correlated with intelligibility for a small minority of listeners; this may be partially attributed to the limited variability in the intelligibility scores, which were generally very high. There were striking differences between the findings here and

https://doi.org/10.1017/S0272263197001010 Published online by Cambridge University Press

those of the Mandarin study, however, in the correlation patterns between accent features and the ratings data for accent and perceived comprehensibility. The majority of listeners (over 70%) in the Mandarin study showed significant correlations between their accent ratings and the accent features of phonemic errors, grammatical errors, and nativeness of prosody; there were far fewer significant correlations in the current study. One possible explanation for this difference has to do with the percentage of errors in the data. When grammar and phonemic errors are compared across the two studies, for instance, the number of errors per utterance is 0.6 and 0.9, respectively, in the Mandarin study, whereas in the current study there are 2.52 and 2.60 grammar and phonemic errors, respectively, per utterance. Presumably the larger number of errors in the current study may have led to competing accent features and lower correlations overall. Interestingly, even though phonemic errors were correlated with accent and perceived comprehensibility ratings for only 15% of the listeners, the participants believed that segmental deviations are the primary source of a foreign accent. Twelve of the 13 listeners who responded to the open-ended question on the causes of accent cited individual sounds, but none of the 13 listeners asked to comment on factors affecting comprehensibility identified segmental errors. Clearly, the errors are perceptually salient but have a limited effect on comprehensibility. The fact that the number of grammar errors correlated with perceived comprehensibility rankings for just over half the listeners is mirrored in the responses to the question regarding comprehensibility: Half of the individuals questioned identified grammar as a major factor in determining relative ease or difficulty of understanding, despite the seemingly minor nature of the errors. There are at least two possible explanations for the degree of correlation between grammar errors and perception of comprehensibility. First, it may be the case that grammatical errors are perceived as more problematic when they occur in conjunction with other types of errors (cf. Varonis & Gass, 1982); second, there may be a scapegoat effect resulting from the high saliency of the grammatical errors as opposed to their gravity.

Prosody had an effect on considerably fewer listeners in this study than in Munro and Derwing (1995a), although for a third of them goodness of prosody was correlated with perceived comprehensibility. There are several possible explanations for the difference observed here. First, the Mandarin speakers in the previous study were of a much higher proficiency level than the L2 speakers in the current study. Not only did the speakers here make considerably more grammatical and phonemic errors, as already noted, but the experimenters noted other nonnative-like properties in the stimuli that could not be readily quantified. These included unusual voice quality, inappropriate lexical choice, and various discourse factors. Moreover, the prosody ratings here are more rigorous in that they are based on filtered speech, rather than on unaltered stimuli, to eliminate the possible contaminating effects of other aspects of accent. Clearly, more studies of prosodic variables, such as Wennerstrom's (1994) investigation of ESL learners' intonation patterns in discourse, are needed.

The case of speaking rate is somewhat different from the other variables. Rate was negatively correlated with comprehensibility for 10 of the listeners; in other

https://doi.org/10.1017/S0272263197001010 Published online by Cambridge University Press

words, they felt that it was more difficult to understand faster speech, although rate did not interfere at all with intelligibility (the two significant correlations between rate and intelligibility were positive). Five of the 13 individuals who were asked to comment on factors that interfered with their comprehension complained that some of the speakers spoke too fast. This is reminiscent of Dahl (1981), who asked NNSs to listen to instructions read by NSs. When the NNS listeners reported that a given message was difficult to understand, they usually said that it was delivered at too fast a rate. Actually, those messages were no faster than the ones that were judged as being easy to understand. In Anderson-Hsieh and Koehler's (1988) study, NS listeners reported that heavily accented speech was delivered at a faster rate than less accented samples, although in fact there was no difference in actual rate. It may be that rate sometimes serves as a general scapegoat for perceived comprehension difficulties, whether it is the direct cause or not. Even in situations in which there is no interference with intelligibility, the increased processing costs associated with a particular accent may lead a listener to believe that the speech sample in question is produced at too fast a rate.

In this study, aspects of accent such as the number of segmental errors and prosodic goodness were treated as though they carried equal weight. There is no reason to believe, however, that these features are equivalent; investigation of the relative contribution of accent features using concepts such as functional load (see Catford, 1987) may provide directions for pronunciation instruction.

From a methodological standpoint, it appears that, although the exact word-match technique is more conservative than a measure that allows transcription errors unrelated to intelligibility, the differences between the two are minor. The unambiguous nature of exact word-match makes it an attractive method for assessing intelligibility. Furthermore, it is preferable to simple error counting because it controls for slight differences in utterance length.

The language identification task indicated that the listeners were able to recognize each of the languages at a rate above chance; however, they found Cantonese much easier to identify than the other three languages. The subjects' reported familiarity with various accents predicted their success at language identification. Furthermore, there was a correlation between familiarity and intelligibility scores. This finding extends the evidence of Gass and Varonis (1984) that familiarity with a particular language facilitates comprehension. Gass and Varonis manipulated familiarity within the confines of their experiment; in this study we were able to establish a link between intelligibility and individuals' familiarity with particular accents developed outside the experimental setting.

CONCLUSION

This study extends our understanding of the relationships among accent, perceived comprehensibility, and intelligibility, but there are a number of limitations that should be addressed. First, the utterances used here were relatively short, and in the experimental setting listeners had far more time to process them than is ordinarily available in natural discourse situations. Studies that incorporate assessment of

https://doi.org/10.1017/S0272263197001010 Published online by Cambridge University Press

intelligibility of longer stretches of accented speech at the discourse level are necessary to extend our understanding in this area.

Although we have confirmed the quasi-independence of accentedness, comprehensibility, and intelligibility in the speech of second language learners from several L1 backgrounds, there are some interesting differences in the findings reported here and those reported in previous research. Many of these differences can be accounted for by the lower proficiency level of the speakers. Because their speech contained a much larger number and wider range of errors, the listeners were most likely influenced by competing error types. In other words, there may have been a cumulative effect that is not reflected in individual measures of pronunciation. This may partially explain the overall lower correlations between accent features and task scores. In addition, the listeners in this study probably could not benefit from a familiarity effect to the same degree as listeners in previous studies in which all speakers came from the same L1 background. When assigning ratings to speech samples from several L1 backgrounds, our listeners may have been unable to apply a single set of criteria to all speakers.

Another area of investigation suggested by these results is an examination of individual differences in understanding accented speech. Both this study and that of Gass and Varonis (1984) have shown that familiarity with a particular accent facilitates intelligibility, but differences in familiarity alone cannot account for the range of responses in the accentedness and comprehensibility judgment tasks or the transcription task. To what do individual listeners attend, and what causes them to focus on those elements?

These results are likely to be of interest to second language teachers. Improvement in NNS comprehensibility, at least for intermediate- and high-proficiency learners, is more likely to occur with improvement in grammatical and prosodic proficiency than with a sole focus on correction of phonemic errors.

(*Received 17 August 1996*)

REFERENCES

Abercrombie, D. (1949). Teaching pronunciation. *English Language Teaching, 3,* 113–122.

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of non native pronunciation and deviance in segmentals, prosody and syllable structure. *Language Learning, 42,* 529–555.

Anderson-Hsieh, J., & Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning, 38,* 561–605.

Catford, J. C. (1987). Phonetics and the teaching of pronunciation: A systemic description of English phonology. In J. Morley (Ed.), *Current perspectives on pronunciation: Practices anchored in theory* (pp. 87–100). Washington, DC: TESOL.

Catran, J. (1985). *How to speak standard American English without a foreign accent.* Sherman Oaks, CA: Jade Publications.

Dahl, D. A. (1981). The role of experience for speech modifications for second language learners. *Minnesota Papers in Linguistics and Philosophy of Language, 7,* 78–93.

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America, 97,* 3125–3134.

Flege, J. E., Takagi, N., & Mann, V. (1995). Japanese adults can learn to produce English /r/ and /l/ accurately. *Language and Speech, 38,* 25–56.

https://doi.org/10.1017/S0272263197001010 Published online by Cambridge University Press

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

Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics.* New York: Newbury House.

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition, 12,* 251–285.

Major, R. C. (1987). Measuring pronunciation accuracy using computerized techniques. *Language Testing, 4,* 155–169.

Morley, J. (1991). The pronunciation component of teaching English to speakers of other languages. *TESOL Quarterly, 25,* 481–520.

Morley, J. (1994). A multidimensional curriculum design for speech-pronunciation instruction. In J. Morley (Ed.), *Pronunciation pedagogy and theory: New views, new directions* (pp. 66–91). Alexandria, VA: TESOL.

Munro, M. J. (1995). Nonsegmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition, 17,* 17–34.

Munro, M. J., & Derwing, T. M. (1995a). Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning, 45,* 73–97.

Munro, M. J., & Derwing, T. M. (1995b). Processing time, accent and comprehensibility in the perception of native and foreign accented speech. *Language and Speech, 38,* 289–306.

Munro, M. J., Flege, J. E., & MacKay, I. R. A. (1996). The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics, 17,* 313–334.

Nelson, C. (1982). Intelligibility and non-native varieties of English. In B. B. Kachru (Ed.), *The other tongue: English across cultures* (pp. 58–73). Urbana: University of Illinois Press.

Patkowski, M. (1990). Age and accent in a second language: A reply to James Emil Flege. *Applied Linguistics, 11,* 79–89.

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

Samar, V. J., & Metz, D. E. (1988). Criterion validity of speech intelligibility rating-scale procedures for the hearing-impaired population. *Journal of Speech and Hearing Research, 31,* 307–316.

Sato, C. J. (1991). Sociolinguistic variation and language attitudes in Hawaii. In J. Cheshire (Ed.), *English around the world: Sociolinguistic perspectives* (pp. 647–663). Cambridge: Cambridge University Press.

Smith, L. E. (1992). Spread of English and issues of intelligibility. In B. B. Kachru (Ed.), *The other tongue: English across cultures* (2nd ed., pp. 75–90). Urbana: University of Illinois Press.

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

Tajima, K., Port, R., & Dalby, J. (1994, June). *Effects of temporal correction on intelligibility of foreign-accented speech.* Paper presented at the 127th meeting of the Acoustical Society of America, Cambridge, MA.

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

Tyler, A. (1992). Discourse structure and the perception of incoherence in international teaching assistants' spoken discourse. *TESOL Quarterly, 26,* 713–729.

Van Els, T., & De Bot, K. (1987). The role of intonation in foreign accent. *The Modern Language Journal, 71,* 147–155.

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

Wennerstrom, A. (1994). Intonational meaning in English discourse: A study of non-native speakers. *Applied Linguistics, 15,* 399–420.

Wingstedt, M., & Schulman, R. (1984). Comprehension of foreign accents. In W. Dressler (Ed.), *Phonologica 1984: Proceedings of the 5th international phonology meeting* (pp. 339–344). Cambridge: Cambridge University Press.

https://doi.org/10.1017/S0272263197001010 Published online by Cambridge University Press