*Processing Time, Accent, and Comprehensibility in the Perception of Native and Foreign-Accented Speech**

MURRAY J. MUNRO
Simon Fraser University

TRACEY M. DERWING
University of Alberta

KEY WORDS

foreign accent

intelligibility

Mandarin

second language acquisition

speech processing

ABSTRACT

In this study, a sentence verification task was used to determine the effect of a foreign accent on sentence processing time. Twenty native English listeners heard a set of English true/false statements uttered by ten native speakers of English and ten native speakers of Mandarin. The listeners assessed the truth value of the statements, and assigned accent and comprehensibility ratings. Response latency data indicated that the Mandarin-accented utterances required more time to evaluate than the utterances of the native English speakers. Furthermore, utterances that were assigned low comprehensibility ratings tended to take longer to process than moderately or highly comprehensible utterances. However, there was no evidence that degree of accent was related to processing time. The results are discussed in terms of the "costs" of speaking with a foreign accent, and the relevance of such factors as accent and comprehensibility to second language teaching.

INTRODUCTION

Foreign-accented speech may be defined as non-pathological speech that differs in some noticeable respects from native speaker pronunciation norms. Evidence indicates that, in adult second language (L2) learners, non-native patterns of production are pervasive, affecting large portions of the segmental inventory (Munro, 1993) as well as prosodic aspects (see, e.g., Anderson-Hsieh, Johnson, & Koehler, 1992). Furthermore, accentedness is nearly inevitable in the speech of late L2 learners, even among those immersed for many years in the L2 environment (see Flege, Munro, & MacKay, 1995).

*This research was funded by the Support for the Advancement of Scholarship Fund, University of Alberta, and by the Social Sciences and Humanities Research Council of Canada. The authors gratefully acknowledge the assistance of L. Karra and many useful comments from J. Bürki-Cohen, B. L. Derwing, J. E. Flege, S. Rice, M. H. Southwood, and an anonymous reviewer.

Address for correspondence: Murray J. Munro, Department of Linguistics, Simon Fraser University, Burnaby, BC, V5A 1S6, CANADA. E-mail: mjmunro@sfu.ca

From the perspectives of both speaker and listener, a foreign accent may entail a variety of communicative "costs". In some instances, utterances may be partially or completely misunderstood because listeners are unable to recognize phonetic segments, words, or larger units that are pronounced with an accent. In such cases, the amount of information lost is presumably related to the type, severity and frequency of divergences from the norms (see also Flege, 1988).

Even when an L2 speaker's message is fully understood, however, an accent may have an impact on communication. Listeners sometimes exhibit prejudice against particular groups of L2 speakers or against non-native accents in general.[1] In fact, a number of researchers have noted irritation, a downgrading of attitudes towards speakers, or outright discrimination because of a non-native accent or nonstandard dialect (Albrechtsen, Henriksen & Faerch, 1980; Anisfeld, Bogo & Lambert, 1962; Brennan & Brennan, 1981a,b; Cunningham-Andersson, 1993; Fayer & Krasinski, 1987; Gumperz, 1982; Gynan, 1985; Johannson, 1978; Kalin & Rayko, 1978; Sato, 1991).

A further, more subtle, effect of an accent on comprehension was noted by Munro and Derwing (1995). They observed that Mandarin-accented utterances that were transcribed perfectly after a single hearing by native English listeners were sometimes judged by the same listeners as difficult to understand. It might be hypothesized that the listeners' tendency to assign low comprehensibility scores to some accented utterances was partly due to increased processing difficulty, which might manifest itself as increased processing time. To our knowledge there has been no investigation thus far of the role of processing time in the perception of accented speech. Yet there are several reasons to expect that accented speech should take longer to process than native-produced speech. For instance, the time required for recognition of accented consonant and vowel segments may be greater if those segments differ considerably from category prototypes. This may lead to increased recognition time for larger units such as syllables and words. Increased processing time may also result from a lack of comprehension or miscomprehension of lexical items, which might necessitate special top-down processing. Even though the speaker's message may ultimately be understood, the listener may have to work especially hard to decode it, perhaps even by "replaying" it from short-term memory.

In this study, a sentence verification task was carried out (cf. Pisoni & Dedina, 1986) in which listeners assessed the truth or falsity of a set of 40 sentences read by both native English and L2 speakers. Response latencies were measured in order to estimate processing time. In addition we asked the listeners to rate the speakers' utterances for accentedness and comprehensibility, in order to determine how such ratings might be related to the response latency data.

Research on the speech of second language learners has indicated that non-native utterances can be evaluated along several dimensions. Because there are some inconsistencies in the ways in which such dimensions have been interpreted and in the methods used to rate L2 learners' speech, we define three dimensions relevant to this study: intelligibility,

[1] Of course, such prejudice should not be viewed as the "fault" of the speakers. Clearly a number of social, economic and political factors may influence attitudes towards accented speech, but we will not attempt to address these here.

comprehensibility and accentedness. *Intelligibility* refers to the extent to which an utterance is actually understood. The intelligibility of both normal and pathological speech may be assessed by presenting listeners with words, sentences, or longer units, and asking them to write, in standard orthography, what they have heard. The resulting transcriptions are subsequently scored for accuracy (see Barefoot, Bochner, Johnson & von Eigen, 1993; Brodkey, 1972; Kent, Miolo, & Bloedel, 1994; Lane, 1963). We use the term *comprehensibility* to refer to listeners' perceptions of difficulty in understanding particular utterances. Speech samples have often been rated for comprehensibility with Likert-type scales (e.g., Munro & Derwing, 1994; Schairer, 1992; Varonis & Gass, 1982). *Accentedness* refers to how strong the talker's foreign accent is perceived to be. Ratings of accentedness have been obtained using a variety of techniques, including judgements on Likert scales (Munro & Derwing, 1994; Oyama 1982), assessments on a quasi-continuous scale, using an adjustable lever on a response box attached to a computer (Flege et al., 1995); and binary (good/bad) judgements (Varonis & Gass, 1982). The three dimensions of intelligibility, comprehensibility and accentedness are related but partially independent. Although heavily accented speech tends to be lower in comprehensibility and intelligibility than unaccented speech, this is not always the case. Munro and Derwing (1995) observed, for instance, that English utterances from native Mandarin speakers that were rated as moderately or heavily accented were often perfectly intelligible and highly comprehensible. Parallel findings have been reported in studies of extemporaneous utterances from other L1 groups (Derwing & Munro, 1995) and of identifications and ratings of L2 speakers' vowel productions (Munro, Flege, & MacKay, in press) and consonant productions (Flege, Takagi, & Mann, 1995).

The sentence verification and rating tasks in this study were designed to examine some of the costs of having a foreign accent. Here we are concerned with the question of whether or not a foreign-accented message is understandable, how difficult that message is to understand, and how much time it takes a listener to understand it. We also consider the relationships among verification times, comprehensibility and accent for highly intelligible utterances controlled for content and form.

METHODS

Talkers

The talkers were ten (five male, five female) native speakers of Mandarin and ten (five male, five female) native speakers of English between 25 and 41 years of age. The Mandarin speakers were graduate students at the University of Alberta in Edmonton, who had arrived in Canada after the age of 18. They had been living in Canada for two to six years ($M = 4$ years). All had scored a minimum of 550 on the Test of English as a Foreign Language (TOEFL)[2], and all used both English and Mandarin in Canada. In an informal assessment, four expert raters, all of whom had extensive experience working with non-native speakers of English, independently evaluated the Mandarin talkers' accents. All four

[2] These scores probably underestimate the talkers' English competence, given that they had lived in Canada for some time after taking the TOEFL.

agreed that the speakers' accents ranged from mild to strong. The native English talkers spoke a variety of English that was free from strong regional characteristics. Five spoke general Canadian English, and five spoke a Midwestern American variety of English.[3]

Stimulus Sentences

In this study, processing times were estimated by measuring the time needed for listeners to assign true/false judgements to a number of short sentences. A list of 25 true and 25 false sentences was created, the truth values of which could be easily determined by North American listeners on the basis of everyday knowledge. (E.g., *Elephants are big animals. Most people wear hats on their feet.*) Each item was a single-clause sentence of four to eight words. All lexical items, with the exception of a few well-known proper nouns, were listed as high frequency words by Sakiey and Fry (1979). As indicated below, some sentences were eliminated during the piloting phase of the experiment. The final list of items used in the study is given in the Appendix. The mean length of the sentences actually used was 5.9 words.

Recordings

Individual recordings were made with high-fidelity audio equipment in a sound-treated booth. Immediately before the recording session, each talker was given the list of sentences, instructed to read through it silently, and (in the case of the Mandarin talkers) invited to ask how to pronounce any unfamiliar words. The talker then practiced reading the full list aloud for one of the experimenters (TMD), who provided assistance with the pronunciation of a small number of lexical items. This was done to ensure that pronunciation errors would not be attributable to orthographic cues. During the recording phase, each talker read two repetitions of each of the 50 sentences. In the small number of cases (fewer than 20 from all talkers) in which noticeable hesitations or lexical substitutions occurred, the talker was asked to produce the utterance a third time.

So that response latencies could be measured, the stimuli had to be presented on a microcomputer. Therefore, the 1000 utterances were digitized at 10 kHz (16-bit resolution) with a Kay Computerized Speech Lab (CSL) and stored as audio files. Each item was then edited visually and aurally from a waveform display to ensure that it had been digitized correctly and that there was no silence after the last word in the utterance. The latter precaution was taken to ensure that response latencies would be as accurate as possible.

In a pilot task designed to eliminate any stimuli with ambiguous truth value, the full set of 50 statements, half of which were produced by randomly-selected Mandarin talkers, and half by English talkers, was presented to three native English listeners, two of whom were phonetically trained. They were asked to evaluate each statement as true or false by pressing a button on a response box. Each participant was then debriefed. On the basis of these participants' comments and careful consideration of all statements, nine items were eliminated because they were judged as potentially ambiguous, misleading, or noticeably

[3] Eight of the native English talkers were recorded in the Department of Biocommunication, University of Alabama at Birmingham. All other talkers were recorded in Edmonton. Each talker received $10 (either US or Canadian) for participating in the study.

more difficult to evaluate than the other items. A tenth sentence was excluded because it contained a lexical item that some of the Mandarin talkers had reported as unfamiliar.

Listeners

All listening sessions were held at the University of Alabama at Birmingham (UAB). The listeners were 20 native speakers of English (9 female, 11 male) all of whom spoke without a strong regional accent; six were from Canada and the remainder were mainly from the US Midwest. All were either students or staff at UAB. Their ages ranged from 19 to 45, and all passed a pure-tone hearing screen (250 – 8000 Hz) prior to performing the listening tasks. None had any prior knowledge of the list of statements to be used for verification. Ten of the listeners reported regular contact with non-native speakers of English (though not necessarily Mandarin speakers); ten had only occasional or no contact with non-native speakers. The listeners were each paid $10 US for participating in an individual session of approximately 40 minutes.

Procedure

Each of the 40 statements could be verified only once by each listener. (If a listener heard a statement for a second time, he or she might not evaluate it on-line, but rather from memory, and the response latency would be meaningless.) Accordingly, 20 different randomized stimulus lists of 40 items each were prepared, one for each of the listeners. In each list, half the utterances were produced by native Mandarin and half by native English talkers. Items were selected so that each listener would hear each talker exactly twice, once producing a true statement and once producing a false one. All listeners heard the full set of 20 talkers and the full set of 40 different stimulus sentences; however, no two listeners ever heard the same talker producing the same sentence. Each listener heard the statements in a different random order.

Individual listening sessions were held in a sound-treated booth. Stimuli were presented at a comfortable listening level through high-fidelity headphones, and verification data were obtained via a custom response box connected to a microcomputer. The entire stimulus set was played three times. During the first pass, the listeners were instructed to place their dominant hand at a point midway between two buttons, one marked "True" and the other marked "False," on the response box. They were told to listen carefully to each statement and to press the appropriate response button as quickly as possible when they had determined the truth value of the statement. They were asked to strive for both speed and accuracy. Response times were measured with one millisecond resolution from the beginning of the utterance. After evaluating a statement, the listeners were asked to write out in standard orthography exactly what they had heard. They then pressed the button a second time so that the computer would play the next stimulus.

During the second and third passes, the listeners used a response sheet to rate each utterance on a scale of 1 to 9 for either comprehensibility or foreign accentedness. Ten listeners rated comprehensibility before accentedness, and ten performed the ratings in the opposite order. For the comprehensibility ratings, the listeners were instructed to decide how difficult each talker was to understand. A rating of "1" was to be used for talkers who were "not difficult to understand at all," while "9" was to be used for talkers who were "very difficult to understand." For the foreign accent ratings, the listeners were instructed to decide how strong each talker's foreign (i.e., non-English) accent was. A rating of "1" was to be used if the talker had "no foreign accent at all," and "9" was to be used if the talker had a "very strong foreign accent." Numbers in between were to be used for intermediate degrees of accent. For both rating tasks, the listeners were reminded that they had already heard the entire stimulus set, so they would have some idea of the range of comprehensibility or accent in the speech samples. They were instructed to keep this range in mind and to use the full rating scale in their judgements, assigning some scores of "9" and some scores of "1."

RESULTS

Verification and Transcription Scores

Verification scores (i.e., the number of times each subject correctly assigned truth value) were tallied automatically by the computer, and each transcription was scored by an assistant as either correct or incorrect. To be considered correct, a transcription had to correspond exactly to the actual utterance, except that a few minor errors were permitted (i.e., the use of a singular form instead of a plural or the omission of the determiners "a" and "the"). High rates of accuracy were observed: Of the 800 utterances, 96% were verified correctly and 94% were transcribed correctly. The verification scores from the twenty listeners were submitted to a repeated measures ANOVA with L1 of talker (Mandarin, English) and Truth Value (True, False) as within-subject factors. (Unless otherwise noted, an alpha level of 0.01 was adopted in the ANOVAs reported here.) A significant effect of L1, $F(1, 19) = 13.571$, was observed, indicating that the Native English talkers' utterances were correctly verified more often (98% correct) than those of the Mandarin talkers (93% correct). The effect of Truth Value, $F(1, 19) < 1$, was not significant, nor was the interaction between L1 and Truth Value, $F(1, 19) < 1$.

The pattern of results in the transcription task was similar to that observed in the verification task. Of the 400 sentences produced by the Native English talkers, 99% (i.e., all but four) were transcribed correctly. An ANOVA revealed that significantly fewer of the Mandarin talkers' utterances (89%) were transcribed correctly, $F(1, 19) = 32.624$. Again the effect of truth value failed to reach significance, $F(1, 19) = 1.753$, as did the two-way interaction, $F(1, 19) = 3.289$.

Response Latency Data

In analyzing the response latencies (RLs), we followed the standard procedure (cf. Pisoni & Dedina, 1986) of including data from only those sentences that were both verified correctly and transcribed correctly. This procedure was necessary because data from incorrectly verified or incorrectly transcribed utterances would be uninterpretable. In some instances, listeners may have accidentally pressed the wrong button, even though they actually understood the utterance. On the other hand, when listeners failed to understand an utterance (as indicated by a wrong transcription or, in some cases, a wrong verification) the RL might reflect the time it took to guess rather than the time it took to determine the meaning of the utterance.

To minimize the effect of outliers on the analyses, we excluded RL values falling outside

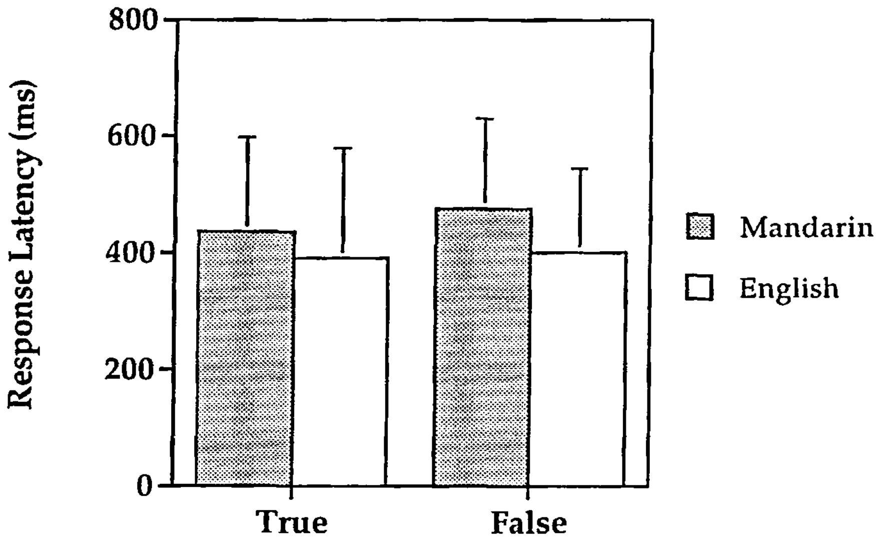

Fig. 1

Mean response latencies (with standard deviations) for true/false judgements by 20 listeners on the 20 true and 20 false statements produced by 10 native Mandarin and 10 native English talkers.

the range of ± 2 standard deviations from the mean response time of 469 ms. Also, in the statistical analyses reported here, we computed the RLs separately for the true and the false statements in case there was a difference in verification times for these two types of utterances.

Prior to analyzing the response time data, we examined utterance durations. The mean sentence durations for the ten Mandarin talkers (M = 2.29 s, *s.d.* = 0.25) and the ten native English talkers (M = 1.77 s, *s.d.* = 0.13) were submitted to a one-way ANOVA with Native Language of talker as a between-subjects factor. The native English talkers produced sentences of shorter duration than the native Mandarin talkers, $F(1, 19) = 33.35$. To eliminate the possible confounding effects of differences in utterance duration on the RLs in the analyses (and because durations varied from item to item anyway), for each stimulus item the total utterance duration was subtracted from the total response time. The RL value for any item was therefore a measurement of the amount of time it took the listener to press the response button after the *end* of the utterance.

Analysis by Listener. The mean RLs for each listener were submitted to a two-way repeated measures ANOVA with Truth Value (two levels) and Native Language of Talker (two levels) as factors. Figure 1 illustrates the four sets of response times. A significant effect of Native Language was observed, $F(1, 19) = 27.107$, because the sentences produced by the Mandarin talkers took longer to verify (M = 457 ms) than did the sentences produced by the native English talkers (M = 395 ms). Although there was a tendency for the false statements to require longer verification times than the true statements, the effect of Truth Value failed to reach significance, $F(1, 19) = 2.129$, as did the interaction of Native Language and Truth Value, $F(1, 19) < 1$.

An examination of the data from the 20 individual listeners revealed that 17 of them took longer to verify the statements produced by the native Mandarin talkers than those produced by the native English talkers. The remaining three showed a slight (M = 11 ms), nonsignificant trend in the opposite direction. The difference in RLs for the two sets of utterances varied considerably from listener to listener, however. The smallest effect was only 3 ms, while the largest was 137 ms.

Analysis by Utterance. The listeners' performance on the 40 individual statements was also considered. Mean RLs for each statement (pooled over talkers and listeners) were computed for both native Mandarin and native English talkers. A repeated measures ANOVA with Native Language of Talker and Truth Value as factors revealed that individual statements took significantly longer to verify when uttered by Native English talkers than when produced by Mandarin Talkers, $F(1, 38) = 8.711$. The effect of Truth Value and the interaction of the two factors were both nonsignificant, $F(1, 38) = 1.646$, and $F(1, 38) = 1.025$, respectively. For 29 of the 40 items, the Mandarin talkers' utterances took longer to verify than those of the native English talkers. In the remaining instances (marked with an asterisk in the Appendix), there was a tendency in the opposite direction.

We could identify no commonalities in the sentences exhibiting the reversed pattern. It is possible, however, that differences in speaking rates played a role. For instance, if the Mandarin speakers produced some sentences particularly slowly, it is possible that the listeners may have already processed substantial parts of these utterances *prior* to hearing the end of the sentence. This might result in shorter measured response times. On the other hand, fast utterances produced by Native English speakers may have taken longer to process, because processing did not start until the sentence was complete or nearly complete. To test this hypothesis we looked for correlations between utterance durations and response times in the productions from the two groups of talkers. Because response latency data are subject to considerable noise (due to such factors as variation in speed of hand movement and straying of attention on the part of listeners), we used mean data from several trials rather than the original raw data set. The RT data were rank ordered and divided into "bins" of 15 data points. For each bin, mean RT values and corresponding mean utterance durations were computed. For the Mandarin speakers' productions the Pearson correlation between response time and utterance duration was –.463, and for the native English talkers' productions it was –.411 ($p < .05$ in both cases). It appears, then, that faster utterances tended to take longer to verify, regardless of the L1 background of the speaker. However, the fact that the Mandarin talkers' productions took significantly longer overall to verify than those of the native English talkers, even though the latter productions were slower, indicates that the effect of an accent on verification times was more important than the effect of speaking rate.[4]

[4] The finding that faster speaking rates were associated with longer RLs actually strengthens the case for the effect of accent on RLs. However, as a cautionary measure, we recomputed the ANOVA on the 40 items, this time using normalized response times: We divided all response times by the durations of the corresponding utterances to remove the effect of utterance duration. Even with this adjustment, the effect of native language on response times continued to be significant, $F(1, 38) = 7.126$.

Analysis by Talker. The RL data were also examined from the perspective of individual talkers. Response times were pooled over listeners and sentences to give mean times for each talker. A mixed-design ANOVA with L1 of Talker as a between-subjects factor and Truth Value as a within-subjects factor again revealed a significant effect of L1, $F(1, 18) = 9.555$, but not of Truth Value, $F(1, 18) = 7.849$. Nor was the interaction effect significant, $F(1, 18) = 3.176$. The range of mean response times for the Mandarin talkers was from 377 ms to 522 ms. For the native English talkers, the range was from 325 to 445. All but one of the Mandarin talkers had an average RL greater than the native English mean value.

Rating Data

Comprehensibility and foreign accent scores were tabulated for each of the 800 stimulus items. As expected, the native English talkers' productions received much lower accent ratings than did those of the native Mandarin talkers: $M = 1.5$ versus 6.3; $t(798) = 46.03$, $p < .05$. The same was true for the comprehensibility ratings: $M = 1.5$ versus 5.4; $t(798) = 28.82$, $p <.05$.

Previous studies have shown that comprehensibility tends to improve with increased exposure to foreign-accented speech (Gass & Varonis, 1984), although a foreign accent may tend to be heard as stronger with increased experience (Flege & Fletcher, 1992; Munro & Derwing, 1994). In the present experiment the stimuli were heard three times. It was our expectation that the first pass, during which the sentences were verified and transcribed, would serve as a familiarization set and that there would be a minimal effect, if any, of differences in familiarity between the second and third passes, during which the comprehensibility and accent scores were assigned. To verify this, we compared the mean comprehensibility scores (on the Mandarin talkers' productions only) from the listeners who rated comprehensibility during the second pass with scores from those who assigned these ratings during the third pass. The difference was nonsignificant, $t(18) = 1.61$, as was the difference between the accent scores obtained during the second and third passes, $t(18) = 0.48$.

Figures 2 and 3 are histograms showing the distributions of the comprehensibility and accent ratings assigned to all the Mandarin talkers' productions. These data show quite different patterns: The comprehensibility ratings are fairly evenly distributed across the full range, while the distribution of accent scores is sharply skewed, with a peak at the high end of the scale. Figure 4 shows the accent ratings of the 72 Mandarin-accented utterances that received comprehensibility ratings of 1 or 2 (easiest to understand). The distribution of accent ratings shows peaks at "1" and at "6," with other scores at all points between "1" and "9." As in the Munro and Derwing (1995) study, these data support the hypothesis that a non-native speaker may be rated as highly comprehensible but still be heard to have a moderate or relatively strong foreign accent.

The Pearson correlation between the mean accent and comprehensibility scores for the Mandarin talkers was computed at .624 ($p < .01$). However, for individual listeners the strength of the correlation varied considerably, ranging from .140 to .917, and for six listeners the correlation failed to reach significance at the .01 level. Once again these findings confirm the observation by Munro and Derwing (1995) that accentedness and comprehensibility are only partially-correlated dimensions of L2 speech. Furthermore, the

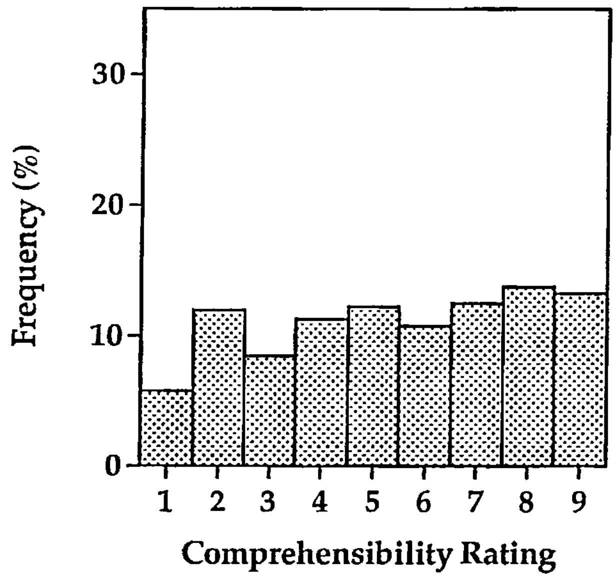

Fig. 2

Distribution of comprehensibility ratings for the 400 utterances produced by the native Mandarin talkers. The ratings ranged from "1 – not difficult to understand at all" to "9 – very difficult to understand."

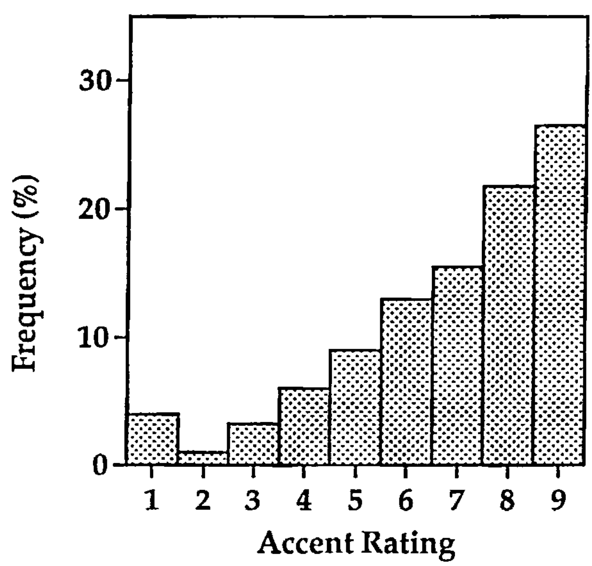

Fig. 3

Distribution of accent ratings for the 400 utterances produced by the native Mandarin talkers. The ratings ranged from "1 – no foreign accent at all" to "9 –very strong foreign accent."

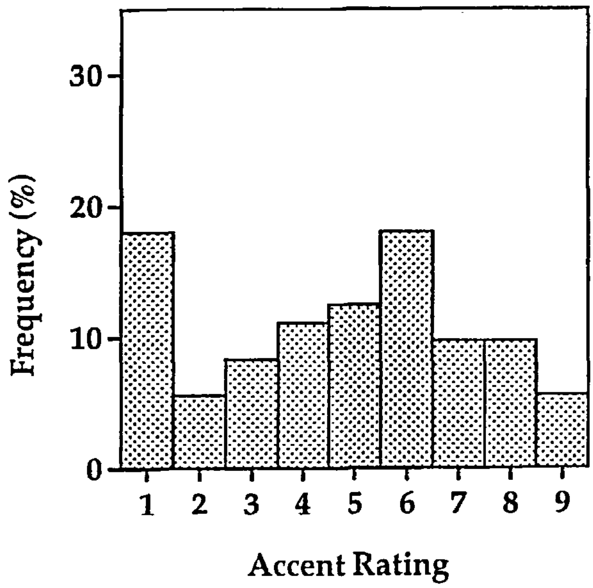

Fig. 4

Distribution of accent ratings for the 72 utterances produced by the native Mandarin talkers that received comprehensibility ratings of "1" or "2."

strength of the relationship between these dimensions varies from listener to listener. Some listeners appear to assign considerable importance to comprehensibility when making accent ratings while others do not.

An additional question to be considered here was whether utterances evaluated as heavily accented or as low in comprehensibility would tend to require longer processing times than utterances receiving more favourable evaluations. Once again, because of the problems inherent in using raw latency data from several listeners, we grouped the RLs (with outliers removed) for analysis according to comprehensibility and accent scores. Items with comprehensibility ratings of "1," "2," or "3" were classified as "High comprehensibility tokens," those with ratings of "4," "5," or "6" as "Moderate comprehensibility tokens," and those with ratings of "7," "8," and "9" as "Low comprehensibility tokens." An analogous classification scheme was used to group the RLs according to accent ratings. Mean response latencies for each of the groups are given in Figure 5. A one-way ANOVA revealed a significant effect of comprehensibility group on the RL data, $F(2, 309) = 6.672$. Post hoc Tukey HSD tests indicated that the Low comprehensibility tokens had longer RLs than either the High comprehensibility or the Moderate comprehensibility tokens. Although there was a parallel tendency for heavily accented utterances to have longer RLs, the effect of accent group on RL scores failed to reach significance, $F(2, 309) = 1.557$. In short, the results suggest that degree of comprehensibility had an impact on response times, while no evidence was found that accentedness in itself had such an effect.

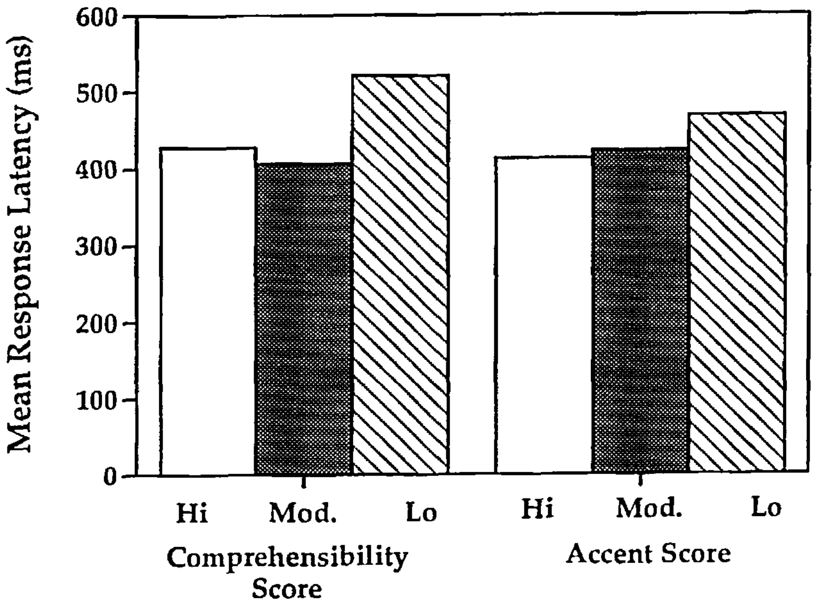

Fig. 5

Mean response latencies for utterances receiving "high" ratings ("1," "2," or "3"), "moderate" ratings ("4," "5," or "6") and "low" ratings ("7," "8," or "9") on the comprehensibility and accent scales.

Effects of familiarity with accented speech

The final issue addressed in this study was whether experience with foreign-accented speech would have an effect on the dependent variables under consideration here. It might be expected that listeners with regular exposure to accented speech would exhibit faster response times to the accented utterances than would listeners with little or no such experience. They might also obtain better transcription and verification scores and differ in their accent and comprehensibility ratings. Several one-way ANOVAs were performed to determine whether self-reported exposure to accented speech (i.e., regular contact vs. little or no contact with non-native speakers) influenced any of the dependent measures. However, the effect of exposure on response times, $F(1, 19) = 1.626$, verification scores, $F(1, 19) = 1.409$, transcription scores, $F(1, 19) = 0.171$, mean comprehensibility ratings, $F(1, 19) = 0.449$, and mean accent ratings, $F(1, 19) = 0.663$, was nonsignificant.

DISCUSSION

The results of this study illustrate two types of "cost" associated with a foreign accent. First, the listeners made slightly more errors in verifying and transcribing the Mandarin-accented utterances than the statements produced by the NSs of English. In some instances, then, full comprehension was apparently blocked because of the presence of a foreign accent.

However, errors in the verification and transcription of the Mandarin-accented utterances were still fairly rare, occurring 5 – 10% more often than for native English utterances. Second, the Mandarin speakers' productions took significantly longer to verify (by about 50 ms on average) than those of the native English talkers. It is likely that the difference in verification times was due to greater processing time for the Mandarin-accented utterances.

While nearly all the listeners showed at least a tendency to require more time to verify the Mandarin-accented utterances than the unaccented ones, an examination of the listeners' performance on individual items indicated that Mandarin-accented utterances were not always verified more slowly than native English ones. We can offer no definite explanation for why some productions failed to show the same pattern. However, our data suggest that at least two factors may have interacted to influence the verification times of the utterances under consideration here. On the one hand, the presence of a Mandarin accent apparently resulted in longer processing times. On the other, the duration data reported here indicate that the Mandarin talkers generally spoke more slowly than the native English talkers. Also, within talker groups, slower speaking rates were associated with faster response times. The effect of a slower speaking rate may have been to decrease the amount of time required to verify an utterance after its completion, because more of the utterance may have been processed *prior* to its completion. While the increase in processing time due to the presence of an accent was the prevailing factor in most of the verification times, it is possible that for some reason in some utterances, speaking rate had a stronger effect.

The comprehensibility and accentedness data collected here largely confirm the findings of Munro and Derwing (1995). In that study of extemporaneous speech samples, it was observed that while comprehensibility and accentedness were indeed related, listeners frequently assigned moderate or high accentedness scores to speech samples that were perfectly intelligible and highly comprehensible. In the earlier study, an overall Pearson correlation of .601 was observed between comprehensibility and accent scores assigned by 18 listeners, indicating that a strong accent tended to be associated with reduced comprehensibility. However, the strength of the relationship varied considerably from listener to listener. The data from the present study indicate very similar patterns for single-sentence utterances read by the talkers. The overall correlation between comprehensibility and accent was nearly the same (.624), but once again the strength of the relationship varied greatly from listener to listener. For those utterances that had been transcribed correctly and assigned good comprehensibility scores, a wide range of accent scores was once again observed. This finding confirms that, as was the case for extemporaneous speech, a short utterance read aloud may be highly intelligible and comprehensible, yet rated as moderately or heavily accented.

In both the Munro and Derwing (1995) study and the present one, the relationship between comprehensibility and accentedness appears to be somewhat weaker than that reported by Varonis and Gass (1982). They obtained a rank order correlation of .899 between these dimensions for grammatically correct statements. We propose that the differences between the results of these studies are at least partly due to the fact that we computed Pearson correlations based on data from two similar 9-point scales, while Varonis and Gass used rating scale data for their comprehensibility scores, but binary (good/bad) judgements for the accent ratings.

An analysis of the response latency data in terms of the comprehensibility and accent ratings revealed a relationship between comprehensibility ratings and response times. However, there was no statistical evidence of a relationship between accentedness and response times. Not only does this finding give additional support to our earlier observation that comprehensibility and accent are partially-independent dimensions, but it also suggests that when listeners evaluate the comprehensibility of non-native speech samples, they take processing time into account. Perhaps when they are aware that an utterance takes extra time to process, they are inclined to rate it as less comprehensible, even when they ultimately determine the intended message. There was no evidence here to indicate that judgements of accentedness were influenced by the same considerations.

CONCLUSIONS

The findings reported here demonstrate the importance of distinguishing among three dimensions of L2 speech: intelligibility, comprehensibility, and accentedness. Assessments by expert raters as well as naive listeners indicated that some of the talkers who participated in this study exhibited strong foreign accents. Yet when their accented productions were presented without a meaningful context to untrained listeners, they were nearly always intelligible and were often rated as highly comprehensible. While accented productions generally took longer to evaluate than unaccented ones, there were some exceptions to this rule. Furthermore, the fact that processing times were related to comprehensibility ratings but not to accent ratings suggests that the degree of accentedness in an utterance is not a reliable predictor of processing time. In short, while this study did indeed identify some of the costs of having a foreign accent, it also showed that an accent – even a strong one – is by no means an inevitable barrier to communication.

The findings reported are useful in conceptualizing the effect of a foreign accent on speech perception by native listeners. Accented speech may be regarded as speech that deviates in various non-pathological ways from native speaker norms. The deviations may include phone substitutions, phonetic distortions, and non-native prosodic patterns. While some deviations may be so severe as to result in an unintelligible or partially-intelligible message, others may have only the more subtle effect of requiring listeners to "work harder" to understand the message. While the exact nature of the extra work can only be speculated upon, it may entail special top-down processing because a particular phonetic segment, word, or phrase was misunderstood. Alternatively, "replaying" of all or part of the message from short-term memory may occur. Awareness of extra processing time may cause listeners to evaluate some accented messages as harder to understand than unaccented ones, even when full comprehension eventually does occur. One hypothesis to be explored in future research is that the listener "irritation" reported in previous studies may result from reduced intelligibility as well as from awareness of increased processing demands.

These results may also be interpreted in terms of their relevance to L2 pedagogy. In particular, it should be stressed that a foreign accent, in itself, is not necessarily an impediment to communication. Therefore, the notion that foreign accent "reduction" automatically entails improved comprehensibility is quite incorrect. If improved communicative competence is a primary goal in second language teaching, then attention ought to be directed to those specific aspects of the learner's speech that most affect comprehensibility and intelligibility and not to those aspects simply associated with accent. There is a clear need for further research on which aspects of a foreign accent may have the strongest impact on comprehensibility and intelligibility. Another pedagogical implication of this study concerns the finding that, within L1 groups, sentences that were uttered more slowly tended to be verified more quickly. This may mean that reduced speaking rate could serve as a compensatory strategy for L2 learners whose speech is of reduced comprehensibility when uttered at a normal rate.

It should be noted that the data obtained here were collected under ideal listening conditions. It is important that future studies examine ratings of L2 speech under a variety of realistic conditions during which communication might take place. For instance, it is not known whether the effects of noise or filtering (e.g., in telephone or radio transmissions, in noisy rooms, or at variable loudness levels) have the same degree of impact on the processing time or comprehensibility of accented speech as on native-produced speech, or whether the effects of such conditions vary as a function of degree of accent.

Although familiarity with accented speech did not appear to influence response times, or any of the other dependent measures in this study, we did not control for the specific type of L1 accent to which the listeners had been exposed. In view of Gass and Varonis' findings that familiarity with a particular accent and a particular speaker improved comprehensibility, it seems reasonable to investigate the relationship between familiarity and processing times further. A better understanding of the range and scope of effects produced by foreign accented speech will have many practical benefits.

Revised manuscript received: October 26, 1995; accepted: October 31, 1995

REFERENCES

ALBRECHTSEN, D., HENRIKSEN, B., & FAERCH, C. (1980). Native speaker reactions to learners' spoken interlanguage. *Language Learning*, **30**, 365 – 396.

ANDERSON-HSIEH, J., JOHNSON, R., & KOEHLER, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, **42**, 529 – 555.

ANISFELD, M., BOGO, N., & LAMBERT, W. (1962). Evaluational reactions to accented English speech. *Journal of Abnormal Social Psychology*, **69**, 89 – 97.

BAREFOOT, S., BOCHNER, J., JOHNSON, B., & VON EIGEN, B. (1993). Rating deaf speakers' comprehensibility: An exploratory investigation. *American Journal of Speech-Language Pathology*, **2**, 31 – 35.

BRENNAN, E., & BRENNAN, J. (1981a). Accent scaling and language attitudes: reactions to Mexican American English speech. *Language and Speech*, **24**, 207 – 221.

BRENNAN, E., & BRENNAN, J. (1981b). Measurements of accent and attitude toward Mexican American speech. *Journal of Psycholinguistic Research*, **10**, 487 – 501.

BRODKEY, D. (1972). Dictation as a measure of mutual intelligibility: A pilot study. *Language Learning*, **22**, 203 – 220.

CUNNINGHAM-ANDERSSON, U. (1993). Stigmatized pronunciations in non-native Swedish. *PERILUS*, **17**, 81 – 106.

DERWING, T., & MUNRO, M. (1995). Accent, intelligibility and comprehensibility: Evidence from four L1s. Manuscript submitted for publication.

FAYER, J., & KRASINSKI, E. (1987). Native and nonnative judgements of intelligibility and irritation. *Language Learning*, **37**, 313 – 326.

FLEGE, J. (1988). The production and perception of speech sounds in a foreign language. In H. Winitz (Ed.), *Human Communication and Its Disorders, A Review 1988*, (pp. 224 – 401). Norwood, NJ: Ablex.

FLEGE, J., & FLETCHER, K. (1992). Talker and listener effects on the perception of degree of foreign accent. *Journal of the Acoustical Society of America*, **91**, 370 – 389.

FLEGE, J., MUNRO, M., & MACKAY, I. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97, 3125 – 3134.

FLEGE, J., TAKAGI, N., & MANN, V. (1995). Japanese adults can learn to produce English /ɹ/ and /l/ accurately. *Language and Speech*, **38**, 25 – 56.

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

GUMPERZ, J. (1982). *Discourse Strategies*. Cambridge, U.K.: Cambridge University Press.

GYNAN, S. (1985). Comprehension, irritation and error hierarchies. *Hispania* 68, 160 – 165.

JOHANSSON, S. (1978). Studies in error gravity: Native reactions to errors produced by Swedish learners of English. *Gothenburg Studies in English 44*. Department of English, University of Gothenburg, Sweden.

KALIN, R., & RAYKO, D. (1978). Discrimination in evaluative judgements against foreign accented job candidates. *Psychological Reports*, **43**, 1203 – 1209.

KENT, R.; MIOLO, G., & BLOEDEL, S. (1994). The intelligibility of children's speech: A review of evaluation procedures. *American Journal of Speech-Language Pathology*, **3**, 81 – 95.

LANE, H. (1963). Foreign accent and speech distortion. *Journal of the Acoustical Society of America*, **35**, 451 – 453.

MUNRO, M. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech*, **36**, 39 – 66.

MUNRO, M., & DERWING, T. (1994). Evaluations of foreign accent in extemporaneous and read material. *Language Testing*, **11**, 253 – 266

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

MUNRO, M., FLEGE, J., & MACKAY, I. (In press). The effects of age of second-language learning on the production of English vowels. *Applied Psycholinguistics*, **17**.

OYAMA, S. (1982). A sensitive period for the acquisition of a nonnative phonological system. In S. Krashen & M. Long (Eds.), *Child-adult differences in second language acquisition*. Rowley, MA: Newbury House.

PISONI, D., & DEDINA, M. (1986). Comprehension of digitally encoded natural speech using a sentence verification task (SVT): A first report. *Research on speech perception progress report No. 12*. Indiana University.

SAKIEY, E., & FRY, E. (1979). *3000 Instant words*. Highland Park, NJ: Dreier Educational Systems.

SATO, C. (1991). Sociolinguistic variation and language attitudes in Hawaii. In J. Cheshire (Ed.) *English around the world: Sociolinguistic perspectives*. Cambridge, U.K.: Cambridge University Press.

SCHAIRER, K. (1992). Native speaker reaction to non-native speech. *The Modern Language Journal*, **76**, 309 – 319.

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

APPENDIX

List of Stimulus Sentences for Verification Task

These sentences were presented in a different random order for each listener. Items marked with an asterisk showed no tendency to require more response time when produced by Mandarin talkers.

*1. Gasoline is an excellent drink.
*2. Elephants are big animals.
3. The Queen of England lives in Washington.
*4. Spaghetti grows on tall trees.
5. Hot and cold are opposites.
*6. The sun always sets in the north.
7. The inside of an egg is blue.
8. August is a winter month.
9. It always snows in July.
10. March has thirty-eight days.
11. Most people wear hats on their feet.
12. The stars come out in the day.
13. Exercise is good for your health.
*14. Japan is a wealthy country.
*15. Wednesday is the first day of the week.
*16. All men can have babies.
17. All dogs have fifteen legs.
18. Shakespeare wrote many fine plays.
19. Most teenagers like rock and roll.
20. Some people love to eat chocolate.
21. Some people keep dogs as pets.
22. Young children can be very noisy.
23. Some roses have a beautiful smell.
24. Hungry cats like to chase mice.
*25. People eat through their noses.
26. A nickel is worth twenty-five cents.
27. You can start a fire with a match.
28. Red and green are colours.
*29. Many houses are made of bricks.
*30. There are many cities on the moon.
31. Italy is a country in Europe.

32.	Many people drink coffee for breakfast.
33.	You can buy beer at church.
34.	The American flag has stars and stripes.
35.	Gold is a valuable metal.
36.	Milk comes from yellow chickens.
37.	Most swim suits have long sleeves.
*38.	A monkey is a kind of bird.
39.	Ships travel on the water.
40.	You can buy a burger at McDonalds.