Munro, M. J., Derwing, T. M., & Holtby, A. K. (2012). Evaluating individual variability in foreign accent comprehension. In. J. Levis & K. LeVelle (Eds.). *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference*, Sept. 2011. (pp. 233-239). Ames, IA: Iowa State University.

EVALUATING INDIVIDUAL VARIABILITY IN FOREIGN ACCENT COMPREHENSION

Murray J. Munro, Simon Fraser University

Tracey M. Derwing & **Amy K. Holtby**, University of Alberta

How much do people vary in their ability to understand foreign-accented speech? Answering this question may lead to useful insights into communication processes among interlocutors from diverse linguistic backgrounds. Although some speakers are often described as being easier to understand than others, listener factors also play a role in whether a particular accented utterance is understood. Such influences include prior experience with the speaker's accent and sharing the same first language as the speaker. Age may also be important, in that young children may have less robust perceptual representations that compromise processing of unfamiliar speech patterns. This paper describes the development of a pilot tool for assessing listeners' comprehension abilities. Six groups of listeners differing in age and L1 background completed a true-false judgment task in which they responded to sentence-length items produced in English by Cantonese speakers. The resulting comprehension scores indicate effects of listener age (teens performed better than younger children) and listener proficiency; however, they also indicate shared comprehension across listeners, even among those from diverse L1 backgrounds.

INTRODUCTION

In the study of second language (L2) phonetic learning, *intelligibility* and *comprehension* are closely related phenomena. Intelligibility is often seen as a characteristic of speakers, utterances, and accents, and it is common to describe people and their speech productions in terms of how intelligible they are. However, intelligibility, as most speech researchers understand the term, refers to the degree to which a speaker's communicative intent matches the listener's response (Schiavetti, 1992). For that reason, it can be assessed only by reference to what listeners understand. It cannot be evaluated by enumerating phonetic errors; nor can we measure it with acoustic analysis software, no matter how sophisticated our approach might be. This is because intelligibility does not reside exclusively in the speaker or in the utterance itself. Rather, in real-world situations it is the result of a speaker's attempt to communicate with an audience. Assessing speech intelligibility therefore requires that we measure the comprehension[1] of listeners. However, the complexity inherent in the notion of comprehension makes it impossible that any single measure could suffice. In the speech sciences, the most common approach to intelligibility assessment is to have listeners transcribe (in standard orthography) the utterances

[1] It is important here not to confuse *comprehension* with *comprehensibility*. We use the latter term to refer to a listeners' perception of the degree of difficulty in understanding an utterance.

they hear and to count the number of correct words in their transcriptions. While there is no question that determining the words in an utterance is a prerequisite for understanding, it is often not clear how such a simple measurement corresponds to the listener's appreciation of such matters as the speaker's attitude or intent, or the relevance of the utterance to the speaking context. Understanding these phenomena is clearly a part of comprehension, and any full characterization of intelligibility must take them into account.

The fact that some listeners succeed more than others at comprehending L2 speech raises intriguing issues in the measurement of intelligibility and in the teaching of L2 pronunciation. One fundamental problem facing researchers is establishing the sources of inter-listener variability. Although many anecdotally-based opinions exist, actual research on this topic remains limited. Gass and Varonis (1984) observed a positive effect on listeners of familiarization with a particular accented speaker, the accent being spoken, accented speech in general, or the topic being spoken about. Other work indicates that sharing the L1 background of the speaker (i.e., the same accent) can offer comprehension benefits for speech in noise (Bent & Bradlow, 2003). However, this effect may occur primarily in low-proficiency listeners (Hayes-Harb et al., 2008). In a study of speech without noise, Munro, Derwing & Morton (2006) assessed high-proficiency L2 listeners' comprehension of English produced with their own and other accents. Although Japanese listeners did understand slightly more than other listeners when presented with Japanese-accented speech, Mandarin-speaking listeners showed no such advantage for their own accent. Given that other studies have yielded similar weak findings, there is no indication that sharing an L1 background offers large or consistent benefits for comprehension. Rather, listeners – even from diverse backgrounds – tend to comprehend similarly: a speaker who is not understood by one listener will tend to be understood poorly by others (Munro, in press).

Another source of variability in the comprehension of L2 speech is the listener's age. Burda, Scherz, Hageman, & Edwards (2003) observed poorer comprehension by geriatric listeners compared with younger adults. Although comprehension of accented speech by children has received little attention from researchers, older children might be expected to have an advantage over younger children. Because L1 phonetic development is a matter of language experience, older children may process speech more efficiently, giving them greater capacity to adjust to speech patterns with which they are not familiar.

The goal of this exploratory project is to determine the feasibility of constructing an easy-to-administer tool for comparing listeners' comprehension of L2 speech. Such a tool could be used to systematically compare different groups of listeners to evaluate the effects of such factors as L1 background and age on L2 speech comprehension. For this purpose we propose a sentence verification task consisting of utterances representing a range of speakers who vary in intelligibility. To be useful as an assessment tool, the task must be straightforward for listeners to complete and sensitive enough to yield differences in performance when listeners genuinely differ in comprehension. Our approach is to pilot a 40-item test in which we compare listeners varying in L1 background and age.

METHOD

The verification task used here is based on previous work by Munro & Derwing (1995), who developed a set of short statements that could be readily judged as true ('Some people have sandwiches for lunch') or false ('April is the first month of the year') using general knowledge.

This task assesses listener comprehension (and therefore intelligibility) in that listeners must process a complete utterance in order to determine the correct response. It is a convenient alternative to counting correct words in a transcription task and tends to yield very similar results (Munro & Derwing, 1995). Individual listeners' scores on the 40-item test were used to assess their comprehension. In both the original study and the current one, speech representing a single foreign accent was used to minimize effects of potential prejudices against particular accents.

Stimulus Preparation

The test stimuli were drawn from a set of over 3200 utterances recorded by 81 Cantonese speakers, all of whom were high-proficiency English speakers enrolled in post-secondary institutions in Canada. Other aspects of their backgrounds are largely irrelevant to this study because our objective was to obtain a set of speech items varying in intelligibility. Success in that respect could be determined only after completion of the experiment.

During individual recording sessions in a sound-treated room, the speakers read aloud a set of 40 sentences similar to those used by Munro and Derwing (1995). Prior to actual recording they practiced the full set of items once. High fidelity recording equipment was used, and all productions were saved in digital format with a sampling rate of 22.05 kHz and 16-bit resolution. Repetitions were elicited for productions that included false starts, self-corrections or hesitations; only fluently-produced utterances were retained.

The multiple steps used to reduce the full set to a single 40-item test will not be described here in detail. The process entailed screening of the full set for low-, medium- and high-comprehensibility speakers by two phonetically-trained research assistants. Subsequently, three complete sets of all 40 sentences were assembled by randomly choosing items from high, medium, and low comprehensibility speakers. These sets were piloted informally on native English listeners to ensure an intelligibility range of 0 to 100% on the items. A final selection of 40 sentences was then made, consisting of 20 true items and 20 false items, each spoken by a different Cantonese speaker.

Listeners

Here we report on six groups of listeners who completed the test. Table 1 provides basic information about each group according to native language, age, and listening conditions. The EA, EG4, and EG groups – all native speakers of Canadian English – were tested to permit an examination of the effects of age on L2 speech comprehension. While EA consisted of adults, EG4 and EG10 were native Canadian English children in grades 4 and 10 respectively. The CA and MA groups were adult native speakers of Cantonese and Mandarin, and the SA (Slavic-speaking) group consisted of adult native speakers of Russian and Ukrainian. The CA group was tested so as to allow a comparison between adults sharing the same L2 accent as the speakers (i.e., Cantonese), with native English listeners (the EA group). Listeners from the CA group were all high-proficiency speakers of English studying at Canadian post-secondary institutions. The MA and SA groups, however, had only intermediate-level English speaking skills.

Table 1.

The Six Listener Groups in the Study

Group	L1	Age	N	Task Conditions
EA	English	Adult	26	sound booth, headphones
EG4	English	8-10 years	22	quiet room, loudspeaker
EG10	English	14-16 years	27	quiet room, loudspeaker
CA	Cantonese	Adult	7	sound booth, headphones
MA	Mandarin	Adult	21	quiet room, headphones
SA	Russian/ Ukrainian	Adult	23	quiet room, headphones

Procedure

Because the test was administered in different locations it was not possible to use identical listening conditions for all participant groups. The EA and CA groups completed the test individually in an audiometric booth in the first author's research lab. Stimuli were presented randomly in a self-paced task through high-quality headphones. After hearing each stimulus item, the listeners responded by selecting "True," "False," or "Don't know" from buttons on a computer screen. They were instructed to use the third choice if they could not understand the item. Once a response was registered, the computer played the next item.

The MA and SA listeners were participants in a larger study reported elsewhere. Because it was impossible for them to visit the same lab used for the EA and CA groups, they were tested individually in a quiet room, wearing headphones. Their task was similar to that of the EA and CA groups, except they used a pen and paper to circle their responses. The children performed the task during group listening sessions in quiet rooms in their schools. Stimuli were presented free-field through high-fidelity audio equipment connected to loudspeakers. After each item, playback was paused so that that the listeners could record their responses on paper by circling one of the same three choices described above. The experimenter controlled the presentation rate to ensure that all listeners stayed in step.

RESULTS

Mean scores for the six groups are presented in Figure 1. The EA group score (M = 87%) serves as a reference for comparison with the other groups. Although the EA group performed well above chance, scores ranged from 75% to 95%. No listener scored perfectly. "Don't know" responses made up 6% of the total, and 7% of the responses were incorrect.

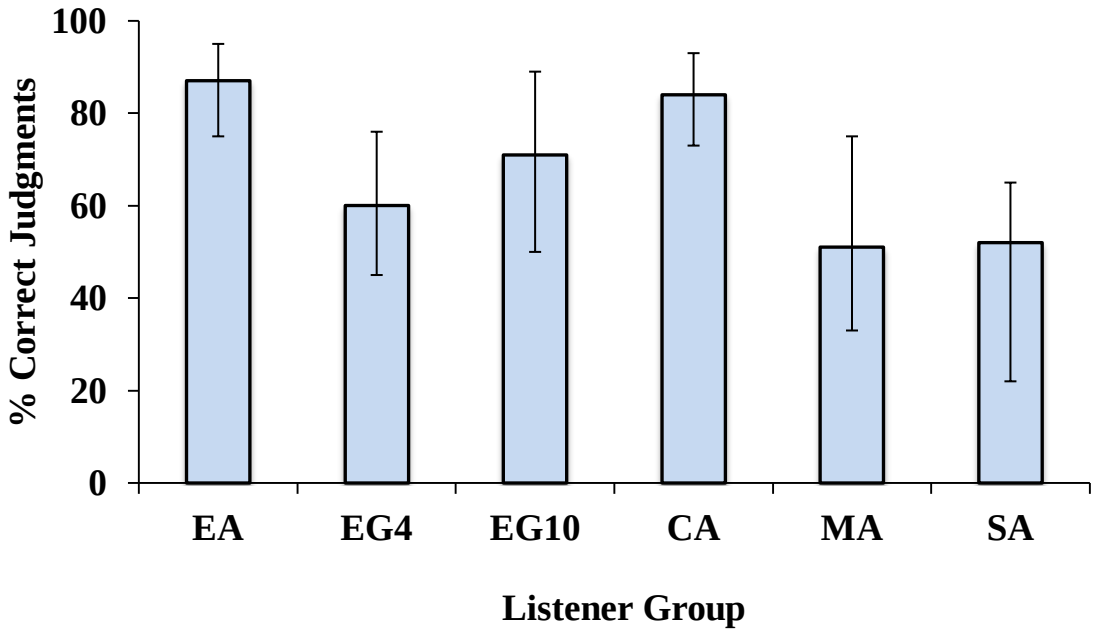

Figure 1. Mean scores for each of the six listener groups: English Adults (EA), English Grade 4 (EG4), English Grade 10 (EG10), Cantonese Adults (CA), Mandarin Adults (MA) and Slavic-speaking Adults (SA). Error bars indicate score ranges.

Effect of age

The effect of age was evaluated by comparing the EG4 and EG10 scores using an independent samples *t*-test. A significant between-group difference, $t(47) = 4.471$, $p < .001$ was observed, indicating better performance on the part of the EG10 group. Although the EA group performed numerically better than both the EG4 and EG10 groups, we did not carry out statistical tests because the listening conditions for the EA group were different than for the two non-adult groups.

Effect of shared L1 background

To determine whether sharing the same L1 background as the speakers offered any benefits for the CA group relative to the EA group, a further independent samples *t*-test was computed. The difference in scores between these two groups proved non-significant, $t(31) = 1.007$, $p = .322$.

Effect of variable non-shared L1 background

The performance of the MA and SA groups was compared in order to determine whether a non-English L1 background in phonologically different languages would affect comprehension. An independent samples *t*-test yielded no significant effect $t(42) = .375$, $p = .710$. Moreover, scores on the 40 items for the two groups were highly correlated, $r(38) = .828$, $p < .0001$. Figure 2 illustrates this relationship.

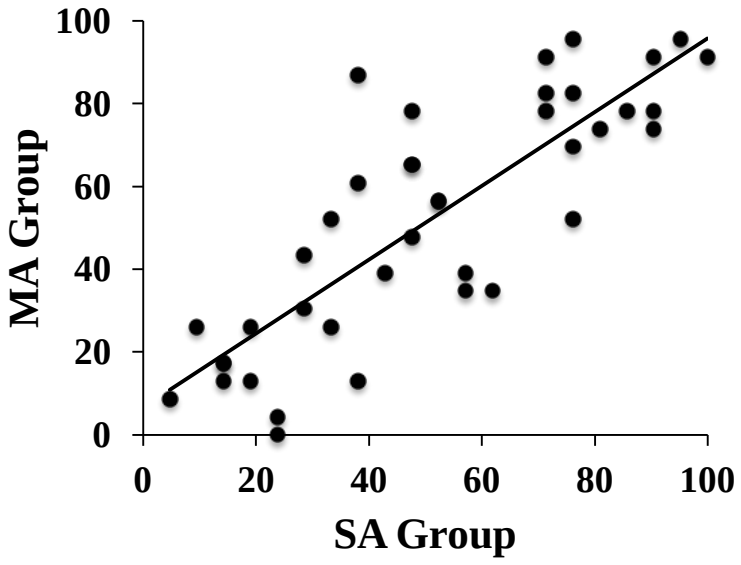

Figure 2. Relationship between the Mandarin (MA) and Slavic (SA) listeners' scores.

DISCUSSION

The results of this investigation point to the feasibility of a verification task as a means of evaluating sources of variability in L2 speech comprehension. In addition to being short and easy to administer, the test was sensitive to between-group differences. The statistically better performance by teens than younger children suggests a beneficial effect of cognitive maturity on the comprehension of accented speech. The (informally assessed) better performance of native English adults than of both groups of children lends additional support to such an account.

As in some previous investigations, sharing an L1 background with the speakers appeared to afford no benefit to the high English-proficiency Cantonese listeners in this study. This outcome is consistent with Hayes-Harb et al.'s (2008) finding that such effects were more evident in lower-proficiency speakers.

Because the Cantonese phonological system is closer to the Mandarin system than to Slavic phonology, one might have predicted better performance by the Mandarin listeners than the Slavic listeners on this test. However, these two listener groups did not differ from each other in terms of mean scores. Moreover, an analysis of item scores revealed a strong correlation between the two groups. Thus, as in previous work (Munro et al., 2006) utterances that were difficult for one group of listeners to understand tended to be comparably difficult for another group from an entirely different linguistic background. This finding supports the view that intelligibility is not merely a highly subjective phenomenon that differs dramatically from one listener to another. Rather, it is often a shared experience for listeners from diverse L1 backgrounds. While intelligibility does not reside entirely within the speech stream, it is also incorrect to assume that it lies completely 'in the ear of the beholder.'

Future work should explore the usefulness of this task in evaluating other sources of listener variability in L2 comprehension. Among the factors worth considering are listener attitudes and amount of L2 exposure. A further matter to investigate is whether some listeners are simply more adept at understanding L2 accents, even when age, L1 background, attitudinal factors, and

exposure are taken into account. If so, research to identify the underlying cognitive influences on L2 speech comprehension will be needed.

ACKNOWLEDGMENTS

This project was completed with invaluable assistance from Marilyn Abbott, Cliff Burgess, Beth Holtby, Susan Morton, Kathleen Nickle, Natasha Penner, and Ron Thomson. Data from the GR4 and GR10 listeners were collected as part of Amy Holtby's MEd project. This research was supported by grants to the first two authors from the Social Science and Humanities Research Council of Canada.

ABOUT THE AUTHORS

A former ESL teacher, Murray Munro is a Professor of Linguistics at Simon Fraser University. His applied phonetics research has appeared in a variety of international journals. Email: mjmunro@sfu.ca

Tracey Derwing is a Professor in the TESL program at the University of Alberta. She has conducted numerous studies of pronunciation and oral fluency development in L2 learners. Email: tracey.derwing@ualberta.ca

Amy Holtby (MEd TESL) is currently teaching at the College of the North Atlantic in Qatar. She is a graduate of the University of Alberta TESL program. Email: aholtby@hotmail.com

Mailing address: Prof. Murray J. Munro, Department of Linguistics, Simon Fraser University, 8888 University Drive, Burnaby BC, Canada, V5A 1S6.

REFERENCES

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America, 114,* 1600–1610.

Burda, A. N., Scherz, J. A., Hageman, C. F., & Edwards, H. T. (2003). Age and understanding of speakers with Spanish or Taiwanese accents. *Perceptual and Motor Skills, 97,* 11–20.

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

Hayes-Harb, R., Smith, B. L., Bent, T., & Bradlow, A. R. (2008). The interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English word- final voicing contrasts. *Journal of Phonetics, 36,* 664–679.

Munro, M. J. (In press) Intelligibility. In Chapelle, C. (Ed.) *The Encyclopedia of Applied Linguistics*. Wiley-Blackwell.

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

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition, 28,* 111–131.

Schiavetti, N. (1992). Scaling procedures for the measurement of speech intelligibility. In R. D. Kent (Ed.) *Intelligibility in speech disorders,* pp. 11–34. Amsterdam: John Benjamins.