

THE INTERLANGUAGE SPEECH INTELLIGIBILITY BENEFIT: THE CASE OF ARABIC-ACCENTED ENGLISH

[Ghazi Algethami](#),¹ University of York

[Dr. John Ingram](#), University of Queensland

[Dr. Thu Nguyen](#)

This study attempted to further examine the so-called *Interlanguage Speech Intelligibility Benefit*, which refers to the intelligibility advantage L2 listeners have over native listeners when they listen to L2 speech produced by speakers who share the same native language. 19 native speakers of Australian English and 19 L2 speakers of English whose native language is Saudi Arabic listened to English utterances produced by ten L2 Saudi speakers of English who were of two groups; high pronunciation proficiency and low pronunciation proficiency. Utterances from three native speakers of Australian English were also included as controls. The listeners were asked to transcribe what they thought the speaker had said in English orthography. The percentage of words identified correctly by each listener for each utterance is taken as the intelligibility score. Although the L2 Saudi listeners did slightly better than the native listeners when they listened to Saudi accented English, both listener groups did not differ significantly. The results corroborate previous findings suggesting that the intelligibility advantage for the non-native listeners is small, if any, and the phonetic properties of the L2 speech itself are strong determinants of how it is perceived regardless of the listeners' native language.

INTRODUCTION

It has long been suggested that non-native listeners have an intelligibility advantage over native listeners when they listen to non-native speech produced by L2 speakers who share the same L1 (Weinreich, 1953). This means that the non-native listeners could do better than the native listeners at understanding L2 speech produced by those who share with them the same L1. Recent studies have attempted to test this intuition empirically (Bent & Bradlow, 2003; Hayes-Harb, Smith, Bent, & Bradlow, 2008; Imai, Walley, & Flege, 2005; Ingram & Nguyen, 2007; Major, Fitzmaurice, Bunta, & Balasubramanian, 2002; Munro, Derwing, & Morton, 2006; Stibbard & Lee, 2006; van Wijngaarden, 2001; van Wijngaarden, Steeneken, & Houtgast, 2002). Generally, these studies found very small and inconsistent evidence for this advantage.

For example, Stibbard and Lee (2006); Major et al. (2002); and Bent and Bradlow (2003) showed that L2 listeners did not outperform native listeners in recognizing words produced by L2 speakers who share the same L1 background. In contrast, Imai et al. (2006) reported that L2 Spanish listeners yielded higher intelligibility scores than native listeners when they were both presented with Spanish-accented English. Munro et al. (2006) found that while the Japanese listeners found Japanese-accented English more intelligible than did the native listeners, the

¹ Authors ordered by percentage of contribution to the current study.

Cantonese listeners did not report such an advantage when listening to Cantonese accented English.

It was also suggested that this intelligibility advantage might be mediated by the pronunciation proficiency level of the L2 speakers, which has been mostly operationalized as the degree of foreign accent they speak with (Hayes-Harb et al., 2008). For example, Hayes-Harb et al. (2008) showed that low proficiency L2 Mandarin listeners did better than high proficiency L2 Mandarin and native English listeners when they listened to speech produced by low proficiency Mandarin speakers.

Bent and Bradlow (2003) called this advantage *The Interlanguage Speech Intelligibility Benefit*. However, it should be noted that they used it to mean that the native and non-native listeners were equal in understanding the non-native speech. Stibbard and Lee (2006) argued, however, that the non-native listeners have to outperform their native counterparts in understanding the L2 speech to be called an advantage. The current study follows Stibbard and Lee.

A major limitation in most previous studies is that they used speech which was elicited from L2 speakers by explicit tasks of reading and pronunciation assessments, a method which would encourage the L2 speakers to monitor their production and hence undermine some of the L1 interference, which has been assumed to enhance the intelligibility for the non-native listeners who share the same L1. However, reliance on read speech is almost inescapable as it offers control over the content to be compared. An exception to this limitation is Munro et al. (2006). They extracted utterances from narratives obtained by asking participants to describe extemporaneously a cartoon story. Nevertheless, this method does not offer us complete control over what lexical items would be included in the utterances to be presented to the listeners. Thus, differences between intelligibility scores given to the listeners might be affected by the differences between the utterances. In addition, Munro et al. (2006) did not examine how the proficiency level of the L2 speakers or the degree of their foreign accents affects their intelligibility to the L2 listeners who share the same L1.

The current study implemented an elicitation method, first used in Ingram and Thu (2007), to both offer control over the content or the lexical items to be included in the utterances and to deflect the L2 speakers from monitoring their L2 production. The speakers were asked to paraphrase a number of English sentences within a specified time-duration, and then write and read out their responses into a microphone. The purpose of the time-allotted paraphrase task was to place a moderate cognitive load on the L2 speakers so that they would be preoccupied with formulating the sentences rather than with monitoring their pronunciation. It also offered a way to control the lexical items to be included in the listening task. The complexity of the paraphrase task was sufficient to engage the L2 speakers and deflect their attention from the pronunciation side of the task.

Ingram and Thu (2007) did not find an intelligibility benefit for L2 Vietnamese learners of English when they listened to Vietnamese accented English, but rather a disadvantage as the native English listeners found the Vietnamese accented English more intelligible than did the Vietnamese L2 listeners. However, they did not examine the effect of the proficiency level of the L2 speakers which has been shown to affect the degree of intelligibility among listeners (Hayes-Harb et al., 2008).

The current study examines whether L2 Saudi listeners outperform native English listeners in understanding Saudi Arabic-accented English. A few studies have examined the production of English speech by L2 Saudi speakers. Flege and Port (1981) studied the production of English stop consonants by L2 Saudi speakers. Their results generally showed that the L2 Saudi speakers used their L1 Arabic temporal acoustic correlates of stop consonants when they were asked to produce English stop consonants. However, except for /p/, they showed that this acoustic deviation from the native norms did not pose much difficulty for native listeners when they were asked to identify which stop consonant was produced. Munro (1993) studied the production of English vowels by 23 L2 Arabic speakers, of whom only three were Saudis. He pointed out that “the acoustic properties of Arabic-accented vowels were: exaggerated tense-lax duration differences, small voicing-conditioned duration differences, a very short /a/, a tendency toward a low F1 in /ε/ and /a/, a tendency toward a low F2 in most back vowels, and generally less formant movement in F1 and F2 than in the native speakers’ productions” (p. 51). However, it is somewhat difficult to attribute these acoustic properties to the Saudi-accented English because the participants were of different Arabic dialects which are phonologically different (Watson, 2002). Other studies have used only auditory analysis to examine the English pronunciation mistakes made by L2 Saudi speakers. Altaha (1995) pointed out some phonemes which were difficult for his students to produce: /p/, /v/, /ε/, /ɜ:/ and /æ/. Basalamah (1990) showed that his L2 Saudi participants used their L1 Saudi Arabic stress pattern when speaking English. Except for Flege and Port (1981), none of these studies has shown how these deviations from the native norms observed in the speech of L2 Saudi speakers would affect their intelligibility. It is also worth pointing out that deviations from the native phonetic norms perceived in the speech of second language speakers could cue our perception of foreign accentedness, but they might not necessarily reduce its intelligibility (Derwing & Munro, 1997). Therefore, it is still not clear what pronunciation mistakes or deviations from the native speech would make the L2 speaker harder to understand. The investigation of these mistakes is beyond the scope of the current study, though.

Before describing the study, some related terminological issues should be explained. Derwing and Munro (1997) defined *intelligibility* as the actual understanding of the message. Most previous studies measured *intelligibility* by counting the number of correct words identified by the listeners. However, the correct identification of words does not necessarily mean *understanding the message* as listeners might be able to identify all the words while still puzzling over what the speaker is trying to communicate (Zielinski, 2004). For convenience and consistency with previous studies, the current study uses *intelligibility* to mean the correct identification of the words uttered by the speaker.

The pronunciation proficiency level of L2 speakers was used in some previous studies to mean the degree of foreign accent they speak with (Hayes-Harb et al., 2008; Imai et al., 2005; Stibbard & Lee, 2006). In these studies, L2 speakers were grouped into two subgroups according to the degree of foreign accent. Generally, those with stronger foreign accents were categorized as *Low Proficiency* L2 speakers and those with less foreign accent were categorized as *High Proficiency* L2 speakers. However, the term *proficiency* in pronunciation might also entail other dimensions that are not directly related to foreign accentedness, such as fluency and grammaticality (Derwing & Munro, 1997). For the sake of consistency with previous studies, the current study uses the terms *high proficiency* and *low proficiency* to refer to the degree of foreign accentedness, as measured by the ratings given by the native listeners in this study.

METHOD

First, a description of how the stimuli used in the listening test were constructed is given. Then, the construction and delivery of the listening test which provided the data for the current study are described.

Stimulus Construction

Twenty three English sentences were adopted from Ingram and Thu (2007) (see Appendix A). There was no basis for choosing these utterances other than that of convenience.

Speakers

The speakers who supplied the listening materials for the present study were full-time international Saudi postgraduate students at Australian universities in Brisbane. They were in the age range of 20-35 and ranged in residency in Australia from a period of 1 to 4 years. All of them had attained written and spoken English proficiency that enabled them to undertake Master and PhD degrees at the University of Queensland, Griffith University and Queensland University of Technology. Three native speakers of Australian English were also recruited to provide native English utterances to be used as controls.

Speech Elicitation Method

The current study attempted to create a speaking situation where there is a moderate cognitive load involved and where the speakers are more preoccupied with the formulation of the message rather than with self-monitoring of their pronunciation. A grammatical paraphrase task was found to meet these requirements and at the same time provide control over the lexical selection of items among the speakers.

Grammatical Paraphrase Task

The grammatical paraphrase task required subjects to transform sentences, which were constructed previously, presented in spoken and written form (over headphones and a computer screen) into a meaning-equivalent form. The materials were presented via a spoken Language Assessment Program (Harrington & Ingram, 2003). The participants typed in the paraphrase in response to an initial prompt word and when satisfied with their construction, read out the sentence that they had formed. The linguistic aspects of the task were sufficiently complex to engage the subjects (who were all L2 speakers of English, with the exception of three controls) and to deflect their attention from the pronunciation aspect of the task. A sample of this production is presented below:

Example Stimulus: *The soldier's face and mouth were covered by a mask.*

(Audio + written)

Visual Prompt: *A mask*

Paraphrase response: *A mask covered the soldier's face and mouth* (typed, and then spoken).

After speaking their paraphrase responses into a headset microphone, subjects pressed a button for presentation of the next item in the set, randomly selected without replacement until all the items had been presented. The typed response and the audio signal were saved to a database from which the listening items were selected. Because some of the L2 speakers were not able to come to the phonetic lab at the University of Queensland, and access to the Language Assessment Program (Harrington & Ingram, 2003) was not possible, recording took place in a quiet furnished room in their homes using a personal laptop. The paraphrase test was presented via a word document, where the participants were asked to click on each sentence to hear it, paraphrase it according to a prompt word, and then read it aloud into a microphone. The participants were told that they had thirty minutes to complete the test.

Construction of the Listening Test

Most of the participants for the listening test were drawn from an introductory linguistics class at the University of Queensland. They participated in the listening experiment for course credit. More L2 Saudi participants in the listening test were needed. They were drawn from the Saudi Students Society in Brisbane. All of them were full-time postgraduate students at the University of Queensland, Griffith University and Queensland University of Technology. Their range of residency in Australia was from 1 to 4 years. The experimental task was time-constrained (approx. 20 minutes) to avoid fatigue and flagging interest.

Due to likely practice effects if any test utterance was heard repeatedly in the course of the experiment, it was necessary to block items in such a way that no listener heard a given utterance more than once (with the exception of four control utterance at the end). It was also necessary to expose each listener to the full range of speaker variation in the test sentence set.

For this reason, each listener was assigned to one of 18 overlapping blocks of 23 items. The four control items (spoken by native speakers of Australian English) which involved repetition of test utterances were always presented as the last four items in a block, so the perception of accented English items would not be affected by previous exposure to the same utterance produced by different speakers.

Subjects for the Listening Experiment

The first group of listeners was comprised of native-born speakers of Australian English (N=19), who reported no familiarity with Arabic-accented English. The second group of listeners was made up of Saudi students whose native language is Saudi Arabic (N=19).

Procedure for the listening experiment

Participants in the listening experiment were randomly and evenly allocated to one of 18 blocks of items (versions of the experiment). As there were 19 listeners of each group and 18 blocks of utterances, two listeners in each group had to listen to the same block of utterances. The eighteen overlapping blocks of the listening test were needed because no listener could hear a given utterance more than once in an attempt to expose each listener to the full range of speaker variation in the test sentence set and to avoid the effect of repetition in the listening experiment. A version of the experiment and instructions for undertaking it were e-mailed to the subjects as word-files containing links to sound-files for each utterance uploaded in the University of

Queensland web-domain (Appendix B). Subjects had the option of doing the experiment at their time of choosing, using the multi-media facilities of their home computers or using a University machine. Response sheets were e-mailed to the experimenters.

Subjects were instructed that they could play each item up to four times. They were told to transcribe what they thought the speaker said into standard orthography and then to rate the original utterance for accentedness on a 5-point scale: (1) no detectable foreign accent, (2) mild accent, (3) moderate accent, (4) strong accent, (5) very strong accent.

RESULTS

The percentage of words identified correctly by each listener for each utterance was taken as the intelligibility score. Errors including word omissions, word substitutions, and incorrectly transcribed inflections, but excluding spelling errors, were also counted but not included in the analysis for this study. The L2 Saudi speaker group was divided into two groups- high proficiency group and low proficiency group- based on mean ratings of the degree of foreign accent given to them by the native listeners. L2 speakers who received mean rating of 2.5 or less were considered high-proficient while those who received mean rating above 2.5 were considered low-proficient. The non-native listeners' ratings of foreign accentedness were not taken into account when dividing the L2 speakers into high and low proficiency groups as the reliability of their ratings might be affected by their L1. Figure 1 shows the mean foreign accent ratings given to both the native and L2 speakers by the native listeners.

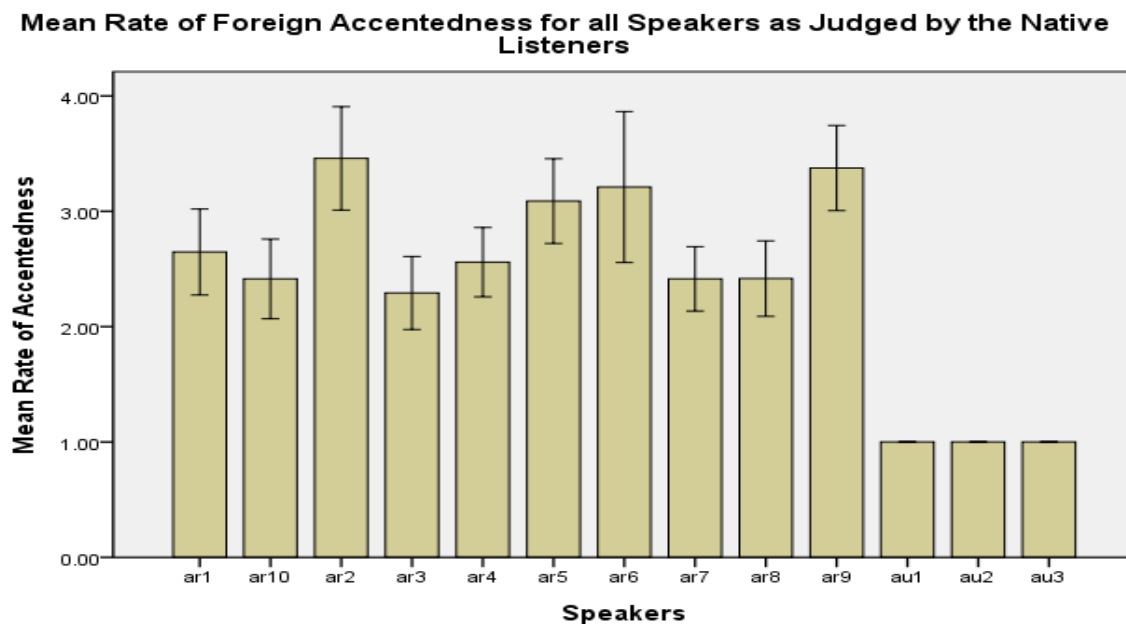


Figure 1. Mean rating of the degree of foreign accent for both the native speakers (au) and the L2 speakers (ar).

Unless stated, the significance level reported here is $p < .05$. A two-way analysis of variance with listener group (two-levels: Saudi and English) as a between-subject factor and speaker

group (Native: N; High Proficient: HP; and Low Proficient: LP) as a within subject factor showed a significant main effect of speaker group ($F(2,873) = 7.574, p < .05$), and no significant effect of listener group ($F(1,873) = 2.846, p > .05$). However, there was a significant interaction between listener group and speaker group ($F(2,873) = 5.682, p < .05$).

Post-hoc pairwise t-tests showed no significant difference between the two listener groups when they listened to both high and low proficiency L2 Saudi speakers of English. In contrast, the two listener groups differed significantly when they listened to the native English utterances. Table 1 shows the average intelligibility scores given for each listener group when listening to both the native and non-native speech. Figure 2 also shows the results graphically.

Table 1. Average intelligibility scores given for each listener group

Listener	Speaker	% Intelligibility Score
Saudi	LP	91.722
	HP	94.494
	Native	92.533
Native	LP	90.478
	HP	93.752
	Native	100.000

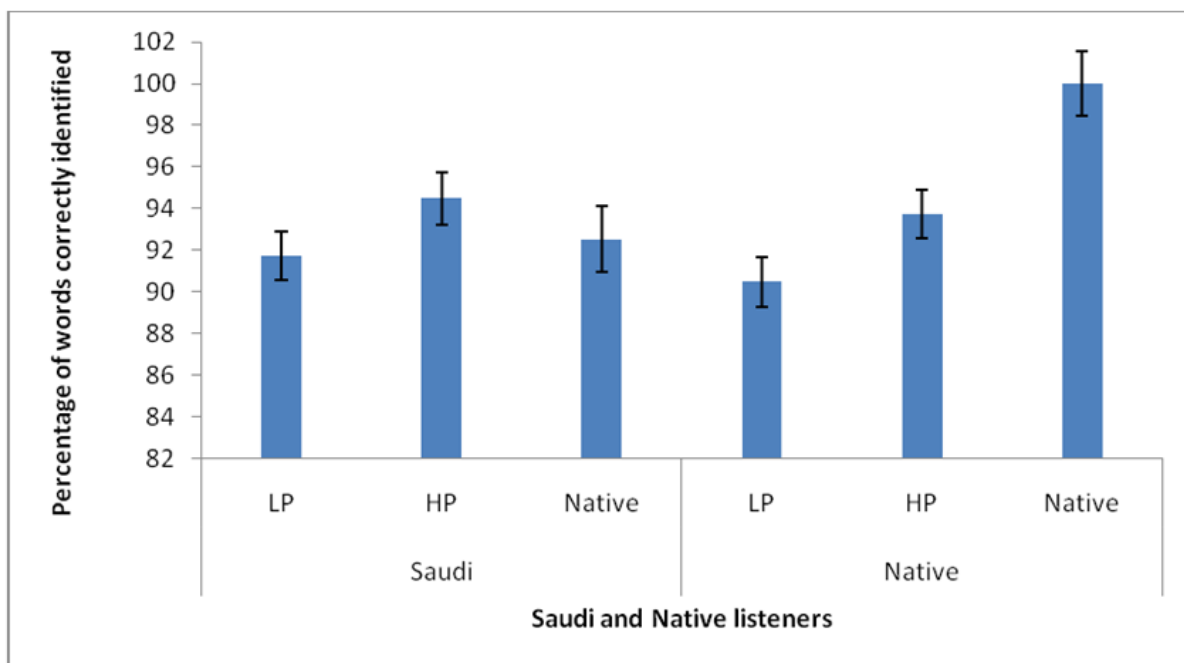


Figure 2. Average intelligibility score for both listener groups for each speaker group.

DISCUSSION AND CONCLUSION

Previous studies found small or no intelligibility advantage for non-native listeners over native listeners when they listened to L2 speech from speakers who share the same L1. It has been suggested that the L2 speakers who share the same native language also share the same interlanguage, a fact assumed to offer them an intelligibility advantage over the native listeners (Bent & Bradlow, 2003). We hypothesized that previous studies did not find such an advantage because the speech stimuli used in the listening experiments were elicited by explicit ways of reading and pronunciation assessments. The current study used a methodology to try to deflect the L2 speakers from monitoring their speech, a fact that might undermine some of the L1 interferences which might facilitate the intelligibility of L2 speech for L2 listeners who share the same L1.

Although the non-native listeners yielded higher intelligibility scores than the native listeners for both L2 speaker groups, the difference was quite small and did not reach statistical significance. This gives further credence to the previous studies that showed small or no intelligibility benefit for non-native listeners when they listened to L2 speech produced by L2 speakers who share the same native language.

One might conclude that despite the degree of foreign accent L2 speakers have, there are certain aspects of L2 speech that seem to be detrimental to intelligibility for both native and non-native listeners. Future research on L2 speech must seek what these features are (Munro, 2008). These features will be appreciated by L2 teachers and curriculum designers as intelligibility places a high priority in teaching and learning a second language.

Although it is not the purpose of the current study, it might be useful to mention some of the pronunciation mistakes made by the L2 Saudi speakers which impeded intelligibility. We noticed that they, especially the low proficiency group, tended to produce /b/ as /u/ or sometimes /o:/. Both L2 groups showed less diphthongization than the native speakers when they produced diphthongal segments. The stop consonant /p/ was in many cases produced as /b/. Stress placement and timing also played a role in reducing the intelligibility of the L2 speech.

ABOUT THE AUTHORS

Ghazi Algethami is a PhD candidate at the University of York. His general research interest is the production and perception of second language speech.

Affiliation: University of York.

Address: United Kingdom, York, Heslington, YO10 5DD.

Email: ghazi.algethami@gmail.com

Dr. John Ingram is a senior lecturer at the University of Queensland. His research interests include Phonetics, Phonology, Second language speech, Psycholinguistics, and Neurolinguistics.

Affiliation: University of Queensland.

Address: Australia, University of Queensland, School of Languages and Comparative Cultural Studies, St Lucia, 4072.

Email: j.ingram@uq.edu.au

Dr. Thu Nguyen's research interests include second language speech, Vietnamese accented English, L2 prosody, and Vietnamese phonology.

Email: nguyenthianhthu@email.com

REFERENCES

- Altaha, F. (1995). Pronunciation errors made by Saudi university students learning English: analysis and remedy. *International Review of Applied Linguistics*, 109-110, 110-123.
- Basalamah, M. (1990). *A Study of Certain Aspects of the Suprasegmentals of Arabic and their Influence on the English Pronunciation of Arab Teachers of English in the Western Zone of Saudi Arabia*. Unpublished PhD thesis, University College London, London.
- Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America*, 114, 1600-1610.
- Derwing, T. M. & Munro, M. J. (1997). Accent, comprehensibility and intelligibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 20, 1-16.
- Flege, J. E. & Port, R. (1981). Cross-language phonetic interference: Arabic to English. *Language and Speech*, 24(2), 125-146.
- Hayes-Harb, R., Smith, B., Bent, T., & Bradlow, A. (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.
- Imai, S., Walley, A. C., & Flege, J. E. (2005). Lexical frequency and neighbourhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. *Journal of the Acoustical Society of America*, 117, 896-907.
- Ingram, J. & Nguyen, T. (2007). *Intelligibility of Vietnamese accented English*. Paper presented at the TESOL in the internationalization of higher education in Vietnam, Hanoi, Vietnam.
- Harrington M. & Ingram J. (2003). LanguageMAPTM: A multi-media language assessment program, <http://www.languagemap.com>.
- Major, R., Fitzmaurice, S., Bunta, F., & Balasubramanian, C. (2002). The effects of nonnative accents on listening comprehension: Implications for ESL assessment *TESOL Quarterly*, 36(2), 173-190.
- Munro, M. (2008). Foreign accent and speech intelligibility. In J. Edwards & M. Zampini (Eds.), *Phonology and second language acquisition* (pp. 193-218). Philadelphia: John Benjamins.

- Munro, M., Derwing, T., & Morton, S. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, 28, 11-131.
- Stibbard, R. M., & Lee, J.-I. (2006). Evidence against the mismatched interlanguage intelligibility benefit hypothesis. *Journal of the Acoustical Society of America*, 120, 433-442.
- van Wijngaarden, S. J. (2001). Intelligibility of native and non-native Dutch speech. *Speech Communication*, 35, 103-113.
- van Wijngaarden, S. J., Steeneken, H. J. M., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *Journal of the Acoustical Society of America*, 111, 1906–1916.
- Watson, J. (2002). *The phonology and morphology of Arabic*. Oxford: Oxford University Press.
- Zielinski, B. (2004). *Measurement of speech intelligibility: What are we actually measuring?* Paper presented at the Conference of American Association for Applied Linguistics, Portland, USA.

Appendix A

Australia is the world's driest continent
 It is cold soup
 Jack tried to steal the jewels
 Ann will paint the room
 The strong box was where John kept his money
 The window will be washed by Nick
 Because it was dark John stumbled at the blackboard
 Elephants used to be hunted for their tusks and hides
 John was admired by the other students
 The room will be painted by Ann
 Without his glasses, John could not see the blackboard
 Alan folded the newspapers
 Kim wanted a box of chocolate, not a bunch of flowers
 The other students admired John
 Dark clouds were gathering as they came across the bridge
 He needs a strong box to carry all those books
 The bluebells and tulips were hung from the eaves of the green house
 The soldiers face and mouth were covered by a mask
 The plate ran away with the spoon
 Bill's life savings were invested in a friendly society
 Where the queen was sleeping was the royal tent
 What they wanted was to migrate to a friendly society

Appendix B

Accented English Experiment - Group 1 (version 1)

The aim of this experiment is to assess the effects of non-native accent upon spoken sentence comprehension. We are interested in finding out how the strength of a non-native accent is related to how easy/difficult particular spoken sentences are to understand. Please listen to each sentence (up to, but **not** more than **four** times). Type what you think that the speaker said in the space provided. Then rate the sentence for the strength of the speaker's non-native accent. Type in the words you think you recognize even if the sentence as a whole may not make sense to you.

Your last name:	First name:
Student number:	Gender (M, F): Age:
Your native language:	Mother's native lang.:
Father's native lang.:	Years lived in Australia:
Other languages you speak/understand well:	
Places you have lived for more than 6 months outside Australia:	
Main language spoken at home:	
Accented Englishes that you are familiar with hearing (check with X): Chinese () Thai () Vietnamese () Japanese () Indian () Arabic () Other () None of these ()	

Please fill in the following information:

Instructions:

The first column of each row in the table below contains an **Item**, a short spoken sentence. If you hold down the **Ctrl** key (bottom left of your keyboard) and move the mouse cursor over the item a small hand will appear. Right click on the cursor to hear the item. You may play each item *up to four (4)* times, but **not** more than that. When you have finished listening to an item, close the ‘sound play’ window and answer the following questions on the questionnaire:

In the first column type in what you think the speaker said, or at least those words that you thought you heard, even if the sentence as a whole does not make sense.

In the second column make a rating as to how strong the speaker’s accent appeared to be:

Strength of non-native English accent

(1) no foreign accent (2) mild accent (3) moderate accent (4) strong accent (5) very strong accent

When doing this experiment, please work entirely on your own. Work through each item in the order given in the table. Having completed an item, move to the next one. Do not go back and change any of your answers. Do not play any item more than four times. It is your first listening impression that we are after. So do not spend much time on any given item.

Item No.	Type what you think that the speaker may have said.	Accent Rating
item01		
item02		
item03		
item04		
item05		
item06		
item07		
item08		
item09		
item10		
item11		
item12		
item13		
item14		

Item No.	Type what you think that the speaker may have said.	Accent Rating
item15		
item16		
item17		
item18		
item19		
item20		
item21		
item22		
item23		

Thank you!

When you have finished filling in this form, please save it with a file name that includes your name and student id: (e.g.: 'accent_xpt_Joe_Bloggs_129457').

Having saved the file, please send it back as an email attachment.