The EMIME Mandarin Bilingual Database

Mirjam Wester and Hui Liang. The EMIME Mandarin Bilingual Database. Technical Report EDI-INF-RR-1396, The University of Edinburgh, 2011.

The EMIME Mandarin Bilingual Database

Mirjam Wester[1], *and Hui Liang*[2]

[1]Centre for Speech Technology Research, University of Edinburgh, UK
[2]Idiap Research Institute, Martigny, Switzerland
mwester@inf.ed.ac.uk, hui.liang@idiap.ch

Abstract

This paper describes the collection of a bilingual database of Mandarin/English data. In addition, the accents of the talkers in the database have been rated. English and Mandarin listeners assessed the English and Mandarin talkers' degree of foreign accent in English.

Index Terms: evaluation, accent rating, cross-lingual

1 Introduction

The motivation for recording a bilingual database arose from the EMIME speech-to-speech translation task. In this project, we are aiming for personalized speech-to-speech translation such that a user's spoken input in one language is used to produce spoken output in another language, while continuing to sound like the user's voice[1]. However, how do we measure whether our modeling attempts are successful or not - that is how are we to measure whether or not a user sounds similar in two different languages? Aside from the complications associated with asking listeners to compare natural speech to synthetic speech there is an even more fundamental question we would like to see answered first. How well do listeners judge speaker similarity across language boundaries when the stimuli consist of natural speech. To investigate this we needed a database of bilingual data. This paper describes the design and collection of this database.

In designing the bilingual database for our talker discrimination experiments we have two assumptions. First of all, we assume talker discrimination is easier when the different languages spoken by individual talkers are from the same language family. That is, listeners should be able to judge more accurately whether or not a talker is the same when the talker is speaking two closely related languages. Secondly, if bilingual talkers are highly fluent in their two languages, talker discrimination should be more difficult. Anecdotal evidence seems to suggest that proficient non-native talkers of English do not necessarily sound like the same person when speaking their native language.

The languages under consideration in EMIME are Japanese, Mandarin, Finnish and English. In [1] the English/German and English/Finnish portions of the EMIME database are described. This report covers the English/Mandarin portion of recordings. The aim of the experiment described in this paper is to select talkers with the least degree of perceived foreign accent[2] because, as stated above, we expect that the more native the bilingual talker sounds in English, the more difficult it will be for listeners to recognize them as the same talker in both their native language (L1) and their second language (L2). This paper addresses the following question. Which Mandarin talkers in the EMIME database have the least degree of perceived foreign accent?

[1]http://www.emime.org
[2]The definition we take of accent is a manner of pronunciation of a language.

2 Data collection

2.1 Stimulus materials

English and Mandarin prompts were used. Each set contains 25 Europarl sentences, 100 English news sentences and 124 Mandarin news sentences, respectively, and 20 semantically unpredictable sentences (SUS). The 25 Europarl sentences were selected from the ACL WMT 2008 test set of the Europarl (proceedings of the European Parliament) parallel corpus [2], for Mandarin these sentences were translated from English. The news sentences for English were taken from the Wall Street Journal 1 corpus [3], comprising 40 enrolment sentences and 60 test set sentences. The Mandarin sentences were selected from the Speecon corpus [4]. The 20 SUS for English and Mandarin were taken from the Blizzard 2009 set [5].

2.2 Talkers

In total 17 talkers were recorded. Seven male and ten female talkers The bilingual talkers were recruited via the Edinburgh University Careers Services. Bilingual in this context means a person who has the ability to speak and read two languages.

The same native English talkers as in [1] were used. Work by Flege & Fletcher [6] has shown that it is important to include native talkers in a non-native accent rating task. It was found in [6] that the degree of foreign accent is influenced by the proportion of native (or near-native) speakers included in the test set. Increasing the proportion of native speakers in the stimulus set caused the non-native speakers to be rated as more accented. It also ensured that listeners used a wider range of the rating scale. Furthermore, native controls serve to confirm that listeners are correctly performing the task by testing that they can distinguish native from non-native speech.

2.3 Recording procedure

Recordings were carried out using ProTools HD hardware and software in a hemi-anechoic chamber. Two different microphones were used, a close-talking DPA 4035 mounted on the subjects headphones and a Sennheiser MKH 800 p48 microphone placed about 10cm from the subject using an omnidirectional pattern. The speech was sampled at 96kHz 24bit depth and stored directly to a computer. These recordings were subsequently downsampled, using Pro-Tools to 22 kHz 16bit and segmented into sentence level chunks.

The recordings took between 1 and 2 hours per person to complete. The bilingual talkers first read the English sentences and then their native language with a short break between the two sessions. When they made an error, the talkers were asked to re-read the sentence. A remuneration of £20 was given to the bilingual subjects for their time and effort. Two of the female Mandarin talkers were more fluent in reading English than reading Mandarin and chose to first read the Mandarin prompts followed by the English. One of the female participants was not a competent reader of Mandarin which wasn't spotted until she started reading the prompts. We thanked her for her time and didn't continue recording after about 20 minutes.

For the accent rating experiment, three of the female Mandarin talkers' data was not included. The talkers were omitted for the following reasons: Mandarin not fluent enough, Mandarin data wasn't recorded, one of the talkers had a rather distinctive voice.

3 Accent rating experiment

This section describes the accent rating experiment: the materials used, the design of the experiment and the participants.

3.1 Stimuli

The accent rating experiment was performed on English sentences. Three sentences (short, medium, long) were selected. For all talkers the same sentences were used. The selected sentences are:

- Sometimes it helps to take a step back. (9 syllables)
- A second meeting is reportedly scheduled for today. (15 syllables)
- Microbiology is the study of organisms that cannot be seen by the naked eye. (25 syllables)

3.2 Design

There are six different test conditions:1) Mandarin and English females, 2) Mandarin and English males, 3) Finnish and English females and 4) Finnish and English males, 5) German and English females, 2) German and English males. (for more details on the German and Finnish data see [1])

Each test condition consists of 84 trials: 14 talkers x 3 sentences x 2. Within a test condition there are six blocks of 14 talkers reading the same sentence. The order of the talkers is different for every block to control for any possible effect of talker order. There are six different orders for the three sentences. For each of the four testing conditions six versions were generated to control for sentence order. Each listener was assigned a different selection of the six test conditions, always alternating between male and female sets, always including Mandarin female and male test sets as well as then either the Finnish test sets or the German test sets.

3.3 Listeners

Accent ratings were collected from 12 native monolingual English listeners and 12 Mandarin listeners. Listeners were recruited at the University of Edinburgh. None of them had any known hearing, speech or language problems. Subjects were paid for their participation.

The subjects were asked to score the degree of foreign accent for each utterance on a scale from 0 to 6, with "0" = no foreign accent at all and "6"= strong foreign accent. They were told the native speakers were from various different English speaking backgrounds, and that none of these native English accents should be considered foreign.

The listening experiments in Edinburgh were carried out in sound isolated booths. Audio was presented from a Mac mini computer using Beyerdynamic DT 770 PRO headphones. The experiment was conducted via a web interface. The subjects' task was to click on an audio file, listen to the sentence stimulus and click with a mouse on one button in the range from 0 to 6 to indicate their judgement of degree of accent. Subjects were free to listen to the utterance as often as he/she needed to to make a judgement. The experiment can be found at `http://homepages.inf.ed.ac.uk/mwester/accent_rating/start_accent.html`

4 Results

This section shows the results of the accent-rating task. Only results for Mandarin and English talkers are presented, from the English listeners. Listener ratings were converted to normalized z-scores. The boxplot in Figure 1 shows the overall z-score results for the female talkers in the top half and the male talkers in the bottom half, as judged by all English listeners. Larger z-score ratings indicate a greater degree of foreign accent. Abbreviations in Figure 1 and the rest of this paper are: English female = EF, Mandarin female = MF, English male = EM, Mandarin male = MM.

On the basis of the data in Figure 1 the selected female talkers for the talker discrimination experiments are talker MF1, MF2, MF4, MF5 and MF7. For the male talkers we can see that there is a smaller degree of variation in mean degree of foreign accent for the Mandarin male talkers compared to the other talker groups

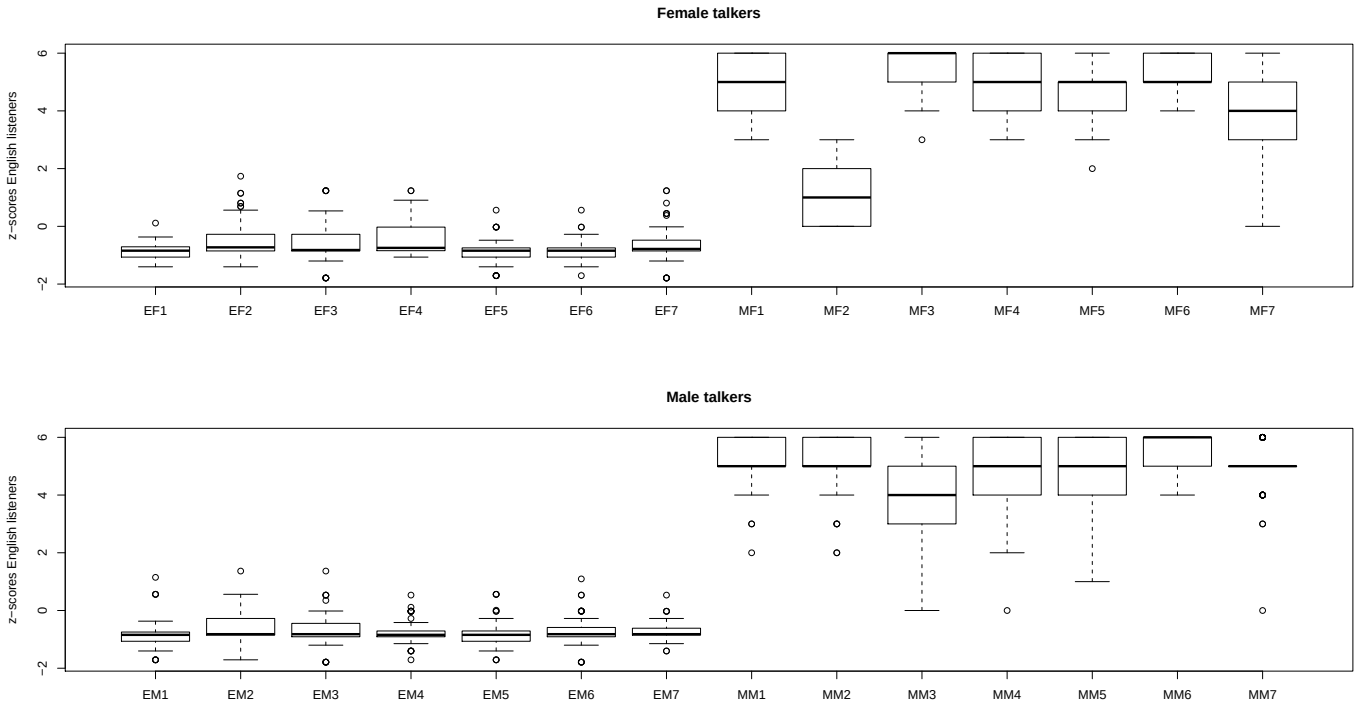

Figure 1: Female and male talkers' z-scores based on 12 English listeners' judgements.

[1]. All male Mandarin talkers are judged to have a strong foreign accent. The male talkers selected for further discrimination experiments are MM1, MM3, MM4, MM5, and MM7.

5 Discussion and Conclusions

This paper has described the collection of "The EMIME Bilingual Mandarin/English Database". This database has been made available under the Open Database License: http://opendatacommons.org/licenses/odbl/1.0/. In addition, the results of an accent-rating task have been reported. The set of five male and five female Mandarin talkers with the least degree of foreign accent as judged by native monolingual English listeners have been highlighted.

6 Acknowledgements

The research leading to these results was partly funded from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement 213845 (the EMIME project).

References

[1] M. Wester, "The EMIME Bilingual Database," The University of Edinburgh, Tech. Rep. EDI-INF-RR-1388, 2010.

[2] P. Koehn, "Europarl: A parallel corpus for statistical machine translation," in *MT Summit 2005*, 2005.

[3] D. Paul and J. Baker, "The design for the Wall Street Journal-based CSR corpus," in *Proceedings of the DARPA Speech and Natural Language Workshop*, 1992.

[4] D. Iskra, B. Grosskopf, K. Marasek, H. van den Heuvel, F. Diehl, and A. Kiessling, "Speecon-speech databases for consumer devices: Database specification and validation," in *Proc. LREC*, 2002.

[5] S. King and V. Karaiskos, "The Blizzard Challenge 2009," in *Proc. Blizzard Challenge Workshop*, Edinburgh, U.K, 2009.

[6] J. Flege and K. Fletcher, "Talker and listener effects on degree of perceived foreign accent," *J. Acoust. Soc. Am.*, vol. 91, no. 1, pp. 370–389, 1992.