Mutual intelligibility of American, Chinese and Dutch-accented speakers of English

Hongyan Wang and Vincent J. van Heuven

Phonetics Laboratory,
Universiteit Leiden Centre for Linguistics, Leiden, The Netherlands
{h.wang,v.j.j.p.van.heuven}@let.leidenuniv.nl

10.21437/Interspeech.2005-356

Abstract

This paper presents the results of a comprehensive study of the mutual intelligibility of Chinese, Dutch (both foreign-language learners) and American (native language) speakers of English. Intelligibility is tested at the level of the segment, word and sentence, after careful selection of representative speakers from the three language backgrounds. The results show that production and perception skills are generally correlated at all levels, that both speakers and listeners are more successful in the order Chinese < Dutch < American. Against this background, however, intelligibility is unexpectedly good when speakers and listeners share the same mother tongue.

1. Introduction

We are interested in the communicative problems that crop up when members of speech communities with other native languages than English communicate with each other using English as a lingua franca. Such situations typically arise when delegates from all over the world meet at an international conference, such as Interspeech, where the official language is English. If the listener is not a native speaker of English, are native English speakers easier to understand than non-native speakers? Is the non-native listener at an advantage when he listens to English spoken by someone who has the same mother tongue as the listener? Is it true that non-native English by speakers of a language that is genealogically close to English (e.g. Dutch), is easier to understand than English produced by speakers with a genealogically unrelated mother tongue (e.g. Chinese)? These are the types of question that we aim to answer in our research on mutual intelligibility of Chinese, Dutch and American speakers of English.

Although there is a considerable body of research on the intelligibility of foreign-accented language, the comparisons are almost invariably limited to two languages. For instance, [1] determined the intelligibility of Dutch-accented English for English and Dutch listeners, and compared this with the intelligibility of English-accented Dutch for the same two listener groups. Intelligibility of non-native accents in English was studied extensively, e.g. by [2, 3] but only for native English listeners. The first systematic study of mutual intelligibility of non-native speakers of English [4] involved small numbers of speakers from many different language backgrounds. Intelligibility was determined at the sentence level such that no detailed diagnostics were available to pinpoint possible causes of poor intelligibility. It also remains unclear to what extent the speakers that represented the various background languages were comparable in their experience with, and prior exposure to, native English. Nevertheless, the study showed that non-native communication in English is more successful if speaker and listener share the same mother tongue. In a pilot study we replicated this overall effect in a detailed study of mutual intelligibility of Chinese, Dutch and American speakers, comparing production and perception of vowels, consonants, clusters and words in meaningful and meaningless context sentences [5, 6]. Because of the size of the materials the number of speakers in [5, 6] had to be kept very small, i.e. one male and one female speaker for each language background. Unfortunately, the speakers were chosen on the basis of their availability in the Netherlands, so that there is no guarantee that the speakers are at all representative of their peer groups, i.e. young well-educated adults (university students not specializing in English). Moreover, since the Chinese and American listeners in [5, 6] resided in the Netherlands, they were used to Dutch-accented English, which may have unduly boosted the intelligibility of the Dutch speakers of English. In order to remedy these defects, the present follow-up study was set up. It replicates the pilot study, but this time the speakers were carefully selected, through a screening experiment, so as to be truly representative of their peer group, and all listeners were tested in their own country.

We will now first describe the construction of the stimulus materials used both in the pilot study [5] and in the present replication. Next, we will explain how optimally representative speakers were selected for the main experiment. The remainder of the paper will then deal with the procedures and results of the main experiment

2. Materials

Basic materials. Three groups of speakers produced speech materials of five different types. In our materials we included five tests, probing aspects of intelligibility at the lowest (phoneme) level, at the intermediate (word) level, and at the highest (sentence) level.

1. **Vowel list**: words containing 19 different full vowels and diphthongs (excluding schwa) in identical /hVd/ contexts. This consonant frame is fully productive in English, allowing all the vowels of English to appear in a meaningful utterance, either a word or a short phrase [7]. Yet the listeners will get no lexical information from the consonantal context when they have to identify the vowel.
2. **Consonant list**: nonsense words /aCa/ containing 24 intervocalic English single consonants. The sole purpose of this list was to elicit the 24 English consonants in a symmetrical, identical vowel frame. The use of nonsense items was unavoidable.

3. **Cluster list**: 21 CC or CCC clusters in /aCC(C)a/ nonsense sequences. The list more or less exhausts the English inventory of initial consonant clusters.
4. **SUS-list**: 30 Semantically Unpredictable Sentences with high-frequency words occurring in syntactically correct but semantically nonsense sentences [8]. The SUS sentences were distributed over five different syntactic frames, as in, for instance *The state sang by the long week.*
5. **SPIN list**: fifty short sentences, with a contextually predictable or unpredictable target word in final position [9]. As in the SUS test, all words were common, high-frequency English monosyllables. In the unpredictable contexts the final target words were (more or less) used in citation forms, as in *We should consider the **map***. Predictable contexts occurred in sentences such as *Keep your broken arm in the **sling***.

Speakers were twenty native Dutch students, twenty Chinese, and twenty American students at Leiden University. Within each nationality there were 10 male and 10 female speakers. Speakers had not specialised in English language beyond the secondary-school level. Non-native speakers did not have, or never had in the past, regular contact with English-speaking friends or relatives, nor did they ever live in an English-speaking country. Native American speakers, rather than British – or some other Anglo-Saxon nationality – speakers were used, as the pronunciation norm of English taught in the People's Republic of China is American rather than British.

Speakers read the materials from paper in individual sessions while seated in a sound-insulated recording booth. Their vocal output was recorded through a Sennheiser MKH-416 microphone on a DAT recorder, and later downsampled (16 KHz, 16 bits) and stored on computer disk.

Speaker-screening test. Materials were then constructed for a speaker-screening test. Using the results of the earlier pilot study [7, 8] we determined for each speaker-listener combination sharing the same native language the subset of the ten most confusable vowels and consonants. We then constructed separate vowel and consonant identification tests for Chinese, Dutch and American speakers. Thus the American tests comprised 20 (speakers) × 10 (vowel types) = 200 items and the same number of consonant items. The Dutch and Chinese versions were constructed analogously. The Dutch items were presented to 20 native Dutch listeners, drawn from the same population as the speakers (but different individuals). The American items were presented to 20 Americans living in the Netherlands (different individuals but same peer group as speakers) but the Chinese items were presented to 20 native Chinese listeners in their own country (students at Jilin University, Changchun, P. R. China) so as to ensure that they were indeed representative of the Chinese student population (Chinese students who are selected to be sent abroad are preselected on the basis of above-average command of English).

Materials were presented over good-quality headphones to listeners individually or in small groups. On the basis of percent correct vowel and consonant identification scores (giving equal weight to both parts of the test) one speaker was selected from each group of 10 defined by nationality and gender. The single speaker was selected such that s/he was closest to the middle of the ranges established for each gender-by-nationality group.

Main experiment. The main experiment comprised the full sets of items for all five parts, but spoken only by the six optimally representative speakers – as determined by the screening procedure. Part 1 contained the 19 /hVd/ words for all six speakers in random order (across speakers), preceded by six practice items, yielding a total of 120 items. Part 2 contained the 24 /aCa/ items in random order across speakers, yielding 150 items (including 6 precursor practice items). Part 3 contained the six (speakers) × 21 /aCC(C)a/ items in random order, preceded by four practice items (130 in all). In part 4 a selection of SUS was presented such that each speaker contributed one lexically different sentence in each syntactic frame, so that the test comprised 5 (frames) × 6 (speakers) = 30 sentences (containing 112 content words in all) in random order across frames and speakers (preceded by five practice sentences, one for each different frame). Since part 4 involved word recognition, it was necessary to prevent learning effects by blocking sentences over speakers. Part 5, finally, comprised 50 SPIN sentences. Each of the six speakers contributed eight different sentences. The set of 48 was preceded by two practice sentences (one high predictable, one low predictable), yielding a total of 50 sentences in the test.

The materials were presented to 36 native Dutch listeners (tested in Leiden, the Netherlands), 36 Chinese listeners (tested in Changchun) and 36 American listeners (tested at the University of California at Los Angeles, USA). Within each group there were 18 male and 18 female listeners. Listeners volunteered, had no self-reported hearing problems, and were paid (the equivalent of) 10 Euros.

Stimuli were presented in a small lecture room over headphones. In parts 1, 2, and 3 the listeners were instructed to make a single forced choice from the 19 (part 1), 24 (part 2) or 21 (part 3) response alternatives, which were printed on their answer sheets. Subjects were told to gamble in case of doubt. Each item was presented just once with an inter-stimulus interval (offset to onset) of 7 seconds during the first half of each part, which was reduced to 5 seconds in the second half (when the listeners were highly familiar with the layout of the answer sheet). In part 4, the entire sentence was made audible once. Then the utterance was incrementally repeated such that the utterance was truncated after the first content word on the first repetition, after the second content words in the second repetition, and so on, until the final content word was made audible. The listeners had answer sheets before them with the functions words printed for each sentence but with the content words replaced by a line of constant length, as follows: *Why does the ___ ___ the ___ ___?* After each repetition the listener was given 3 seconds to fill in the next content word in the sentence. Then the entire sentence was repeated one more time to allow the listener to make any last-minute changes that he deemed necessary. In part 5 the listeners' task was just to fill in the last word of each successive sentence. No printed version of the sentences was provided. The entire listening session took 90 minutes, with a break in between.

3. Results

Figures 1, 2, and 3 plot percent correctly identified vowels (part 1), single consonants (part 2) and clusters (part 3), respectively, broken down by nationality of the listeners and broken down further by nationality of the speaker group.

Overall, the Chinese listeners have the lowest vowel identification scores (around 30% correct). Dutch listeners are intermediate (40–60% correct), and the American listeners perform best (50–70% correct vowels). Chinese-accented vowels are most difficult for both Dutch and American listeners but they are not identified significantly more poorly for Chinese speakers. American listeners are most successful when listening to native L1 American English; Chinese-accented vowels lead to significantly more perceptual errors, and the Dutch-accented vowels are in between. Dutch listeners have severe problems in identifying the Chinese-accented vowels, but are equally successful with Dutch and American-accented English vowels. Generally, then, each listener group is relatively most successful when having to identify English vowels when these were produced by speakers who have the same language background as the listener.

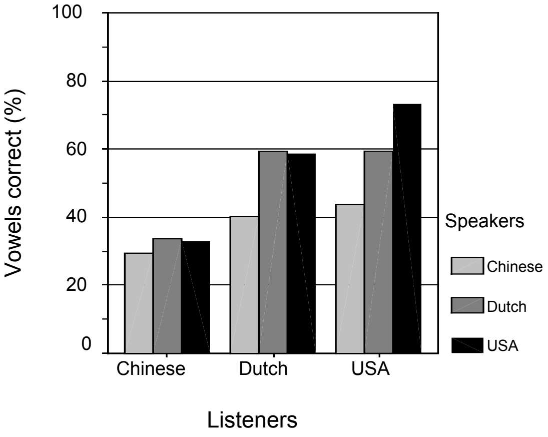

Figure 1: Percent correctly identified vowels broken down by listener group and by nationality of the speaker

Turning now to single consonant identification (figure 2), we observe, first of all, that overall consonant identification is more successful than vowel identification. Again, Chinese listeners have poorer scores than either the Dutch or American listeners. Dutch listeners do not show any disadvantage compared with American native listeners. Chinese-accented consonants are the most difficult for both Dutch and American listeners, but they are better recognized than by the Chinese listeners themselves. Dutch-accented consonants are poorly recognized by Chinese listeners.

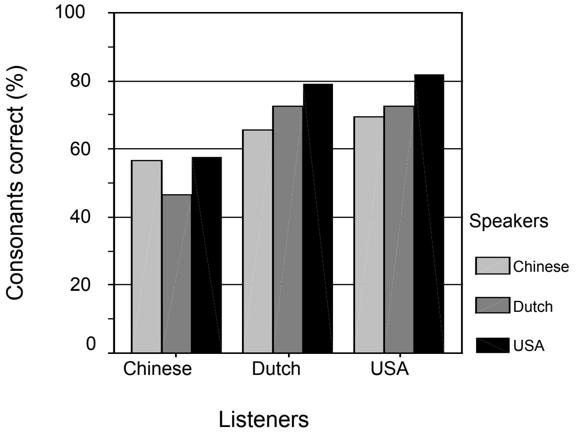

Figure 2: Percent correctly identified single consonants.

Identification scores for the cluster test (figure 3) are better still than those for the vowels and close to those found for single consonants. Chinese listeners score around 50% correct for Chinese and American speakers but only 35% for Dutch speakers. Dutch and American listeners are very close to each other, with scores between 80% and 90% correct. There is very little difference between the Dutch and American speakers (as was also observed for simplex consonant identification). Clusters produced by Dutch speakers are identified by Chinese listeners significantly more poorly than those spoken by American speakers.

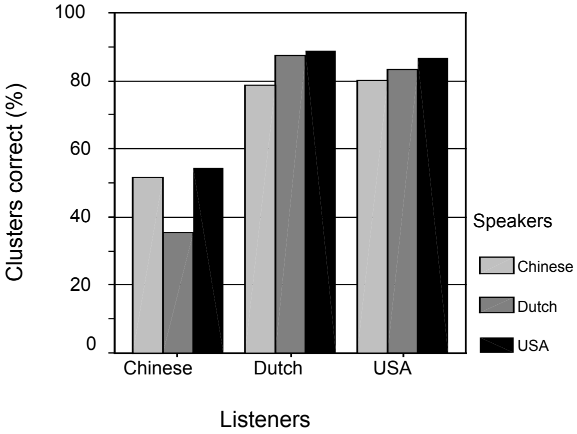

Figure 3: Percent correctly identified two and three-member consonants.

The scores for the SUS test are presented in figure 4. Poorest word-recognition is obtained for the Chinese listeners: around 35% correct, irrespective of the speaker's nationality. Relatively speaking, Chinese listeners identify the words better when these are Chinese accented. The configuration of results is roughly the same for Dutch and American listeners. These obtain close to perfect word-recognition scores if the speakers are either Dutch or American (with a small advantage for American speakers). The Chinese speakers' intelligibility is poorer by some 30 per cent.

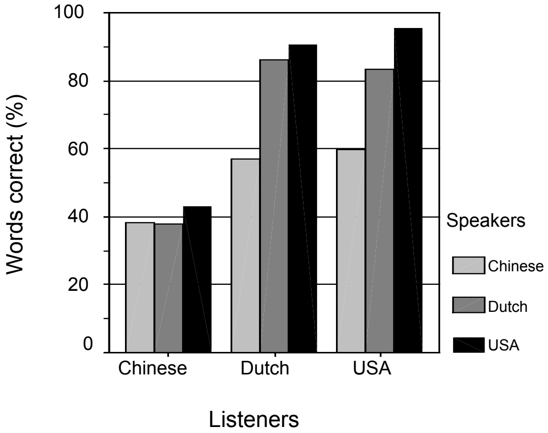

Figure 4. Percent correctly identified words in SUS test.

The results for part 5 (word recognition in meaningful utterances) are displayed in figure 5 for words in high-predictable (HP, top) and low-predictable (LP, bottom) contexts.

Generally, Chinese listeners have poor word-recognition scores (around 20% correct) even for HP words; curiously enough they perform better (40% correct) when the speakers are Dutch. American listeners clearly outperform their Dutch counterparts, especially in HP sentences. There is no difference in intelligibility between Dutch and American speakers when the targets are in HP contexts; in LP contexts a Dutch accent is a handicap for American but not for Dutch listeners.

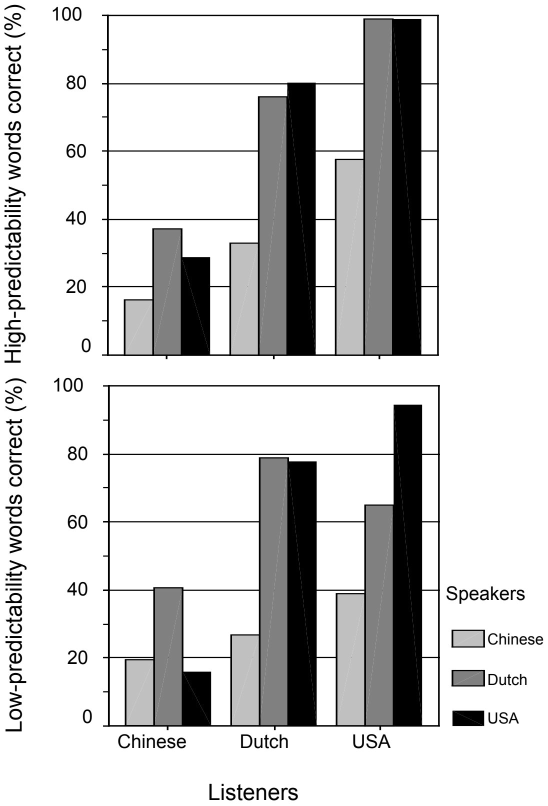

Figure 5. Percent correctly identified words in meaningful sentences, with high-predictability (top panel) and low-predictability (bottom panel) contexts.

4. Conclusions

The results of our study indicate, firstly, that Chinese speakers are more difficult to understand, and have more difficulty making themselves understood in English than their Dutch counterparts, whose production and perception of English is not much poorer than that of American native speakers and listeners. Note that the comparison was made between groups of university students (not specializing in English) in comparable stages of their academic training. The difference in proficiency may be due to either the closer genealogic distance between Dutch and English or because Dutch nationals get more English exposure through education and the media (or both).

Native speakers/listeners of English are always at an advantage; on average, they understand all types of speakers best, and they are understood better by all groups of listeners. However, there is a clear tendency in the results that when listeners and speakers share the same language background, communication is more successful than could be predicted from additive speaker and listener effects.

The three types of speakers, Chinese and Dutch L2 and native L1 American speakers of English, are most effectively discriminated by the SPIN words-in-context recognition test, but only in the part using words in low-predictability contexts. Note that, counter what the name of the test (Speech Perception in Noise [9]) suggests, we merely used SPIN sentences but did not add noise. The result indicates, then, that even the foreign accent induced by a genealogically closely related language severely reduces intelligibility, at least when lexico-syntactic contextual cues are absent.

Surprisingly, Chinese listeners obtain the (relatively) best SPIN scores when the speakers are Dutch. Since Chinese listeners were tested who had never been outside of China, the effect cannot be due to exposure to Dutch-accented English (as we could in [5]). Also, the effect cannot be predicted from the scores on the lower-level segment identification tests.

5. Acknowledgements

We are grateful to Robert S. Kirsner for providing facilities for administering the test at UCLA, and to Valerie Hazan for making the SUS materials available to us.

6. References

[1] Wijngaarden, S. J. van "Intelligibility of native and non-native Dutch speech", *Speech Comm.* 35:103–113, 2001.

[2] Bradlow, A.R., and D. Pisoni "Recognition of spoken words by native and non-native listeners: talker-, listener- and item-related factors", *J. Acoust. Soc. Am.*, 106:2074–2085, 1999.

[3] Flege, J. E., Bohn, O.-S., and Jang, S. "Effects of experience on non-native speakers' production and perception of English vowels", *J. Phonetics* 25:437–470, 1997.

[4] Dent, T., and Bradlow, A. R. "The interlanguage speech intelligibility benefit". *J. Acoust. Soc. Am.* 114:1600–1610, 2003.

[5] Wang, H., and Heuven, V. J. van "Mutual intelligibility of Chinese, Dutch and American speakers of English". In L. Cornips, P. Fikkert (eds.) *Linguistics in the Netherlands 2003*, Benjamins, Amsterdam, 213–224, 2003.

[6] Wang, H., and Heuven, V. J. van "Cross-linguistic confusion of vowels produced and perceived by Chinese, Dutch and American speakers of English". In L. Cornips, J. Doetjes (eds.) *Linguistics in the Netherlands 2004*, Benjamins, Amsterdam, 205–216, 2004.

[7] Benoît, C., Grice, M., and Hazan, V. "The SUS test: A method for the assessment of text-to-speech synthesis intelligibility using Semantically Unpredictable Sentences". *Speech Comm.* 18:381–392, 1996.

[8] Peterson. G. E., and Barney, H. L. "Control methods used in a study of the vowels", *J. Acoust. Soc. Am.*, 24:175–184, 1952.

[9] Kalikow, D. N., Stevens, K. N., and Elliott, L. L. "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability", *J. Acoust. Soc. Am.* 61:1337–1351, 1977.