Is a shared interlanguage beneficial?
Mutual intelligibility of American, Dutch and Mandarin speakers of English.

Hongyan Wang* & Vincent J. van Heuven
Leiden University Centre for Linguistics
Cleveringaplaats 1
PO Box 9515
2300 RA Leiden, The Netherlands
{h.wang, v.j.j.p.van.heuven}@hum.leidenuniv.nl

* Now at Dept of Foreign Languages, Shenzhen University,
3688 Nanhai Dadao, Nanshan District
Shenzhen, Guangdong Province
P R China
Wanghongyan0069@hotmail.com

Abstract

We determined the mutual intelligibility Mandarin Chinese, Dutch and American speakers of English in all nine logically possible combinations of speaker and listener native language backgrounds. Designated speakers (one male, one female per language group) were selected from larger sets of 20 speakers so as to be optimally representative of their peer groups. All non-native speakers and listeners were university students who did not specialize in English and had never lived in an English speaking community. Intelligibility was tested in separate tests targeting vowels, onset consonants, onset consonant clusters, words in syntactically correct but semantically empty sentences (SUS test), and words in meaningful sentences in which they appeared in either low or high predictability contexts. We test the hypotheses that mutual intelligibility between speaker and listener is better as (i) their native languages resemble each other more, and (ii) if speaker and listener share the same native language. In order to test the second hypothesis we propose a new method for quantifying the so-called interlanguage speech intelligibility benefit (ISIB).
1. Introduction

In the past 100 years, English has involved into the *lingua franca* of the world. By now it is the language *par excellence* of international politics, of the financial and business world and of the scientific community. Where artificial languages such as Esperanto have failed, English now fulfills a worldwide need of a universal language, but there are also downsides to this development. The world sees itself confronted with a bewildering variety of Englishes which are not only used by native speakers of English. One sometimes speaks – somewhat disparagingly – of Frenglish (French-accented English), Spanglish (Spanish English), Dunglish (Dutch English), Japlish (Japanese English) and Chinglish (Chinese English) to indicate these non-native varieties. Every person from the western world who has attended an international conference with English as the official language, can bear testimony to how difficult and tiring it can be to listen to non-native presentations. And where this leads to just a rather harmless irritation, human lives can be at stake when, for example, a Spanish flight controller has to communicate in English over a noisy radio transmission with a Dutch airline pilot.

There is only little consolidated knowledge of the specific pronunciation and speech recognition problems in so-called non-native communication, that is in situations where at least one of two interlocutor does not use his/her mother tongue. Even in the case of English, the most widely spoken second or foreign language, the number of studies into intelligibility problems in which either speakers or listeners are not native speakers of English is limited. Hardly anybody has done systematical research into intelligibility problems that exists when two non-English foreigners with different native language backgrounds have to communicate with each other in English. The few studies that exist in this area are methodologically non-optimal (e.g. Smith and Rafiqzad 1979, Bent and Bradlow 2003), which renders interpretation of the results hazardous.

If we take the communication between a speaker and a listener sharing the same mother tongue as a baseline condition, then we know that communication suffers when one of the interlocutors is replaced by a foreigner (e.g. Van Wijngaarden, 2001). Moreover, communication turns out to be more cumbersome if (ceteris paribus) only the listener is a foreigner than when only the speaker is foreigner (Wang 2007). A native listener is flexible and adjusts quickly to a deviant speaker (Cutler 2012, and references therein). A native speaker often has no idea what he should do to in order to become more intelligible to a foreign listener.

In the study which we report here, the situation is even more complicated because both the speaker and the listener may have to use a language which is not their mother tongue. Concretely, we will be dealing with the mutual intelligibility in English of Chinese, Dutch and American speakers of English. The target language is English. One source language, Dutch, is genealogically and geographically close to English. The geographic area of the other source language, Chinese, is thousands of miles away from the Anglo-Saxon world, and the typological differences between Chinese on the one hand and English and Dutch on the other, are enormous. Intelligibility is defined here in a rather technical sense as the recognizability of the words in the order in which the speaker placed them (e.g. Gooskens et al. 2010). Correct recognition of a reasonable number of words in the right order is a pre-condition to understanding the message. We also want use the results of our study in a diagnostic way. That is why it does not suffice to determine intelligibility at the sentence level, but we also examine the identifiability of smaller units such as individual vowels, consonant and consonant clusters. In this way we come to a number of questions, such as:
How well are the English vowels identified (and what is their confusion structure)?
(2) How well are single consonants and consonant clusters are identified (and what is their confusion structure)?
(3) How intelligible are the words in various types of sentences?
(4) What linguistic level (vowel identification, consonant identification, cluster identification, word recognition) yields the most sensitive measure needed to quantify differences in intelligibility?
(5) Can we predict intelligibility of words in various types of sentence contexts from the identifiability of individual vowels and consonants?
(6) Is the intuition correct that the mutual intelligibility of speaker and listener increases as the differences between the respective native languages are smaller? In other words, are American and Dutch speakers of English more mutually intelligible than either Chinese and American speakers or Chinese and Dutch speakers of English?
(7) Is it an advantage if the non-English speaker and listener share the same mother tongue?

In this short article we can only address part of these questions, namely (1) through (3), without going into confusion structures, and the last two questions. For elaborate and statistically supported answers to all these questions we refer to the first author’s dissertation (Wang, 2007).

2. Methodological complications

Because of the complexity of the research question and the degree of detail of the materials on the basis of which this type of research has to be carried out, researchers in practice always economize either on the number of speakers, or on the quantity of the materials per speaker. In the earlier mentioned research by Van Wijngaarden (2001), Dutch and English speakers were compared, who produced the complete set of vowels and consonants. Van Wijngaarden used two Dutch speakers of English (who are also native speakers of Dutch materials) and two English speakers of Dutch (who are also the native speakers of the English materials). It was not specified how the two speakers per language background were selected, nor how warranted their comparability is. We do know that the intelligibility of native speakers, even when listened to by native speakers, may differ considerably. When we are dealing with foreign-language speakers the differences in intelligibility between speakers will be even larger. The research done by Bent and Bradlow (2003) used two Chinese and two Korean speakers of English who were matched on the basis of comparable intelligibility scores obtained with American-English listeners. This, of course, does not answer the question whether Korean English is more difficult or easier to understand than Chinese English: the differences of intelligibility have been leveled out a priori through the selection procedure.

In our own research, we took a different course of action. It is unavoidable that we had to make do with just two speakers per language background. But we strived to select these speakers on the basis of their mean intelligibility within their peer group. The peer group is defined in all cases as young speakers of English, university students who did not study English as their major subject. In this way we compare the speaking and listening skills of the next generation of academics in three countries, i.e., China, the Netherlands and the United States.

A second precaution was that the non-native speakers and listeners should not have any special familiarity with the target language. That is why we took care that none of the speakers or
listeners who participated in the experiments, had any experience with English other than what they had learnt at school, excluding, for instance, subjects who had stayed in an English-speaking country or had had English-speaking family members or acquaintances. In the same vein, we excluded any Dutch listeners who might have been used to Chinese-accented English, and made sure that none of the Chinese listeners had ever heard English spoken with a Dutch accent.\footnote{In the remainder of this contribution we will describe first the composition of the materials, next the selection of two optimally representative (i.e. average) Dutch, Chinese and American speakers of English and finally we will describe the main experiment.}

In the remainder of this contribution we will describe first the composition of the materials, next the selection of two optimally representative (i.e. average) Dutch, Chinese and American speakers of English and finally we will describe the main experiment.

3. Composition of the materials

Three groups of speakers produced speech materials at five different linguistic levels, each of which probed different aspects of intelligibility, viz. at the lowest level of the sounds (phonemes), at the intermediate level of the words and at the top level of the sentence.

1. **Vowel List.** Nonsense words that contained the 19 different full vowels and diphthongs of English (with the exception of schwa) in identical /hVd/ environments such as in the words *heed, hid, head, had*, etc. This consonant frame is fully productive in English so that all English vowels may occur as a word or short phrase (Peterson and Barney 1952). Therefore, the consonant environment does not provide the listener with any useful information about the identity of the vowel.

2. **Consonant list.** This comprised nonsense words /\textipa{CA}/ (i.e. a consonant spoken in between two vowels /\textipa{A}/ (the vowel in father) which contained all inter-vocalic single consonants. The only purpose of this list was to obtain the 24 English consonants in a symmetrical vowel environment. The use of nonsense words was unavoidable here.

3. **Cluster List.** This was a set of 21 CC or CCC clusters /\textipa{C(C)A}/ nonsense strings. This is a non-exhaustive list of word-initial consonant clusters in English.

4. **SUS sentences.** 30 so-called Semantically Unpredictable Sentences with high-frequency words that were embedded in syntactically legal but semantically uninterpretable sentences (Benoit et al. 1996). The SUS sentences were equally divided over five different syntactic frames, e.g. *The state sang by the long week.*

5. **SPIN sentences.** Fifty short sentences with contextually predictable or unpredictable target words in final position (Kalikov et al. 1977). As was the case with the SUS test, all words were everyday, high-frequency English monosyllables. In the low-predictability contexts the words were (more or less) contained in citation form, as in *We should consider the map.* High-predictability (or ‘pregnant’) contexts were sentences such as *Keep your broken arm in a sling,* whether the preceding context constrained the target word to just one (or just a few) possibilities. In the original application of the test was for determination of the degree of hearing loss. In the original application, materials were presented in background noise (SPIN = Speech Perception in Noise). No added noise was needed in our own experiment as the materials were difficult enough because the non-native accents and reduced listening skills of the experimental subjects.
There were three groups of speakers. This first group comprised twenty Dutch-speaking students from the Randstad (City Belt) in the Western part of the Netherlands, ten male and ten female speakers. The second group consisted of twenty native speakers (ten male, ten female) of Mandarin (Standard Chinese). These speakers were recorded at Jilin University in Changchun in the Northern part of China. The third group included twenty American students (ten male, ten female) of the University of Leiden. None of the non-native speakers studied English or possessed any special knowledge of English beyond the secondary-school level. The non-Americans had never had any regular contact with English-speaking family members, friends or acquaintances, and had never spent a longer period in an English-speaking country. As native speakers of English we used Americans instead of British speakers because American is the pronunciation norm in the Chinese educational system, while the Dutch pronunciation of English is American rather than British. The speakers read the materials from paper in individual recording sessions in a quiet room. The speech was recorded on digital audio tape (DAT) with the aid of a Shure SM10A close-talking microphone and was later re-sampled (16 KHz, 16 bit) and stored on the hard disk of a computer.

4. Speaker selection test

Next, a speaker-selection experiment was carried out on the basis of the results of a earlier pilot study (Wang and Van Heuven 2003, 2004). We determined for every combination of speaker and listener with the same language background which were the ten most difficult (most confusing) vowels and consonants. Based on these results, separate intelligibility tests were constructed for Chinese, Dutch and American speakers. The selected vowels and consonants of every speaker were presented only to listeners who shared the language background of the speaker. For example, the American selection of test vowels contained 20 (speakers) × 10 (vowel types) = 200 items. In the same way a test set of 200 consonants was compiled. These 400 test items were presented for identification to 20 native American listeners only (who, like the speakers, lived in the Netherlands at the time the test was conducted). The Dutch and Chinese versions were constructed in the same way. The Dutch items (i.e., English spoken with Dutch accent) were presented to 20 Dutch listeners who were sampled from the same population as the speakers (but different individuals). The Chinese items were presented to 20 Chinese listeners in their own country (again, students at Jilin University, but different individuals than the speakers) in order to make sure that they were representative of the Northern Chinese (Mandarin speaking) student population.

The materials were presented to the listeners over headphones, individually or in small groups. For each listener group the vowel test was presented first, followed by the consonant test. In the vowel tests listeners chose the response alternative on their answer sheet they believed was intended by the speaker, with a forced choice from all 19 vowels of American English. Similarly, in the consonant test listeners chose one and only one alternative from a list of the 21 possible intervocalic single consonants. On the basis of the percentage of correctly identified vowels and consonants (with equal weight for both sound types) from every group of ten men and ten women per nationality, a single speaker was located who deviated least from the average of his/her group. The results of this selection test are presented graphically in figures la-b-c, for the Chinese, Dutch and American speaker/listener combinations, respectively. In each panel the
vowel identification scores are plotted (horizontally) against the consonant identification scores (vertically) for each individual speaker. The single male and female speaker who were located closest to the centroid (at the intersection of the crosshairs) of the two-dimensional distribution of means were selected as the optimally representative members of their peer group. For each speaker group the male-female pair with the shortest mean Euclidean distance to the centroid of the scatter cloud was selected. The scatter of individual speakers in the vowel-consonant space is roughly the same for each of the three speaker/listener groups. This fits with our earlier claim that there exist considerable differences in intelligibility even within a peer group of native speakers. We may observe that there is no correlation between the vowel scores and the consonants scores in any of the three groups, so that speakers who produce good-quality vowels may nevertheless produce poorly identifiable consonants and vice versa. Also, the range of variability in vowel identification and consonant identification if roughly the same in each language group. For this reason we decided to give equal weight to vowels and consonants in the selection procedure.

Figure 1. Location of 20 Chinese (panel A), 20 Dutch (panel B) and 20 American (panel C) speakers of English in a two-dimensional plot of correct vowel identification (horizontally) against consonant identification (vertically) scores. The crosshairs define the means of the vowel and consonant identification scores. The two individuals identified by a number (filled markers) were selected as the optimally representative speakers within each panel (further see text).

Informal inspection of the three panels reveals that the American speakers are located in the top right triangle of the two-dimensional space, indicating that the group as a whole produced highly identifiable vowels and consonants. The non-native speakers are found rather more towards the left-bottom part of the segment-identification space. The mean percentage of correctly identified consonants was the same for Chinese and Dutch speakers and listeners (66%), which is substantially poorer than the consonant scores in the American native group. The Dutch and Chinese speakers do differ, however, in the vowel scores, which are 11 percentage points better for the Dutch group – though still substantially poorer than in the American native group. Since different selections of vowels and consonants were presented depending on the language
background of the speakers and listeners, no definitive conclusions should be drawn from these observations.

5. Main Experiment

The main experiment contained the complete sets of items for all five test parts, but only those spoken by the six optimally representative speakers, as identified in the earlier speaker selection test. Part 1 included the 19 /hVd/ words of all six speakers in random order (across speakers) and preceded by six practice stimuli, which yielded a total of 120 items. Part 2 contained the 24 /aCa/ items in random order (across speakers) at a total of 150 items (including six preceding practice stimuli). Part 3 contained the six (speakers) × 21 /aCC(C)a/ items in random order, preceded by four practice items (130 all together). In Part 4, a selection of SUS sentences was presented such that each speaker contributed a single, lexically different, sentence in each syntactic frame so that the test contained 5 (frames) × 6 (speakers) = 30 sentences (with a total of 112 content words) in random order across frames and speakers (and preceded by five practice sentences, one for each different syntactic frame). Because Part 4 is a word recognition task, in which a word that has been recognized earlier would have an advantage when presented the second time due to learning effects (so-called ‘priming’), it was necessary to block sentences over the speakers such that the same content word was never presented twice to the same listener. Part 5, finally, contained SPIN sentences. Each of the six speakers contributed eight different sentences. The same sentence was never presented more than once to the same listener (blocking). The set of 48 sentences was preceded by two practice sentences (one high predictable, one low predictable), which yields a total of 50 SPIN sentences.

The materials were presented to 36 native listeners of Dutch (Leiden, from the Randstad), 36 Chinese listeners (Mandarin-speakers in Changchun) and to 36 American listeners (South Californian English-speaking, tested at the University of California at Los Angeles, USA). Each group of listeners comprised 18 men and 18 women. Listeners participated in the experiment on a voluntarily basis, had no self-reported hearing deficiencies, and received (the equivalent of) €10 for their services.

The stimuli were presented in small lecture rooms over headphones. In Parts 1, 2 and 3 were the listeners were instructed to make a forced choice from the 19 (part 1), 24 (part 2) or 21 (part 3) response alternatives, which were printed on their answer sheets. Listeners had to make a single choice at all times or gamble in case of doubt. Each item was offered only once, with a pause of 7 seconds in between items in the first half of every part of the test and of 5 seconds pause in the second half of every part (because the listeners could then find their way on the answer sheet more quickly). In Part 4, the entire sentence was made audible just once. Then the sentences were repeated incrementally such that the sentence was truncated after the first content word during the first repetition and after the second content word on the second repetition and so on, until at last even the final content word was made audible. Listeners had answer sheets in front of them with the function words printed per sentence while the content words had been replaced by a line of uniform length, as in *Why does the _____ ______ the _____?* After each repetition listeners were given 3 seconds to fill in the next content word in the sentence. Then the whole sentence was repeated one more time to allow the listeners a final opportunity to make changes. In Part 5, the listeners’ task was just to write down the final word of each following sentence. The subjects did not receive a printed version of the spoken sentences. The entire experiment took about 90 minutes, with a pause in the middle.
6. Results

Figures 2, 3 and 4 present respectively the percentage of correctly identified vowels (part 1), single consonants (part 2) and consonant clusters (part 3), broken down by native language background of the listeners and broken down further by native language background of the speakers. Speaker groups which are marked in the figures by the same number over the bar do not differ significantly within the listener group concerned ($\alpha = 0.05$) on the basis of a Scheffé post-hoc test following an Analysis of Variance (ANOVA).

On average, the Chinese listeners obtained the lowest vowel identification scores (about 30% correct). Dutch listeners assume an intermediate position (40-60% correct) and American listeners get the highest scores (50-70% correct). Chinese-English vowels are the most difficult both for Dutch and for American listeners, but they are not identified significantly more poorly by the Chinese listeners. American listeners are most successful when they listen to native speakers of American English, Chinese-English vowels do not lead to significantly more confusion and Dutch-English vowels assume an intermediate position. The Dutch listeners experience severe problems if they have to identify Chinese-English vowels but in general they are equally successful with Dutch and American realizations of the vowels. By and large, it would appear that every listener group operates more successfully, relatively speaking, with English vowels if these are spoken by a speaker who shares the language background with the listener. This effect has been called the ‘the interlanguage speech intelligibility benefit’ and was first described by Bent and Bradlow (2003). These authors, however, never quantified the benefit. In the next section, therefore, we will propose a computational method that quantifies the benefit.
(see also Van Heuven and Wang 2007), and will then examine the size of the effect for each of
the test components in our study.

In Figure 3 we see first of all that the identification of consonants is better overall than
that of vowels. Again, the score of the Chinese listeners is lower than that of Dutch and
Americans, while the Dutch listeners do not seem to experience any disadvantage relative to the
Americans. Chinese English consonants are the most difficult for both Dutch and American
listeners, but they are recognized more accurately by these listeners than by Chinese listeners
themselves. Dutch-English consonants are recognized poorly by Chinese listeners.

![Figure 4. Correct consonant cluster identification for Chinese, Dutch and American listeners broken down by accent of speaker. For further details see figure 2.](image1)

![Figure 5. Correct recognition of key words in Semantically Unpredictable Sentences for Chinese, Dutch and American listeners broken down by accent of speaker. For further details see figure 2.](image2)

The identification scores for the cluster test (Figure 4) are again better than those of the vowel
test and come close to the single-consonant level. Chinese listeners score 50% correct if the
speakers are Chinese or American, but they do not get a score higher than 35% for the Dutch
speakers. Dutch and American listeners are close to each other, with scores between 80 and 90%
correct. There is only a little difference between the Dutch and American listeners, as we have
already observed in the identification of single consonants. Clusters produced by the Dutch
speakers are identified significantly more poorly than when spoken by Americans.

The scores for the SUS-test are shown in Figure 5. The poorest word recognition is found
for the Chinese listeners: ca. 40% correct irrespective of the mother tongue of the speaker.
Relatively speaking, Chinese listeners recognize the words better when they are spoken with a
Chinese accent. The results are roughly the same for Dutch and American listeners. These
listener groups obtained almost perfect word recognition scores, with a minimal advantage for
the Americans. The intelligibility of the Chinese speakers is about 30% poorer. The intelligibility
of the Chinese and American speakers as determined here for American listeners is fully in line
with what has been reported by Rogers et al. (2004), who used (meaningful) Harvard sentences (Egan 1948) as test materials.

The results for part five (word recognition in meaningful sentences) are shown in Figure 6 for high-predictability (HP, left panel) and low-predictability (LP, right panel) contexts.

Figure 6. Percent correctly recognized keywords in SPIN-test for Chinese, Dutch and American listeners broken down by accent of speaker. Low-predictability (LP, A: left) and high-predictability (HP, B: right) contexts are shown in separate panels. For further details see figure 1.

Roughly speaking, the Chinese listeners have the poorest word recognition scores (about 20% correct), even for highly predictable words. Curiously enough they do better (40% correct) when the speakers are Dutch. American listeners obtain better scores than their Dutch counterparts, especially in the high-predictability sentences. There seems to be no difference in intelligibility between Dutch and American speakers if the target words are presented in high-predictability contexts. Only in the low-predictability context does a Dutch accent appear to be a handicap for American but not for Dutch listeners.

7. Quantifying the interlanguage speech intelligibility benefit

As stated above, it has been known for some ten years now that foreign-accented speech need not be less intelligible than native-accented speech. How intelligible a speech sample is, depends on the specific combination of speaker and listener language background. Obviously, the highest intelligibility levels can be expected when both speakers and listeners belong to the same homogeneous native language community. When the listener is non-native, however, it may happen that a non-native speaker is more intelligible than a native speaker, as was suggested
several decades ago by Smith and Rafiqzad (1979). This effect was first attested in a strictly
controlled experiment by Van Wijngaarden (2001). The phenomenon was christened the
interlanguage speech intelligibility benefit (ISIB) by Bent and Bradlow (2003), who tested two
versions of the effect, the weak and the strong version. The strong version of the benefit
hypothesis holds that any foreign-accented speaker is more intelligible to a non-native listener
than a native speaker is, possibly because foreign speakers talk slowly and typically use simple
words and sentence structures. The weak version states that foreign-accented speech is more
intelligible to a non-native listener only if speaker and listener share the same native-language
background. In the latter case, speaker and listener have access to the same underlying
phonological system so that deviations from the norms of the target language are predictable.
Unfortunately, Bent and Bradlow (2003) never quantified their results in terms of the benefit, so
that their claim that their data support the weak hypothesis better than the strong hypothesis is in
need of further testing.

Table 1 lists the segment identification or word recognition scores for the six tests broken
down by the nine combinations of listener and speaker language backgrounds. For each of 18
situations the listener group with the best score is bolded.

Table 1. Summary of test results. Percent correct on each of six tests broken down by native language (L1) of
listener and broken down further by L1 of speaker. Each mean is based on 36 listeners. The listener group with the
best performance is represented in bold face in a shaded cell.

<table>
<thead>
<tr>
<th>L1 of Listener</th>
<th>Test</th>
<th>Vow</th>
<th>Cons</th>
<th>Clus.</th>
<th>SUS</th>
<th>LP</th>
<th>HP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mandarin</td>
<td>Mandarin</td>
<td>29.7</td>
<td>57.2</td>
<td>52.8</td>
<td>39.3</td>
<td>19.4</td>
<td>16.7</td>
</tr>
<tr>
<td></td>
<td>Dutch</td>
<td>33.5</td>
<td>46.8</td>
<td>36.9</td>
<td>39.0</td>
<td>38.9</td>
<td>37.8</td>
</tr>
<tr>
<td></td>
<td>Am. English</td>
<td>33.1</td>
<td>58.2</td>
<td>56.0</td>
<td>44.2</td>
<td>17.9</td>
<td>31.8</td>
</tr>
<tr>
<td>Dutch</td>
<td>Mandarin</td>
<td>40.3</td>
<td>66.6</td>
<td>78.8</td>
<td>57.1</td>
<td>26.9</td>
<td>33.1</td>
</tr>
<tr>
<td></td>
<td>Dutch</td>
<td>59.3</td>
<td>73.7</td>
<td>87.8</td>
<td>86.2</td>
<td>81.3</td>
<td>76.1</td>
</tr>
<tr>
<td></td>
<td>Am. English</td>
<td>58.6</td>
<td>80.6</td>
<td>89.1</td>
<td>90.5</td>
<td>77.8</td>
<td>84.9</td>
</tr>
<tr>
<td>Am. English</td>
<td>Mandarin</td>
<td>44.9</td>
<td>72.5</td>
<td>82.5</td>
<td>59.5</td>
<td>39.4</td>
<td>57.8</td>
</tr>
<tr>
<td></td>
<td>Dutch</td>
<td>61.0</td>
<td>76.1</td>
<td>85.7</td>
<td>83.0</td>
<td>67.7</td>
<td>99.4</td>
</tr>
<tr>
<td></td>
<td>Am. English</td>
<td>75.3</td>
<td>85.7</td>
<td>89.3</td>
<td>95.5</td>
<td>95.2</td>
<td>99.1</td>
</tr>
</tbody>
</table>

Table 1 shows that the American listeners outperform the other two listener groups (Chinese,
Dutch) in 15 out of 18 situations. Yet, in three test × speaker L1 conditions, the American listen-
ers are outperformed, in one condition even substantially, by Dutch listeners. Crucially, in each
of these three exceptional situations, the Dutch listeners respond to Dutch speakers of English.
These three exceptional conditions, then, are examples of what we might call an absolute
interlanguage benefit. However, the Chinese speaker-listener combinations do not seem to reap
any benefit from their matched interlanguage. Yet, closer inspection of the data reveals that the
Chinese listeners perform relatively better when the speaker is also Chinese than when the
speaker has a different native-language background.

Van Heuven and Wang (2007) proposed a fairly simple computational method based on
linear modeling to quantify the magnitude of the ISIB. Their measure, called relative ISIB (or
RISIB) is basically the interaction component that remains after the main effects of speaker
language and listener language have been added together. An illustration of the method, based on the scores obtained in the vowel identification test (figure 2) is given in table 2. Expected and observed scores are listed in this table. The RISIB is the difference between the two ($\Delta$, residual or prediction error).  

Table 2. Expected vowel identification scores (% correct) on the basis of grand mean = 49% and main effects for Listener and Speaker L1. Observed scores (Obs.) and residuals ($\Delta$) are indicated. Bolded delta’s represent the interlanguage (or native language) benefit. All percentages have been rounded off to the nearest integer.

<table>
<thead>
<tr>
<th>Language background of</th>
<th>Exp.</th>
<th>Obs.</th>
<th>$\Delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listener</td>
<td>Speaker</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Chinese</td>
<td>–16</td>
<td>Chinese</td>
<td>–10</td>
</tr>
<tr>
<td>2. Chinese</td>
<td>–16</td>
<td>Dutch</td>
<td>+3</td>
</tr>
<tr>
<td>3. Chinese</td>
<td>–16</td>
<td>Am. English</td>
<td>+7</td>
</tr>
<tr>
<td>4. Dutch</td>
<td>+4</td>
<td>Chinese</td>
<td>–10</td>
</tr>
<tr>
<td>5. Dutch</td>
<td>+4</td>
<td>Dutch</td>
<td>+3</td>
</tr>
<tr>
<td>6. Dutch</td>
<td>+4</td>
<td>Am. English</td>
<td>+7</td>
</tr>
<tr>
<td>8. Am. English</td>
<td>+12</td>
<td>Dutch</td>
<td>+3</td>
</tr>
<tr>
<td>Grand mean</td>
<td>0</td>
<td>0</td>
<td>49</td>
</tr>
</tbody>
</table>

The observed scores are the mean percent correct vowel identification scores for each of the nine combinations of speaker and listener language background, which were graphically shown in figure 2. When the listeners are Chinese, Dutch and American, the expected mean score is –16, +4 and +12 relative to the grand mean; for the three speaker language backgrounds the expected mean should be additionally corrected with –10, +3 and +7, respectively. Note that the size of the increments/decrements is larger for listener language background than for speaker language background, i.e. the listener effect is larger than the speaker effect.

Generally, the observed scores are correctly predicted or even underestimated by the linear addition of the two main effects. Only in three combinations of factor levels is the observed score substantially better than the prediction. These are precisely the conditions in which the listeners are confronted with vowel tokens spoken by their fellow countrymen (shaded rows in Table 1). This native or interlanguage benefit is 4 to 8 percentage points better than the expected score. RISIB is most negative when the speaker-listener combination is American-Chinese or vice versa (5 to 6 points below the expected mean). RISIB is slightly negative (between 1 and 2 points below the expected mean) whenever one of the interactants is Dutch (whether speaker or listener in a mixed-language pair). The size of the RISIB is plotted in Figure 7 for each of the six segment identification or word recognition tests and broken down by the language background of the listeners.
Figure 7. Relative native/interlanguage benefit (percentage points) for Chinese, Dutch and American speaker-hearers of English, for six tests. For further details, see text.

In Figure 7 we have only plotted the RISIB-values for combinations of speakers and listeners sharing the same native language. In 16 out of 18 conditions there is weak to strong relative interlanguage or (in the case of the American listeners) native-language speech intelligibility benefit. If we had also plotted the results for the speaker and listener groups not sharing the same native language, we would have seen the exact mirror image, i.e., negative residuals with just a few exceptions. It would seem that, overall, the Chinese speaker-hearer combinations benefit most from the shared interlanguage. Possibly, the deviations from the (American) English pronunciation norms are largest for the Chinese group, which therefore benefits most by having access to the underlying phonological system causing the deviations. The discrepancies between Dutch English and American English are smaller and can be overcome more readily.
8. Conclusions

The results of this study show, first of all, that Chinese speakers are more difficult to understand in English, and understand English more poorly themselves, than their Dutch counterparts, whose production and perception of English is not much poorer than that of American native speakers and listeners. Again we point out that the comparison is based on the performance of groups of university students (who have not specialized in English) in a comparable stage in their academic education. The difference in their performance is either caused by closer genealogical relationship between Dutch and English than between Chinese and English, or it is a consequence of greater familiarity with English on the part of Dutch listeners, because of more intensive contact with English through education and media (or both).

Furthermore, it seems that native speakers and listeners of English always have an advantage. On average, they understand all types of speakers best, and they are best understood by all groups of listeners. Native listeners possess highly detailed knowledge of the system of their language. On account of that they can successfully understand deviant speech (e.g. with foreign accent) and successfully deal with difficult listening circumstances (e.g. in noise or over the telephone) when even highly advanced non-native listeners fail.

Independently of the native-language effect, our results show a clear tendency for the transmission of information to be more successful than can be predicted by a linear combination of the main effects of speaker and listener language background, when speaker and listener share the same language background. Again and again we see in the results that Chinese, respectively. Dutch, listeners enjoy an advantage over listeners with one of the other nationalities, when they listen to speakers with the same foreign accent that they produce themselves. At the same time this finding should be a warning to all non-native speakers of English. Because of this shared-interlanguage benefit, foreign speakers of English (or of any other foreign language) may have an overoptimistic view of their own intelligibility in English, both because they understand their own type of foreign-accented English relatively well, and because they are insufficiently aware of the difficulties they create for either native listeners of the target language or for non-native listener with a different language background.

The three types of speakers, Chinese and Dutch (both foreign-language learners) and the American native speakers, are distinguished from each other most effectively by the SPIN test, that is a word recognition task in which words are presented in meaningful contexts but, crucially, only in that part of the SPIN-test that presents target words in the high-predictability context. Native-language listeners can generate specific expectations about the target word on the basis of words they recognize earlier in the context. Foreign-language listeners have great difficulty processing the sentence so that they are not (or less) able to remember previous words and at the same time develop hypotheses about possible following words. Moreover, it is striking that even the American native listeners experience a considerable problem with Dutch English if the target words are at the end of a low-predictability context. Our results therefore show that a slight foreign accent as in the case of Dutch speakers of English by itself substantially reduces intelligibility if the choice of the target word is not constrained by the words in the preceding context and by the syntactic structure of the sentence.

An interesting conclusion is that the test that makes the sharpest distinction between mother-tongue and foreign-language speakers (and listeners) uses precisely everyday materials, that is short, simple sentences with highly frequent words. In all other tests the materials being presented are more or less exotic. Although the results of these latter tests may have diagnostic
value, they are less suited as an immediate measurement of intelligibility of speakers and listening ability of listeners.

It is surprising, finally, that the Chinese listeners obtained the highest SPIN-scores comparatively when the speakers are Dutch. Since we tested Chinese speakers who had never lived outside the People’s Republic of China, this effect cannot have been a consequence of adaptation to Dutch English. This finding is in line with other findings that Dutch English should be considered internationally as one of the easier and more intelligible varieties of English (e.g. Rogerson-Revell 2007).

9. Acknowledgments

We thank Robert S. Kirsner for making facilities available to conduct the tests at UCLA, and Valerie Hazan for her help in constructing the SUS sentences.

10. References


Notes

1. Results of a previous study of us were largely uninterpretable because we worked with Chinese speakers of English who had spent a year in the Netherlands on average, and were therefore accustomed to Dutch English (see Wang and Van Heuven 2003, 2004).

2. Dutch speakers of English increasingly switch to the American pronunciation norm (although most teaching materials used in the secondary school class room are British). Forty percent of the tested pronunciation variables was found to reflect the American norm (van der Haagen, 1998).

3. Chinese students who are sent abroad are selected – among other things – on the basis of the quality of their English (minimum IELTS score of 6.5). Such Chinese speakers/listeners of English would not be representative and were therefore not used in our study.

4. The numbers presented in this table deviate slightly from what was published in Wang (2007) and Van Heuven and Wang (2007). The present numbers are correct.