Mutual intelligibility of Chinese, Dutch and American speakers of English

Hongyan Wang and Vincent J. van Heuven
Jilin University, Changchun & Universiteit Leiden /
Universiteit Leiden

1. Introduction

Over the last fifty years, English has emerged as the world's lingua franca. It has become the language of international politics, trade, finance, and science. With the advent of English as the language of the world, a language barrier is about to disappear. Where artificial languages such as Esperanto have failed, English has stepped into the void and fills the need for a universal language. There is also a downside to the story. There is now a bewildering variety of forms of English with foreign accents characteristic of the various nations on this earth. These varieties of English are sometimes mockingly referred to by portmanteau designations such as *Spanglish* (Spanish English), *Dunglish* (Dutch English), and *Chinglish* (Chinese). Whoever has been to international conferences where the official language was English, will recall how difficult it was to follow the presentations of delegates from, especially, East Asia. The communicative problems are often no more than a mild nuisance, but human lives may be at stake when the air-traffic controller is a native speaker of Spanish and has to understand English messages spoken by a Dutch airline pilot (and vice versa) in a noisy cockpit. The practical solution adopted in the airline industry is basically a matter of fighting the symptoms rather than killing the disease. Typically, exchanges between pilot and tower proceed by a strict protocol, allowing only a small set of highly predictable words and phrases at any given point in the discourse. Moreover, the words themselves, and the sounds within them, have been chosen so as to be maximally distinct. A more principled alternative would be to find out what it is that causes the foreign accent in the non-native forms of English, and to develop a successful pedagogy that would allow anyone in the world to learn to speak English without even the slightest trace of a foreign accent.

As matters stand today, however, the discipline of (applied) linguistics is nowhere near developing a successful pedagogy for teaching foreign-language

Linguistics in the Netherlands 2003, 213–224.
ISSN 0929–7332 / E-ISSN 1569–9919 © John Benjamins Publishing Company

pronunciation, although interesting — and sometimes even promising — attempts are being made towards computer-assisted pronunciation coaches. But even the first stage of tackling the foreign-accent problem, i.e. determining extent of intelligibility loss, and pinpointing the specific sounds and sound combinations that cause the communicative problems, has received only scant attention in the literature. Typically, such studies address loss of intelligibility or comprehensibility of native versus foreign-accented speech for native listeners of L1. The general finding is that foreign-accented speech is more difficult to understand (see e.g. Wingstedt and Schulman 1984, Munro and Derwing 1995).

As a case in point Van Wijngaarden (2001) recorded short Dutch sentences spoken by advanced English learners and by Dutch native speakers. Intelligibility was determined by presenting the materials to groups of English learners of Dutch as well as native Dutch listeners. The materials were presented repeatedly such that the speech-noise distance was improved on each successive presentation in steps of 2 decibels (dB). The speech-to-noise distance (SND) at which 50% of the sounds were correctly identified, was computed. This measure, which is a standard diagnostic in audiology, served as the intelligibility threshold. Intelligibility thresholds were best when native Dutch listeners heard Dutch spoken by native speakers of Dutch, and worst when English learners were presented with English-accented realisations of Dutch, the difference amounting to some 5 dB poorer SND. An intelligibility loss of 3 to 4 dB was incurred when either the speakers or the listeners where English.[1]

It is generally believed that listeners have amassed over the course of their life an immense knowledge of the sound system of their native language. This knowledge comprises the structure of the lexicon (which words exist, and how often they are used in various types of language use), how the sounds pattern into syllables and syllables into words, and what lawful variations exist in the realisations of the sounds in the language. This knowledge of the redundancy in the lexicon and the phonology of the language enables the native listener to identify sounds and recognize words even when the speech stimulus is severely impoverished or incomplete. It has been shown, time and again, that native speakers are better than foreign learners at recognizing words that are swamped by noise (Ringeling 1984), or that are spoken by a (imperfect) text-to-speech computer (Pisoni et al. 1985). Similarly, native listeners recognize words from shorter fragments than foreign listeners in gating experiments (Nooteboom and Truin 1980), and native listeners are faster than foreigners in phoneme detection (Cutler et al. 1986, Bradlow and Pisoni 1999) and in lexical decision tasks (Poelmans 2003).

When Dutch and English listeners are exposed to Dutch-accented English, two sources of redundancy compete. Since the words are selected from the English lexicon, superior knowledge of the lexical redundancy of the English system is a distinct advantage to the native English listeners. However, given that the English sound system is distorted by interference patterns typical of the Dutch pronunciation, implicit knowledge of the phonetics and phonology of Dutch might be an advantage to the Dutch listener of English.[2]

An L2 speaker typically superimposes his native (L1) language system of sound categories onto the sounds of the L2. As a result the L2 learner may divide up the vowel space, to give just one example, in ways that deviate systematically from the native L1 division of the space. English, for instance, distinguishes two front low-mid vowels, i.e. /e/ as in *bet* versus /æ/ as in *bat*. Dutch (and Chinese) have only one low-mid front vowel /ɛ/. The beginning and intermediate learner fails to pick up the distinction between /e/ and /æ/ and replaces both by the single /ɛ/ of his native language. As a result of this the native English listener will not know which of the two alternatives the L2 speaker intended. In this example, the L2 is under-differentiated relative to L1 (Weinreich 1963); the L2 learner has to set up new sound categories (Flege 1987). This is a process of years, although it can be speeded up through explicit instruction.

In the present research we will complicate the situation considerably relative to, for instance, van Wijngaarden's (2001) study by not comparing speakers and listeners across just two languages but across three. Our target language is English. One interfering language is genealogically and geographically close; this language is Dutch. The second interfering language is spoken thousands of kilometres away from either England or from the United States of America, and belongs to a completely different language family, i.e. Mandarin Chinese. Our research addresses the mutual intelligibility of Chinese speakers of English compared with the English of (L2) Dutch and of native American (L1) speakers. Specifically we ask the following:

1. How well are English vowels identified in /hVd/-sequences (and what is the structure in their perceptual confusions)?
2. How well are English consonants and C-clusters identified in intervocalic position in the syllable onset (and what is their confusion structure)?
3. What is the intelligibility of words in various types of sentences?
4. Which linguistic aspect (vowel identification, consonant identification, cluster identification, word recognition) provides the most sensitive measuring tool to determine differences in intelligibility?

2. Method

Given our aim to determine the mutual intelligibility of Chinese, Dutch and American speakers of English, it follows that we need to record speakers belonging to each of the three nationalities when they speak English. In the next stage of the study the recordings will be presented to listeners belonging the same three nationalities in order to determine the relative intelligibility of the three types of speech for listeners of each nationality. In all, then, nine different combinations of speaker and listener groups will be involved in the experiment, as indicated in (1).

(1)

	Nationality of listeners		
Nationality of speakers	**Chinese**	**Dutch**	**American**
Chinese	(1)	(2)	(3)
Dutch	(4)	(5)	(6)
American	(7)	(8)	(9)

In principle, the listeners will be confronted with recordings produced by all speakers. This quickly leads to an unmanageably large set of listening materials. To keep the experiment within bounds, it was decided to limit the speaker group to just two individuals per nationality, i.e., one male and one female speaker of each nationality, totalling six speakers across the three nationalities. The number of listeners of each nationality was twelve, with a total of 36 listeners across the three nationalities.

2.1 Speakers

The six speakers (see above) were young adults, university students or staff at Universiteit Leiden, who had not specialised in English language beyond the secondary-school level. They did not possess any explicit knowledge of the English sound structure. Speakers did not have, nor ever had in the past, regular contact with English-speaking friends or relatives, nor had they ever lived in an English-speaking country. The Dutch speakers were Dutch nationals who spoke Standard Dutch. The Chinese speakers were students at the Medical faculty, who had not lived in the Netherlands for more than twelve months at the time the recordings were made. When in the Netherlands, they communicated with Dutch nationals in English, so that the Chinese speakers of English were, if anything, more acquainted with Dutch-accented English than with the native (American) English pronunciation. The two native American speakers of English had been in the Netherlands for no more than four months at the time of the recording. They, too, were students at Leiden University. Native American speakers, rather than British — or some other Anglo-Saxon nationality — speakers were used, as the pronunciation norm of English as taught in the People's Republic of China is American rather than British.[3]

2.2 Materials

The three groups of speakers produced speech materials of five different types, which will be illustrated below. The materials are not normally used in the context of second-language acquisition teaching or research. They were typically adopted from the field of quality assessment of talking computers (speech output assessment, cf. van Bezooijen & van Heuven 1997) or from speech audiology. In both fields one of the partners in the communication process is defective, either the speaker (i.e. the talking computer is not unlike a speaker with a foreign accent) or the listener. In speech technology and in audiology graded sets of materials have been devised in order to determine at what level of textual difficulty the communication process breaks down (intelligibility threshold). In our materials we included five such tests, probing aspects of intelligibility at the lowest (phoneme) level, at the intermediate (word) level, and at the highest (sentence) level.

1. **Vowel list**: a list of words containing the 20 full vowels and diphthongs (so 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 (Peterson & Barney 1952).[4] The general idea is that speakers have no special problem in reading out these sequences (since they are meaningful). Yet the listeners will get no information from the consonantal context when they have to identify the vowel.
2. **Consonant list**: a list of 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**: a list of 20 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 (Benoît et al. 1996). The SUS sentences were distributed over five different syntactic frames, as in, for instance *The state sang by the long week* or *Why does the range watch the fine rest?*
5. **SPIN list**: fifty short sentences, with a contextually predictable or unpredictable target word in final position (Kalikow et al. 1977). 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***.

The six speakers read the materials (in the order list 1 through 5) from paper in individual sessions while seated in a sound-insulated recording booth. Their vocal output was recorded through a Sennheiser MKH-416 unidirectional condenser microphone on a DAT recorder, which was placed outside the recording booth.

After the recording sessions the materials were downsampled (16 KHz, 16 bits) and stored on computer disk. Materials were then constructed for a listening experiment comprising five parts. Part 1 contained the 20 /hVd/ words for all six speakers in random order (across speakers), preceded by 10 practice items, yielding a total of 130 items. Part 2 contained the 24 /aCa/ items in random order across speakers, yielding 160 items (including 16 precursor practice items). Part 3 contained the six (speakers) × 20 /aCC(C)a/ items in random order, preceded by 10 practice items (130 in all). In Part 4 a selection of SUS sentences 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 111 content words in all) with a random order across frames and speakers (preceded by 5 practice sentences, one for each different frame).[5] Part 5, finally, comprised 50 SPIN sentences. Each of the six speakers contributed eight different sentences. The set of 48 was preceded by just two practice sentences (one high predictable, one low predictable), yielding a total of 50 sentences in the test.

2.3 Listeners

Listeners were drawn from the same pool of individuals from which the speakers had been drawn. Twelve listeners were Chinese, twelve Dutch, and twelve American. Across nationalities, subjects were roughly evenly divided over the sexes. Most listeners were students at Leiden University. However, some of the Dutch listeners had left university some years earlier and now held jobs in business or administration. The Chinese and American listeners were in the Netherlands for short period of study. Neither group had taken special courses on Dutch nor had the Americans any knowledge of Chinese.

Listeners took the tests in small groups, no more than three at a time. The stimuli were presented in a sound-insulated lecture room over Sennheiser HD 424 headphones being played back digitally at a comfortable loudness level from a notebook computer. The presentation was divided into five parts. Prior to each part the listeners read standardized written instructions, and listened to a series of practice items in order to get familiar with their task, the layout of the answer sheets, and with the time constraints of the stimulus presentation. In Parts 1, 2, and 3 the listeners were instructed to make a single forced choice from the 20 (Parts 1 and 3) or 24 (Part 2) 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 (so that the length of the line provided no clue as to the missing word's identity), 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.

In each part of the test we gave the listeners ample time to study the layout of the answer sheets (except Part 5, which was self-explanatory), before any practice items were played to them. At no time during the presentation of the materials was any feedback given to the listeners. The entire listening session took 75 minutes, with a short coffee break after either Part 2 or Part 3.

3. Results

The results for Part 1 (vowels), Part 2 (single consonants) and Part 3 (clusters) are presented in Figures 1, 2, and 3, respectively, which plot percent correctly identified vowels, single consonants and clusters, respectively, broken down by nationality of the listeners and broken down further by nationality of the speaker group.

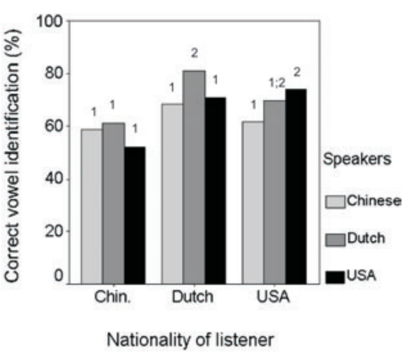

Figure 1. Percent correctly identified vowels broken down by listener group and by nationality of the speaker.

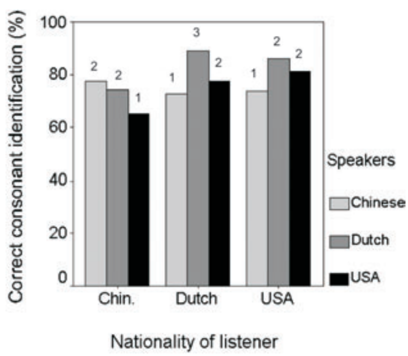

Figure 2. Percent correctly identified single consonants (further see Figure 1).

Overall, the Chinese listeners have the lowest vowel identification scores (between 50 and 60% correct). Dutch listeners perform best (between 65 and 80% correct), and the American listeners are intermediate (between 60 and 70% correct vowels). Chinese-spoken vowels are most difficult for both Dutch and American listeners but not for Chinese speakers. Generally, listeners obtain the highest identification scores when responding to materials produced by speakers of their own native language.[6]

Turning now to the single consonant identification test (Figure 2), we observe, first of all, that overall consonant identification is more successful than vowel identification. Again Chinese listeners have slightly lower scores than either the Dutch or native American listeners. As before, Chinese listeners perform best for Chinese speakers of English, and Dutch listeners score best for Dutch speakers of English. However, the Americans obtain slightly better scores for Dutch English than for American English.[7]

Identification scores for the cluster test are better still than those of either the vowels or single consonants. Chinese listeners score between 80 (for Dutch and American speakers) and 90% (for Chinese speakers) correct clusters. Dutch and American listeners are very close to each other, with scores around 90% correct. There is very little difference between the three speaker groups here.[8]

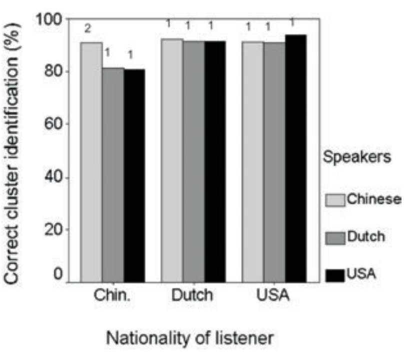

Figure 3. Percent correctly identified two and three-member consonant clusters (further see Figure 1).

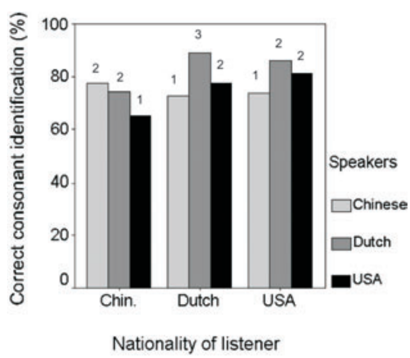

Figure 4. Percent correctly identified words in S.

The scores for the SUS test are presented in Figure 4. The (relatively) poorest word-recognition scores are obtained for the Chinese listeners (around 70% correct). The Dutch and American listeners score over 90% correct word recognition, except when the speakers are Chinese. For the Dutch and American listeners, Chinese speakers have reduced intelligibility by some 20 per cent. However, there is no difference in intelligibility between the three speaker pairs for Chinese listeners. Chinese English, then, is relatively more intelligible to Chinese than to Dutch and American listeners.[9]

The results for Part 5 (word recognition in meaningful utterances) are displayed in Figure 5 for words in high-predictable (left) and low-predictable (right) contexts.

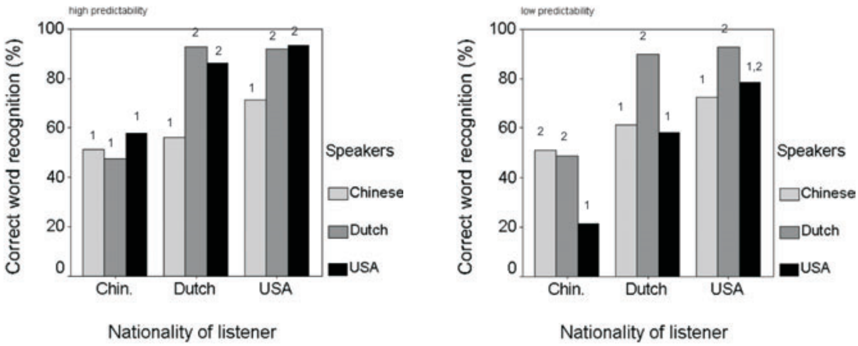

Figure 5. Percent correctly identified words in meaningful sentences, with high-predictability (left panel) and low-predictability (right panel) contexts.

Generally, the Chinese listeners have (very) poor word-recognition scores (around 40% correct) even for words in high-predictability contexts. Dutch and, especially, American listeners obtain much better recognition scores, especially when the speakers are Dutch or American nationals. There are no effects of contextual predictability for Chinese and Dutch speakers. However, when the speakers are American, there is a large drop in intelligibility of the unpredictable targets vis à vis their predictable counterparts, irrespective of the nationality of the listener. For some reason, the American speakers pronounced the unpredictable sentence-final target words (and only these) consistently less clearly than the 'Dunglish' speakers, and in one condition even less clearly than the Chinese speakers.[10] We can only speculate on the reason for this unexpected effect. Possibly, the mono-syllabic targets are highly familiar everyday words in English, but not for foreigners. Therefore it might be the case that these items were pronounced less carefully by the native speakers, leading to poor word recognition in the absence of a useful preceding context.

4. Conclusions

The first questions we asked (see introduction) concerned the differences in intelligibility of native versus foreign-accented speech to native and foreign listeners for each of five linguistically different aspects of speech processing. Simplifying grossly, it appears that Chinese listeners performed most poorly in all of the five tests. There were no differences in overall performance between the Dutch and

American listener groups when intelligibility was measured at the phoneme level (i.e., the vowel, consonant, and cluster identification tests). However, when the tests involved word recognition, the native American listeners proved clearly superior to Dutch listeners, who in turn outperformed the Chinese listeners.

For the three tests at the phoneme level, differences in intelligibility due to the speakers' native-language background were generally small, certainly smaller than the differences between listener groups. Yet, we observe a clear tendency for each speaker group to be understood somewhat better when the listener has the same nationality as the speaker. That is to say, Chinese-accented vowels, consonants and clusters are identified relatively more successfully by Chinese listeners of English than by the two other listener groups. Dutch-accented phonemes are identified most successfully by Dutch listeners of English. This finding supports our prediction that implicit knowledge of the interfering L1 sound system might be an advantage when having to decode L2 speech spoken with an L1 accent. However, it contradicts earlier results reported by van Wijngaarden (2001) for Dutch as the target language.

Overall, the tests were unable to establish a difference in intelligibility between Dutch and native American speakers of English. In fact, in some of the tests the Dutch speakers were understood even better than the Americans. This unexpected result was obtained in the consonant tests as well as in the recognition of words in low-predictability sentences. More detailed analyses of the confusion patterns in the perception of vowels, consonants, and clusters — not only in the phoneme-level tests but also in the word recognition tests — may afford a better separation of foreign versus native speakers and listeners even in the phoneme-level intelligibility tests.

Notes

1. Similar effects were observed in a second experiment in which the intelligibility of vowels and initial consonants in CVC words was determined for a range of speech-to-noise distances.

2. Note, however, that the results reported by van Wijngaarden (2001) would seem to indicate that knowledge of the interfering native (in his study: English) sound system does not afford the English listeners of Dutch to obtain better intelligibility scores for Dutch words spoken with an English accent than for the same words when spoken with the proper Dutch accent.

3. 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).

4. The original list of items was developed for British English, which is non-rhotic. When pronounced by American speakers, the so-called centering diphthongs will be followed by a (frictionless continuant) /r/ sound. This was an unavoidable consequence of using American speakers of English.

5. Since Part 4 involved word recognition, it was necessary to prevent learning effects by blocking sentences over speakers. Therefore, six versions of Part 4 were created such that sentences were rotated over speakers according to a Latin Square design. As a result, each unique combination of a sentence and a speaker was heard by six different listeners (two Chinese, two Dutch, two American), and no listener heard the same sentence more than once.

6. The data were aggregated over items (vowels) and subjected to a two-way Analysis of Variance (ANOVA) with nationality of speaker and nationality of listener as fixed factors. The effect of speaker nationality is significant, $F(2,99)=3.4$, $p=.038$, as is the effect of listener nationality, $F(2,99)=16.6$ ($p<.001$) and the interaction between the two, $F(4,99)=2.8$ ($p=.031$). Post-hoc comparisons (Scheffé, $p<.05$) reveal that Chinese listeners had poorer scores overall than American and Dutch listeners, who did not differ from each other. Separate one-way ANOVAs were then run on the results per listener nationality. When two or three speaker nationalities are identified by the same digit over the bars in Figures 1 through 5, they do not differ significantly from each other within the triplet (Scheffé, $p<.05$).

7. The ANOVA shows significant effects for nationality of speaker, $F(2,99)=15.3$ ($p<.001$), and of listener, $F(2,99)=15.1$ ($p<.001$) as well as for the interaction between the two, $F(4,99)=10.1$ ($p<.001$). Dutch speakers have significantly better consonant identification scores than American and Chinese speakers, who do not differ from each other; Chinese listeners have poorer scores than American and Dutch listeners, who do not differ from each other.

8. The ANOVA indicates a significant effect of listener nationality, $F(2,99)=22.0$ ($p<.001$), but not for speaker nationality. The listener × speaker interaction is significant, $F(4,99)=5.9$ ($p<.001$). Overall, the Chinese speakers have better cluster identification scores than Dutch and American speakers, who do not differ from each other.

9. The ANOVA shows significance for nationality of speaker, $F(2,99)=23.3$ ($p<.001$), and of listener, $F(2,99)=46.6$ ($p<.001$), as well as for the interaction, $F(4,99)=7.5$ ($p<.001$). Chinese speakers and listeners have poorer scores than American and Dutch speakers and listeners, who do not differ from each other.

10. A three-way ANOVA assuming fixed effects indicates significance for all three factors, $F(2,198)=18.6$ ($p<.001$) for speaker nationality, $F(2,198)=148.0$ ($p<.001$) for listener nationality, and $(F1,198)=11.9$ ($p=.001$) for predictability. The interaction of speaker nationality × listener nationality and of speaker nationality × predictability is significant, $F(1,198)=11.2$ ($p<.001$) and 17.5 ($p<.001$), respectively. Overall, each of the three listener groups differs significantly from the other two (Scheffé, $p<.05$); Dutch speakers are more intelligible than either Chinese or American speakers, who do not differ significantly from each other.

References

Benoît, C., M. Grice & V. Hazan (1996) 'The SUS test: A method for the assessment of text-to-speech synthesis intelligibility using Semantically Unpredictable Sentences'. *Speech Communication* 18, 381–392.

Kalikow, D.N., K.N. Stevens & L.L. Elliott (1977) 'Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability'. *Journal of the Acoustical Society of America* 61, 1337–1351.

Bezooijen, R. van & V.J. van Heuven (1997) 'Assessment of speech synthesis'. In D. Gibbon, R. Moore & R. Winksi (eds.) *Handbook of standards and resources for spoken language systems,* 481–653. Mouton de Gruyter, Berlin/New York.

Cutler, A., J. Mehler, D. G. Norris & J. Seguí (1986) 'The syllable's differing role in the segmentation of French and English'. *Journal of Memory and Language* 25, 385–400.

Flege, J. (1987) 'A critical period for learning to pronounce foreign languages?' *Applied Linguistics* 8, 162–177.

Haagen, M. van der (1998) *The English pronunciation of Dutch learners caught between two norms.* LOT dissertation series 12. LOT, Utrecht.

Peterson, G. E. & H. L. Barney (1952) 'Control methods used in a study of the vowels'. *Journal of the Acoustical Society of America* 24, 175–184.

Pisoni, D. B., H.C Nusbaum & B. G. Greene (1985) 'Perception of synthetic speech generated by rule'. *Proceedings of the IEEE* 73, 1665–1676.

Poelmans, P. (2003) *Developing second-language listening comprehension: Effects of training lower-order skills versus higher-order strategies.* LOT dissertation series 76. LOT, Utrecht.

Ringeling, J. C. T. (1984) *Reducing redundancy in normal, soft and whispered speech: A study on native and near-native perception.* Doctoral dissertation, RU Utrecht.

Nooteboom, S. G. & P. G. M. Truin (1980) 'Word recognition from fragments of spoken words by native and non-native listeners'. *IPO Annual Progress Report* 15, 42–47.

Weinreich, U. (1963) *Languages in contact.* Mouton, The Hague.

Wijngaarden, S. J. van (2001) 'Intelligibility of native and non-native Dutch speech'. *Speech Communication* 35, 103–113.