English as a lingua franca: mutual intelligibility of Chinese, Dutch and American speakers of English

Wang, H.

English as a lingua franca: Mutual intelligibility of Chinese, Dutch and American speakers of English

Published by
LOT
Janskerkhof 13
3512 BL Utrecht
The Netherlands

phone: +31 30 253 6006
fax: +31 30 253 6406
e-mail: lot@let.uu.nl
http://www.lotschool.nl

Cover illustration: Plot of vowels in formant space produced by American speakers of English (see Chapter five)

ISBN: 978-90-78328-20-9
NUR 632

English as a lingua franca: Mutual intelligibility of Chinese, Dutch and American speakers of English

PROEFSCHRIFT

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus Dr. D.D. Breimer,
hoogleraar in de faculteit der Wiskunde en
Natuurwetenschappen en die der Geneeskunde,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 10 januari 2007
klokke 13.45 uur

door

WANG HONGYAN

geboren te Tongliao, China
in 1967

Promotiecommissie

promotor:	prof. dr. V.J.J.P. van Heuven
referent:	prof. dr. ir. L.C.W. Pols (Universiteit van Amsterdam)
overige leden:	prof. dr. A.P.A. Broeders
	prof. dr. C.J. Ewen
	prof. dr. Liu Yi (Shenzhen University, P. R. China)
	dr. J.M. van de Weijer

The first year (2002/03) of the research reported in this dissertation was financially supported by a grant from the China Scholarship Council (12-months stay at the LUCL phonetics laboratory). During the second year (2003/04) the author was financially supported by a Delta scholarship from the Leiden University Fund (LUF). During the final two years of the research the author received a scholarship from the Leiden University Centre for Linguistics (LUCL).

Table of contents

Chapter seven: Intelligibility of intervocalic consonants

Chapter eight: Intelligibility of intervocalic consonant clusters

Chapter nine: Intelligibility of words in sentences

Chapter ten: Conclusions

Preface

One day in the first month of my stay at LUCL I was asked at a lunch table how well Dutch and Chinese speakers could understand each others' English and how it would be if compared with American speakers and listeners. I did not expect then that this would be the topic of my dissertation and that it would take me four years to answer these questions. Neither did I realize at the time how much these questions would broaden my mind as a researcher.

This dissertation would never have become what it is now if I had not received so much support, help and love from so many people around me. Saying "Thank you" from the bottom of my hearth will only begin to express the gratitude I feel towards the people who have contributed to the realization of this book.

Academically, I would like to thank Dr. Valérie Hazan from University College London, who sent me materials for my experiments and provided comments on the results; Prof. Jan Hustijn from Amsterdam University, who gave me one year's lectures on second-language acquisition and many light-shedding questions; Prof. Jim Flege from the University of Alabama, Birmingham and Prof. Ann Cutler from the Max Planck Institute for Psycholinguistics at Nijmegen for pointers to relevant literature and for personal encouragement; Prof. Jason Zhang from Shanghai Normal University for help with Chinese phonology.

My main experiment was run in three different places, i.e., in Jilin University, China, Leiden University, the Netherlands and UCLA in the USA. I very much appreciate the help from Prof. Bob Kirsner at UCLA and from Prof. Fu Guifang at Jilin University. Thank you both for contacting informants for me. I also thank all my informants in these three universities. Without their help this book could never have been written

These acknowledgement may not end without expressing my heart-felt gratitude to my second home in the past four years, the simple but cozy phonetics lab and to all my friends there working together with me, helping me and keeping me inspired. I have never been in an office for such a long time before and I have never so much enjoyed the friendship from everyone there. I thank Liang Lei, who guided me on my first days. I thank Lilie Roosman, who was like a sister to me in hard times doing hard work. I thank Jason Zhang, who helped me with many problems in life and in research; he left the lab before I did but he has never been away from my heart, which is still full of gratitude. I thank my roommates Rob Goedemans and Elisabeth Mauder, my dearest hero and "witch". I thank my dear Ellen, Vincent, Maarten, Jos, Gijs, Jurgen and Johanneke. I thank you and your families for your care and encouragement.

A special word of thanks is for my Dutch friend Marieke, who used to encourage me by saying "The last steps are made of lead" – which adequately captures my current feeling. Thanks to all the people who are helping me now during the final stages of the work, my aunts, my cousins, my sisters and brothers in Shenzhen, Tongliao and Changchun.

Thanks are also due to my parents, who gave me a strong character. I thank my daughter, Ziru, the angel of my life; thank you for being with me and always encouraging me.

Last but not least, I would like to thank the China Scholarship Council, the Leiden University Fund for its Delta scholarship and LUCL for your financial support during the years on my research.

Chapter one

Introduction

Now the whole earth had one language and few words. And as men migrated from the east, they found a plain in the land of Shinar and settled there. And they said to one another, 'Come, let us make bricks, and burn them thoroughly.' And they had brick for stone, and bitumen for mortar. Then they said, 'Come, let us build ourselves a city, and a tower with its top in the heavens, and let us make a name for ourselves, lest we be scattered abroad upon the face of the whole earth.' And the LORD came down to see the city and the tower, which the sons of men had built. And the LORD said, 'Behold, they are one people, and they have all one language; and this is only the beginning of what they will do; and nothing that they propose to do will now be impossible for them. Come, let us go down, and there confuse their language, that they may not understand one another's speech.' So the LORD scattered them abroad from there over the face of all the earth, and they left off building the city. Therefore its name was called Babel, because there the LORD confused the language of all the earth; and from there the LORD scattered them abroad over the face of all the earth. (Genesis 11:1-9)

1.1 English as a lingua franca

It is suggested in the Bible that the ideal state of the world would be, and at some stage was, one in which all mankind spoke the same language. However, God punished mankind for its arrogance with the multiplicity of languages, or the 'confusion of tongues'. Although a blessing for professional linguists, language teachers, translators and interpreters alike, the fact that there exist some 6,000 languages on the face of this earth which are mutually unintelligible, has been a matter of enormous financial consequences. It has been estimated, for instance, that the cost of having all documents translated in all the languages spoken in the European Community for the transactions of the European Parliament are in excess of 1 billion Euros a year.[1]

Over a century ago the Polish ophthalmologist Zamenhof devised the artificial language Esperanto, in an attempt to provide the world with a common language

[1] James Owen in London in the National Geographic News (February 22, 2005): 'The European Union has been operating in 20 official languages since ten new member states joined the legislative body last year. With annual translation costs set to rise to 1.3 billion dollars (U.S.), some people question whether EU institutions are becoming overburdened by multilingualism'.

that would be easy to learn and use, so that the confusion of tongues could be overcome. Although Esperanto has had numerous speakers, it never rose to the status of a lingua franca of the world. If any language may aspire to that status today, it would have to be English.

Indeed, English has become the language of international politics, trade, finance, and science. This comes with mixed blessings. On the one hand it brings the convenience of global communication, but the downside is that we now face a bewildering variety of forms of English ('Englishes') with foreign accents characteristic of the various nations on this earth, which are difficult to understand – for native listeners of English and even more so for non-native listeners. These varieties of English are sometimes mockingly referred to by portmanteau designations such as Spanglish (Spanish English), Dunglish (Dutch English), and Chinglish (Chinese English). Often the problem of non-native communication is no more than a mild nuisance, but human lives may be at stake when, for instance, an 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.

1.2 Topic of the dissertation

The topic of the present dissertation is the mutual intelligibility of speakers of English from diverse native-language backgrounds. As will be explained in greater detail in Chapter two, when a person speaks a language that is not his mother tongue, the language produced deviates in many respects from that of its native speakers. The most noticeable deviation of this so-called interlanguage is in the way the foreigner pronounces the target language. In fact, the foreign speaker's approximation to the target language will have a large number of sound properties, not only in the pronunciation of the vowels and consonants but also in the realization of the speech melody and rhythm, that seem to be copied from the speaker's mother tongue. Generally, this native-language interference is so strong that the foreign speaker's mother tongue can be established just by listening to his pronunciation of the foreign language. Native listeners are sensitive to the deviation from the native norm in the speech of foreign learners but, normally, communication does not break down on account of this. However, it has been shown that foreign-accented speech is highly vulnerable to background noise; it is clearly a less optimal code than speech between two native speakers (see Chapter two). Be this as it may, the native listener is normally able to cope with deviant speech and reconstructs the foreign speaker's intentions in spite of the suboptimal signals. The communicative problems will be severely aggravated when both interactants, i.e. speaker and listener, are non-native speakers, especially when they do not share the same mother tongue. In such situations the speaker produces distorted sound patterns (reminiscent of his mother tongue) which the listener cannot interpret because they do not conform to the patterns needed for the target language nor to the patterns in the mother tongue of the listener.

1.3 Approach

The basic problem that this thesis addresses, then, is to establish how difficult is it for speakers and listeners to understand each other when using English as a lingua franca, when the interactants do not share the same mother tongue. We will compare the results with several 'control' conditions. In one, both speakers and listeners are native users of English – which, of course, is the situation where optimal communication is expected. In a second control condition, either speakers or listeners, but not both, use English as a foreign language, and in a third control condition the subjects are neither native speakers nor native listeners of English but share the same mother tongue.

These conditions were obtained by having Chinese, Dutch and American speakers of English produce English words and sentences and offering the recordings to listeners with the same three native-language backgrounds. This yields nine combinations of speaker and listener nationalities:

Native language of speaker	Native language of listener		
	American	Chinese	Dutch
American	1	2	3
Chinese	4	5	6
Dutch	7	8	9

Communicative problems are expected to be greatest in combinations 6 and 8, which involve non-native speakers and listeners with different mother tongues. Optimal communication is predicted for combination 1, which contains native speakers and listeners of English. Comparisons will be made of combinations 4 and 7, with native listeners and foreign speakers as opposed to 2 and 3, with native speakers and foreign listeners. This comparison will tell us whether non-native communication is better when the speakers are native or when the listeners are native. A possible 'interlanguage benefit' (see below) in non-native communication can be tested in the combinations 5 and 9, where non-native speakers and listeners of English have the same native-language background, i.e. Chinese in 5 and Dutch in 9.

1.4 Goal of the study

This type of research has not been done before. To be true, there has been a wealth of research on the intelligibility of foreign-accented speech for native listeners of the target language (for a survey with emphasis on English as the target language, see Chapter two) and on the intelligibility of English for foreign listeners relative to native listeners. The point is, of course, that in all these studies there is always one party that uses English as the native language. The problem we address in the present study is more complicated, viz. the mutual intelligibility in English of non-native speakers with different source-language backgrounds. In fact, I am aware of just one (recent) study that addresses part of the issues raised here. Bent and Bradlow (2003) determined sentence intelligibility scores for a large number of foreign learners of American English of diverse linguistic backgrounds (Chinese, Korean, Japanese, Rumanian and many other nationalities). Not only did their results bear out that intelligibility was best between American speakers and listeners, but they also showed the existence of what they called an interlanguage speech intelligibility benefit, that is, that intelligibility between foreign learners of English sharing the same mother tongue was demonstrably better than between learners with different native languages.

In this thesis I want to study these matters in greater detail, using a much smaller variation of language backgrounds of speakers and listeners, but targeting the intelligibility not only at the sentence level but also at the lower levels of individual vowels and consonants, and of consonant clusters. Such a detailed study might allow us to pinpoint the problematic sounds, separately for speakers and for listeners, and from that to understand why intelligibility at the sentence level is successful to the degree that it is.

Concretely, we have asked the following questions for each of the nine combinations of speaker and listener nationality (or rather: native language backgrounds):

(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 (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?

The non-native speakers and listeners used by Bent and Bradlow (2003) differed considerably in their English proficiency. The Korean learners of English, for instance, were much better than the Chinese learners. It is unclear in their study, however, if the difference between the two Korean and two Chinese speakers was due to longer length of residence in the USA, to younger age of learning, or whether Korean learners have an edge over Chinese learners because the Korean sound system is more like that of English than the Chinese sound system is. In our study we have made an effort to select learners of English in the Netherlands and in China

that were representative of their populations. Specifically, we targeted young adult learners of English as a foreign language, i.e. in a situation of supervised learning in an environment where English is not the dominant language, nor the language of instruction. The learners were university students who do not specialize in English language and/or literature, and they did not have any regular contact with native speakers of English. The speakers we selected were in the middle of their peer groups, and represent the English proficiency of the typical young academically trained user of English as a foreign language in China and in the Netherlands.

1.5. Effect of linguistic distance

We have studied the mutual intelligibility of Chinese, Dutch and American (native) speakers of English. Dutch and English are West Germanic languages which are genealogically quite close and typologically similar. The two languages share a large number of cognates in their vocabularies, have many similarities in word and sentence structures, use comparable prosodic systems (both languages are of the stress accent type) and have highly similar segmental sound systems (phonetics and phonology). Chinese is a completely different language, typologically a poly-synthetic language with simple syllable structures, a complex lexical tone system, and a smaller vowel inventory than English. Detailed comparisons of the segmental sound systems of the three languages will be given in Chapter three. In our thesis we test the – obvious – hypothesis that mutual intelligibility in a lingua franca situation increases as the native languages of the interactants are more similar. We predict, accordingly, that Dutch-accented English is more intelligible than Chinese-accented English. This will not only be the case when the listeners are American but also when the listeners are Chinese or when they themselves are Dutch (the latter advantage would be due to the interlanguage benefit).

Potentially, comparing mutual intelligibility of non-native speakers and listeners of English may be used as a method to establish linguistic distance between any two languages in the world. There has been an upsurge of research activity in dialectometry on establishing the degree of similarity (and by implication linguistic distance) among dialects of a language or among languages within a language family. The research methodology does not rely on linguists' (or even naïve language users') intuitions of linguistic distance between varieties, but quantifies linguistic distance in terms of the number of symbolic operations, i.e. deletion, addition or substitution of phonemes in a transcription of word pairs (Levenshtein distance metric, see for instance Heeringa and Nerbonne, 2001; Heeringa, 2004; Gooskens and Heeringa, 2004). The method works quite well, even when there is a fair number of non-cognate word pairs between the two languages under comparison. The method has been verified against both judged and functionally determined communicative distance measures, such as intelligibility scores (percentage of correctly translated words) and opinion scores on intelligibility. The results indicate that the distance metric makes an accurate prediction of subjective and objective intelligibility scores.

However, the method breaks down when two unrelated languages are compared. First of all, when there are no cognate word pairs shared between the languages, then the number of symbolic operations that have to be carried out to map a word

onto its counterpart in the other language is determined by chance (and will be very large). Secondly, when two languages are not related to each other mutual intelligibility will be zero, so that no correlation can be established between the distance metric and the practical intelligibility measures. And yet linguists will readily agree that some sound systems are more like each other than others, even if all the languages belong to different families. Here, I would argue, we could fruitfully turn to the mutual intelligibility of these languages if their speakers and listeners use English. The more distant two languages, the smaller the mutual intelligibility when these speakers and listeners use English.

1.6 Contrastive analysis

As matters stand today, it is not possible to express the distance (difference) between two languages such as Dutch and English, or between Chinese and English, numerically. The differences are multidimensional and it is unclear how the various dimensions should be weighed against each other. All we can say, or rather assume, is that Dutch and English are closer than Chinese and English, but not how much closer. Nor would the (differences in) distance measures allow us to make a prediction of specific learning problems. Assuming that non-native communication is more problematic than communication between native speakers, can we predict specific difficulties from a comparison of the two sound systems? If the communication is between two non-native speakers of English who do not share the same source language, can specific problems be predicted by comparing the sound systems of all three languages involved? In this thesis we will attempt to make such predictions, based on various models of positive and negative transfer from the mother tongue to the foreign language, and test these against the observations in our experiments.

A literature survey in Chapter two will show that generally, contrastive analyses of the sound systems of source and target language have not been very successful in predicting learning problems. Sounds and contrasts that should be problematic proved easy in practice, and unexpected learning problems have been observed where the contrastive analysis predicted none. We will use the contrastive analysis only as a frame of reference in order to facilitate the presentation of the results. At best, it will allow us to show that certain views on native language interference in the foreign language provide better explanations than some other views. A second benefit of contrastive analyses is that may fulfill a useful role in interpreting findings post hoc, and in classifying types of errors (confusion patterns) post hoc.

1.7 Structure of the dissertation

After this short introductory chapter, the thesis is structured as follows.

Chapter two provides extensive background on the production and perception of non-native speech, models of second language acquisition, and techniques for measuring intelligibility at the full range of linguistic levels.

Chapter three contains a rather traditional overview of the sound systems of the three languages involved in the study, viz. Chinese, Dutch and English, as well as a comparative analysis of the languages in order to predict specific pronunciation and/or perception problems for the various combinations of speaker and listener nationalities.

In Chapter four I will outline the overall setup of the experimental work undertaken in the thesis, and provide a motivation for the choices we made. The chapter then describes the basic materials we collected from groups of 20 speakers for each language background, and how two optimal speakers (one male, one female) were selected from each set of 20 for the definitive tests.

In Chapter five I will present an acoustic analysis of English vowels spoken by Chinese, Dutch and American speakers, and consider how distinct the vowels in the English inventory are from each other, in terms of spectral and temporal properties. We will do this by applying a statistical technique called Linear Discriminant Analysis. The results of the analysis may be used as a prediction of perceptual confusions in the English vowel system as produced and perceived by the three groups of speakers.

In Chapters six through nine I present detailed results for the production and perception of vowels (Chapter six), simple consonants (Chapter seven), consonant clusters (Chapter eight) and for words in meaningless as well as meaningful sentences (Chapter nine).

In Chapter ten I will consider the relationships between the lower (word) and higher (sentence) levels, and try to establish which of the six tests we used affords the clearest separation of the various groups of speakers and listeners. I will then summarize the results, and draw overall conclusions.

Chapter two

Background

2.1 Foreign accent

Languages differ, and people from different places speak differently. Everyone may have had the experience, when listening to a foreigner speaking his/her own language, of having great difficulty in understanding what he is trying to say, not because of the speaker's lack of knowledge of vocabulary and language structure but because the sounds he produced seemed peculiar and because his voice rose and fell in unexpected places.

With the development of globalization and internationalization there is more and more communication involving speakers from many different linguistic and cultural backgrounds. Internet and cheap intercontinental telephony make oral communication feasible between people from anywhere in the world. Internet conferencing would be an ideal way for researchers to exchange ideas and to save time, money and energy as well, if they could really talk to each other without problems. Unfortunately, on many occasions, communication breaks down because the listener cannot get a clear idea of what his interlocutor is trying to say, due to his deviant pronunciation and speech melody. The consequences of such non-native communication may be severe if it happens in the air traffic control tower, or hospital emergency room, when people from different language backgrounds who need urgent information or help, cannot make themselves understood.

2.1.1 What is a (foreign) accent?

As a distinctive manner of oral expression, the notion of accent has two uses in linguistics. On the one hand accent refers to the way a speaker uses to make a syllable stand out in a word (word stress) or to make a word stand out in a constituent or sentence (sentence stress) so as to mark the syllable or word as communicatively important in the spoken utterance. To this effect the speaker may employ a variety of phonetic means, such as more careful pronunciation, greater loudness, longer duration and a relatively sudden change in vocal pitch (see for instance Van Heuven and Sluijter, 1996; Nooteboom, 1997). On the other hand, accent may refer to the way of speaking that is characteristic of a specific group of people from a regional background. What both readings of the term accent have in common is that some entity, be it a syllable, a word, or a speaker, stands out from its background. This thesis is about the second meaning of accent, i.e. deviant pronunciation rather than prosodic prominence.

People in different regions speak differently even in the same country in the same language. A regional variety of a language differing from the standard language is called a **dialect** when it is distinguished by differences at several linguistic levels, e.g. in pronunciation, grammar and vocabulary. When there are no differences in grammar and vocabulary but only the pronunciation (including the rhythm and melody) differs, the language variety is called an **accent** or **local/regional accent (L1 accent)**. Everybody speaks with some sort of an accent as a pattern of speech production. It betrays the speaker's geographical background, socio-economic class, ethnic identity, educational level, etc. Normally, the more distant the speaker's region is from that of the listener, the more different the accents of the interlocutors are, and the more difficult it is for them to understand each other.

When people learn a foreign language (L2), especially after puberty, they do not normally acquire native pronunciation in the new language. They will typically speak the foreign language with an accent, which is often the result of substituting phonemes and/or allophones of the native language (L1) for sounds that are needed in the foreign language. This kind of accent is called **foreign accent (L2 accent)**. Broadly speaking, then, foreign-accented speech is non-pathological speech produced by second-language users that sounds noticeably different from the speech of native speakers of the target language. It is probably true that there is little or no principled difference between speaking a language with a regional (native) accent or with a foreign accent. In both cases structures from the native dialect or language are transferred to the target language – be it the standard variety of one's native language or to a foreign language.

2.1.2 Linguistic levels in foreign accent

Even though we have restricted the notion of (foreign) accent to non-native language varieties that differ from the native norm only in terms of the sounds, it is not unusual to subdivide this area into the more abstract, representational aspects called **phonology** versus the more concrete aspects of the implementation of the abstract categories which are subsumed under the heading of **phonetics**. Phonologically foreign accent is often seen as wrong / missing representations of phonemes in the second language; phonetically, foreign accent is primarily the incorrect phonetic output routine which is employed to implement a correct phonological representation. Phonetic deviance is readily detectable by native listeners and can arise from phonemic, subphonemic, or suprasegmental differences in speech production (Flege, 1995). A phonemic difference would be the failure to distinguish between two members of a contrast in the target language because there is no such contrast in the learner's mother tongue, cf. Chinese (and Dutch) learners of English do not have distinct sound categories for the phonemes /ɛ/ and /æ/. An example of a subphonemic difference would be the failure to observe certain positional allophonic variants of a phoneme, such as the use in English of clear /l/ in the onset versus dark /l/ in the coda, when the learner's native language does not feature this allophonic difference, as would be the case for a French learner of English. A suprasegmental difference at the level of phonetics would be, for instance, the way Japanese learners of English would fail to mark English stressed syllables by greater duration and loudness, as their native language marks stress by pitch only (Beckman 1986).

2.1.3 Relative importance of pronunciation, morpho-syntax and vocabulary for intelligibility and comprehensibility

The question that we raise in this section is whether foreign-accented speech is indeed more difficult to understand than native speech. It should be stated at the outset that communication between a foreign speaker and a native listener is generally unproblematic as long as the foreign accent is relatively mild and the communication channel is noiseless.[1] For instance, Munro and Derwing (1995a) showed that the word error rate of their Mandarin learners of English was 11% against 4% for native English control speakers.[2] It is not easy to interpret such a finding. On the one hand the intelligibility of the foreign speakers is still quite high, since nine out of every ten words are correctly recognized. On the other hand, the error rate of the Mandarin learners is three times as high as that of the native speakers. Munro and Derwing used studio-quality recordings played back to listeners under high-fidelity conditions, unrealistic of real-life communication. Also, the Mandarin learners of English were immigrants to Canada with a minimum length or residence in excess of one year. The quality of their pronunciation must have been a lot better than that of the more typical Chinese speaker of English without any experience in an English-speaking environment.

Under more averse communicative circumstances, predictably, the intelligibility of foreign-accented speech deteriorates relative to native speech. As a case in point, Van Wijngaarden (2001) showed the effect of native versus non-native speech by adding noise to the communication channel. He defined an intelligibility threshold ('Speech Reception Threshold' or SRT) at 50% correct sentence recognition. His results showed that intelligibility was at threshold at a –6dB speech-to-noise ratio between native L1 Dutch speakers and listeners. When the speakers were English learners of Dutch, communication was less robust, at an SNR of –2dB, indicating that non-native speech is clearly less resistant to noise. This ties in with the subjective opinions of native listeners when exposed to samples for foreign-accented and native speech. The former type is uniformly judged to more difficult to understand. It seems to be the case, then, that when judging the difficulty of accented-speech; the judges have a clear conception of how well speech samples will hold up under averse listening circumstances.

---

[1] Lane (1963) seems to have been the first to establish that word recognition by native listeners is poorer for foreign-accented than for native-accented speech utterances. He found that word recognition for Serbian-, Japanese- and Punjabi-accented English was approximately 36% poorer than for native-English speech in a range of signal-to-noise ratios and filtering conditions. Lane's results, then, also indicated that the effect of foreign accent was greatly reduced as the speech channel was relatively noiseless.

[2] It is not easy to compute the word error rates from the data presented by Munro and Derwing (1995a). Given a mean utterance length of 10.7 words and three utterances contributed by each of ten Mandarin learners and two native control speakers, which were orthographically transcribed by 18 native listeners I divided the total number of word errors obtained for the Mandarin learners (636) by 5,778 and that of the control speakers (44) by 1,156.

Foreign learners of a target language deviate from the native norm not only in terms of pronunciation but also in their use of words and morpho-syntactic structure (so that it would be more apt to speak of foreign dialect, see above). We might therefore ask the question whether getting the pronunciation right should be a greater or lesser concern for the foreign learner than getting the lexis and morpho-syntax right. For several decades, pronunciation experts have stressed improving intelligibility as the most important goal of pronunciation teaching. As early as 1949, Abercrombie argued that most "language learners need no more than comfortably intelligible pronunciation" (p.120). This view has been echoed more recently by Gilbert (1980), Pennington and Richards (1986), Crawford (1987) and Morley (1991). However, this does not necessarily mean that improving one's pronunciation is the only – or even the most important – way to become a more intelligible speaker of a foreign language.

Several researchers have attempted to isolate the role of pronunciation, as compared to other linguistic features, in speech understanding. Gynan (1985) found that listeners judged that the phonology of Spanish non-native speakers of English interfered with their comprehensibility to a greater extent than grammatical errors did. Ensz (1982), on the other hand, found that grammar was more important than pronunciation for speech understanding when American non-native speakers were judged by native speakers of French. In a study of English-accented German, Politzer (1978) found that vocabulary errors affected listening comprehension most significantly, followed by grammar and then by pronunciation. In the study by Munro and Derwing (1995a) discussed above, the authors correlated the number of pronunciation errors and syntactic errors with objective (word error rate) and subjective (opinion scores) intelligibility measures. Grammatical errors correlated more strongly than phonemic errors with subjective intelligibility whilst the reverse was true for the objective word error measure. Later the same year Munro and Derwing (1995b) tested comprehension (by a sentence verification task) and processing time (for correct verification only) of 20 native English listeners who were exposed to the production of 10 Mandarin-accented speakers of English and 10 native English control speakers. Ninety-nine percent correct verification was obtained for the control speakers against 93% for the Mandarin speakers. Moreover, correct verification took about 60 ms longer for the Mandarin-accented utterances than for the control utterances. The native English utterances received much better accent ratings (mean = 1.5 on a scale from 1 to 9) than the Mandarin-accented counterparts (mean = 6.3). The same was true for subjective comprehensibility ratings (1.5 versus 5.4 for native versus foreign-accented tokens). In both studies (Munro and Derwing 1995a, b), judged comprehensibility and accentedness correlated around $r = 0.624$, which correlation is similar to the $r = 0.580$ reported by Van Heuven, Kruyt and De Vries (1981) but considerably less than the $r = 0.889$ that was found by Varonis and Gass (1982).

It should be noted that the articles reviewed here studied the relative strength of pronunciation versus morpho-syntactic errors as determinants of intelligibility through a correlational approach. To the best of my knowledge only Van Heuven (1986) varied morpho-syntactic and phonemic errors in an orthogonal experimental design. In his study of the intelligibility of native versus Turkish-accented Dutch, he varied the quality of the pronunciation (native, foreign) independently of the

morpho-syntactic properties (native, foreign) and found that the effects of pronunciation were roughly twice as large as those of morpho-syntactic deviations.[3] Native Dutch pronunciation resulted in 23% more correctly understood utterances (and 145 ms faster reaction times) than the Turkish-accented counterparts. The effects of Dutch versus Turkish morpho-syntax were a difference of 12 % and 93 ms, about half as large as the effect of pronunciation.

One could also argue on logical grounds that a good-quality pronunciation in a foreign language has higher priority than proper grammatical and morphological structure. Generally, a speaker can make himself understood in a foreign language as long as the content words are intelligible; the exact order in which the morphemes and words reach the listener would seem to be of secondary importance. After all, for word-order to have a (positive or negative) effect on intelligibility, the listener should first recognize the words: without any words there would be no word order to begin with.

2.1.4 Relative importance of various aspects of pronunciation (vowels, consonants, stress, accentuation, melody, rhythm)

A number of findings in foreign-accented speech research have emerged over the years with respect to those characteristics of speakers that were associated with either a greater or lesser degree of perceived foreign accent. Specific characteristics of the tokens produced by speakers have been associated, in various studies, with degrees of perceived foreign accent. Little is known about the relative importance of errors at each of the various linguistic levels on intelligibility of foreign-accented speech and perceived strength of foreign accent. Moreover, it may well be the case that particular errors are highly conspicuous and yet do not interfere with intelligibility, whereas other errors may go more or less unnoticed but are quite harmful to intelligibility. As a case in point, it has often been found that deviations in vowel quality and duration are very noticeable in foreign-accented speech. Yet, native listeners of languages such as English and Dutch are extremely flexible when they process utterances with incorrectly pronounced vowels. Van Ooijen (1994) showed that when confronted with nonwords that differed from their nearest lexical word in either one vowel or one consonant, listeners were much quicker to correct the vowel than the consonant. It has been argued (e.g. Best, 1993) that errors in vowels, which have greater intensity and duration than consonants, should be more detrimental to intelligibility than consonant errors.

There are indications that incorrect placement of word stress in English is highly detrimental to intelligibility. In good-quality native speech stress errors are not a problem, but stress is very important in speech of poor segmental quality, such as computer speech, speech in noise, and foreign-accented speech. It would appear that the stress pattern serves to limit the lexical search space for the native listener. When the stress pattern is incorrect, the listener will reinterpret the segments so that a word is found within the incorrectly constrained sublexicon. Examples that speak to the issue are given by Bansal (1966), quoted in Cutler (1983: 79), for Indian

[3] Van Heuven (1986) is a summary in English of earlier work reported in more detail in Dutch by Van Heuven, Kruyt and De Vries (1981) and Van Heuven and De Vries (1981, 1983).

English. In the Indian pronunciation of English the stress is perceived by English listeners one syllable later than where the Indian speaker intends it to be. As a result *character* was perceived as *director* and *written* as *retain*. However, it would seem to me that such effects will be restricted to languages that have contrastive stress. If a language has fixed stress or no stress at all, deviations from the canonical stress pattern will not greatly interfere with speech intelligibility.

It has often been said that speech melody has little impact on speech intelligibility.[4] The relative unimportance of melody is also suggested by the practice of state-of-the-art speech recognition software. There is not a single automatic speech recognizer that uses melodic information; the words can be recognized quite well just by identifying their constituent vowels and consonants. Nevertheless, Van Wijngaarden (2001) showed that the intelligibility of electronically monotonized Dutch speech (as defined by the Speech Reception Threshold, i.e. the signal-to-noise ratio at which 50% word recognition was still possible) was more difficult than the same utterances with melody intact (a 2-dB change in SRT).[5]

When dealing with speech melody, one has to make a clear distinction between melody at the level of the sentence (as in the preceding paragraph) and at the word level. It would seem obvious that incorrect word tones would greatly reduce the intelligibility of monotonous speech in tone languages, especially when the language has a predilection for short, monosyllabic words with a simple CV structure and has a large inventory of lexical tones – such as Chinese languages (with at least four tones, as in Mandarin, up to ten or more as in Cantonese). We are not aware of any studies of the intelligibility of monotonized speech in tone languages.

The upshot of the above is that it is very difficult to make generalizations as to the relative importance of specific levels in the linguistic hierarchy for intelligibility. Too much depends on the structural differences between source and target languages at each of the levels; therefore, what seems a clear difference of one level in favour of another in one language pair may be reversed in another pair.

More detailed studies have addressed the relative importance of specific types of error for the detection of foreign accent. Magen (1998) edited Spanish-accented English phrases so as to correct elements thought to be associated with the foreign accent. Adjustments to syllable structure, consonant manner of articulation and word stress were found to produce the most substantial effects in decreasing degree of perceived foreign accent. Adjustments to voice onset time (VOT), on the other hand, had little effect. Gonzalez-Bueno (1997) considered the role of stop voicing by manipulating the voice onset time of the initial segment /k/ in the Spanish word *casa* 'house' spoken by a native speaker of English. In judging the foreign accentedness of the single word token, raters identified those instances where the VOT of the /k/

[4] Obviously, speech melody is much more important for speech understanding. The chunking of the stream of speech in phrases and the highlighting of important words within the phrases, as well as the signalling of clause type, depend on the intonation pattern. Since the present thesis is about speech intelligibility rather than understanding, this role of intonation will not be considered.

[5] Here Dutch listeners heard native Dutch speakers. The effect of monotonization was a 3-dB poorer SRT when English-accented speech was presented to Dutch listeners.

was between 15 and 35 ms as most native-like, suggesting that VOT may indeed influence the degree of perceived accentedness. Using natural stimuli collected in a longitudinal study of English pronunciation by Japanese learners, on the other hand, Riney and Takagi (1999) only found limited support for a correlation between the VOT of stop segments and global foreign-accent ratings.

The accuracy of liquid pronunciation has also been considered within this context, though again, the relationship between segmental accuracy and global accent has not been firmly established

Major (1986) found that among native speakers of Brazilian Portuguese learning English, higher rates of epenthesis were significantly correlated with stronger global foreign accent. The use of epenthetic /i/ as opposed to schwa was particularly indicative of stronger accent.

Prosodic aspects of speech have also been demonstrated to correlate with global foreign accent. Magen (1998) and Major (1986) found that when all segmental information was removed from the speech stream judges were able to distinguish between English passages spoken by native speakers of English and native speakers of Mandarin. Jilka (2000) found similar results, with the accuracy of sentence level intonation being significantly correlated with the degree of perceived foreign accent in German speech of native speakers of English. Anderson-Hsieh, Johnson and Koehler (1992) echo the importance of prosodic factors in influencing perceived foreign accent, identifying them as more important than segmental and syllable structure factors in their study of English learners from range of L1 backgrounds.

The influence of speech rate has also been considered by some researchers. MacKay, Meador and Flege (2001) found that among late Italian-English bilinguals shorter sentences were perceived to be less foreign accented. Munro and Derwing (1998) found that the English speech of native speakers' of Mandarin was deemed to be more accented when slowed down and that at least some speakers' accents were found to be less strong when their speech was speeded up. Munro and Derwing (2001) suggest that the natural speaking rate of non-native speakers is typically somewhat slower than optimal (i.e. native) and found that when foreign-accented speech needed only slight speeding-up in order to be perceived as less foreign.

Interestingly, however, comprehensibility and intelligibility have been found to be only moderately correlated with global foreign accent scores (Munro and Derwing 1995a, 1999) in the English speech of native speakers of Mandarin. Munro and Derwing (1995b) pointed out that even highly accented speech can still be intelligible and comprehensible to native speakers.

2.1.5 Attitudes towards foreign accent

We have seen in the preceding section that, although the effects of foreign accent may be relatively small in terms of intelligibility and comprehensibility of speech utterances communicated through a virtually noiseless channel, native listeners seem to hear immediately that a speaker has an accent. Given, then, that foreign accent is readily detectable even when it does not overtly influence intelligibility, we may ask if native listeners are annoyed by foreign-accented speech or even discriminate against speakers with a foreign accent. Indeed, there has been a long tradition of research on attitudes towards foreign accent, as one of the salient characteristics of

L2 learners. A wide range of studies has shown that listeners often evaluate foreign accented speech negatively (Brennan and Brennan, 1981a, b; Fayer and Krasinski, 1987; Kalin and Rayko, 1978; Ryan and Carranza, 1975).

Aristotle (384-322 B.C.E.) believed that the type of language which speakers use has an effect upon their credibility or *ethos* (trans. Cooper, 1932). A similar idea is apparent in the Renaissance rhetoricians' preoccupation with the details of verbal expression. Research by dialect geographers in the early twentieth century called attention to language varieties which were stigmatized or, on the other hand, accorded prestige (Bloomfield, 1933). The earliest contemporary research on language attitudes towards language varieties was done by Lambert, Hodgson, Gardner and Fillenbaum (1960). In the 1970's, researchers continued to study attitudinal consequences of ethnically and regionally determined language variation. These numerous studies have shown that native listeners tend to downgrade non-native speakers simply because of foreign accent (Lambert, Hodgson, Gardner and Fillenbaum, 1960; Anisfeld, Bogo and Lambert, 1962; Ryan and Carranza, 1975; Kalin and Rayko, 1978; Brennan and Brennan, 1981a, b).

A significant body of research shows that foreign-accented speakers may be viewed as less intelligent, less competent, and even less attractive than native speakers. Rubin and Smith (1990) conducted a matched-guise study in which two Chinese women produced mini-lectures in both moderately and highly accented English. Intelligibility was functionally tested using a Cloze Blank-filling test. In subsequent opinion tests among various dimensions, one was to rate the instructor in terms of accentedness and teaching ability. Crucially, during the listening session half of the students saw a picture of a Caucasian woman and half saw a picture of a Chinese woman. The results showed that the students did not distinguish between the highly and moderately accented conditions but were affected by the suggested ethnicity of the speaker. Objective intelligibility scores and perceived accentedness were poorer when the Asian speaker was suggested; moreover, the impression of the instructor's teaching ability was negatively correlated with perceived accentedness. It seems to me that the listeners held a stereotypical expectation of the Chinese speaker being poorly intelligible (quite probably based on real-life experience), which caused them to make a less motivated effort to understand the speaker, i.e. the listeners gave up the attempt even before they had really tried.

Schinke-Llano (1983, 1986) noted that classroom teachers are often reluctant to engage English L2 students in conversation beyond basic classroom management exchanges. All these findings suggest that early intelligibility problems with foreign-accented speakers may have negative attitudinal and communicative effects on later exchanges with similarly accented speakers.

The negative effects of foreign accent have been found to extend beyond the classroom. Some evidence indicates that people in English-speaking regions in Canada have been denied housing or employment simply because of a French accent (from Munro's website).[6] The discrimination of foreign accent appears to have catalyzed the rise of accent-reduction programs which aim to reduce or eliminate foreign accents altogether.

[6] http://www.sfu.ca/%7Emjmunro/research.htm.

It would be wrong to conclude from the evidence above that foreign accent is always a social handicap. Research in the Netherlands (Doeleman, 1998) presents a more balanced view. Some foreign accents were found to be prestigious (Dutch spoken with an American, or better still, British accent) whereas other accents were attributed low prestige (e.g. Surinam, Moroccan and Turkish accent). The status of the accent seemed tied to the status of the community whose language is the source of the accent.

2.2 Causes of foreign accent

Now that we have defined what we mean by foreign accent and have briefly considered (some of) its communicative effects, let us consider factors that may cause foreign accent. Given that foreign accent in speech production is tantamount to saying that the sounds produced by the learner are off-target, we may ask what factors limit the phonetic accuracy in foreign language speech production. Moreover, it has often been noted that some learners have a stronger, more noticeable foreign accent than others. What, then, makes one L2 speaker have a more or less heavy accent than another? What factors contribute (most) to cross-language variation in foreign accent?

2.2.1 Age effects (AOA and AOL)

Age of arrival (AOA) and age of learning (AOL) are important factors for foreign-accented speech. AOA refers to the first arrival time of the L2 learner in a predominantly target language speaking country. AOL refers to the chronological age at which an individual first begins receiving massive input from native speakers of an L2 in a naturalistic context. Although very young immigrant children may arrive in the new country a few years earlier than they are exposed to the L2 (typically not before they go to school), AOA and AOL generally coincide. We will therefore no longer distinguish between them.

Taking a cue from Lorenz' (1961) work on imprinting in ducks and geese, Lenneberg (1967) introduced the critical period concept to research in native-language acquisition and claimed that foreign accent in an L2 cannot be overcome easily after puberty, because after puberty the ability for self-organization and adjustment to the physiological demands of verbal behavior quickly declines.[7] The brain behaves as if it has become set in its ways and primary, basic skills not acquired by that time usually remain deficient for life. Flege, Yeni-Komshian and Liu (1999) suggest that age affects phonology more than morphosyntax.

Many researchers support the view that age of learning is a very significant determinant of the degree of foreign accent. Long (1990) concluded that the L2 is

[7] Originally the phrase 'critical period' was used in ethologists' studies of species-specific behavior. It is the period when imprinting is observed in certain species such as young birds and rats. For example, geese isolated from their parent birds since the hatching react to and follow the moving object they see first. This kind of behavior can be learnt only during a short period of time after hatching (Lorenz, 1961, quoted in Clark and Clark, 1977).

generally spoken without an accent up to an AOL of 6 years, with a foreign accent by nearly all subjects having AOLs greater than 12 years, and either with or without foreign accent by subjects in the intermediate AOL range. Flege and Fletcher (1992) provided indirect evidence that foreign accent may be evident in the speech of adults who began learning their L2 as early as 7 years of age. As far as the pronunciation of an L2 is concerned, many studies have shown that earlier is usually better, i.e., people who arrive in a target language community at an early age have an advantage over those who arrive as adults (Asher and Garcia, 1969; Selinger, Krashen and Ladefoged, 1975; Oyama, 1976; Suter, 1976; Purcell and Suter, 1980; Tahta, Wood and Lowenthal, 1981a, b; Flege, 1988; Patkowski, 1990; Thompson, 1991; Flege and Fletcher, 1992; Flege, Yeni-Komshian and Liu, 1999). Both the proportion of individuals observed to speak their L2 with a detectable accent, and the strength of perceived foreign accent among individuals with detectable foreign accent have been found to increase as the age of learning the L2 increased. Results of Flege and co-workers show that in the production of several English consonants, Italian bilinguals whose AOL was earlier than 11 years generally performed better than those whose AOL was later than 21 years (Flege and Fletcher, 1992; Flege, Munro and Mackay, 1995; Piske, Mackay and Flege, 2001). These researchers proposed that even when other variables such as length of residence are partialed out, age of learning remains the most critical predictor of degree of foreign accent.[8]

2.2.2 Experience effects (LOR and L2 USE)

Two more factors that often come up as potential determinants of degree of accent are Length of Residence (LOR) and intensity of L2 use (USE). LOR is defined as the number of years spent by the learner in a country where the L2 is the predominant language. USE refers to how much/how often the learners use their L2 in daily life. Researchers have largely failed to reach agreement on the existence of a significant correlation between the accuracy of L2 pronunciation and either LOR or USE (Oyama, 1976; Flege and Fletcher, 1992; Piske, Mackay, and Flege, 2001).

Nevertheless, many studies (e.g. Tahta et al., 1981a) show that (frequency of) L2 use is significantly associated with foreign accent: the more the L2 is used, the better is the pronunciation of the L2. For example, Flege, Munro and Mackay (1995) found that language use at work, at home, or with friends was the second major factor in accentedness (after AOL).[9]

---

[8] Adult speakers can also attain native-like pronunciation. In Ioup et al. (1994), two adult participants were rated as natives in the production and perception of Arabic. Obler (1989) also reports an exceptional speaker who learned several different languages after puberty and attained native-like proficiency. Finally, Bongaerts (1999) and Bongaerts et al. (1997, 2000) investigated L2 adult speakers of different L1 backgrounds, and reported that some speakers attained native-like pronunciation in sentence reading tasks and spontaneous conversation. It seems that there are some, but not many, such exceptional speakers. However, Birdsong (1999) claims that almost 30 % of his participants in French speech tests reached native-like proficiency, and we cannot ignore these participants as outliers.

[9] In earlier studies (Flege and Fletcher, 1992; Oyama, 1976), L2 use was not found to be a significant factor. Closer reading of Flege and Fletcher (1992), however, shows that reported L2 use is significantly correlated with judged accentedness of the speaker ($r = .431$) at the .05

Piske, Mackay and Flege (2001) showed that LOR was no longer significantly correlated with perceived accentedness when L2 use was partialed out in the analysis. The reason that LOR was not found to be a significant factor was speculatively accounted for as follows: First, after a certain age, the amount of input does not affect the L2 proficiency significantly, which supports the critical period hypothesis. Second, the amount of L2 use varies greatly among learners. Third, the quantity of input might not be as important as the quality of input. In Flege and Liu (2001), late Chinese bilinguals were cross-classified into a group with short LOR (less than 3.8 years) versus long LOR (more than 3.8 years), and a student versus non-student group. Participants took several tests, including an English stop identification test and a listening test. The results showed that long LOR only guaranteed success for students but not for non-students. Furthermore, the student group as a whole performed better than the non-student group.

The conclusion seems warranted, therefore, that experience with the L2 is indeed an important determinant of degree of accentedness in the L2. Length of residence and frequency of L2 use, however, are only rough statistical indicators of experience. More accurate predictions could probably be made if the details of the learning situation were taken into account.

2.2.3 Transfer from the native language

The pronunciation of sounds in adults' native language and the differences between those sounds often interfere with recognizing a foreign speech sound or distinguishing one foreign speech sound from another. Weinreich (1953) defines interference phenomena as "those instances of deviation from norm of either language which occur in the speech of bilinguals as a result of their familiarity with more than one language," and then adds that "the greater the difference between the two systems, the more numerous the mutually exclusive forms and patterns in each, the greater is the learning problem and the potential areas of interference." Since then interference has been attributed to the fact that between any two languages there are similarities and differences on all levels of analysis. As Weinreich implies that the degree of interference that would ensue from the partial similarities and the complete differences between the two competing categories, one is in the learner's native language and the other in the target language. Linguists assume that by comparing the relevant categories in L1 and L2 the area of interference between L1 and L2 can be predicted.

Linguists attribute the ease or difficulty of learning L2 phonological categories to (i) the competing phonemic categories of the L1 and L2 systems, (ii) the allophone membership of the phonemic categories and (iii) the distribution of the categories within their respective system (Brière, 1968).

Lado (1957) argues that there is a hierarchy of difficulties in learning the phonological categories of a foreign language. He defines the area of difficulties in terms of:

---

level, and even at the .01 level if one-tailed testing is accepted. So there seems to be no conflict between these and later publications on the topic.

(1) The distinctive versus the non-distinctive features of the two systems: Does the native language have a phonetically similar phoneme?
(2) The allophonic membership of the phonemes: Are the variants of the phonemes similar in both languages?
(3) The distribution of phonemes: Are the phonemes and their variants similarly distributed?

According to Lado's contrastive hypothesis, similar sounds are physically similar to those of the native language, that pattern similarly to them, and that are similarly distributed. These similar sounds will be easily learnt by simple transfer without any difficulty (positive transfer). On the other hand, sounds that are physically different from the L1 system, that structure differently, and that are distributed differently, will be the most difficult for L2 learners (negative transfer).

In the next three subsections we will summarize and briefly discuss three current views on transfer from the native (source) language to the foreign (target) language in so far as they relate to the acquisition of the L2 sound system. All three models address the issue to what extent foreign accent, and learning problems, can be predicted by comparing the sound systems of source and target language.

2.2.3.1 Flege's Speech Learning Model (SLM)

By comparing the systematic similarities and differences between the actual pronunciations of foreign and native sounds, Flege (1987) defines L2 sounds which have no direct equivalent in L1 as "new" sounds and equivalent sounds which differ acoustically from their counterpart in L1 as "similar" sounds. Typically, a new sound is transcribed with an IPA basic symbol that differs from the symbol used to denote the equivalent sound in the native-language inventory. For instance, Dutch /ɛ/ is used as a substitute for the more open sound /æ/ in British English. After prolonged exposure to the foreign language, the learner will come to realize that the substitution is inadequate, and he will gradually form a new category for the foreign sound. Similar sounds are typically transcribed with the same base symbol from the IPA inventory and may differ only in diacritics (if at all). The auditory discrepancies between the native and foreign sounds are so small that the learner will never realize the substitution is harmful – even though his realizations of the target sounds may be noticeably incorrect when judged by native listeners of the target language. Flege's Speech Learning Model (SLM) predicts that the similar sounds are less easily produced and perceived in a native-like way than are new sounds, because the similar sounds in L1 have perceptual equivalence and merge into the same category in L2. This model predicts a greater (and more permanent) degree of difficulty for acquiring L2 sounds. The closer a target language sound is to the L1 sound, the more difficult it is to set up a new category for it (but also, the less the need for it, as the difference between source and target sound becomes negligible). Crucially, SLM makes the explicit claim that setting up new categories for the sounds in the L2 will go to the detriment of the categories in the L1, which will become less well-defined.

2.2.3.2 Kuhl's Native Language Magnet model (NLM)

In her Native Language Magnet (NLM) model, Kuhl (1991) proposes that native language categories are prototypes. Each prototype occupies a specific location in a space whose dimensions are the phonetic properties that define that class of categories, as, for example, the vowel space is defined by vowels' formant frequencies. Tokens near a prototype are perceptually drawn towards it. This is why Kuhl refers to the prototypes as 'magnets'. Foreign as well as native sounds are drawn more strongly to these prototypes as a function of their proximity from them in the phonetic space. More distant foreign sounds either assimilate to another prototype if they are closer to it, or do not assimilate if there is no nearby prototype. Newly born infants come into the world with a fixed and large set of prototypes for all sort of vowels and consonants. As a result, sounds in the infant's language environment are perceptually attracted to some of the prototypes, and after six months or so certain prototypes have received ample reinforcement whilst others that have no function in the infants native language, have attrited due to lack of activation. In a sense, learning a first language is a matter of unlearning certain prototypes and at the same time tuning the activated prototypes. When at a later stage in life a second language has to be learned, the learner will assimilate the foreign sounds to the existing prototypes in his native language inventory, so that it is very difficult to perceive any difference between the foreign sounds and their equivalent native sounds, as they are all assimilated to the same prototype and therefore sound alike. The second-language learner's main task, then, is to set up new prototypes – by reactivating and tuning attrited prototypes between existing native-language prototypes to account for the foreign sounds. NLM holds that the new prototypes will be set up without degrading the prototypes that were already in place for the native language. Like SLM, NLM is primarily a perception-driven model of language learning.

2.2.3.3 Best's Perceptual Assimilation Model (PAM)

Best analyses foreign accent in terms of similarity and difference between articulatory gestures across languages. Articulatory gestures refer to articulatory organs (active articulator, including laryngeal gesture), constriction locations (place of articulation), and constriction degree (manner of articulation). Different phonetic segments in different languages have different gestural constellations (in Best's terminology). Because all human languages draw upon the same set of gestural possibilities of the human vocal tract, there is usually a great deal of overlap among languages in the gestures and constellations contained within their individual phonological spaces, at least at segmental level. Non-native (foreign-accented) segments are those whose gestural elements or intergestural phasing do not match precisely with any native constellations (Best, 1995). Note that PAM does not appeal to perceptual characteristics of the sounds; perception is necessarily mediated through articulation, which makes the model a reincarnation of the motor theory of speech perception ('direct realism').

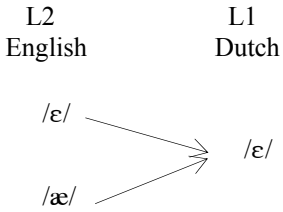
Focusing on these non-native segments, Best and co-workers try to determine to what extent the non-native segments are perceived in terms of the structures of the

source language, according to their similarities to, and discrepancies from, the native segmental constellations. For instance, if the listener's language has no ejective stops but does have voiceless aspirated and prevoiced stops, the glottal gestures and phasing of ejectives will be more similar gesturally to the voiceless aspirates than to the prevoiced stops, given that both the glottal closure and glottal widening prevent voicing and the glottal gestures of both are phased so as to reach their peaks at the release of the oral closure with which they are linked. Similarities between non-native segments and native gestural constellations determine the listeners' perceptual assimilation of the non-native phones to native categories. The listener is expected to detect gestural similarities and discrepancies to native phones. It is also expected that the listener will detect deviations from the gestural properties of native constellations as well. When the non-native sounds are very different from the native phones, they may be perceived only as having speech-like properties but may not assimilate strongly to any particular native category. In the extreme case they may not be recognized as speech (i.e. in terms of gestural constellations) but may instead be heard as non-speech sounds, such as clapping hands or flicking fingers. Best predicts that the non-native phones will be perceived in three ways:

(1) assimilated to a native category;
(2) perceived as an uncategorized speech sounds (this happens when the non-native phone falls in between two native categories (i.e. similar to at least native phones);
(3) perceived as a non-speech sound, which arguably happens when the non-native phone bears no resemblance to any phone in the native system.

PAM predicts several pairwise assimilation types. When non-native phones are phonetically similar to two different native phonemes and assimilate separately to them, the assimilation pattern is termed Two Category assimilation (TC). Flege calls it 'old contrast', e.g. the perception of English /p^h/ ~ /b̥/ and Spanish /o/ ~ /u/ by Dutch listeners (Escudero, 2001). In this pattern, the learner associates a binary contrast in the L2 with a binary contrast in L1. According to Best (1995) and Flege (1995), this leads to good category differentiation. Yet, this pattern may cause a perceptual problem, namely a boundary mismatch in the learner's L2 perception system, leading to problems with lexical access (Escudero, 2005). The learner should solve this problem by *shifting* the boundary between the categories in her L2 to match that of the target language.

Next, two sounds in the target language may, instead, assimilate equally well or poorly to a single native phoneme, termed Single Category assimilation (SC). Or both might assimilate to a single native phoneme, but one may fit better than the other, termed a Category Goodness difference (CG). An example of this assimilation pattern is given here:



Alternatively, one non-native phone may be Uncategorized as defined above, while the other is Categorized, forming an Uncategorized-Categorized pair (UC). Or both non-native phones might be Uncategorized speech segments (UU). Two phones' articulatory properties may both be quite discrepant from any native phonemes, and be perceived as Non-Assimilable (NA) non-speech sounds.

Discrimination of non-native contrasts can be hindered, aided or unaffected by native phonology, depending on how the non-native phones relate to native phonemes and contrasts. Native phonology should aid discrimination when the two phones are separated by native phonological boundaries, but should hinder it when both phones assimilate to the same native phoneme. TC and UC contrasts should be discriminated quite well, because in both cases the contrasting phones fall on opposite sides of a native phonological boundary. On the other hand, with the CG and SC types, both phones assimilate to the same native phoneme, so discriminability is impeded by native phonology. If one phone is good and the other is poor, discrimination will be very good (CG difference), but not as good as in a TC contrast because it is hindered by assimilation to a single native phoneme. In the SC case, both non-native phones are equivalent in phonetic goodness, hence discrimination is poor, hindered both by lack of phonological contrast and by lack of difference in fit.

PAM makes predictions about how listeners will recognize (or: assimilate) non-native phones with respect to the phonological categories of their native language, and how they will discriminate non-native contrasts.

When surveying the various models discussed above, it seems that it is very difficult to predict the specifics of foreign accent from a systematic comparison of phonetic and/or phonological properties of the source and target language. In spite of Flege's heuristic (same symbols different = new sounds, only diacritics different = similar sounds), learners sometimes have no problems where they are predicted or experience great learning problems where the model predicts their absence. Often, the researchers were honest enough to own up that their classification of sounds and sound contrasts in second language learning were based on existing pedagogical wisdom, and did not directly follow from any contrastive analysis. Given this unsatisfactory state of affairs, it might be worthwhile briefly considering an alternative option, which basically is a formalized procedure of being 'wise after the event'.

2.2.4 Alternative approach

Since a few years an alternative for contrastive analysis has been to use techniques from speech technology and apply these to the problem of foreign language learning. Automatic speech recognizers can be trained to recognize the sounds in the native language. Once the system is properly trained on a sufficient number of tokens of the target sounds (vowels and consonants) as spoken by a homogeneous group of first-language speakers, it will be able to successfully classify any new token as a sound in terms of the inventory of the target language, as long as the sounds are being spoken by a member of the same linguistic community as the training set. The recognizer technology as such is not directly relevant to the methodology; it may be Hidden Markov Modelling, Neural Networks, some hybrid mixture of both or even a more traditional multivariate statistical analysis such as Linear Discriminant Analysis. The point is that whatever the nature of the classification algorithm used, it can be applied to the task of classifying the sounds in another language that deviates to a greater or lesser degree from the training language. The algorithm will then misclassify the sounds in the foreign language by the same system of the training language, yielding crucial errors in the classification pattern. In fact, the confusion structure that we obtain in the vowels after they have been classified by the algorithm will be close to the misclassification we will obtain when foreign-language learners were asked to the same thing, i.e. the classification results/errors are a good prediction of the learning problems of human second language learners. This procedure has been used with reasonable success by Strange, Akahane-Yamada, Kubo, Trent, Nishi and Jenkins (1998) and by Strange, Bohn, Trent and Nishi (2004).

2.3 Measurement of intelligibility

2.3.1 Terminological preliminaries

Let us first consider what we mean by intelligibility. The exchange of ideas between speaker and listener has often been described in terms of a series of processes which together make up the so-called speech chain. When the stream of sounds impinges on the listener's eardrum, the listener will recognize linguistic units in the stream of sounds, viz. words, which appear in a particular order. A stretch is perfectly intelligible if all the words in the utterance are correctly recognized in the correct order. This is not the same as speech understanding (also: comprehension). Although the recognition of the words and their linear order is a precondition, comprehension is obtained if the listener correctly reconstructs the speaker's intentions, i.e. if the listener understands what the utterance means. A nonsensical utterance such as the beginning of Lewis Carroll's (1872) 'Jabberwocky':

> 'Twas brillig, and the slithy toves
> Did gyre and gimble in the wabe:
> All mimsy were the borogoves,
> And the mome raths outgrabe

is technically intelligible – as we can tell exactly what the words are (how many, and how they relate to each other) but we cannot reconstruct the writer's original intention as the words do not exist in the lexicon; it is therefore not comprehensible.

As a consequence of this we distinguish between the processing modules for word recognition (in sentential context) and higher-order integration of the word meaning into multi-sentence understanding (comprehension). Separate tests are required to test intelligibility and comprehension. Typically, intelligibility tests determine the number of correctly reproduced linguistic units (sounds, words) without ever asking if the listener understood the meaning of the utterance. Comprehension tests, on the other hand, ma check whether the listener has understood the meaning of utterances, for instance by asking the listener to choose whether the sentence is true, unlikely or nonsense (sentence verification). However, it is never an explicit concern for a comprehension test to check whether the listener has recognized all the words.

A second distinction we need to make is that between functional testing and opinion testing. Whether we are dealing with the testing of intelligibility or of comprehension, two types of test methodology are possible. Opinion tests ask the listener to subjectively rate a stretch of speech along one or many rating scales. For instance, an opinion test of intelligibility might ask the listener to assign a score to a foreign-accented utterance between 1 and 7 along a scale of intelligibility, where '1' would mean 'I think it is impossible to even recognize a single word' and '7' might mean 'I think it would be very easy to recognize all the words in this utterance perfectly'. Intermediate values would represent intermediate degrees of difficulty (i.e. intelligibility). Research has shown that native listeners have excellent and reliable intuitions on the relative intelligibility of (foreign-accented) speech utterances, with high within and between-rater agreement. Such a procedure may allow us to rank order foreign-accented utterances or speakers, but it will not tell us what the percentage of correctly recognized words will actually be. This is why we need functional tests. Functional tests require the listener to recognize words (when we are interested in intelligibility) or to actually grasp the meaning of the sentence(s) (when we are targeting comprehension). In our research we will only deal with functional tests of intelligibility. Nevertheless, in this introductory chapter we will briefly deal with both techniques for the testing of intelligibility as well as comprehension – but the emphasis will always be on functional testing.

When it comes to functional testing, a further split in tests has to be made in terms of on-line versus off-line techniques. On-line tests require the subject to respond when they are still processing the auditory stimulus, i.e. the process is tapped while it is still in full swing. Off-line tests allow the subject time to reflect before issuing the response. In the intelligibility tests used in this dissertation we only used off-line measures. This was not a principled choice but merely one based on convenience. Especially because a large part of the testing had to be done in the field (in China and in the United States), where we had no easy access to laboratories with possibilities to run on-line tests efficiently (using multiple work-stations), we were content to use more traditional off-line tests. These can be administered to small groups of listeners in parallel, without the need of sophisticated equipment.

2.3.2 Functional tests of intelligibility

As explained above, intelligibility involves the correct recognition of linguistic units in their linear order. Units exist at various levels of the linguistic hierarchy. The lowest level, with the smallest units, is that of the individual speech sounds or phonemes, i.e. the vowels and consonants. These are in principle units that carry no meaning of their own. This is a characteristic they share with consonant clusters. The smallest linguistic unit with a meaning of its own is the morpheme. Since morphemes often coincide with short words, we will collapse the word and morpheme levels for the purpose of the present study. In the following subsections we will briefly present test methods that have been devised to determine the intelligibility of speech at the level of the phoneme (consonants, vowels, clusters) and at the word level.

2.3.2.1 Intelligibility of consonants (onset, coda)

When discussing techniques that have been developed for the testing of the intelligibility of consonants, we will not deal with perceptual experiments that target single contrasts. For instance, there is a very substantial literature on the perception of the /r/ ~ /l/ contrast in English by Japanese learners. Rather, we will survey techniques that determine the intelligibility of all the consonants in the target-language system, so that the results may be used diagnostically to determine which consonants, and which contrasts between them, are a learning problem and which ones are easy. Sounds in language naturally occur in the context of a word. So it would seem reasonable to present listeners with words containing the various consonants that make up the phoneme inventory of the target language. When doing so, however, one runs the risk that the (foreign) listener will correctly determine the identity of a consonant without actually having heard the sound correctly. This may happen as a result of lexical redundancy. For instance, when the listener hears the rhyme portion of a monosyllabic word such as /…ɔp/, only some consonants can be considered, viz. /p, t, k, m, ʃ, tʃ, h/ as in *pop, top, cop, mop, shop, chop* and *hop*, respectively. All other singleton consonants would not qualify as they do not combine with the rhyme /…ɔp/ to make up existing words in the English lexicon. As a result, getting the identity of the onset consonant in /…ɔp/ right would be a mixture of using bottom-up information provided by the consonant signal and top down information supplied from the lexicon. Unfortunately, there is not a single word frame in the English language that would allow each of the consonants in the language to appear in the onset position (let alone in intervocalic position). Therefore three other ways are commonly used to overcome the confound with lexical redundancy in intelligibility testing. The first is to present rhymes that allow only a small set of consonants to be filled in and list a closed set of printed alternatives (typically four) for the listener to choose from (MRT or Modified Rhyme Test). We refer to Van Heuven and Van Bezooijen (1995) for a more elaborate survey of testing procedures. Although this procedure does not eliminate lexical redundancy as a source of extra information, at least its effect is kept

constant. The second possibility is to have the full set of consonants embedded in a _VC frame, and force the listener to choose from the full inventory whether the response would be sense or nonsense. This creates the risk that the listener may have a lexical bias such that response alternatives that are words will be favored over alternatives that do not yield an existing word. The third possibility is to embed the consonants in fixed V_V structures that always result in nonsense items, so that the risk of lexical bias does not arise. In our experiments we have chosen for the latter option, as it is a highly efficient solution that does not involve the risk of lexical bias.

The intelligibility of consonants may differ substantially depending on their position in the syllable. It has been found that, generally, consonants in onset position are more difficult to identify correctly than the same consonants in coda position. One reason for this asymmetry might be that, across languages, the number of different coda consonants is smaller than the number of onset consonants. Mandarin Chinese, for example, has a set of 21 consonants which may appear in the onset, of which only the nasals /n/ and /ŋ/ remain as possible coda consonants. In Dutch and English the distribution is less lopsided but still asymmetrical: Dutch has 17 onset consonants against 11 in the coda, and English has 23 versus 21 (lacking /w, j, h/ but including /ŋ/, which is not an onset consonant), respectively. In our intelligibility tests of consonants we concentrated on the onset position.

2.3.2.2 Intelligibility of vowels

When it comes to testing the intelligibility of vowels in languages such as English and Dutch, there is generally no need to resort to the use of nonsense items. As it happens, there are fully productive consonant frames that allow the insertion of any vowel in the inventory of the language and still yield a meaningful, existing word or phrase. The most widely used context for vowels in English is the /hVd/ frame. This frame was first used in the classical study by Peterson and Barney (1952), and has been used over and over again in later studies. Because of its wide-spread acceptance, we decided to follow established practice here, and adopt the same methodology.[10] This obviates the need for rather cumbersome and time-consuming tests such as the Minimal Pairs Intelligibility Test, which does test minimal vowel pairs in sentence context (Van Santen, 1993).

[10] In the classical study of Dutch vowel formants, Pols, Plomp and Van der Kamp (1974) used the /hVt/ frame. Given that Dutch has final devoicing, this seems a reasonable substitute for the English /hVd/ frame, were it not that the /hVt/ leaves several accidental gaps, so that the stimulus set is a mixture of sense and nonsense words. There is in fact only one fully productive consonant frame for Dutch, which is /rVt/ - thanks to the existence of proper nouns such as *Ruud* /ryt/ (short for Rudolph) and /rət/ *Ruth*). The problem here, however, is that the /r/ has many allophones (differing among other things in place of articulation, i.e. apical versus uvular) and that it is difficult to segment from the vowel.

2.3.2.3 Intelligibility of clusters

The importance of consonant clusters in English should not be underestimated. About 40% of one-syllable words in English begin and 60% end with consonant clusters (Spiegel, Altom, Macchi and Wallace, 1990). The Bellcore Test and the CLID Test have been developed to fill this gap in the test batteries. The CLID Test (CLuster IDentification Test, Jekosch, 1994) is a very flexible architecture which can be used for generating a wide variety of monosyllabic stimuli (e.g. CCV, VCCC, CCCVVC) in an in principle unlimited number of languages, as long as matrices are available with the phonotactic constraints to be taken into account. Both the intelligibility of initial and final consonants and of (sequences of) medial vowels can be tested. In contrast to the CLID Test, the Bellcore Test (Spiegel et al., 1990) has a fixed set of stimuli, comprising both meaningless and meaningful words. Sequences of consonants are tested separately in initial and final position. The test has been applied to assess the intelligibility of two speech synthesizers compared with human speech presented over the telephone. The syllable score for human telephone speech was 88% correct.

In our experiments we only tested the intelligibility of consonant clusters in onset position preceded and followed by the vowel /ɑ/. We included 17 double consonant clusters /aCCa/ and supplemented these with three triple consonant clusters /aCCCa/. In this way the same format could be used as the one we employed in the case of single consonants. All the stimuli used were nonsense items. As a result, we did not test the intelligibility of clusters in coda positions, nor could we test for possible interactions of consonant articulation and the coarticulated vowel segments. It was felt, however, that the materials selected covered a sufficiently wide range of potential pronunciation problems of foreign learners of English. Including an even larger set of materials would have rendered the experiment unmanageable.

2.3.2.4 Word recognition tests (on-line, off-line)

The (monomorphemic) word is the smallest unit in the language that links a sound shape with a meaning. We assume that words are stored in the mental lexicon, where they are specified, among other things, in terms of their sounds and the order of the segments them, rhythmic structure (number of syllables and the position of the stressed syllable, in so far as the language has stress), syntactic properties, and meaning. Intelligibility was defined above as the extent to which the words in an utterance can be recognized in the same order as they were produced by the speaker. Word recognition tests therefore play a prominent role in intelligibility testing.

Word recognition can be restricted to isolated target words that are presented without any spoken context. This is convenient for diagnostic purposes. If the listener fails to recognize the word, the problem should be located in the word itself. However, words in everyday communication hardly ever occur in isolation. Therefore, word recognition in connected speech is a more realistic test. When the recognition of a target word fails, however, it is difficult to determine whether the cause is in the target word itself or whether the failure is due to the fact that some earlier words were not recognized and failed to constrain the identity of the later

target word. A solution to this dilemma has been found in presenting the same target both in context and in isolation. This is a laborious solution, however, as the stimuli must be blocked over two groups of listeners (who have to be equally proficient) in order to prevent learning effects (priming).

In our experiments we used two types of word-recognition tests. The first used so-called Semantically Unpredictable Sentences (SUS-test, Benoît, Grice and Hazan, 1996), in which simple (often monosyllabic) words were presented in sentences but where the sequence of words made no sense, as in *The state sang by the long week* (for more details on the SUS test, see § 4.2.4). The content words are not made more predictable by earlier content words in the utterance, so that the test is actually an accumulation of single word recognition items. The second word recognition is the Speech In Noise (SPIN) test. This test requires listeners to recognize sentence-final words which are or are not predictable from the earlier context (Kalikov, Stevens and Elliot, 1977). This test will be discussed below in § 4.2.5. Both tests require the listener to write down the target words by way of dictation. There are no severe time constraints on the task performance, so that these are basically off-line word recognition tests, which inform us to what extent the word recognition was a problem for the listener, but disclose nothing about the ongoing word recognition process.

There are several on-line techniques that tap the word recognition process in real time. Since we chose not to use on-line techniques (see above) we will be brief about them. Most on-line techniques require listeners to detect the presence of some feature of a word, by pressing a response key as quickly as they can manage. The response time is indicative of the moment the word was recognized. The features to be detected can be of several kinds. In phoneme detection, listeners are instructed to press a button as soon as they detect its presence in the stimulus. Generally, each individual sound making up a word is acoustically unreliable. Therefore listeners wait until they have recognized a word in a sequence of sounds. Phoneme detection is therefore indicative of the time it takes to recognize the word that harbors the target. The more difficult the word is to recognize, the longer the target detection will take. Alternatively, listeners may be instructed to detect the presence of some semantic feature, e.g. press a button when they hear the name of an animal, or when they hear a word that expresses a tangible object (rather than an abstraction). Here the rationale behind the test is that the listener may only access the meaning of a word in the mental lexicon after the word has been recognized, so that semantic property detection again is indicative of word recognition time. An interesting alternative that was found to discriminate quite clearly between native and non-native listeners (Poelmans, 2003) is the lexical decision technique. Here the listeners are presented with sound sequences that either constitute a word or a nonword. The subject is instructed to press one of two buttons marked ‘word’ or ‘nonword’, depending on his decision. Obviously, the decision that the stimulus is a word can only be made once the word is recognized, so that this task, too, indicates the time needed (and thereby the difficulty) for word recognition.[11]

[11] Interestingly, discrimination of native and non-native listeners was most clearly achieved in the (correct) rejection of non-words rather than in the (correct) acceptance of words in Poelmans (2003).

2.3.3 Functional tests of speech understanding (comprehension tests)

Listeners have understood (or: comprehended) a sentence or longer stretch of speech if they have grasped the meaning of the passage. There are several functional tests that have been employed to test the quality of the listener's understanding or comprehension. Broadly, these methods can be of four types: (i) having listeners answer questions on the contents of the passage, (ii) verifying the truth of sentences, (iii) verifying on-line descriptions of still pictures or video footage, and (iv) carrying out spoken instructions.

Comprehension questions. Either before or after the stimulus speech passage is presented, listeners are given a specific question that can be correctly answered only if they understood the contents of the passage. Asking the question before presenting the passage diminishes memory load and tests intentional listening. Asking the question post hoc makes heavier demands on memory and may therefore be less desirable. The comprehension questions can be of the open or closed type. Open questions ask listeners to formulate and write down their answers from scratch, closed questions present the listeners with two, three or four alternative answers from which they have to choose the correct one. Answering comprehension questions are off-line tests. Listeners have ample time to think or recall the speech and give the answer.

Sentence verification tests. Here listeners are asked to judge whether a sentence they hear, expresses a truth, is unlikely, or nonsense, by pressing one of two keys marked 'true' or 'false' immediately after they hear a sentence. For instance, a stimulus *People wear their hats on their feet* should be responded to by pressing 'false'. In this kind of test, listeners can make the right choice only if they understood the sentence correctly.[12] Sentence verification can be used as an on-line comprehension testing technique. In order to do so, subjects should be asked to press the response key as fast as they can manage, if possible even before the end of the speech utterance has been reached.

Descriptive language. One way to test comprehension of speech is to ask a listener to indicate whether a spoken description does or does not match a visually presented scene, by pressing one of two keys marked 'correct' or 'wrong'. The visual presentation can be a in the form a still picture or it can be a scene from a movie. By aligning the spoken description with the development of the scene (only possible in video footage) on-line comprehension can be tapped.

Carrying out instructions. A last test we want to mention here relies on carrying out spoken instructions. The listener is asked to carry out certain actions following spoken instructions recorded on tape. Obviously, listeners can only carry out the instructions if they understood them. The implementations of this technique range from crude to sophisticated. A crude but effective use of the technique is the Token Test, which has been in service for decades to test speech understanding with patients suffering from brain lesions. The patient has several geometric objects on

[12] Given just two response alternatives (true, nonsense) the chance of getting the correct response by guessing is 50%. The role of guessing quickly diminishes with the number of items in the test. When a binary verification test comprises 50 items, the chance of answering all items correctly by pure guessing is very small.

the table in front of him (circle, square, triangle) which may have different colors. Instructions are of the type: 'Put the red square on top of the yellow triangle'. Van Heuven (1986) describes a similar technique used to determine differences in comprehensibility of several types of deviant (foreign-accented) Dutch speech, and shows that the technique is very sensitive. More recent versions of the technique no longer require the physical manipulation of objects in space but instruct the listeners to manipulating objects on a screen by moving a joystick or computer mouse, which also affords easy measurement of response time. The instruction tests can be conceived of as on-line tests.

Comprehensibility and intelligibility have been found to be related with global foreign accent scores, but since we will test intelligibility rather than comprehension, we will not further discuss comprehension tests in the present study.

2.3.4 Information reduction techniques

Especially when the intelligibility of speech produced by native speakers is heard by native listeners, performance levels tend to be close to ceiling, so that small differences in proficiency between speakers or between listeners are hard to detect reliably. An often used solution to ceiling effects is the use of information reduction in the stimulus. The underlying idea is that speech is a highly redundant code, and that native listeners may use the redundancies better than non-native listeners. There are several signal degradation techniques that can be used to reduce the redundancy in the spoken word forms. First, we may obscure the speech signal with noise so that some of the distinguishing properties of the word are no longer audible. Second, we may eliminate certain frequencies or frequency bands from the signal, such that important distinguishing frequencies are no longer available. Third, we may simply eliminate complete segments or larger parts of the speech signal by replacing them by stretches of silence or by noise, without changing the temporal relationships among the sounds that remain. We will briefly discuss these techniques and review how they have been used in the study of intelligibility of (foreign-accented) speech.

2.3.4.1 Speech in noise

One of the earliest applications of speech in noise has been the development of testing materials for audiological purposes. In order to determine the extent of hearing loss with hard-of-hearing patients the threshold of hearing may be determined by asking listeners to recognize words presented to them in noise. On the first pass the noise is much stronger than the speech, so that the word cannot be recognized, not even by a healthy listener. On successive following passes the noise level is reduced in steps of, say, 3 decibels, until the spoken word is loud enough – relative to the reduced noise level – to be recognized. The signal-to-noise ratio (expressed in dB) at which the word can be recognized is the intelligibility threshold. Using this measure, differences in intelligibility of different words, spoken by different speakers, or heard by different individuals (whether native or non-native, whether hearing-impaired or healthy) can be determined. A well-known

set of audiology sentences to be presented in noise was developed for the SPIN test (Speech In Noise) by Kalikow et al. (1977). In this test, simple monosyllabic target words were presented at the end of simple short sentences, which came in two varieties. In one type of sentence the final target word was used in citation form, such that its identity was not constrained by the preceding context, as in *They are discussing the (map)*. In the other condition the target word was strongly constrained by the context as in *She cooked him a hearty (meal)*. Results show that the predictable words were recognized at more severe signal-to-noise ratios than the unpredictable words. The same techniques, and even the same test materials, have later been used to test differences in intelligibility of synthetic speech (Van Bezooijen and Van Heuven, 1997 and references therein) and non-native speech production and perception (e.g. Van Wijngaarden, 2001).

It has been shown that speech is more or less effectively masked depending on the specific type of noise. For instance, white noise (in which all frequencies occur with equal chance and with equal amplitude) is a less effective masker than noise which has roughly the same spectral distribution as speech, i.e. with emphasis on the lower part of the spectrum. Thus, pink noise (−3dB/octave) and even more strongly, ANSI noise are more effective maskers. The most effective type of all is so-called speech noise (also called babble noise) which is actually speech produced by the same speaker as the individual who spoke the target stimulus, dubbed several times with different phasing, see Eggen (1989). A second parameter in using speech in noise is whether the noise has constant intensity (so that loud sounds exceed the noise but low-intensity sounds – typically consonants – are completely masked) or whether the noise is modulated so as to follow the intensity contour of the speech stimulus.

In our experiments we used SPIN sentences but presented them without any added noise. As it turned out, the quality of the non-native speech (and of the non-native listeners) was so poor that the intelligibility was evenly distributed in the 30 to 90-% range.

2.3.4.2 Filtering

It is a well-known phenomenon that a foreign speaker may successfully communicate with a native listener under ideal circumstances but that communication tends to break down when the telephone is used. The reason for the breakdown is that the telephone filters speech such that only frequencies in the restricted band between 300 and 3300 Hz are transmitted. The impoverished signal contains enough information for successful communication between two native speakers, who know the code, but when either the speaker or the listener is foreign, the signal is too poor to allow full intelligibility. For this reason filtered speech (high-pass, low-pass and band-pass) has been used as a means of degrading the speech stimulus in an attempt to make fine-grained determinations of differences in intelligibility of various types of materials; for instance French and Steinberg (1947), Hirsh, Reynolds and Joseph (1954), Miller and Nicely (1955), and others after these pioneering studies, used filtering in intelligibility testing.

2.3.5 Gating

The classical gating study was done by Grosjean (1980). In this study listeners were first presented with a short initial portion of a target word and asked to guess what the word would be. On second and later passes the audible portion of the target word was lengthened by one phoneme at the time, until the listener could supply the entire word. Highly frequent words and words in more constraining preceding contexts could be finished from short onset fragments than low-frequency words in less constraining contexts. Nooteboom and Truin (1980) used the same technique and showed that native Dutch listeners could recognize target words from shorter onset fragments than non-native (English) listeners of Dutch. Smeele (1985) showed that native-accented Dutch words were recognized by Dutch listeners from shorter onset portions that German-accented Dutch words.

Variations on the gating paradigm abound. Instead of suppressing the final portion of the target word, some researchers have suppressed the initial portion (replacing it by either silence or by noise – the latter condition proves more conducive to intelligibility). Moreover, the portions of the signal that are suppressed (zeroed) or replaced by noise need not be contiguous.

2.4 Can higher-order performance skills be predicted from lower-order components?

Sentences are made up of words, and words are made up of syllables, which in turn are composed of vowels and consonants. From the perspective of the speaker, being able to pronounce the sounds in a word correctly is a skill that has relatively little to do with the skill of arranging the words to form a syntactically appropriate sentence. Pronunciation would seem to involve a good deal of motor skills, whilst arranging the morphology and syntax is a much more cognitive skill. In previous sections we discussed the notion of a critical period which allows the formation of native or near-native pronunciation skills in a second language. No mention is ever made of a critical period needed for the acquisition of the morpho-syntax of a second language. This in it self would seem to suggest that the lower-level phonetic skills have little in common with the higher-order syntactic skills. On the strength of this argument we would expect a weak correlation between lower and higher-order skills when it comes to speech production in a second language.

When we approach the problem from the perspective of the listener, the argument will be different. It would make sense to predict that words with poorly articulated sounds will be hard to recognize, and sentences made up of poorly recognized words will not be understood. Accordingly, we may ask how well word recognition scores can be predicted from the consonant and vowel identification scores obtained for the same speaker. If the sounds can be successfully identified then we expect the word recognition scores for the same speaker to be high, too. We may also ask the question which of the two sets of sounds would be the better predictor of word recognition, vowel identification or consonant identification scores.

In order to answer such questions one needs data for a fair number of speakers, who range between poor and good. In the data to be collected in the present project we recorded speech from groups of 20 American, 20 Dutch and 20 Chinese speakers of English, who pronounced vowels, consonants and words in context. Vowel and consonant identification scores were obtained for all 60 speakers but only with a subset of the materials. Full vowel and consonant identification scores as well as word recognition scores were collected from only six speakers (two Chinese, two Dutch and two Americans). No word recognition was tested for the remaining 54 speakers. Given the extremely small number of speakers, trying to predict word recognition in SUS sentences and in SPIN sentences is hazardous. We will nevertheless present an analysis of the relationships.

2.5 Problems at the phonological level vs. phonetic level

When analyzing problems in the acquisition of the sound system of a second language, it has been customary to distinguish phonological problems from phonetic ones. It is not always clear if there is a boundary between these two disciplines, and if the distinction is useful at all when applied to foreign language acquisition. We will make an effort to separate the two, and consider what the distinction may contribute to our understanding of the problems.

2.5.1 Phonology

By phonology we mean the abstract structure of the sound system of a language, in the abstraction of the precise phonetic implementation of the sound categories. Properties that can be studied in terms of abstract structure are the number of sounds in the inventories of source and target languages, the distinctive features needed to organize the sounds in the inventory in contrasts (oppositions) and the constraints on the formation of legal syllables. Differences between positional allophones within the same phoneme category are also subsumed under the heading of phonological structure.

2.5.1.1 Differences in inventories (number of sounds, oppositions)

Obviously, two languages differ in their sound system if there is a difference in the number of sounds between the two languages. In our research we will study the type of English spoken by speakers with Dutch and Chinese as their native language. The number of phonemes in English, Dutch and Chinese differ considerably. Generally, the Dutch inventory is more like the English system than the Chinese system is. For instance, in both Dutch and English the number of vowels is much larger than in Chinese. In order to divide the vowel space among the vowels in the inventory, English and Dutch vowels must differ along more parameters than the vowels in the Chinese inventory. Typically, then, Dutch and English vowels are differentiated in addition to other parameters, by their length (or tense~lax). Chinese does not employ a length (or tense~lax) contrast, so that presumably Chinese learners of English will have at least one problem to overcome that should be easy for Dutch learners, i.e. learning to use the length contrast in English. In Chapter three we will present an

overview of the inventories of the three languages involved in the present study, and indicate by what articulatory features the various contrasts in the systems can be accounted for. We will make an effort to predict what kind of learning problems will be seen in the results of our experiments, following the models of Flege (SLM) and Best (PAM).

2.5.1.2 Differences in syllable structure (no clusters, simpler clusters, no coda)

Even if we know the size and internal structure of the phoneme inventories of source and target languages, there are quite a few systemic properties that we cannot yet deal with. Languages differ widely in the way they build syllables from vowels and consonants. The simplest syllable type across languages is composed of a consonant (C) followed by a vowel (V). More complex syllable types can be formed either by omitting the initial consonant or by augmenting the number of consonants in the onset (i.e. preceding the vowel) or in the coda (i.e. following the vowel). Chinese has a predilection for simple syllables, whereas English and Dutch allow rather complex syllable types with clusters of multiple consonants in onset and/or coda.

Syllable types across languages are implicationally ordered, such that all the simple syllable types allowed in strongly constrained languages (such as Chinese) will also occur in less constrained languages with many complex syllable types (such as English and Dutch) but not vice versa (Blevins, 1985). Therefore, learners with a constrained source language will have a problem in producing the more complex syllable types in a less constrained target language. So we predict that Chinese learners of English will experience considerable problems when having to produce (and perceive) the complex English syllable types. Typically clusters will be broken up into several syllables by inserting epenthetic vowels between consonants that are not legal clusters in the source language. Dutch is probably a less constrained language than English, allowing more and more varied consonant clusters both in the onset and in the coda, so that generally we do not predict any problems with English syllable types for Dutch learners.

2.5.1.3 Positional allophones (final devoicing)

In the preceding subsection we introduced the difference between onset and coda. It happens very often that a language uses clearly distinct allophones for the same phoneme such that one allophone occurs only on the onset position whereas the other appears in the coda only. English /l/ has two varieties, called clear (or light) and dark (also: dull or velarized). The clear /l/ is restricted to onset positions whilst the dark allophone is bound to coda positions. In this respect English and Dutch behave identically, so that we predict no learning problem here for Dutch learners of English. Chinese learners of English have a double problem. First they do not normally allow syllable types with a coda consonant – the only exception being codas with a nasal in them – so that learning to pronounce an /l/ in the coda is not only a problem as such, but also one that is aggravated by the fact that Chinese learners will have to learn to pronounce a clear /l/ in the onset and differentiate this from a dark /l/ in the coda.

A second, related problem is that certain sounds may occur on the onset that are not allowed in the coda. As a case in point, Dutch has an opposition between voiced and voiceless obstruents, as does English, but only in onset position. The

voiced~voiceless opposition is neutralised in coda position, where only the voiceless member may occur. In English, however, voiced obstruents abound in coda position. This presents a great learning problem for Dutch learners of English. Even though the source language has both voiced and voiceless obstruents in the onset position, Dutch learners find it very difficult to pronounce a voiced obstruent in the coda, as they should when they speak English. Apparently, then, sound contrasts do not generalise to other syllable positions than those they occupy in the source language. The Chinese learning situation presents an interesting test case. Chinese has voiced and voiceless obstruents in the onset, just like Dutch and English). However, Chinese has no coda consonants at all (with the exception of the nasal coda) so that, possibly, the Chinese learner of English is not impeded in setting up the voiced~voiceless contrast in the coda. It is unclear if this structural property would give the Chinese learner an advantage over his Dutch counterpart when learning English.

2.5.2 Phonetics

What is left for phonetics is the implementation of the categories and the contrasts between them. If two systems, say in source and target language, have the same number of sounds, organized in terms of the same oppositions, there may still be considerable differences in the division of the articulatory space between the various categories, and in the way the boundaries between the categories are cued by acoustic features

.

2.5.2.1 Same oppositions but different boundaries

Source and target languages may have the same opposition along the same parameter with the same number of categories along the parameter, and yet differ in the phonetic implementation of the contrast. As an example, consider the phonetic implementation of the tense~lax contrast in stops in English, Dutch and Chinese. In each of the three languages stop consonants in the onset fall in one of two phonological categories, probably best characterized as tense (fortis, voiceless) as opposed to lax (lenis, voiced). The phonetic parameter that carries the contrast is often called Voice Onset Time (VOT). This is a complex parameter in that it is defined as the time interval between the moment of the release of the stop consonant (coinciding with a short noise burst and a sudden increase in energy at higher frequencies) and the moment the vocal cords start vibrating. In Dutch the lax member of the contrast has negative VOT, meaning that the vocal cords begin vibrating some 50 ms before the stop is released, i.e. there is glottal pulsation while the mouth is closed (‘prevoicing’). The Dutch fortis member has 0 VOT, so that glottal pulses are produced as soon as the mouth opens. The English tense~lax opposition is phonetically different. Here the lax member has 0 VOT, while the tense counterpart has positive VOT, meaning that the vocal cords do not start vibrating until well after the consonant is released. The time interval between the release burst and the onset of glottal pulsation is filled up with a whispered (voiceless) vowel sound called aspiration. From this difference in phonetic implementation we predict that Dutch speakers of English have a problem: they will use the Dutch implementation of the contrast when speaking English, with the result

that an English listener may well mistake the Dutch /p, t, k/ for English /b, d, g/, respectively, since all these sounds have 0 ms VOT. Chinese has roughly the same phonetic implementation of the tense~lax opposition in onset stops as in English, so that Chinese speakers of English would not have any problems here.

2.5.2.2 Different cue tradings

Typically a phonological contrast is phonetically cued not along a single acoustic parameter but along multiple parameters simultaneously. For instance, the difference between tense and lax members of vowel contrasts in Dutch and English are cued by differences in duration (the tense members are some 50% longer than their lax counterparts) as well as by differences in vowel quality (timbre). Typically, the lax vowels assume more centralized positions in the articulatory vowel space than the tense counterparts. In German, however, the difference between the members is cued only by a difference in duration while the vowel quality differences are negligible (Strange et al. 2004). To compensate for the lack of a qualitative difference, the German tense and lax vowels have a larger difference in duration than in English. English speakers of German may apply the English implementation of the tense~lax contrast so that the durational difference in their German vowels is too small to be effective, while the Germans will be relative insensitive to the quality differences in English-accented German. A similar problem was noted by Van Heuven (1986) in the perception of Dutch vowels by Turkish learners. The Turkish learners revealed a very clear tense~lax contrast in the Dutch open vowel pair /a: ~ ɑ/ but were sensitive only to the duration cue and failed to pick up the quality cue (the tense member have higher first and second formant values than the lax counterpart) and consequently misclassified about half of the vowel tokens.

2.6 Concluding remarks

The purpose of this study is not to carry out any systematic testing of theories of second-language acquisition of sound systems. The above survey of such theories was merely offered as background information enabling the reader to understand why certain choices were made in the experimental part of my project. Nevertheless, whenever my results give rise to reflection on theoretical positions, I will do so and refer back to sections of the present chapter.

Chapter three

Contrastive analysis

A classic question in phonetic theory is: "What sounds can a language have?" It has been asked about vowel inventories (Lindblom, 1986) and consonant inventories (Lindblom and Maddieson, 1988). Every speech sound belongs to one or other of the two main classes known as vowels and consonants. Describing the vowel and consonant inventories is a start in describing the salient phonetic structure of a language. The standard view is that the sound inventory in a language is the result of two competing forces: one favors sounds that are easy to produce while the other force pulls the system towards more distinctiveness, i.e. maximal contrast between elements of the system (Lindblom, 1986). The native language (source language) of L2 learners plays the role of a 'Phonological filter' which deeply influences the L2 learners' pronunciation so that it deviates from that of native speakers of the target language (Polivanov, 1931; Trubetzkoy, 1939/1969). As a result L2 learners have perceptual blind spots which lead to perceptual errors. These blind spots prevent the L2 listeners from identifying the foreign phonemes correctly. Instead, they substitute their own L1 sounds for the foreign phonemes.

In this chapter I will give the sound inventories of three languages, General American English (GA),[1] Standard Dutch (Algemeen Beschaafd Nederlands, ABN)[2] and Standard Mandarin Chinese (Putonghua).[3] These are the native languages of the three groups of speakers and listeners that will be studied in the present dissertation. More details on the choice of experimental subjects will be provided in Chapter four.

3.1 Vowels

A vowel is defined as a typically voiced sound in the production of which the air issues in a continuous stream through the pharynx and mouth, there being no obstruction and no narrowing such as would cause audible friction. Vowels are

[1] The American subjects for the final experiment in this research are from Los Angeles, California, which is generally regarded as a place where GA is used. Californian English is a dialect of the English language spoken in the U.S. state of California. As a variety of American English, Californian English is similar to most other forms of American speech in being a rhotic accent, which is historically a significant marker in differentiating different English varieties.

[2] The Dutch subjects come from the cities around Leiden called 'city belt' (Dutch: Randstad), where people speaking standard Dutch, ABN, can easily be found.

[3] The language for the Chinese subjects in our experiment is Standard Chinese, which is the present-day dialect of Beijing promulgated as a standard language in Mainland China, Taiwan and Singapore. Our Chinese subjects come from the northeast of China, Changchun, where people speak the Northeast dialect, which is very close to Standard Mandarin Chinese.

usually described in terms of quality and duration. Since vowels are distinguished from one another chiefly by whether they are produced in the front, centre, or back of the mouth, whether the tongue position is high, mid or low, and whether the lips are spread or rounded, the basic building blocks of most vowel systems are the three qualities, as many of the vowels of the world's languages can be described simply by the three traditional dimensions high-low, back-front, and rounded-unrounded.

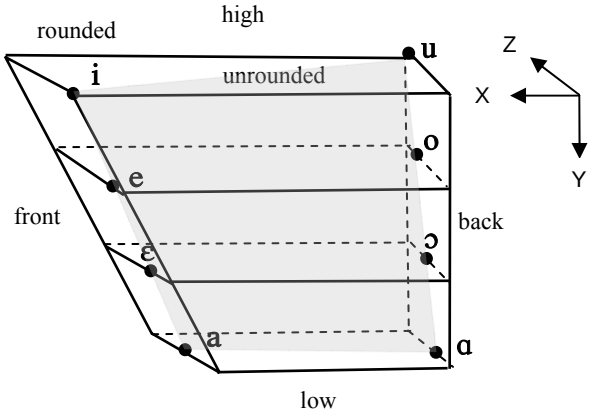


Figure 3.1. The location of the eight cardinal vowels in a three-dimensional articulatory vowel space defined by backness (X dimension), height (Y dimension), and rounding (Z dimension) [after Ladefoged (1971: 72), Ladefoged and Maddieson (1990: 94)].

This figure shows the location of a set of reference vowels, i.e., the cardinal vowels described by Jones (1956), within a space defined by these dimensions. What Jones effectively gave phonetics in his CV system (Cardinal Vowel system) was a mapping system which presented what is essentially auditory and acoustic information in a convenient visual form.[4] It is the only widely used system for vowel description. It gave phoneticians a yardstick for measuring the vowel quality which is invaluable in phonetic description.

Another element which is considered by some to be of importance in determining vowel quality is the state of the tongue and lips as regards muscular tension. Those who consider that vowels may be differentiated by degrees of muscular tension distinguish two classes, tense vowels and lax vowels. Tense vowels are supposed to require considerable muscular tension on the part of the tongue; in lax vowels the tongue is supposed to be held loosely. The difference in quality between the English vowel *seat* and *sit* is described as a difference in tenseness: the vowel in *seat* is considered tense and the vowel in *sit* lax (Jones, 1956).

[4] Jones's system has been be criticized by Collins and Mees (1981) as follows: '[Jones] took no account of the significance of the root of the tongue and its relationship to the pharynx wall. Indeed, he disregarded the pharynx cavity altogether, mentioning only tongue height in his theory. Later research has shown that it is the relative sizes of the oral and pharyngeal cavities which are the crucial factors in vowel quality'.

In some languages there are vowels which are distinguished by duration alone. For instance, in Danish, there is an opposition between long and short vowels, and in Estonian even between short, long and superlong (Lehiste 1970). In many languages, similar oppositions between sets of vowels are also marked by differences in vowel quality. Such combinations of duration and vowel quality are employed in Dutch and English.

3.1.1 Vowel inventories in the three languages

In the following sections I will compare the vowel inventories of English, Dutch and Mandarin Chinese. I will present literature data consisting of structural vowel tables published for the languages and of formant measurements. Formant measurements have been used since 1950 as an semi-objective way to determine vowel quality. The technique will be discussed at greater length in Chapter four; for the purpose of the present chapter it is sufficient to know that the centre frequency of the lowest resonance in the speech signal (first formant, F1) varies with the degree of mouth opening and that the second-lowest formant (F2) corresponds inversely with the degree of backness. Acoustic vowel charts plot F1 from top to bottom against F2 from right to left; in this way the configuration of vowel points assumes the same orientation as in a traditional articulatory vowel chart, with /i/ in the top left-hand corner, /u/ in the top right-hand corner, and /a/ at the bottom. In the charts we present below, we did not plot the formant frequencies in hertz but transformed the hertz-values to Bark units. Equal distances in the Bark space correspond to equal differences in perceived timbre (or: vowel quality). For details we refer to Chapter four.

3.1.1.1 English vowels

The vowel system of General American English (GA, as exemplified for instance in the American English pronouncing dictionary by Kenyon and Knott, 1944, see also Gussenhoven and Broeders, 1976: 186-195) is best described as composed of four vowel heights and three degrees of backness. Height is a four-level parameter with high, high-mid, low-mid and low as the phonetically relevant degrees. Backness has three degrees, viz. front, centre and back. English has a split in its vowel system such that most vowels are tense (long duration, peripheral articulation) but some are lax (short duration, more centralized pronunciation). The four degrees of height are defined on the tense vowel set; the back vowels require four degrees. When tense and lax vowels are kept apart, three degrees of height suffice for the front vowels (high/close for /iː/ (*heed*, *bead*), mid for /eː/ (*hayed*, *stayed*) and low/open for /æ/ (*had, mad*). For the back vowels, however, we have to distinguish between high/close /uː/ (*who'd*, *mood*), high-mid /oː/ (*hoed*, *showed*), low-mid /ɔː/ (*hawed*, *clawed*), and low/open /ɑː/ (*father*). In many American dialects and probably also in General American, /ɔː/ and /ɑː/ have merged (Wells, 1982; Labov, Ash and Boberg, 2006), simplifying the vowel system to three degrees of height, and restoring symmetry between front and back vowels. The lax vowel set comprises just two degrees of height (high vs. low). Textbooks on British English often mention so-called centring diphthongs as an extra set of vowel phonemes, as in *fear*, *fair*, *poor*.

These vowels could be claimed to be phonemes on the strength of such minimal triplets as *bead* ~ *beard* ~ *bid*. It seems to me, however, that these vowels can be treated as positional allophones of tense vowels followed by coda-/r/. The reason why the underlying vowel should be tense rather than lax has to do with the phonotactics of the centring diphthongs: they cannot be followed by any other consonants than the alveolars (/t, d, s, z/), which is the same environmental constraint that applies to other tense vowels; lax vowels can be followed by a larger variety of consonants and clusters, which are not possible as codas after murmur diphthongs. For instance, centring diphthongs cannot be followed by coda clusters except when the last consonant is one of the set /t, d, s, z/, i.e., the set covering the suffixes used to code plural, past tense, or third person singular after stems with either voiceless consonants (/-t, -s/) or with voiced sounds (/-d, -z/). Quite probably also [ɔ:] should be analysed as a surface phenomenon in non-rhotic varieties of (British) English. In rhotic varieties, and especially in General American, this vowel sound can be analyzed as /ʌ/ followed by coda-/r/.

GA has two diphthongs, /ai, ɑu/, which start at an open position and glide towards a close position along the front and back side of the vowel space, respectively. The third diphthong is /ɔi/, which runs from back to front in the mid part of the space. Table 3.1 summarizes the vowel inventory of GA. The unstressed neutral vowel schwa (/ə/) is not included in table 3.1.

Table 3.1. The General American vowel inventory. Vowels in parentheses are allophones in GA before /r/, but have surface-phonemic status in RP English.

	front		central		back	
	tense	lax	tense	lax	tense	lax
High	i: (ɪə^{r})				u: (ʊə^{r})	
hi-mid	e: (ɛə^{r})	ɪ			o: (ɔə^{r})	ʊ
lo-mid		ɛ	ɜ:^{r}	ʌ	ɔ:, ɔi	ɔ
Low	ai	æ			ɑ:, ɑu	

Figure 3.2 presents the classical formant data collected for American English by Peterson and Barney (1952) drawn separately for male (squares) and female (circles) speakers, and broken down by the tense (solid lines) versus lax (dotted lines) subsystems.[5] The F1 and F2 frequencies have been transformed to Bark units, in

[5] Peterson and Barney (1952) identified the primary acoustic features of the American English vowels on the basis of /hVd/ productions by 28 women, 33 men, and 15 children (ages not specified). They found a general correspondence between vowel type and frequencies of the first and second formants (F1 and F2). Hillenbrand, Getty, Clark and Wheeler (1995) replicated and extended the Peterson and Barney study. Hillenbrand et al. sampled 45 men, 48 women, and 46 10- to 12-year-old children. Analysis of formant data by Hillenbrand et al. showed differences from the formant data in the Peterson and Barney study, both in terms of mean frequencies of F1 and F2, and the degree of overlap among adjacent vowels. However, the data were similar to Peterson and Barney regarding vowel-specific formant frequencies, as well as change in formant values according to vocal tract size and shape.

order to create a visual display in which equal distances between vowels represent auditorily equal differences in vowel quality (timbre). Note that the vowel /æ/, which is a lax vowel in terms of its distributional properties (may not occur in an open syllable at the end of a word) is considered a tense vowel (see § 5.2). Also, the vowel /ɔ/ is treated as tense (see § 5.3).

The American English vowel system consists of 11 distinct vowels (or monophthongs) /i, ɪ, e, ɛ, æ, ʌ, u, ʊ, o, ɔ, ɑ/ (Peterson and Barney, 1952). Categorization of vowels according to features of tongue articulation reveals a vocal tract vowel space which consists of four distinct corners corresponding to a quadrilateral shape. Vowels identified for each corner are /i/ (high-front), /æ / (low-front), /u/ (high-back), and /ɑ/ (low-back).

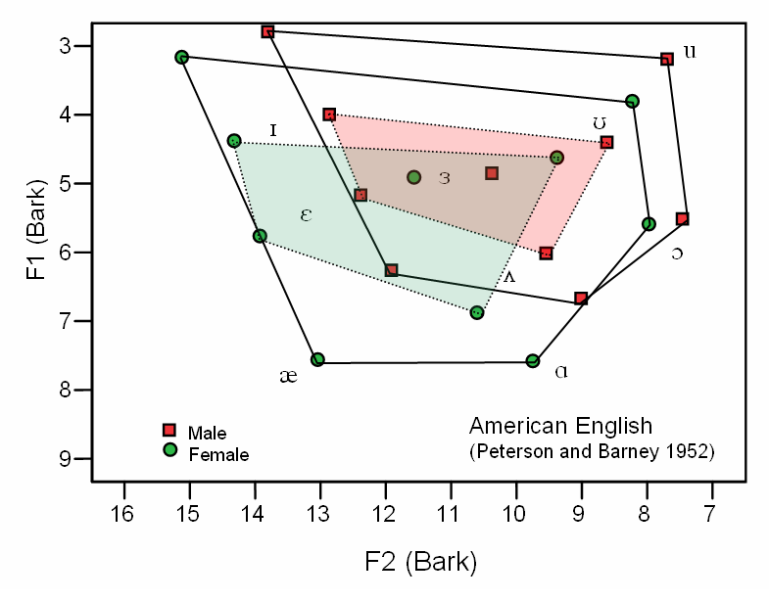


Figure 3.2. The tense (solid lines) and lax (dotted lines) vowels of General American plotted in an F1 (top to bottom) by F2 (right to left) display. Male (squares) and female (circles) vowels have been plotted separately. (After Peterson and Barney, 1952, with Bark-transformed frequency values for F1 and F2).

3.1.1.2 Dutch vowels

The Dutch vowel system (Table 3.2) is in many respects similar to English. It also has tense and lax vowels, and distinguishes four degrees of height and three degrees of backness. However, the central part of the vowel space is more densely filled as Dutch has (rounded) central high and high-mid vowels. In the lax front vowels Dutch distinguishes two degrees of height for the /ɪ ~ ɛ/ contrast, where English has three: /ɪ ~ e ~ æ/. Dutch is underdifferentiated relative to English in the high back vowels, where English has the tense ~ lax opposition /uː ~ ʊ/ while Dutch only has

/uː/. Dutch also has a number of vowels that are absent in the English system; such overdifferentiation is hardly ever a source of confusion (cf. Lado, 1957).

Dutch has three full diphthongs, /ɛi, œy, ɑu/, the first two of which have their starting point at a low-mid vowel height and the latter at a fully open position. Also, Dutch has some degree of diphthongization on the tense high-mid vowels, so that /eː/, /øː/ and /oː/ are realized as [e^{i}], [ø^{y}] and [o^{u}], respectively. There is a sixteenth vowel, schwa, which is not included in the table. This neutral vowel cannot be stressed; if it is, it will change to /ɵ/, the rounded lax central vowel.

Table 3.2. The basic Dutch vowel inventory (Rietveld and Van Heuven 2001).

	front		central		back	
	tense	lax	tense	lax	tense	lax
high	iː		yː		uː	
hi-mid	eː	ɪ	øː	ɵ	oː	
lo-mid	ɛi	ɛ	œy			ɔ
low			aː		ɑu	ɑ

Figure 3.3 gives the arrangement of the twelve monophthongs of Dutch (excluding schwa) in an acoustic vowel diagram. Dutch also has tense and lax vowels, but the lax subsystem seems reduced along the height dimension only, not also along the backness parameter, as it is in English.

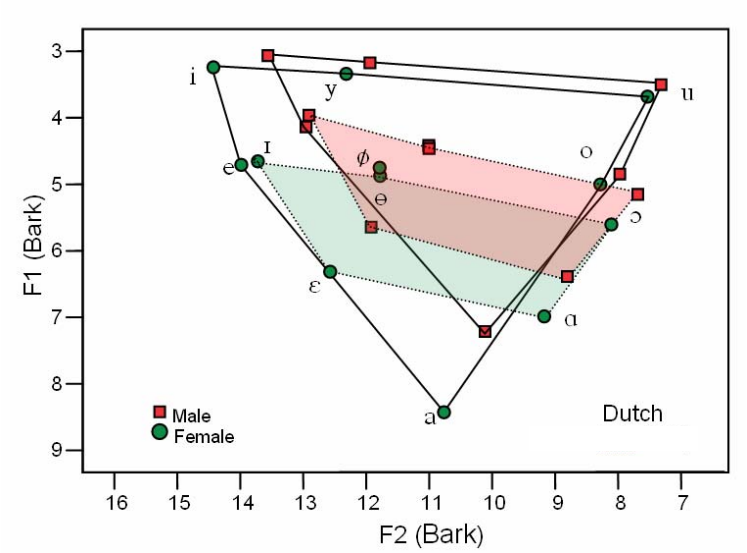


Figure 3.3. Dutch monophthongs plotted in an F1 by F2 plane (Barks). Male data (50 speakers) adapted from Pols et al. (1970), female data (25 speakers) from Van Nierop, Pols and Plomp (1973).

3.1.1.3 Chinese vowels

There has been a longstanding controversy in the literature on the number of underlying vowel categories in Mandarin, and the relationship of the myriad of surface vowel forms to these phonemic categories (e.g., Chao, 1934, 1968), R. Cheng, 1966; C. Cheng, 1973; Pulleyblank, 1984; Lin, 1989; Wang, 1993; Wu, 1994). The reason for this controversy is that most phonetic manifestations of vowels in Mandarin occur in a fairly narrow range of contexts, which suggests that they probably can be reduced to a smaller set of basic vowel categories. There is disagreement both on the number of surface (phonetic) vowels in Mandarin as well as on the number of underlying, abstract (phonological) vowels. Surface vowels can be as many as twelve or thirteen; the number of underlying vowels varies between four and six (Wan and Jaeger, 2003). The large majority of sources distinguish twelve surface vowels (see also Flege et al., 1997; Li and Thompson, 1981; Light, 1976; Maddieson, 1984; Wu 1964), which can be reduced to a smaller number of underlying vowels in different ways, yielding different numbers. We assume that positive and negative transfer of vowels from L1 to L2 is located towards the surface level rather than at some deep level of representation. Cheng (1966) relates the twelve surface vowels to their underlying forms as follows:

/i/ → [i], [ɾ],[ɪ]
/y/ → [y]
/u/ → [u]
/ə/ → [e], [ə], [o], [ɤ]
/a/ → [a], [ɑ], [ɛ]

The twelve surface vowels can be represented in a structural way as exemplified in Table 3.3.

As the table shows, Mandarin has no length (i.e. no tense ~ lax) contrast; contrasts such as /iː ~ ɪ/, /uː ~ ʊ/, /ɔ ~ ɔː ~ oː/, and /e ~ æ/ do not occur.

Table 3.3. The Mandarin surface vowel inventory.

	front		central		back	
	–round	+round	–round	+round	–round	+round
high	i	y	ɪ			u
high-mid	e		ɚ		ɤ	o
low-mid	ɛ					ɔ
low			a			ɑ

Chinese is different from both Dutch and English because Chinese is a tone language, in which tones are lexically specified. In general, all full syllables carry a lexical tone, whereas weak syllables have the neutral tone (or are ‘toneless’). As far as we know, however, the tones of Chinese do not interfere in any way with the production or perception of English sounds by Chinese learners. This does not rule

out the possibility that tonal interference may be found in the learning of English (sentence) prosody, but this is outside the scope of the present research.

Studies examining acoustic characteristics of vowel production in Mandarin are limited. Wu (1964) examined vowels produced by Mandarin speakers (four male adults, four female adults, four children). His measurements included, among other properties, F1, F2, F3 frequencies of formants for six standard vowels /i, e, y, u, o, ɑ/, as well as of allophones of /i/ and /e/. A later study by Howie (1976) acoustically analysed these six vowels produced by two male speakers.

Figure 3.4 presents formant measurements of F1 and F2 plotted in the same way as we did for English and Dutch. These formant values were published by Li, Yu, Chen and Wang (2004) for five male and five female speakers of Mandarin producing seven monophthongs /i, y, u, e, ɚ, o, ɑ/. Five of these have a fairly unrestricted distribution; /e/ and /o/, however, may be considered allophones of /i/ and /u/, respectively, which surface in specific environments only.

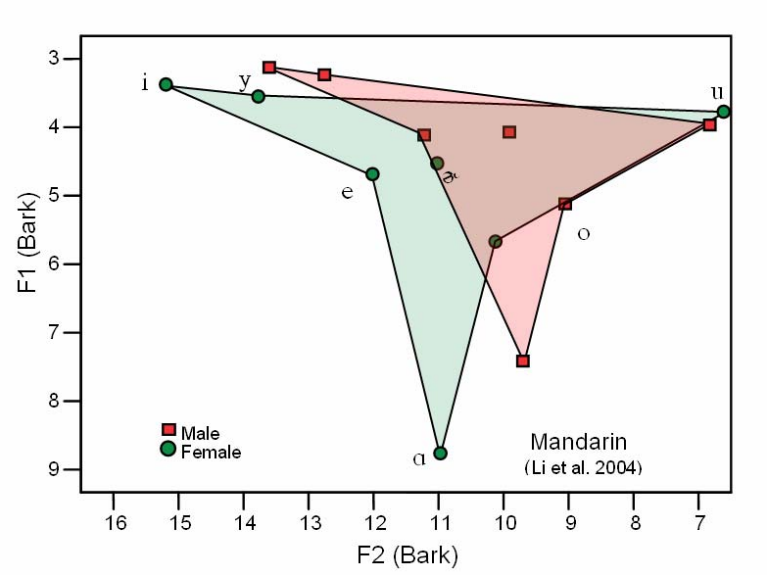


Figure 3.4. F1 versus F2 (Bark) for seven monophthongal vowels of Mandarin (Beijing dialect) spoken by five men and five women (adapted from Li et al. 2004). [6]

3.1.2 Prediction of pronunciation problems in vowels

In Table 3.4 below, I have attempted to present together the vowel inventories of Dutch, Mandarin and (American) English in a crude contrastive analysis. Here I use the principles that were advocated by Lado (1957), and which also underlie the

[6] F2 for male /u/ is specified by Li et al. (2004: 257) as 9.147862 Bark. I assume that the first digit is in error and corrected it to 7. This decision is supported by Figure 3 in Li et al.

categories of Flege's (1987) Speech Learning Model (SLM), in order to define three classes of speech sounds in a target language. The first is the category of identical sounds. These are sounds that are transcribed with the same narrow IPA symbol in source and target language; they should constitute no learning problem. In the table they have been left unmarked. The second category are sounds in source and target language that are written with the same IPA base symbol but differ in diacritic marks. These sounds are phonetically similar but not identical; such similar sounds are predicted to constitute long-term learning problems in second-language acquisition. In the table, similar sounds are indicated in grey cells. The third type are new sounds. Here a sound is needed in the target language which does not occur in the source language. The sound in the source language that is phonetically closest to the target is written with a different base symbol in the IPA. The prediction is that such new sounds constitute a learning problem in the initial stages of the acquisition process, but sooner or later the new category will emerge, and that it will be quite authentic. In the tables, new sounds are printed in white against a black background.

Table 3.4. Contrastive vowel analysis of Dutch and English (upper panel) and of Mandarin and English (lower panel). Grey cells in source languages denote source sounds that are not needed in the target language. White, grey and black cells in the target language represent identical, similar and new sounds, respectively.

V-height	Place of Constriction					
	Front		Central		Back	
Source: Dutch						
	Tense	Lax	Tense	Lax	Tense	Lax
High	i		y		u	
High-mid	e:	ɪ	ø:	ɵ	o:	
Low-mid		ɛ				ɔ
Low			a:			ɑ
Diphthong	ɛi		œy			ɑu
Target: English						
	Tense	Lax	Tense	Lax	Tense	Lax
High	i:				u:	
High-mid	e:	ɪ			o:	ʊ
Low-mid		ɛ				
Low	æ			ʌ		ɒ
Diphthong	ai		ɔi			ɑu

Table 3.4. Continued.

V-height (down)	Place of Constriction (across)					
	Front		Central		Back	
Source: Mandarin						
	−round	+round	−round	+round	−round	+round
High	i	y	ɿ		u	
High-mid	e		ɚ		ɤ	o
Low-mid	ɛ					ɔ
Low			a			
Diphthong	ai		ɔi			ɑu
Target: English						
	Tense	Lax	Tense	Lax	Tense	Lax
High	i:				u:	
High-mid	e:	ɪ			o:	ʊ
Low-mid		ɛ				
Low	æ			ʌ		ɒ
Diphthong	ai		ɔi			ɑu

It is rather unclear how realistic the predictions of SLM are. The vowels in the Dutch inventory have an unrestricted distribution, and can readily be employed in English. Some of the Mandarin vowels are highly context-sensitive allophones, which may or may not generalize to English. Moreover, Mandarin has no length (or tense~lax) contrast. The lax members of the opposition in English are transcribed with separate base symbols – and are therefore new sounds. The tense (long) members differ from the Mandarin counterparts in a diacritic only (length mark) and are therefore similar sounds.

In the next sections we will review comments made by experts on English pronunciation teaching to Dutch (§ 3.1.2.1) and Mandarin (§ 3.1.2.2) learners. These comments are not predictions based on an a priori comparison of source and target sound systems but summarize classroom experience.

3.1.2.1 Dutch ~ English

This section summarizes comments made in pronunciation text books at the university level for Dutch learners of English (e.g. Gussenhoven and Broeders; 1976, 1981; Collins, Hollander and Rodd, 1977; Collins and Mees, 1981). When in these comments Dutch and English are called similar, the term does not necessarily have the same status it has in Flege's SLM. The authors of the textbooks, who are accomplished phoneticians with a keen ear for minute phonetic differences between sounds, hardly ever call a pair of sounds in source and target language identical or the same. Therefore 'similar' sounds may refer to pairs of Dutch/English sounds that are written with the same base symbol and diacritics. The summary is presented in Table 3.5.

Table 3.5 Survey of pronunciation problems with vowels by Dutch learners of English, derived from Collins and Mees (1981) and Gussenhoven and Broeders (1976: 88). D = Dutch, E = English.

English target	Dutch substitutions/typical errors/comments
/iː/	Absent in D; similar sound D /i/ is the usual replacement. Too close, too front, especially too short. The articulation is also considerably tenser than E /iː/.
/ɪ/	Absent in D; D /ɪ/ is similar to this sound. Generally D learners have no problem with this sound.[7]
/eː/	Similar to D /eː/. Both E and D /eː/ are phonetically diphthongized, E /eː/ has a slightly lower onset and a stronger glide element. D /eː/ is within the range of acceptable pronunciations of E /eː/.
/ɛ/	D learners use similar D /ɛ/ as a replacement. D learners are generally unaware of the E /æ~e/ contrast so that perceptual confusion may result.
/æ/	Absent in D; most D learners will substitute D /ɛ/. Perceptual confusion is predicted between E /æ~e/.
/ɑː/	Absent in D; typically replaced by D /aː/, which varies considerably in quality. There is considerable overlap between D /aː/ and E /ɑː/.
/ɔː/	Absent in D; there appears to be no regular substitution from Dutch speakers. Some use an extended D /ɔ/, whilst others use the marginal vowel D /ɔː/ (as in French loan words). Others use the allophone of D /oː/ that occurs before D /r/.
/oː/	E /oː/ and D /oː/ are phonetically realised as diphthongs. E /oː/ has lower onset and somewhat stronger diphthongization but no perceptual confusion will arise with any other E vowel.
/ʊ/	Absent in D. E /ʊ/ is perhaps the most difficult vowel for D learners. There is no D vowel near E /ʊ/. Most D speakers confuse E /ʊ/ and E /uː/, hearing both in terms of D /u/.
/uː/	Absent in D, but similar to D /u/. Some speakers substitute D /u/. This vowel is closer to E /uː/, and the D sound is shorter (except before /–r/). The D articulation is also tenser. Most D speakers regularly confuse E /uː ~ ʊ/.
/ʌ/	Absent in D; D learners tend to substitute D /ɵ/ for E /ʌ/. More advanced students sometimes substitute D /ɑ/ for this sound.
/əː/	Absent in D; usually replaced by D /ɸːr/ or /ɵr/. Neither substitution is acceptable, having inappropriate lip-rounding, and too close a tongue position.

[7] Speakers from The Hague, Rotterdam, Amsterdam, Antwerp, may confuse E /iː~ ɪ/. Speakers from Dordrecht, Nijmegen, Noord-Brabant and Limburg may have a very open quality, which may give rise to confusion with English /e/.

Table 3.5. Continued

/ə/	The quality of D /ə/ is similar to that of E /ə/ in most contexts and transfers well into E. Difficulty may arise word-finally. D final allophone tends to be closer and is rounded, giving a markedly different effect from the very open word-final E. E /ə/ is similar to /ʌ/ in syllable final-position, D /ə/ is closer to D /ɵ/.
/ɒ/	Absent in D. The usual D substitution for E /ɒ/ is D /ɔ/, which is too close, too round and generally over-tense. Tenseness of D realisation of E /ɒ/ is especially noticeable before fortis plosives. A less common error is to pronounce /ɒ/ too front and unrounded, so losing contrast with /ʌ/. Mispronunciation of /ɒ/ is a very persistent error, often heard from otherwise proficient speakers. It appears to be difficult for D native speakers to detect.
/ai/	Absent in D; is often replaced by D VC sequence /a:j/, whose vowel part is too long, esp. before voiceless plosives, so that confusion may arise with voiced plosive (e.g. *tight ~ tide*).
/ɑu/	Similar sound D /ɑu/ is substituted; its onset may be too rounded but no perceptual confusions arise.
/ɔi/	Absent in D; D vowel+glide sequence /o:j/ is often substituted, whose onset is too close but does not lead to perceptual confusion

3.1.2.2 Chinese ~ English

Although many textbooks have been produced describing the differences between the sound systems of Dutch and English (see above) and giving detailed analyses of pronunciation errors of Dutch learners of English, such studies are virtually non-existent for Chinese learners of English. In fact, I know of just one pedagogical study by Zhao (1995), which makes a comparison between the sounds of Mandarin and of English and contains a discussion of pronunciation errors of Chinese learners of English. Much of what will be discussed in the following paragraphs has been taken from Zhao (1995); it should be borne in mind that her comments, too, relate to the sounds of British English, specifically RP. This is not a great concern as long as we are dealing with the consonants, since these do not differ very much between British and American English. It is a major concern when dealing with the vowel system.

The Chinese sound system that Zhao (1995) uses as her reference is that of Mandarin (also called Putonghua or Common Speech), which is comparable in status to RP in England. Like RP in English, there is also a standard form of pronunciation in modern Chinese. This pronunciation, which is being popularized throughout the P.R. China, is based on the northern dialect family, with Beijing speech sounds as the norm. In China, TV and radio announcers use the Common Speech. Teachers and students in school are required to use it, too. It is the main language spoken in China and one of the world's major languages, ranking among

the official working languages at the United Nations and other international organizations.

According to Zhao (1995), experience shows that Chinese learners of English who speak the Common Speech have fewer difficulties in acquiring a good English pronunciation than those who speak with broad local accents, who often have many difficulties to overcome before they can pronounce English acceptably. This would be because there are more similarities between the pronunciation of Common Speech and that of English. These claims seem rather speculative and remain to be tested in future research; such testing is clearly beyond the scope of the present dissertation.

Figure 3.5 has been copied from Zhao (1995). It is a traditional cardinal vowel chart with the RP-English vowels drawn as solid black circles and the Chinese vowels as open circles.

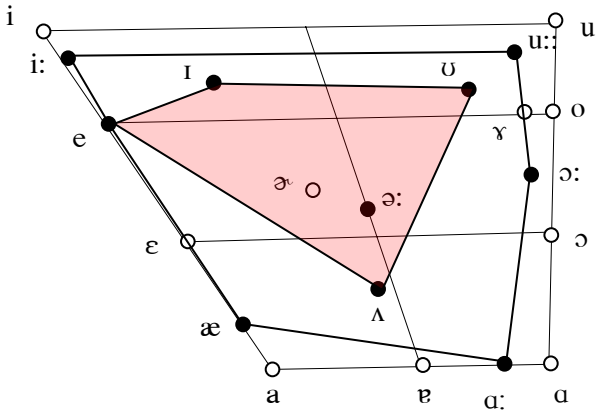


Figure 3.5. Comparison of RP English (solid markers) and Mandarin (open markers) vowels in a traditional Cardinal Vowel diagram (after Zhao, 1995).

For the purpose of the present dissertation Figure 3.5 has to be interpreted with some caution, as we will use General American as the pronunciation norm for English. As will be shown in more detail in Chapter five, the vowel system of English, whether British or American, can be conceived of as two subsystems, one of which is peripheral, with so-called tense (and long) vowels along the outer edge of the vowel diagram, and the other is more centralized, with four vowels configured along an inner circle. Zhao (1995) does not treat tense /e:/ and /o:/ as monophthongs. Rather, she deals with these vowels as half diphthongs, which is why they have not been included in Figure 3.5. The positions of the RP vowels seem quite reasonable; However, I would question the locations of the Chinese vowel sounds. Zhao seems to suggest that Chinese /i, a, ɑ, u/ are identical to cardinal vowels 1, 4, 5 and 8, respectively. It would seem rather unlikely that a language such as Mandarin,with a smaller vowel inventory than English, would have its vowels in more peripheral

positions.[8] The high central vowel /y/, which is one of the vowels of Chinese, has been omitted from the chart, most likely as Zhao believes that this vowel is never a reasonable substitute for any vowel of English. The half-close unrounded vowel /ɤ/ is given in the figure; Zhao claims that it is used as a substitute for English /ə:/.

I will now present a list of vowel pronunciation errors as identified by Zhao (1995). Later, in Chapter six, we will have occasion to check the predictions in this list with the confusion data collected in our own experiments. We will then be able to either confirm or disconfirm whether such errors do indeed occur. Moreover, we will examine our data to see whether there are any systematic errors that were not predicted by Zhao. If such errors should be found, the added value our experimental approach would be shown: we predict that even a trained teacher of English as a foreign language may well miss systematic pronunciation errors in foreign-accented English (especially when the teacher is a native speaker of the same language as that of the learners), that can only be brought to light through experimental methods.

The following table is a summary of Zhao's treatment of the English vowel sounds by Chinese learners. It lists all the vowel phonemes of (RP) English in the left-hand column. In the right-hand column I first specify if the particular sound has no counterpart in Chinese. When no remark is made as to the absence of the vowel in Chinese, Zhao implicitly claims that there is some vowel in Chinese that Chinese learners of English will use as a reasonable substitute for the target sound in English. Sometimes the substitute is a good match for the target sounds, in which case no further comments are made. Most of the time, however, the substitute differs from the target; the table will then specify how the substitute differs, and what perceptual confusions are likely to arise as a result of the substitution. The perceptual consequences are sometimes explicitly mentioned by Zhao, but when she makes no explicit claims, I have derived the predictions myself.

Zhao (1995) makes two claims with respect to the diphthongs of English. The first is that Chinese learners tend to reduce the contrast between long and short vowels in English, which would follow from the fact that Chinese does not use length as a distinctive feature. She then goes on to say that diphthongs are like long vowels, implying that Chinese-accented diphthongs will be too short. Chinese has both falling and rising diphthongs. A falling diphthong has its most prominent element first and the less prominent (semivowel, glide) element last, while a rising diphthong has the more prominent element last. English has falling diphthongs only.[9] Zhao adds a warning that there are rising diphthongs in Chinese and the beginning of these rising diphthongs is less prominent than the end. She seems to

---

[8]At first glance one would be tempted to believe that the open circles in figure 3.5 are in fact the cardinal vowel positions, given as reference points. However, these are explicitly the articulatory positions indicated by Zhao (1995) for the vowels of Mandarin. It is unclear from her description how these positions were determined, nor did she supply any references.

[9] It would be possible, however, to analyze the realization of tense /u:/ as [ju:] after certain consonants as a rising diphthong. Examples would be: *puke* [pju:k], beauty [bju:ti], *mew* [mju:], *tune* [tju:n], *dune* [dju:n], *new* [nju:], *cue* [kju:], and many others. Since the glide [j] *only* occurs in combination with tense /u:/ (including its centring diphthong allophone [ʊə], there is no point in increasing the set of onset clusters with a large number of /Cj/ sequences.

imply, therefore, that Chinese learners tend to substitute rising diphthongs for English target diphthongs. However, given the large inventory of falling diphthongs in Chinese we do not think it very likely that Chinese learners of English will ever use a rising diphthong as an approximation to an English target – except perhaps for /Cju:/ (also see note 2), but then the substitution would be highly felicitous.

Table 3.6. Survey of pronunciation problems with vowels by Chinese learners of English, derived from Zhao (1995). M = Mandarin, E = English.

English target	Mandarin substitutions/typical errors
/i:/	M /i/, too short, not tense enough, not high enough; confusion with E /ɪ/
/ɪ/	absent in M, M /i/ substituted, too long, too tense, too high, confusion with E /i:/
/e/	Generally no problem, but Northern speakers may substitute [ai] or [ei], yielding confusion with E /ai/ and with E tense /e:/
/æ/	Absent in M, pronunciation will be too close, confusion with E /e:/
/ɑ:/	M has three allophones: [a] (open syll.), [ɐ] (closed syll.) and [ɑ] (before nasal coda). Realization not open enough, confusion with E /ʌ/.
/ɔ/	Sound does not exist in M. Diphthong [au] substituted with glide and not enough lip-rounding. Confusion with E /au/
/ɔ:/	M [o)] substituted, too open but quite similar to modern (closer) British E pronunciation for /ɔ:/
/ʊ/	M [u] substituted. Too long, confusion with E /u:/
/u:/	M [u] substituted. Too short, confusion with E /ʊ/
/ʌ/	Sound does not exist in M. M [ɐ] substituted. Too open, confusion with E /ɑ:/
/ə:/	[ɤ] and [o] substituted. Too short, too close, too backward, confusion with /ɔ/ and/or /ɔ:/ The central vowel /ə˞/ is actually a retroflex [ə]; this sound would be quite similar to the American E realization of /ə:/
/ə/	[ɤ] and [o] substituted. Too long, too close, too backward, confusion with /ɔ/ and/or /ɔ:/

3.2 Consonants

Consonants are made by causing a complete or partial obstruction in the mouth or pharynx, and are usually described in terms of where the obstruction is made in the mouth (or: place of articulation), how the sound is made (or: manner of articulation), and whether or not the vocal cords vibrate (or: voicing). Consonants, therefore, all differ from each other in at least one of these ways.

In terms of the size of the inventories, Chinese has the largest variety with 26 different onset consonants, but only two of these may occur in the coda (while some

consonants can be considered variants of each other). English has a slightly less rich inventory of 24 consonants including three that are exclusively found in the coda. Dutch has the smallest inventory with 21 consonants, nine of which cannot occur in the coda.

3.2.1 Dutch consonants vs. English consonants

The classification of consonants involves at least three factors, the state of glottis, the place of articulation, and manner of articulation. The two charts 'Dutch vs. English' (Table 3.7) and 'Chinese vs. English' (Table 3.8) include all the consonant symbols in English, Dutch and Chinese. The horizontal axis shows the various places of articulation, the vertical axis the various manners of articulation, while the voiceless consonants are distinguished from voiced ones by placing the former on the left in any box and the latter on the right. Consonants that occur both in the Dutch and in the English inventory are in white cells. Dutch sounds in grey-shaded cells are absent in English, English (target) sounds in black cells are absent in Dutch. Grey cells in the English panel contain target sounds that occur also in Dutch but which have different phonetic realizations. These would be transcribed with the same broad phonemic symbol but differ from their Dutch counterparts in phonetic detail, i.e. in diacritic marks. These 'similar sounds', as they would be classified by Flege, are indicated in the bottom panel against a gray background. Here we simply count 24 consonants in the English inventory, six of which do not occur in the Dutch inventory and seven of which differ in phonetic detail from their Dutch counterparts. Specific predictions of learning problems will be discussed later.

Table 3.7. Consonant of Dutch (upper panel) versus English (lower panel) in a manner (down) by place (across) table. Further see text.

Manner	Place of Articulation								
	labial	labial-dental	dental	Alveolar	Alveolar-palatal	(retroflex)	palatal	velar	glottal
Source: Dutch									
Stop	p b			t d				k	
Nasal	m			n				ŋ	
Fricative		f v		s z	ʃ ʒ			χ ɣ	
Affricate									
Approx.		ʋ		r			j		ɦ
Lateral				l					
Target: English									
Stop	p^h b̥			t^h d̥				k_h g̊	
Nasal	m			n				ŋ	
Fricative		f v	θ ð	s z	ʃ ʒ				
Affricate					tʃ dʒ				
Approx.	w			ɹ			j		h
Lateral				l					

3.2.2 Chinese consonants vs. English consonants

Table 3.8 presents a contrastive listing of the consonants of Chinese and English arranged by manner (down) and place (across). When a table cell contains two sounds, the one on the left represents the fortis (aspirated) and the one on the right the lenis (unaspirated voiceless) member of a pair of obstruents. Grey cells in the Chinese panel denote sounds that do not occur in English, black cells in the bottom panel represent sounds that occur in English but are absent in Chinese.

Table 3.8. Consonant sounds of Chinese (upper panel) versus English (lower panel) in a manner (down) by place (across) table. Further see text.

Manner	Place of Articulation								
	labial	labial-dental	dental	Alveolar	Alveolar-palatal	(retroflex)	palatal	velar	glottal
Source: Chinese									
Stop	p^h b̥			t^h d̥				k^h g̊	
Nasal	m			n				ŋ	
Fricative		f		s		ʂ	ɕ		
Affricate				ts^h ts		$tʂ^h$ tʂ	$tɕ^h$ tɕ		
Approx.	w				ɻ		j	χ	
Lateral				l					
Target: English									
Stop	p^h b̥			t^h d̥				k^h g̊	
Nasal	m			n				ŋ	
Fricative		f v	θ ð	s z	ʃ ʒ				
Affricate					tʃ dʒ				
Approx.	w			ɹ			j		h
Lateral				l					

The table reveals that of the 24 English consonants ten do not occur in Chinese; however, the remaining 14 should be quite similar to their Chinese counterparts.

3.2.3 Prediction of pronunciation problems in consonants

As we did for vowels, we will now present tables containing the most likely errors in the production and perception of English consonants by Chinese and Dutch learners. The consonant data are largely based on Zhao (1995) for Chinese learners of English; the Dutch data are based on Collins and Mees (1981) and Gussenhoven and Broeders (1976). Again, since both textbooks deal with pronunciation difficulties of British English (RP) sounds we adapted some of the claims so as to be applicable for American English.

3.2.3.1 Dutch-English consonant transfer

Table 3.9 presents a summary of remarks and observations made by Collins and Mees (1981) on differences between the Dutch and English consonants. A pervading problem in the pronunciation of English consonants by Dutch learners is that Dutch does not allow voiced (lenis) obstruents in coda positions; in such positions the voiced ~ voiceless (or lenis ~ fortis) opposition is neutralized, and only the voiceless

(fortis) member of the pair will be realized. We will not discuss this problem in the table below; rather we consider this a consequence of a rule difference between Dutch and English depending on a sound's position in the syllable; the matter will therefore be discussed in § 3.3.

Table 3.9 Survey of pronunciation problems with consonants by Dutch learners of English, derived from Collins and Mees (1981) and Gussenhoven and Broeders (1976: 142−143). D = Dutch, E = English.

English target	Dutch substitutions/typical errors
/f/	D /f/ is identical to E /f/.
/v/	E /v/ has less friction than its D counterpart. Most D speakers substitute D /v/ for E / v/; these are similar sounds. [10]
/θ, ð/	Both are absent in D. These two sounds pose major problems of recognition and articulation for the learner. /ð/ is far harder for D learners than /θ/. Replacement of /ð/ by /d/ is one of the most common and persistent D errors. /θ/ is easier for D learners; the traditional instruction of tongue between teeth obtains the slit tongue shape characteristic of /θ/, which distinguishes it from /s/.
/s, z/	The articulation of /s, z/ is different from that of E. D /s, z/ are typically articulated with a portion of the tongue between front and blade whilst the tip is kept down behind the front teeth. With some speakers there may also be some lip-rounding. D /s/ has less firmly held stricture than E /s/; the jaw is more open with a laxer articulation. As a result, the friction of D /s/ is *graver* than the *sharp* friction which characterises the English sound. Some D speakers produce a D /s/ which is acceptable if transferred into E, whilst others produce a sound which is between E /s/ and E /ʃ/. Some of the D accents lack a contrast /s ~ z/. Other accents have no contrast /s ~ sj/ and /z ~ zj/.
/ʃ, ʒ/	The D sequence /sj/ in *chef* has more obvious palatal off-glide than its E counterpart. The articulation is often unrounded; the effect of this is to make D /sj/ sharper in friction.
/h/	E /h/ tends to have somewhat stronger glottal friction than D /h/, and voiceless *pharyngeal* friction can be heard from some speakers. E /h/ is only voiced between some voiced sounds, whereas D /h/ tend to have breathy voice in all contexts. Breathy (voiced) /h/ does not compromise its identity in E.

[10] The Dutch labio-dental semivowel /ʋ/ would be a better substitute but Dutch speakers do not do this.

Table 3.9. Continued.

/m, n, ŋ/	These are identical in D and E. One difficulty likely to arise is excessive nasalization of preceding vowels (plus deletion of the conditioning nasal). This is especially noticeable in open vowels. Nasal release of /t, d/ may provide problems for D students, particularly into syllabic consonants. D learners tend to insert /ə/ between stop and nasal, e.g. *rotten* /ˈrɔtən/.
/l/	The distribution of clear [l] and dark [ɬ] is similar in D and in E, though for many D speakers intervocalic /l/ is dark. Many D accents (Rotterdam and Amsterdam) have dark [ɬ] in all contexts including initial position. Articulation of clear [l] is similar in D and in E. Dutch /l/ is not devoiced following fortis plosives, compare E *plan* [pl̥æn] and D *plan* [plɑn]. D dark [ɬ] is significantly different from E dark [ɬ]. No perceptual confusions will arise from the differences.
/j, w/	D /j/ is similar to the E sound, but is often realized with friction, thus giving a voiced palatal fricative [j̝]. Because of the similarities of E /j/ and D /j/ there are few significant problems for the learner.
/w/	For E bilabial /w/ the typical substitution is D labio-dental /ʋ/. /w/ presents a major problem for Dutch learners both in terms of articulation and in confusion of E /w ~ v/ contrast.
/r/	D onset /r/ is either an alveolar or uvular trill (or fricative in clusters). E /r/ is a retroflex approximant. D coda /r/ may also be an approximant. Although substitution of trill and fricative may sound foreign, no perceptual confusion will arise.
/p, b/ /t, d/ /k/	D /p, t, k/ have very short VOT and are not aspirated. These realisations are substituted for E /p^h, t^h, k^h/, and may be confused with E /b̥, d̥, g̊/. D lenis stops /b, d, (g)/ have negative VOT (prevoicing); no perceptual confusion should arise when these sounds are substituted for their E counterparts.
/g/	Absent in D; D /g/ occurs mostly in loanwords or as an allophone of /k/. It is not available as a substitute for E /g/. D /k/ may be substituted, even in the onset. Perceptual confusion with /k/ is expected.
/tʃ, dʒ/	Absent in D; these affricates are either replaced by/confused with the fricatives /ʃ, ʒ/ or by some sequence of /t(s)j/, /d(z)j

3.2.3.2 Chinese – English consonant transfer

The following Table 3.10 summarizes the typical errors and substitution patterns observed for English consonants spoken by Chinese learners of English. Again our main source of information is Zhao (1995).

Table 3.10. Survey of pronunciation problems with simplex consonants by Chinese learners of English, derived from Zhao (1995). M = Mandarin, E = English.

English target	Mandarin substitutions/typical errors
/b, d, g/	E voiced (lax) plosives have 0 VOT (and therefore have no voice lead nor voice lag). This is as in M; no problems are predicted.
/p, t, k/	Voiceless (tense) plosives are aspirated both in M and in E but more strongly in E. Confusion with E /b, d, g/ may result as a result of insufficient aspiration
/f, v/	/v/ is absent in M, /f/ exists. /f/ is not a problematic target sound but /w/ and /f/ are substituted for /v/ (the latter especially in the coda)
/θ, ð/	Both are absent in M. /t/, /s/ and /f/ are substituted for /θ/, and /d/, /dz/ and /v/ for /ð/
/s/	/s/ in M is articulated with the tongue blade against the back of the upper teeth, in E with blade against alveolar. Substitution is either unnoticed or confusion with /θ/ arises
/z/	/z/ does not exist in M; the unaspirated voiced affricate /dz/ is substituted, which may be confused with E /dʒ/ or even with /tʃ/
/ʃ, ʒ/	These fricatives do not exist in M. No substitutes are given; no confusions are predicted.
/tʃ, dʒ/	/tʃ/ is approximated by M [ts^{h}] and /dʒ/ by [ts]. No specific confusions are predicted.
/w/	M and E /w/ are quite similar. M /w/ is in free variation with /v/. As a result /w/ is often incorrectly replaced by /v/ (and vice versa). /w/ ~ /v/ confusion is predicted.
/j/	M /j/ is similar to E. No problems predicted
/h/	M no /h/; the uvular fricative [χ] is substituted. This will not lead to confusions but the substitution will be unacceptable.
/l/	The clear /l/ is exactly the same as the M lateral /l/ The dark /ɬ/ is a more difficult sound for M learners, because in M, the lateral consonant never occurs in the coda
/m, n, ŋ/	English /m/, /n/, /ŋ/ are quite similar to M /m/, /n/, /ŋ/. However, M /m/ and /n/ never appear in the coda; M learners tend to pronounce the last phoneme unclearly, or even omit it unintentionally. [11] Word-medial /ŋ/ is claimed to be difficult for M learners.
/r/	E onset /r/ is replaced by the M fricative /z̧/ which is quite similar to the target but has slight friction; confusion with E /z/ is predicted. [12] No problems are predicted with coda-/r/; here the Chinese retroflex vowel is an adequate substitute.

[11] Some Chinese learners, especially people from Hunan, Sichuan, Fujian and Anhui provinces, may replace /n/ with /l/ or /l/ with /n/, as these sounds are free variants in the local dialects.

[12] One common error among Southern Chinese learners of English is the confusion of /r/ with /l/ and also with /n/. They produce *right* /rait/ as *light* /lait/ or *night* /nait/. Since this is not a problem for Northern Chinese (Mandarin) speakers, we have not included this confusion in the table.

3.3 Syllable structure

Human speech is basically spoken as a sequence of opening gestures of the mouth. Of course, once the mouth has been opened, it has to be closed before it can be opened a second time. One cycle of opening and closing the mouth produces a phonetic syllable. The alternation of opening and closing gestures takes place at a rate of some five cycles per second. When the mouth is maximally open, vowel sounds are produced; when the mouth is completely or partially closed, consonants are produced. The segments (consonants and vowels) within a syllable are subject to the sonority principle: louder and more sonorous sounds are produced in the middle of the syllable when the mouth is maximally open, and sounds of decreasing sonority are produced as they are closer to the edges of the syllable.

Languages differ widely in the complexity of syllable structures they allow. The simplest type of syllable structure is a regular alternation of a single consonant (C) and a single vowel (V). Many languages only allow regular CVCV alternation and in all languages CV is the most frequent syllable type. Mandarin comes rather close to such a CV language. English and Dutch have a richer variety of syllable types, and they allow up to three consonants in sequence in the beginning of a syllable and up to four in the final part of the syllable. Many consonants have rather different pronunciations depending on whether they precede the vowel or follow it within the syllable. Research has indicated that positive transfer of consonants is limited to source and target segments that have the same position in the syllable (Lado, 1957; Flege, 1995). Also, speakers of a language that has a simple CV structure find it difficult to produce sequences of consonants that are not interspersed with vowels. It is therefore important to review some of the differences in syllable structure among the three languages under consideration.

3.3.1 English

English is a language that allows complex syllable structures. Syllables are split up in an onset and a rhyme portion; the rhyme is further subdivided into the vocalic nucleus and the coda, which contains all postvocalic consonants. Onsets in English may vary in length from zero to three consonants. If the onset has its maximal length, i.e. three segments, the very first segment must always be /s/. Given this severe restriction the /s/ is considered to be outside the onset and given special appendix status. The vocalic nucleus either contains a long (or tense) vowel or a short (or lax) vowel. A word (or syllable) may not end in a lax vowel; lax vowels have to be followed by at least one coda consonant. Tense vowels may occur at the end of a word (or syllable). Given that diphthongs may occur at the end of a word, it follows that a diphthong functions as a tense vowel in English. The maximal number of consonants that can follow the vowel is three if the vowel is lax and two if the vowel is tense. In maximally long coda strings the last consonants are restricted to {t, d, s, z}, on the grounds of which this final constituent has been given appendix status. These can only occur as realisations of some suffix (past tense, past participle, plural, third person singular, as in *milked*, *ranged*, *milks*, *fields*). The velar nasal takes up the position of two coda consonants. Semivowels (glides) /j, w and h/ cannot occur

in the coda; they are restricted to the onset. Voiced (lenis) and voiceless (fortis) obstruents may occur in the onset and in the coda.

3.3.2 Dutch

The syllable structure of Dutch, which is closely related to English, has much in common with the English system. The syllable is hierarchically subdivided in much the same way. Dutch has zero to three consonants in the onset, with special appendix status for initial /s/. The vocalic nucleus contains either a short/lax vowel or a long/tense vowel, which again is functionally equivalent to a diphthong. Lax vowels may not occur at the end of a word or syllable; they have to be followed within the rhyme by at least one coda consonant. The maximum number of coda consonants is four, which can only occur after a lax vowel and then contains appendix consonants {s, t, st, ts} as in *herfst* /hɛrfst/ 'autumn'. The velar nasal counts as two consonants; /h/ cannot occur in the coda. However, other than in English, semivowels /w, j/ may occur in the coda but only after a long/tense vowel, as in *haai* /ha:j/ 'shark', *geeuw* /ɣe:w/ 'yawn'. Voiced as well as voiceless obstruents occur in the onset; in coda position voiced obstruents are impossible; these are neutralized to their voiceless counterparts.

Coda clusters are often broken up in Dutch by the insertion of an epenthetic vowel schwa. The insertion typically takes place when two adjacent consonants in the code do not differ enough in sonority, as in *melk* > [mɛlək], herfst > [hɛrəfst]. No vowel epenthesis takes place before obstruents which may occur in the appendix (i.e. /s/ and /t/) (see e.g. Van der Hulst, 1984).

3.3.3 Chinese

Traditional Chinese phonology divides the syllable into an Initial and Final. The Initial is the way a syllable begins, usually with a consonant. The Final is the syllable minus the Initial. For example, in *ta*, *chi*, *jin*, *chuang*, the Finals are *a*, *i*, *in*, and *uang*, respectively. The longest form of a Final consists of three parts: a medial (or: semivowel), a main vowel (or: head vowel), and an ending (or, in the case of retroflex suffixes, sometimes two endings, as in the *er*-sound *ming'er* 'tomorrow').

A Final in Mandarin comprises one of four medials: ∅ (empty), /i/, /u/, or /iu/ (= [y], one of three vowels: /a/, /e/, or /o/, and one of six endings: ∅, -i, -u, -n, -ŋ, and [ɻ] (phonetically -r).[13] Actually, there are only 40 different Finals (if Finals involving retroflex suffixes are not counted). As a result of these very severe restrictions on possible syllables in Mandarin, no obstruent clusters are possible in the onset (Initial) nor in the coda (Final). Onset clusters can maximally have a length of two segments, in which case the consonant closest to the vocalic nucleus must be a semivowel. Coda clusters are disallowed; in fact, syllables are generally open, i.e. end with a vowel. The only possible coda consonants are the nasals /n/ and /ŋ/. In compound vowels with /a, e, o/ as the first segment and /i, u/ as the second element, the latter are phonetically realised as semivowels, creating a diphthong. Phonetically,

[13] This gives rise to 4 × 2 × 6 = 48 possible Finals, since *a* and *o* count as allophones of one phoneme.

the retroflex approximant [ɻ] could also be considered a coda but this sound functions as a vowel.

3.3.4 Dutch versus English syllable structures

Generally, the syllable structures of Dutch and English are highly similar; in fact, Dutch syllable structure seems even less constrained than English, given that Dutch allow onsets such as /kn, pn, ɣn/ (in written obstruent+/n/ clusters the obstruent is not pronounced in English). As a result, Dutch speakers of English are expected to have few problems in realizing the complex syllable structures of English.

Complex clusters are no problems as such. However, due to some language-specific restrictions and peculiarities of Dutch some interference phenomena may arise. A very serious difficulty for Dutch speakers of English is to maintain the fortis ~ lenis (voiceless ~ voiced) contrast in coda obstruents. Lax/voiced coda obstruents are consistently realised as their fortis/voiceless counterparts, which may lead to perceptual confusion in English in minimal pairs such as *bad ~ bat*, *lies ~ lice*, *ridge ~ rich*, *leave ~ leaf*, *mouth* (verb) ~ *mouth* (noun), and many more.

Dutch speakers have a predictable tendency to break up English coda clusters, using their epenthetic vowel rule. Although the pronunciation of *milk* as [mɪɬək] sounds foreign, intelligibility will not be compromised by the epenthetic vowel.

3.3.5 Chinese versus English syllable structures

Since Mandarin allows no onset clusters except C+glide, Chinese speakers of English are predicted to have problems with the pronunciation of all other CC and CCC clusters of English. They are expected to break up awkward clusters by inserting an epenthetic vowel. Examples given by Zhao (1995: 95) indicate that /ə/ is inserted in between the members of CC clusters (*spy* > /səpai/, *pray* > [pəre^{i}]. No examples are given of pronunciation problems involving CCC onset clusters.

Even more problems are expected in the realisation of English coda clusters. Given that Mandarin only allows /n/ and /ŋ/ in the coda, any other consonant in that position will be awkward. Problems will increase when the coda contains two or more consonants. Chinese learners of English employ two strategies to cope with coda consonants. One is to add an epenthetic vowel [ə] after the coda consonant, which is then resyllabified to the onset of a separate syllable; this is what often happens in single C codas. When the coda is a cluster, it is often simplified by deleting one of the members of the cluster (after which epenthesis and re-syllabification may take place). Given the absence of obstruents in Mandarin codas and the absence of coda clusters, it is an open question how Chinese learners of English will deal with the fortis ~ lenis opposition in English codas. English is one of a minority of languages that maintains this contrast in coda position; in the majority of the world's languages the contrast is neutralised and only the voiceless member surfaces. One would predict that the realisation of marked phenomena in the target language (English) are a learning problem when these phenomena are absent in the source language (Mandarin). This prediction follows from the Markedness Differential Hypothesis (MDH, Eckman, 1977).

Zhao mentions one special strategy whereby Chinese learners arguably substitute Mandarin onset affricates [ts^{h}] and [ts] for English coda clusters /ts/ and /dz/, respectively. Since the place of articulation of the Mandarin affricates (tip of the tongue against the back of the upper teeth) is not the same as that of the English targets (tongue blade and the teeth ridge), this strategy will only be partially successful.

3.4 Concluding remarks

In this chapter we have reviewed the extensive literature on differences in sound structures between Chinese, Dutch and English. Some of the literature, especially that relating to the acoustical properties of vowels, was experimental in nature. The vast majority of the sources consulted, however, is based on observations made by teachers of English as a foreign/second language or by linguistic phoneticians using observation unaided by instrumental analysis. We will not be able to test each individual observation against experimental data to be collected in the next chapter(s). However, Chapter three will provide a database of observations we may turn to when discussing our experimental results. Very often we will point out correspondences between observations made in Chapter three and later experimental results, and on a few occasions we will also discuss experimental findings that have gone unnoticed in the (pedagogical) literature.

Chapter four

Data collection

In Chapter four I will outline the overall setup of the experimental work undertaken in the thesis, and provide a motivation for the choices we made. The chapter then describes the basic materials that were collected from groups of 20 speakers for each of three language backgrounds, i.e., Chinese, Dutch and American English, and how we selected two optimal speakers (one male, one female) from each set of 20 for the definitive tests.

4.1 Introduction

The purpose of the present thesis is to study the mutual intelligibility in English of speakers whose native language is Chinese, Dutch or American English. As was explained in Chapter one, the reason for choosing Dutch and Chinese as the source languages was that we wished to compare the role of transfer from a language that is closely related to the target language, English, and one that has no genealogical relationship with English at all. Dutch and Mandarin seem adequate representatives of these two categories. As for the variety of English we target in our research, we decided to work with General American (see Chapter three), rather than British English. American English is the model for English as a Second Language (ESL) in the educational system of the People's Republic of China. In the Dutch educational system the official norm is British English, but this norm is not strictly adhered to. In the teaching practice at Dutch secondary schools, hardly any attention is paid to matters of pronunciation. Moreover, the type of pronunciation more or less spontaneously adopted by Dutch learners of English resembles American rather than British English. Dutch-accented English, especially when spoken by university students and graduates, is rhotic, with a very strong approximant /r/ in the coda, which is also widespread in the present-day Dutch of the younger generations (see Van Bezooijen, 2005). In a recent study (Van der Haagen, 1998) it was shown that 40 % of the pronunciation variables in the English of Dutch secondary school pupils reflect the American-English pronunciation standard.[1] Given that Chinese ESL speakers adhere to the American pronunciation norm, and that Dutch learners

[1] It would appear that the language variety spoken in the media sets the norm here. English-spoken Dutch television programs and movies in theatres are not dubbed but subtitled. It has been estimated that four times as many programs are broadcast in American English than in British and/or Australian English (Van der Haagen, 1998).

vacillate between British and American norms, we decided that American English would be the target variety in our study.

A second problem was to decide on the type of learner to be studied. In earlier research (e.g. Bent and Bradlow, 2003; Van Wijngaarden 2001) the choice of speakers was more or less arbitrary or left unmotivated. We know, however, that there are sizeable differences in intelligibility among native speakers, so that the choice of the speakers to be included in our study is not arbitrary. In the type of study we have undertaken, there is no room for large numbers of speakers, so that the one or two speakers per language background that are included in the sample have to be truly representative of their peer group. In this chapter we will describe how we started with groups of 10 male and 10 female speakers from each of three language communities, and then selected one male and one female speaker from each group for inclusion in the final experiment such that these would be optimally representative of the larger group.

The third problem is what level of English proficiency should be adopted in the comparison of speaker and listener groups. Our research is concerned with English learnt as a foreign language, i.e. in a school setting where the language of instruction (and the daily language) is not English but either Dutch (in the Netherlands) or Mandarin (in the People's Republic of China). We decided to target groups of comparable ESL speakers in each of these two countries. The groups should comprise ESL speakers who need English professionally, and use the language for complex verbal messages, clearly beyond the needs of, say, tourists. However, we explicitly did not want to target specialists in English such as teachers of English as a second language, university students majoring in English language and literature, and the like. We therefore selected as our speaker and listener population the group of advanced students or graduates at the university level, specializing in any academic discipline other than English language and literature. Moreover, speakers and listeners should not have stayed in English-speaking countries for a long period of time, and have had no regular contact with English speaking friends or relatives. Although – presumably – the level of English proficiency will be better for Dutch than for Chinese nationals, the number of teaching hours will be comparable in the two countries. In both systems, English is first taught in the final forms of primary school, and is extended throughout secondary school with an intensity of two to three hours in the weekly curriculum. No further teaching of English is required once Dutch students enter university. In the PR China English skills are also part of the university curriculum of undergraduates. In spite of the possible effects of the diverging educational practice in the two countries, we decided to target non-specialist university students and graduates, since these are the typical professionals who attend international English-spoken meetings and conferences.

In the remainder of this chapter we will, first of all, describe the materials we have collected at the level of meaningless sounds (vowels, consonants, consonant clusters) and at the level of the word, in meaningful as well as in meaningless sentences (§ 4.2). These materials were then recorded from 20 speakers of English (10 male, 10 female) in the Netherlands (with Dutch as the L1), in China (with Mandarin as the L1) as well as from 20 native speakers of American English residing in The Netherlands (§ 4.3-4). The most difficult vowels and consonants were then selected (on the basis of pilot experiments conducted a year earlier) and

submitted for auditory identification by 20 listeners from the same language background as the speakers. On the basis of percent correctly identified vowels and consonants, one speaker was then selected from each group of ten (male or female; Dutch, Chinese or American background) for inclusion in the final experiment (§ 4.5). The data collection methods for the final experiment are described in § 4.6. No detailed results will be reported in this chapter; these will be presented in Chapters five (vowels), six (consonants), seven (clusters), and eight (words).

4.2 Materials to be collected

Our 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 and 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.[2] 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.

4.2.1 Vowels (/hVd/ list)

A list of words was compiled containing the 19 full vowels and diphthongs of English (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 and Barney, 1952). The listeners will get no structural information from the consonantal context when they have to identify the vowel. The consonants cannot help to reduce the set of recognition candidates in the lexicon, so that word recognition depends solely on vowel recognition and vice versa. The list of 19 vowels is shown in Table 4.1.

[2] An even more extreme viewpoint was adopted by Chen et al. (2001) in their study of American English vowels produced by Chinese learners. Since the article was submitted and published in the journal *Clinical Linguistics and Phonetics*, the authors by implication consider a foreign accent a disease that has to be treated by therapy.

Table 4.1. The 19 vowel sounds of English in /hVd/ context, plus phonemic transcription and sample words.

Vowel		Trans.	Ref. words	Vowel		Trans.	Ref. words
1.	h**ee**d	/hi:d/	f**ee**d, n**ee**d	11.	h**ar**d	/hɑ:d/	c**ar**d, b**ar**red
2.	h**i**d	/hɪd/	m**i**d, k**i**d	12.	h**u**d	/hʌd/	m**u**d, bl**oo**d
3.	h**ay**ed	/he:d/	pl**ay**ed, st**ay**ed	13.	h**ear**d	/hə:d/	b**ir**d, w**or**d
4.	h**ea**d	/hed/	r**e**d, b**e**d	14.	h**i**de	/hɑid/	sl**i**de, r**i**de
5.	h**a**d	/hæd/	b**a**d, s**a**d	15.	h**oy**ed	/hɔɪd/	t**oy**ed, empl**oy**ed
6.	wh**o'**d	/hu:d/	gl**ue**d, r**u**de	16.	h**ow**'d	/haud/	l**ou**d, all**ow**ed
7.	h**oo**d	/hud/	g**oo**d, w**oo**d	17.	h**ere**'d	/hɪəd/	b**ear**d, sn**eer**ed
8.	h**oe**d	/ho:d/	r**oa**d, sh**ow**ed	18.	h**oore**d	/huəd/	t**our**ed, m**oor**ed
9.	h**aw**ed	/hɔ:d/	s**aw**ed, fr**au**d	19.	h**air**ed	/hɛəd/	sh**are**d, c**are**d
10.	h**o**d	/hɔd/	g**o**d, n**o**d				

The original list of items was developed for Southern British English, which is non-rhotic. When pronounced by American speakers, the so-called centering diphthongs (ending in a schwa-like element) will often be monophthongs followed by a (frictionless continuant) /r/ sound. Also, the contrast between vowels 9, 10 and 11 (the latter as in *father*) may be neutralized in American English. We decided to run the full set of potential contrasts, but kept post-hoc pooling of vowels (as stimulus and as response categories) as an option.

4.2.2 Consonants (Consonant lists)

I targeted the full set of 24 intervocalic English single consonants, which were included in a list of nonsense words /aCa/. 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. No indications of stress position were included. We assumed that (native) speakers would generally pronounce these sequences with stress on the final syllable while reducing the first vowel to schwa; only for the two cases where the consonant is illegal in the onset (/ɑ:ʒɑ:/, /ɑ:ŋɑ:/), would it be reasonable for speakers to stress the first syllable and reduce the final vowel.

Table 4.2. The 24 syllable-initial simplex consonants of English used intervocalically in /ɑːCɑː/ environments, plus phonemic transcription and sample words.

Consonants		Trans.	Ref. words	Consonants		Trans.	Ref. words
1.	a**p**a	/ɑːpɑː/	**p**en, **p**ea	13.	a**h**a	/ɑːhɑː/	**h**e, **h**i
2.	a**b**a	/ɑːbɑː/	**b**ee, **b**y	14.	a**r**a	/ɑːrɑː/	**r**ed, **r**ose
3.	a**t**a	/ɑːtɑː/	**t**ea, **t**o	15.	a**f**a	/ɑːfɑː/	**f**at, **f**oot
4.	a**d**a	/ɑːdɑː/	**d**esk, **d**id	16.	a**v**a	/ɑːvɑː/	**v**ase, **v**est
5.	a**k**a	/ɑːkɑː/	**k**iss, **k**ey	17.	a**ch**a	/ɑːtʃɑː/	**ch**air, **ch**eese
6.	a**g**a	/ɑːgɑː/	**g**ate, **g**o	18.	a**j**a	/ɑːdʒɑː/	**j**am, **j**ar
7.	a**s**a	/ɑːsɑː/	**s**ea, **s**ee	19.	a**m**a	/ɑːmɑː/	**m**um, **m**y
8.	a**z**a	/ɑːzɑː/	**z**oo, **z**ero	20.	a**n**a	/ɑːnɑː/	**n**ice, **n**ight
9.	a**sh**a	/ɑːʃɑː/	**sh**y, **sh**e	21.	a**ng**a	/ɑːŋɑː/	ha**ng**er
10.	a**zh**a	/ɑːʒɑː/	plea**s**ure, A**s**ia	22.	a**l**a	/ɑːlɑː/	**l**ie, **l**ay
11.	a**th**a	/ɑːθɑː/	**th**in, **th**ink	23.	a**y**a	/ɑːjɑː/	**y**es, **y**et
12.	a**dh**a	/ɑːðɑː/	**th**at, **th**ose	24.	a**w**a	/ɑːwɑː/	**w**as, **w**ar

4.2.3 Consonant clusters (Cluster lists)

A compilation of 21 CC or CCC clusters in /aCC(C)a/ nonsense sequences was made. The list more or less exhausts the English inventory of initial consonant clusters. Given that onset clusters in English typically mark a stressed syllable, the second syllables in this list are always stressed. The initial vowel /ɑː/ was most easily read as schwa, which is what most native speakers intuitively did.

Table 4.3. A selection of 21 English CC(C) intervocalic onset clusters plus phonemic transcription and sample words.

Clusters		Trans.	Ref. words	Clusters		Trans.	Ref. words
1.	a**pl**a	/əˈplɑː/	**pl**ane, **pl**ay	11.	a**spr**a	/əˈsprɑː/	**spr**ing, **spr**ead
2.	a**bl**a	/əˈblɑː/	**bl**ue, **bl**ow	12.	a**spl**a	/əˈsplɑː/	**spl**it, **spl**endid
3.	a**pr**a	/əˈprɑː/	**pr**ay, **pr**ice	13.	a**scr**a	/əˈskrɑː/	**scr**eam, de**scr**ibe
4.	a**br**a	/əˈbrɑː/	**br**ead, **br**ing	14.	a**sp**a	/əˈspɑː/	**sp**eak, **sp**eed
5.	a**tr**a	/əˈtrɑː/	**tr**ee, **tr**y	15.	a**st**a	/əˈstɑː/	**st**ar, **st**ay
6.	a**dr**a	/əˈdrɑː/	**dr**y, **dr**iver	16.	a**sc**a	/əˈskɑː/	**sc**ale, **sch**ool
7.	a**cr**a	/əˈkrɑː/	**cr**y, **cr**eam	17.	a**sm**a	/əˈsmɑː/	**sm**all, **sm**art
8.	a**gr**a	/əˈgrɑː/	**gr**ey, **gr**een	18.	a**sn**a	/əˈsnɑː/	**sn**ake, **sn**eeze
9.	a**cl**a	/əˈklɑː/	**cl**ass, **cl**ean	19.	a**sl**a	/əˈslɑː/	**sl**ow, **sl**im
10.	a**gl**a	/əˈglɑː/	**gl**ass, **gl**ue	20.	a**sw**a	/əˈswɑː/	**sw**eat, **sw**im
				21.	a**thr**a	/əˈθrɑː/	**thr**ough, **thr**ow

4.2.4 Words in meaningless sentences (SUS-lists)

A set of 30 Semantically Unpredictable Sentences was complied with high-frequency words occurring in syntactically correct but semantically nonsense sentences (Benoît et al., 1996).[3] 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?* The five different syntactic frames are illustrated in Table 4.4. The full set of 30 SUS sentences used in the experiment is given in appendix A4.1.

[3] I thank Valérie Hazan of the Phonetics Department at University College London for her kind assistance in this matter.

Table 4.4. Examples of SUS sentences representing each of five different syntactic frames.

Structure			Examples
1.	Intransitive	Subj. – V – Adv.:	*The state sang by the long week.*
2.	Transitive	Subj. – V – Dir. Obj.:	*The real field made the vote.*
3.	Imperative	V – Dir. Obj:	*Use the game or the hair.*
4.	Interrogative	Q. word – V – Subj – Dir Obj:	*When does the charge like the late plane?*
5.	Relative	Subj. – V – Complex Dir Obj:	*The farm meant the hill that burned.*

SUS-sentences have the appearance of normal sentences. They can be pronounced fluently with appropriate accentuation, rhythmic structure and intonation. The words are only syntactically but not semantically constrained by their context, so that the listener must search the full set of words in a particular lexical category for each slot in the structure. This, of course, eliminates a lot of redundancy from the sentences and poses a severe challenge for listeners.

4.2.5 Words in meaningful sentences (SPIN-lists)

Fifty short sentences, with either a contextually predictable or unpredictable target word in final position, were selected from the original SPIN materials (Kalikow et al. 1977). The SPIN test (SPeech In Noise) was originally developed as a diagnostic instrument in audiology. The materials are normally presented for recognition with variable signal to noise ratios in order to determine a speech recognition threshold (50% word recognition scores). 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***.

Given that the target words are in the same category as their counterparts in the SUS materials, we predict that word recognition should be easier in the SPIN sentences than in the SUS materials, ceteris paribus. Of course, within the category of SPIN materials the targets in the unpredictable contexts should be more difficult to recognize than in the predictable contexts. The SPIN test is less efficient than the SUS test, as the former yields just one score for each sentence, whilst the latter contains up to five target words in one sentence. The complete set of SPIN sentences used in our materials is provided in appendix A4.2.

4.3 Speakers

Three groups of 20 speakers were recorded. Within each group ten speakers were male and another ten were female. Before making the recordings potential speakers filled in a questionnaire which asked them about their language background, contacts with native speakers of English, etc. The questionnaire is included in appendix A4.3. Although the answers were not analyzed systematically, they were used to ascertain that all speakers met the requirements (cf. § 4.1). In a few cases potential speakers were not recorded, for instance when it became clear that the speaker did not speak the standard variety of his/her language.

One group of 20 were native speakers of Dutch, students at Leiden University of any discipline except English Language and Literature. All spoke Standard Dutch of the Western (City Belt) variety. Table 4.2 presents the demographic data on the Dutch speakers.

The second group of 20 comprised native speakers of Chinese. All were second-year students at Jilin University, preparing towards a BA degree in various disciplines (mainly Psychology), with the exception of English Language and Literature.[4] All were speakers of North-East Mandarin. This is a variety of Mandarin which is very close to official Standard Chinese.[5] Demographic data on the speakers were collected through a Chinese version of the questionnaire, and are presented in the second part of Table 4.2.

The third group of speakers were American nationals who temporarily lived in the Netherlands in and around Leiden. They were either students at Leiden University or professionals working in Dutch branches of American (multinational) companies in the Leiden area. Since these were native control speakers, no requirements were made with respect to English training and overall educational level. Moreover, the American speakers hailed from various parts of the United States. Demographic data are provided in part three of Table 4.2. The speakers did not speak Dutch regularly. Their length of residence in the Netherlands was never more than three years, and none of the speakers planned to settle permanently in the Netherlands. They generally lived in American communities, and spoke their own language on a daily basis. It is safe to assume, therefore, that their pronunciation (and perception) of English was unaffected by their stay abroad.

---

[4] In a pilot study (Wang and Van Heuven, 2003, 2004, 2005) we recorded two Chinese speakers who lived in Leiden, The Netherlands, at the time of the recording. Although there are many native speakers of Chinese in Leiden, we decided to record our speakers for the final experiment in China, for two reasons. First, it would have been very difficult to find a sufficiently large group of Chinese speakers of English in the Netherlands with a homogenous language background. Second, Chinese graduates who are selected to be sent abroad for specialization have an above-average command of English that is not representative of the academic population in China at large. The results of the pilot experiment were used to single out the ten most confusable vowels and consonants in English spoken and identified by Chinese nationals. This selection was used in the present thesis to determine the most representative male and female speakers within the larger groups of ten.

[5] Standard Chinese is spoken by less than one percent of the Chinese population.

Speakers took part in the experiment on a voluntary basis. They were approached through advertisements on notice boards and on the intranet, or through the mediation of a colleague/lecturer who was asked to make an announcement in class. Participants were paid a fee of € 7 for their services.

4.4 Recording procedures

The 20 Dutch and 20 American speakers of English read the materials (in the order list 1 through 5) from paper in individual sessions while seated in a quiet lecture room in the Leiden University Phonetics Laboratory. Their vocal output was digitally recorded through a Shure SM10A close-talking microphone on the hard disk of a computer (44.1 KHz, 16 bits). Both speaker and experimenter were present in the room. During the recordings all other computers in the room had been switched off. Some background noise was generated by the computer on which the signals were recorded, which was effectively reduced by our use of a close-talking microphone.[6] The Chinese speakers of English were recorded in Jilin University in Changchun, PR China. These recordings took place in a small quiet room with only the speaker and the experimenter present, using the same microphone as in Leiden. Signals were digitally recorded directly onto the hard disk of a notebook computer.

4.5 Selecting representative speakers

The total set of materials recorded comprised a very large collection of speech materials. It would have been impossible for listeners to be confronted with the full set of materials spoken by each of our 60 speakers. It was necessary, therefore, to severely reduce the size of the materials for the final experiment. It had been our intention all along to include in the final experiment one male and one female speaker of English from each of the three language backgrounds, Chinese, Dutch and American. We therefore needed a procedure to select the optimal representative from each of the six groups of speakers, so that we would effectively reduce the size of the materials for the final test to one-tenth.

4.5.1 Set-up of the speaker-selection test

As the most representative male and female from each group of 20 we considered that we should locate neither the best nor the poorest but the most typical, i.e. average, speakers within the peer groups. The most typical speaker can be located

[6] This solution was preferred over the use of professional, high-quality recording equipment on the grounds of the argument that we needed recordings of uniform quality regardless whether these were made in The Netherlands or in China. Since we knew beforehand that no recording studio and professional equipment would be available at Jilin University, we decided to downgrade the Leiden recording environment so as to be comparable to the Chinese facilities.

only through comparing his/her intelligibility with that of the other members in the group, so that again a very large, in fact unmanageably large, experiment would have to be run. We therefore decided to base our search for the most typical speakers only on the first two datasets we recorded, i.e. the vowel test and the simplex consonant test, since these are arguably the severest tests on the quality of the speaker's pronunciation. These two tests present the stimuli without any lexical redundancy, i.e., knowledge of the lexicon or of sentence-level constraints does not help the listener here at all. The same would apply to the consonant cluster set, but preliminary experiments had already indicated that clusters were more easily identified by all groups of listeners than simplex consonants (Wang and Van Heuven, 2003). In order to reduce the materials further, and at the same time make the screening test more efficient, we decided not to include all the 19 vowels and 24 simplex consonants, but restrict the presentation to the ten most difficult vowels and ten most difficult consonants within each speaker group.

The preliminary experiment (Wang and Van Heuven, 2003, see also footnote 4) produced complete confusion matrices for the vowels and simplex consonants for each of the nine combinations of speaker and hearer nationalities. We decided to select only the confusion matrices obtained for speaker-hearer groups that shared the same native language. As a result, the optimally representative Chinese speaker of English will be selected oh on the basis of his/her intelligibility in English for fellow Chinese listeners. The same principle, *mutatis mutandis* of course, was applied to the selection of the Dutch and American speakers. The original confusion structures in the pilot experiments can be consulted in the literature, be it for the vowels only (see Wang and Van Heuven, 2004). It is clear from these confusion matrices that the order of difficulty, as evidenced by the error percentages in the identifications, is not the same for the three speaker/listener groups.

4.5.2 Stimuli

Tables 4.1 and 4.2 present the subsets of the ten most difficult vowels and consonants, respectively, for each of the three nationalities. In principle, the ten vowels or consonants selected are among the top-10 error percentages, but on some occasions we had to replace one or two sounds with high error percentages by alternatives with much lower error percentages; this was necessary in order to include attractive distractors in the list of ten. For instance, /f/ turned out to be an easy consonant for Chinese speakers/listeners but was included in the set of ten in order to provide an attractive response alternative for /v/ − which was a very difficult sound indeed. Moreover, in the selection of vowel sounds (full) diphthongs and /r/-colored vowels were excluded, so that only monophthongs could be selected.

Table 4.1. Percent error in vowel identification in pilot experiment for Chinese, Dutch and American speakers of English. Listeners shared the language background of the speaker. Vowels marked with an asterisk were selected for the screening test.

	Vowel	**Chinese**		**Dutch**		**American**	
1.	iː	12	*	0		13	
2.	ɪ	44	*	0	*	38	*
3.	eː	12	*	11	*	19	*
4.	ɛ	65	*	0	*	12	*
5.	ɑː	21		6		13	
6.	æ	82	*	50	*	13	*
7.	uː	76	*	50	*	19	*
8.	ʊ	56	*	6	*	56	*
9.	ɔː	71	*	83	*	75	*
10.	ɔ	21		0	*	50	*
11.	oː	50	*	33	*	19	*
12.	ʌ	76	*	28	*	63	*
13.	əː	24		11		13	
14.	ai	12		6		13	
15.	ɔɪ	29		6		13	
16.	au	35		0		12	
17.	ɪə	26		22		31	
18.	ʊə	24		17		6	
19.	ɛə	6		50		13	
	Total	**41**		**20**		**26**	

Table 4.2. Percent error in consonant identification in pilot experiment for Chinese, Dutch and American speakers of English. Listeners shared the language background of the speaker. Consonants marked with an asterisk were selected for the screening test.

	Consonants	**Chinese**		**Dutch**		**American**	
01	p	3		0		0	
02	b	3		0		6	
03	t	6		0	*	0	*
04	d	15		17	*	6	
05	k	6		0		6	
06	g	18		0		0	
07	s	41	*	17	*	56	*
08	z	47	*	22	*	19	*
09	ʃ	6	*	0	*	31	*
10	ʒ	47	*	6	*	94	*[7]
11	θ	44	*	33	*	94	*
12	ð	76	*	39	*	75	*
13	h	12		0		6	
14	r	15		0		6	
15	f	0	*	6		0	*
16	v	74	*	17		12	
17	tʃ	21	*	0	*	0	
18	dʒ	21	*	17	*	13	
19	m	0		0		0	
20	n	6		44		6	
21	ŋ	15		11		0	
22	l	15		6		0	*
23	j	21		6		0	
24	w	35		0		25	*
	Total	**23**		**10**		**19**	

Summary statistics on the subsets of ten vowels and ten consonants are provided in Tables 4.3 and 4.4, respectively.

[7] In the pilot experiment the consonants /θ/ produced by the American female speaker and /ʒ/ produced by the American male speaker were both strongly confused by the listeners with /t/ and /f/. This depressed the consonant identification scores for this group. We may have recorded very poor native speakers for these two consonants, but we interpret the confusion structures in the pilot experiment such that these two consonants may be the most confusing consonants for the vast majority of American listeners. In order to enable these confusions we chose /t/ and /f/ as the contrast consonants to compare with /θ/ and /ʒ/.

Table 4.3. Percent identification error obtained in preliminary experiment for the selection of ten most problematic vowels produced in /hVd/ frames. Mean, standard deviation, minimum, maximum and range of error percentage are indicated. The mean error percentage for the full set of 19 vowels is given in parentheses.

Speakers/listener group	Mean		SD	Min.	Max.	Range
Chinese	58.8	(41.3)	20.8	11.8	82.4	70.8
Dutch	26.1	(19.9)	28.2	0	83.3	83.3
American	36.3	(25.7)	23.2	12.5	75.0	62.5

Table 4.4. Percent identification error obtained in preliminary experiment for the selection of ten most problematic simplex consonants produced in /ɑːCɑː/ frames. Mean, standard deviation, minimum, maximum and range of error percentage are indicated. The mean error percentage for the full set of 24 consonants is given in parentheses.

Speakers/listener group	Mean		SD	Min.	Max.	Range
Chinese	25.0	(22.7)	25.9	20.6	76.5	55.9
Dutch	13.7	(10.0)	13.9	5.6	44.4	38.8
American	28.9	(19.0)	37.7	6.3	93.4	87.1

As can be seen in Tables 4.3 and 4.4, the mean difficulty (percent error obtained in the preliminary study) was greater for the vowel test than for the consonant test. Also the level of difficulty was not uniform across the three speaker/hearer groups. These differences, of course, do not invalidate the screening test; they just show that what is difficult in one group may not be difficult for another group. What is important is that the overall level of difficulty in the selections was closer to 50% error than the means found in the pilot experiment; on account of this, the selections provide a more efficient and discriminating testing instrument than when the full set of 19 vowels and 24 consonants had been included.

Two separate tests were constructed from the selections for each of the three listener groups. For each listener group, the first test comprised the ten hVd tokens for the ten male and ten female speakers sharing the same language background as the prospective listeners, in quasi random order. Immediate succession of the same vowel type or tokens produced by the same speaker were systematically excluded. This resulted in a vowel identification test for each listener group comprising 20 (speakers) × 10 (vowel types) = 200 stimuli. These were preceded by ten practice items, randomly chosen from the set of 200.

Three consonant identification tests, one for each listener group, were compiled in analogous fashion, yielding 20 (speakers) × 10 (consonant types) = 200 stimuli, again preceded by ten practice items.

4.5.3 Listeners

For the screening test we enlisted the services of 20 Chinese listeners at Jilin University, Changchun, PR China, 20 Dutch listeners at Leiden University, the Netherlands and 20 American listeners, who also listened to the materials at Leiden University.

Listeners were drawn from the same population as the speakers. They were university students or professionals with a university education (or comparable), with normal hearing, with no special qualifications in English. They did not specialize in English Language and Literature, and had not had regular contact with native speakers of English.

Listeners were found through advertisements on public notice boards, through e-mail messages, etc., as described in § 4.3. They were paid a fee of € 5 for their services.

4.5.4 Procedure

The stimuli were played back over good quality headphones (Sennheiser HD 424) from a notebook computer to listeners individually or in small groups of up to six seated at tables in a small lecture room. Dutch and American listeners were tested in a lecture room of the Leiden University Phonetics Laboratory. Chinese listeners were tested in a comparable room at Jilin University, Changchun. Listeners were issued instructions and separate answer sheets for the two parts of the experiment. On the answer sheet for the vowel identification test, the ten possible response categories were listed from left to right, exemplified by sample words. The subjects were asked to tick the response category they thought was intended by each following item played to them. Subjects were told to tick one and only one response alternative; they were not allowed to leave an item blank, and were told to gamble in case of doubt. The response alternatives were different for the three versions (Chinese, Dutch, American listeners) as the sets of most confusable vowels and consonants differed per listener nationality (cf. § 4.5.2). Verbatim instructions and copies of the answer sheets (English listeners only) are included in appendix A4.4.

For each part of the screening experiment (vowels, consonants) the subjects heard ten practice items, included to allow them to get familiar with the temporal structure of the stimulus presentation and the visual layout of the answer sheets. The practice items were followed without a break by the 200 vowel or consonant items, with inter-stimulus intervals of 5 seconds (offset to onset) and with a short beep separating blocks of ten stimuli. A short break was observed between the vowel and the consonant identification test. The whole test for both parts took about 90 minutes.

4.5.5 Results

The results of the speaker-selection test are presented in Table 4.5. Percentages of correct vowel identification and consonant identification were determined for each of the 60 speakers, and listed per language in ascending order of correct vowel identification. Summary statistics, i.e. mean, SD, minimum and maximum score and range, are given per language at the bottom of the table.

The results reveal, quite clearly, that within each of the three language groups individual intelligibility of speakers may differ substantially. This confirms the need for carefully selecting speakers within their peer groups for inclusion in a cross-linguistic intelligibility study. Overall, the Dutch speakers were less intelligible in English for Dutch listeners, than American speakers were for American listeners. Intelligibility was poorest among Chinese speakers and listeners of English. The mean differences between the three groups of speakers are not relevant to our purpose, which is solely to locate the most typical speakers within the peer groups.

Interestingly, for the entire group of 60 speakers, the female speakers turned out to be more intelligible, at least in terms of their vowel and consonant identification scores, than the male speakers. It has been suggested that women have more intelligible voices than men (Tielen, 1992 and references given therein), but so far results have been inconsistent. Also, there are persistent claims that women should have a greater talent for learning foreign languages. Figure 5.1 plots mean percent correct vowel identification against consonant identification for male and female speaker groups separately (but accumulated over all ten speakers per group) for the three nationalities. The figure shows that there is a small (but significant) superiority of women along both vowel and consonant dimensions within the American group, which would support the claim that women have more intelligible voices than men. The advantage of the female voices, however, is clearly larger for the non-native speakers (Chinese and Dutch nationals), which would indicate that there is a second effect, possibly due to women's greater gift for language learning. The superiority of the females within the Chinese and Dutch groups would then be the compound result of the inherent advantage of the female voice and the greater gift for foreign language learning. Be this as it may, the clear difference in performance of the male and female subgroups should play a role in the selection of the optimal speakers.

Table 4.5. Vowel identification (% correct) and consonant identification (% correct) for individual speakers (S#) broken down by native language background and gender. Within each category results are listed in ascending order of vowel identification. Summary statistics are provided at the bottom of the table.

Speakers	Language background of speaker-hearer group								
	Chinese			Dutch			American		
Male	S#	V	C	S#	V	C	S#	V	C
1.	4	45.0	77.0	18	61.5	57.5	8	61.3	85.0
2.	7	49.0	61.5	10	63.0	70.5	14	61.9	87.5
3.	8	49.5	65.0	7	65.0	70.5	4	68.1	81.9
4.	18	53.0	64.5	1	67.5	52.5	11	80.0	92.5
5.	1	54.5	67.5	4	68.0	63.0	6	81.3	86.3
6.	19	55.5	60.5	11	71.0	67.0	17	84.4	87.5
7.	2	56.5	59.5	6	71.5	65.0	19	85.6	89.4
8.	3	57.0	57.0	2	72.0	66.0	2	90.6	85.6
9.	20	58.0	69.0	12	74.0	53.5	1	90.6	86.9
10.	21	63.0	71.5	13	75.0	73.0	20	91.3	86.4
	Mean	54.1	65.3		68.9	63.9		79.5	86.9
	SD	51.8	6.1		4.6	7.2		11.7	2.8
	Min	45.0	57.0		61.5	52.5		61.3	81.9
	Max	63.0	77.0		75.0	73.0		91.3	92.5
	Range	18.0	20.0		13.5	20.5		30.0	10.6
Female									
1.	14	47.5	61.0	3	52.5	72.0	15	68.8	73.8
2.	9	49.0	71.5	20	58.5	72.0	3	71.3	86.9
3.	6	50.5	65.5	5	63.5	67.0	10	75.6	88.1
4.	12	53.0	62.0	15	65.5	67.0	12	76.3	85.0
5.	15	60.5	75.0	14	66.5	64.0	13	77.5	83.8
6.	16	61.5	60.0	19	67.0	59.5	9	81.9	83.1
7.	10	61.5	70.0	16	67.0	67.0	16	85.0	65.6
8.	13	63.5	66.0	17	68.5	75.5	7	85.6	81.9
9.	17	64.0	62.5	9	69.5	64.5	5	86.8	85.0
10.	11	64.0	77.0	8	71.0	65.5	18	86.9	85.0
	Mean	57.5	67.1		65.0	67.4		79.6	81.8
	SD	6.7	6.0		5.6	4.7		6.6	6.9
	Min	47.5	60.0		52.5	59.5		68.8	65.6
	Max	64.0	77.0		71.0	75.5		86.9	88.1
	Range	16.5	17.0		18.5	16.0		18.1	22.5
All	Mean	55.8	66.2		66.9	65.6		79.5	84.3
	SD	6.1	6.0		5.4	6.2		9.2	5.7
	Min	45.0	57.0		52.5	52.5		61.3	65.6
	Max	64.0	77.0		75.0	75.5		91.3	92.5
	Range	19.0	20.0		22.5	23.0		30.0	26.9

4.5.6 Selection of optimally representative speakers

Closer inspection of the data in Table 4.5 shows that the correlation between percent correct vowel and consonant identification is relatively poor. That is to say, speakers with high vowel intelligibility need not have a correspondingly high consonant identification score. Table 4.6 presents the correlation coefficients between vowel and consonant identification for each of the three groups of speakers separately and across all speakers.

Table 4.6. Pearson correlation coefficients for vowel and consonant identification for Chinese, Dutch and American speakers of English (language background of speaker and listeners is shared).

Language group	Chinese	Dutch	American	All
r =	0.092	–0.248	0.045	0.584
N =	20	20	20	60
p	= 0.701	= 0.291	= 0,852	< 0.001

At first sight there appears to be a fairly strong correlation between vowel and consonant identification scores. However, this correlation is merely caused by the fact that vowel and consonant identification are higher, on average for the American native speakers than for the Dutch learners, and these are better again than those of the Chinese L2 speakers. Crucially, within each of the three speaker/hearer groups no correlation remains. This is shown by Table 4.6, and is graphically illustrated below in Figures 4.2 to 4.4.

Given the low correlation between vowel and consonant identification scores, we decided to give equal weight to both parameters in the process of selecting the most representative male and female speakers within each language group. Figures 4.2 to 4.4 plot the 10 male and 10 female speakers in the Chinese, Dutch and American groups, respectively, as points in a two-dimensional space defined by the correct vowel (vertical) and consonant identification scores (horizontal). In each figure the mean vowel and consonant identification score is indicated by a horizontal and vertical line, respectively; the centroid of the scatter clouds is defined as the crossing point of the two lines representing the mean scores. The most typical male and female speakers are the individuals with the closest Euclidean distance from the centroid. These individuals have been marked with solid symbols in the figures, as opposed to the less typical speakers who have been marked with open symbols.

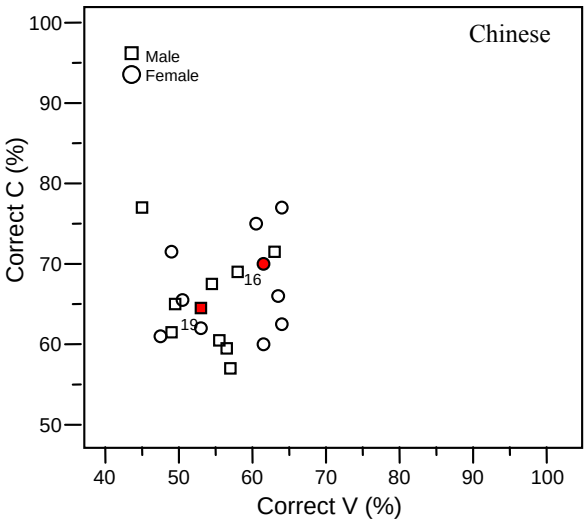


Figure 4.2. Ten male (squares) and ten female (circles) Chinese speakers of English plotted as a function of correct vowel identification (horizontal) and correct consonant identification (vertical) scores.

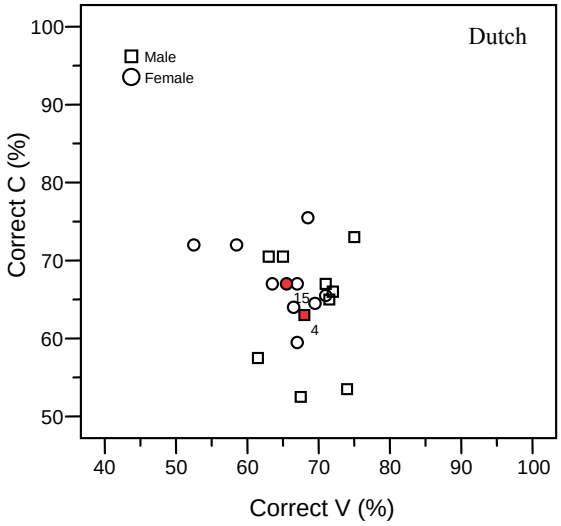


Figure 4.3. As Figure 4.2 but for Dutch speakers of English.

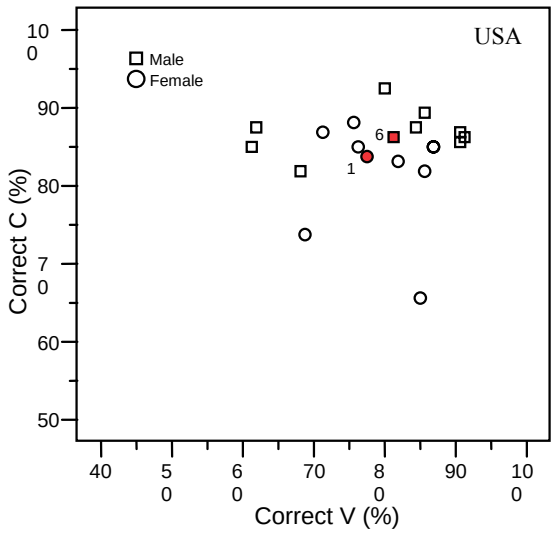


Figure 4.4. As Figure 4.2 but for American speakers of English.

4.6 Final experiment

After the optimally representative male and female speakers were selected for each of the three speaker groups, we set up the final listening experiments in order to determine mutual intelligibility among the nine possible combinations of speaker and hearer nationalities involved in this study.

The materials we used for the final tests were the same as those described in § 5.2-4. This time, however, all the materials were used, i.e. 19 /hVd/ items (vowel identification), 24 /aCa/ simplex consonants (consonant identification), 24 /aCC(C)a/ clusters (cluster identification), 30 SUS sentences and 48 SPIN sentences. Only the materials of the most representative male and female speaker were included for each of the three speaker nationalities, yielding a total of six speakers.

4.6.1 Preparation of stimulus materials for final tests

After the recording sessions the materials were downsampled (16 KHz, 16 bits) and stored on computer disk. Materials were then constructed for the final listening experiment comprising five parts. Part 1 contained the 19 /hVd/ words for all six speakers in random order (across speakers), preceded by ten 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) × 21 /aCC(C)a/ items in random order, preceded by four 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). 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. 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.

4.6.2 Listeners of final tests

Three groups of listeners were used in the final run of the experiments. One group comprised 36 Dutch listeners, 18 male and 18 female, drawn from the same population from which the Dutch speakers had been selected (cf. § 4.3). These listeners heard the stimulus materials in the Leiden University Phonetics Laboratory (see below). The Dutch listeners were paid a fee of € 10 for their participation in the experiment.

The second group of final listeners were students at Jilin University.[8] They belonged to the same population (but were different individuals) and were selected according to exactly the same criteria as the Chinese speakers of the stimulus materials. The subjects studied at my home university, and could be persuaded to take part in the experiment through advertisements on notice boards and by asking colleagues (fellow teachers) in the faculty to instruct their students to contact me. Half of the Chinese listeners were male, the other half female. They were paid the equivalent of € 5 in Chinese national currency.

The third group of listeners did the final experiments in Los Angeles, USA. These were 18 male and 18 female students at the University of California at Los Angeles (UCLA). Students with prior exposure to Dutch and/or Chinese accented English were not admitted as subjects. Listeners were found through advertisements in the student newspaper (UCLA Daily Bruin), through advertisements on public notice boards and on the internet, and through personal contacts with my host at UCLA.[9] American listeners received a compensation of $ 10 for their participation in the experiment.

---

[8] Obviously, we could not use Chinese listeners who resided in the Netherlands, as we needed Chinese listeners who had not been exposed earlier to Dutch-accented English. Taking this precaution we eliminated a basic flaw from the experimental design that may have compromised the results of our pilot studies – which did indeed use Chinese and American listeners residing in the Netherlands (Wang and Van Heuven, 2003, 2004).

[9] I gratefully acknowledge the material and moral support given to me by Dr. Robert S. Kirsner, Professor of Dutch and Afrikaans at UCLA, who made facilities available for running the experiments and who was instrumental in finding the required number of qualified listeners, and obtaining formal permission from the UCLA human subjects' ethics

4.6.3 Procedure of final tests

Listeners took the tests in small groups, no more than three at a time. The stimuli were presented in a quiet 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. Response alternatives were exemplified on the answer sheets, as well as in the instructions by common English words in ordinary spelling with the target sound(s) underlined. The written instructions and the answer sheets have been reproduced in Appendix A4.4. 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. The verbatim text of the instructions is provided in Appendix A4.4.

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 instructions for this part of the experiment are included in Appendix A4.4.

In each part of the test we gave the listeners ample time to study the layout of the answer sheets (except for 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 two or part three.

At the end of the session listeners filled in a questionnaire providing information on their linguistic background and their prior exposure to English (for Dutch and Chinese listeners) or to Dutch and Chinese-accented English (for American listeners). The text of the questionnaires is included in Appendix A4.3.

---

board to use them in my experiments. My two-weeks' stay at UCLA was funded in part by professor Kirsner.

The results have not been analyzed systematically but were used on the spot to determine whether or not a listener was indeed an admissible subject.

4.6.4 Data presentation in the next chapters

In this chapter we have described the procedures observed to collect the materials for our study on the mutual intelligibility of Chinese, Dutch and American speakers of English. In so far as we presented results in this chapter we did so as part of the selection process needed to locate the optimally representative male and female speaker within each of the three language groups. I will provide a detailed presentation of the results in the next four chapters. In Chapter five, we will present the mutual intelligibility in the nine combinations of speaker and listener nationalities in terms of vowel identification. Chapter six will do the same for simplex consonants and consonant clusters. Chapter seven presents the results for the word recognition tests, both in meaningless and in meaningful (low and high predictability) sentences. In Chapter eight, we return to the vowel and consonant identification scores obtained for the full sets of ten male and ten female speakers per nationality. We will examine in that chapter to what extent the variability in the vowel and consonant identification scores can be explained by acoustical properties (or the lack thereof) in the tokens produced. Such an acoustical analysis might reveal systematic differences in the way the sounds of English are produced by native speakers of (American) English, and how these sounds differ from the realizations produced by Chinese and Dutch ESL speakers. We predict, of course, that perceptual confusions can be related to lack of acoustical contrast between the sounds concerned, whether in terms of quality (vowel formants), temporal structure (vowel and consonant duration), or voice onset time (VOT).

Chapter Five

Acoustic analysis of vowels[1]

5.1. Introduction

In this chapter we will provide an acoustical analysis of the vowel tokens produced by the 20 Chinese, 20 Dutch and 20 American speakers of English, which were recorded in the course of the project. A description of the materials and the method of data collection were given in Chapter four.

As was explained in Chapter three, the vowel systems of (Mandarin) Chinese, Dutch and American English differ considerably, both in the number of vowels in the inventory and in the details of their position within the articulatory vowel space, and possibly also in terms of their durational characteristics. Although the phonetic differences are typically described in terms of articulatory properties, we have not tried to determine articulatory properties if the vowels through physiological measurements – as we had no recourse to the type of equipment needed, such as X-ray photography, Magnetic Resonance Imaging (MRI) or Electromagnetic Mid-sagittal Articulography (EMMA). Rather we used acoustic measurements that are known to have rather clear correspondences with articulatory properties of vowels. How this is done will be explained briefly in the next section.

5.1.1. Objective measurement of vowel quality

There is agreement among experimental phoneticians that vowel quality can be quantified with adequate precision and validity by measuring the center frequencies of the lower resonances in the acoustic signal. Specifically, the center frequency of the lowest resonance of the vocal tract, called first formant frequency or F1, corresponds closely to the articulatory and/or perceptual dimension of vowel height (high vs. low vowels, or close vs. open vowels). For an average male voice, the F1 values ranges between 200 hertz (Hz) for a high vowel /i/ to some 800 Hz for a low vowel /a/. The second formant frequency (or F2) reflects the place of maximal constriction during the production of the vowel, i.e., the front vs. back dimension, such that the F2 values range from roughly 2200 Hz for front /i/ down to some 600 Hz for back /u/. For female voices the formant frequencies are some 10 to 15% higher, on account of the fact that the resonance cavities in the female vocal tract are smaller (shorter) by 10 to 15% than those of a male speaker.

---

[1] This chapter is a slightly adapted version of H. Wang and V. J. Van Heuven (2006) Acoustical analysis of English vowels produced by Chinese, Dutch and American speakers. In J. M. Van de Weijer and B. Los (eds.) *Linguistics in the Netherlands 2006*. Amsterdam/ Philadelphia: John Benjamins, 237–248.

The relationship between the formant frequencies and the corresponding perceived vowel quality is not linear. For instance, a change in F1 from 200 to 300 Hz brings about a much larger change in perceived vowel quality (height) than a numerically equal change from 700 to 800 Hz. Over the past decades experimental phoneticians and psycho-physicists have developed an empirical formula that adequately maps the differences in hertz-values onto the perceptual vowel quality domain, using the so-called Bark transformation (for a summary of positions see Hayward, 2000). Using this transformation, the perceptual distance between any two vowel qualities can be computed from acoustic measurements.

We used the Bark formula advocated by Traunmüller (1990):

$$\text{Bark} = [(26.81 \times F) / (1960 + F)] - 0.53,$$

where F represents the measured formant frequency in hertz.

For many languages formant measurements have been published, so that an adequate determination can be made of the vowel systems of those languages. Probably the best known set of formant measurements was produced for American English, in the early fifties by Peterson and Barney (1952) for male and for female speakers separately (see also Chapter three). These authors used the same stimuli that we used, i.e. vowels embedded in a /hVd/ consonant frame. Similar vowel sets were recorded for 50 male and 25 female speakers of Dutch by Pols and co-workers in the seventies (Pols, Van de Kamp and Plomp, 1973 and Van Nierop, Pols and Plomp, 1973, respectively). Formant measurements for the vowels of Mandarin (Beijing dialect) became available only recently (Li, Yu, Chen and Wang, 2004).

Formant measurements for Chinese-accented English (aiming at the American pronunciation norm) were published by Chen, Robb, Gilbert and Lerman (2001). The authors recorded a subset of the American English vowels (eleven monophthongs) in the same /h_d/ monosyllables that we used ourselves. However, their speakers (20 male and 20 female adults) had been living in the USA for at least two years after having received intensive exposure to spoken English in China in order to qualify for the TOEFL test required to enter a university in the USA. This is clearly a different type of ESL speaker than we target in our study, so that it makes every sense that we should measure the formants in our speaker group separately. We would predict, of course, that certain vowels that are acoustically indistinct in our dataset will be more clearly differentiated in Chen et al.'s (2001) data but not so clearly as when spoken by American native speakers. Moreover, no data are available in Chen et al. (2001) on the perception of the ESL tokens; so that it is unclear to what extent the vowels produced by their advanced Chinese learners of English were correctly identified by either Chinese or American listeners.

No formant data have ever been published for Dutch-accented English vowels. However, several studies have been done on the perceptual mapping of English vowels by Dutch ESL speakers. In such studies, a large number of vowel tokens were generated by speech synthesis covering the acoustical vowel space according to a finely-meshed grid. Listeners, whether native or foreign, were then instructed to indicate for each artificial vowel sound which of the vowels in the target language would be most compatible with it (often with a goodness or typicality rating). The responses allow the researcher to reconstruct the perceptual vowel space of the

listener in terms of the prototypical vowel exemplar for each perceptual category and some area of tolerance around the prototype, where more or less acceptable tokens of the category may occur. Unfortunately, the Dutch ESL listeners were all university students of English (Schouten, 1975) or Dutch-English bilinguals (Broerse, 1997), and therefore cannot provide a basis for comparison for our study. [2] There is no alternative, then, but to measure the acoustical properties of Dutch-accented English vowels ourselves, using the data we collected in the present study.

5.1.2. The problem of vowel normalization

Unfortunately, formant values measured for the same vowel differ when the vowels are produced by different individuals. The larger the differences between two speakers in shape and size of the cavities in their vocal tracts, the larger the differences in formant values of perceptually identical vowel tokens are. Given that the vocal tracts of women are some 15 percent smaller than those of men, comparison of formant values is especially hazardous across speakers of different sex. Numerous attempts have been made, therefore, to factor out the speaker-individual component from the raw formant values such that phonetically identical vowels spoken by different individuals would come out with the same values. None of these vowel normalization procedures have proven fully satisfactory (Adank, Van Heuven and Van Hout, 1999; Labov, 2001: 157-164; Rietveld and Van Heuven, 2001). Broadly, two approaches to the normalization problem have been taken in the literature (for a detailed discussion of the issue of vowel normalization, see also Nearey, 1989). The first approach, called intrinsic normalization, tries to solve the problem by considering only information that is contained in the single vowel token under consideration, typically by computing ratios between pairs of formant values such as F1/F0 and F2/F1.[3] The alternative, extrinsic normalization, looks at tokens of all the vowels in the phoneme inventory of a speaker and expresses the position of one vowel token relative to the other tokens within the individual speaker's vowel space.

In the present study we have opted for a straightforward extrinsic vowel normalization procedure, first used by Lobanov (1971), which is simply a z-normalization of the F1 and F2 frequencies over the vowel set produced by each individual speaker. In a z-normalization, the F1 and F2 values are transformed to z-scores by subtracting the individual speaker's mean F1 and mean F2 from the raw formant values, and then dividing the difference by the speaker's standard deviation. Z-transformed F1 values less then 0 then correspond to relatively close (high) vowels, values larger than 1 refer to rather open vowels. Similarly, negative z-scores for F2 refer to front vowels, whilst positive z-scores for F2 represent back vowels. In our case we applied the Lobanov normalization after first transforming the hertz values to Bark values.

---

[2] Also, in Broerse's study only the perceptual norms were determined for the checked (short, lax) vowels in the inventory.

[3] When formant values are rescaled to Bark, the numerical difference (F1–F0; F2–F1, etc.) is preferred over the ratio.

5.1.3 Vowel duration

The vowels of English and Dutch can be divided into two major groups on the basis of their phonological behavior, which largely correspond with phonetically short (and lax) versus long (and tense) vowels (for details, see Chapter three). Typically, the short/lax and long/tense vowels are in paired oppositions. In English, examples of such pairs are /i: ~ ɪ/ and /u: ~ ʊ/. Since vowel duration plays an important role in marking the contrast, next to vowel quality differences, we also measured the vowel duration in the tokens recorded in our dataset. Since some speakers speak faster than others, raw vowel duration cannot be used in the comparison. Rather, durations should be normalized within speakers. Here, too, we used a simple z-normalization procedure (see above) so that negative normalized durations refer to relatively short vowel tokens, and positive values represent relatively long vowel durations.

Chinese does not use exploit length as a vowel feature at the phonological level. We would predict (see Chapter three) that Chinese ESL speakers will make less difference between the short (lax) and long (tense) vowels of English – whether as subsets in the vowel inventory or in pairwise oppositions – than Dutch ESL speakers, and certainly less clearly so than native speakers of English.

5.1.4 Selecting vowels for analysis

Our recordings contain tokens of 19 vowel types, that is, if the speakers had indeed spoken British English. Given that our speakers, including the Dutch speakers, without having been instructed to do so used an American-style of pronunciation, without r-coloured vowels (so-called murmur diphthongs), there seems little point in measuring the vowels that were followed by /r/. Therefore we eliminated the tokens representing *here'd, haired, hard, hoored* and *heard*. Next, we decided not to include any full diphthongs as these would introduce the complication of having to trace the spectral change over the course of the vowels. This eliminated the types *hide, how'd* and *hoyed*. What remained is precisely the set that was also measured in Chen et al. (2001). We finally decided also to eliminate the /ɔ:/ type. It appeared that our speakers did not systematically differentiate between this vowel and /ɔ/. Moreover, quite a few of our L2 speakers pronounced *hawed* as /haud/.

5.2 Formant plots

Using the Praat speech processing software (Boersma and Weenink, 1996) the beginnings and end points of the target vowels were located in oscillographic and/or spectrographic displays. Formant tracks for the lowest four formants, F1 through F4, were then computed using the Burg LPC algorithm implemented in Praat, and visually checked by superimposing the tracks on a wideband spectrogram. Whenever a mismatch between the tracks and the formant band in the spectrogram was detected, the model order of the LPC-analysis was changed *ad hoc* until a proper match was obtained between tracks and spectrogram. Once a satisfactory

match was obtained, the values for F1 and F2 were extracted at 25, 50, and 75% of the duration of the target vowel, as well as the vowel duration as such, and stored for off-line statistical processing.

Formant values were then converted to Bark (see § 5.1.1) and averaged over the ten male and ten female speakers in each speaker group separately. These mean F1 and F2 values are plotted in acoustical vowel diagrams in figures 5.1a-f for male and female Chinese, Dutch and American speakers of English. Each plot contains the position of the ten monophthongs selected as explained in § 5.1.4.

The **Chinese** ESL speakers' vowels show tight clustering, and therefore little spectral distinction between intended /i:/ and /ɪ/ (see figure 5.1a-b for the Chinese speakers). This result was predicted from the contrastive analysis of the Chinese and American English vowel systems in Chapter three. The lack of differentiation between the two vowels is very clear for the male speakers; there is some measure of spectral distinction in the Chinese female tokens. Similarly, there is hardly any spectral difference between intended /ɛ/ and /æ/, nor between /u:/ and /ʊ/. The lack of distinction in these two vowel pairs was also predicted by the contrastive analysis.

In spite of the lack of distinctive vowel pairs, we may observe that the Chinese ESL speakers spread their vowels over a large portion of the acoustical vowel space. Although the number of (phonological) vowels in Chinese is relative small (between seven and ten, see Chapter three), this does not prevent Chinese ESL speakers from using a very large vowel space. Probably, this is a consequence of the much larger number of distinct vowel allophones in Chinese, which gives Chinese ESL speaker an advantage. The substitution of context-dependent allophones is not predicted, however, by Flege's Speech Learning Model (Chapter two).

We divided our American English inventory of ten monophthongal vowels into two subsets, corresponding to five tense vowels and five lax vowels. Here, the vowel /ɔ/ is classed as a tense vowel on the grounds that it is a merger of tense /ɔ:/ and lax /ɔ/. Its location in the vowel space (see figure 5.1e-f for the American speakers) motivates this choice quite clearly. Also, we classified the open front vowel /æ/ as tense, though not on phonological or distributional grounds (it would be phonologically lax since it cannot occur at the end of a word, see Chapter three). Phonetically, however, there is good reason to consider American /æ/ a tense vowel: it is clearly longer than all other lax vowels, and is in fact as long as any tense vowel in the system, and it is also peripheral, that is, on the outer edge of the vowel space. This must also have been the (implicit) reason prompting Strange, Bohn, Nishi and Trent (2004) to classify American /æ/ as tense.[4] In figure 5.1 the five tense vowels have been linked with a solid line; the lax vowels have been linked with a dotted line. We observe, in figure 5.1a-b that the tense and lax vowel polygons largely overlap, indicating that the Chinese ESL speakers basically fail to spectrally distinguish between the spectrally more peripheral tense set and the spectrally reduced (centralised) lax set.

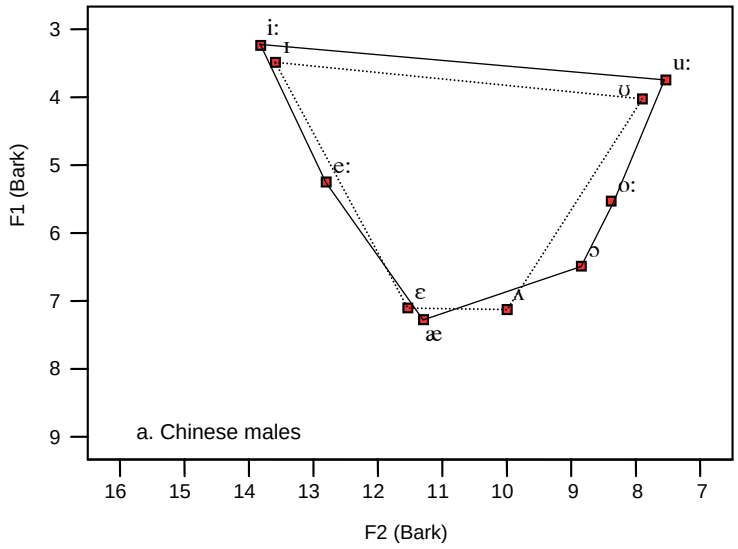
The ESL tokens produced by the **Dutch** speakers are generally distributed over a much smaller portion of the vowel space than the Chinese ESL tokens. One reason for the apparently shrunken vowel space in Dutch ESL may be that Dutch speakers

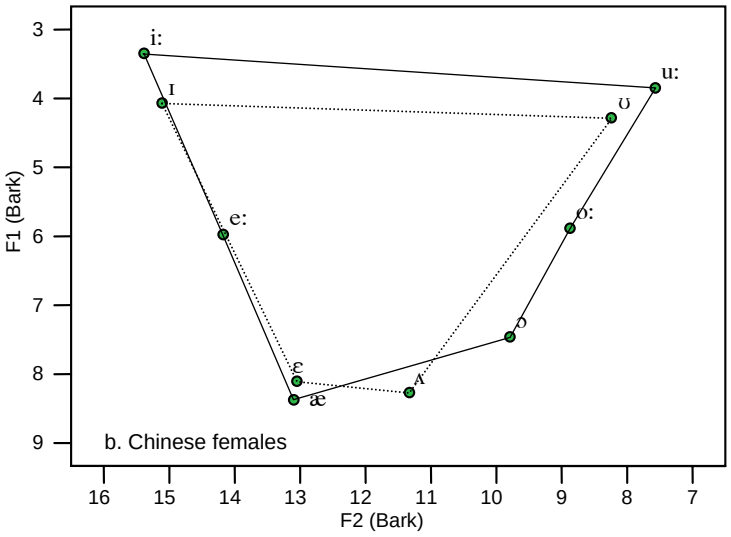
[4] Strange et al. (2004) plot the eleven monophthongal vowels of American English (same as in Chen et al., 2001) recorded by four male speakers in disyllabic /hVba/ frames.

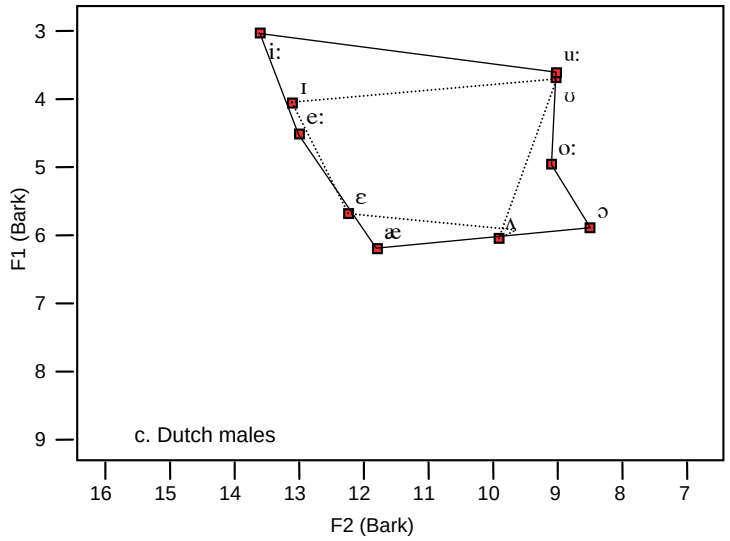
reserve the most open part of their vowel space for the vowel /a:/, as in Dutch *taak* /ta:k/ 'task', for which they have no use in English. In spite of the rather contracted vowel space, the vowels within the space seem spectrally more distinct than those of the Chinese speakers. There is a clear spectral difference between intended /i:/ and /ɪ/, which is predicted as positive transfer should occur from Dutch to English (see Chapter three). There is a fair degree of separation between intended /ɛ/ and /æ/. Although the separation is not as large as in the native American speech (see below), the success on the part of the Dutch speakers is unexpected, and in fact runs counter to the prediction from the contrastive analysis in Chapter three. The /ɛ/ ~ /æ/ contrast is typically listed as a cause for the formation of new sounds (Flege) and we are surprised to find that in our group of ESL speakers some notion of the difference has already been established. Interestingly, the other vowel pair that has traditionally been mentioned as a cause for the formation of new sounds, /u:/ ~ /ʊ/, remains completely undifferentiated in the Dutch ESL speakers – as predicted by the contrastive analysis in Chapter three.

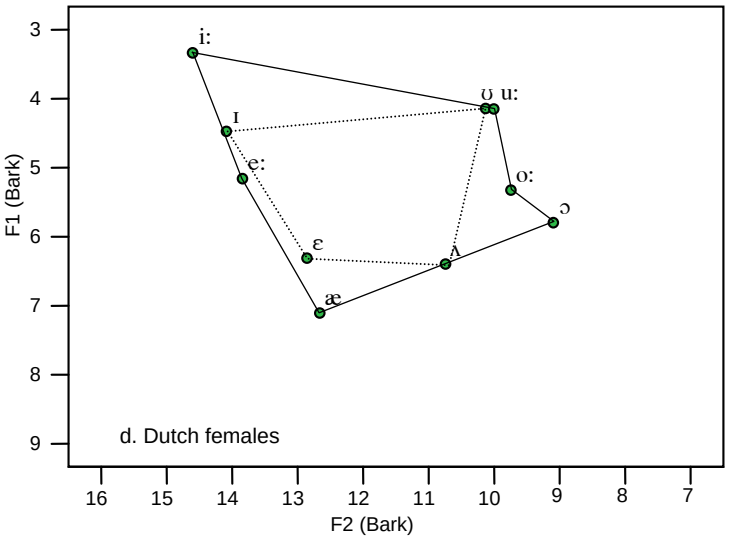
Dutch and English both have tense and lax vowel subsets. Inspection of figure 5.1c-d, however, shows that the tense and lax subsets are not very clearly separated in Dutch ESL. One reason for the relatively poor separation between the subsets is the lack of a /u: ~ ʊ/ contrast in Dutch. The Dutch speakers do not spectrally distinguish between the two, so that here the two subsystems merge. Also, at the lower edge of the vowel space there is little differentiation between more centralized (half) open lax vowels and peripheral open tense vowels as the Dutch ESL speakers do lower /ɔ/ as much as they should for American English, and at the same time observe insufficient contrast between /ɛ/ and /æ/.

If we now turn to the **American** native realisation of the vowels, in figure 5.1e-f, we notice that the vowel spaces are larger than those found for the Dutch ESL speakers, but much smaller than those of the Chinese ESL speakers. Nevertheless, the American native vowels are spectrally much more distinct than those produced by the Dutch speakers, and even more so than the Chinese ESL vowels. There are very large spectral differences between the members of the pairs /i: ~ ɪ/, /ɛ ~ æ/ and /u: ~ ʊ/. Moreover, the figure illustrates quite convincingly that the tense and lax vowel subsets are organised in terms of an outer (peripheral) and an inner (more centralised) area. In this respect, too, the L1 speakers clearly differ from both the Dutch and (even more) from the Chinese ESL speakers


i:
ɪ
u:
ʊ
e:
o:
ɔ
ɛ
ʌ
æ
a. Chinese males
F1 (Bark)
F2 (Bark)
3
4
5
6
7
8
9
16
15
14
13
12
11
10
9
8
7


i:
ɪ
u:
ʊ
e:
o:
ɔ
ɛ
ʌ
æ
b. Chinese females
F1 (Bark)
F2 (Bark)
3
4
5
6
7
8
9
16
15
14
13
12
11
10
9
8
7


i:
u:
ɪ
ʊ
e:
o:
ɛ
ɔ
ʌ
æ
F1 (Bark)
F2 (Bark)
3
4
5
6
7
8
9
16
15
14
13
12
11
10
9
8
7
c. Dutch males


i:
ʊ u:
ɪ
e:
o:
ɔ
ɛ
ʌ
æ
F1 (Bark)
F2 (Bark)
3
4
5
6
7
8
9
16
15
14
13
12
11
10
9
8
7
d. Dutch females

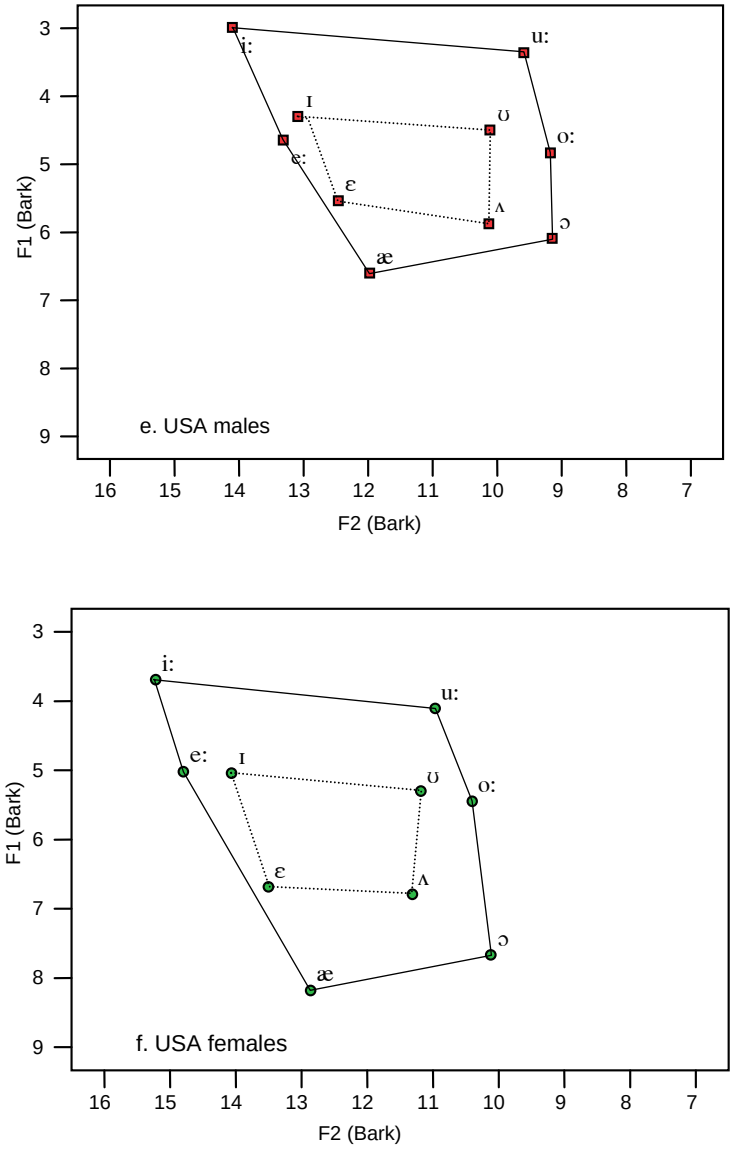


Figure 5.1. The mean values of F1 and F2 (in Bark) of the ten American English monophthongs plotted separately for tense (solid polygons) and lax (dotted polygons) vowels for six groups of speakers (indicated in each panel). Male speakers are represented by squares, female speakers by circles.

5.3 Vowel duration in Chinese, Dutch and American English

The vowels of American English are often separated into two length categories, short and long (see Chapter three). Phonetically, the four short vowels, /ɪ, ɛ, ʌ, ʊ/ do not only have short durations, they also take up more centralized positions in the vowel space. For this reason, the set of four may be called lax as well. The other vowels of American English are long and have positions along the outer perimeter of the vowel space. These are, in the present restricted dataset, the vowels /i:, e:, æ, ɔ, o:, u:/.

Since vowel duration may be expected to contribute to the perceptual identification of vowel tokens by English listeners, we measured vowel duration in each of the 600 tokens in our dataset, and plotted mean vowel duration for each of the ten types, separately for lax and tense categories in figure 5.2a for Chinese ESL speakers, in panel b for the Dutch speakers and in panel c for the American L1 speakers.

Taking the native speakers as our starting point, figure 5.2c clearly shows that the four lax/short vowels have much shorter duration (with means between 169 and 185 ms) than the six long/tense vowels (with means between 225 and 266 ms). As a result of this, vowels that are spectrally close to each other, such as /e:/ (266 ms) and /ɪ/ (184 ms), are yet acoustically distinct. Note also that when the vowels are ordered from short to long, as has been done in figure 5.2c, the increment between adjacent vowels in the figure is never more than 14 ms (which is the difference in mean vowel duration between /o:/ and /æ/). However, the discrepancy between the longest of the short vowels (/ɛ/, 185 ms) and the shortest of the long vowels (/u:/, 225 ms) is 40 ms. These results can be taken in evidence of the phonetic correctness of the subdivision of the American English vowels into the short and long categories made here.

If we now consider the vowel durations produced by the Chinese ESL speakers (figure 5.2a) we note that the short vowels are roughly within the duration range of the American L1 speakers. Also, the long vowels are generally within the native range for long vowels, with the exception of the vowels /æ/ and /ɔ/. Interestingly, these are precisely the vowels that distributionally pattern with the short vowels, as they cannot occur at the end of a word in English. When foreign learners are trained to pronounce English according to British (RP) norms, short vowel duration for /æ/ and /ɔ/ could reasonably be expected. However, given the fact that the Chinese ESL speakers were taught according to American pronunciation norms, this explanation is ruled out. We must assume, therefore, that the vowel duration of /æ/ and /ɔ/ has an incorrect perceptual representation in Chinese ESL speakers.

The Dutch ESL vowel durations are surprisingly similar to the Chinese realisations. Again, there are two gross duration categories, one for short vowels with durations less than 200 ms, and one for long vowels with durations in excess of 240 ms. As in the Chinese ESL tokens, the Dutch speakers make the long vowels /æ/ (208 ms) and /ɔ/ (172 ms) too short by American-English standards. Moreover, the Dutch speakers, who did not differentiate between /u:/ and /ʊ/ in spectral terms (see figure 5.1c-d), also have a tendency to make the short /ʊ/ too long (202 ms) – even though this is still some 40 ms shorter than their mean duration for long /u:/. Unexpectedly, then, it seems as if the Dutch ESL speakers are not more successful

in keeping the American-English lax and tense vowels distinct than the Chinese speakers, even though Dutch is language with a tense ~ lax subdivision, which is not the case for Mandarin.

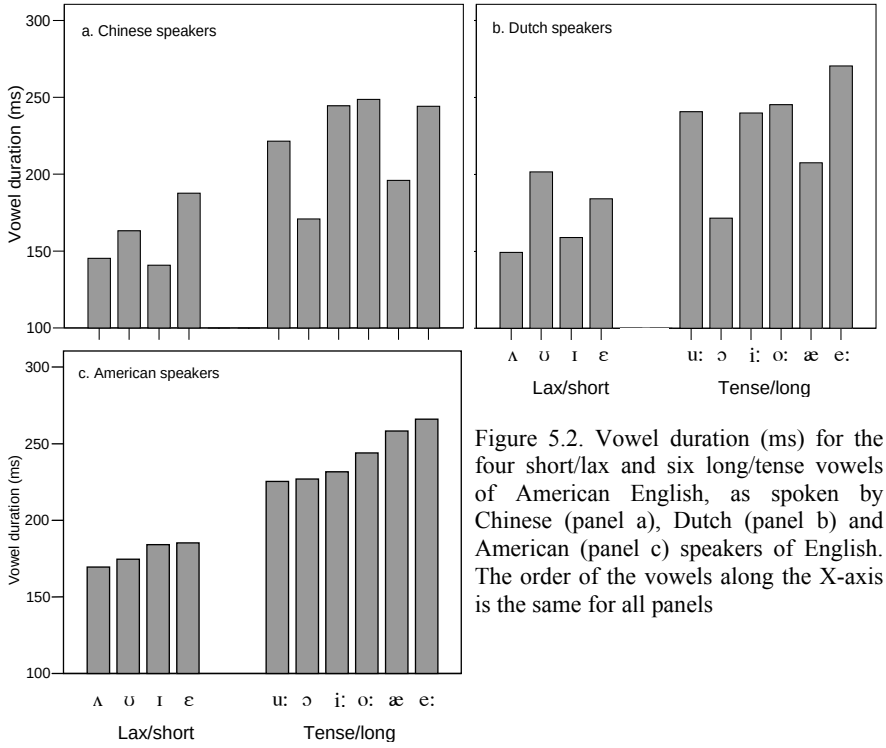


Figure 5.2. Vowel duration (ms) for the four short/lax and six long/tense vowels of American English, as spoken by Chinese (panel a), Dutch (panel b) and American (panel c) speakers of English. The order of the vowels along the X-axis is the same for all panels

5.4 Automatic vowel classification

So far we have only considered the means of the realisations of the vowels – in terms of vowel quality (F1 and F2) and of duration – averaged over groups of ten male and ten female speakers. The means do not tell us anything about how well the individual speakers keep the vowels distinct in their pronunciation of English. Figure 5.3a-c plot the individual realisation of the vowels in the F1 by F2 plane as scatter clouds, enclosed by spreading ellipses. These ellipses are drawn along the principal component axes, optimally capturing the directionality of the scatter of the vowel tokens within one vowel type. The ellipses have been plotted at + and – 1 SD away from the F1-F2 centroids. Before computing the scatter points and the ellipses based upon them, however, speaker normalization had to be carried out – as explained in § 5.1.2 – in order to make the vowel tokens produced by different individuals of different genders comparable.

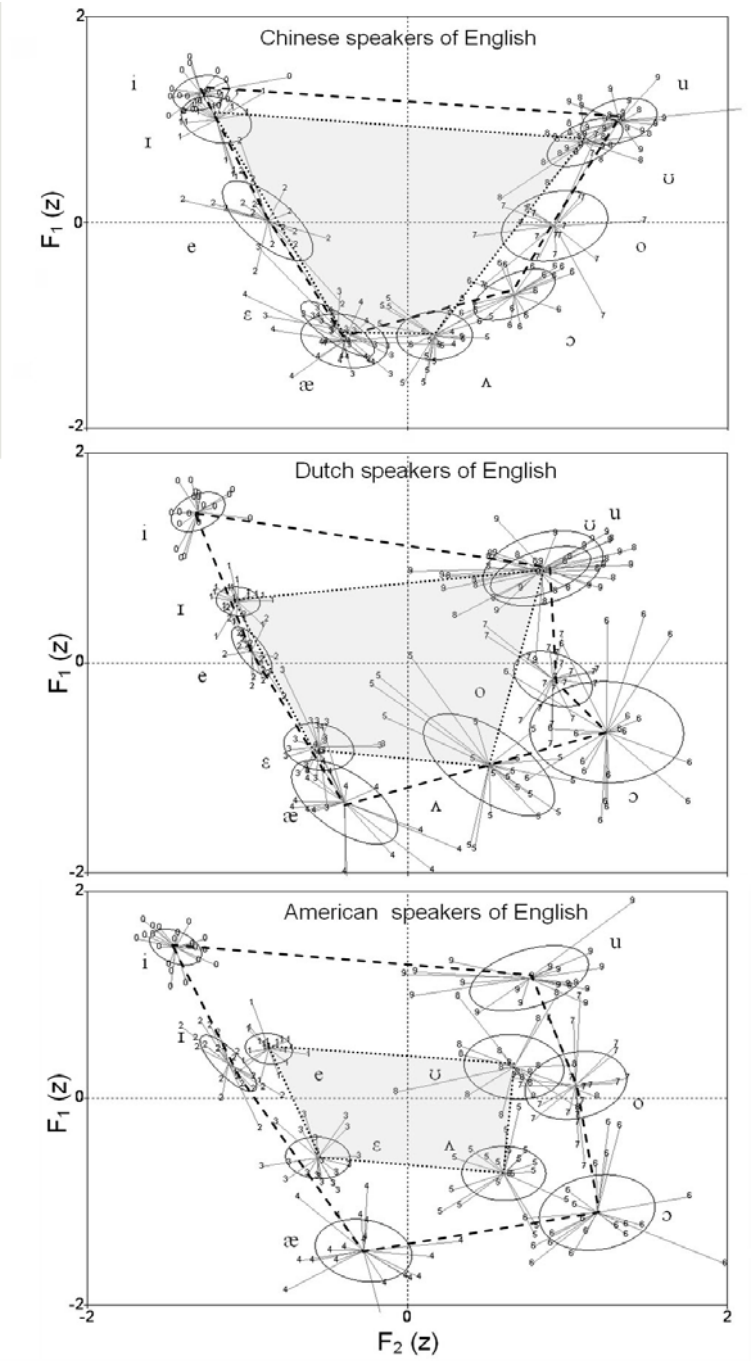


Figure 5.3a-c. Individual vowel points for Chinese speakers of English (after speaker-individual z normalization) plotted in the F1 by F2 plane, with spreading ellipses drawn at +/−1 SD away from the centroid along the first two principal component axes of the scatter clouds.

The figures show that, generally, the Chinese speakers (Figure 5.3a) have more overlap between the ellipses of neighbouring vowels than is the case in the Dutch ESL realizations (Figure 5.3b). The American native L1 speakers have the smallest (Figure 5.3c).

We will now attempt to quantify the difference between the three speaker groups in terms of the degree of success in keeping the ten vowels distinct. We have used Linear Discriminant Analysis (LDA) for this purpose. LDA is an algorithm that computes an optimal set of parameters (called discriminant functions) which automatically classifies objects in pre-established categories. For a comprehensive treatment of LDA in research on vowel identification, see Weenink (2006). The more distinct the categories are in the dataset, the fewer the number of classification errors yielded by the algorithm. In the case at hand, the discriminant functions are based on linear combinations of weighted acoustic parameters F1, F2 and duration. Again, before running the LDA, speaker normalization was carried out using the z-transformation on the formant frequencies in Barks. The results of the LDA are presented in terms of confusion matrices (see Table 5.1 on the next page), which show the intended vowels in the rows against the vowels as classified by the algorithm in the columns. Correctly classified vowel tokens are in the cells along the main diagonal. All off-diagonal cells contain confusions.

I will first present the overall percentage of correctly classified vowel tokens of Chinese, Dutch and American speakers of English. Moreover, we ran the LDA twice. The first time we just included the two spectral parameters as possible predictors of vowel identity, i.e. F1 and F2 (converted to Bark and z-normalized within individual speakers). The second time we extended the set of predictors by also including (z-normalised) vowel duration. Figure 5.4 presents these results.

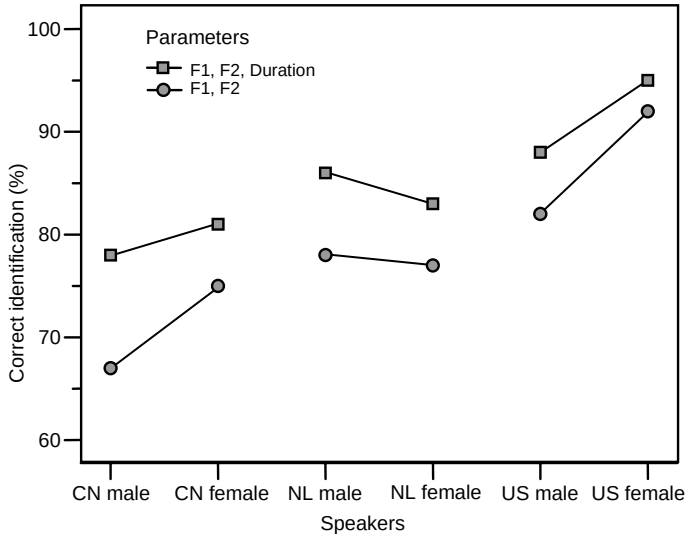


Figure 5.4. Percentage of correctly classified vowel tokens by Linear Discriminant Analysis with F1 and F2 as predictors, and with duration added as a third predictor for six groups of speakers (male and female Chinese, Dutch and American speakers of English).

Figure 5.4 shows at once that the vowels as spoken by the native speakers afford the best automatic identification, those spoken by the Dutch learners can be less successfully identified, and the Chinese ESL tokens are poorest. Adding duration to the set of predictors boosts the correct identification by some 10 percentage points (a little less for the American L1 vowel tokens, possibly due to a ceiling effect). Finally, the vowel tokens produced by the female speakers tend to be more distinct, and therefore better identified, than those spoken by the males. However, there is no such gender effect in the Dutch vowel set.

A more detailed view of the LDA results is presented in Table 5.1, where percent predicted vowel identity is crosstabulated against the actual vowel identity for Chinese, Dutch and American native speakers in the upper, middle and lower panels, respectively. The results presented in this table were based on the output of the LDAs which used F1, F2 and vowel duration as predictor variables.

The results obtained for the Chinese-accented vowel tokens reveal two major problems, viz. the more or less symmetrical confusion of /ɛ/ and /æ/ and an asymmetrical confusion of lax /ʊ/ with tense /u/ (but not vice versa). These pronunciation errors follow from a traditional contrastive analysis, and were also noted in a pedagogical textbook (Zhao, 1995).

In the results for the Dutch speakers of English we find two symmetrical error patterns, i.e. /ɛ/ ~ /æ/ and /ʊ/ ~ /u/ and their counterparts, all of which were predicted by contrastive analyses (Table 3.4) and were noted in the pedagogical literature (Tables 3.5 and 3.6 for Dutch and Chinese speakers, respectively). One incorrect classification type was never predicted, however. This is the incorrect classification of intended vowel /ʌ/ as a front vowel /ɛ/.

We will have occasion to review the LDA results from a different perspective in Chapter ten, where we will make an attempt to use the LDA to make predictions of cross-linguistic vowel perception, thus simulating for instance the perception of Dutch-accented vowel tokens by Chinese listeners of English. Before we discuss such attempts, we will first deal with the results of human perception of vowels, consonants, consonant clusters and words in meaningless and meaningful contexts in Chapters six through nine.

Table 5.1. Classification matrices of observed and predicted vowel identity of English vowel tokens produced by Chinese (upper panel), Dutch (middle panel) and American native speakers. Prediction of vowel identity made by Linear Discriminant Analysis using F1, F2 and vowel duration as predictors. Percent correct in parentheses.

Presented vowels	Vowel identity predicted from Chinese production data (80.0%)										
	æ	e:	ɛ	i:	ɪ	ɔ	o:	ʊ	ʌ	u:	Total
æ	**60**		30						10		100
e:		**85**	10	5							100
ɛ	20	5	**65**						10		100
i:				**95**	5						100
ɪ					**100**						100
ɔ						**80**	5		15		100
o:						20	**70**			10	100
ʊ							5	**70**		25	100
ʌ			5			10			**85**		100
u:								10		**90**	100

Presented vowels	Vowel identity predicted from Dutch production data (83.0%)										
	æ	e:	ɛ	i:	ɪ	ɔ	o:	ʊ	ʌ	u:	Total
æ	**75**		25								100
e:		**95**			5						100
ɛ	15		**80**		5						100
i:				**100**							100
ɪ					**100**						100
ɔ						**80**	5	10	5		100
o:						5	**90**	5			100
ʊ								**60**		40	100
ʌ			15			10			**75**		100
u:							5	20		**75**	100

Presented vowels	Vowel identity predicted from American production data (92.5%)										
	æ	e:	ɛ	i:	ɪ	ɔ	o:	ʊ	ʌ	u:	Total
æ	**100**										100
e:		**95**			5						100
ɛ			**100**								100
i:				**100**							100
ɪ					**100**						100
ɔ						**85**	10		5		100
o:						5	**80**	5		10	100
ʊ			5				5	**80**		10	100
ʌ						10		5	**85**		100
u:										**100**	100

Chapter six

Intelligibility of vowels[1]

6.1 Introduction

In this chapter we will present the results for intelligibility of vowels of the groups of listeners. These results are from 36 Dutch listeners in Leiden, 36 Chinese listeners in Changchun and 36 Americans in Los Angeles listening to the six selected, optimally representative speakers. We would like to find out, among other things, how the classification errors we found in the previous chapter by applying a Linear Discriminant Analysis to the acoustical properties of the vowel tokens are different from the actual human perception results. The correctness and classification matrixes presented in Chapter five will be the reference data for the present chapter.

As we predicted in Chapter three, the differences in the sound systems in the three native languages will lead to a foreign accent for the Chinese and Dutch speakers of English which consist in deviations from the generally accepted pronunciation norm of English that are reminiscent of the native language of the learners, either Chinese or Dutch. The established structures of the Chinese/Dutch representation must be confronted with speech data from the target language, English. As a source of variability in speech, can Dutch/Chinese-accented English be detrimental to speech perception? When listeners are unable to recognize phonetic segments, words or larger units, will the result be partial or complete misidentification? If so, how well are English vowels identified by native American, Chinese and Dutch listeners? What is their confusion structure? Can we relate the confusions to specific interference patterns that reflect structural properties of the mother tongue of the non-native speaker and/or listener (Chapter three)? Will the confusion structure be different from the automatic classification results in Chapter five? This chapter may provide answers to these questions.

---

[1] Summaries of Chapters six to nine have appeared in English as H. Wang and V. J. Van Heuven (2005) Mutual intelligibility of American, Chinese and Dutch-accented speakers of English. *Proceedings of Interspeech 2005*, Lisbon: ISCA, 2225–2228 and in Dutch as V.J. Van Heuven and H. Wang (2006) Onderlinge verstaanbaarheid van Chinese, Nederlandse en Amerikaanse sprekers van het Engels. In T. Koole, J. Nortier, B. Tahitu (eds.) *Artikelen van de vijfde sociolinguistische conferentie*, Delft: Eburon, 257–266.

6.2 Results

Our research focuses on English as the target language and Dutch and Chinese as the source languages. We compare the intelligibility of Chinese-accented English, Dutch-accented English and native American English in an attempt to clarify how well these people understand each other and themselves when they are speaking English with their respective accents.

We hypothesize that foreign-accented English must be more difficult for English listeners as the source language deviates more from English, but native listeners still have strategies which non-native listeners lack for coping with all sorts of non-optimal speech, including foreign accents. Generally, then, native English listeners will be at an advantage over foreigners when listening to non-native English. There may just be one exception to this rule: non-native listeners may understand their own accented English better than native English listeners do. Since the foreign listeners are acquainted with the interfering native language, they may be sensitive to cues in the source language that native English listeners fail to pick up. This is what was called the interlanguage benefit by Bent and Bradlow (2003). Provisional data showing that this effect does apply to the present problem were presented earlier by Wang and Van Heuven (2003, 2004).

6.2.1 Overall results

The overall results for vowel intelligibility are presented in Figure 1, broken down by nationality of listeners and broken down further by nationality of speakers.

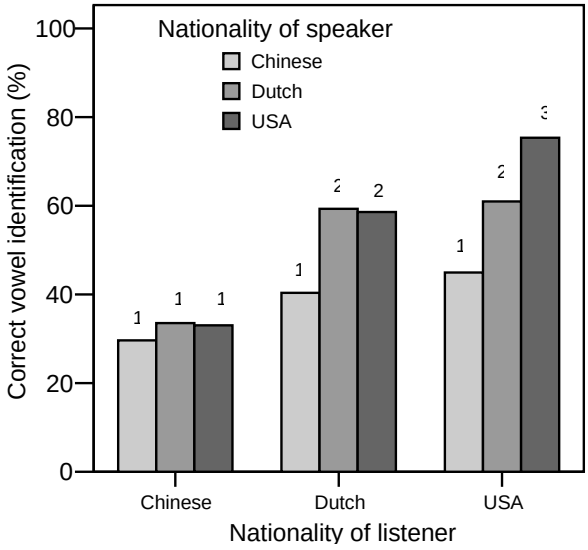


Figure 6.1. Percent correctly identified vowels for Chinese, Dutch and American listeners broken down by accent of speaker. Numbers above the bars indicate subgroup membership as determined by the Scheffé procedure (see text). Means and standard deviations are numerically specified in Appendix A6.1.

The data were submitted to an Analysis of Variance (ANOVA) run on the mean percent correct scores for each listener with nationality of speaker and nationality of listener as fixed factors.

Across speaker groups, the Chinese listeners have the lowest vowel identification scores (29–34% correct, mean = 32%). Dutch listeners perform at an intermediate level (40–59% correct, mean = 53%), and the American listeners are the best (45–75% correct, mean = 60%).[2] The effect of listener group was significant, $F(2, 315) = 204.9$ ($p < .001$). A post-hoc test (Scheffé procedure with $\alpha = .05$) indicates that all three speaker nationalities were different from each other.

Across listener groups Chinese speakers obtained the lowest vowel identification scores (38%). The Dutch and American speakers' vowels were identified with 51% and 56% correct, respectively. The effect of speaker nationality is significant, $F(2, 315) = 77.7$ ($p < .001$). The Chinese speakers are significantly poorer that the other two nationalities, which do not differ from each other. We may note that the effect of listener nationality is almost three times larger than the effect of speaker nationality.

Crucially, the interaction between listener and speaker nationality also reaches significance, $F(4, 315) = 17.0$ ($p < .001$). This implies that the mean scores obtained for specific combinations of listener and speaker nationalities cannot be computed by simply adding or subtracting a term for each factor level. Specifically, it can be shown that listeners obtain higher vowel identification scores when responding to materials produced by speakers of their own native language. This can be shown by computing the expected scores for each of the nine possible combinations of listener and speaker nationality and then comparing this expected score with the observed score. Mean percent correct vowel identification equals 50. When the listeners are Chinese, Dutch and American, the expected score is –18, +3 and +10 below or above the mean; for the three speaker nationalities the mean should be corrected with –12, +1 and +6, respectively. The expected and observed scores are listed in table 6.1 together with the difference between the two (delta or prediction error).

Generally, the observed scores are correctly predicted or even overestimated 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 6.1). This native or interlanguage benefit is 5 to 10 percentage points better than the expected score.

---

[2] This result is different from a pilot test which showed that Dutch listeners performed best. In the pilot (Wang and Van Heuven, 2003) the Chinese listeners had the lowest vowel identification scores (50 to 60% correct). Dutch listeners performed best (65 to 80% correct), and the American listeners were intermediate (60 to 70% correct). Chinese-spoken vowels were most difficult for both Dutch and American listeners but not for Chinese speakers. Generally, listeners obtained the highest identification scores when responding to materials produced by speakers of their own native languages. This small advantage of Dutch-accented English for Chinese listeners may have been caused by the circumstance that our Chinese listeners had lived in the Netherlands for some six months, and therefore had had more exposure to Dutch-accented English than to L1 American English.

Table 6.1. Expected vowel identification scores (% correct) on the basis of grand mean = 50% and main effects for Listener and Speaker nationality for each combination of factor levels. Observed and error scores are indicated. Bolded delta's represent the interlanguage or native language benefit.

	Listener nationality		Speaker nationality		Expected	Observed	Δ
1.	Chinese	–18	Chinese	–12	20	30	**+10**
2.	Chinese	–18	Dutch	+1	33	34	+1
3.	Chinese	–18	American	+6	38	34	–4
4.	Dutch	+3	Chinese	–12	41	40	–1
5.	Dutch	+3	Dutch	+1	54	59	**+5**
6.	Dutch	+3	American	+6	59	59	0
7.	American	+10	Chinese	–12	48	45	–3
8.	American	+10	Dutch	+1	61	61	0
9.	American	+10	American	+6	66	75	**+9**

To conclude this part of the data presentation, we ran separate one-way ANOVAs in order to determine to what extent the three speaker nationalities differed within each of the three listener groups. Within the Chinese listeners, speaker was not a significant effect, $F(2, 105) = 1.4$ (ins.). In the Dutch listener group the Chinese speakers were more difficult to understand than either the Dutch or the American speakers, $F(2, 105) = 40.7$ ($p < .001$), who did not differ from each other (Scheffé). For American listeners the Chinese speakers were more difficult to understand than the Dutch speakers, who in turn were more difficult to understand than fellow Americans, $F(2, 105) = 69.9$ ($p < .001$), where all three speaker groups differed significantly. Significant differences between speaker groups have been indicated in figure 6.1 with superscript numbers over each bar.

6.2.2 Overview of the sound system

The experimental literature on foreign-language interference typically addresses one specific contrast at a time. For instance, there is a vast literature on the acquisition of the English /r ~ l/ contrast by speakers of Asian backgrounds (where the contrast is no part of the phonology). In the area of vowels much effort has been made to study the details of the acquisition of 'new' contrasts such as English /e ~ æ/ by Germans, or the English /iː ~ ɪ/ contrast by Hispanic learners (Flege, 1995). However, experimental studies targeting the confusion structure in an entire vowel inventory in a cross-linguistic setting are few and far between.

Before we present and analyze the confusion structure in the Chinese, Dutch and American tokens of English vowels, let us briefly recapitulate, in Table 6.2, the comparison of the three vowel systems as provided in Chapter three, in an attempt to derive specific predictions as to where confusions may arise in Chinese and Dutch-accented varieties of English.

Table 6.2. Summary of vowel systems of Mandarin Chinese, Dutch and English.

Chinese (source language)						
	Front		Central		Back	
High	i		y		u	
Mid	e		ɚ		o	
Low			ɑ			
Dutch (source language)						
	Front		Central		Back	
	Tense	Lax	Tense	Lax	Tense	Lax
High	i		y		u	
Hi-mid	e:	ɪ	ø:	ɵ	o:	
Lo-mid	ɛi	ɛ	œy			ɔ
Low			a:		ɑu	ɑ
English (target language)						
	Front		Central		Back	
	Tense	Lax	Tense	Lax	Tense	Lax
High	i:, ɪə^{r}				u:, ʊə^{r}	
Hi-mid	e:, ɛə^{r}	ɪ			o:, (ɔə^{r})	ʊ
Lo-mid		ɛ	ɚ:^{r}	ʌ	ɔ:, ɔi	ɔ
Low	ai	æ			ɑ:, ɑu	

6.2.3 Correct vowel identification

In order to obtain an overview of which vowels are more difficult than others, for each combination of speaker and listener nationality, we present percentages of vowels correctly identified by Chinese, Dutch and American listeners in separate panels for Figure 6.2. In each panel the results have been broken down by nationality of the speakers. We have simplified the presentation rather drastically by listing the results only for those vowels that can be considered monophthongs. All full diphthongs, vowels followed by /r/ and the strongly confused /ɔ:/ have been omitted from the figures (for details on this data reduction, see below). In the panels the ten monophthongs have been ordered in descending order of correct identification when the speakers are American. Generally we would expect the results for the non-native speakers, i.e. by Chinese and Dutch speakers, to fall below the percentage correct of the American speakers. Only in exceptional cases do we expect the non-native vowels to be identified better than the native vowels.

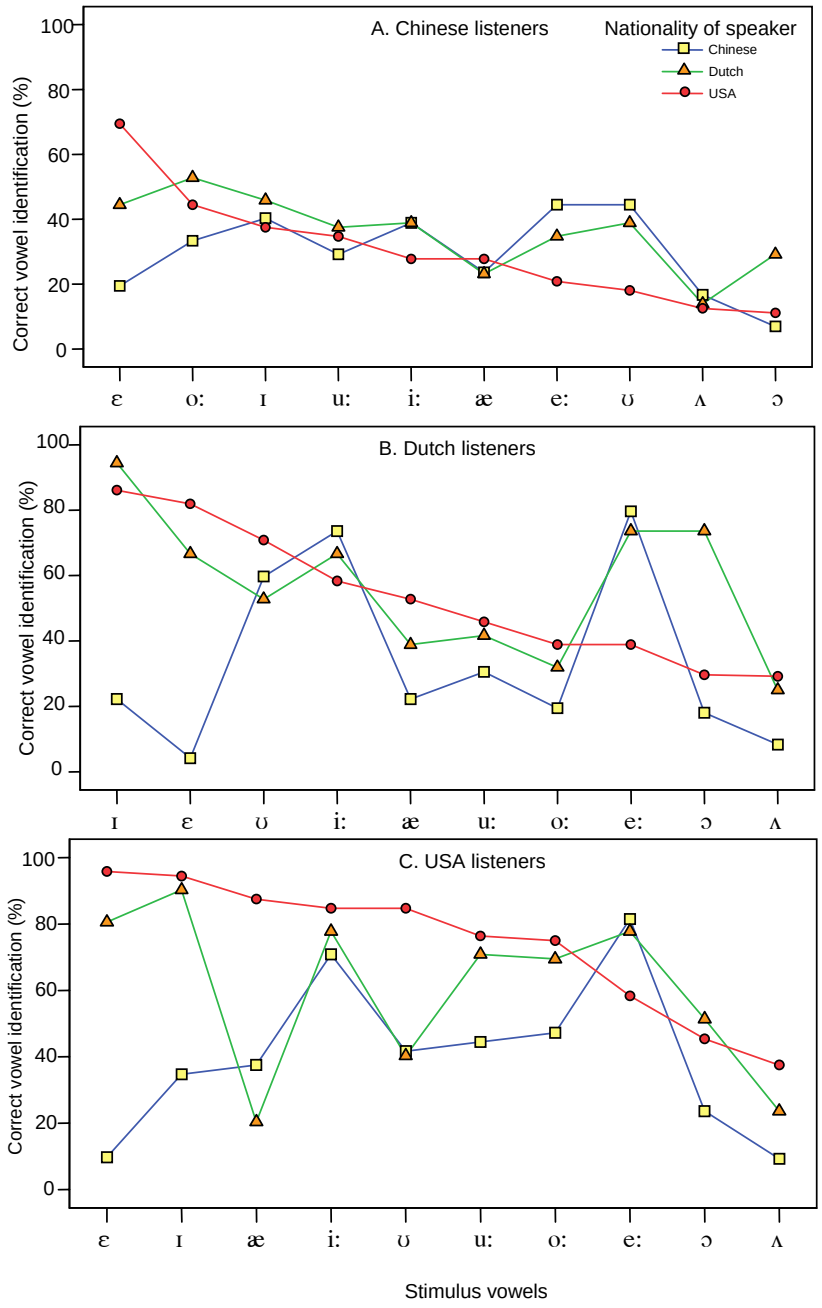


Figure 6.2 Correct vowel identification (%) for ten phonological English monophthongs produced by Chinese, Dutch and American speakers. Panels A, B and C present the results for Chinese, Dutch and American listeners, respectively.

Obviously, some vowels are more difficult than others. Moreover, there is hardly any correlation in the percentages correct identification of the vowels spoken by Chinese, Dutch and American speakers. That is to say, when a vowel spoken by an American speaker is easy to identify, it does not mean that the same vowel is also easy when it is produced by a Chinese or a Dutch speaker. Table 6.3 lists the correlation coefficients. In only one situation is the correlation significant, i.e., the identification of American and Dutch-spoken vowels is correlated for Chinese listeners.

Table 6.3. Pearson's correlation coefficients for identification of vowels produced by Chinese, Dutch and American speakers broken down by nationality of the listeners.

Listener nationality	Speaker nationalities		
	CN ~ NL	CN ~ US	NL ~ US
CN	r = 0.495	r = 0.032	r = 0.654*
NL	r = 0.277	r = –0.002	r = 0.513
US	r = 0.373	r = 0.118	r = 0.407

*: $p < 0.05$

Figure 6.2a shows that, relative to the American native speakers, the vowels /e:/ and /ʌ/ are easier to identify for **Chinese listeners** when the speakers are either Dutch or Chinese. The vowel /ɛ/, however, is clearly more difficult relative to the American pronunciation when it is spoken by a Dutch speaker, and even more so when the speaker is Chinese. In order to understand why the vowel /ɛ/ is a special source of difficulty we will have to examine its confusion structure, which we will defer to the next section.

When the **listeners are Dutch** (Figure 6.2b), we may observe that generally the American-spoken vowels are easier to identify correctly than the non-native tokens. Remarkably, the non-native tokens of /e:/, and for Dutch speakers also /o:/, are easier to identify than their American-accented counterparts. On the other hand, several non-native vowels are clearly more difficult than the native vowels. When the speakers are Chinese there are great difficulties with /ɪ/ and /ɛ/ as well as less severe problems with many of the other vowels: /ɛ, u:, o:, ɔ, ʌ/. When the speakers are Dutch themselves, there seem to be no specific difficulties.

American listeners (Figure 6.2c) are much better off listening to vowels spoken by fellow Americans than to foreign-accented vowels. Still, the non-native tokens of /e:/ are identified better by American listeners than their own tokens of /e:/. Also, Dutch-accented /o:/ is better than the American counterpart. Non-native /æ/ and /ʌ/ stand out as especially difficult vowels, as does lax /ɛ/ pronounced by Chinese speakers. Again, in order to better understand why certain vowels present special problems we need to know more about the specific confusion patterns in the vowel identifications, which is the topic of the next subsection.

6.2.4 Vowel confusion structures

6.2.4.1 Confusion matrices

Confusions in an identification task are customarily presented in a confusion matrix. Here the rows list the intended (stimulus) categories, while the columns represent the perceived categories. Correctly perceived stimuli appear in the cells along the main diagonal from top-left to bottom-right; errors are in the off-diagonal cells. As an illustration, Table 6.4 presents the confusion matrix for the 19 English vowels as produced by Chinese speakers and identified by American listeners.

In order to improve legibility, the cells in Table 6.4 have been shaded such that cells with larger numbers of observations in them have darker grey shades. Generally, the darkest cells find themselves along the main diagonal, indicating that very often the vowels as intended by the Chinese speakers were correctly identified by the American listeners. The values in the cells are percent correctly identified vowels relative to the row marginals, i.e. percentages should add up to 100 in each row of the matrix. Grey cells that are off the main diagonal represent substantial amounts of error or confusion. There are several concentrations of confusion in the table. For instance, tense /i:/ and lax /ɪ/ are strongly confused: /i:/ is misperceived by the American listeners (i.e. mispronounced by the Chinese speakers) in 25% whereas /ɪ/ is mistaken for /i:/ in more than half of the cases (53%). A similar confusion pattern can be seen further down the diagonal where there is a similar confusion pattern for tense /u:/ and lax /ʊ/, indicating that possibly all tense~lax contrasts are a source of error in the communication between Chinese speakers of English and American listeners. Interestingly, it also seems a recurring pattern that the tense vowel is dominant in the confusion pattern: the lax counterpart is confused more often with the tense vowel than vice versa. Such asymmetrical confusion patterns are often found in vowel perception studies.

There are several more concentrations of confusion in the table, which we will not analyze here. The point at issue here is that we need some method to extract and highlight the confusion structure in tables such as 6.4. Several methods have been proposed and applied in the literature. We will briefly review these, and then decide not to use any of these. Instead, we will propose a more practical analytical tool for our purpose, and then use this tool when analyzing the confusion structures in each of the nine combinations of speaker and listener nationalities. The full set of confusion matrices has been included in Appendix A6.2. In the main text of this chapter we will present a selection of the most obvious confusions in confusion graphs.

Table 6.4. Sample confusion matrix for 19 American English vowels produced by Chinese speakers and identified by American listeners.

		Response vowel (American)																		
		i:	ɪ	e:	ɛ	ɑ:	æ	u:	ʊ	ɔ:	ɔ	o	ʌ	ʌr	ai	ɔi	ɑu	iə	uə	ɛə
Stimulus vowel (Chinese)	i:	**71**	25	1		1					1									
	ɪ	53	**36**	1					6		3								1	
	e:	1		**85**	4		1				1				2	2	4		1	
	ɛ		3	8	**10**	3	7						1	1	63	1	1			
	ɑ:			1		**52**	10	1		7	13		3	9					3	
	æ			6	4		**39**	3		1					46			1		
	u:		1	1				**45**	39			3	1				3	1	3	
	ʊ				1	1	4	46	**42**								4			
	ɔ:		1	3		10	1	3		**18**		1		4	25	3	29			
	ɔ			2	2	6	2			18	**26**					2	42	2		
	o	3			1	1	3	3	29	1	4	**49**			1		3			
	ʌ		1	1	1	6	69	1		1	8		**9**		1	2				
	ʌr		1			14			4			1	1	**68**	1	1		1	6	
	ai		1	1		4	20			4	3		1		**61**		1	1		
	ɔi			1	3	1						4			37	**51**			3	
	ɑu	1			1		1	1	1	21	19	1		1	1	3	**46**			
	iə	1	1											21	1			**70**	1	3
	uə			1		3	1		1					7		6	1		**79**	
	ɛə					42	4			1				25	1	4		1		**20**

6.2.4.2 Extracting confusion patterns

Hierarchical cluster schemes (HCSs, Kruskal, 1964) have often been advanced as an analytic tool for extracting confusion structures from tables such as 6.5. The output of an HCS is a tree structure that visualizes which vowels constitute highly confusable subsets in the table. Alternatively, data reduction can be attempted by Multidimensional Scaling (MDS, Kruskal and Wish, 1978). We feel, however, that neither HCS not MDS do full justice to what actually goes on in the data. Both techniques presuppose a symmetrical confusion matrix, that is, the likelihood of vowel *x* to be confused with vowel *y* must be equal to that of *y* being confused with *x*. As Table 6.5 shows, this is not generally the case. Perceptual asymmetries such as those shown between tense and lax counterparts cannot be expressed in HCS or MDS; for instance, the asymmetrical confusions between /i:/ and /ɪ/ would average to a symmetrical 39%. For the sake of illustration we present just one HCS dendrogram and explain what features of the confusion structure are overlooked by the technique.

The dendrogram shows that the most confusable vowel pairs are /u:/ and /ʊ/. At approximately the same high level there is confusion between the pairs /ɛ/ and /ai/, and between /ɔ/ and /ɑu/. The tree also shows that there is just a little less confusion between the vowels /i:/ and /ɪ/ and between /ɑ:/ and /ɛə/. The /ɛ+ai/ cluster is joined at the next level by /æ/, indicating that /æ/ constitutes a more cohesive cluster with /ɛ+ai/ than any other vowel(s). In this way the tree structure seems to reveal the existence of roughly four more or less cohesive groups of vowels, plus a number of

isolates. The groups would be the high back vowels (/u:, ʊ, o:/, the low front vowels /ɛ, ai, æ, ɔi/, and the low back vowels /ɔ, ɑu, ɔ:/. The fourth group is not phonetically interpretable /ɑ:, ɛə, ʌr/.

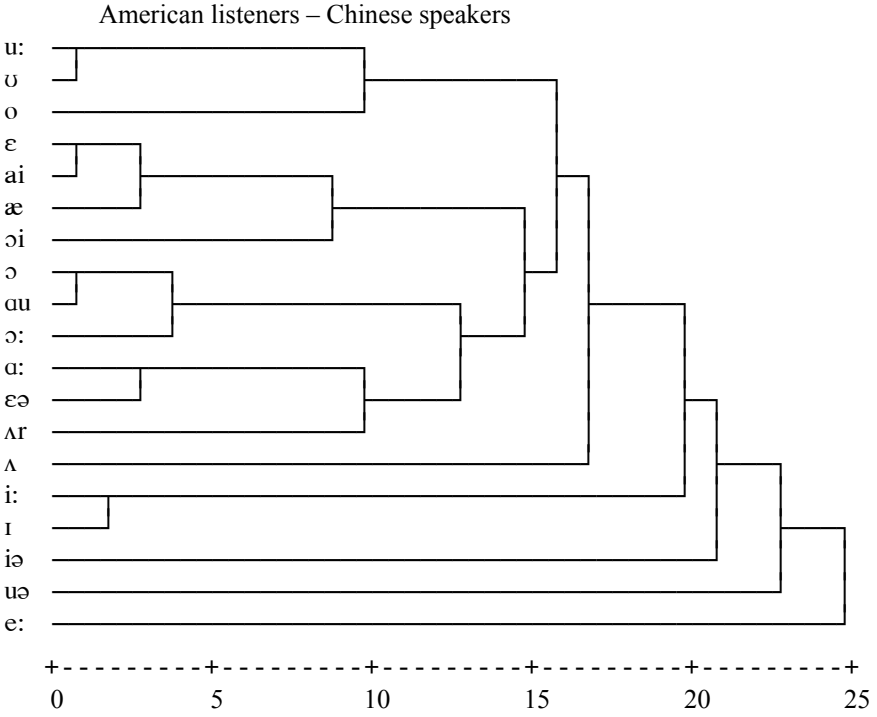


Figure 6.3. Hierarchical cluster scheme (average linkage between groups) for all 19 vowel stimuli in American responses to Chinese-accented English.

One reason why the groups are difficult to interpret phonetically is that both monophthongs and diphthongs are response categories, as are vowels followed by /r/. The vowel groups tend to be more coherent if only monophthongs are included in the trees. For the sake of completeness, all nine dendrograms (average linkage, ten phonological monophthongs only) for vowel confusions are presented in Appendix A6.3, but we will not discuss them in the text. Discussion will take place on the basis of the confusion graphs, which contain more information in a more insightful manner.

6.2.4.3 Design of the confusion graphs

We present confusion structures in the English vowels as produced by American, Chinese and Dutch speakers and as perceived by listeners of the same language backgrounds. Vowels are arranged according to the 4 (height) × 3 (backness) vowel

quality grid, with a finer distinction between tense, lax and r-colored vowels by means of superposed 'concentric' rings. The tense vowels are located on the outer ring. Note that we have placed the vowel /æ/ more or less on the outer ring, indicating that is precisely at the boundary between tense and lax vowels. This arrangement would seem to do justice to the fact that this vowel behaves as a lax vowel in the phonological system of American English (where it is not allowed at the end of a word) and as a tense vowel from a phonetic viewpoint (long duration and extreme vowel quality). In the diagrams tense /ɔ:/ and lax /ɔ/ are kept separate, in order to demonstrate that these vowels are extremely confusable (to the point that they can be considered merged in the sound system of General American). Long /ɑ:/ has been located on the outer ring, even though in our stimulus word it was followed by /r/, because its quality appeared not to be centralized at all. Although we cannot be sure, we assume that this vowel would merge completely with /ɔ:/; the only reason why the listeners have kept the two vowels apart is because /ɑ:/ was followed by /r/ in the word *hard* which made it audibly quite distinct from *hawed*.

Confusion between any two vowels is indicated by an arrow from the intended to the non-intended vowel. The confusion percentage is indicated at the tip of the arrow. Arrows were drawn only for 'problematic' vowel pairs, defined as pairs that were confused in at least 20% of the responses. This is different from what did in the pilot test in which we defined problematic vowels as those vowels that were confused with some other vowel in more than 10% (Wang and Van Heuven, 2004). Since in this final test we had the Chinese subjects in China and the test was done with selected speakers, who were less proficient than those we used in the pilot test, the percentage of correctly identified vowels is lower than in the pilot test. Maintaining the more relaxed inclusion criterion of 10% vowel confusion would have yielded overly complicated and messy structures. That is why we now adopted the 20%-criterion as the definition of problematic vowels.

In the next sections I will present nine confusion graphs, one for each combination of speaker and listener nationality. The first three confusion graphs will contain the structures obtained for Chinese listeners, exposed to Chinese, Dutch and American speakers. Then I will repeat the set of three speaker nationalities for Dutch listeners, and I will conclude with the three sets obtained for American speakers.

In the confusion graphs the results for the stimulus type /ɔi/ has been omitted. This was done, firstly, to avoid visual clutter in the graphs. There seems to be no obvious place in the table where yet another (mid) low diphthong can be accommodated. A second reason why this particular stimulus category could be sacrificed, is that listeners either confuse this diphthong very strongly but in a non-systematic fashion (as happened in the case of Chinese listeners confronted with Chinese speakers) or they have no problems with this vowel at all. The problem of the Chinese speaker/listener group is not limited to this particular vowel but can be generalized to all three diphthongs /ai, ɔi, ɑu/. As a compromise, therefore, we will present the results in all the confusion graphs for the low diphthongs /ai, ɑu/ and decided to omit /ɔi/. For the full set of confusions I refer to Appendix A6.2.

6.2.4.4 Confusion structures of Chinese listeners

When Chinese listeners have to identify the English vowels spoken by fellow **Chinese speakers**, systematic confusion is found for no fewer than 15 stimulus vowels (in fact 16, if /ɔi/ had been included). The most frequent vowel confusions are /ʌr > ɑ:/ by 61%, /i: > ɪ/ by 50% and /ɔ > ɑ:/ by 46%. It shows /ɑ:/ as a big problem, not only because /ɑ:/ itself is confused with /ʌ/ in 37% but even more so because /ɑ:/ functions like a magnet, attracting massive confusions from neighboring vowels. As a result we may set up the larger group of (mid) low back vowels as a highly confusable vowel cluster. Highly problematic is also the front diphthong /ai/. It is strongly misperceived as either /i:/ or /ɪ/. It would appear, then, that the Chinese speakers emphasize the second part of this diphthong (and reduce the onset portion of the diphthong) so that it sounds rather like /i:/ or /ɪ/. Given also that Chinese listeners do not differentiate between the tense and the lax counterparts within this pair renders this a plausible confusion pattern. Interestingly, as we will see below, Dutch and American listeners do not have this problem with the Chinese /ai/. This suggests that the problem resides not so much with the Chinese speakers but with the Chinese listeners.

Although most of the confusion pairs are unidirectional (or 'asymmetrical'), there are some pairs which are confused in both directions. These are /i: ~ ɪ/, /ɛ ~ æ/ and /u: ~ ʊ/. This is what we can predict from the sound systems in the Chinese and English inventories, as the confusions are within spectrally adjacent members of tense~lax oppositions.

When Chinese listeners listen to **Dutch speakers**, they have 10 pairs of confusing vowels with no confusion pairs higher than 46%, which is lower than when they listen to their own speakers. The most frequent confusions are /ɛ ~ æ/ by 46% (the highest) and /ʌ ~ ɑ:/ by 40%. Symmetrical confusion pairs are /ɛ ~ æ/ and /u: ~ ʊ/.

When the **American speakers**' vowels are identified by Chinese listeners, there is much confusion as well. When Chinese listeners respond to American speakers there are asymmetrical confusions only. They confuse /i:/ with /ɪ/ by 50% and /ɔ/ with /ɑ:/ by 46%. The /u: ~ ʊ/ confusion is also unidirectional at 32%. The unidirectional confusion structure for the high vowels may be the result of the fact that Chinese speakers pronounce tense /i:/ shorter than the American and Dutch speakers do. Assuming that Chinese listeners attend to the duration cue rather than to spectral cues in the tense~lax contrast, there will be a bias towards perceiving the lax counterpart in these contrasts.

Chinese listeners have problems in front vowels in both Dutch and American speakers and their own speakers as well. Back vowels are difficult for Chinese listeners. The vowel /ɑ:/ remains a problem for all three groups of speakers.

Note, finally, that there are virtually no confusions that cross the boundary between front vowels and back vowels. That is to say, when a front vowel is confused, it is always with some other front vowel, and back vowels are confused with back vowels only. This also means that confusions take place mainly along the dimensions of vowel height and tenseness.

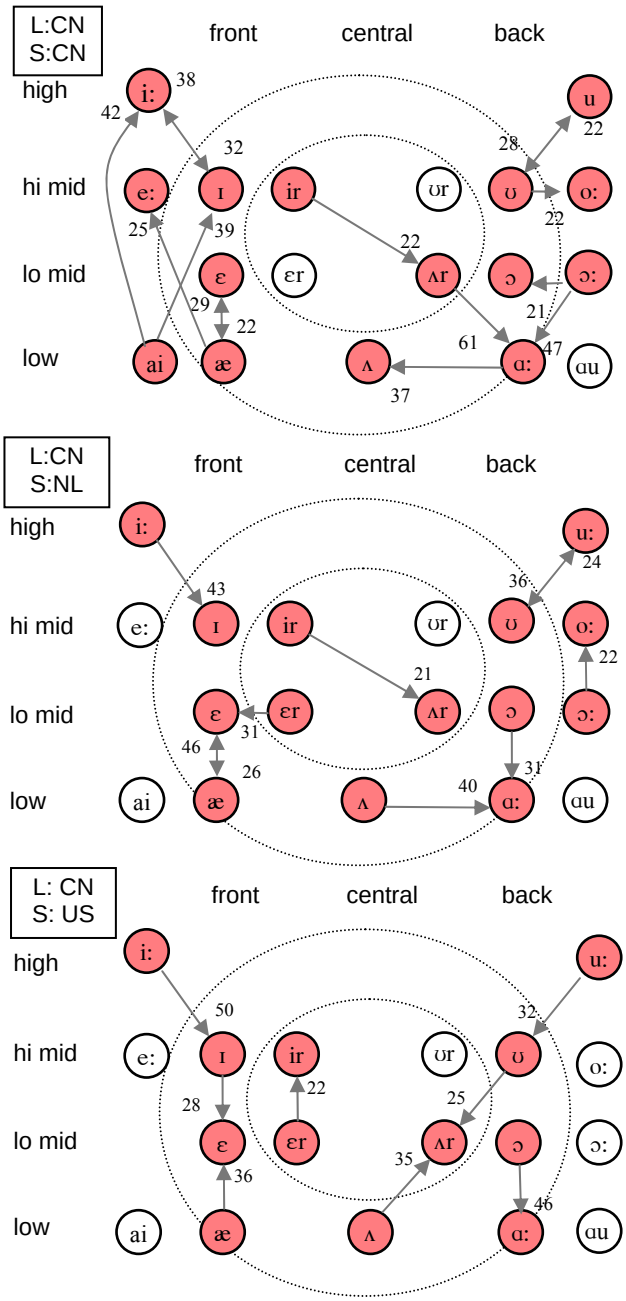


Figure 6.4. Confusion graphs for Chinese listeners, exposed to Chinese (CN), Dutch (NL) and American (US) speakers (from top to bottom). Only confusions ≥ 20% are indicated by arrows. L = listeners, S = speakers.

6.2.4.5 Confusion structures of Dutch listeners

Figure 6.5 presents the confusion structure in the vowels heard by Dutch listeners. When the speakers are **Chinese**, massive vowel confusion is observed. Lax /ɪ/ is strongly confused with tense /i:/; there is bidirectional confusion between and /u:~ ʊ/. The open front vowel /æ/ is unidirectionally confused with /ɛ/, and also with the diphthong /ai/, as is /ɛ/. In the ear of the Dutch non-native listener these three vowels spoken by Chinese learners are very poorly distinguished. Also Chinese /ʌ/ is unidirectionally mistaken for /æ/. Finally, the four low back vowels are strongly confused – and therefore poorly differentiated in the ear of the Dutch listeners. There is one confusion across front and back vowels: /ɔ:/ > /ai/.

When Dutch listeners respond to Dutch speakers of English the confusions are restricted to just four pairs, which seem rather predictable from a contrastive analysis of the Dutch and English vowel systems; these are the pairs /ɛ ~ æ/ and /u: ~ ʊ/. In addition to these there is unidirectional confusion of /ɔ/ to /o/ and of central /ʌ/ to /æ/. This latter confusion is unexpected but does in fact mirror the same confusion when the speakers are Chinese. It would indicate that both the Chinese and the Dutch speakers realize some of the /ʌ/ tokens rather too front and too low.

When the speakers are **Americans** the number of confusions is minimal. Dutch listeners confuse both open /æ/ and the long half-closed vowel /e:/ for /ɛ/, the first because they have no clear category boundary between /æ/ and /ɛ/, the latter possibly because the American speakers make their tense /e:/ too short for the Dutch norm of a tense vowel, or because the American onset of /e:/ is lower than what is expected by a Dutch listener. American /ʌ/ is confused with lax /ʊ/. This would indicate that Americans pronounce /ʌ/ more back than the Dutch and Chinese speakers do, and that the Dutch target of this vowel has a more front and low position. Long /u:/ is unidirectionally confused with lax /ʊ/, and, finally, /ɔ:/ in *hawed* is strongly confused for /ɑ:/ as in *hard*. This is quite likely due to the fact that American speakers tend to neutralize the contrast between these vowels.

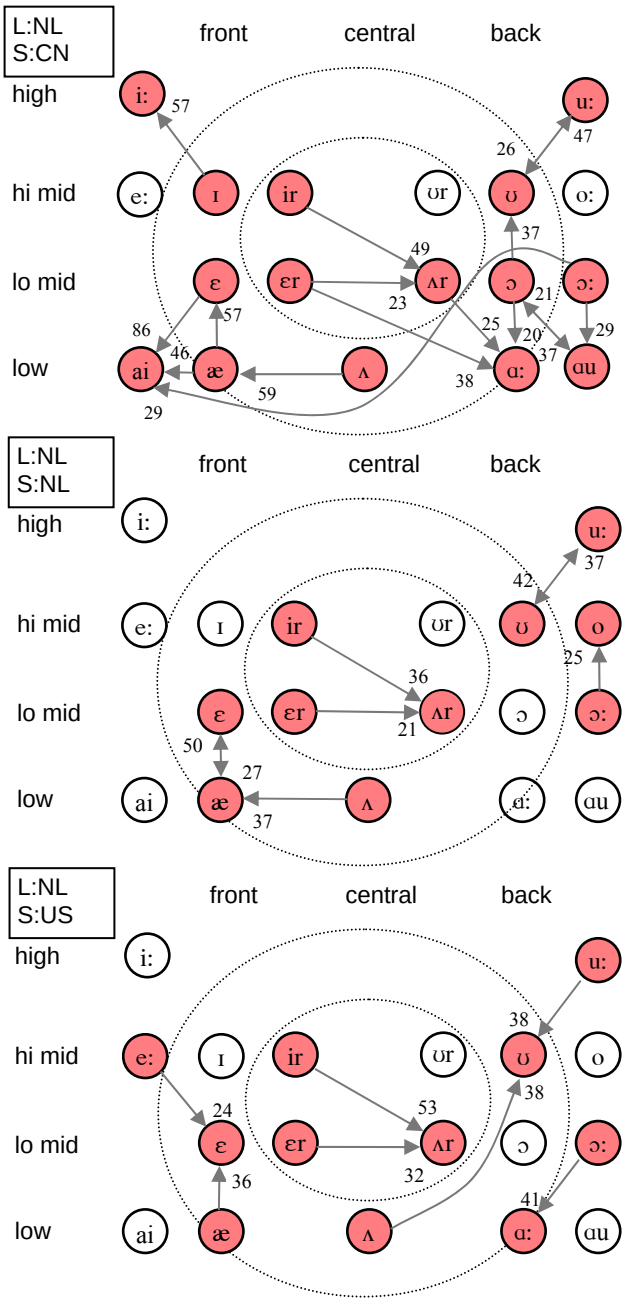


Figure 6.5. Confusion graphs for Dutch listeners, exposed to Chinese (CN), Dutch (NL) and American (US) speakers (from top to bottom). Only confusions ≥ 20% are indicated by arrows. L = listeners, S = speakers.

6.2.4.6 Confusion structures of American listeners

When American listeners listen to **American speakers** there is relatively little confusion in the vowels. The literature shows that even in such situations vowel identification is far from perfect, with scores ranging between 54 and 88 % correct (Peterson and Barney, 1952; Strange, Verbrugge, Shankweiler and Edman, 1976 and references therein). Our results are no exception since the identification performance ranges between 38 and 96%, depending on the vowel type. Still, only two vowel pairs were confused in more than 20%. These are the confusion of lax /ʌ/ > /ʊ/ and the partial merger of tense and lax /ɔ/.[3] The latter confusion was also the most frequent one in the classical study by Peterson and Barney (1952) but the /ʌ/ > /ʊ/ confusion, although it did occur in the classical data, only ranked third there (after /ʌ/ > /ɔ:/ and after /ʌ/ > /ɔ/). It would seem, therefore, that our American speakers pronounced their /ʌ/ somewhat differently than the Peterson and Barney speakers did. The most confusing pairs for both Chinese and Dutch listeners, /ɛ ~ æ/ and /u:~ ʊ/ (see below), do not constitute any problem for American listeners when they listen to their own speakers, so that, clearly, the Dutch and Chinese speakers fail to observe a contrast here.

The vowels produced by the **Dutch speakers** show some confusions that are also mentioned in the literature, i.e. /æ/ > /ɛ/ and /ʊ/ > /u:/ (see Table 3.5). These confusions are unidirectional: clearly the Dutch way of pronouncing /æ/ is not open enough, hence the unidirectional confusion with /ɛ/. The data also suggest that Dutch /u/ resembles American tense /u:/ more than its lax counterpart. There is considerable confusion of /ʌ/ > /ɔ/. Dutch-accented /ʌ/ tends to be too far back, causing unidirectional confusion with /ɔ/. This confusion was foreshadowed in Table 3.5, where the observation was made that (advanced) Dutch learners often substitute their /ɑ/ for /ʌ/. Not expected from Table 3.5 would be the remaining confusion, i.e. /ɔ:/ > /o:/ This confusion would indicate that the Dutch speakers tend to make the English /ɔ:/ too close.

When Americans listen to **Chinese speakers** there is confusion of height and tenseness in the high-front as well as in the high-back vowels. The American listeners have problems in identifying high and low front vowels and high and low back vowels with 15 confusion pairs (with 69% confusion for the most problematic vowel pair). There are four bidirectional pairs /i: ~ ɪ/, /u: ~ ʊ/, /i ~ æ/, and /ɔ: ~ au/. The former two confusions were also listed as pronunciation problems for Chinese learners of English in Table 3.6; the latter two confusions have not been noted in the pedagogical literature.

This configuration is also isomorphic to the pattern found for Dutch listeners exposed to Chinese speakers. This conformity shows that the source of the problem resides in the pronunciation of the Chinese speakers rather than in the perception of the Dutch listeners.

---

[3] See the following quote from Peterson and Barney (1952: 178): "The very low scores on [ɔ:] and [ɔ] … undoubtedly result primarily from the fact that some members of the speaking group and many members of the listening group speak one of the forms of American dialects in which [ɔ:] and [ɔ] are not differentiated ."

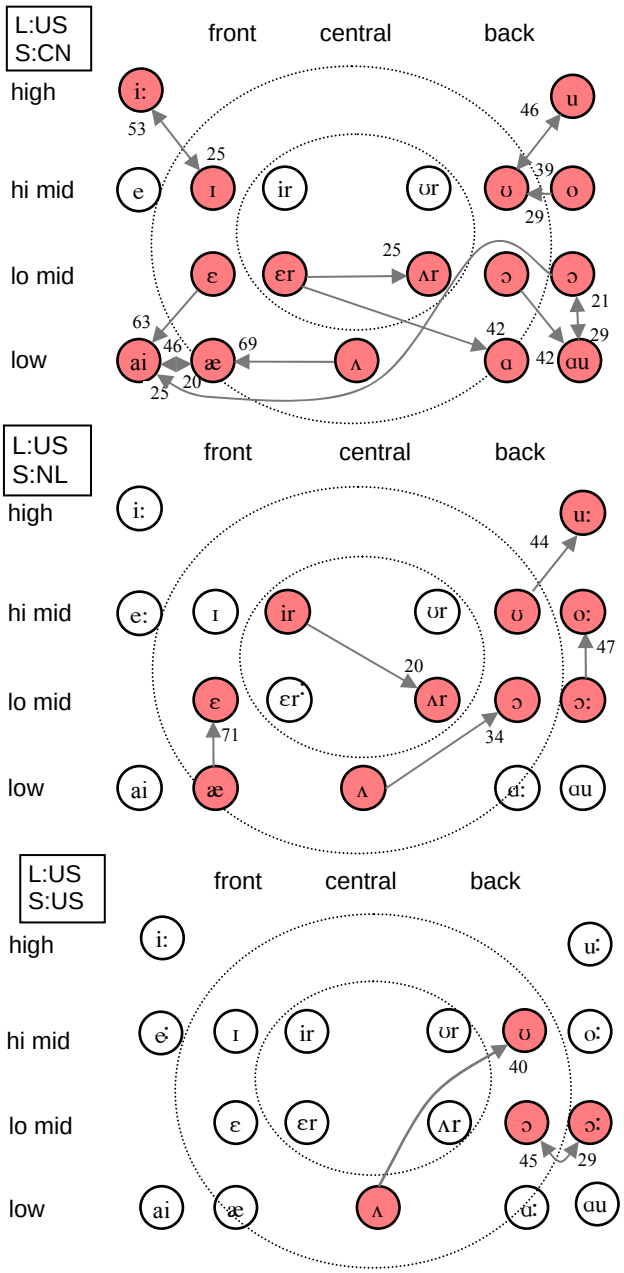


Figure 6.6. Confusion graphs for American listeners, exposed to Chinese (CN), Dutch (NL) and American (US) speakers (from top to bottom). Only confusions ≥ 20% are indicated by arrows. L= listeners, S = speakers.

6.3 Summary

By way of summary Table 6.5 lists the numbers of problematic vowels in the data. Here a problematic vowel is more or less arbitrarily defined as a vowel which in any speaker-hearer combination is identified correctly in less than 75%. The numbers are broken down for the nine combinations of speaker and listener language background.

Table 6.5. Number of problematic vowels (see text) broken down by nationality of speaker and of listener.

speaker	listener			
	Chinese	Dutch	USA	Total
Chinese	**19**	15	17	51
Dutch	19	**13**	11	43
USA	19	12	**4**	35
Total number	57	40	32	139

Table 6.4 shows that, overall, American native listeners have fewer problems with the English vowels than L2 listeners. Dutch listeners are a good second, and Chinese listeners clearly have problems. More generally, the language background of the listener exerts a stronger influence on the number of confused vowel pairs than the L1 of the speaker.

6.4 Conclusion and discussion

Our first hypothesis was that English vowels will be more difficult to identify as the foreign speaker's native language is more unlike English by the interference of their L1s. We predict, then, that Chinese-accented English vowels will be more difficult to identify for native English listeners than, for instance, Dutch-accented vowels. Conversely, English vowels produced by native English speakers should then be more intelligible to Dutch listeners than to Chinese listeners. Both predictions were clearly borne out by the experimental results. Although these results can indeed be seen as experimental support for our typological distance hypothesis, it should be pointed out that cultural and educational differences between the People's Republic of China (with little exposure to English) and the Netherlands (with an abundance of English) may also have contributed to the difference in intelligibility.

The confusion structure in the foreign-accented Englishes can partly be accounted for by a contrastive analysis of the vowel inventories of the target and source languages involved. For Dutch-accented English, we predicted problems with the non-high lax front vowels /ɪ ~ e ~ æ/ and with the /uː ~ ʊ/ contrast. The results show that these were, indeed, the most frequent confusion types, not only when L1 English listeners identified Dutch-accented vowels, but also when Dutch L2

listeners identified native English vowel tokens. Moreover, our contrastive analysis predicted that Chinese-accented English would have all the problems of Dutch English but would additionally suffer from massive tense~lax vowel confusion, both in production and in perception. The experimental results show that this prediction is correct.

On the other hand, we found a number of problematic vowel contrasts that are not easily predicted from a contrastive analysis, e.g. the /uː > oː/ and /ɔː > oː/ confusions for Dutch speakers identified by American listeners. We did not encounter any cases where predicted problems did not arise. Our results, then, provide partial support for the transfer hypothesis in foreign language learning, which claims that L2 learners will not distinguish between contrasts in the target language that do not occur in their native tongue. At the same time, a weaker version of the transfer hypothesis seems called for, in that, although it makes no false predictions, it predicts only a subset of the L2 vowel learning problems.

Many of the confusions found in this chapter were mentioned as learning problems in the pedagogic literature on the learning of English as a second language for Dutch and Chinese learners, in Tables 3.5 and 3.6, respectively. However, we noted some pronunciation problems that were not mentioned in the tables, indicating that occasionally pronunciation problems escape the trained ears of foreign-language teachers. I would claim that such problems can only be brought to light by experimental methods such as those used in the present study.

Chapter seven

Intelligibility of intervocalic consonants

7.1 Introduction

In this chapter I will present the results for intelligibility of intervocalic consonants for three groups of listeners. The results are from the same groups of listeners as in the previous chapter. As we analyze the sound system of the consonants in the three languages, we predict consonants will be more difficult for Chinese than for Dutch English L2 learners. The results we are going to present on the one hand will represent the actual intelligibility of consonants for Chinese and Dutch listeners, which may partially support the predictions in Chapter three derived from models of L2 perception, and on the other hand, the results may raise new questions which cannot be explained from by these theories.

7.2 Results

7.2.1 Overall results

The overall results for consonant intelligibility are presented in Figure 7.1, broken down by nationality of the listeners and broken down further by nationality of speakers. As before (§ 6.2.1), the data were submitted to an Analysis of Variance (ANOVA) run on the mean percent correct scores for each listener with nationality of speaker and nationality of listener as fixed factors.

Across speaker groups, the Chinese listeners have the lowest consonant identification scores (47 to 58% correct, mean = 54%). Dutch listeners perform at an intermediate level (67 to 81% correct, mean = 73%), and the American listeners are the best (71 to 83% correct, mean = 78%). The effect of listener group was significant, $F(2, 315) = 186.7$ ($p < .001$). A post-hoc test (Scheffé procedure with $\alpha = .05$) indicates that all three speaker nationalities were different from each other.

Across listener groups, Chinese and Dutch speakers obtained the lowest vowel identification scores (65%). The American speakers' vowels were correctly identified in 75 percent of the cases. The effect of speaker nationality is significant, $F(2, 315) = 35.8$ ($p < .001$). The American speakers are significantly better than the other two nationalities, which do not differ from each other. As before, we may note that the effect of listener nationality is much larger, in fact more than five times larger in the present case, than the effect of speaker nationality. Again, the interaction

between listener and speaker nationality also reaches significance, $F(4, 315) = 8.3$ ($p < .001$), indicating interlanguage or native language benefit (see Chapter six).

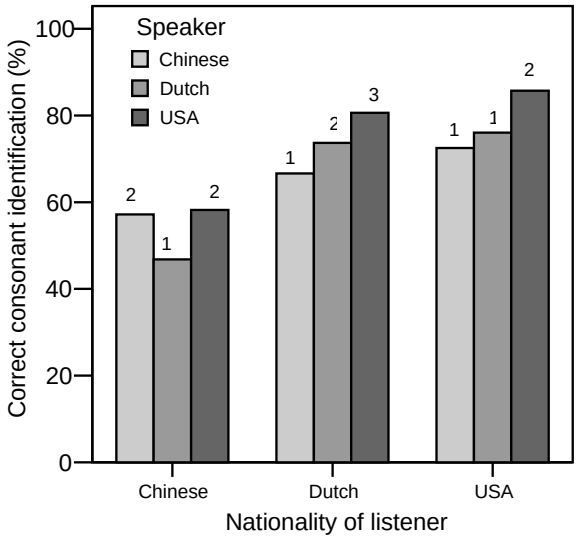


Figure 7.1. Percent correctly identified consonants for Chinese, Dutch and American listeners broken down by accent of speakers. Numbers above the bars indicate the subgroup membership as determined by the Scheffé procedure. Numerical values of means, N, SD and Se are included in Appendix A7.1.

Figure 7.1 shows overall correct consonant identification. It does not allow us to identify individual consonants that represent special difficulties. Therefore, we ask, first of all, which are the problematic consonants for each group of listeners? This question will be taken up in the following section (§ 7.2.3). Secondly, if a sound is massively misidentified, then what is it heard as instead? This question will be dealt with later when we examine the confusion structure in the consonant data (§ 7.2.4).

7.2.2 Correct consonant identification

In order to get an overview of which consonants are more difficult than others, for each combination of speaker and listener nationality, we present percentages of consonants correctly identified by Chinese, Dutch and American listeners in separate panels for Figure 7.2. In each panel the results have been broken down by nationality of the speakers. In each panel the 24 consonants have been ordered in descending order of correct identification when the speakers are American. Generally we would expect the results for the non-native speakers, i.e. by Chinese and Dutch speakers, to fall below the percent correct of the American speakers. Only in exceptional cases do we expect the non-native vowels to be identified better than the native vowels.

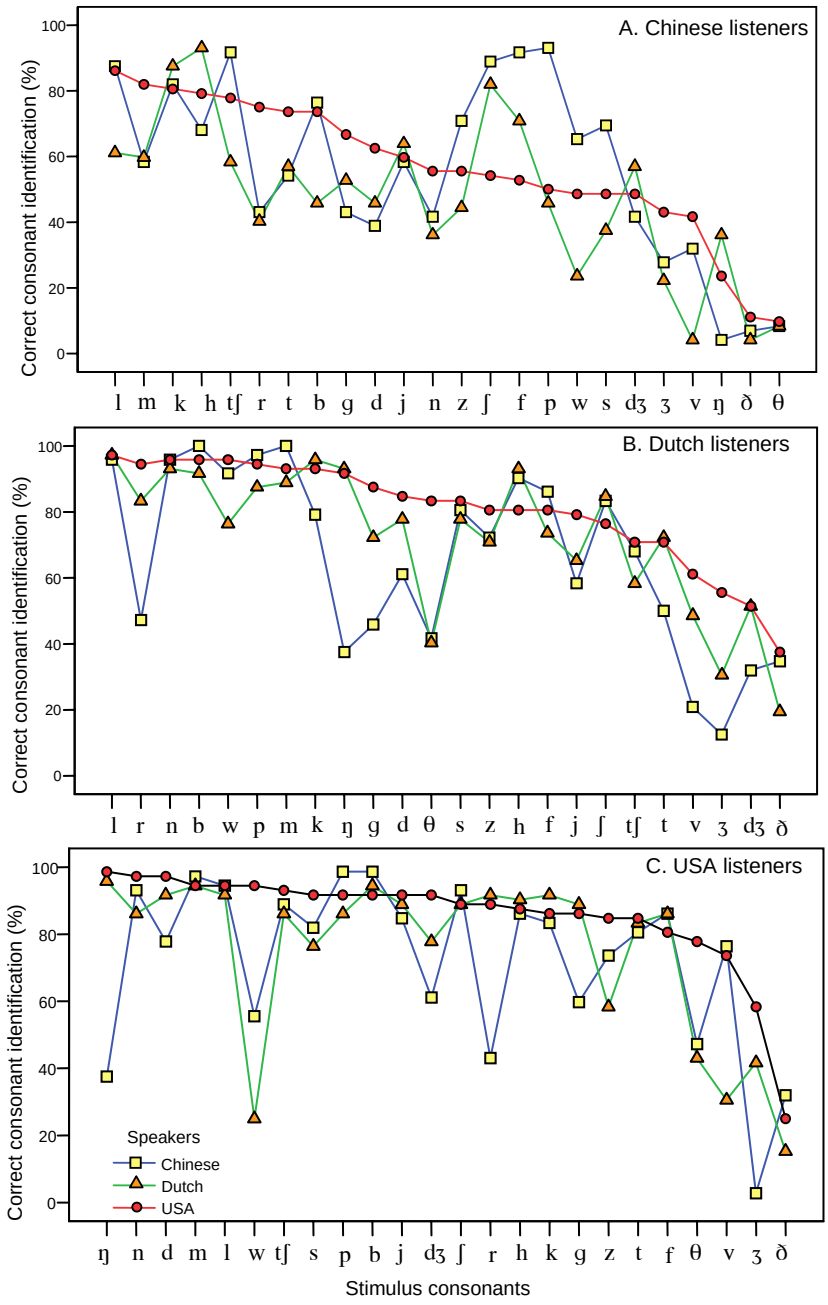


Figure 7.2. Correct identification (%) for all 24 single English consonants produced by Chinese, Dutch and American speakers. Panels A, B and C present the results for Chinese, Dutch and American listeners, respectively.

Again taking the American speakers as the norm, we may observe that there is a wide range of correct consonant identification scores with over 80% correct for /l/ and /m/ going down to less than 20% correct for /θ/ and /ð/. In contradistinction to the vowel data (Chapter 6), there is an overall trend for consonant identification scores to run parallel regardless of the language background of the speakers. As a result, correlation coefficients for correct consonant identification scores for pairs of speaker nationalities are all significant (see Table 7.1).

Table 7.1. Pearson's correlation coefficients for identification of consonants produced by Chinese, Dutch and American speakers broken down by nationality of the listeners.

Listener nationality	Speaker nationalities		
	CN ~ NL	CN ~ US	NL ~ US
CN	r = 0.679**	r = 0.667**	r = 0.711**
NL	r = 0.744**	r = 0.680**	r = 0.848**
US	r = 0.564**	r = 0.565**	r = 0.704**

**: $p < 0.01$

Figure 7.2-A shows the Chinese listeners' identification of the 24 consonants of Chinese, Dutch and American speakers. The correct identification rate for American speakers runs from more than 80% down to almost l0%. It is not the case that the American speakers' consonant tokens are more intelligible than the non-native tokens. Seven Chinese-accented consonants are clearly better identified by Chinese listeners; although these seven do not form a natural class, the set would appear to comprise labials and fricatives. These, of course are the types of consonants that also occur in the Chinese consonant inventory. Curiously enough, there are also a number of Dutch-accented consonants that are clearly better identified than the native American tokens, viz. /h, ʃ, f/. This is hard to explain, given that Dutch /h/ is often voiced, whilst Chinese listeners would expect /h/ to be voiceless (as it should be both in Chinese and in English), and /ʃ/ is not a phoneme of Dutch at all.

Figure 7.2-B shows the Dutch listeners' identification of the 24 consonants of Chinese, Dutch and American speakers. The correctness of American consonant tokens covers a range from 96% to 38%. In this figure we can see that the American speakers' tokens almost invariably obtain all the highest identification scores, with no significant exceptions. When Dutch listeners listen to their fellow speakers, there are just a few consonants that are identified clearly more poorly than the American counterparts, viz. /w, θ, ʒ, ð/. The latter three are not phonemes of Dutch, so that their difficulty can be explained as cases of negative transfer. The high incidence of /w/ errors may be due to the incorrect labio-dental articulation of this glide, so that it gets confused with /v/ which would not happen in the case of either the Chinese or American tokens of /w/, which would be bilabial. We will consider this explanation later on. Chinese-accented consonants are obviously the most difficult tokens for Dutch listeners. A very substantial loss of consonant identification is incurred for the Chinese-accented consonants /v, ʒ, dʒ, r, ŋ, g/. These are the voiced fricatives, the

voiced nasals and the /r/. Voiced fricatives are absent in the onset inventories of both Chinese and Dutch, which would account for their low identification rates.[1] Chinese /r/ is typically pronounced as a (voiced) fricative, so that it will be confused with /ʒ/.

Figure 7.2-C shows American listeners' identification of the 24 consonants of Chinese, Dutch and their own speakers. The percentage is presented in the order of correctness from high to low of every consonant produced and identified by American native speakers. American listeners have the highest identification for the velar nasal /ŋ/ and dental nasal /n/ (99% correct) produced by their own speakers and the lowest identification for the voiced palatal fricative /ʒ/ (59%) and for the voiced dental fricative /ð/ (25%). This indicates that native American listeners have problems with their own speakers for certain consonants. Things are roughly the same when American listeners respond to Dutch speakers with the exception that the bilabial approximant /w/ is now very poorly identified (25% correct). As mentioned before, bilabial /w/ is a new sound for the Dutch learners of English; its Dutch counterpart is a labio-dental approximant /ʋ/, which sounds very much like the English voiced fricative /v/. We will take this matter up below, when we discuss the confusion structure within the consonant set. However, when the American listeners react to Chinese speakers, they seem to have more difficulties in identifying the Chinese-accented English production. Some of the sounds which are no problem when they are produced by American themselves or by Dutch speakers are problems when they are produced by Chinese: /ʤ, w, j, r, g, θ/. Possible reasons for the poor identification of these sounds will be discussed when we review the confusion patterns below.

7.2.3 Consonant confusion structure

Full 24 × 24 consonant confusion tables were generated for all nine combinations of speaker and listener nationalities. These have been included in Appendix A7.1, together with hierarchical cluster schemes (dendrograms, Appendix A7.2) computed according to the method of average between-group linkage. In these raw materials it is rather difficult to observe clear confusion structures as relatively few consonants cluster. Just as we did in Chapter six with the vowels, we will therefore present and analyse the confusion structure in the consonants by means of confusion graphs. In these graphs the consonants have been arranged roughly by manner (plosive, fricative, semivowel, liquid, nasal from left to right along the horizontal dimension) and by place (labial, dental, alveolar, palatal, velar from top to bottom). In order not to overly complicate the graphs, the affricates /ʤ/ and /tʃ/ have been entered as plosives with a palatal place of articulation. Within the set of obstruents there is a further split between voiced and voiceless counterparts; these are listed side-by-side nested under manner.

[1] Voiced fricatives are completely absent from Chinese phonology. In Dutch they have to be assumed to be present at the abstract phonemic level but the voiced ~ voiceless distinction is neutralised (to voiceless) in the word onset in most varieties of Dutch (e.g. Van de Velde, 1996; Slis and Van Heugten, 1989).

In the confusion graphs arcs have been drawn linking confused consonants. The arrowheads point to the target of the confusion; as before, the number printed at the tip of the arrow indicates the percentage of the cases in which the source sound was confused with the target. In order to be able to clearly identify obvious clusters of confusable consonants, only confusion pairs with a relative frequency ≥ 20% have been identified. Such clusters are indicated by a darker grey shade.

I will now present nine confusion graphs, one for each combination of speaker and listener nationality. The first three confusion graphs (Figure 7.3A-C) will contain the structures obtained for Chinese listeners, exposed to Chinese, Dutch and American speakers. Then I will repeat the set of three speaker nationalities for Dutch listeners (Figure 7.4A-C), and I will conclude with the three sets obtained for the American speakers (Figure 7.5A-C).

7.2.3.1 Confusion structures of Chinese listeners

When Chinese listeners identify the English consonants spoken by fellow **Chinese speakers**, there are 14 pairs of confusions across place and manner. The most frequent confusions are /v > w/ (40%), /d > b/ (43%) and /θ >s/ (43%).

When Chinese listeners listen to **Dutch speakers**, there are fewer confusion pairs but the confusion rates in individual consonant pairs tend to be higher than for the corresponding Chinese-accented tokens. Chinese listeners have six pairs of confusion consonants when responding to Dutch speakers but two of these pairs have higher confusion rates: 50% and 47%. When Chinese listeners identify consonant tokens produced by fellow Chinese speakers, confusion tends to occur across place of articulation rather than across manner. When responding to Dutch speakers of English, all confusion occurs across manner, with just one exception for the pair /θ > f/, which is across place (from dental to labial fricative).

When Chinese listeners respond to **American speakers**, there are eight pairs of confusions, with 35% as the highest percentage. The consonants /v/ and /w/ are strongly and symmetrically confused, also when the speakers are Chinese. In spite of what the figure seems to suggest, Chinese-accented /ŋ/ is very poorly identified irrespective of the language background of the listeners (3, 38 and 38% correct for Chinese, Dutch and American listeners, respectively, see Appendix A7.1). However, confusions are widely scattered for the Chinese listeners (no confusion ≥ 20%) but are somewhat more systematic for Dutch and American listeners.

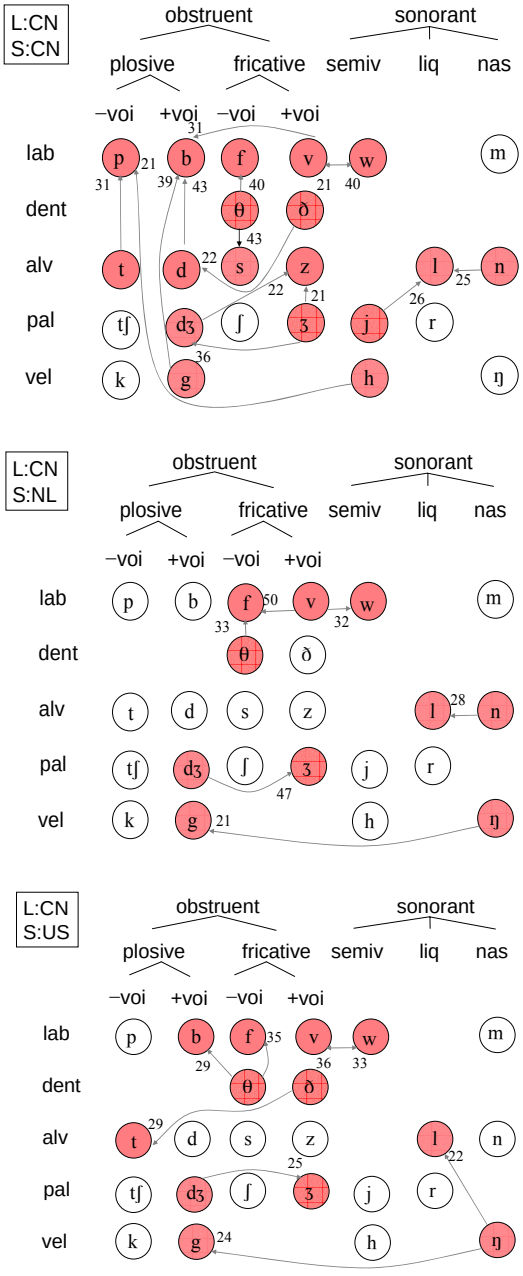


Figure 7.3A-C. Confusion graphs for Chinese listeners, exposed to Chinese (CN), Dutch (NL) and American (US) speakers (from top to bottom). Only confusions ≥ 20% are indicated by arrows. L = listeners, S = speakers.

7.2.3.2 Confusion structures of Dutch listeners

Figure 7.4 presents the confusion structure in the consonants heard by Dutch listeners. When the **speakers are Chinese**, massive consonant confusion is observed. The voiced fricative /v/ is strongly confused with the labial semivowel /w/ (71%). In the ear of the Dutch non-native listeners the velar nasal /ŋ/ remains a problem when it is spoken by Chinese learners. It is insufficiently distinguished from /g/. The pairs /j > dʒ/ and /ʒ > dʒ/ are confused by manner but place is preserved; /r > z/, /t > θ > s/ and / ð > d/ are manner confusions but both place and voicing are preserved. We observed before that Chinese /r/ is pronounced as a voiced fricative. This is also relevant here since Dutch listeners confuse it with /z/. Interestingly, Chinese speakers poorly distinguish the three palatal consonants: both /j/ and /ʒ/ are confused with /dʒ/. The palatal fricative /ʒ/ does not occur in Chinese at all but /dʒ/ does; this would account for the confusion as a speaker error induced by negative transfer. The confusion involving the palatal approximant /j/ is more difficult to explain. Phonetically, /j/ does occur at the beginning of syllables. In the conception of the Chinese language user, however, [j] should be parsed as belonging to the vocalic nucleus rather than to the onset; therefore, Chinese words (or syllables) beginning with [j] would have to be preceded by an empty onset, i.e. a glottal stop to fill the empty onset. If the habit of inserting a stop-like feature before /j/ carries over into English, then we would expect a more stop-like percept, which is compatible with perceived /dʒ/.

When Dutch listeners respond to **Dutch speakers** of English, the confusions are restricted to just four clusters. The first pair is /v > f/, which is predictable given that Dutch initial /v/ typically loses its voicing. Next, the voiced dental fricative /ð/ is either pronounced without voicing and is heard as /θ/ or with the wrong manner and is heard as /d/. Similarly, the palatal fricative /ʒ/ either loses its voicing and is confused with /ʃ/ or it is weakened to an approximant and shows up as /j/. The approximant /j/, in turn, seems to get strengthened as is often is misperceived as the affricate /dʒ/. This also happened with Chinese-accented tokens of /j/. This time, however, no explanation of the confusion seems possible from the phonology of the source language.

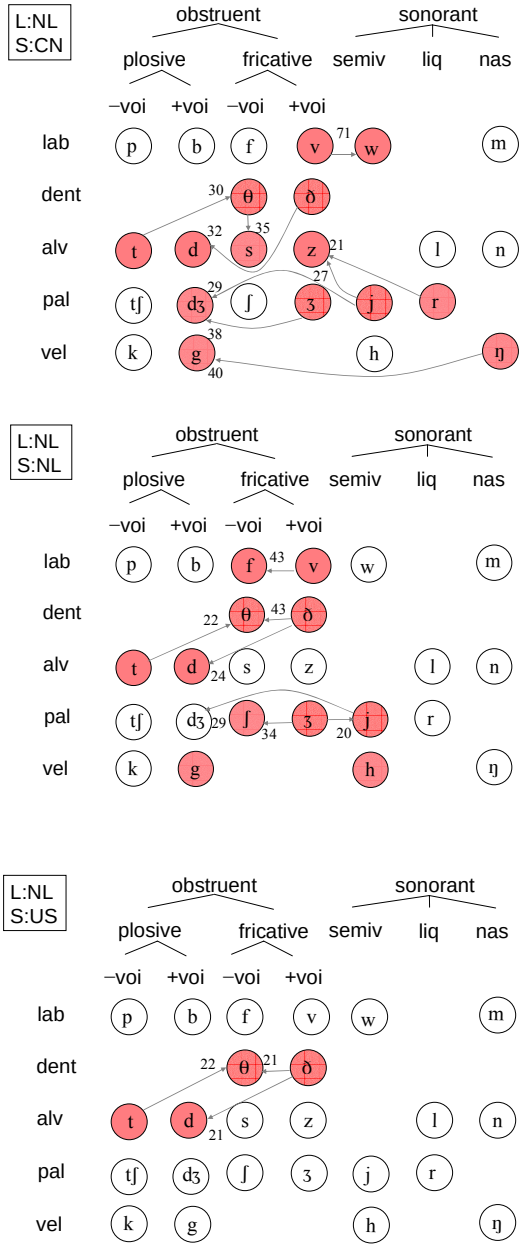


Figure 7.4A-C. Confusion graphs for Dutch listeners, exposed to Chinese CN), Dutch (NL) and American (US) speakers of English (from top to bottom). Only confusions ≥ 20% are indicated by arrows. L = listeners, S = speakers.

When the **speakers are American,** the number of confusions is minimal. No more than three confusion pairs occur. Note that the confusions found here are a proper subset of the confusions we found when the speakers were Dutch. Again, /ð/ is misheard as either /θ/ or /d/, and /t/ goes to /θ/. This would seem to suggest that the confusion is at least partly due to perceptual uncertainty on the part of the Dutch listeners.

7.2.3.3 Confusion structures of American listeners

When American listeners respond to **Chinese speakers**, confusions are limited to consonant pairs within the set of voiced plosives (/d, dʒ, g/) and voiced fricatives (/v, ð, ʒ/). There is no systematic confusion along the voicing dimension. Chinese (unlike Dutch) has clearly voiced affricates, and the tense~lax contrast in stops uses the same phonetic parameters as in English, viz. aspiration for the tense stops (positive VOT) and absence of prevoicing in the lax counterparts (0 VOT). These findings largely reflect the observations made by Zhao (1995), which were summarized in Table 3.10. However, counter to Zhao we find no indications that /p, t, k/ are pronounced with insufficient aspiration. The systematic confusion of /r/ with /ʒ/ indicates that Chinese /r/ is pronounced as a fricative. This confusion was also mentioned by Zhao (see Table 3.10). The strongest confusions are /ʒ/ > /dʒ/ (75%) and /ð/ > /d/ (40%). These errors were also observed by Zhao. Neither /ʒ/ nor /ð/ occur in Chinese; it seems that these targets are realized with stop-like characteristics. On the basis of this, one would expect the third voiced fricative that is absent from Chinese, i.e. /v/, to be systematically confused with its stop counterpart /b/. However, Chinese-accented /v/ is primarily confused with /w/ (33%); the predicted confusion with /b/ is the second-most frequent confusion (10%). As was observed in Chapter three, Chinese has no voiced fricatives but does use voiced affricates. The systematic confusion of stop/affricate manner for voiced fricatives may then be accounted for as negative transfer from the source language. We have no clear explanation, finally, for the confusion of the velar nasal /ŋ/ with its oral counterpart /g/. The problematic nature of onset /ŋ/ was noted by Zhao, but she never explicitly stated what confusion would arise. The problem may have its origin in the use of /ŋ/ in onset position. This is not impossible for Dutch and American speakers as /ŋ/ may surface intervocalically in the onset after lax vowels (as in *si<u>ng</u>ing*, *lo<u>ng</u>ing*, *ha<u>ng</u>ing*). In the sound system of Chinese, however, /ŋ/ is strictly limited to the coda position; possibly, when a Chinese learner is forced to pronounce /ŋ/ in onset position, there is a tendency to substitute the most similar sound that is allowed in the onset, which would be /g/.

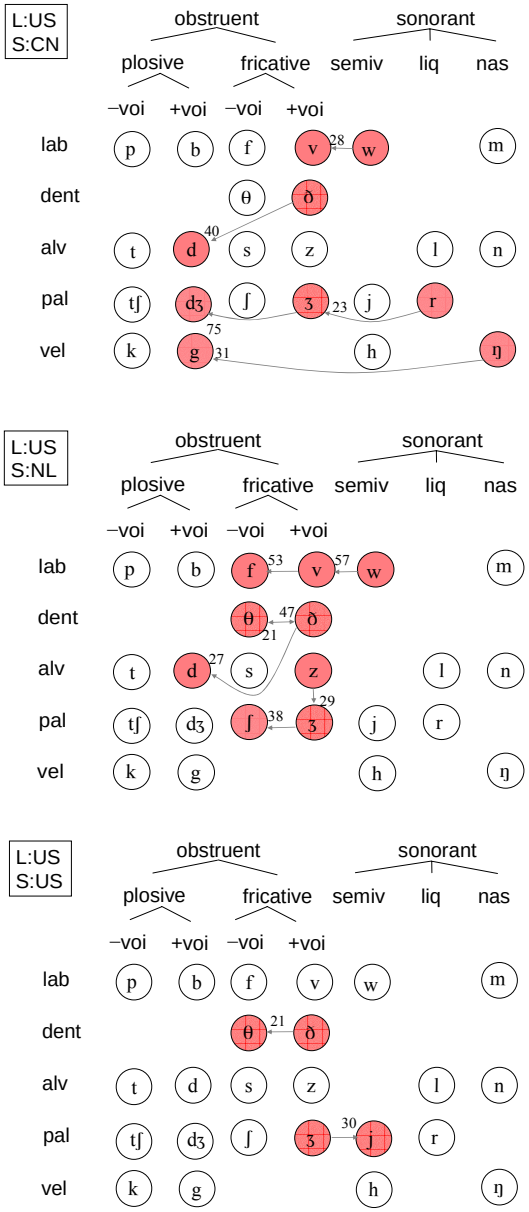


Figure 7.5A-C. Confusion graphs for American listeners, exposed to Chinese (CN), Dutch (NL) and American (US) speakers (from top to bottom). Only confusions ≥ 20% are indicated by arrows. L = listeners, S = speakers.

When American listeners respond to **Dutch speakers**, consonants are confused within three groups only, viz. /f, v, w/, /θ, ð, d/ and /z, ʃ, ʒ/. American listeners confuse Dutch speakers' /w/ with /v/ (57%) and /v/ with /f/ (53%). These confusions can be accounted for as negative transfer. Dutch /w/ is a labio-dental and therefore resembles English /v/ rather closely. Also, Dutch /v/ is very often devoiced and therefore identical to /f/. The dental fricatives /θ/ and /ð/ are confused symmetrically and also with the alveolar plosive /d/. Dutch has no dental fricatives and the voicing contrast is often lost. Dutch speakers of English have a tendency to replace /ð/ by its stop counterpart. In the last confusion cluster, the voiced alveolar fricative /z/ is confused with the palatal fricatives /ʒ/ and /ʃ/. It has been observed before that the Dutch alveolar fricatives lack the characteristic high-frequency noise components of English /s/ and /z/.[2]

The above results largely follow the observations found in the pedagogical literature on the pronunciation problems of Dutch learners of English, which were summarized in Table 3.9.

The last confusion graph shows the errors American listeners make when exposed to fellow **American speakers**. There are only two pairs of confusions: /ð/ > /θ/ (21%) and /ʒ/ > /j/ (30%). No other consonants are systematically confused, if at all.

7.3 Summary

Table 7.2 lists the number of problematic consonants in the data. A problematic consonant is defined as a consonant which in any speaker-hearer combination is identified correctly in less than 75%. The numbers are broken down for the nine combinations of speaker and listener language background.

Table 7.2. Number of problematic consonants (see text) broken down by nationality of speaker (down) and of listener (across).

speaker	listener			
	Chinese	Dutch	USA	Total
Chinese	**18**	13	8	39
Dutch	21	**11**	6	38
USA	18	6	**2**	26
Total number	57	30	16	103

[2] Flege (1984) lists English /s/ as a 'similar sound' for Dutch learners, indicating that the difference between Dutch and English /s/ escapes the Dutch listener but contributes to the perception of foreign accent by native English listeners. Pre-palatal fricatives /ʒ/ and /ʃ/ do not occur in the phonology of Dutch (they only occur in loanwords or surface as a result of coalescence of either /s/ of /z/ with /j/), which may be a reason that Dutch /s/ and /z/ are realized with less emphasis on the high-frequency components: there is no risk of confusion with /ʒ/ and /ʃ/.

Table 7.2 shows that, overall, native American listeners have fewer problems with the English consonants than L2 listeners. Dutch listeners are a good second, and Chinese listeners clearly have problems. More generally, the language background of the listener exerts a stronger influence on the number of problematic consonants than the L1 of the speaker. This matter will be discuss at greater length in Chapter ten.

7.4 Conclusions and discussion

We hypothesized that English consonants would be more difficult to identify as the sound system of the L2 speaker's native language deviates more from English. The differences between the Dutch and Chinese consonant inventories are relatively small, and both languages have roughly the same number of consonants that would be reasonable substitutes for English targets. In this respect the prediction is rather different than in the case of the vowel systems. The results show two things. First, Chinese and Dutch accented consonants are relatively well identified by all groups of listeners, and certainly better than the vowels (Chapter six). Moreover, the difference in intelligibility between Chinese and Dutch accented consonants is very small, which would seem in line with the above hypothesis.

In spite of the overall high level of intelligibility of the non-native consonants, we observed that there are a number of consonants that are clearly less intelligible than their native American counterparts. Most of these cases could be accounted for in terms of negative transfer from the mother tongue. In several of these cases, however, the account could only be given in retrospect – there seems no reasonable way to predict the intelligibility problem a priori.

Importantly, we also found a number of non-native consonants that were identified better as the intended targets than was the case for native American tokens of these consonants. This situation, however, was encountered only when the listeners had the same language background as the speakers. These cases, then, are concrete instances of interlanguage benefit in intelligibility.

Rather than drawing more, and more detailed, conclusions, we will now first present and analyze the intelligibility of English consonant clusters in Chapter seven, and then discuss the intelligibility of consonants in more general terms.

Chapter eight

Intelligibility of consonant clusters

8.1 Introduction

In this chapter I will present the results for intelligibility of intervocalic consonant clusters for three groups of listeners. The results are from the same groups of listeners as in the previous chapters. Since simplex consonants and consonant clusters are all composed of the same phonetic substance, i.e. consonants, we might expect few or no significant differences between the results of the previous and the present chapters. However, consonant clusters are conspicuously absent in Mandarin Chinese. We know from work on other languages (see Chapter two) that clusters constitute a source of difficulty – and such problems may also be found for Chinese learners of English. The format of our experiment does not allow us to test such escape strategies as vowel insertion to break up awkward clusters (as happens in the English of Japanese or Indonesian learners). Nor can the alternative, deleting one of the consonants from the input cluster, be checked in our data, unless a three-member cluster were simplified to a two-member cluster. We must bear in mind that the forced-choice paradigm used in our experiment may have led to an overestimation of the quality of the pronunciation (and identification) of consonant clusters.

8.2 Results

8.2.1 Overall results

The overall results for cluster intelligibility are presented in Figure 8.1, broken down by nationality of the listeners and broken down further by nationality of the speakers. As in Chapters six and seven, the data were submitted to an Analysis of Variance (ANOVA) run on the mean percent correct scores for each listener with nationality (or: language background) of speaker and nationality of listener as fixed factors.

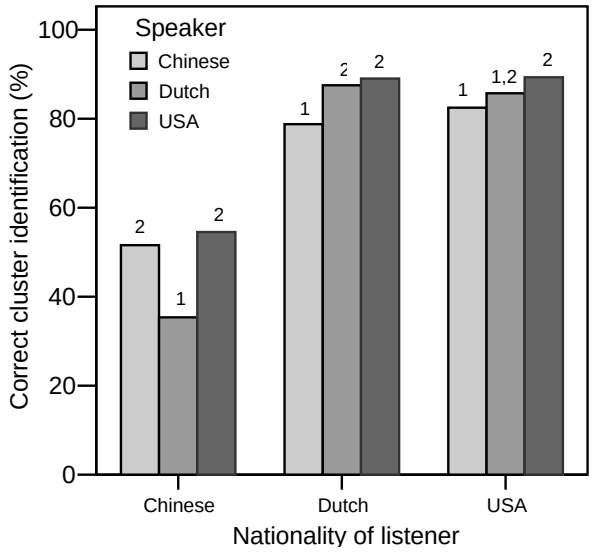


Figure 8.1. Percent correctly identified consonant clusters for Chinese, Dutch and American listeners broken down by accent of speakers. Numbers above the bars indicate the subgroup membership as determined by the Scheffé procedure. Numerical values of means, N, SD and Se are included in Appendix A8.1.

Across the speaker groups, the Chinese listeners have the lowest consonant cluster identification scores (53 to 56% correct, mean = 48%). Dutch listeners perform very closely to American listeners (79 to 89% correct, mean = 85%), and the American listeners are the best (82 to 89% correct, mean = 86%). The main effect of listener is highly significant, $F(2, 315) = 371.0$ ($p < .001$). Scheffé post hoc tests reveal that the Chinese listeners differ from Dutch and American listeners, who do not differ from each other.

The effect of speaker nationality is also significant but much less so than the effect of listener, $F(2, 315) = 15.9$ ($p < .001$). In fact, the listener effect in the cluster data is more than 20 times stronger than the speaker effect. The Dutch (mean = 70%) and the Chinese (mean = 71%) speakers do not differ from each other, but both are poorer than the American speakers (mean = 78%).

Figure 8.1 shows overall correct consonant cluster identification. It does not allow us to identify individual clusters that represent special difficulties. Therefore, we ask, firstly, which are the problematic clusters for each group of listeners? This question will be taken up in the following section (§ 8.2.2). Secondly, if a sound is massively misidentified, then what is it heard as instead? This question will be dealt with later when we examine the confusion structure in the cluster data (§ 8.2.3).

8.2.2 Correct cluster identification

In order to get an overview of which clusters are more difficult than others, for each combination of speaker and listener nationality, we present the percentages of clusters correctly identified by Chinese, Dutch and American listeners in separate panels in Figure 8.2. In each panel the results have been broken down by nationality of the speakers. Per panel, the 21 consonant clusters have been ordered in descending order of correct identification, when the speakers are American. The intelligibility for specific consonant clusters may differ widely between speaker nationalities. Table 8.1 lists Pearson's r for percent correct cluster identification in the three pairs of speaker nationalities for each of the three listener groups. The r-values are low and do not reach statistical significance, except those between Dutch and American speakers when the listeners are not Chinese; here the coefficients are between .5 and .6, which is significant at the $p < .05$ and $p < .01$ levels, respectively. Apparently, the consonant clusters spoken by native English and non-native Dutch speakers are to some extent (relatively) equally difficult. This would seem to make sense, since the English and Dutch sound systems have a large inventory of (often the same) consonant clusters, whilst Chinese has no consonant clusters at all.[1]

[1] Chinese has consonant clusters on the surface. These are combinations of some onset consonant followed by a glide /j/ or /w/, which in the phonology of Chinese are not counted as part of the onset but are parsed with the vowel. Only one such cluster was included in our test materials, viz. /sw/, which happens to be a combination that does not occur in the Chinese inventory.

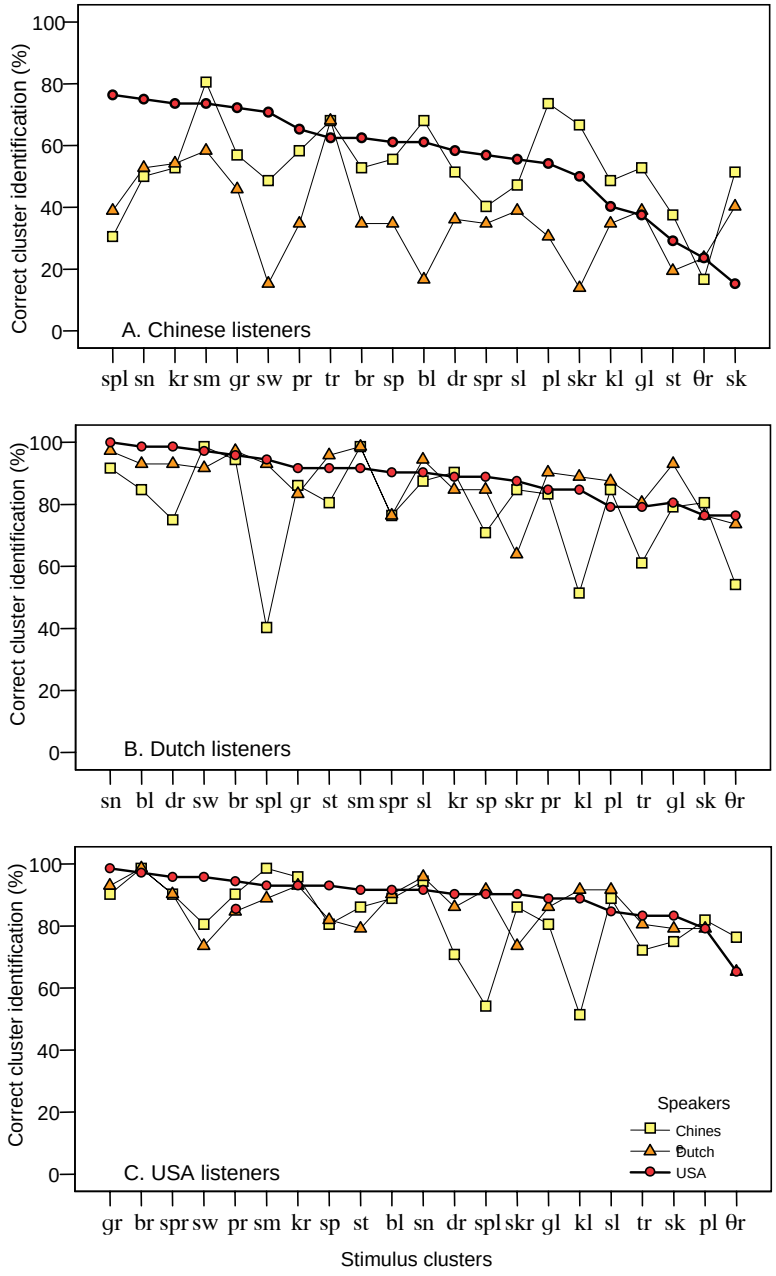


Figure 8.2. Correct identification (%) of 21 English intervocalic consonants produced by Chinese, Dutch and American speakers. Panels A, B and C present the results for Chinese, Dutch and American listeners, respectively.

Table 8.1. Pearson's correlation coefficients for identification of consonant clusters produced by Chinese, Dutch and American speakers broken down by nationality of the listeners.

Listener nationality	Speaker nationalities		
	CN ~ NL	CN ~ US	NL ~ US
CN	.268	.330	.362
NL	.255	.346	.545*
US	.217	.340	.606**

*: $p < .05$; **: $p < .01$

Figure 8.2-A shows the Chinese listeners' identification of the 21 consonant clusters of Chinese, Dutch and American speakers. The correct identification rate for American speakers runs from more than 80% (for /gl/) down to 16% (for /sk/). It is not the case that the American speakers' cluster tokens are more intelligible than the non-native tokens as has happened in the results for the simplex consonants. Six Chinese-accented consonant clusters are clearly identified better by Chinese listeners; these are /pl, skr, gl, kl, st, sk/. Dutch-accented clusters are extremely difficult for Chinese listeners. Almost all the clusters are identified more poorly than either Chinese-accented clusters or than the native American tokens. Especially the clusters /sw, bl, skr/ are poorly (< 20% correct) identified.

Figure 8.2-B shows the Dutch listeners' identification of the 21 consonant clusters of Chinese, Dutch and American speakers. The correctness of American consonant tokens covers a range from 99% to 80%. In this figure we can see that the American speakers' tokens almost invariably get the highest identification scores, with no significant exceptions. When Dutch listeners listen to their fellow speakers, there are just few clusters, /skr, spr/, that are identified clearly more poorly than the American counterparts. Chinese-accented consonant clusters are obviously the most difficult tokens for Dutch listeners as also happened in the case of the simplex consonants (Chapter seven). Chinese-accented /spl, kl, tr, θr/ are especially difficult for the Dutch listeners (between 40 and 60% error).

Figure 8.2-C shows American listeners' identification of the 21 consonant clusters of Chinese, Dutch and American speakers. The percentage is presented in the order of correctness from high (97%) to low (70%) of every consonant produced and identified by American native speakers. American listeners have the highest identification score for /br/ produced by their own speakers and the lowest identification for /θr/ (67%). This indicates that native American listeners have problems with their own speakers for certain consonant clusters. Nevertheless, there is substantial native language benefit, as the scores for other speaker nationalities are poorer overall. Dutch-accented clusters and Chinese-accented clusters are both poorly identified by American listeners but the figure reveals that the problematic consonant clusters may differ between the two non-native varieties of English. In responding to Dutch accented clusters, American listeners have clear difficulties in listening to /tr, θr, sw/; difficult Chinese-accented clusters are /kl, gl, st, sk, spl/.

We will now examine the confusion structures among the sets of consonant clusters, for each combination of listener and speaker nationality, in an attempt to understand why certain clusters present specific problems.

8.2.3 Confusion structure in consonant clusters

The clusters have been arranged in a matrix-like structure such that place of articulation appears along the vertical axis, with labials at the top, alveolars in the middle and velars at the bottom of the matrix. Each category along the vertical dimension has a top row and a bottom row. The top rows exclusively list two-member clusters; three-member clusters are on the bottom rows (midway between voiced and voiceless). The horizontal axis of the matrix is composite. The first two columns comprise /sC/ clusters, where C is a plosive in the first column and a sonorant in the second column. The third and fourth columns list obstruent + /r/ clusters, while the fifth and sixth columns list obstruent + /l/ clusters. Within each pair of columns, voiceless obstruents appear on the left (in odd-numbered columns) and their voiced counterparts to the right in even-numbered columns. Voicing is not contrastive in three-member clusters; hence these have been listed in between the odd and even-numbered columns, when applicable. As before, in order to avoid visual clutter, only confusion pairs have been indicated with arrows when the specific confusion occurred in 20% or more of the responses to the stimulus.

8.2.3.1 Cluster confusion for Chinese listeners

Figure 8.3A-C shows the cluster confusion structure for Chinese listeners responding to Chinese, Dutch and American speakers of English, respectively. The graphs show that the number of strongly confused cluster types is small, much smaller than was the case for either vowels or simplex consonants.

In Figure 8.3 A, Chinese listeners confused Chinese-accented /st/ with /sp/ (31%), /spl/ with /spr/ (31%) and /sl/ with /spl/ (21%). When listening to Dutch-accented clusters (figure 8.3B) /spl/ is identified as /spr/ (33%), /tr/ as /θr/ (23%) and /gr/ as /kr/ (25%). There are only two confusion pairs when Chinese listeners respond to American speakers /spl > spr/ (26%) and /kl > kr/ (33%).

Interestingly, the /spl > spr/ confusion pair is a problem for Chinese listener irrespective of the nationality of the speakers. This would indicate, of course, that the source of the confusion is not so much in a speaker defect but in the perception. Chinese listeners are relatively insensitive to the /r/ ~ /l/ contrast, and it does not matter very much whether the contrast is properly marked in the stimulus. It would seem, moreover, that the confusion is restricted to three-member clusters only; /l/ and /r/ were not confused as simplex consonants (Chapter seven). It is not clear why the confusion is directional from /spl/ to /spr/ only. The same directionality is observed in /kl > kr/; never do we find a confusion from /r/ to /l/.

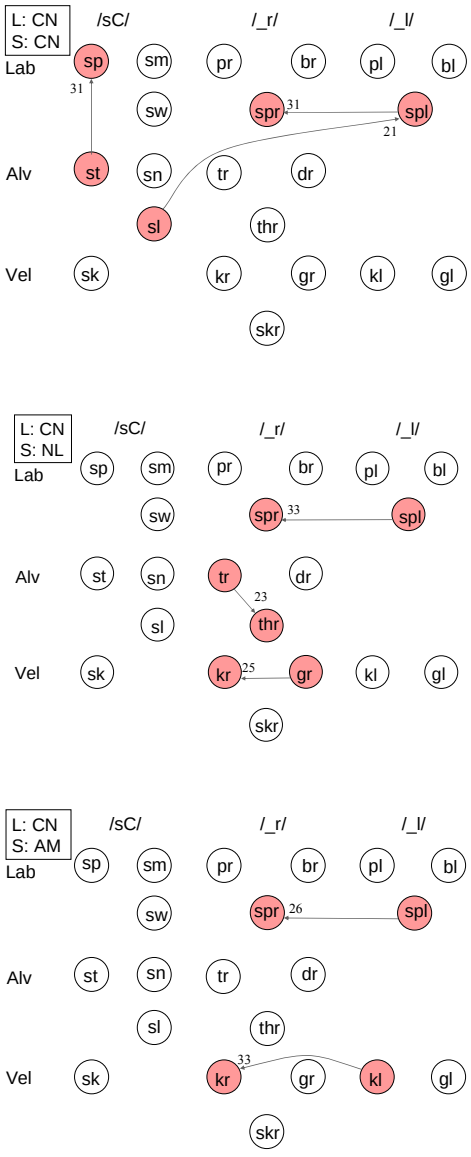


Figure 8.3A-C. Confusion graphs for Chinese listeners, exposed to Chinese (CN), Dutch (NL) and American (US) speakers (from top to bottom). Only confusions ≥ 20% are indicated by arrows. L= listeners, S = speakers.

8.2.3.2 Cluster confusions for Dutch and American listeners

Figure 8.4A-B lists the confusion pairs for consonant clusters for Dutch and American listeners, respectively, when exposed to Chinese-accented consonant clusters. When the speakers are either American or Dutch, no confusion pairs were obtained with a frequency ≥ 20%, which is why the confusion graphs involving Dutch or American speakers will not be presented (they would not show any arrows).

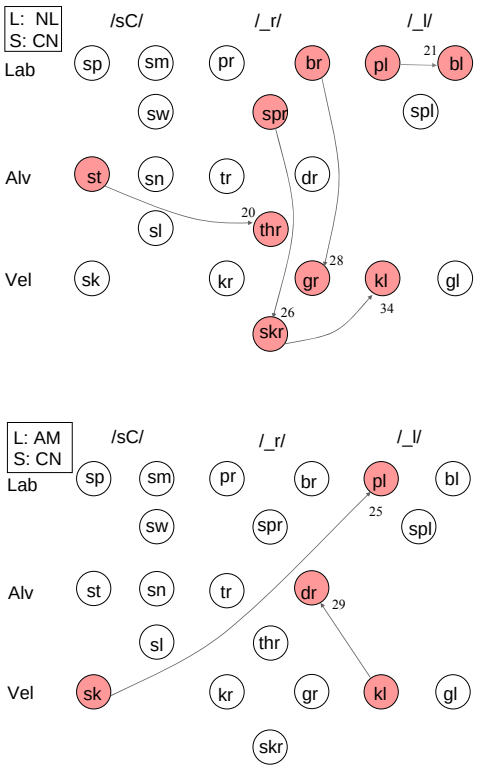


Figure 8.4A-B. Confusion graphs for Dutch (NL) and American (US) listeners, exposed to Chinese (CN) speakers. Only confusions ≥ 20% are indicated by arrows. L = listeners, S = speakers.

As could already be seen in the presentation of the percentages correct (Figure 8.2), consonant clusters are not really a problem between Dutch and American speakers and listeners – at least not when determined in a forced choice paradigm allowing cluster responses only. Apparently, the sound systems of Dutch and English are similar enough to prevent large-scale confusion in the consonant clusters.

However, when the speakers are Chinese we find five confusion pairs. With the exception of one (/pl > bl/) all these confusion pairs involve a cluster – either as source or as target – that contains a consonant or a sequence that is illegal in the sound system of Dutch: /br > *gr/, /*skr > kl/, /st > *θr/ and /spr > *skr/. Since none of these confusions are found when the listeners are American (next section), I suggest that the problem is caused by the Dutch listeners.

Only two confusion pairs remain when the listeners are American. One involves an /l > r/ confusion (/kl > dr/) but not the reverse. The other pair, /sk > pl/, is not part of a recurring pattern.

8. 3 Summary

Table 8.2 lists the number of problematic consonant clusters in the data. A problematic cluster is defined as a cluster which in any speaker-hearer combination is identified correctly in less than 75%. The numbers are broken down for the nine combinations of speaker and listener language background.

Table 8.2. Number of problematic consonant clusters broken down by nationality of speaker and of listener.

speaker	listener			
	Chinese	Dutch	USA	Total
Chinese	**20**	5	18	43
Dutch	21	**1**	0	22
USA	19	0	**1**	20
Total number	60	6	19	85

Table 8.2 shows that, overall, Dutch listeners have the least number of problematic consonant clusters in the three groups of listeners. American listeners are a good second, and Chinese listeners clearly have the most problems. The number of problematic clusters is 60 out of 85 in the Chinese listener group (75%) and 43 out of 85 in the Chinese speaker group (51%).

8.4 Conclusions and discussion

We hypothesized that English consonant clusters would be more difficult to identify, as the sound system of the L2 speaker's native language deviates more from English. The differences between the Dutch and Chinese consonant inventories are relatively small, and both languages have roughly the same number of consonants that would be reasonable substitutes for English targets, but there are no consonant clusters in Mandarin Chinese. In this respect the prediction is rather different than either in the case of the vowel systems or in the case of the consonant systems. The results show that Chinese-accented consonant clusters are relatively well identified

by all groups of listeners, and certainly better than the vowels (Chapter six) and simplex consonants (Chapter seven). Dutch-accented consonant clusters are very well identified by American listeners and by the Dutch listeners themselves, but not by Chinese listeners. Dutch-accented consonant clusters are the most difficult for Chinese listeners. The difference in intelligibility between Chinese and Dutch-accented consonants is relatively large.

In spite of the intelligibility of some Chinese-accented consonant clusters, we observed that there are very few clusters that are clearly more intelligible than their native American counterparts. This is in contrast with our earlier findings for simplex consonants, where we found a range of Chinese-accented consonants which were better identified by Chinese listeners than American native tokens of the same consonants: /p, f, w, s, z, ʃ, tʃ/.

Chapter nine

Intelligibility of words in sentences

9.1 Introduction

In the preceding chapters we were concerned with the intelligibility of the smallest building blocks of language, i.e. the vowels and the consonants, in either meaningless sound sequences or in existing words and short phrases constructed such that the identification of target segments was not made (more) predictable by context. This is a listening situation that occurs only very rarely in everyday life. Normally sounds occur in meaningful words. Typically, when sounds occur in the context of a word in a sentence, the listener needs to get only a few of the constituent segments to piece the word together, using lexical redundancy. For instance, the last two sounds in the word *elephant* are perfectly predictable once the listener has heard /ɛləfə/; there are simply no other words in the English lexicon than *elephant* that begin with this sequence. When the target word is embedded in a meaningful context sentence, segments in short, monosyllabic words will also be predictable. If the listener misses the initial consonant in *I heard the _at mew*, the listener will know from his knowledge of the world that the entity that produced the mewing sound must be a *cat* rather than a *rat* (or *bat* or *gnat*), let alone a *mat*. In the present chapter we will deal with the intelligibility of meaningful words in several kinds of sentence contexts.

The first type of sentence context is a syntactically correct structure, but the words that are filled in the various slots in the structure do not yield a meaningful sequence. For instance, in *The state sang by the long week*, it is at least odd that an inanimate subject *The state* should perform an action normally only manageable by humans (i.e. singing); also, the choice of the preposition *by* would seem to be ungrammatical. These sentences were called Semantically Unpredictable Sentences (Benoît, Grice and Hazan, 1996; see also Chapter two) or just SUS sentences. They were originally constructed for the purpose of evaluating the quality of text-to-speech systems. The claim would be that the SUS test will discriminate in a highly sensitive way between small differences in speech quality, when the subjects are native listeners of the stimulus language. The test was not developed to discriminate excellent from not-so-excellent speakers and listeners.

The second type of test we used in our materials is the SPIN test, which stands for SPeech In Noise test (Kalikov, Stevens and Elliot, 1977). The SPIN test (see also Chapter two) requires listeners (patients with hearing loss, in the original application) to fill in the last word of a short sentence; the final word is either highly predictable (HP) from the preceding words in the sentence (e.g. *She put her broken arm in a*

sling) or not predictable from the context (low predictability, LP, e.g. *We should consider the map*). The SPIN LP sentences are more or less comparable with the SUS sentences in that the target words appear in grammatically correct word sequences, may benefit from the presence of a precursor utterance (phonetic adaptation to phonetic quality, melody, rhythmic structure and coarticulation) but not from any semantic constraints. Earlier comparisons of SUS sentences with SPIN sentences (Hazan and Shi, 1993), using normal English listeners, brought to light that the SUS sentences were much more difficult (12% correct on average) than the SPIN sentences (HP 84% correct, LP 48% correct).

We decided to include all three types of sentences in our test battery (i.e., SUS, SPIN-LP, SPIN-HP,) precisely because together they would seem to cover a very large range of listener abilities, large enough to adequately discriminate all nine combinations of speaker and listener nationalities in our study. More specifically, since the purpose of the SPIN audiology test was to discriminate between listeners from a wide range of hearing ability and that of the SUS test was to differentiate between better and poorer talking machines, one would expect therefore that the SPIN test will be rather more sensitive to differences between listeners, whilst the SUS test would be susceptible to differences between speakers.

I will now present the results of word recognition in sentences for each of the nine combinations of speaker and listener groups (Chinese, Dutch, and American). The results of the SUS sentences will be presented first (§ 9.2), followed by the SPIN sentences (§ 9.3). Within each test, I will first present the intelligibility scores in terms of percent correctly reported words. Here a word will be counted as incorrectly reported even if just one phoneme within the word was incorrectly reported. In a second analysis I will present a more refined scoring method where onsets, vocalic nuclei and codas are scored separately so that each target word may have a score of 0, 33, 67 or 100% correct. In between the two word recognition analyses, I will present and analyze the results for onsets, nuclei and codas separately. This latter breakdown of the data will afford a direct comparison with the vowel, consonant and cluster identification results in Chapters six, seven, and eight, respectively. The same sequence of results will then be repeated for the SPIN sentences.

9.2 Intelligibility in SUS sentences

Every listener heard 30 SUS sentences. These were evenly distributed over five different syntactic frames (see Chapter four, § 4.2.4) with each speaker (i.e., one male and one female Chinese, Dutch and American speaker) donating one sentence to each syntactic frame. Speakers were blocked over sentences such that any listener heard each sentence only once, and every speaker donated each sentence as often as any of the other speakers.

9.2.1 Overall result

A broad phonemic transcription was produced for all the stimulus (input) and response (output) forms. To this effect the orthographic input and output forms were

converted to broad IPA by hand. In the case of the input forms this could be done efficiently, since the same words occurred in the same order for all of our 108 listeners; once the input list was transcribed it could simply be copied. The responses required much more work. All transcriptions were checked by an independent expert; whenever discrepancies were found between the transcribers, these were discussed and checked against a pronouncing dictionary (Kenyon and Knott, 1944). Stress marks were not included in the transcriptions of either input or output forms.

Figure 9.1 presents the overall percentages of correctly reproduced words broken down first by nationality of the listener and broken down further by nationality of the speaker.

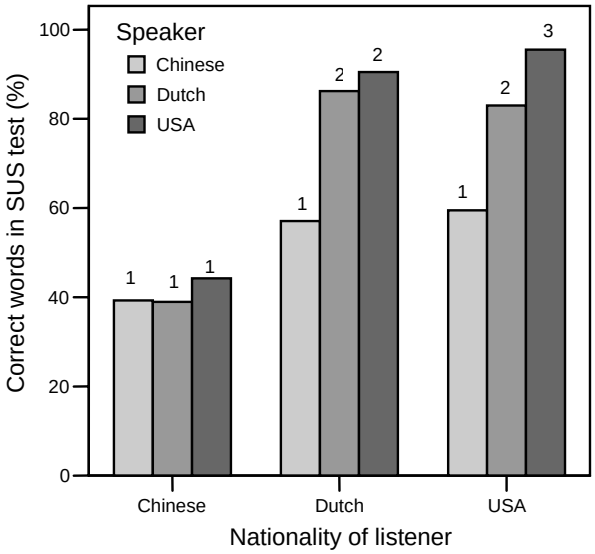


Figure 9.1. Percent correct word identification in SUS test for Chinese, Dutch and American listeners broken down by accent of speakers. Numbers above the bars indicate the subgroup membership as determined by the Scheffé procedure. Numerical values of means, N, SD and Se are included in Appendix A9.1.

The effect of listener nationality is highly significant by a two-way ANOVA with listener and speaker nationality as fixed factors, $F(2, 312) = 669.0$ ($p < .001$).[1] Post-hoc Scheffé tests reveal that the Chinese listeners (mean = 41% correct) performed more poorly than the Dutch (78%) and the American (79%) listeners, who did not differ from each other. There is a smaller effect of speaker nationality, $F(2, 312) = 240.0$ ($p < .001$) by which Chinese speakers are poorest (52%), Dutch speakers are intermediate (70%) and Americans are best (77%). All three speaker nationalities

[1] Unfortunately, the responses of one Chinese listener were missing, so that the number of valid listeners in this group was 35 instead of the nominal 36. This is reflected in the smaller number of degrees of freedom in the error terms in the ANOVAs.

differ from each other (Scheffé, $p < .05$). As was also observed in earlier chapters, the effect of listener nationality is appreciably stronger than that of speaker nationality (here roughly in a 3:1 ratio). As before, the speaker × listener interaction also reached significance, $F(4, 312) = 45.9$ ($p < .001$). The interaction is clearly the result of what we have called the interlanguage benefit in earlier chapters. For Dutch and American listeners, Chinese speakers are difficult to understand but Chinese listeners have word-recognition scores for fellow Chinese speakers which are not less than for the Dutch or American speakers. By the same token, Dutch listeners do relatively better for Dutch speakers than for speakers of other nationalities. Similarly, even American speakers have a small advantage when listening to their own speaker type.

Figure 9.2 lists the percentage of correct word recognition for each of the nominally 36 listeners in each nationality. The figure shows very clearly how sensitive the SUS test is. There is a clear gap in the distribution of the scores at 60%. Chinese listeners never obtain scores of 60% or more, while no Dutch or American ever gets a score below 60. There is virtually no difference between the Dutch and the American listener groups.

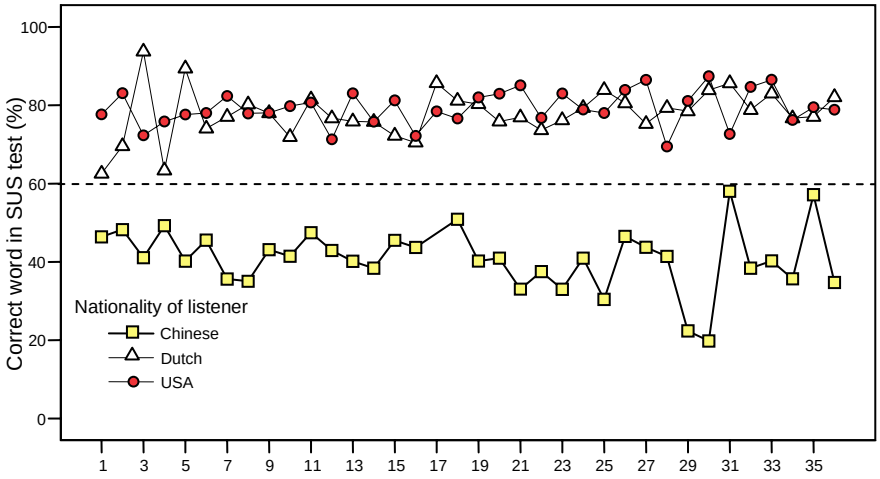


Figure 9.2. Correct identification (%) for words in SUS test by Chinese, Dutch and American listeners. Note that listener 17 is absent from the Chinese set of subjects (see also note 1).

9.2.2 Intelligibility of subsyllabic constituents

So far, we have merely analyzed the results in terms of the percentage of correctly recognized words. In order to obtain a more refined view of the specific difficulties we will now present percent correctly reported subsyllabic units, i.e., onsets, vocalic nuclei and codas. Since the consonant inventories of Chinese, Dutch and English differ less in size and complexity than the vowel inventories, we would predict that the effect of speaker and listener nationality will be greater for nuclei (vowels) than

for onsets (simplex initial consonants and clusters). Moreover, since both Dutch and English allow a wide variety of coda consonants and clusters, whilst Chinese only allows nasals in the coda, we predict large differences in percent correctly identified codas when speaker and/or listener nationality is Chinese. Once the scores are broken down by subsyllabic constituent, we may also derive a more refined overall word recognition score by counting correct onsets, nuclei and codas together.

Figure 9.3A-C presents the percentages of correctly identified onsets, nuclei and codas, respectively, broken down by nationality of listener and by nationality of speaker, as was done for overall word recognition in Figure 9.1. Figure 9.3D presents the composite word recognition scores, where each word can be recognized at 0, 33, 67 of 100% correct, depending on the number of subsyllabic constituents reported correctly.

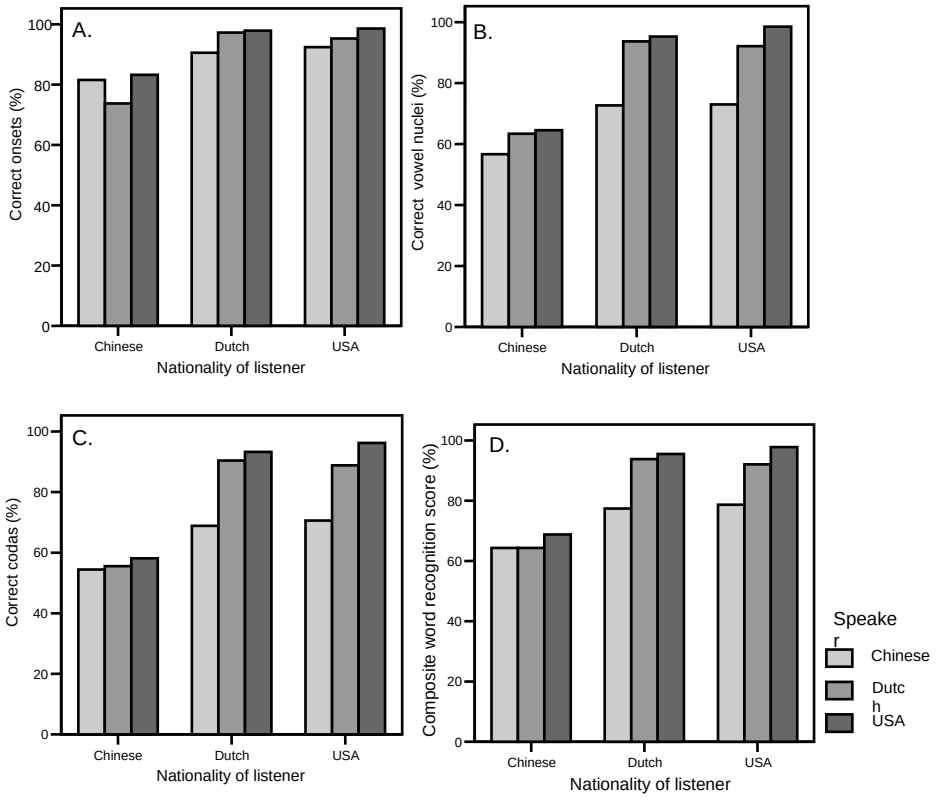


Figure 9.3. Percent correctly identified onsets (A), vocalic nuclei (B), and codas (C) in word identification in SUS test for Chinese, Dutch and American listeners broken down by accent of speakers. Panel D plots a composite word-recognition score (see text).

Interestingly, the difference between the Chinese, Dutch and American listeners in Figure 9.3A is relatively minor (80, 95, 95% correct), $F(2, 312) = 302.2$ ($p < .001$), for the onsets. It is greater for vocalic nuclei (Figure 9.3B, 68, 83, 86% correct), $F(2, 312) = 469.1$ ($p < .001$), and greatest for the codas (Figure 9.3C, 56, 84, 85% correct), $F(2, 312) = 527.1$ ($p < .001$). The prediction formulated above, i.e. that the difference between the three listener nationalities would be smallest in the onsets, intermediate in nuclei and greatest in the coda, is borne out by these data.

Observe that there is a striking resemblance between Figure 9.3A and Figures 7.1 (simplex onset consonants) and 8.1 (complex onsets). In all three figures the Chinese listeners exhibit considerable interlanguage benefit, which graphically shows up in the poorer identification rates for Dutch-accented onsets. This would show that the detailed tests using nonsense sound sequences in Chapters seven and eight make valid predictions of the listeners' behavior in meaningful words. A more detailed error analysis may be done to examine to what extent the consonants and clusters that were problematic in the nonsense materials are also problematic in the meaningful words.

Figure 9.3D shows that the more refined scoring mechanism is beneficial to the poorer speaker and listener groups. The percentages of correct scores for the Chinese listeners are elevated from ca. 40% to ca. 60%. Similarly, the scores for the Chinese speakers are raised from 60 to 80% when the listeners are either Dutch or American. Clearly, the composite word recognition scores discriminate less effectively between poorer and better combinations of speaker and listener groups.

9.3 Intelligibility in SPIN sentences

9.3.1 About the SPIN test

The SPIN test (Speech In Noise) was developed as a diagnostic tool to determine the severity of hearing loss in audiological settings. The article in which the concept of the SPIN test was introduced (Kalikov et al., 1977) mentions that the test had not been administered systematically to patients but data were presented to a normal-hearing reference group of American listeners. These data can be used as a background against which some of our own data can be gauged. SPIN sentences should be administered at various signal-to-noise levels. In our application we did not do this, as we noted in pilot versions of our test that the range of intelligibility across the various speaker and listener types was more or less fully covered; had we presented stimuli in noise, some of our listener groups would not have understood a single word.

The SPIN test presents sentence-final target words in high-predictability (HP) and in low-predictability (LP) contexts (see introduction). In the LP contexts the results should be roughly similar to those obtained in the SUS sentences. In both type of tests, the target words have to be understood purely from bottom-up acoustic information contained in the word itself; syntactic and semantic cues in the preceding context are useless. In the HP sentences, the words in the preceding context strongly constrain the identity of the sentence-final target word. In this condition, the SPIN test comes rather close to real-life speech recognition, where the

outcome of the processing task is the result of interaction between acoustic bottom-up information and top-down semantic and syntactic information. It seems a reasonable hypothesis that the interaction between the two information sources makes heavier demands on the listener, so that the native listeners will benefit substantially from the contextual information but that the non-native listeners will be hindered by the dual-processing task – having to attend to two non-automatized processing tasks at the same time and not doing a good job on either.

9.3.2 Overall word recognition in SPIN sentences

We will first present the results in terms of overall word recognition, once across both predictability conditions, and then separately for LP and HP sentences. In this part of the data presentation a word will be counted as an error if any component of it was not correctly reported by the listener, whether a coda consonant, a vocalic nucleus of some part of the coda.

As before, a broad phonemic transcription was produced for all the stimulus (input) and response (output) forms. All the target words were monosyllabic. However, in just a few cases listeners reported a two-syllabic word, e.g. *bet* was twice reported as *better* and *lane* as *today*. In such cases the segments of input and output forms were aligned manually such that the best match was obtained for an onset, nucleus and a coda. In the examples just given the first three segments of *better* were aligned with *bet* (also respecting the stress location); in *today* the /d/ was aligned with /l/ of *lane* (error), the two stressed vowels with each other (correct) and an empty coda was matched with the /n/. Non-aligned (extra) phonemes were not included in the analysis. Differences between the aligned input and output transcription were detected automatically; the scoring of the responses was done by computer. When even a single mismatch was found between input and output form, the entire word was scored as an error. In other words, every single segment in the word had to be reported correctly or else the word was not counted as a correct response. This is a very strict scoring principle. A more sophisticated scoring system in which errors in onsets, nuclei and codas were counted separately will be presented in a later section (§ 9.3.3).

Figure 9.4 presents the percentages of correctly recognized target words as defined here, broken down by nationality of listener and of speaker. The data have been accumulated over the two predictability conditions.

The data in Figure 9.4 were subjected to a three-way ANOVA with predictability (LP versus HP) of the targets, nationality of speaker and nationality of listener as fixed effects. The effect of listener was largest, $F(2, 630) = 807.6$ ($p < .001$), with Chinese listeners scoring 27% correct word recognition, Dutch listeners 63% and Americans 77%. All three listener groups differed significantly from each other (Scheffé, $p < .05$). A smaller effect was obtained for speaker nationality, with Chinese speakers performing significantly poorer (32%) than the Dutch and American speakers (both at 67%), $F(2, 630) = 500.4$ ($p < .001$). The effect of contextual predictability is much smaller, with 52 versus 60% correct words for LP and HP, $F(1, 630 = 58.8$ ($p < .001$). There was significant interaction between speaker and listener nationality, $F(4, 630) = 71.7$ ($p < .001$), which to some

extent reflects interlanguage or native language benefit. However, there is one remarkable instance of foreign-language benefit: the Chinese listeners perform significantly better when the speakers are Dutch than when the speakers are either Chinese or American. Possibly, the Dutch non-natives speak more slowly and deliberately than the American native speakers, which may have helped the Chinese listeners to get more useful information from the signal than with other speaker nationalities.

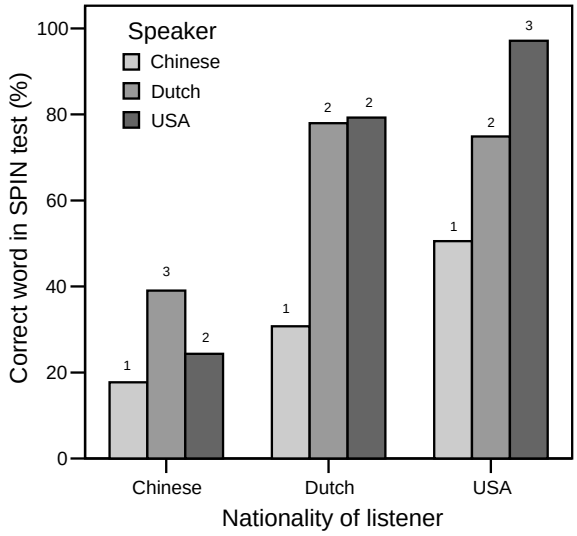


Figure 9.4. Percent correct word identification in SPIN test for Chinese, Dutch and American listeners broken down by accent of speakers. Numbers above the bars indicate the subgroup membership as determined by the Scheffé procedure. Numerical values of means, N, SD and Se are included in Appendix A9.2.

There is also significant interaction between the predictability condition of the targets and listener nationality (but not with speaker nationality), $F(2, 630) = 22.6$ ($p < .001$). We will analyze the interaction in the next paragraph. Also the three-way interaction was significant, $F(4, 630) = 18.0$ ($p < .001$). We will first analyze the two-way interaction (in Figure 9.3), and then we will analyze the three-way interaction by presenting the results for LP and HP separately (in figure 9.6A-B).

Figure 9.5 shows the interaction between predictability and listener nationality in detail. The figure shows that there is no effect of contextual predictability for the non-native listening groups, whether Chinese or Dutch. However, the difference is significant for the American listeners; here HP targets get better recognition scores than their LP counterparts. It seems, therefore, as if only the Americans profit from the contextual information. This would be in line with our suggestion above that non-native listeners do not recognize enough of the context to use it to their advantage.

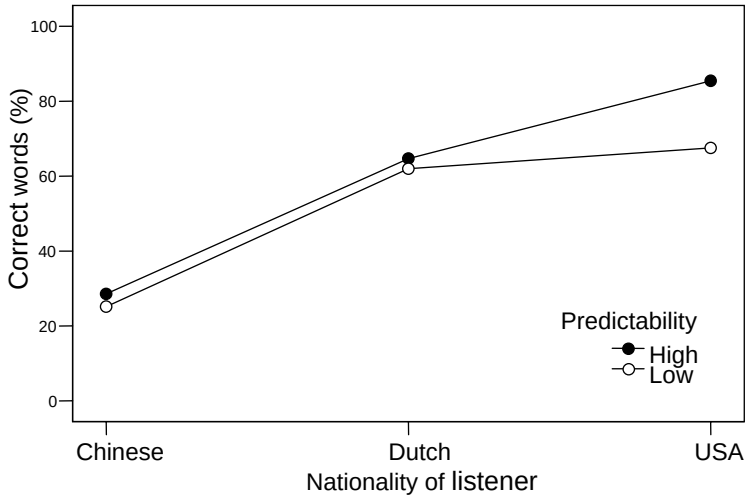


Figure 9.5. Percentage of correctly recognized words in SPIN test broken down by listener nationality and by contextual predictability of targets.

We will now present the word recognition scores for the LP and HP conditions separately. This is done in figure 9.6A-B.[2]

Comparing the two panels by listener nationality, we may observe, first of all, that the Chinese listeners benefit from HP words somewhat but only if the speakers are American. Also the gain in percent correct is counteracted by a small loss of intelligibility in the HP utterances of Chinese and Dutch speakers. The Dutch listeners have no advantage of HP words at all. Apparently, they fail to use the semantic information contained in the meaningful context preceding the targets. The American listeners present an altogether different configuration of scores. If the speakers are American it does not really matter whether the words are LP (95% correct) or HP (97% correct). The quality of the pronunciation is such that recognition is close to ceiling in both conditions; there is no room for improvement due to HP. However, when the speakers are non-native, the pronunciation is relatively poor, in fact very much poorer for Chinese speakers (37% correct) and rather poorer for Dutch speakers (77%). When these speakers are tested with HP words, the Americans get so much useful information from the context that they improve their recognition scores by roughly 25 percent for Dutch speakers and by 20 percent for Chinese speakers. So, the significant three-way interaction mentioned

[2] There are minor differences in the word recognition scores in figures 9.6A-B and earlier reports of the data (e.g. Wang and Van Heuven, 2005). The reason for the small discrepancies is that some errors in the database (wrong alignments of input and output transcriptions were corrected in the present final analysis.

above is due to the fact that contextual information is only used by native listeners, and only if there is room for improvement, that is, when the speakers are foreign.

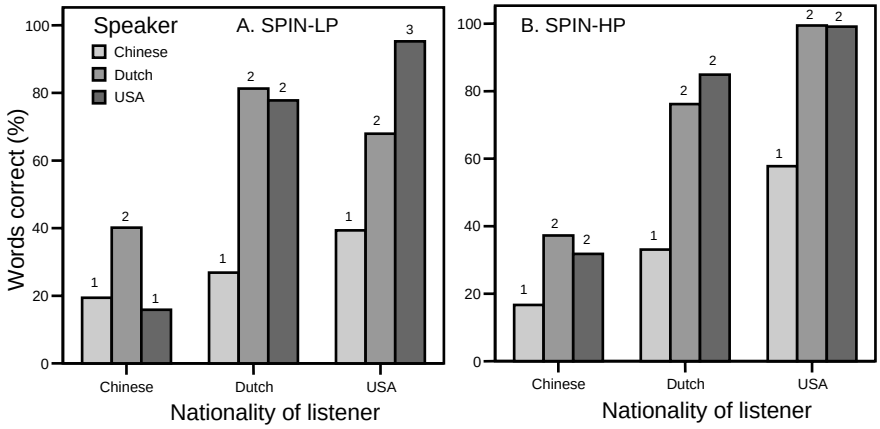


Figure 9.6. Percentage of correct word identification in SPIN test for Chinese, Dutch and American listeners broken down by accent of speakers, for low-predictability words (panel A) and for high-predictability words (panel B). Numbers above the bars indicate the subgroup membership as determined by the Scheffé procedure. Numerical values of means, N, SD and Se are included in Appendix A9.2.

9.3.3 Recognition of subsyllabic units in SPIN sentences

We will now examine the intelligibility of the subsyllabic components of the LP and HP target words. Figures 9.7A-B-C present percent correctly reported onsets, nuclei and codas for the LP words; figure 9.8A-B-C will do the same for the HP words.

As also appeared in Figure 9.3A for the SUS sentences, the difference between the Chinese, Dutch and American listeners in Figure 9.6A is relatively minor (73, 92, 95% correct), $F(2, 630) = 263.1$ ($p < .001$; all listener groups differ significantly, Scheffé) for the onsets. The difference is greater for vocalic nuclei (Figure 9.6B, 56, 78, 83% correct), $F(2, 630) = 286.4$ ($p < .001$; all listener groups differ significantly, Scheffé), and greatest for the codas (Figure 9.3C, 47, 71, 81% correct), $F (2, 630) = 354.9$ ($p < .001$; all listener groups differ significantly, Scheffé). Again, the hypothesis that the difference between the three listener nationalities would be smallest in the onsets, intermediate in nuclei and greatest in the coda (see also § 9.2.3) is borne out by these data. The resemblance also shows that the SUS sentences are highly comparable to the SPIN-LP sentences.

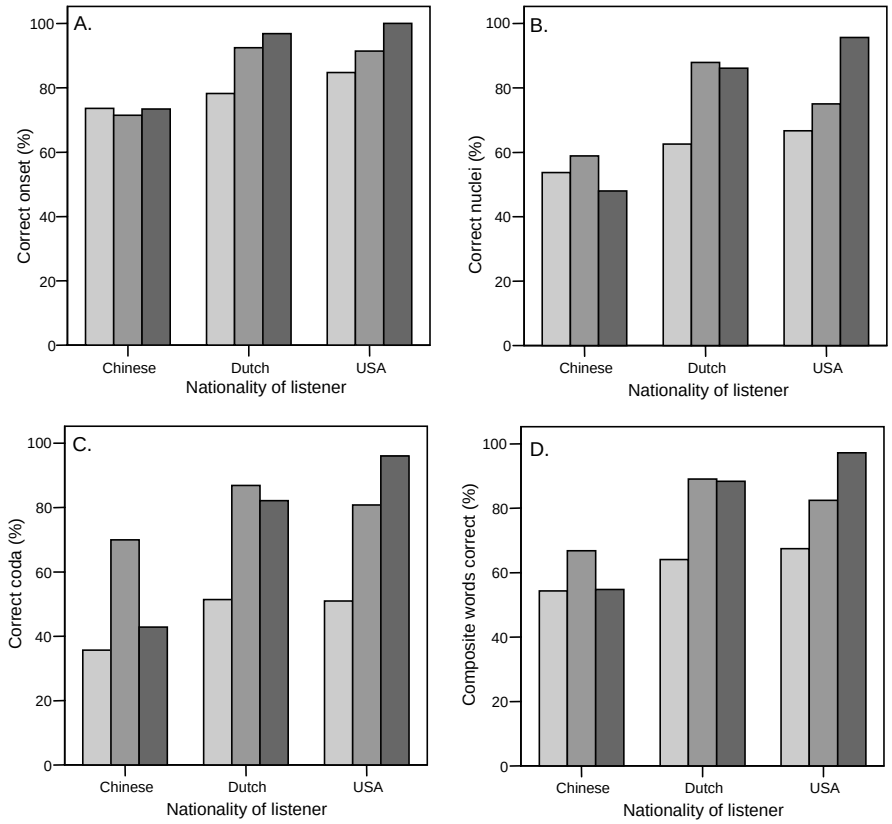


Figure 9.7. Percent correctly identified onsets (A), vocalic nuclei (B), and codas (C) in word identification in SPIN-LP test for Chinese, Dutch and American listeners broken down by accent of speakers. Panel D plots a composite word-recognition score (see text).

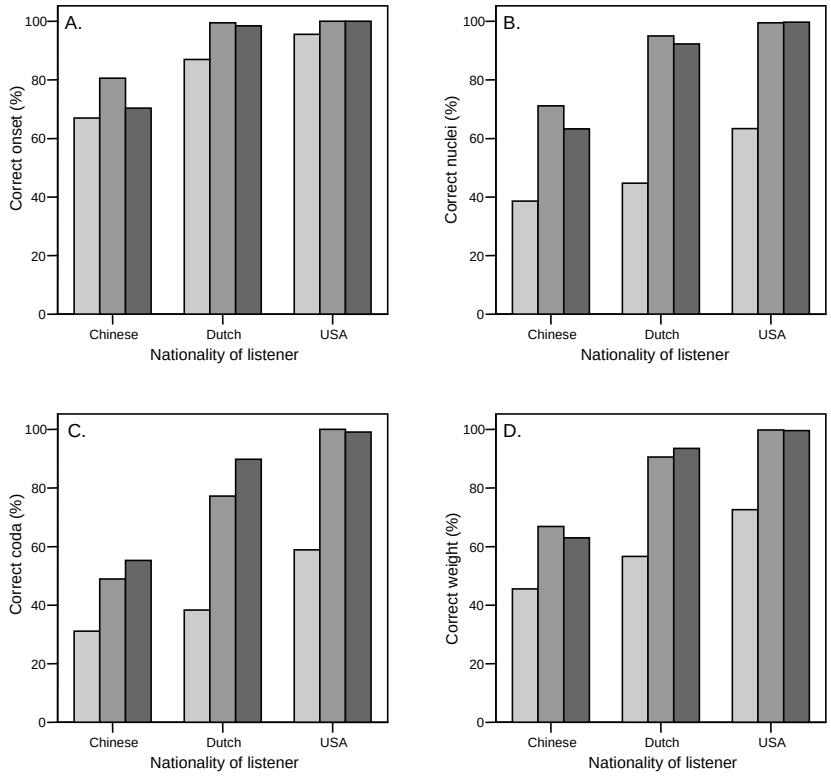


Figure 9.8. Percent correctly identified onsets (A), vocalic nuclei (B), and codas (C) in word identification in SPIN-HP test for Chinese, Dutch and American listeners broken down by accent of speakers. Panel D plots a composite word-recognition score (see text).

9.4 Conclusions

In this chapter we focused on the intelligibility of words spoken in the context of sentences rather than on the intelligibility of individual vowels and consonants in informationless contexts. Two types of sentence test were used: SUS and SPIN. The first test presented words in syntactically correct but semantically anomalous sentences, in which the function words correctly constrained the content words in terms of part of speech category but not in terms of meaning. One would expect words in such sentences to be difficult to understand. The second test contained syntactically and semantically correct sentences, which were constructed such that the sentence-final target word was either highly predictable from the preceding context (HP) or not. In the low-predictability sentences (LP) the context was neutral as to the identity of the targets, i.e. they were neither made more nor less predictable than when they had been presented as citation forms. All else being equal, the order of difficulty between the three types of sentences would be SUS > SPIN-LP > SPIN-

HP. Table 9.1 summarizes the scores for the three tests, overall and broken down by speaker and listener groups.

Table 9.1. Word recognition scores for SUS, SPIN-LP and SPIN-HP sentences broken down by nationality of listener and of speaker. See appendices A9.1-2 for number of listeners, and values of SD and Se.

Nationality of		SUS scores by		SPIN scores	
Listener	Speaker	word	sentence	LP	HP
Chinese	Chinese	39	5	19	17
Chinese	Dutch	39	6	39	38
Chinese	American	44	5	18	32
Dutch	Chinese	57	17	27	33
Dutch	Dutch	86	60	81	76
Dutch	American	91	71	78	85
American	Chinese	60	18	39	58
American	Dutch	83	52	68	99
American	American	96	85	95	99
Overall		**66**	**36**	**52**	**60**

The table shows that the overall prediction does not hold: the SUS sentences are the easiest type. However, within the two types of SPIN sentences the prediction is correct: words in HP sentences are easier than words in LP sentences but the difference is rather small (but significant, cf. § 9.3.2). Reasons why the SUS sentences obtained better scores than either of the SPIN sentence types will be discussed in § 9.5.

The overall word recognition scores tend to be more extreme for the SPIN sentences than for the SUS sentences. The least and most favorable speaker/listener combinations in the SUS test are Chinese/Chinese and American/American with 39 and 96% correct, respectively. The comparable numbers for the SPIN-LP test are 19 and 95%, and for the SPIN-HP test 17 and 99. The discriminatory power of the various types of tests used in this dissertation will be examined in more detail in the next chapter. For now it will suffice to say that tests seem to discriminate better as they come closer to real-life speech perception, i.e. words in normally constrained, meaningful sentences. Interestingly, although the SPIN sentences were developed as audiological test materials to be presented in a range of signal-to-noise ratios, no degradation by added noise was needed in order to create a sufficiently wide range of scores in the present application of the test. Clearly, the suboptimal performance of the non-native speakers and listeners compensated for the absence of added noise.

For all three types of test (SUS, SPIN-LP, SPIN-HP) we find that the largest effect is that of listener nationality. It is stronger than the effect of speaker nationality by a factor 3. For both listener and speaker effects we find that the Americans obtain the highest scores, closely followed by the Dutch nationals, while the Chinese subjects performed much more poorly. The effects of context, as determined by comparing the SPIN-LP and HP sentences, are generally minimal,

except for American native listeners; only native listeners use the information contained in earlier words in the sentence to predict the identity of the sentence-final target word.

Again we observed clear effects of the interlanguage benefit, showing that listeners who hear speakers of their own nationality obtain better scores than when they are exposed to speech of speakers from a different nationality.

In Chapter three we predicted that coda consonants would present problems especially for Chinese speakers and listeners, as Chinese does not have any coda consonants, except the nasals /m/ and /n/. In the earlier chapters on the production and perception of vowels, consonants and clusters, no materials were included on codas. The only possibility to test the effects of onset versus coda consonants and clusters is to examine the scores in the present word recognition tests. The results in Figures 9.3 for the SUS sentences and 9.6 and 9.7 show, for SPIN-LP and SPIN-HP sentences, respectively, that the greatest differentiation between the Chinese and the other listener nationalities is found in the coda consonants; differentiation is somewhat poorer in the vowels and least in the onsets.

The last conclusion we will draw from this chapter is that not much is gained by computing a partial word recognition score based on correct identification of sub-word constituents. Generally, the scores for partial word recognition show the same tendencies as those for the constituent parts; moreover, when compared with the overall word recognition scores the results show the same order among the nine speaker/listener combinations but in a more compressed range, i.e. with poorer differentiation among the nine combinations.

9.5 Discussion

There is a remarkable discrepancy between our results and those reported by Hazan and Shi (1993) (see also § 9.1). In both studies a comparison can be made of the results obtained with SUS sentences and with SPIN sentences. Hazan and Shi found word recognition scores of 12, 48 and 84 percent correct for SUS, SPIN-LP and SPIN-HP sentences, respectively. My results reveal not the slightest difference between the scores on the SUS sentences and those on the SPIN-LP materials. Moreover, although the overall effect of LP versus HP sentences in the SPIN test is preserved in my study, the effect of context was only found for American listeners when the speakers were non-native.

Hazan and Shi (1993) recorded the materials from one male British English speaker and presented the materials to 50 native listeners. The materials were presented with a signal to noise ratio of 6 dB. It is possible, therefore, that the degradation due to the poorer signal-to-noise ratio (SNR) caused the enormous differentiation between the three tests in Hazan and Shi. We presented all our materials in quiet. As a result percent correct word recognition is close to ceiling in all three tests – but only if American native listeners respond to American speakers. When our speakers and/or listeners are non-native, the scores are rather more in the middle of the range. However, in our edition of the tests, there was virtually no difference between the LP and the HP word in the SPIN sentences (except when American listeners responded to American speakers) and the SUS sentences were

some 10% better than the SPIN sentences for all conditions involving a non-native party. We must assume that the relative ease of the SUS test was caused by the way we presented the materials, i.e. not just once but repeatedly using a gating method incrementing the utterance in word-sized chunks.

The most important reason, however, why the mean SUS scores in Hazan and Shi were so low would seem to lie in the fact that these authors used the sentence as the scoring unit, whereas I computed word-recognition scores. In Hazan and Shi (1993), if even one word in a SUS sentence was wrong, then the entire sentence was wrong. In order to check whether my results would be more comparable to those of Hazan and Shi, I recomputed the SUS scores using the sentence as the scoring unit. The results in terms of the sentence-based scores have been listed in Table 9.1, along with the word recognition scores, as well as in Appendix A9.1.

Overall, the SUS scores drop from 66 to 36% when the sentence is used as the scoring unit instead of the word. As a result of this, the SUS scores are closer to those reported by Hazan and Shi (18% correct sentence recognition) but they are still considerably better. Moreover, the discriminatory power of the SUS sentence-based scores is better than that of the word-based scores. This property of the SUS test has been reported earlier by the designers of the SUS test (Benoît et al., 1996: 388). I will come back to the issue of discriminatory power of the tests used in my research in Chapter ten.

Chapter ten

Conclusions

10.1 Introduction

In this final chapter I will recapitulate the more general questions underlying the present study. I will identify research questions one by one, in separate sections, and consider what evidence has been obtained in the dissertation, and formulate (tentative) answers to the questions.

The first question, or rather group of questions, relates to the general issue of what determines the success of the communication between speaker and hearer. Given that Chinese is not related to English but that Dutch and English are closely related West Germanic languages, we would expect Dutch speakers and hearers of English to be more successful in the communication process than Chinese interactants. This leads to the following questions:

1. Is it true that speaker/hearers with an L1 that is close to the target language have an advantage over learners with a more distantly related L1?

In order to answer this question we will have to review the scores obtained for Chinese, Dutch and American speakers (averaged over listener groups) and listeners (averaged over speaker groups) at each of the six linguistic levels tested, i.e. vowels, consonants, clusters, words in nonsense sentences, and words in low and high-predictability meaningful sentences.

2. To what extent do separate tests at the lower levels (vowels, consonants, clusters) and at the higher levels (word recognition in nonsense sentences, and low/high predictability meaningful sentences) contribute independent information to the measurement of mutual intelligibility?

And related to question (2) there is the complementary question:

3. Can word recognition be predicted from success in identification of vowels, consonants and clusters at the lower level? What, more generally, is the correlation between the various types of test results?

To answer these two questions we will run regression analyses in which we enter the scores on the lower-order tests, i.e. vowels, consonants and clusters, as predictors and the three types of word recognition scores as criterion variables. We will first run the analyses in an integrated fashion across all 108 listeners involved in the experiments, without compartmentalizing the scores per listener nationality. As a

next step in the analysis we will run regression analyses for each listener nationality separately.

4. Which tests are most successful in discriminating the better from the poorer listeners?

The discriminatory power of the six tests employed in the present study can be defined as the ratio of the between-group and the within-group variance of the identification/recognition scores for the three listener groups, i.e. the F-ratio in one-way analyses of variance with the native language of the listener as the factor. We will supplement this quantification with results of Linear Discriminant Analyses, which we claim provide a better indication of the discriminatory power of the tests at issue.

In the next set of questions we target the predictability of performance scores from either (i) a linguistically motivated a priori contrastive analysis of the sound systems of source and target language of the speakers and listeners, or (ii) an acoustical analysis of the vowel productions of the three speaker/hearer groups involved in our study. Obviously, the latter analysis makes sense only for the vowel identification part of our results.

5. Can vowel and consonant errors/confusions be predicted from a contrastive analysis of the sound systems of source and target language?
6. Can vowel perception and confusion structure be predicted from an acoustical analysis? Does an LDA on F1, F2 and duration measurements yield the same types of errors as in human perception?

The last set of questions relates to the role of speaker and listener nationality (or language background) in determining the success of the communication process.

7. Which factors contribute most to mutual intelligibility? Is the quality of the speaker more or less important to the effectivity of the communication process than the quality of the listener?
8. Is the native listener always the best performer?
9. Do our results support the hypothesis that native/interlanguage benefit exists?

10.2 Effect of genealogical relationship between source and target language

The first question we will address here is whether our experiments support the general hypothesis that L2 learners with a native language that resembles the L2 in many ways, as a result of a close genealogical relationship such as exists between Dutch and English, have a better proficiency in the L2 than learners whose native tongue is not related to the target language, as is the case for Chinese learners of English. Figure 10.1 is a summary of the scores obtained by Chinese, Dutch and American listeners (accumulated over speakers, left-hand panel A) and by Chinese,

Dutch and American speakers (accumulated over listeners, right-hand panel B) for each of the six tests administered in our research.

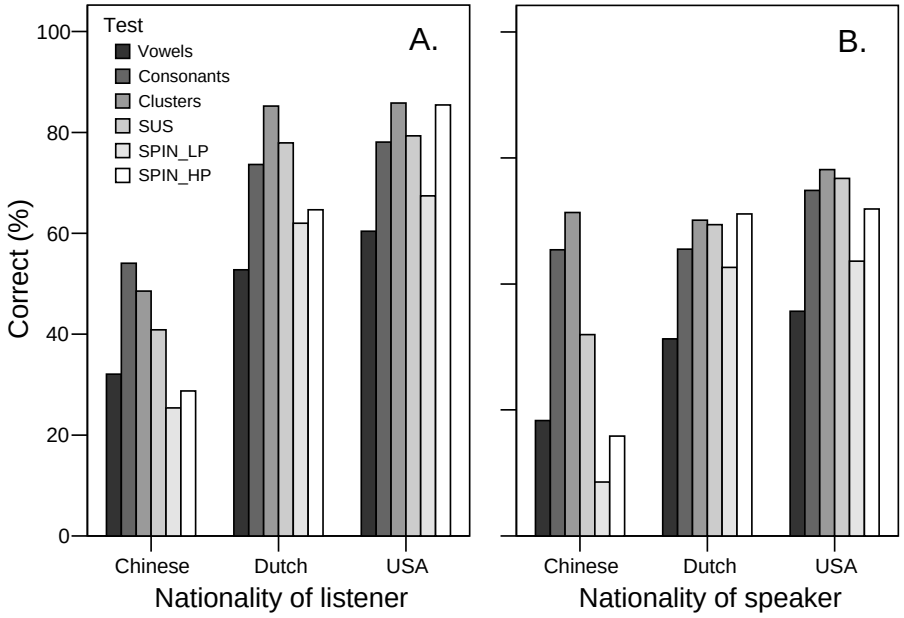


Figure 10.1. Summary of test scores obtained on six tests, broken down by nationality of listener (panel A) and by nationality of speaker (panel B).

The results show quite clearly that, overall, the difference between the Dutch and the American listeners and speakers is smaller than that between the Chinese and the Americans. Generally, also, the effect of listener nationality is much larger than the effect of speaker nationality (for a more detailed analysis of this difference, see § 10.6.1). However, there is substantial interaction between the role of the interactant and the type of test employed. Chinese listeners are much poorer than Dutch listeners. This is also true of Chinese speakers, who are generally poorer than Dutch speakers, except in two tests, i.e. consonant and cluster identification: here the two speaker groups are roughly equal.

These overall conclusions are supported by a three-way ANOVA with the six tests, speaker nationality and listener nationality as fixed factors. The effect of listener nationality, $F(2, 1890) = 2058.8$ ($p < .001$), is much larger than that of speaker nationality, $F(2, 1890) = 643.8$ ($p < .001$). The overall scores averaged over all six tests are 38% for the Chinese listeners, 69% for the Dutch listeners and 76% for the American listeners. The three listener nationalities differ significantly from each other by a Scheffé post-hoc test. The effect of the factor test as such is not relevant but the interaction of test and listener nationality is significant, $F(10, 1890) = 36.2$ ($p < .001$).

On the basis of these results we tentatively conclude that, indeed, genealogical proximity between source and target language is a considerable advantage for the L2 learner. It should be pointed out, however, that other factors may have contributed to the overall effect, such as the possibly greater exposure of Dutch learners to English through the media and the potentially better quality of the pronunciation of English by Dutch secondary-school teachers.

10.3 Correlations among tests at various linguistic levels

The second set of questions I raised in § 10.1 relates to the correlations among the six tests we used in our research. To what extent does each test contribute independent information to the quality assessment of an individual listener? Table 10.1A contains a correlation matrix of the scores of each of the six tests. Since the overall performance on the tests depends strongly on the nationality of the listener (see above), I have also computed correlation matrices for each of the three nationalities separately. It is to be expected that the correlation coefficients drop considerably as a result of the breakdown by listener nationality.

Table 10.1. Correlation coefficients r for all six tests for three listener groups combined (A) and for each listener nationality separately (B-C-D). Each cell contains 324 (A) or 108 (B-C-D) measurement points. Coefficients are significant for $r \geq .170$ ($p < .05$) and $r \geq .248$ ($p < .01$).

A. All listener groups

	Vow	Cons	Clust	SUS	S-LP
Cons	.744				
Clusters	.701	.815			
SUS	.770	.692	.731		
SPIN-LP	.736	.612	.563	.814	
SPIN-HP	.754	.679	.648	.817	.814

B. Chinese listeners

	Vow	Cons	Clust	SUS	S-LP
Cons	.253				
Clusters	.320	.596			
SUS	.248	.162	.246		
SPIN-LP	.159	−.156	−.272	−.096	
SPIN-HP	.164	.005	−.054	.061	.417

C. Dutch listeners

Cons	.604				
Clusters	.550	.721			
SUS	.629	.421	.440		
SPIN-LP	.632	.397	.386	.819	
SPIN-HP	.661	.558	.525	.788	.784

D. American listeners

Cons	.666				
Clusters	.514	.650			
SUS	.689	.430	.250		
SPIN-LP	.700	.556	.380	.862	
SPIN-HP	.646	.398	.285	.822	.754

When the three listener groups are combined, correlations among the six tests are substantial, with r-values between .563 and .817. This means that the six tests provide parallel information to some extent, with overlap between 37 and 67% (i.e. the square of the correlation coefficient). It also means that there is room for

improvement such that two or more tests provide more information as to the quality of an individual listener than each test on its own.

The correlations differ considerably when the results are broken down by separate listener nationalities. Correlations remain high for the American listeners, with r ranging between .250 and .862, as well as for the Dutch listeners, with values between .386 and .819, but they are low for the Chinese group, with r-values between .005 and .596.

Generally, the highest correlations are found for pairs of word recognition tests, i.e. the SUS test and either version of the SPIN test. Correlations are also high for consonant and cluster results.

We will now consider how well the listener's performance on the higher-order listening skills can be predicted from the results obtained by the same individual on the lower-order skills, i.e. the identification of vowels and consonants (and clusters). In order to answer this question we performed multiple linear regression analyses with the word recognition tests as criterion variables, and with the three lower-order tests as predictors, which were entered into the analyses simultaneously. We computed the results for the three listener groups combined (with the risk of inflated results) as well as for each of the listener nationalities separately. The results of the twelve analyses are summarized in Table 10.2, which lists both multiple R and R^2, as well as the beta weights for each of the three predictors.

Table 10.2. Summary of multiple regression analyses predicting word-recognition test results from identification scores on vowel, consonant and cluster identification tests, across all listeners combined (N = 108) and for each listener nationality separately (N = 36 per nationality). Significant beta weights for predictors are indicated by an asterisk.

Listener group	criterion	R	R^2	β_V	β_C	β_{CC}
All	SUS	.816	.666	.496*	.032	.358*
	SPIN-LP	.742	.551	.625*	.134	.016
	SPIN-HP	.778	.605	.528*	.179*	.132*
Chinese	SUS	.304	.092	.188	.005	.183
	SPIN-LP	.376	.141	.275*	−.017	−.349*
	SPIN-HP	.201	.040	.199	.039	−.141
Dutch	SUS	.639	.409	.563*	−.028	.150
	SPIN-LP	.634	.402	.606*	−.014	.063
	SPIN-HP	.697	.485	.484*	.165	.139
American	SUS	.700	.490	.746*	.034	.156
	SPIN-LP	.711	.506	.601*	.190	−.052
	SPIN-HP	.649	.421	.693*	−.030	−.521

Given that the simple correlation coefficients were higher (see Table 10.1) when the three listener groups are combined (N = 108) than when computed for separate listener groups (N = 36), it comes as no surprise that the multiple R values in the regression analyses are higher for the combined listener groups than for each group

separately. Also, predictions of word recognition scores are more successful for the Dutch and American listener groups than for the Chinese group. If fact, no significant R-values were found for the Chinese listeners when the criterion was the SUS or the SPIN-HP score. R was significant for the SPIN-LP result, but here the significant contribution of the consonant cluster identification score was negative, indicating that a poorer identification result correlated with a better word-recognition score.

For the Dutch and the American listeners, the vowel identification scores carry much more weight in predicting word-recognition performance than either the correct identification of simplex consonants or of consonant clusters. This could be construed as an indication that word recognition depends more on vowels than on consonants, which would be in contradiction with results from the literature that suggest that word recognition in English depends more on correct consonant than vowel identification (e.g. van Ooijen, 1994 and references therein). However, this conclusion should be viewed with some caution. First, the contribution of the predictor with the highest simple correlation with the criterion is inflated, since the second-best predictor is then stripped of its intercorrelation with the best predictor. Also, the individual vowel scores have a greater variance than the consonant identification scores, so that the smaller contribution of the consonant (cluster) scores may be the result of a restricted range effect.

Interestingly, for the American listeners only, the prediction of word recognition from lower-order skills is better when the target words do not benefit from contextual predictability. Prediction of word recognition is poorest for the SPIN-HP sentences. I suggest that this is because the American listeners, and only these (see Chapter nine), strongly rely on top-down information obtained from preceding words when recognizing the last word in the sentence, leaving less room for a contribution of bottom-up skills such as vowel and consonant identification.

10.4 Discriminatory power of tests at various linguistic levels

In this section we will consider the question which of the six tests we used in our study affords the best separation between the three listener groups. Assuming for the moment that American native listeners should be superior to all non-native listeners, and that L2 learners with a native language that is genealogically close to the target language (i.e. Dutch listeners) should do better than learners with a non-related L1 (i.e. Chinese listeners), we would expect tests to be able to differentiate between these three types of listener.

In order to answer this question I computed, for each of the six tests employed, the overall score of each individual listener, i.e. accumulated over all items in the test, and over all six speakers. I then ran separate one-way ANOVAs for each listening test, with listener nationality as a single fixed factor.[1] The magnitude of the

[1] These one-way ANOVAs were also run in the preceding Chapters six through nine. Table 10.3 is simply a summary of earlier results. In the SUS test the missing Chinese listener #17 (cf. Chapter eight) was given a mean value but adjusted such that the value reflected this listener's overall ranking on the other tests (mean z-score across all valid test scores).

F-ratio may serve as a first approximation of the discriminatory power of the test. The results can be seen in Table 10.3, which specifies the F-ratio for the six tests, as well as the grouping that can be made among the three listener nationalities on the basis of the Scheffé post-hoc procedure ($\alpha = 0.05$).

Table 10.3. F-ratio of effect of listener nationality and post-hoc grouping for each of six listening tests employed in this study. Percent correct classification by Linear Discriminant Analysis is indicated in the rightmost column (see text).

Test	F-ratio	Post-hoc grouping	LDA % correct
1. Vowels	98.0	CN, NL, US	65.7
2. Consonants	78.7	CN, {NL+US}	71.3
3. Clusters	153.1	CN, {NL+US}	68.5
4. SUS sentences	428.3	CN, {NL+US}	69.4
5. SPIN-LP sentences	200.4	CN, {NL+US}	71.3
6. SPIN-HP sentences	324.0	CN, NL, US	85.2

The results are not immediately interpretable. The test that reveals the largest effect of listener nationality is the SUS test but in spite of the large F-ratio this test fails to discriminate between Dutch and American listeners. The test with the second-largest F-ratio, based on the scores obtained for the SPIN-HP sentences, affords a better separation of the three groups. More generally, it appears that the discriminatory power of the tests based on word recognition is better than that of phoneme identification tests.

Since it is difficult to interpret the results in Table 10.3, I made a second attempt at establishing the discriminatory power of the six tests. This time I ran Linear Discriminant Analyses (LDAs) for each of the six tests. In the LDA I predicted listener nationality from the test scores, and computed percent correct classification across all 108 listeners (three nationalities represented by 36 listeners each). Clearly, the higher the percentage of correctly classified listener nationalities, the better the discriminatory power of the test at issue. The results of the LDA are presented in the rightmost column of Table 10.3. This time it is quite obvious that the greatest discriminatory power is attained by the high-predictability SPIN sentences. The 108 listeners are correctly classified for nationality in 85% of the cases, which is 15 percentage points better than the second-most sensitive test, i.e. the low-predictability SPIN test, with 71% correct classification. The confusion matrix for the automatic classification of the three listener nationalities from the results of the SPIN-HP sentences is as in Table 10.4.

Table 10.4 shows that the Chinese and the American listeners are generally classified correctly with less than 10% error. The Dutch listeners, with SPIN-HP scores in between those of the Chinese and American listeners, are incorrectly classified in nearly 40% of the cases. Their performance overlaps more with that of the American listeners than with that of the Chinese group, with the result that incorrect classification is asymmetrically distributed with roughly a 2:1 bias towards the American group.

Table 10.4. Confusion matrix of listener nationality predicted from SPIN-HP test scores of individual listeners. N = 36 listeners (= 100%) per nationality. Correct classifications are along the main diagonal, indicated in bold face.

Nationality of listener	Predicted Group Membership			Total
	Chinese	Dutch	USA	
Chinese	**97.2**	2.8	0.0	100.0
Dutch	11.1	**63.9**	25.0	100.0
USA	0.0	5.6	**94.4**	100.0

Interestingly, a similar analysis of SUS scores using the sentence as the scoring unit (see § 9.5) yielded only 69% correct classification of listener group. Chinese listeners were perfectly classified (100% correct), as they have very low sentence-based SUS scores, but there was considerable confusion between Dutch and American listeners (27 and 48% correct classification, respectively). In fact, the discrimination of the three listener groups was virtually identical for the word-based and sentence-based scoring methods of the SUS test.[2]

These results confirm the rather more intuitive impression of the sensitivity of the tests, as expressed in the earlier chapters. On the basis of the above result we would – once again – recommend that SPIN-HP sentences be used for fast and sensitive listening ability testing in the area of applied linguistics.

10.5 Predicting performance

In this section we will examine the results of our experiments in order to determine how well success in the communication from speaker to listener can be predicted, either from a structural comparison of (aspects of) the disparate sound systems of the speaker and the listener (contrastive analysis) or from the acoustic structure of the sounds in the L1 and in the L2. We know from the literature (see Chapter two) that contrastive analysis of the sound systems of L1 and L2 often fails to make the right predictions but at least we will try to determine how (un)successful the predictions are. Predicting vowel identification from acoustical analyses of the vowels in L1 and L2 is a more promising approach since it uses detailed and fine-grained acoustical information that the traditional, basically impressionistic, contrastive analysis has no access to.

[2] The one-way ANOVA for the sentence-based SUS scores yielded an F-ratio of $F(2, 105) = 326.7$ ($p < .001$), which is in fact poorer than the result reported in Table 10.3 for the word-based SUS scores. The post-hoc Scheffé test indicated that only the Chinese listeners differed from the Dutch and American listeners, who did not differ from each other.

10.5.1 Contrastive analysis

In Chapter three we reviewed an admittedly dated view on foreign language learning that states that shared phones between source and target language have positive transfer, i.e. do not cause a learning problem (Flege's identical sounds). The target language may also have phonemes that do not occur in the source language; these would cause an initial learning problem for the foreign language learner but these problems will be overcome with sufficient exposure and practice (Flege's 'new sounds'). There is a third category of comprising sounds that are almost the same between source and target language but differ in small but noticeable phonetic feature. These so-called similar sounds will be the most persistent sources of error. The categorization of English vowels and consonants in terms of identical, new and similar sounds has been given in Chapter three in Table 3.6 for vowels and in Tables 3.7-8 for consonants (Dutch and Chinese learners, respectively). It is not entirely clear how this classification translates into predictions of specific confusion patterns for vowels and consonants. To reduce the complexity of the analytic problem, I will restrict the analysis to only the communication of vowels and consonants between one non-native learner group and native speakers, that is, we will only consider four combinations of speaker and listener nationalities, viz. Chinese-American (and vice versa) and Dutch-American (and vice versa). I will assume that identical sounds are never a problem but that the production or perception of any English sound in the new or similar category will in some way result in perceptual confusion between the target sound and its immediate competitors, i.e. will lead to lower percentage of correct identifications of the target sound.

To simplify matters further, I decided to operationalize production errors as perceptual confusions obtained when the speakers are non-native and the listeners are native. Conversely, perception errors are defined as confusions found for non-native listeners when they are exposed to native sounds. In the following tables I have listed percent correct identification of all the vowels and consonants of English separated into three categories, i.e. identical, new, and similar (as defined in Tables 3.6-8) for each of the four possible combinations of native and non-native listener and speaker nationalities. Of course, the classification of target sounds in terms of the three categories differs depending on the non-native language involved.

No similar consonants exist between Mandarin and English. This category therefore remains empty. As a result of these missing data no two-way repeated-measures ANOVA can be performed over all data. Instead we ran separate one-way repeated-measures ANOVAs on each column in Table 10.5.

The effect of type of learning problem for vowels spoken by Chinese learners of English is highly significant by a one-way ANOVA with problem type as a within-listener factor, $F(2, 59.3, \text{Huyhn-Feldt corrected}) = 18.9$ ($p < .001$). All differences between pairs are significant by paired t-tests, even though the difference between identical and similar sounds is significant only in one-tailed testing (assuming that identical sounds should be transmitted more successfully than either similar or new sounds). For Dutch-accented vowels the effect of learning problem is also highly significant, $F(2, 70) = 52.8$ ($p < .001$). Paired t-tests indicate that every difference between pairs of means is significant at $p < .05$.

Table 10.5. Percent correctly perceived vowels and consonants produced by Chinese and Dutch learners of English and perceived by American listeners.

Type	Speaker nationality			
	Chinese		Dutch	
	Vowels	Consonants	Vowels	Consonants
Identical	51	83	82	76
Similar	44	---	55	90
New	28	60	46	56

The effect of learning problem for Chinese-accented consonants is significant by a paired t-test, $t(35) = 15.2$ ($p < .001$). For Dutch-accented consonants the overall effect of learning problem is significant, $F(2, 70) = 114.6$, with significant differences between each pair.

The overall picture that emerges from table 10.5 is that identical sounds are transmitted from speaker to listener more successfully than similar sounds. New sounds are least successfully transmitted.

This finding is in partial conflict with the predictions made by Flege's Speech Learning Model. The model is supported by the results in so far as identical sounds are indeed transmitted most successfully. However, counter to the model's prediction, it is not the case that new sounds are less problematic than similar sounds. The latter result does not necessarily mean that the SLM is wrong. Quite likely, our learners of English have not had enough exposure to native English in real-life communicative situations to discover that they need certain new sound categories.

Let us now briefly examine perception problems on the part of Chinese and Dutch learners, when confronted with American vowels and consonants. The results are presented in Table 10.6.

Table 10.6. Percent correctly perceived vowels and consonants produced by American speakers and perceived by Chinese and Dutch learners of English.

Type	Listener nationality			
	Chinese		Dutch	
	Vowels	Consonants	Vowels	Consonants
Identical	39	63	64	81
Similar	22	---	38	88
New	20	50	60	71

Vowels spoken by American native speakers are correctly perceived by Chinese listeners in the order identical, similar and new with 39, 22 and 20% correct, respectively. The effect of learning type is significant by RM ANOVA, $F(2, 70) = 14.6$ ($p < .001$); however, similar and new sounds do not differ from each other by a paired t-test. For Dutch listeners the effect of learning type is also significant, F(2,

65.8 Huyhn-Feldt corrected) = 31.6 ($p < .001$); identical and new sounds do not differ from each other by a paired t-test.

New consonant sounds are more difficult to perceive than identical sounds. For consonants perceived by Chinese learners the difference is significant by a paired t-test, $t(35) = 5.5$ ($p < .001$). For Dutch listeners the effect of learning difficulty is significant by a one-way RM ANOVA, $F(2, 65.6$ Huyhn-Feldt corrected$) = 26.7$ ($p < .001$). Paired t-tests show that all pairs of sound types differ from each other. Note, however, that similar sounds are perceived more adequately than either identical or new sounds. SLM would predict that similar and identical sounds do not differ perceptually from the point of view of the learner; both types of target sound are believed to be equivalent to sounds in the source language.

Again, we may conclude that identical sounds are transmitted more successfully from (American native) speaker to non-native listener than either similar or new sounds. These latter two do not differ systematically. We have to conclude, provisionally, that Flege's SLM is only partially successful in predicting learning problems. Especially the predicted difference between similar and new sounds could not be found in the results, so that SLM in the present case does not do any better than Lado's older transfer model, which predicts positive transfer (no learning problem) for identical sounds, and negative transfer for any target sounds that do not occur in the source language (negative transfer).

10.5.2 Predicting vowel perception from acoustic analyses

Now that we have seen that contrastive analysis is only moderately successful at predicting problems in production and perception of contrasts in a foreign language, let us consider an alternative possibility of predicting perceptual confusions by listeners with a particular L1 (Chinese, Dutch, American English) who are exposed to English sounds, specifically monophthongal vowels, spoken with a Chinese, Dutch or American accent. Can we actually predict the results we obtained in Chapter six for human perception of these vowels from the results of automatic classification (as done in Chapter five) of the same sounds by an algorithm such as Linear Discriminant Analysis (LDA)?

I ran three LDAs. In the first, the classification algorithm was based on the discriminant functions derived from the Chinese-accented vowel tokens; it was applied to all vowel tokens, including the Dutch-accented and the General American tokens. Here we expect the Chinese-accented tokens to be classified better, since the training data are the same as the test data. This would be a modeling of the interlanguage benefit for the Chinese speaker-listener group. When the Chinese-based discriminant functions are applied to Dutch and American-accented vowel tokens, we simulate the situation where a Chinese listener has to identify these vowel tokens. Here we predict poorer classification results. In the second application of the LDA, the training data were the Dutch-accented vowel tokens; the difference in percent correct vowel identification between the Dutch tokens and the Chinese or American tokens would be a quantitative approximation of the interlanguage benefit for the Dutch speaker-listener combination. In the third run the LDA was trained on the American vowel tokens. The native-language benefit for the American speaker-

listener combination should show up in superior classification of the American tokens.

The LDAs were run on the ten monophthongs of American English only (see Chapter five), i.e. excluding the vowels followed by /r/, and excluding diphthongs. The vowel in *hawed* was also omitted from the analysis, as it typically merges with the vowel in *hod*. This selection then leaves the vowels in the words *heed*, *hid*, *hayed*, *head*, *had*, *hud*, *hod*, *hoed*, *hood* and *who'd*. For the sake of comparability the vowel identification by human listeners, as reported in Chapter six, was recomputed such that the set of response vowels was identical to the set of stimulus vowels, i.e. the same set of ten. Within the restricted set this selection resulted in only minor discrepancies with the full vowel identification results reported in Chapter six.

Figure 10.2A-B presents percent correctly identified vowel tokens, in two panels. The left-hand panel A displays the classification by human listeners, broken down by listener nationality and within each cluster by nationality of the speaker. The right-hand panel B presents the percentages of correct classification of vowel tokens by LDA broken down first by the L1 of the speakers who supplied the training data (mimicking the effect of listener), and with the clusters broken down further by the native language of the speaker.

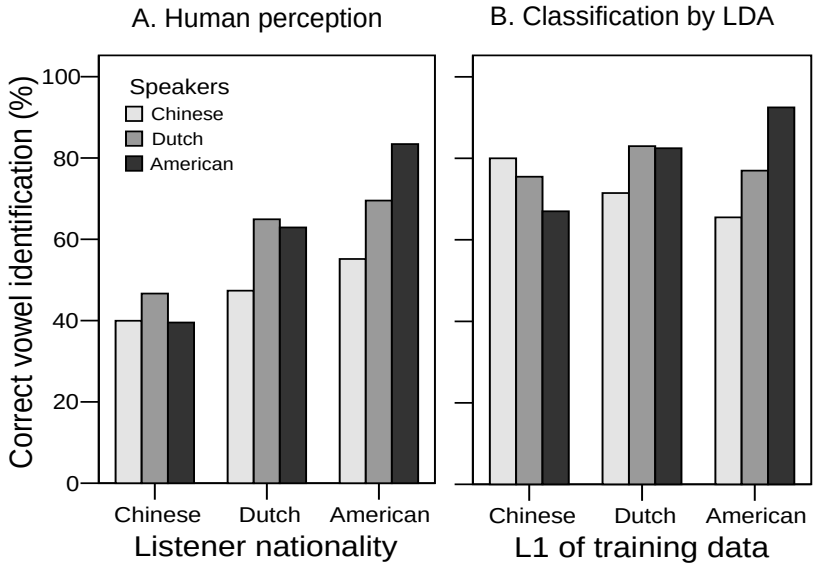


Figure 10.2. Correct identification (%) within a restricted set of ten vowels by human listeners broken down by L1 of listener and of speaker (panel A), and by Linear Discriminant Analysis broken down by L1 of training data and by nationality of speaker (panel B).

As said, the human vowel identification (panel A) shows virtually the same scores as for the full vowel data reported in Chapter six. This indicates that the performance of the speaker and listener groups is not unduly affected by the selection of the ten

target vowels from the larger set of 19. The automatic classification by LDA (panel B), on the basis of acoustic properties of vowel tokens produced by ten male and ten female speakers (after z-normalisation within individual speakers of vowel duration and Bark-transformed first and second formant values; see Chapter five), yields higher percent correct scores. If we abstract from the absolute difference in scores, we may observe that the configuration of scores for American native and Dutch listeners and their respective simulation in the LDA are quite similar. The configuration for the Chinese listeners, however, is rather different. Not only is the mean percent correct classification much better in the LDA, also the interlanguage benefit is so large here that the automatic classification of Chinese-accented vowels from Chinese-accented training data attains the highest score, even in absolute terms. This finding indicates, once more, that there is a lot of information in the Chinese-accented English vowels which might be used profitably in the process of vowel identification. Clearly, Chinese listeners are better tuned in to this information, but even they do not exploit the acoustic cues to the maximum.

The configuration of scores in panels A and B of figure 10.2 are correlated at r = 0.698 (p = 0.036). Figure 10.3 is a scatterplot of the nine pairs of scores obtained in human and machine identification of the ten English monophthongs.

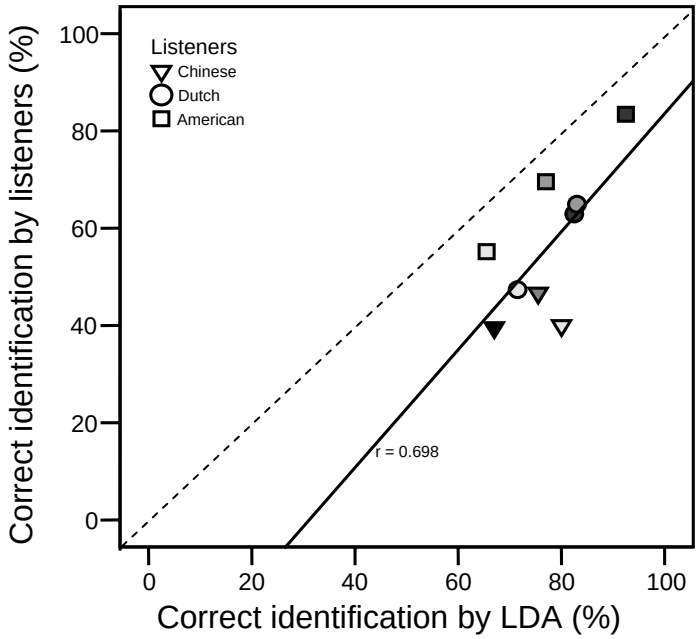


Figure 10.3. Correct classification by Linear Discriminant Analysis plotted against human perception of the same vowel tokens. Nationality of listeners (and L1 of the speakers supplying the training data for the LDA) is indicated in the legend. Nationality of the speaker group is coded in the grey shades of the markers (black: American speakers, dark grey: Dutch speakers, light grey: Chinese speakers).

The figure shows that high correlations exist between percent correct vowel identification by LDA and by human listeners as long as the listeners are not Chinese. Correlation between LDA and human perception is weaker for the Chinese listeners, as the Chinese listeners fail to use substantial acoustic information in the Chinese-accented vowel tokens, which is picked up by the LDA.

The above analysis has shown that cross-language vowel perception can be predicted from an acoustical analysis of the (native and non-native accented) vowel tokens produced by speakers of the nationalities involved, at least as long as we only want to predict mean percent correct classification across the vowel inventory. In what follows now, we will examine to what extent the human identification of individual vowels within the inventory can be predicted by LDA.

Table 10.7 presents correlation coefficients computed for pairs of correct vowel identification scores obtained from human listeners and from LDA, using – as before – the nationality of the speakers as a simulation of the human listener – in each combination of speaker and listener nationality. The correlation coefficients are based on ten pairs of scores (for ten vowel types) in the cells of the matrix, on 30 pairs for the marginals and on 90 pairs for the overall dataset. Both Pearson's r and Spearman's rho were computed.

Table 10.7. Correlation coefficients (r: upper, rho: lower line per cell) for human and machine correct vowel identification in nine combinations of speaker and listener nationality.

Speakers	Listeners / LDA train set			
	Chinese	Dutch	American	All
Chinese	**.181** **.146**	–.019 .018	–.144 .079	–.135 .038
Dutch	.387 .494	**.673**** **.803****	.689* .782**	.561** .598**
American	.159 .122	.166 .423	**.310** **.378**	.501** .552**
All	.234 .180	.278 .404*	.424* .565**	.332** .404**

The table shows, first of all, that the non-parametric rho coefficients are higher than parametric r. This indicates that the relationship between correct human vowel identification scores and the results of the LDA is not linear.

Generally, the success in correct vowel identification by human listeners is better predicted from the LDA if the listener group (and the nationality of the training set for the LDA) is the same as that of the speakers; this is, again, the influence of the interlanguage language benefit. The correlation between the LDA and human vowel identification is poor and insignificant for the Chinese listeners and speakers. It is better for American listeners and speakers, and best for the Dutch speakers and listeners. The results indicate that, at least for Dutch and American combinations of speakers and listeners, the LDA provides a rough indication of which vowels will be error-prone and which will be less so.

The strictest test on the adequacy of predicting problems in cross-language vowel identification would be to use the LDA to predict specific vowel confusion errors. Table 10.8 presents a survey of the top-ten vowel confusions as found in the human identification of the ten stimulus vowels broken down by nationality of speakers and listeners. In the same table I have listed which of these confusions was predicted to be in the top ten by the corresponding LDA. Depending on the particular combination of speaker and listener nationality, the LDA successfully predicts a confusion pair in the top-10 list between three (Chinese listeners, American speakers) and eight (American listeners, Dutch speakers) times. Even in the poorest speaker-listener combination the result is significantly better than chance, using Cohen's kappa as the measure of agreement in a series of 90 binary judgments (top-10 ~ lower) by two independent judges (human perception ~ prediction by LDA) with $\kappa = 0.212$ ($p = 0.044$). For the most felicitous speaker-listener combination we obtain $\kappa = 0.775$ ($p < 0.001$). The κ-values and their probabilities have been indicated for all nine combinations of speaker and listener nationalities in Table 10.8.

We conclude that cross-linguistic human perception of vowels can be predicted, with varying success but invariably (much) better than chance, from the acoustic properties of the vowel tokens as produced by native speakers and foreign learners, using Linear Discriminant Analysis. This technique has been used before but only in the comparison of two languages, e.g. English and German (Strange et al. 2004) or Japanese and English (Strange 1999) (see also Chapter two). We have now shown that the technique may also be used to predict (part of) the confusion structure of (English) vowels in non-native communication with either or both speaker and hearer having a different language than English and even a different native language.

Table 10.8. Ten most frequent vowel confusions (Cnf) for nine combinations of listener (Lis) and speaker nationality. The columns marked H list percent confusion by human listeners, L is percent confusion found by LDA (see text), R is the results in terms of success (h = hit) or failure (m = miss); Ncor presents the number of confusion types correctly predicted by the LDA in the top ten of human vowel confusions. Kappa and associated p-values are indicated.

	Chinese speakers					Dutch speakers					American speakers				
Lis	Cnf	H	L	R	Ncor	Cnf	H	L	R	Ncor	Cnf	H	L	R	Ncor
CN	æ>ɛ	46	30	h	5	æ>ɛ	57	30	h	5	iː>ɪ	55	15	h	3
	ɛ>æ	42	20	h		uː>ʊ	39	30	h		uː>ʊ	37	55	h	
	uː>ʊ	34	10	h		ɛ>æ	31	15	h		ʌ>ɔ	29	85	h	
	ʊ>uː	25	25	h		ʌ>ɔ	30	60	h		æ>ɛ	43	5	m	
	ɔ>ʌ	23	15	h		ʊ>uː	27	20	h		ɪ>ɛ	31	0	m	
	ɪ>iː	43	0	m		iː>ɪ	49	5	m		ɔ>ʌ	28	0	m	
	iː>ɪ	39	5	m		eː>iː	21	0	m		eː>ɪ	26	5	m	
	ʌ>ɛ	33	5	m	κ = .437	eː>ɪ	21	0	m	κ = .437	ʊ>ʌ	26	0	m	
	oː>ʊ	28	0	m	p <. 001	ɪ>iː	20	0	m	p <. 001	eː>ɛ	25	0	m	κ = .212
	ʌ>æ	18	0	m		ʊ>ʌ	17	0	m		oː>ɔ	24	5	m	p = .044
NL	ʌ>æ	67	10	h	7	æ>ɛ	54	25	h	7	uː>ʊ	40	35	h	5
	ɪ>iː	64	10	h		uː>ʊ	46	20	h		iː>ɪ	21	5	h	
	uː>ʊ	51	35	h		ʊ>uː	34	40	h		ʊ>ʌ	15	15	h	
	æ>ɛ	42	30	h		ɛ>æ	27	15	h		ɔ>ʌ	12	20	h	
	ɛ>æ	40	45	h		ʌ>ɔ	22	10	h		ʌ>ɔ	11	5	h	
	ʌ>ɛ	18	15	h		oː>ɔ	18	5	h		ʌ>ʊ	42	0	m	
	ɔ>ʌ	17	50	h		oː>ʊ	13	5	h		æ>ɛ	39	0	m	
	oː>ʊ	43	0	m	κ =.663	ʌ>æ	45	0	m	κ =.663	eː>ɛ	30	0	m	κ = .437
	ʊ>uː	28	5	m	p <.001	oː>uː	10	0	m	p <.001	ɔ>æ	21	0	m	p < .001
	ɛ>ɪ	20	0	m		iː>ɛ	10	0	m		ɛ>æ	13	0	m	
US	ʌ>æ	76	5	h	4	æ>ɛ	75	30	h	8	ʌ>ʊ	45	5	h	5
	ɪ>iː	54	45	h		ʊ>uː	48	65	h		eː>ɪ	10	5	h	
	ʊ>uː	49	45	h		ʌ>ɔ	42	15	h		oː>ɔ	8	5	h	
	ɛ>æ	24	40	h		uː>ʊ	17	10	h		ɔ>ʌ	5	5	h	
	uː>ʊ	43	0	m		ɔ>ʊ	11	10	h		ʌ>ɔ	5	10	h	
	oː>ʊ	31	0	m		ɛ>æ	10	15	h		eː>ɛ	13	0	m	
	ɛ>eː	29	5	m		ɔ>oː	9	15	h	.775	ɔ>æ	11	0	m	
	iː>ɪ	25	0	m		ɔ>ʌ	6	15	h	.001	uː>ʊ	11	0	m	
	æ>eː	11	0	m	κ = .325	ʌ>æ	21	0	m	κ =	iː>ih	8	0	m	κ = .437
	ɛ>ɪ	10	0	m	p = .002	iː>ɪ	11	0	m	p <	eː>æ	3	0	m	p < .001

10.6 Role of speaker and listener nationality in determining the success of the communication process

10.6.1 Speaker versus listener

The next question we will try to answer is whether the native language of the speaker or that of the listener is more important in predicting the success of the communication between speaker and hearer. In the discussions in Chapters six through nine, a consistent result was that the effect of speaker nationality on the effectiveness of the communication between speaker and listeners was smaller than that of listener nationality. The findings are summarized in Table 10.9, which presents the size of the speaker and listener effects in each of the six tests we administered.

In the columns headed 'Speaker effect' the term is indicated which has to be added to (or subtracted from) the mean score on the test when the speaker is Chinese, Dutch or American. Similarly, the columns headed 'Listener effect' list the increment or decrement that has to be applied to the mean test score for each of the three listener nationalities. The F-ratio is the direct measure of the size of the speaker or listener effect in a two-way ANOVA performed on each of the six tests.

Table 10.9. Summary table of size of speaker and listener effects on effectiveness of communication between speaker and listener.

Test	Speaker effect				Listener effect			
	CN	NL	US	F-ratio	CN	NL	US	F-ratio
Vowels	−10.1	2.9	7.3	77.7	−16.3	4.3	12.0	204.9
Cons.	−3.2	−3.1	6.2	33.4	−14.5	5.0	9.5	185.8
Clusters	−1.8	−3.1	4.9	15.3	−24.6	12.0	12.6	372.4
SUS	−14.1	3.4	10.7	244.5	−25.2	11.9	13.3	716.9
SPIN-LP	−23.1	11.0	12.0	238.5	−26.1	10.4	15.8	312.4
SPIN-HP	−23.8	11.5	12.3	261.3	−30.9	5.1	25.8	506.4

Table 10.9 shows unequivocally that the effect of listener nationality (or: native-language background) is stronger than the effect of speaker nationality. The overriding importance of the listener effect is found in each of the six tests administered in our battery.

In the context of communication in English between speakers and listeners from diverse language backgrounds, it seems that training listeners so as to get tuned in to the peculiarities of some foreign accent in English would be a more fruitful approach to improving non-native communication than training speakers to acquire a better pronunciation. This recommendation would be applicable especially when the number of different foreign accents the listeners have to get used to is limited.

10.6.2 Is the native listener always superior?

The question whether the native listener is always superior to non-native listeners seems trivial at first sight. Nevertheless, we need to consider this question since recent literature makes the claim that under special circumstances it may happen that native listeners are outperformed by foreign learners, specifically if the foreign learner is exposed to English speech produced by someone who has the same native-language background as the foreign listener; the L2 listener may then have an advantage over the L1 listener.

In order to answer this question I have summarized the results of the six tests in the preceding chapters in Table 10.10. This table lists percent correct responses arranged by test in columns and broken down by speaker nationality and then by listener nationality. In total there are 6 (tests) × 3 (speaker nationalities) = 18 conditions for which we may determine whether it is indeed true that American native listeners obtain the highest scores.

Table 10.10. Summary of test results. Percent correct on each of six tests broken down by nationality of speaker and broken down further by nationality of listener. Each mean is based on 36 listeners. The listener group with the best performance is represented in bold face.

Speakers	Listeners	Tests					
		Vowels	Consonants	Clusters	SUS	SPIN_LP	SPIN_HP
Chinese	Chinese	29.7	57.2	52.8	39.3	19.4	16.7
	Dutch	40.3	66.6	78.8	57.1	26.9	33.1
	USA	**44.9**	**72.5**	**82.5**	**59.5**	**39.4**	**57.8**
Dutch	Chinese	33.5	46.8	36.9	39.0	38.9	37.8
	Dutch	59.3	73.7	**87.8**	**86.2**	**81.3**	76.1
	USA	**61.0**	**76.1**	85.7	83.0	67.7	**99.4**
USA	Chinese	33.1	58.2	56.0	44.2	17.9	31.8
	Dutch	58.6	80.6	89.1	90.5	77.8	84.9
	USA	**75.3**	**85.7**	**89.3**	**95.5**	**95.2**	**99.1**

Table 10.10 reveals that in the large majority of the cases the American listeners outperform the other listener groups, i.e. in 15 out of the 18 text × speaker nationality conditions (for the sake of simplicity we ignore here the matter of statistical significance of the difference between the American listeners and the second-best group). In three situations, however, the native listeners do not end up with the highest score. The three situations invariably involve Dutch listeners who respond to Dutch speakers. These, then, are examples of interlanguage benefit in an absolute sense. Such absolute interlanguage benefit is found for the Dutch speaker-listener combination on the cluster identification test, the SUS test and the SPIN-LP test. In the remaining three tests, the difference between the Dutch and the American listeners is very small, and statistically insignificant (see Chapters six and seven) for

the lower-order segment identification tests but the American listeners are vastly superior when it comes to word recognition in meaningful high-predictability context. We reiterate, on the strength of this finding, that the most sensitive and valid test of receptive spoken language proficiency would be the type of test exemplified by the SPIN-HP model (see also § 10.4).

10.6.3 Relative interlanguage benefit

It has been suggested in the recent literature that a situation may arise in which the native listener could be outperformed by non-native listeners. This situation would be found when a non-native listener is confronted with an L2 speaker who has the same native-language background as the listener. In this case, the shared knowledge of the interfering L1 might give the L2 listener an edge over the native listener of the target language. This advantage due to shared speaker-hearer background language has been termed 'interlanguage benefit' (Bent & Bradlow, 2003). We have seen, in the preceding section, that absolute interlanguage benefit does occur, but not in a pervasive manner. It was found in three of the six tests for Dutch speaker-listener combinations, but not in the other three tests, and never for Chinese speaker-listener combinations. Does this mean that in these situations the listeners did not benefit at all from have knowledge of the sound system of the interfering L1? Or can we make a case for a more general view that interlanguage benefit is pervasive, if it is construed not in an absolute but in a relative manner?

In Chapter six I have developed a more sophisticated way of examining the contribution of interlanguage benefit. I suggested that the effect be quantified in a more relative way. Specifically I proposed we compute an expected score for some test by adding to the grand mean the speaker effect and the listener effect. The increments or decrements relative to the mean score in each test, for speaker and for listener nationality, can be found in Table 10.5 above. We then subtract this expected score from the observed score in each speaker-hearer combination. The residual score then expresses the relative interlanguage benefit (for a computational example see § 6.2.1).

Figure 10.4 presents the residual scores as a measure of the native language benefit (for American speaker-listeners) or of interlanguage benefit (for Chinese and Dutch speaker-listeners), for each of the six tests administered. I only present the graph for combinations of speaker and listener groups that share the same L1. The exact complement of this graph (the mirror image reflected around the 0-line) would be obtained for the remaining six speaker-listener combinations.

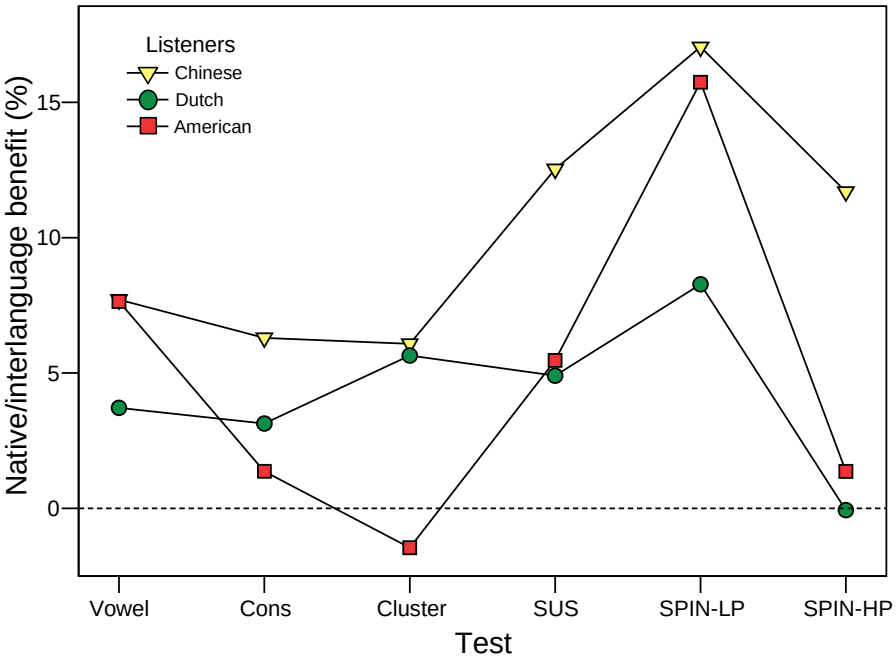


Figure 10.4. Native/interlanguage benefit (percentage points) for Chinese, Dutch and American speaker-hearers of English, for six tests (further see text).

The results show that, with only two exceptions, there is pervasive relative native/interlanguage benefit for each of the six tests. The two exceptions are Dutch listeners in the SPIN-HP test (benefit = 0.1%, i.e. essentially no benefit) and American native listeners in the cluster identification task (a negative residual of −1.5%). In the other 16 situations relative interlanguage benefit is positive. Interestingly, the benefit is consistently largest for the Chinese speaker-listener combination, with a mean of 10.2% across the six tests. The benefit is about half this size for the Dutch and the American speaker-listener groups, with mean values of 4.3 and 5.0%, respectively.

We can only speculate on the reason why the interlanguage benefit should be so much larger for the Chinese speaker-listener combination than for the other two nationalities. It would seem that the Chinese speakers code in their variety of English quite a lot of information that escapes the ear of listeners who are not familiar with the sound structure of Chinese. This is in line with our finding in Chapter five, for instance, where we noted that automatic classification on the basis of the first two formants and duration of the English (monophthongal) vowels was surprisingly successful (ca. 80% correct classification), almost as successful as for the Dutch speakers (ca. 85% correct). We would argue that it is difficult for the Dutch and American listeners to tune in to the subtleties of Chinese-accented English because the Chinese sound system deviates so strongly from that of

Germanic languages. Since the phonetics and phonology of English and Dutch have much more in common, the interlanguage benefit is smaller between these two languages.

A last comment we should make in this context is that there is no difference, in principle, between interlanguage benefit and native language benefit. American listeners benefit from listening to fellow American speakers since both speakers and listeners are thoroughly familiar with the sound system of the native language, as much as the non-native communities are familiar with their respective native sound systems.

By way of conclusion, then, we argue that our experimental results indicate that native and interlanguage benefit is much more widespread than meets the eye. This conclusion hinges on the assumption, which we believe is a correct one, that the benefit should be quantified in relative terms, through linear modeling, rather than in an absolute sense.

References

Abercrombie, D. (1949). Teaching pronunciation. *English Language Teaching*, 3, 113–122.

Adank, P., Heuven, V.J. van & Hout, R. van (1999). Speaker normalization preserving regional accent differences in vowel quality. *Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco*, 1593–1596.

Anderson-Hsieh, J., Johnson, R. & Koehler K. (1992). The relationship between native speaker judgements of non native pronunciation and deviance in segmentals, prosody and syllable structure. *Language Learning*, 42, 529–555.

Anisfeld, M., Bogo, N. & Lambert, W. (1962). Evaluational reactions to accented English speech. *Journal of Abnormal Social Psychology*, 69, 89–97.

Asher, J. J. & Garcia, R. (1969). The optimal age to learn a foreign language. *Modern Language Journa*l, 53, 334–341.

Bansal, R. K. (1966). The intelligibility of Indian English: Measures of the intelligibility of connected speech, and sentence and word material, presented to listeners of different nationalities. Unpublished doctoral dissertation, University of London.

Beckman, M. E. (1986). *Stress and Non-Stress Accent* (Netherlands Phonetic Archives No. 7). Foris. (Second printing, 1992, by Walter de Gruyter.)

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

Bent, T. & Bradlow, A.R. (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America*, 114, 1600–1610.

Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In B. de Boysson-Bardies et al. (eds.) *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*. (pp. 289-304). Amsterdam: Kluwer Academic Publishers.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: a perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (eds.), *The development of speech perception: the transition from speech sounds to spoken words* (pp. 167–224). Cambridge, MA: MIT Press.

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange & J. J. Jenkins (eds.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 171– 204). Timonium, MD: York Press.

Best, C. T., McRoberts, G. W. & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants, *Journal of Experimental Psychology: Human Perception and Performance* 4, 45–60.

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

Bezooijen, R. van (2005). Approximant /r/ in Dutch: Routes and feelings. *Speech Communication* 47, 15–31.

Birdsong, D. (1999). Introduction: Why and why nots of the critical period hypothesis for second language acquisition. In Birdsong, D. (ed.), *Second language acquisition and the critical period hypothesis.* Mahwah, NJ: Lawrence Erlbaum.

Blevins, J. (1985). A metrical theory of syllabicity. PhD dissertation, Massachusetts Institute of Technology.

Bloomfield, L. (1933). *Language*. New York: Henry Holt.

Boersma, P. & Weenink, D. (1996). Praat, a System for Doing Phonetics by Computer. *Report of the Institute of Phonetic Sciences Amsterdam,* 132.

Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In Birdsong, D. (ed.), *Second language acquisition and the critical period hypothesis* (pp. 133–160). Mahwah, NJ: Lawrence Erlbaum.

Bongaerts, T., Mennen, S., & Slik, F. van der (2000). Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as a second language. *Studia Linguistica* 54, 298–308.

Bongaerts, T., Summeren, C. van, Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language, *Studies in Second Language Acquisition* 19, 447–465

Borden, G., Gerber, A. & Milsark, G. (1983). Production and perception of the /r/-/l/ contrast in Korean adults learning English. *Language Learning* 33, 499–526.

Bot., K. de, Gommans, P. & Rossing, C. (1991). L1 loss in an L2 environment: Dutch immigrants in France. In H. W. Seliger & R. M. Vago (eds.), *First Language Attrition*. Cambridge, UK : Cambridge University Press.

Brennan, E. M. & Brennan, J. S. (1981a). Accent scaling and language attitudes: Reactions to Mexican American English speech. *Language and Speech* 24, 207–221.

Brennan, E. M. & Brennan, J. S. (1981b). Measurements of accent and attitude toward Mexican-American speech. *Journal of Psycholinguistic Research* 10, 487–501.

Brière, E. (1968). *A Psycholinguistic Study of Phonological Interference*. The Hague: Mouton.

Broerse, N. N. (1997). Perfect bilinguality – fact or fiction: A comparative study of the perception of checked vowels by early and late English / Dutch bilinguals. Master's thesis, Dept. of English, Leiden University.

Carroll, L. (1872). *Through the looking glass and what Alice found there*. Facsimile edition, New York: Alfred A. Knopf.

Chao, Y. R. (1968) *A Grammar of Spoken Chinese*. Berkeley, CA: University of California Press.

Chen, M. (1985). Beijing Yuyin Yuanliu Chutan: A Preliminary Study of the Origin of Beijing Pronunciation. In Zhigong Zhang, (ed.) *Yuwen Lunji: Collected Essays on Language*. Beijing: Waiyu Jiaoyu yu Yanjiu Chubanshe.

Chen, Y. Robb, M., Gilbert, H. & Lerman, J. (2001). Vowel production by Mandarin speakers of English. *Clinical Linguistics & Phonetics* 15, 427–440.

Cheng, C. C. (1973). A synchronic phonology of Mandarin Chinese. *Monographs on linguistic analysis* No. 4. The Hague: Mouton.

Cheng, R. L. (1966). Mandarin Phonological Structure. *Journal of Linguistics* 2, 135–158.

Chomsky, N. & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.

Clark, H. H. & Clark, C. V. (1977). *Psychology and language*. New York: Harcourt Brace Jovanovich.

Collins, B., Hollander, S. P. den & Rodd, J. (1977). *Accepted English pronunciation*. Apeldoorn: Van Walraven.

Collins, B. & Mees, I. (1981). *The sounds of English and Dutch*. The Hague: Leiden University Press.

Cooper, L. (1932). *The rhetoric of Artistotle*. New York: Appleton-Century-Crofts.

Crawford, W. W. (1987). The pronunciation monitor: L2 acquisition considerations and pedagogical priorities. In J. Morley (ed.), *Current Perspectives on Pronunciation* (pp. 101–121). Washington, DC: TESOL.

Cutler, A. (1983). Speakers' conception of the functions of prosody. In A. Cutler & D. R. Ladd (eds.), *Prosody: Models and Measurements* (pp. 79–92). Berlin: Springer Verlag,.

Doeleman, R. (1998). Native reactions to nonnative speech. PhD dissertation, Tilburg University.

Duanmu, S. (2005). Phonology of Chinese (Mandarin). *Encyclopedia of Language and Linguistics* (2nd edition). Amsterdam: Elsevier Publishing House.

Eckman, F. (1977). Markedness and the Contrastive Analysis Hypothesis. *Language Learning* 27, 315–330.

Eggen, B., & Nooteboom, S. G. (1993). Speech quality and speaker characteristics. In V. J. van Heuven & L. C. W. Pols (eds.), *Analysis and synthesis of speech: Strategic research towards high-quality text-to speech generation* (pp. 279-288). Berlin: Mouton de Gruyter.

Eggen, J.H. (1989). Intelligibility of synthetic speech in the presence of interfering speech. *Speech Communication* 8, 319–327.

Ensz, K.Y. (1982). French attitudes toward typical speech errors of American speakers of French. *Modern Language Journal* 66, 133–39.

Escudero, P. (2005) *Linguistic Perception and Second Language Acquisition: Explaining the attainment of optimal phonological categorization.* LOT dissertation series nr. 113. Utrecht: LOT.

Fayer, J. M. & Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning* 37, 313–326.

Flege, J. E. & Fletcher, K. L. (1992). Talker and listener effects on degree of perceived foreign accent. *Journal of the Acoustical Society of America* 91, 370–389.

Flege, J. E. & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition* 23, 527–552.

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

Flege, J. E. (1987b). The production of 'new' and 'similar' phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics* 15, 47–65.

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America* 84, 70–79.

Flege, J. E. (1995a). 'Second language speech learning: Theory, findings and problems'. In W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in cross-language research* (pp. 233–277). Baltimore: York,.

Flege, J. E. (1995b). Second language speech learning: theory, findings, and problems. In W. Strange (ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research* (pp. 233–277). Timonium, MD: York Press.

Flege, J. E.. Bohn, O. S & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25, 437–470.

Flege, J. E., Munro, M. J. & Mackay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America* 97, 3125–3134.

Flege, J. E., Yeni-Komshian, G. & Liu, S. (1999) Age constraints on second language learning. *Journal of Memory and Language* 41, 78–104.

French, N. R. & Steinberg, J. C. (1947). Factors governing the intelligibility of speech sounds. *Journal of the Acoustical Society of America* 19, 90–119.

Gilbert, J. B. (1980). Prosodic development: Some pilot studies. In R. C. Scarcella & S. D.Krashen (eds.), *Research in second language acquisition* (pp. 110–117). Rowley, MA: Newbury House.

Gimson, A. C. (1980), *An Introduction to the Pronunciation of English* (3rd edition). Edward Arnold, London.

González-Bueno, M. (1997). The effects of formal instruction on the acquisition of Spanish stop consonants. In W. R. Glass & A. T. Pérez-Leroux (eds.), *Contemporary perspectives on the acquisition of Spanish, vol. 2*: *production, processing, and comprehension* (pp. 57–75). Somerville, MA: Cascadilla Press.

Gooskens, C. & Heeringa, W. (2004). Perceptive evaluation of Levenshtein dialect distance measurements using Norwegian dialect data. *Language Variation and Change* 16 189–207.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics* 45, 189–195.

Grosjean, F. (2000). The bilingual's language modes. In J. Nicol (ed.), *One Mind, Two Languages: Bilingual Language Processing* (pp. 1–22). Oxford: Blackwell.

Gussenhoven, C. & Broeders, A. (1976). *The pronunciation of English. A course for Dutch learners*. Groningen: Wolters-Noordhoff-Longman.

Gussenhoven, C. & Broeders, A. (1981). E*nglish pronunciation for student teachers*. Groningen: Wolters-Noordhoff-Longman.

Gynan, S. N. (1985). Comprehension, irritation, and error hierarchies. *Hispania* 68, 160–165.

Haagen, M. J. van der (1998). Ca*ught between Norms: the English Pronunciation of Dutch Learners*. LOT dissertation series nr. 12. Utrecht: LOT.

Hartman, Lawton M. III (1944). The segmental phonemes of the Peiping dialect. *Language* 20, 28–42.

Hayward, K. (2000). *Experimental Phonetics*. Harlow: Pearson Education.

Hazan, V. and Shi, B. (1993). Individual variability in the perception of synthetic speech. *Proceedings of Eurospeech* 1993, Berlin.

Heeringa, W. & Nerbonne, J. (2001). Dialect areas and dialect continua. *Language Variation and Change* 13, 375–400.

Heeringa, W. (2004). *Measuring Dialect Pronunciation Differences using Levenshtein Distance*. Doctoral dissertation. University of Groningen.

Heuven, V. J. van, Kruyt, J. G. & Vries, J. W. de (1981). Buitenlandsheid en begrijpelijkheid in het Nederlands van buitenlandse arbeiders, een verkennende studie [Foreignness and intelligibility of Dutch spoken by foreign workers], *Forum der Letteren* 22, 170–178.

Heuven, V. J. van & Bezooijen, R. van (1995). Quality evaluation of synthesized speech, in W.B. Klein, K.K. Paliwal (eds.), *Speech coding and synthesis* (pp. 707–738). Amsterdam: Elsevier Science,.

Heuven, V. J. van & Sluijter, A. M. C. (1996). Notes on the phonetics of word prosody. In R. Goedemans, H. van der Hulst, E. Visch (eds.) *Stress patterns of the world, Part 1: Background*, HIL Publications (volume 2, pp. 233–269). The Hague: Holland Institute of Generative Linguistics, Leiden/Holland Academic Graphics.

Heuven, V. J. van (1986). Some acoustic characteristics and perceptual consequences of foreign accent in Dutch spoken by Turkish immigrant workers. In J. van Oosten, J. F. Snapper (eds.) *Dutch Linguistics at Berkeley, papers presented at the Dutch Linguistics Colloquium held at the University of California, Berkeley on November 9th, 1985* (pp. 67–845). Berkeley: The Dutch Studies Program, U.C. Berkeley,.

Heuven, V. J. van & Zanten, E. van (1983). A phonetic analysis of the Indonesian vowel system, a preliminary acoustic study, NUSA, *Linguistic Studies of Indonesian and other Languages in Indonesia* 15, 70–80.

Hillenbrand, J., Getty, L.A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America* 97, 3099–3111.

Hirsh, I. J., Reynolds, E. G., & Joseph, M. (1954). The intelligibility of different speech materials. *Journal of the Acoustical Society of America* 26, 530–538.

Hockett, C. F. (1947). *Peiping phonology*. *Journal of American Oriental Society* 67:253-267. Reprinted 1964 in M. Joos (ed.), Readings in Linguistics I (fourth edition, pp. 217–228) Chicago: University of Chicago Press.

Howie, J. (1976). *Acoustical Studies of Mandarin Vowels and Tones*. New York: Cambridge University Press.

Hulst, H. van der (1984). Syllable structure and stress in Dutch. Doctoral dissertation, Leiden University.

Ioup G., Boustagui E., El Tigi M., Moselle M. (1994). Reexamining the critical period hypothesis. A case study of successful adult SLA in a naturalistic environment. *Studies in Second Language Acquisition* 16, 73–98.

Jakobson, R. Fant, C. G. M. & Halle, M. (1952). *Preliminaries to speech analysis*. Cambridge, MA: MIT Press.

Jekosch, U. (1994). Speech intelligibility testing: On the interpretation of results, *Journal of American Voice* I/ 0 Society 15, 63–79.

Jilka, M. (2000). *The contribution of intonation to the perception of foreign accent*. (Ph.D. thesis, University of Stuttgart). Vol. 6, no. 3, *Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung*, University of Stuttgart.

Jones, D. (1956). *Outline of English Phonetics* (8th edition). Cambridge: Heffer.

Kalikow, D. N., Stevens, K. N., Elliott L. L. (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.

Kalin, R., & Rayko, D. S. (1978). Discrimination in evaluative judgments against foreign-accented job candidates. *Psychology Reports* 43, 1203–1209.

Karlgren, B. (1915–1926). *Études sur la Phonologie Chinoise*. Stockholm: P. A. Norsted och Soener.

Kenyon, J. & Knott, T. (1944). *A pronouncing dictionary of American English*. Springfield: Merriam.

Kratochvil, P. (1968). *The Chinese Language Today: Features of an emerging standard*. London: Hutchinson.

Kruskal, J. B. (1964). Non metric multidimensional scaling: a numerical method. *Psychometrika* 29, 115–129.

Kruskal, J B. & Wish, M. (1978). *Multidimensional Scaling*. Thousand Oaks, CA: Sage.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics* 50, 93–107.

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the ''perceptual magnet effect''. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121–154). Timonium, MD: York Press.

Labov, W. (1994). *Principles of Linguistic Change. Internal Factors*, Oxford: Blackwell.

Labov, W. (2001). *Principles of Linguistic Change. Social Factors*. Oxford: Blackwell.

Labov, W., Ash, S., & Boberg, C. (2006). *The Atlas of North American English*. Berlin: Mouton de Gruyter.

Ladefoged, P. & Maddieson, I. (1990). Vowels of the world languages. *Journal of Phonetics* 18, 93–122.

Ladefoged, P. & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Ladefoged, P. (1971). *Preliminaries to linguistic phonetics*. Chicago: University of Chicago Press.

Lado, R. (1957). *Linguistics across cultures*. Ann Arbor: University of Michigan Press.

Lambert, W. E., Hodgson, R. C., Gardner, R. C. & Fillenbaum, S. (1960). Evaluational reactions to spoken language. *Journal of Abnormal and Social Psychology* 60, 44–51.

Lane, H. (1963). Foreign accent and speech distortion. *Journal of the Acoustical Society of America* 35 451–453.

Lenneberg, E. H. (1967). *Biological Foundations of Language*. New York: John Wiley & Sons.

Li, A., Yu, J. Chen, J. & Wang, X. (2004). A contrastive study of Standard Chinese and Shanghai-accented Standard Chinese. In G. Fant, H. Fujisaki, J. Cao & Y. Xu (eds.), *From traditional phonology to modern speech processing, Festschrift for professor Wu Zongji's 95th birthday* (pp. 253–288). Beijing: Foreign Language Teaching and Research Press.

Li, C. N. and Thompson, S. A. (1981). *Mandarin Chinese: a functional reference grammar*. Berkeley, CA: University of California Press.

Li, Wen Chao (1999). *A diachronically-motivated segmental phonology of Mandarin Chinese*. New York: Peter Lang Publishing.

Light, T. (1976). *The Chinese syllabic final*. Ithaca, NY: Cornell University Press.

Lin, Y. H. (1989). Autosegmental treatment of segmental processes in Chinese phonology. Ph.D. dissertation. University of Texas at Austin.

Lindblom B. (1986). Phonetic universals in vowel systems. In J. J. Ohala, & J. J. Jaeger (eds.), *Experimental Phonology* (pp. 113–144). Orlando, FL: Academic Press.

Lindblom, B. & Maddieson, I (1988). Phonetic universals in consonant systems. In L.M. Hyman & C.N. Li (eds.), *Language, speech and mind. Studies in honor of Victoria A. Fromkin* (pp. 62–78). London: Routledge.

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America* 49, 606-608.

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition* 12, 251–285.

MacKay, I. R. A, Meador, D. & Flege, J. E. (2001). The identification of English consonants by native speakers of Italian. *Phonetica* 58, 103–125.

Maddieson, I. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.

Magen, H. S. (1998). The perception of foreign-accented speech. *Journal of Phonetics* 26, 381–400.

Major, R. C. (1987). Measuring pronunciation accuracy using computerized techniques. *Language Testing* 4, 155–169.

Miller, G. A. & Nicely, P. E., (1955). Analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America* 27, 338–353.

Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly* 25, 481–520.

Munro, M. J. & Derwing, T. M. (1998). The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning* 48, 159–182.

Munro, M. J. & Derwing, T. M. (1995a). Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning* 45, 73–97.

Munro, M. J. & Derwing, T. M. (1995b). Processing time, accent and comprehensibility in the perception of native and foreign accented speech. *Language and Speech* 38, 289–306.

Munro, M. J. & Derwing, T. M. (2001). Modelling perceptions of the comprehensibility and accentedness of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition* 23, 451–468.

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America* 85, 2088–2113.

Nierop, D. J. P. J. van, Pols, L. C. W. & Plomp, R. (1973). Frequency analysis of Dutch vowels from 25 female speakers. *Acustica* 29, 110–118.

Nooteboom, S. G. (1997). The prosody of speech: melody and rhythm. In: W. J. Hardcastle & J. Laver (eds.), *The Handbook of Phonetic Sciences* (pp. 640–673). Oxford: Basil Blackwell.

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

Obler, L. K. (1989). Exceptional second language learners. In S. Gass, C. Madden, L. Preston & L. Selinker (eds.), *Variation in second language acquisition: Vol. 2. Psycholinguistic issues* (pp. 141–159). Clevedon: Multilingual Matters.

Ooijen, B. A. van (1994). The processing of vowels and consonants. PhD dissertation, Leiden University.

Oyama, S. (1976). A sensitive period for the acquisition of a non-native phonological system. *Journal of Psycholinguistic Research* 5, 261–283.

Patkowski, M. (1990). Age and accent in a second language: A reply to James Emil Flege. *Applied Linguistics* 11, 79–89.

Pennington, M. C. & Richards, J. C. (1986). Pronunciation revisited. *TESOL Quarterly* 20, 207–225.

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

Piske, T., Mackay, I. R. A. & Flege, J. E. (2001). Factors affecting the degree of foreign accent in an L2: A review. *Journal of Phonetics* 29, 191-215.

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

Politzer, R. L. (1978). Errors of English speakers of German as perceived and evaluated by German natives. *Modern Language Journal* 62, 253–261.

Polivanov, E. (1931). La perception des sons d'une langue étrangère. *Travaux de cercle linguistique de Prague* 4, 79–96.

Pols, L. C. W., van der Kamp, L. J. Th., & Plomp, R. (1969). Perceptual and physical space of vowel sounds. *Journal of the Acoustical Society of America* 46, 458–467.

Pulleyblank, E. G. (1984). Vowelless Chinese? An application of the three-tiered theory of syllable structure to Pekingese. *Proceedings of the XVI International Conference on Sino-Tibetan Languages and Linguistics, Seattle*, 568–610.

Purcell, E., & Suter, R. (1980). Predictors of pronunciation accuracy: A re-examination. *Language Learning* 30, 271–287.

Rietveld, A. C. M. & Heuven, V. J. van (2001). *Algemene fonetiek [General Phonetics]*. Bussum: Coutinho.

Riney, T. J., & Takagi, N. (1999). Global foreign accent and voice onset time among Japanese EFL speakers. *Language Learning* 49, 275–302.

Rubin, D. L. & Smith, K. A. (1990). Effects of accent, ethnicity and lecture topic on undergraduates' perceptions of non-native English speaking teaching assistants. *International Journal of Intercultural Relations* 14, 337-353

Ryan, E. B. & Carranza, M. A. (1975). Evaluative reactions of adolescents toward speakers of standard English and Mexican American accented English. *Journal of Personality and Social Psychology*. 31, 855–863.

Schinke-Llano, L. (1983). Foreigner talk in content classrooms. In H. Slinger & M. Long (eds.), *Classroom centered research in second language acquisition.* Rowley, MA; Newbury House.

Schinke-Llano, L. (1986). *Foreign language in the elementary school: State of the art*. Orlando, FL: Harcourt Brace, Jovanovich.

Schouten, M. E. H. (1975). Native-language interference in the perception of second-language vowels. Doctoral dissertation, Utrecht University.

Seliger, H., Krashen, S. & Ladefoged, P. (1975) Maturational constraints in the acquisition of second languages. *Language Sciences* 38, 20–22.

Slis, I. H. & Heugten, M. van (1989). Voiced-voiceless distinction in Dutch fricatives. In H. Bennis & A. van Kemenade (eds.), *Linguistics in the Netherlands 1989* (pp. 123–132). Dordrecht: Foris,.

Smeele, P. (1985). Effecten van buitenlands accent op de herkenning van gesproken woorden: Duits versus Nederlands [Effects of foreign accent on the recognition of spoken words: Dutch versus German]. MA thesis, Phonetics Laboratory Leiden University.

Spiegel, M., Altom, M., Macchi, M. & Wallace, K. (1990). Comprehensive assessment of the telephone intelligibility of synthesized and natural speech. *Speech Communication* 9, 279–291.

Strange, W. Bohn, S.-O., Trent, S. A. & Nishi, K. (2004). Acoustic and perceptual similarity of North German and American English vowels. *Journal of the Acoustical Society of America*, 115, 1791–1807.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., Nishi, K., & Jenkins, J. J. (1998). Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics* 26, 311–344.

Strange, W., Bohn, O.-S., Nishi, K. & Trent, S.A. (2005). Contextual variation in the acoustic and perceptual similarity of North German and American English vowels. *Journal of the Acoustical Society of America* 118, 1751–1762.

Strange, W., Verbrugge, R. R., Shankweiler, D. P. & Edman, T. R. (1976). Consonant environment specifies vowel identification. *Journal of the Acoustical Society of America* 60, 213–224.

Suter, R. (1976). Predictors of pronunciation accuracy in second language learning. *Language Learning* 26, 233–253.

Tahta, S., Wood, M. & Lowenthal, K. (1981a). Foreign accents: factors relating to the transfer of accent from the first language to a second language. *Language and Speech* 24, 265–272.

Tahta, S., Wood, M. & Lowenthal, K. (1981b). Age changes in the ability to replicate foreign pronunciation and intonation. *Language and Speech* 24, 363–372.

Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning* 41, 177–204.

Tielen, M. T. J. (1992). Male and female speech. An experimental study of sex-related voice and pronunciation characteristics. Doctoral dissertation, University of Amsterdam.

Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America* 88, 97–100.

Trubetzkoy, N. S. (1969). *Principles of phonology* (C. A. M. Baltaxe, Transl.). Berkeley, CA: University of California Press. (Original published 1939)

Varonis, E., & Gass, S. (1982). The comprehensibility of nonnative speech. *Studies in Second Language Acquisition* 4, 114–136.

Velde, H. van de (1996). Variatie en verandering in het gesproken Standaard-Nederlands (1935-1993) [Variation and change in spoken Standard Dutch (1935–1993)]. PhD dissertation, Nijmegen University

Wan, I. P. & Jaeger, J. (1998). Speech errors and the representation of tone in Mandarin Chinese. *Phonology* 15, 417–461.

Wang, H. & Heuven, V. J. van (2003). Mutual intelligibility of Chinese, Dutch and American speakers of English. In L. Cornips & P. Fikkert (eds.), *Linguistics in the Netherlands* 2003 (pp. 213–224). Amsterdam/Philadelphia: John Benjamins,

Wang, H. & Heuven, V. J. van (2004). 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* (pp. 205–216). Amsterdam/Philadelphia: John Benjamins,.

Wang, H. & Heuven, V. J. van (2005). Mutual intelligibility of American Chinese and Dutch-accented English. *Proceedings of the 9th European Conference on Speech Communication and Technology, Interspeech 2005*, Lisbon, 2225–2228.

Wang, H. & Heuven, V. J. van (2006). Acoustical analysis of English vowels produced by Chinese, Dutch and American speakers. In J. M. van de Weijer & B. Los, (eds.) *Linguistics in the Netherlands 2006* (pp. 237–248). Amsterdam/ Philadelphia: John Benjamins.

Wang, J. Z. (1993). The Geometry of Segmental Features in Beijing Mandarin. PhD dissertation, University of Delaware.

Weenink, D. J. M. (2006). Speaker-adaptive vowel identification. Doctoral dissertation, University of Amsterdam.

Wells, J. C. (1982). *Accents of English*. Cambridge: Cambridge University Press.

Weinreich , U. (1953). *Languages in contact*. The Hague: Mouton.

Wiese, R. (1997). Underspecification and the description of Chinese vowels. In J. Wang & N. Smith (eds.), *Studies in Chinese Phonology* (pp. 219–249). Berlin: Mouton de Gruyter.

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

Wu, Y. (1994). Mandarin segmental phonology. Ph.D. dissertation. University of Toronto.

Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.

Zhao, D. (1995). *English phonetics and phonology: as compared with Chinese features*. Qingdao Shi: Qingdao hai yang da xue chu ban she.

Zhong, Q. (1980). *On Chinese phonetics*. Beijing: The Commercial Press.

Appendices

Appendix A4.1 Semantically Unpredictable Sentences (SUS)

Structure 1: Subject – Intransitive Verb – Adverbial Phrase:

1. The state sang by the long week.
2. The man lay through the wide war.
3. The day hung to the great night.
4. The year smiled through the young head.
5. The time ran with the high side.
6. The way ran of the hot room.
7. The thing hung from the small line.
8. The grass lied on the blue night.
9. The school stayed for the new tube.
10. The hand fell of the high form.

Structure 2: Subject – Transitive Verb – Direct Object

1. The real field made the vote.
2. The white home got the art.
3. The clear friend brought the ground.
4. The white sense held the air.
5. The whole month brought the air.
6. The thin job got the road.
7. The poor sense hit the tax.
8. The short field said the air.
9. The full home took the term.
10. The white sense ate the road.

Structure 3: Imperative Verb – Direct Object

1. Use the game or the hair.
2. Ask the trial and the tree.
3. Leave the sport and the thought.
4. Call the club and the growth.
5. Turn the love or the test.
6. Add the sale or the nose.
7. Start the store or the price.
8. Show the plant or the sound.
9. Feel the stock and the list.
10. Live the sport and the fund.

Structure 4: Question word – Verb – Subject – Direct Object

1. When does the charge like the late plane?
2. Where does the band sell the low set?
3. Why does the cell like the deep length?
4. When does the gun like the deep bed?
5. Why does the range watch the fine rest?
6. When does the sign lead the red roof?
7. How does the chance plan the cold fear?
8. How does the chance send the deep roof?
9. Why does the gun bear the red trade?
10. How does the cloud watch the low text?

Structure 5: Subject – Verb – Complex Direct Object

1. The farm meant the hill that burned.
2. The curve helped the blood that won.
3. The hope rode the boat that failed.
4. The crowd heard the moon that lost.
5. The inch paid the branch that passed.
6. The song paid the ball that stopped.
7. The truth rode the hill that died.
8. The lost paid the moon that worked.
9. The aid rode the glass that rose.
10. The truth rode the leg that failed.

Appendix A4.2. Sentences of the Speech-in-Noise test (SPIN)

Low predictability	High predictability
1. Ruth could have discussed the wits.	26. Throw out all the useless junk.
2. We could discuss the dust.	27. She cooked him a hearty meal.
3. We spoke about the knob.	28. Her entry should win the first prize.
4. Paul hopes we heard about the loot.	29. The stale bread was covered with mold.
5. David might consider the fun.	30. The firemen heard her frightened scream.
6. Paul could not consider the rim.	31. Your knees and your elbows are joints.
7. He heard they called about the lanes.	32. I ate a piece of chocolate fudge.
8. They had a problem with the cliff.	33. Instead of a fence, plant a hedge.
9. Harry will consider the trail.	34. The story had a clever plot.
10. We are considering the cheers.	35. The landlord raised the rent.
11. She has known about the drug.	36. Her hair was tied with a blue bow.
12. Bill had a problem with the chat.	37. He's employed by a large firm.
13. We hear they asked about the shed.	38. To open the jar, twist the lid.
14. Jane had not considered the film.	39. The swimmer's leg got a bad cramp.
15. Jane did not speak about the slice.	40. Our seats were in the second row.
16. Paul was interested in the sap.	41. The thread was wound on the spool.
17. I am discussing the task.	42. They tracked the lion to his den.
18. Ruth has discussed the peg.	43. Spread some butter on your bread.
19. Tom is considering the clock.	44. A spoiled child is a brat.
20. He's thinking about the roar.	45. Keep your broken arm in a sling.
21. I should have known about the gum.	46. The mouse was caught in the trap.
22. They heard I asked about the bet.	47. I have got a cold and a sore throat.
23. Betty doesn't discuss the curb.	48. Ruth poured herself a cup of tea.
24. He had a problem with the tin.	49. The house was robbed by a thief.
25. He wants to know about the rib.	50. Wash the floor with a mop.

Appendix A4.3. Questionnaire.

INFORMATION FORM

(Note: personal information contained here will not be released)

Name: _____

Subject Number: _____ Today's date _____

Email: _____ Telephone number _____

Age: ____ Gender: Male / Female

1. Where were you born? (city, state (province), country) _____

2. How long have you lived there? _____

3. Did you move from that place? Y / N
   How old were you then? _____

4. Where did you attend elementary school? _____
   What language did you use at school? _____

5. Where did you attend secondary school? _____
   What language did you use at school? _____

6. Where did you attend college? _____
   What language or languages do you use in class? _____

7. How long have you been in the Netherlands? _____

8. Your native language is _____
   Your parent(s) language is Mother _____
   Father _____

9. Do you have native English speakers in your family? Y / N

10. At what age did you start *learning* English? _____
    In what kind of environment did you start using English?
    at school Y / N
    at home Y / N
    with friends? Y / N

11. At what age did you start *using* English? _____
    In what kind of environment did you start using English?
    at school Y / N

at home Y / N
with friends Y / N

12. Do you have a job? Y / N
What language do you usually use at work? _____

13. Are you a student? Y / N
What language do you usually use in class? _____

14. How many years experience do you have with English? _____

15. Which language do you speak the most often?
at home _____
at school or work _____
with friends _____

16. Do you have any experience living in China? Y / N How long? ____

17. Do you have any experience living in the Netherlands? Y / N How long ? ____

18. Do you have any experience living in the USA? Y / N How long ? ____

19. Do you have any experience living in other English-speaking countries? Y / N
How long? ____

20. Do you think your English is good enough for communication? Y / N

Thank you for your cooperation!

Appendix A4.4. Instructions

Instructions part one: Vowels

In the first part of the test your task is to decide which one of 19 different mono-syllabic words you heard. The words always begin with an *h* and end in a *d*. They differ in the vowel or in the presence of an *r*-sound right after the vowel. Here is a list of the 19 words that you may choose from:

	test word	*rhymes with*		*test word*	*rhymes with*
1.	h**ee**d	f**ee**d, n**ee**d	11.	h**oe**d	r**oa**d, sh**ow**ed
2.	h**i**d	m**i**d, k**i**d	12.	h**u**d	m**u**d, bl**oo**d
3.	h**ay**ed	pl**ay**ed, st**ay**ed	13.	h**ear**d	b**ir**d, w**or**d
4.	h**ea**d	r**e**d, b**e**d	14.	h**i**de	sl**i**de, r**i**de
5.	h**ar**d	c**ar**d, b**ar**red	15.	h**oy**ed	t**oy**ed, empl**oy**ed
6.	h**a**d	b**a**d, s**a**d	16.	h**ow**'d	l**ou**d, all**ow**ed
7.	wh**o'**d	gl**ue**d, r**u**de	17.	h**ere**'d	b**ear**d, sn**eer**ed
8.	h**oo**d	g**oo**d, w**oo**d	18.	h**oor**ed	t**oure**d, m**oore**d
9.	h**aw**ed	s**aw**ed, fr**au**d	19.	h**aire**d	sh**are**d, c**are**d
10.	h**o**d	g**o**d, n**o**d			

In spite of what you may think, each of the 19 words in bold face has a different pronunciation. Please take a minute to study the 19 test-words as they are listed from left to right on your answer sheet (i.e. in the order 1 through 19 in the above table). In order to know how to pronounce the 19 words, carefully study the rhyming words following the test words. Obviously, except for the consonants preceding the vowel, the test words and the rhyming words following it have exactly the same pronunciation.

In the actual test on the tape you will hear six different speakers. Two speakers are native American, two are Dutch, and two are Chinese. Each speaker pronounces each of the 19 test words (or word combinations) that begin with ***h*** and end with ***d***:

heed, hid, head, had, hard, hawed …

We are going to start the tape for a short practice run. You will hear ten words for practice. After each word you should indicate on your answer sheet, by ticking the appropriate box, which of the 19 words you think the speaker intended. **Note that you must make a choice, and one choice only, for each word on the tape.** If you really cannot decide which word you heard, then just gamble.

[…….]

The words are played to you at a rate of one every six seconds and the speakers vary at random from one word to the next.

If you have no further questions with respect to the test procedure, we will switch on the tape for the actual test. To help you keep track on your answer sheets, there will be a short beep after every fifth word on the tape. There will be 120 words all together; this part of the test will take about 15 minutes.

Instructions part two: Consonants

In this part of the test your task is decide which one of 24 different monosyllabic nonsense words you heard. The words always begin with ***a*** and end in ***a***. They differ only in consonants in the middle. Here is a list of the 24 nonsense words with consonants that you may choose from; in order to make you clear about every consonant we provide some real words with the same consonants you are familiar with in the second column:

	test word	*same consonant as in*		*test word*	*same consonant as in*
1.	a**p**a	**p**en, **p**ea	13.	a**h**a	**h**e, **h**i,
2.	a**b**a	**b**ee, **b**y	14.	a**r**a	**r**ed, **r**ose
3.	a**t**a	**t**ea, **t**o	15.	a**f**a	**f**at, **f**oot
4.	a**d**a	**d**esk, **d**id	16.	a**v**a	**v**ase, **v**est
5.	a**k**a	**k**iss, **k**ey	17.	a**ch**a	**ch**air, **ch**eese
6.	a**g**a	**g**ate, **g**o	18.	a**j**a	**j**am, **j**ar
7.	a**s**a	**s**ea, **s**ee	19.	a**m**a	**m**um, **m**y
8.	a**z**a	**z**oo, **z**ero	20.	a**n**a	**n**ice, **n**ight
9.	a**sh**a	**sh**y, **sh**e	21.	a**ng**a	ha**ng**er,
10.	a**zh**a	plea**s**ure, A**s**ia	22.	a**l**a	**l**ie, **l**ay
11.	a**th**a	**th**in, **th**ink	23.	a**y**a	**y**es, **y**et
12.	a**dh**a	**th**at, **th**ose	24.	a**w**a	**w**as, **w**ar

Please take a minute to study the 24 test "words" as they are listed left to right on your answer sheet (in the same order from 1 to 24 as in the table above). Make sure that you understand which consonant sound is intended in each nonsense word, and know (roughly) where each word is in the order from left to right – so that you will be able to work quickly once the tape starts.

In the actual test on the tape you will hear six different speakers. They are the same speakers as in the first part. Each speaker pronounces each of the 24 test words that begin with *a* and end in *a*:

apa, aba, ada, ata,……..

Speakers will alternate randomly on the tape. Your task is to decide for each nonsense word on the tape which consonant occurs between the vowels. Indicate your answer by ticking the appropriate box. Note that you must make a choice, and one choice only, for each word on the tape. If you really cannot decide which consonant you heard, then just gamble.

We will now play the first part of the tape for practice, just to familiarize you with your task and its time constraints.

[…….]

If you have no further questions with respect to the test procedure, we will switch on the tape for the actual test. To help you keep track on your answer sheets, there will be a short beep after every fifth word on the tape. There will be 150 items all together; this part of the test will take just within15 minutes.

Instructions part three: Consonant Clusters

Consonants in English sometimes occur in combinations (pairs or even triplets) at the beginning of words, e.g. in *plane*, *blue*, *pray*, *bread. Pl*, *bl*, *pr* and *br* in these words are called consonant clusters. On the tape you will hear 21 nonsense words with clusters, all of them between vowels *a*. The intended pronunciation of each cluster is also illustrated by words you are familiar with in the second column in the form:

	test word	*as pronounced in*		*test word*	*as pronounced in*
1.	a**pl**a	**pl**ane, **pl**ay	11.	a**spr**a	**spr**ing, **spr**ead
2.	a**bl**a	**bl**ue, **bl**ow	12.	a**spl**a	**spl**it, **spl**endid
3.	a**pr**a	**pr**ay, **pr**ice	13.	a**scr**a	**scr**eam, de**scr**ibe
4.	a**br**a	**br**ead, **br**ing	14.	a**sp**a	**sp**eak, **sp**eed
5.	a**tr**a	**tr**ee, **tr**y	15.	a**st**a	**st**ar, **st**ay
6.	a**dr**a	**dr**y, **dr**iver	16.	a**sc**a	**sc**ale, **sc**hool
7.	a**cr**a	**cr**y, **cr**eam	17.	a**sm**a	**sm**all, **sm**art
8.	a**gr**a	**gr**ey, **gr**een	18.	a**sn**a	**sn**ake, **sn**eeze
9.	a**cl**a	**cl**ass, **cl**ean	19.	a**sl**a	**sl**ow, **sl**im
10.	a**gl**a	**gl**ass, **gl**ue	20.	a**sw**a	**sw**eat, **sw**im
			21.	a**thr**a	**thr**ough **th**row

Please take a minute to study the 21 consonant clusters listed in the nonsense words in the table above and on your answer sheets. Both in the table and on your answer sheets the clusters will be listed in the same order from 1 to 21).

apla, abla, apra, abra,……

In this part of the experiment your task is to indicate which consonant pair or triplet you heard in each of a series of nonsense words.

You will now hear a practice run of 10 nonsense words. Indicate your answer by ticking the appropriate box. Note that you must make a choice, and one choice only, for each word on the tape. If you really cannot decide which consonant you heard, then just gamble.

If there are no further questions regarding the procedure, we will now proceed with the actual test. There will be 130 items; this part of the test will take about 10 minutes.

You will have about 5 seconds to make your choice; there will be a beep after every fifth item.

Instructions part four: Nonsense Sentences

In this part you are going to hear 30 sentences read by the same six speakers as in parts one, two and three. All the sentences are nonsense sentences with very simple words you are familiar with.

e.g. The grass lied on the blue night.
The short field said the air.
Show the plant or the sound.
How does the chance plan the cold fear?
The lost paid the moon that worked.

You can see that in the listed sentences there are no difficult words. In the test we leave the important words in every sentence blank on the answer sheet, e.g. the sentence:

The grass lied on the blue night.

will be printed on the answer sheet as

The _____ _____ on the _____ _____.

Your task is to listen to the tape and fill in the blanks with the words you hear on the tape.

Every sentence will be played three times in a row. During the second presentation there will pause of 3 seconds after ***every*** blanked-out word, which will allow you sufficient time to fill in the blanks. During the third (uninterrupted) presentation you can then check your answers and spelling, and make last-minute changes. Be sure to write clearly, please.

If you have no further questions with respect to the test procedure, I will now switch on the tape for a series of five practice items (fill in the blanks below).

a. The _____ _____ from the _____ _____.

b. The _____ _____ _____ the _____.

c. _____ the _____ or the _____.

d. How does the _____ _____ the _____ _____?

e. The _____ _____ the _____ that _____.

If there are no further questions regarding the procedure, we will now switch on the tape for the actual test. There will be 30 items all together; this part of the test will take just under 10 minutes.

Instructions part five: Meaningful sentences

In this final section you are going to hear 50 sentences read by the same six speakers that you heard before. They are all meaningful sentences with every-day words in them.

In this test, your task is to write down on your answer sheet for each sentence on the tape **only the last word you hear**. Note that last word of any test sentence is **always a one-syllable word**.

Each sentence will be read **only once** with a short pause in between sentences.

Please, write clearly. Do not leave items blank. If you do not recognize a word, then just write down any word that comes close to the sounds you heard on the tape.

If you have no further questions with respect to the test procedure, we will switch on the tape for the actual test. There will be no practice items this time. To help you keep track on your answer sheets, there will be a short beep after every fifth sentence on the tape. This part of the test will take less than 10 minutes.

Appendix A6.1. Percent correct vowel identification broken down by language background of listener and of speaker. Mean, number of observations, standard deviation and standard error of the mean are indicated.

Nationality of		Mean	N	SD	Se
Listener	Speaker				
Chinese	Chinese	29.2	1368	45.5	1.2
	Dutch	33.8	1368	47.3	1.3
	USA	32.9	1368	47.0	1.3
	Total	32.0	4104	46.7	.7
Dutch	Chinese	40.3	1368	49.1	1.3
	Dutch	59.5	1368	49.1	1.3
	USA	58.6	1368	49.3	1.3
	Total	52.8	4104	49.9	.8
USA	Chinese	44.7	1368	49.7	1.3
	Dutch	61.1	1368	48.8	1.3
	USA	75.4	1368	43.1	1.2
	Total	60.4	4104	48.9	.8
Total	Chinese	38.1	4104	48.6	.8
	Dutch	51.5	4104	50.0	.8
	USA	55.6	4104	49.7	.8
	Total	48.4	12312	50.0	.5

Appendix A6.2. Confusion matrices for vowels of each of nine combinations of speaker and listener nationality.

Table A6.2.1. Vowel identification (%): Chinese listeners – Chinese speakers.

		Response vowel																		
		i:	ɪ	e:	ɛ	ɑ:	æ	u:	ʊ	ɔ:	ɔ	o	ʌ	ʌr	ai	ɔi	ɑu	iə	uə	ɛə
Stimulus vowel	i:	**39**	32	4	4	3	1	1		4					4	1		3		1
	ɪ	38	**40**	1	3	1		3		1		1	1		6	1		3		
	e:	11	15	**44**	9	2	1			6	1		1	1	5	2		1	2	
	ɛ	1	3	6	**19**	6	22			11			1	1	19	1	6		1	1
	ɑ:				1	**58**	3	1	1	4	3		10	8	1	1	6		1	
	æ		3	4	29	1	**24**		3	1		1		6	25		1			1
	u:				1			**29**	28	1	8	3	13		1		1	4	10	
	ʊ		4			3		22	**44**		4		15	3		1	1		1	
	ɔ:				3	47	1		4	**11**	7	1	4	1		1	18			
	ɔ		3		4	15		3	3	21	**7**	4	7	1		1	21	3	6	1
	o					1		11	22	1	4	**33**	10			7		1	8	
	ʌ		4	1	19	36	10		2	1	5		**17**	1	2		2	1	1	
	ʌr	1			4	4	1	1		4	1		1	**60**			1	13	3	4
	ai	42	39	4	3						1			1	**3**	1		3	1	1
	ɔi		1	3		34	7	6		4	6		6	1	11	**10**	1	3	1	6
	ɑu		1	3		19		1	3	14	6	3				7	**36**	1	6	
	iə	3			3				1			1		22	1			**60**	1	7
	uə				1	6	1	4	1	15	11	4	4	11		10	4	3	**24**	
	ɛə				3	60	3	1	1	3	4	1	6	4	1	3	6			**4**

Table A6.2.2. Vowel identification (%): Chinese listeners – Dutch speakers.

		Response vowel																		
		i:	ɪ	e:	ɛ	ɑ:	æ	u:	ʊ	ɔ:	ɔ	o	ʌ	ʌr	ai	ɔi	ɑu	iə	uə	ɛə
Stimulus vowel	i:	**39**	43	1	1		1		1					3	4	1		1	1	1
	ɪ	17	**46**	10	6	1	3			1	1		1	1	8			3		1
	e:	18	18	**35**	4		1	3	3	3		3	1	1	3		3	4		
	ɛ	1	4	3	**44**	3	26	1		1		3	1		6		1	1		3
	ɑ:				1	**49**	1	3	3	3	10		18	4	1		3	1	1	1
	æ	3	5		46	3	**23**	1		1		1	2	2	6	2	1		1	4
	u:			1	1		1	**38**	38		3	3	11		1	1			1	
	ʊ			1			1	23	**39**	1	3	1	14	1	1		1		11	
	ɔ:	1			3			10	13	**10**	15	22	3	1	3	8	4	1	4	1
	ɔ			1	1	31		4	3	6	**29**	8	6	3	1	1	4		1	
	o			1		1		3	6	4	11	**53**	1	1		4	13		1	
	ʌ		1	1	3	40	1	3	3	1	13	3	**14**		3	3	7	1	3	
	ʌr	4	1		3	3	1	6	8	3	1	1	8	**50**	1		1	1	4	1
	ai	10	40	14	7	1	3		1	4		1	1	3	**6**	1	1	4	1	
	ɔi			4	4	3	3	1	1	4	6	11	4	6	8	**28**	6	1	6	4
	ɑu			1	1	5	1	2	3	19	9	3	1		1	7	**40**	2	4	2
	iə		3	3	6	3	1	1	6		3	1	1	21			4	**42**		6
	uə	1			4		3	15	8	11	4	1	14	1	1	1	7	4	**19**	3
	ɛə			6	31	1	8							8	3			4	1	**38**

Table A6.2.3. Vowel identification (%): Chinese listeners – American speakers.

Stimulus vowel \ Response vowel	iː	ɪ	eː	ɛ	ɑː	æ	uː	ʊ	ɔː	ɔ	o	ʌ	ʌr	ai	ɔi	ɑu	iə	uə	ɛə
iː	**28**	50	4			1	1			3	1	1		1		1	7		
ɪ	6	**38**	8	28	3	7		1	3		1			3			1		1
eː	7	19	**21**	18	3	3		4	3		1		3	1			11		6
ɛ	1	3	1	**69**	1	14					3		1	3				1	1
ɑː		7		3	**58**	3	1			1	3	10	10	1		1			1
æ	3	3	6	36	15	**28**	1	3			1	3	1						
uː		1	1	1	1		**35**	31	3	7		6	1			1	3	8	
ʊ	1	3		3	1	3	10	**18**	3	7	4	17	25			3	1	1	
ɔː	1			3	68	3		3	**4**	10		4	3					1	
ɔ		3	4	6	46	2		1	3	**11**		10	4	4	3		1		3
o		1		1			3	6	6	18	**44**	1		1	10	3	1	4	
ʌ	1			6	13	1	3	8	3	14	3	**13**	35				1		
ʌr		1	1	8	3					1		1	**64**	1	1	1	10	3	3
ai	8	51	3	24		3		1		1	1	1		**6**					
ɔi		2	2	1		2	2	2	7	2	7	1		2	**59**	3		7	1
ɑu			1	1	4		1	3	29	10	6	4	1	1	6	**25**		6	1
iə	1	3		3	4			3			1	1	18		1	3	**57**		4
uə				1	13	3	6	8	6	7	3	17	8	1	1	4		**22**	
ɛə	3	8	3	10	1	3		1	1		3	1	10	3	1	1	22	1	**26**

Table A6.2.4. Vowel identification (%): Dutch listeners – Chinese speakers.

Stimulus vowel \ Response vowel	iː	ɪ	eː	ɛ	ɑː	æ	uː	ʊ	ɔː	ɔ	o	ʌ	ʌr	ai	ɔi	ɑu	iə	uə	ɛə
iː	**75**	10		3	1	1				3	1		1				4		
ɪ	57	**23**		4				1		1	1		1	3			7		
eː	1		**81**	4	1	1	1		5				1	1	1	2	1		1
ɛ		3	1	**4**		6								86					
ɑː			1		**63**	10	1	3	1	6		3	4		1	3		3	1
æ			3	22		**23**		1		3				46			1		
uː							**31**	47			15		1					6	
ʊ					1		26	**60**	1	3	6	1						1	
ɔː			2		18			2	**6**	3	2	3	2	29	6	29			
ɔ					20	1			13	**19**	1	4	4			37			
o	1			4		1	11	37	4	9	**20**	1			7	1		1	
ʌ		3		16	1	59			1	3		**8**	1	1	3	2	1		2
ʌr					25					1			**59**				4	9	1
ai		6	1	1	3	8			1				3	**74**	3				
ɔi	1	4	4	1	1			3	1	1			1	46	**33**			1	
ɑu				1	10	1	3		7	21		1	1		1	**47**	1	3	
iə									1				49				**44**	1	4
uə					1			3	3	1			3		1			**86**	1
ɛə					38				3			1	23	1	1		4	1	**26**

Table A6.2.5. Vowel identification (%): Dutch listeners – Dutch speakers.

		Response vowel																		
		iː	ɪ	eː	ɛ	ɑː	æ	uː	ʊ	ɔː	ɔ	o	ʌ	ʌr	ai	ɔi	ɑu	iə	uə	ɛə
Stimulus vowel	iː	**70**	7	1	9				1		1	1		1		4		3		
	ɪ	4	**94**												1					
	eː	1	1	**79**	1		1			1			1		1	4	4			1
	ɛ		3		**68**		27		1				1							
	ɑː				1	**58**	18		1	1	15		1	1		1				
	æ		2		50	1	**39**					1	1	1		4	1			
	uː		1					**42**	42			3	3		1		6		1	
	ʊ					3		31	**54**	1	3	4					3			
	ɔː	1						1	12	**24**	16	25		1		3	12		3	1
	ɔ					6	4		3	3	**75**	3	7							
	o							6	7	19	10	**33**				9	16			
	ʌ		1			13	37			1	19		**26**				3			
	ʌr	1			1			1	3			4	7	**78**				3	1	
	ai	3	4	7		1	1					1			**81**	1				
	ɔi	1								10	3				8	**77**				
	ɑu					2	2	2	1	11	2	5				4	**70**	1		
	iə				1	1			1					36				**51**	1	7
	uə			6				1	4			4	1		1	1	7		**73**	
	ɛə			1	3			1						21	1			8		**63**

Table A6.2.6. Vowel identification (%): Dutch listeners – American speakers.

		Response vowel																		
		iː	ɪ	eː	ɛ	ɑː	æ	uː	ʊ	ɔː	ɔ	o	ʌ	ʌr	ai	ɔi	ɑu	iə	uə	ɛə
Stimulus vowel	iː	**60**	19	1	9								1	4	1			3		1
	ɪ		**86**		7	1	3							1	1					
	eː	3	1	**40**	24		9		1	4	3			6				4		4
	ɛ		3		**82**		13						3							
	ɑː				1	**92**	1							3						3
	æ		1		36	3	**53**	1				1		3						1
	uː			3				**46**	38			8					3		3	
	ʊ								**72**		3	6	14	4						1
	ɔː			1		31	15		1	**10**	29	1	7		3		1			
	ɔ		1	2	1	41	11			2	**30**	2	7		2	1	1			
	o				1		1	4	6	13	3	**41**				6	18		6	
	ʌ				4		3	4	38		10	3	**29**	8				1		
	ʌr	1	1		3									**87**				4	1	1
	ai	3	3	1	1							1		3	**86**	1				
	ɔi			1	1			1		1		3	1		1	**91**			1	
	ɑu							6		7		3				4	**78**		1	
	iə				1	1	3					1	1	53		1		**33**	3	1
	uə					6		1	6		6	4		4	1	1	3		**66**	1
	ɛə		1	1	1					1				32				13		**49**

Table A6.2.7. Vowel identification (%): American listeners – Chinese speakers.

		Response vowel																	
		iː	ɪ	eː	ɛ	ɑː	æ	uː	ʊ	ɔː	ɔ	o	ʌ	ʌr	ai	ɔi	ɑu	iə	uə	ɛə
Stimulus vowel	iː	**71**	25	1		1					1									
	ɪ	53	36	1					6		3								1	
	eː	1		**85**	4		1				1				2	2	4		1	
	ɛ		3	8	**10**	3	7						1	1	63	1	1			
	ɑː			1		52	10	1		7	13		3	9					3	
	æ			6	4		**39**	3		1					46			1		
	uː		1	1				**45**	39			3	1				3	1	3	
	ʊ				1	1	4	46	**42**								4			
	ɔː		1	3		10	1	3		**18**		1		4	25	3	29			
	ɔ			2	2	6	2			18	**26**					2	42	2		
	o	3			1	1	3	3	29	1	4	**49**			1		3			
	ʌ		1	1	1	6	69	1		1	8		**9**		1	2				
	ʌr		1			14			4			1	1	**68**	1	1		1	6	
	ai		1	1		4	20			4	3		1		**61**		1	1		
	ɔi			1	3	1						4			37	**51**			3	
	ɑu	1			1		1	1	1	21	19	1		1	1	3	**46**			
	iə	1	1											21	1			**70**	1	3
	uə			1		3	1		1					7		6	1		**79**	
	ɛə					42	4			1				25	1	4		1		**20**

Table A6.2.8. Vowel identification (%): American listeners – Dutch speakers.

		Response vowel																	
		iː	ɪ	eː	ɛ	ɑː	æ	uː	ʊ	ɔː	ɔ	o	ʌ	ʌr	ai	ɔi	ɑu	iə	uə	ɛə
Stimulus vowel	iː	**82**	10		1							3		1				1		
	ɪ	4	**90**											1	4					
	eː		1	**80**	3	1					1	1				1	6		1	3
	ɛ			3	**84**		10									3				
	ɑː			1		**48**	7	1	4	13	14			6	1		1	1		1
	æ	1		1	71		**21**			1	1			1	1	2				
	uː						1	**74**	16			4			1		1		1	
	ʊ	1				1	1	44	**41**		3	1		1			1		4	
	ɔː	4				1		4	13	**16**	3	47		4	1	1	3		1	
	ɔ			1		7	1	1	8	10	**52**	7	4	1	1			3	1	
	o				1						1	**74**	1			10	6		6	
	ʌ		1	1	1	6	17			11	34	1	**24**	1						
	ʌr				2	2		2	6					**74**	2			5	9	
	ai	3	6	4	1		4		1	1	1				**76**					1
	ɔi		1	1	1						1	9		3	4	**78**				
	ɑu					1		1		4						4	**90**			
	iə	1						3						20			1	**67**		7
	uə			5				3		2	2		2			2	8	3	**76**	
	ɛə				6		1				1	1		15	1		3	7		**64**

Table A6.2.9. Vowel identification (%): American listeners – American speakers.

		Response vowel																		
		iː	ɪ	eː	ɛ	ɑː	æ	uː	ʊ	ɔː	ɔ	o	ʌ	ʌr	ai	ɔi	ɑu	iə	uə	ɛə
Stimulus vowel	iː	**86**	7											4	3					
	ɪ		**96**		3							1								
	eː		9	**60**	11	1	3	3	1					1			1	1		7
	ɛ				**96**		1		1					1						
	ɑː			1		**87**	1			1				7				1		
	æ		1	3		3	**89**		1		3									
	uː			1			1	**82**	10				1						3	
	ʊ				1	1			**86**			3		7						1
	ɔː	1		3		3	3	1		**39**	45			3	1					
	ɔ			1	2	1	7		2	29	**46**		3	2	3	1	4			
	o						1			1	7	**81**				6	3			
	ʌ	1	1		3		1		40		4		**38**	7	1			1	1	
	ʌr				1	1					3			**85**	3			3	1	1
	ai	3	3	1		1						1			**87**		1		1	
	ɔi			1	1			2	3	1		2		1		**87**	2		1	
	ɑu									1		1				3	**94**			
	iə		3	1	1	1	3							8	1			**79**	1	
	uə			1		7						1		4	3				**82**	
	ɛə			1		1	1	1						13				3		**79**

	legend
	0%
	1- 10%
	11- 20%
	21- 30%
	31- 40%
	41- 50%
	51- 60%
	61- 70%
	71- 80%
	81- 90%
	91-100%

Appendix A6.3. Dendrograms for 19 vowels for each of nine combinations of speaker and listener nationality.

Figure A6.3.1. Chinese listeners – Chinese speakers

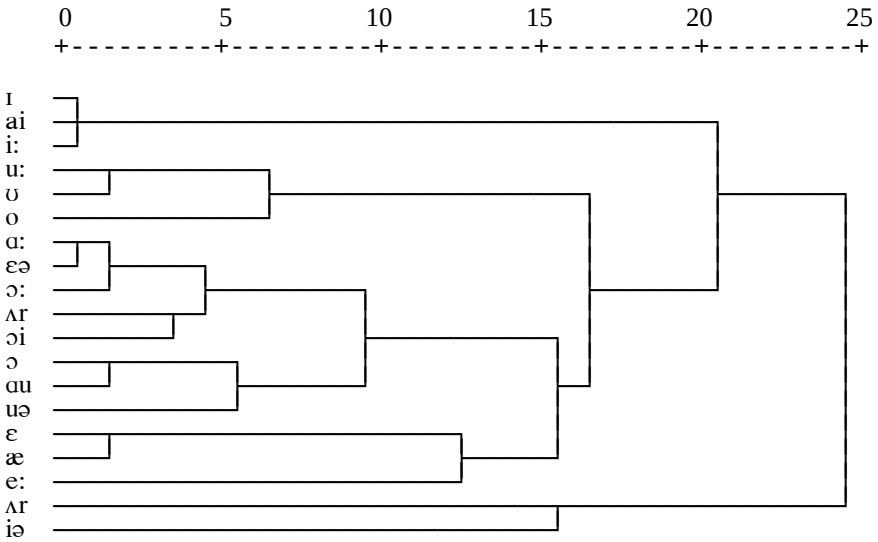


Figure A6.3.2. Chinese listeners – Dutch speakers

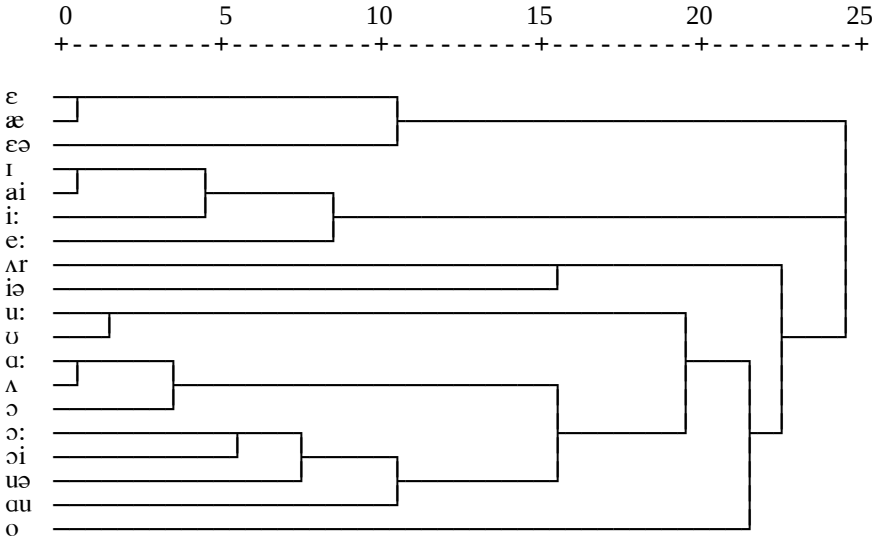


Figure A6.3.3. Chinese listeners – American speakers

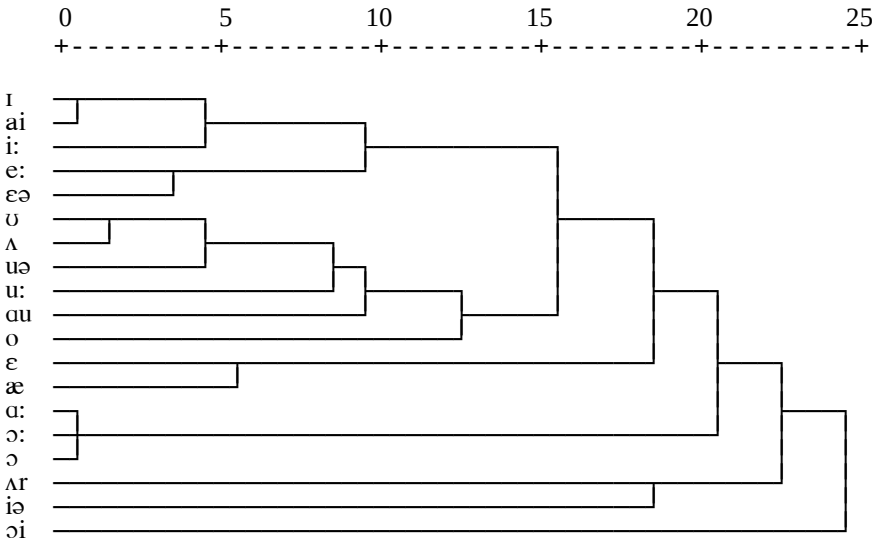


Figure A6.3.4. Dutch listeners– Chinese speakers

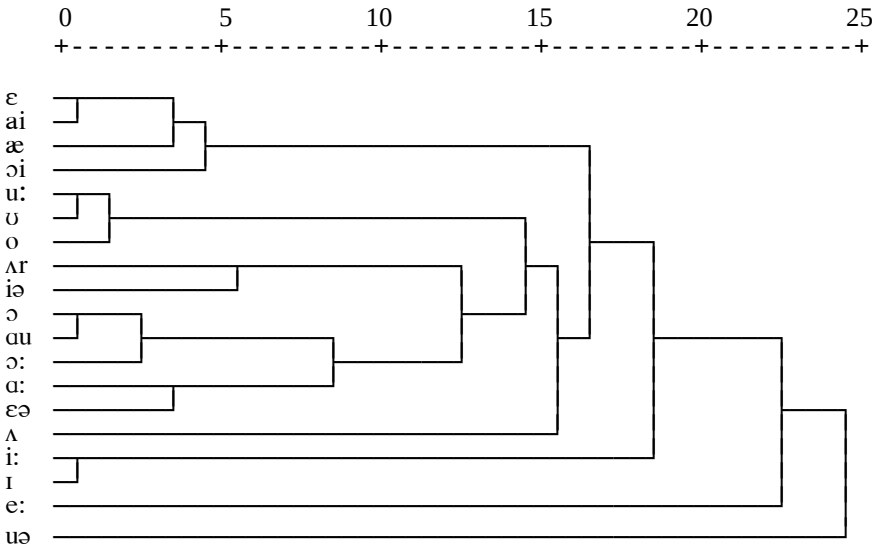


Figure A6.3.5. Dutch listeners – Dutch speakers

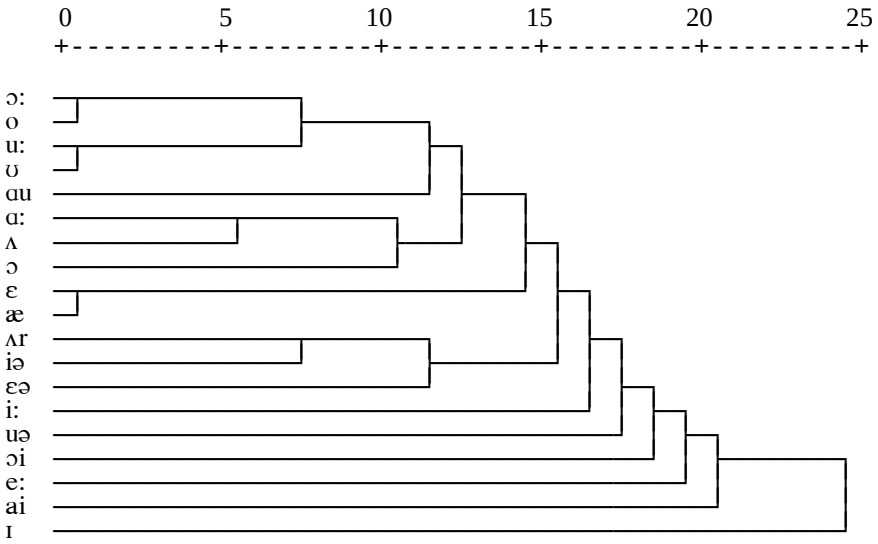


Figure A6.3.6. Dutch listeners – American speakers

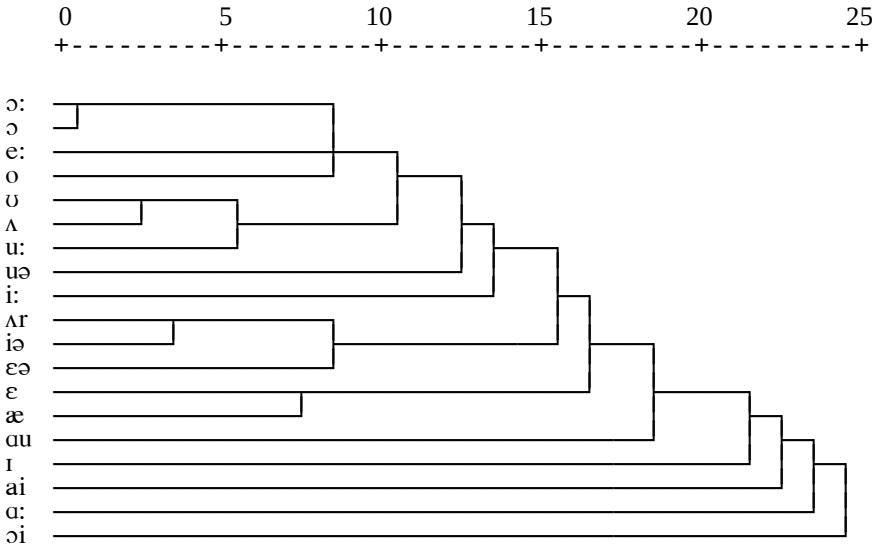


Figure A6.3.7. American listeners – Chinese speakers

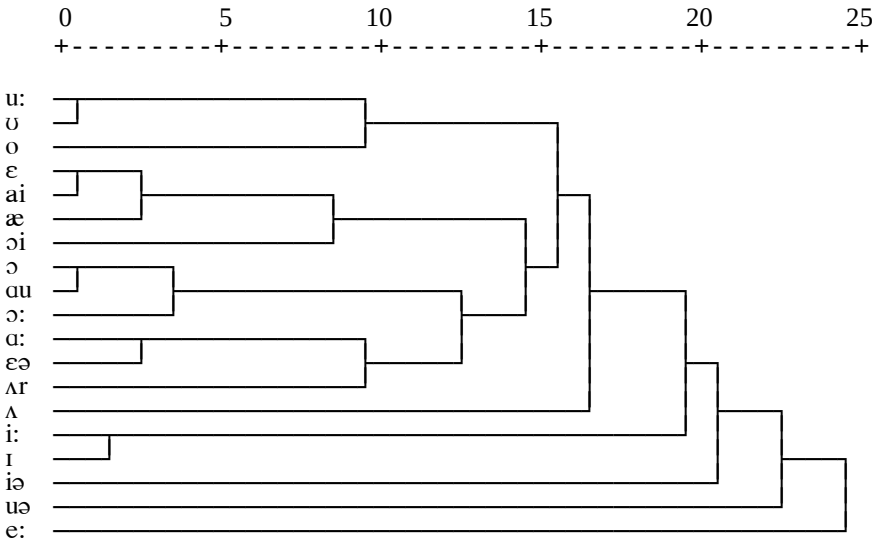


Figure A6.3.8. American listener – Dutch speaker

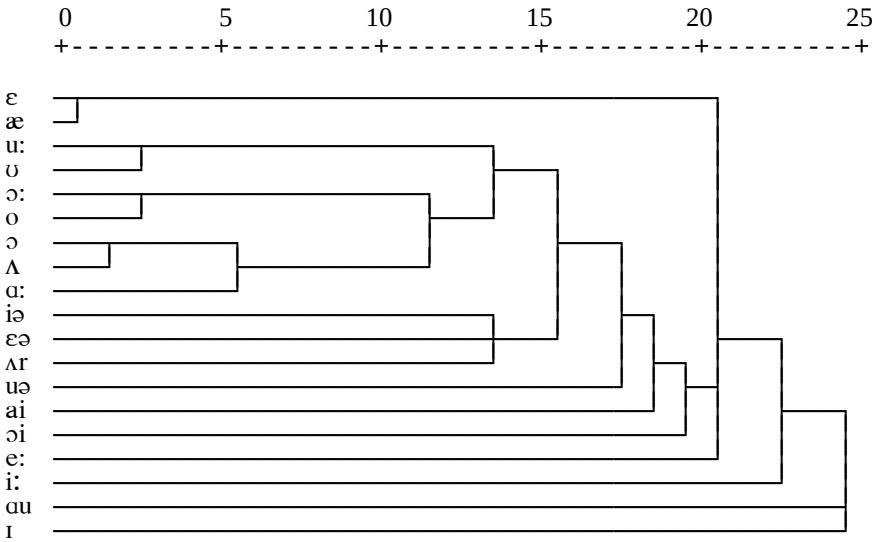
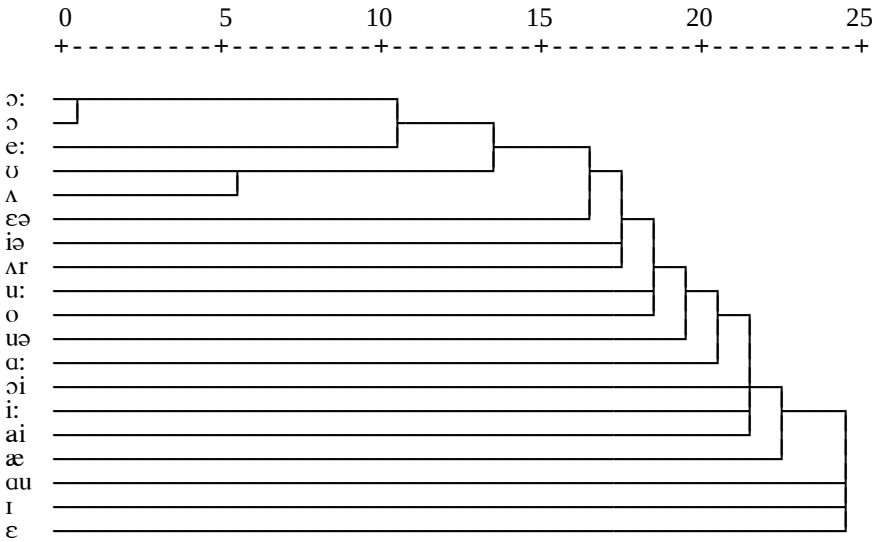


Figure A6.3.9. American listeners – American speakers



Appendix A7.1 Percent correct consonant identification broken down by language background of listener and of speaker. Mean, number of observations, standard deviation and standard error of the mean are indicated.

Nationality of		Mean	N	SD	Se
Listener	Speaker				
Chinese	Chinese	57.2	1800	49.5	1.2
	Dutch	46.8	1800	49.9	1.2
	USA	58.2	1800	49.3	1.2
	Total	54.1	5400	49.8	.7
Dutch	Chinese	66.6	1800	47.2	1.1
	Dutch	73.7	1800	44.1	1.0
	USA	80.6	1800	39.5	.9
	Total	73.6	5400	44.1	.6
USA	Chinese	70.5	1850	45.6	1.1
	Dutch	74.1	1850	43.8	1.0
	USA	83.4	1850	37.2	.9
	Total	76.0	5550	42.7	.6
Total	Chinese	64.8	5450	47.8	.6
	Dutch	64.9	5450	47.7	.6
	USA	74.2	5450	43.8	.6
	Total	68.0	16350	46.7	.4

Appendix A7.2 Confusion matrices for simplex consonants for each of nine combinations of speaker and listener nationality.

Table A7.2.1. Simplex consonants (%): Chinese listeners – Chinese speakers.

		Response Consonants																						
Stimulus Consonants		p	b	t	d	k	g	s	z	ʃ	ʒ	θ	ð	h	r	f	v	tʃ	dʒ	m	n	ŋ	l	y	w
	p	**93**	4							1		1													
	b	3	**76**		1							1			4	1	4						1		7
	t	31	1	**54**	3	3						1		6								1			
	d		43	1	**39**		4		3			3	1				3						1		1
	k	7				**82**	1	1							1	1	3	1				1			
	g		39		6		**43**								3				4			4			1
	s		1			1	1	**78**	3	10	1	3	1				1	1							1
	z							1	**71**		8	6	13								1				
	ʃ							3		**89**		3				1		4							
	ʒ						3	1	21	1	**28**	3	4					1	36			1			
	θ		1					43		4	1	**8**				40	1								
	ð		8		22	1	15		8		3	4	**7**			3	8		4				4		11
	h	21						1					6	**68**		1	1								1
	r								7		17	1	6		**43**		6		3	1			10	1	6
	f		3	1						1						**92**	3								
	v	1	13		1		1	1	1			1	1	1			**32**			1			1	1	40
	tʃ					1			1			1	1	1				**92**							1
	dʒ						1		22		18	1	3	1				6	**42**	3		1	1		
	m		31														3		1	**58**	1		1		4
	n		1	1	8		1		3		1		3				3		1	4	**42**	3	25		3
	ŋ				1	1	14				8	1	3		13	4	8		6	11	3	**4**	15		7
	l		1					1			1	1	1		3						1	1	**88**		
	y					1	1				3				1				7		1		26	**58**	
	w		3			1									1		21			1	3		4		**65**

Table A7.2.2. Simplex consonants (%): Chinese listeners – Dutch speakers

		Response Consonants																						
Stimulus Consonants		p	b	t	d	k	g	s	z	ʃ	ʒ	θ	ð	h	r	f	v	tʃ	dʒ	m	n	ŋ	l	y	w
	p	**46**	15	3	3	6	7				1	1		6		4	1	1		3				3	
	b		**46**	1	1						1	3	4	13		1	11			8	1		1		7
	t		1	**57**	4	1				4	1	6	6	1		1		14			1		1		
	d	1	10	3	**46**						3	4	8	7			1			1	1	11	1		1
	k	1		3		**88**		1	1		1		1				1					1			
	g			4	4	8	**53**		3			1	7	3	1							13		1	1
	s					1		**38**	1	19	6	11	4	3				10	6			1			
	z			1			1	4	**44**	3	15	7	13		1			3	4			3			
	ʃ							3		**82**	1	4	1					8							
	ʒ			1			1	3	1	47	**22**	3	1		4				11			1		3	
	θ	3					1	10	1	1	4	**8**	14	13	1	33	3	1				3	1	1	
	ð	1		7	19		7	1	10	3	4	11	**4**	3	3	4	1		6			6	7	3	
	h				1									**93**				1					1	1		1
	r	1			4		1		3		3	4	4	4	**40**	1	7		3		1	6	7	3	7
	f				1		1	3	1	3		8		4	1	**71**	3					3			
	v	1	15	1							1	4	6	3	3	50	**4**		1			3		1	6
	tʃ	1		1					1	3	7	6			1			**58**	19					1	
	dʒ					1	4		4		18	3		1				8	**57**			3			
	m		2		1					1	2	1	1	18	1		4	1		**47**	1	10	6		4
	n	1			1				3		3	4	10	4	1						**36**	6	28		3
	ŋ				6	1	21	1		1		6		4	3		4		3	1		**36**	7	3	3
	l	3	1								1	3	1	8	1	1	1			3		4	**61**		10
	y										6		1	3	7			1	15			1	1	**64**	
	w		1				1	1		4	3	1	4	14	4	1	32	1		3	1		3		**24**

Table A7.2.3. Simplex consonants (%): Chinese listeners – American speakers

Stimulus Consonants	Response Consonants																						
	p	b	t	d	k	g	s	z	ʃ	ʒ	θ	ð	h	r	f	v	tʃ	dʒ	m	n	ŋ	l	y	w
p	50	13	3	1						1	3		3		6	1			11		1	3	1	3
b	1	74				1				1	1	1			1	8			1	3	1			4
t	8		74	1	1					1	3	4	1		1		4							
d			7	63				6		1	1	13		1		1	1			1		3	1	
k			1	1	81	13							1			1				1				
g			3	8	3	67				1	1	1			1	3		1		1	6	1	1	
s			15		1		49	8	3	3	14	6					1							
z							6	56	1	10	14	10						3				1		
ʃ			1				6	1	54	17	7			3			7	3						1
ʒ						1		4	11	43	6	7		18				6			1	1	1	
θ	1	29		3			7				10	3	1	3	35	7					1			
ð		4	28	14				8		1	6	11	3			8		1		1		7	1	6
h	4	1		1					1		1	4	79	1	1		1				1			1
r		1	1							6	1	3	1	75				4		1		1	4	
f	2	12		1		1			1	1	3	1		1	74	2								
v	3	10		1							3		4	1	1	42			1					33
tʃ			1					1	1	3	4	3					78	6	1		1			
dʒ				3		4		4		25	4	3		3			4	49			1			
m		1								3	1			1		3		3	82	1	3			1
n				4					1	3	3	3	1	11			1		3	56	3	11		
ŋ		3		4	1	24					3			1		3				13	24	22	1	1
l								1		1									1	3		86	3	4
y	1									7	1			10	1	4		10	1			3	60	1
w										1			1	4		36			1	1			6	49

Table A7. 2.4. Simplex consonants (%): Dutch listeners – Chinese speakers

Stimulus Consonants	Response Consonants																						
	p	b	t	d	k	g	s	z	ʃ	ʒ	θ	ð	h	r	f	v	tʃ	dʒ	m	n	ŋ	l	y	w
p	97									1	1													
b		100																						
t	14		51								30	3		1	1									
d	1	1	10	61	1						7	17								1				
k	1			8	80	6				1				1			1							
g	14	7	11		7	46					7	3			1			1			1		1	
s							83	5	5	4	2						1		1					
z							4	72		13	6	4						1						
ʃ							4	1	83	6							4	1						
ʒ				10		7	1	11	3	13		13		1			3	38						
θ			1				35		3	8	42	1			7		3							
ð				32						1	13	35				11	1							6
h					3			1			1	1	90						1					1
r				1				21		13		4		47		1		7						6
f											7				87	4						1		
v								3			1				3	21								71
tʃ			1			1		1	4	4	13	1				3	68	1		1				
dʒ						1	3	27	1	16	1	10					4	33					3	
m																			100					
n		1									1									96		1		
ŋ						40					1						3		3	13	38		1	1
l												1	1									96		1
y									1			1	1		1			29		1		6	58	
w								1					1	1		4								92

Table A7.2.5. Simplex consonants (%): Dutch listeners – Dutch speakers

Stimulus Consonants	Response Consonants																							
	p	b	t	d	k	g	s	z	ʃ	ʒ	θ	ð	h	r	f	v	tʃ	dʒ	m	n	ŋ	l	y	w
p	**91**	3			6																			
b	3	**93**										1	1			1								
t			**72**								22	4					1							
d				**79**	1						4	14							1					
k				3	**97**																			
g					16	**74**	1		1		1						1	1			3			
s						1	**78**	8	8	3	1													
z							10	**71**	4	13								1						1
ʃ							3		**85**	8							4							
ʒ						4		3	34	**31**				1			3	20					4	
θ							6	18	1	6	**40**	7		1	14	7								
ð			4	24		1		4			43	**19**		1	1		1							
h			1							1		1	**94**											1
r					1	1								**85**	1	3								8
f			1		1						14	1			**74**	8								
v			1								3			1	43	**49**								3
tʃ						1			1	3	19	4					**61**	10						
dʒ				3		4		3	1	7	1	16					9	**53**					3	
m						1								1				1	**96**	1	1			
n		1									1					1				**93**		3		
ŋ						3						1	1						1		**93**			
l												1							1			**97**		
y						1			1	1					1			29					**65**	
w		1												1		15	1					1	1	**77**

Table A7.2.6. Simplex consonants (%): Dutch listeners – American speakers.

Stimulus Consonants	Response Consonants																							
	p	b	t	d	k	g	s	z	ʃ	ʒ	θ	ð	h	r	f	v	tʃ	dʒ	m	n	ŋ	l	y	w
p	**94**			1							3	1												
b		**96**		3					1															
t			**71**							1	22	3					3							
d			1	**85**							3	11												
k					**93**	1						1			1		3							
g				1	6	**89**															4			
s						3	**83**	4	4	1	3		1											
z							10	**81**	3	6								1						
ʃ						1	1		**76**	8							10	3						
ʒ						4		6	17	**56**								17						
θ	3		1							1	**83**	4			4									3
ð			3	21			3	3		1	21	**38**				10								1
h										3	1	3	**83**	1	1	1		3			1	1		
r						1					1			**94**								1		1
f	1		1								7			1	**82**	7				1				
v		3		1					1		1	6	3		1	6	**62**							15
tʃ			1			1			1	6	11	6					**72**	1						
dʒ						6		3	3	10		13					6	**52**	1				6	1
m				1															**93**	4			1	
n			1												1					**96**		1		
ŋ	1					4									1		1				**92**			
l										1								1				**97**		
y		1											1			1		15					**80**	
w											1												3	**96**

Table A7.2.7. Simplex consonants (%): American listeners – Chinese speakers.

Stimulus Consonants	Response Consonants																							
	p	b	t	d	k	g	s	z	ʃ	ʒ	θ	ð	h	r	f	v	tʃ	dʒ	m	n	ŋ	l	y	w
p	**99**														1									
b		**99**								1														
t	3		**83**						1		10	3												
d		4	6	**79**	1				1			8												
k	1			1	**86**									3			9							
g	11	8	4	6	3	**61**					4	1											1	
s			1			1	**83**	10	1	1	2			1										
z						1	3	**77**		13	1	3				1								
ʃ							3		**93**	3							1							
ʒ				7		4		3		**3**							6	75	1				1	
θ			1				40		4	1	**47**	1			4									
ð			1	41				3			17	**33**				3							1	
h		1			1	1		1			3	1	**89**										1	
r	1	1					1	6	1	23		4		**44**	1			10					4	3
f			4								1		3	1	**86**	3		1						
v							2					2				**83**	2	3						9
tʃ			3	1					1	3		1					**90**							
dʒ						1		17		14							1	64					1	
m	1																		**97**	1				
n														1		1			1	**96**				
ŋ					1	31	3							1	1	1		3	10	10	**38**	1		
l							1					1							1			**94**	1	
y										1											1	11	**86**	
w		1												7		28			1		1	4		**56**

Table A7.2.8. Simplex consonants (%): American listeners – Dutch speakers.

Stimulus Consonants	Response Consonants																							
	p	b	t	d	k	g	s	z	ʃ	ʒ	θ	ð	h	r	f	v	tʃ	dʒ	m	n	ŋ	l	y	w
p	**87**		1	1	6						3						1							
b		**96**														4								
t			**83**	1						4	7						3	1						
d			1	**93**								6												
k					**93**	1							1				4							
g					4	**90**	1						1				3							
s							**80**	6	10	4														
z						3	4	**58**	3	29	1							1						
ʃ			1					1	**89**	7							1							
ʒ		1				1	1	4	38	**42**							1	10						
θ							8	7		1	**43**	21			18	1								
ð			6	27		3					46	**15**							1					1
h						1			1		1		**93**					1			1			
r														**93**		1			1				1	3
f	1		4				4				1				**86**	1					1			
v		1	6					1		1	1	3		1	53	**31**								
tʃ									1	3							**86**	10						
dʒ						1		1		7			1				10	**79**						
m				1									2			1			**97**					
n	1	1	1			1	1						1							**86**	1	4		
ŋ						1					1										**96**		1	
l				1				1				1	1									**94**		
y			1			1							1					1	1		1		**90**	1
w		3										1		4		57						9		**26**

Table A7.2.9. Simplex consonants (%): American listeners – American speakers.

Stimulus Consonants \ Response Consonants	p	b	t	d	k	g	s	z	ʃ	ʒ	θ	ð	h	r	f	v	tʃ	dʒ	m	n	ŋ	l	y	w
p	**94**	4														1								
b		**93**	3		1	1														1				
t			**85**				1		1		11									1				
d			3	**97**																				
k			1	3	**86**	4					1				1		3							
g			1		7	**86**						1									4			
s			1			1	**93**	4																
z						1	1	**87**		10														
ʃ	1								**89**	6	1						1			1				
ʒ						1		4	4	**59**							1	30						
θ	4						7	3		1	**79**	4			1									
ð		4	18	15				4		3	21	**25**				6		1					1	
h			1	1								3	**89**	3								1	1	
r													4	**90**				1						4
f			3				1		1	1	3	1	1		**88**	1			1					
v		13	1			1		1			3					**76**							1	3
tʃ																	**94**							6
dʒ						3		1	1									**93**					1	
m																3			**96**			1		
n														1						**99**				
ŋ																					**99**	1		
l					1																	**99**		
y	1									1					1			1					**93**	1
w											1									1			1	**96**

legend
0%
1- 10%
11- 20%
21- 30%
31- 40%
41- 50%
51- 60%
61- 70%
71- 80%
81- 90%
91-100%

Appendix A7.3. Dendrograms for consonant confusions for each of nine combinations of speaker and listener nationality.

Figure A7.3.1. Chinese listeners – Chinese speakers

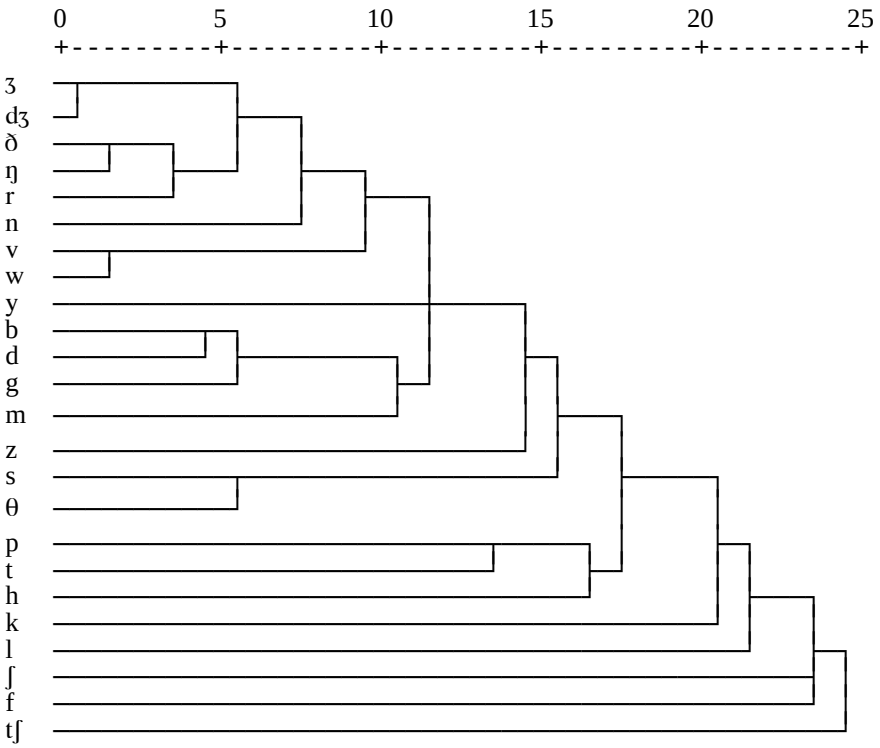


Figure A7.3.2. Chinese listeners – Dutch speakers

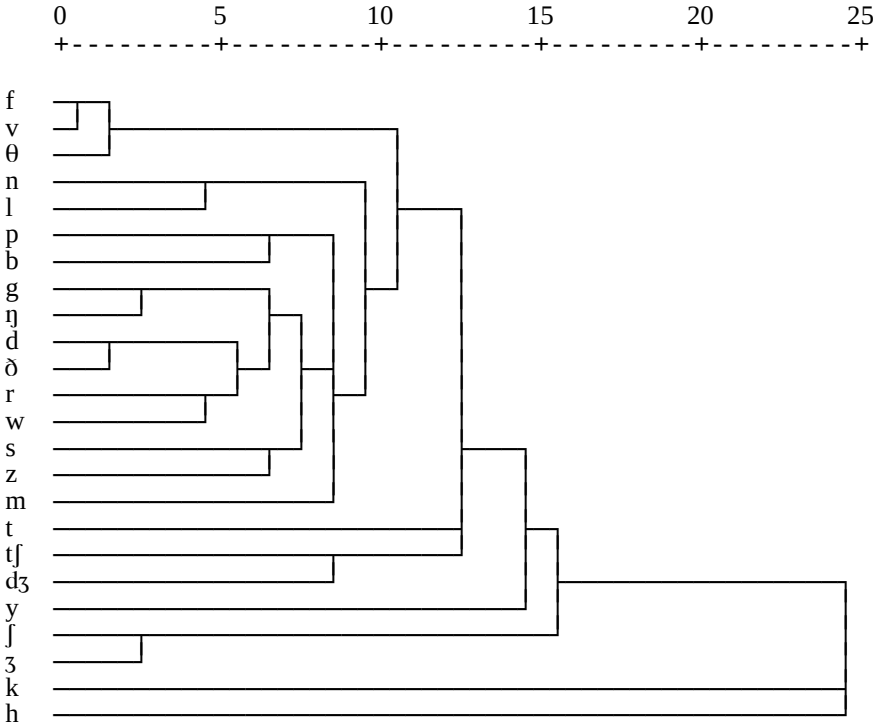


Figure A7.3.3. Chinese listeners – American speakers

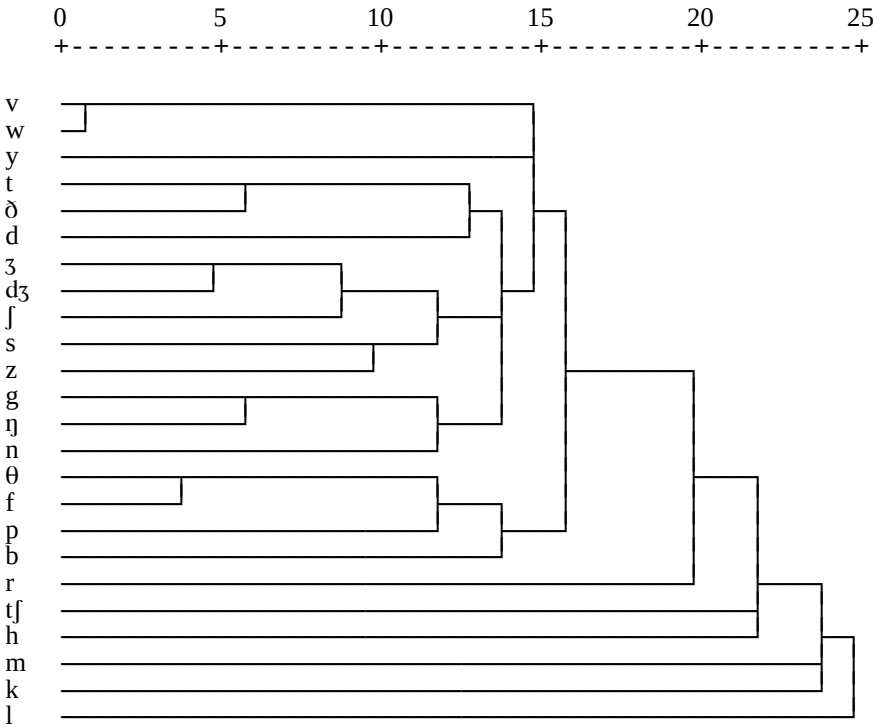


Figure A7.3.4. Dutch listeners – Chinese speakers

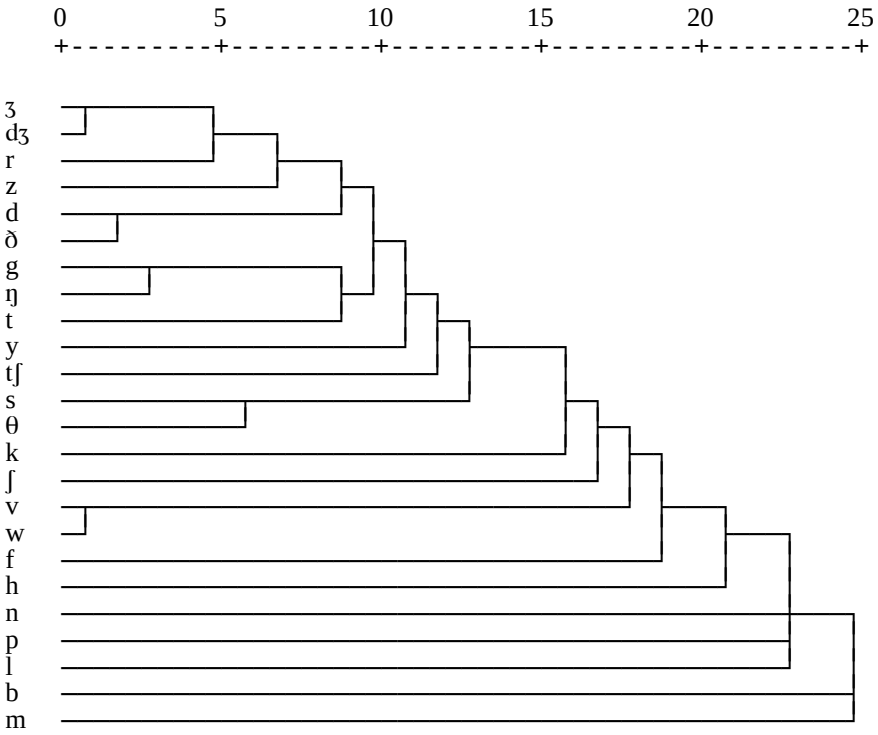


Figure A7.3.5. Dutch listeners – Dutch speakers

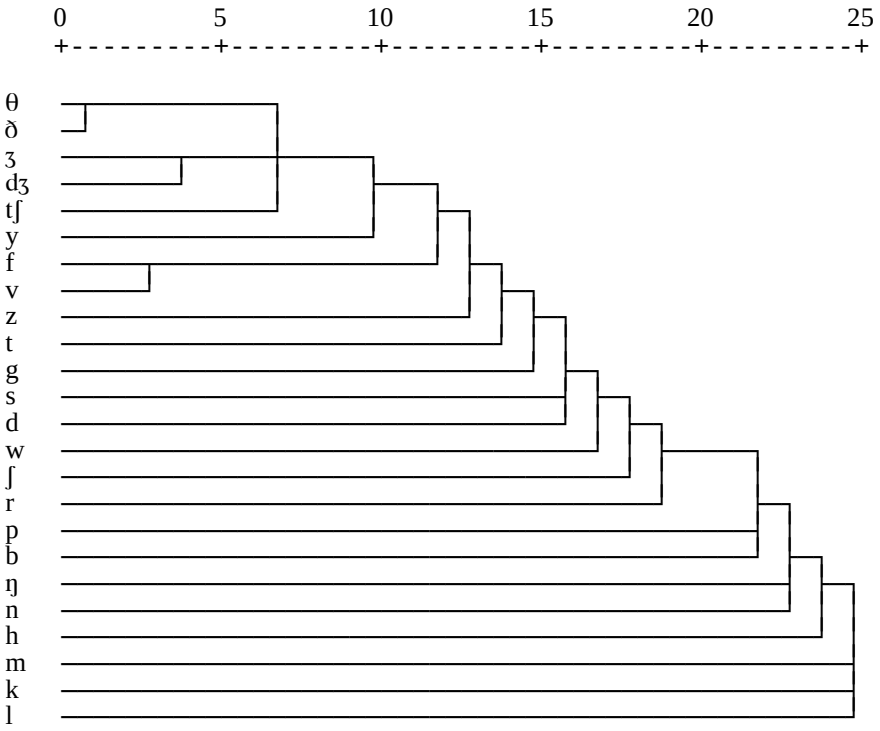


Figure A7.3.6. Dutch listeners – American speakers

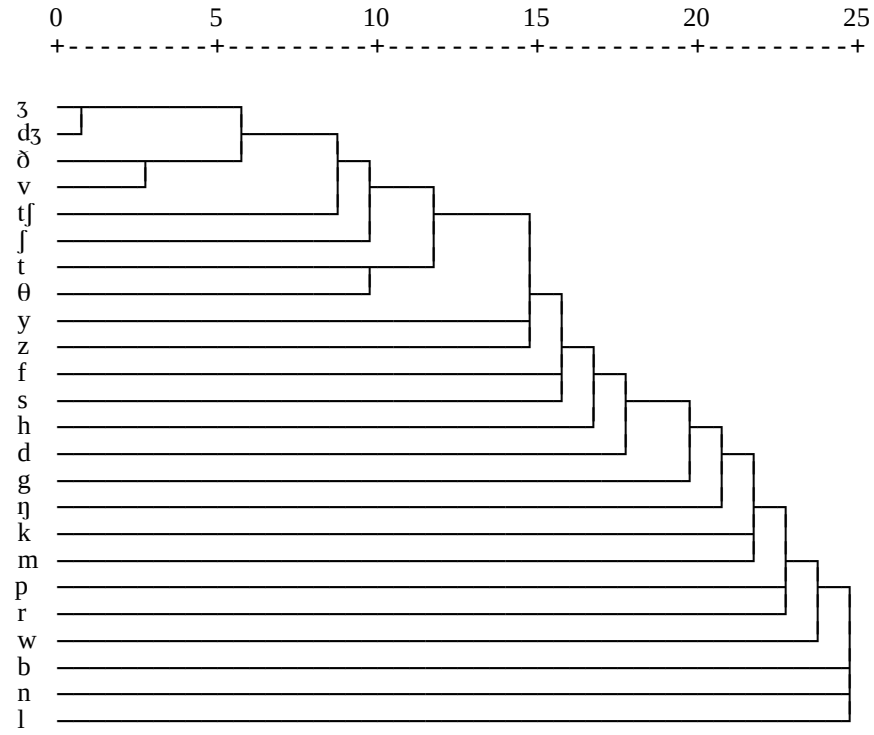


Figure A7.3.7. American listeners – Chinese speakers

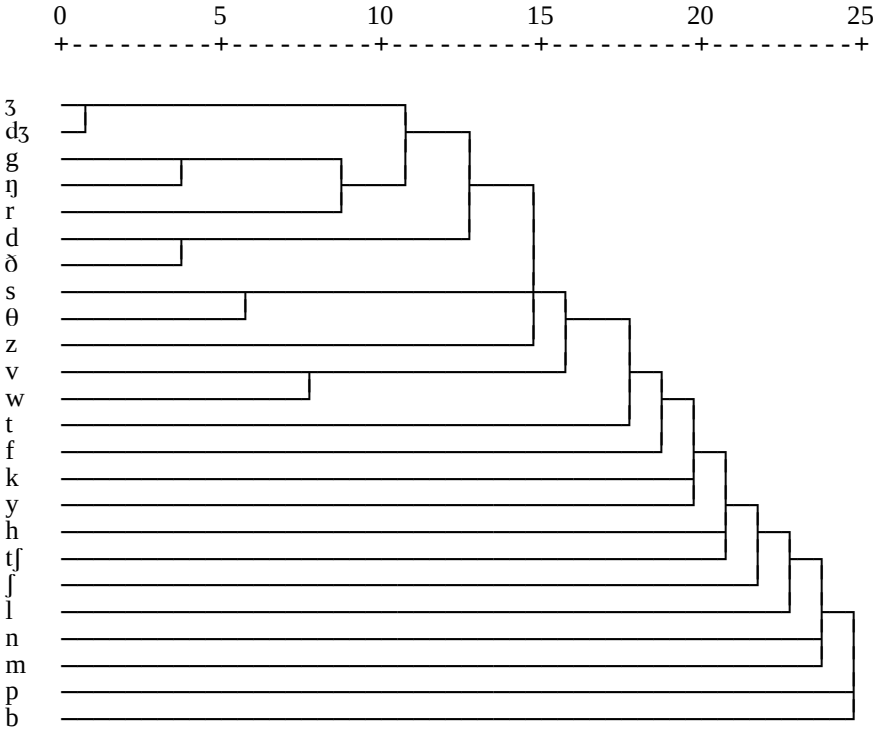


Figure A7.3.8. American listeners – Dutch speakers

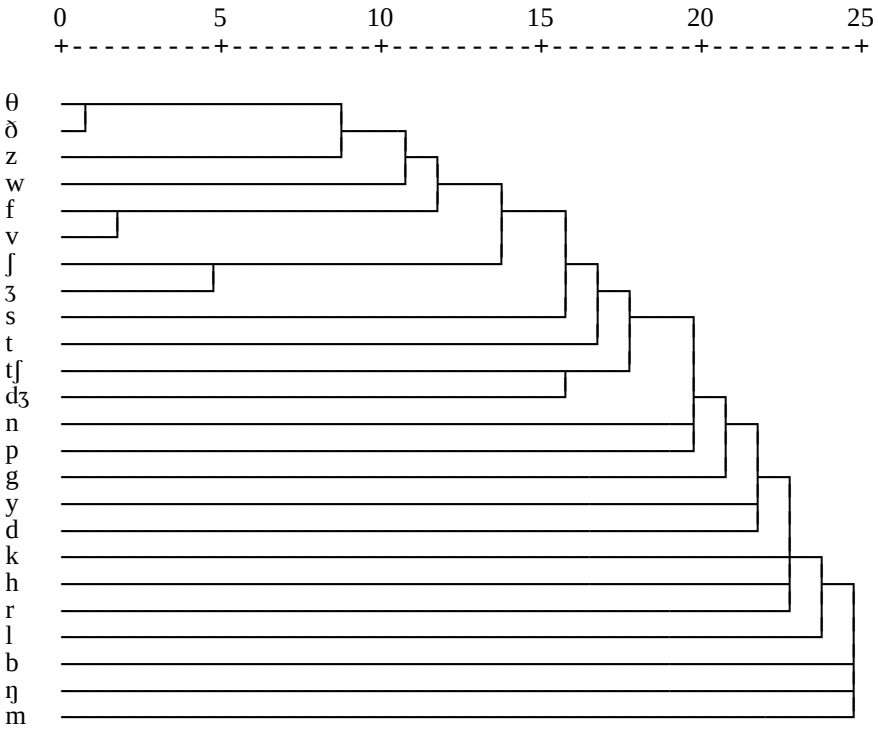
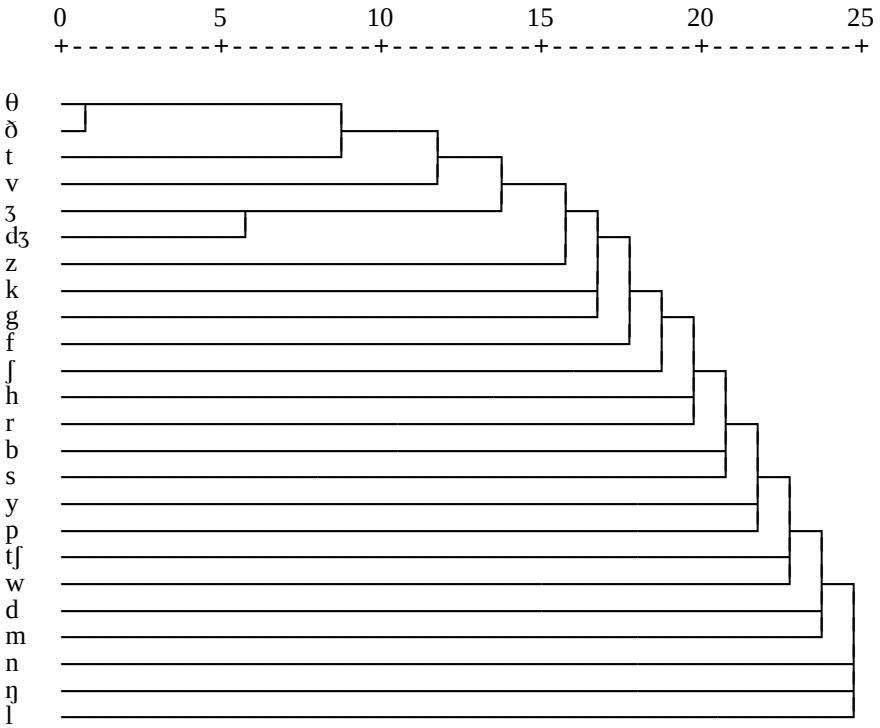


Figure A7.3.9. American listeners – American speakers



Appendix A8.1. Percent correct consonant clusters identification broken down by language background of listener and of speaker. Mean, number of listeners, standard deviation and standard error of the mean are indicated.

Nationality of		Mean	N	SD	Se
Listener	Speaker				
Chinese	Chinese	52.8	36	13.7	2.3
	Dutch	36.9	36	13.3	2.2
	USA	56.0	36	15.5	2.6
	Total	48.5	108	16.4	1.6
Dutch	Chinese	78.8	36	10.8	1.8
	Dutch	87.8	36	11.7	1.9
	USA	89.1	36	9.0	1.5
	Total	85.2	108	11.4	1.1
USA	Chinese	82.5	36	9.1	1.5
	Dutch	85.7	36	9.7	1.6
	USA	89.3	36	8.7	1.4
	Total	85.8	108	9.5	0.9
Total	Chinese	71.3	108	17.4	1.7
	Dutch	70.1	108	26.3	2.5
	USA	78.2	108	19.4	1.9
	Total	73.2	324	21.6	1.2

Appendix A8.2 Confusion matrices for consonant clusters for each of nine combinations of speaker and listener nationality.

Table A8.2.1. Consonant clusters (%): Chinese listeners – Chinese speakers.

		Response clusters																				
Stimulus clusters		pl	bl	pr	br	tr	dr	kr	gr	kl	gl	spr	spl	skr	sp	st	sk	sm	sn	sl	sw	θr
	pl	74	7	7	3					3			4	1						1		
	bl	8	68		10		1		6	1	4		1									
	pr	18		58	6			4	1	1		4	4	1	1							
	br	1	11	6	53		1		11		8	1	4		1	1						
	tr		1		1	68	8	3		1	3	3		1						1		8
	dr	1	3	3		1	52	4	6	1	11	1	3			4	1				4	3
	kr	11		10	4			53	1	3		3	4	4	4		1			1		
	gr				1			1	57		8	3	11	10		1	4		3			
	kl	10		3				14	8	49	7		3	1	1			1	1			1
	gl		6	1					14	3	54		3	1	3	1	1			8	1	3
	spr		4		4		1	3	4		1	40	14	13	7	1	1				3	3
	spl	4	1	3	1	1	1		1		1	32	31	6	6	1	4	3			3	
	skr							10	1		3	6	4	67	6		1	1		1		
	sp	3	3	8	6						1	10	8	1	56			3				1
	st	1				1	4				1	3	3	3	31	38	10	3	1			1
	sk				1		1	3	1	1	3	4	3	14	4	11	51			1		
	sm				1					1			1	1	8			81	3	1		1
	sn						1	3			1			11	1	4	18	3	50	4		3
	sl	1	1							1	6	1	21	8		1	6		1	47		4
	sw									1	3	10	4	1	7	3		8	3	6	49	6
	θr	4	1	3	3		3	1	4	4	4	10	11	15	4	4	3		3	3	3	17

Table A8.2.2. Consonant clusters (%): Chinese listeners – Dutch speakers.

		Response clusters																				
Stimulus clusters		pl	bl	pr	br	tr	dr	kr	gr	kl	gl	spr	spl	skr	sp	st	sk	sm	sn	sl	sw	θr
	pl	31	21	7			4	1	1	1	1	6	7	4	1	1			1	3	3	6
	bl	11	17	6	4		3	1	1		6		3	1		3		7	29	3	1	4
	pr	4	3	35	4	7	1	17	6	3	1	4		1	1	3			4			6
	br	3	6	8	35		1	4	28		7		1	3					1			3
	tr			1		68	3	4	3	1		1		3		1						14
	dr		4	1	1	4	36		22	4	6	3		3		4	1			3	1	6
	kr			10	1	4	4	54		8	3			8	1		1		1			3
	gr	3		7	7		1	8	46	4	7	1	4	1	1	1	3					4
	kl	17	1	3		6	6	10		35	4	4	4	3							3	4
	gl	6	4	4	1	3	11		10	14	39			1			1				1	4
	spr	1	4		3	1	3	3	1	1	1	35	3	26	3	1	3	1		1	3	4
	spl	1		1		1		4	3	11	1	4	39	17		3	6			3		6
	skr			1		3		7	6	34	1	6	7	15	6	1	4				1	4
	sp	4	3		1	3	4		1			10	6	3	35	3	8	6		7	1	6
	st	1		3	1	6	1	1		3	1	14	1	6	14	20	4	1		1		20
	sk					4	1	4	1	4	3	3	4	18		6	40			3	1	7
	sm				1	3	1	3	1	1			1		1	3		58	13	10	1	1
	sn			1	1					1		3	3	1	4			8	53	14		10
	sl			1	3	1	3	1	3	3		3	14	7	3	1	1	4	4	39		8
	sw	4				3	3			1		17	4	15	7	3	1		1	6	15	19
	θr	4		4		11	6	1		3		6	4	11	8	6	3		1	4	4	24

Table A8.2.3. Consonant clusters (%): Chinese listeners – American speakers.

	Response clusters																					
Stimulus clusters		pl	bl	pr	br	tr	dr	kr	gr	kl	gl	spr	spl	skr	sp	st	sk	sm	sn	sl	sw	θr
	pl	**54**	1	4			3	1		7	4		8	1	4		3	1	1	3		3
	bl	7	**61**	1	4		1		6	3	14		1							1		
	pr	11	3	**65**	6			4	1		1	1	4							1		1
	br		10	8	**63**	1		1	8	1	4					1						1
	tr				1	**63**	8		1	1				1	3		1			3		17
	dr		1	1	1	15	**58**	3	3		3			1				1			1	10
	kr	3		4				**74**	1	8				7		3						
	gr				1			1	**72**	3	10	1	1	4	4		1					
	kl				1	4	29	1	3	**40**	3		3	7		1	3			1	1	1
	gl	4	18		3		4	3	13	4	**38**	1	3		1				1	4	1	1
	spr		1		11		1		1		1	**57**	10	4	6	3	1					3
	spl	1		1						3		4	**76**	7	3					4		
	skr			1				6	4	1		1	7	**50**	3	4	17	3	1	1		
	sp		1	7	1					1		8	8	3	**61**	4		3			1	
	st			1			3	4	1	1	3	8	4	3	10	**29**	6	3	1	8		14
	sk	25	6	1		3		1	7	4	6	13	9	3	3		**16**	1				
	sm					1		1				1	1	3	4	1	4	**74**	1	3	1	3
	sn				1	1			3			1	3		1	4		4	**75**	1		4
	sl		1	1	3			1			4	3	4	1	4	4	1		4	**56**	3	8
	sw		1	3		4		1		1	1	1	1				1	3		3	**71**	7
	θr	11	3	3	4		1	1	6	1	1	6	13	1	4	1	3	3		8	6	**24**

Table A8.2.4. Consonant clusters (%): Dutch listeners – Chinese speakers.

	Response clusters																					
Stimulus clusters		pl	bl	pr	br	tr	dr	kr	gr	kl	gl	spr	spl	skr	sp	st	sk	sm	sn	sl	sw	θr
	pl	**85**	1	1							3		7		3							
	bl	6	**85**		8											1						
	pr	7		**85**								6	1		1							
	br		1	1	**94**		1															1
	tr				1	**62**	4							1		4	1				3	23
	dr			3	3		**78**		1		1											13
	kr			3			1	**92**	3		1											
	gr							7	**87**	1	3				1							
	kl							25	12	**54**	7			1		1						
	gl							1	3	15	**79**									1		
	spr			3	10					1		**77**	3		1		1				1	1
	spl		11	1	4							33	**40**		8					1		
	skr							1	7			1		**85**			1	1		1	1	
	sp		1	1								9	9	1	**74**			3	1			
	st					1			1	1				1	6	**84**	1					3
	sk							3		4	1			3	3	1	**82**	1	1			
	sm												1					**99**				
	sn								1				1		1	1			**93**	1		
	sl		4										4			1			1	**89**		
	sw							1													**99**	
	θr					6	9				1	4	6	3	1	4	1	1		3	4	**56**

Table A8.2.5. Consonant clusters (%): Dutch listeners – Dutch speakers.

		Response clusters																				
		pl	bl	pr	br	tr	dr	kr	gr	kl	gl	spr	spl	skr	sp	st	sk	sm	sn	sl	sw	θr
Stimulus clusters	pl	**88**	4	1								1	6									
	bl		**93**				1												4		1	
	pr	4	1	**90**			1		1						1							
	br		1		**97**				1													
	tr					**81**	4												1			14
	dr		1				**93**		1							1						3
	kr							**85**	3	3				8								1
	gr							11	**83**	3				3								
	kl									**90**	10											
	gl									7	**93**											
	spr			6								**76**	3	1	3	3		1		1	1	4
	spl	4											**93**		1					1		
	skr							6	6	3	3	11	3	**64**	1	1	1					1
	sp			1								6	8		**85**							
	st					1				1						**96**	1					
	sk							1		4				7	3	3	**77**	4				
	sm																	**99**	1			
	sn															3			**97**			
	sl						1				1		1				1			**94**		
	sw												1			3	1	3			**92**	
	θr					4	1			1			3	3	4	1		1		3	3	**75**

Table A8.2.6. Consonant clusters (%): Dutch listeners – American speakers.

		Response clusters																				
		pl	bl	pr	br	tr	dr	kr	gr	kl	gl	spr	spl	skr	sp	st	sk	sm	sn	sl	sw	θr
Stimulus clusters	pl	**83**	3			1				3			4							4		1
	bl		**100**																			
	pr	6		**85**	3			1				4			1							
	br		3		**96**				1													
	tr					**79**	3			1						3						14
	dr				1		**99**															
	kr			1				**90**	4	1				1			1					
	gr				1			3	**92**	1				3								
	kl						3	1		**86**	7		1			1						
	gl		1						3	13	**81**			1								1
	spr			1								**92**	4		1					1		
	spl	1										1	**94**							3		
	skr							1	1			4		**88**		1	4					
	sp											1	4	4	**89**	1						
	st									1					1	**92**	4					1
	sk	3								7			1	4	1		**81**				1	
	sm													1				**92**	6		1	
	sn																		**100**			
	sl														1	1	1	3		**90**	3	
	sw				1																**99**	
	θr					15	1			1			1				1				1	**77**

Table A8.2.7. Consonant clusters (%): American listeners – Chinese speakers.

Stimulus clusters	Response clusters																					
		pl	bl	pr	br	tr	dr	kr	gr	kl	gl	spr	spl	skr	sp	st	sk	sm	sn	sl	sw	θr
	pl	**84**	3	1		1							4	3	1			1				
	bl		**89**	1	7					1			1									
	pr			**94**								3		1	1							
	br				**100**																	
	tr		5		2	**81**	2									2					3	6
	dr				2	3	**77**				5						2			2	8	3
	kr							**99**									1					
	gr				1		3		**92**		3								1			
	kl							33	7	**55**	1						1			1		
	gl								7	6	**81**	1		1	1	1	1					
	spr		1		1		1					**90**			4						1	
	spl	3	6	1	1							26	**56**		6					1		
	skr							4	4		1	3		**87**								
	sp	1						1	1			7	6		**82**		1					
	st					3								1	3	**87**	1				3	1
	sk						1	1	3	1	4	1		7		3	**76**		1			
	sm																1	**99**				
	sn																	3	**96**	1		
	sl									1			1	1			1		1	**90**	1	1
	sw				1	1	1	1					3	1	3					1	**85**	
	θr			1		3					1	1		6					1		6	**80**

Table A8.2.8. Consonant clusters (%): American listeners – Dutch speakers.

Stimulus clusters	Response clusters																					
		pl	bl	pr	br	tr	dr	kr	gr	kl	gl	spr	spl	skr	sp	st	sk	sm	sn	sl	sw	θr
	pl	**81**	6		3							1	7	1								
	bl		**90**				1					1	3						1		3	
	pr	3	1	**85**	**1**	1	1	1				4	1									
	br			1	**99**																	
	tr					**83**	7							1			1					7
	dr			1	3	1	**87**		1	3				3								
	kr						1	**93**						4			1					
	gr	1						3	**93**		1			1								
	kl	3						3		**92**	3											
	gl	1		1					4	6	**87**											
	spr											**93**		1			1		3		1	
	spl										1	3	**92**	1	3							
	skr							11		3		7		**76**		1				1		
	sp								1		1	7	7		**82**						1	
	st			4						1	1	1	1		3	**83**	1			3		
	sk							4						10	3	1	**81**					
	sm						1			3	1				1		1	**90**			1		
	sn												1					1	**96**	1		
	sl		1										3						3	**93**		
	sw											1	3	9	4	1				4	**77**	
	θr					7		1	1			3	1	6	3	1	1			4	4	**66**

Table A8.2.9. Consonant clusters (%): American listeners – American speakers.

Stimulus clusters \ Response clusters	pl	bl	pr	br	tr	dr	kr	gr	kl	gl	spr	spl	skr	sp	st	sk	sm	sn	sl	sw	θr
pl	83	1	1			3			3			9									
bl		93	3	3										1							
pr			94								1		1	3							
br		3		97																	
tr					87	3								1	1					1	6
dr						90							4	1			1			3	
kr							94	3	1				1								
gr								100													
kl						1	3	3	90				3								
gl						1		3	3	89	1						1		1		
spr							1		1		96		1								
spl	1		1							1	3	90							3		
skr				1	1		1				1		90				1		1		1
sp											1	3		96							
st												1		1	93	1				3	
sk	1						1					1	6	1	3	86					
sm																1	94	4			
sn										3					1			92	1	3	
sl			1					4			3			1					90		
sw		1									3									96	
θr			1		17	3	1		1		4			1	1	1					67

legend	
	0%
	1- 10%
	11- 20%
	21- 30%
	31- 40%
	41- 50%
	51- 60%
	61- 70%
	71- 80%
	81- 90%
	91-100%

Appendix A9.1. SUS sentences. Percent correct word recognition broken down by language background of listener and of speaker. Mean, number of listeners, standard deviation and standard error of the mean are indicated. Scoring unit is the content word for Word scores. For sentence scores all the content words in a sentence have to be reported correctly for a sentence to be correct.

Nationality of		Word scores				Sentence scores			
Listener	Speaker	Mean	N	SD	Se	Mean	N	SD	Se
Chinese	Chinese	39.3	35	9.5	1.6	4.9	35	7.8	1.3
	Dutch	39.0	35	11.0	1.9	5.7	35	7.0	1.2
	USA	44.2	35	11.3	1.9	4.9	35	7.0	1.2
	Total	40.8	105	10.8	1.1	5.1	105	7.2	0.7
Dutch	Chinese	57.1	36	9.0	1.5	16.7	36	12.6	2.1
	Dutch	86.2	36	8.8	1.5	60.3	36	18.7	3.1
	USA	90.5	36	6.5	1.1	71.1	36	16.9	2.8
	Total	77.9	108	17.0	1.6	49.4	108	28.6	2.8
USA	Chinese	59.5	36	8.1	1.3	18.3	36	9.4	1.6
	Dutch	83.0	36	6.0	1.0	51.9	36	14.9	2.5
	USA	95.5	36	4.1	0.7	85.0	36	13.0	2.2
	Total	79.3	108	16.2	1.6	51.8	108	30.1	2.9
Total	Chinese	52.1	107	12.6	1.2	13.4	107	11.7	1.1
	Dutch	69.7	107	23.2	2.2	39.6	107	27.9	2.7
	USA	77.0	107	24.4	2.4	54.1	107	37.3	3.6
	Total	66.3	321	23.2	1.3	35.7	321	32.4	1.8

Appendix A9.2. Low-predictability (LP) and High-predictability (HP) SPIN sentences. Percent correct word recognition broken down by language background of listener and of speaker. Mean, number of listeners, standard deviation and standard error of the mean are indicated. In the left part of the table the results are given for low-predictability contexts, in the right part the scores obtained for the high-predictability targets are listed.

Nationality of		LP targets				HP targets				All targets			
Listener	Speaker	Mn	N	SD	Se	Mn	N	SD	Se	Mn	N	SD	Se
Chinese	Chinese	19.4	36	15.2	2.5	16.7	36	10.4	1.7	17.7	36	9.3	1.5
	Dutch	38.9	36	15.8	2.6	37.8	36	19.6	3.3	39.2	36	14.9	2.5
	USA	17.9	36	10.5	1.7	31.8	36	12.5	2.1	24.8	36	8.4	1.4
	Total	25.4	108	16.9	1.6	28.7	108	17.1	1.6	27.3	108	14.3	1.4
Dutch	Chinese	26.9	36	16.1	2.7	33.1	36	11.4	1.9	30.7	36	10.4	1.7
	Dutch	81.3	36	12.1	2.0	76.1	36	21.3	3.5	79.7	36	12.2	2.0
	USA	77.8	36	13.4	2.2	84.9	36	14.1	2.3	81.8	36	11.7	2.0
	Total	62.0	108	28.6	2.7	64.7	108	27.8	2.7	64.1	108	26.3	2.5
USA	Chinese	39.4	36	12.7	2.1	57.8	36	11.0	1.8	50.9	36	9.2	1.5
	Dutch	67.7	36	14.5	2.4	99.4	36	3.3	0.6	77.8	36	10.2	1.7
	USA	95.2	36	9.0	1.5	99.1	36	4.1	0.7	97.4	36	5.3	0.9
	Total	67.4	108	26.0	2.5	85.4	108	20.8	2.0	75.3	108	20.9	2.0
Total	Chinese	28.5	108	16.8	1.6	35.8	108	20.1	1.9	33.1	108	16.7	1.6
	Dutch	62.6	108	22.7	2.2	71.1	108	30.5	2.9	65.6	108	22.5	2.2
	USA	63.6	108	35.1	3.4	71.9	108	31.1	3.0	68.0	108	32.5	3.1
	Total	51.6	324	30.6	1.7	59.6	324	32.4	1.8	55.6	324	29.4	1.6

Samenvatting

Engels als lingua franca: Onderlinge verstaanbaarheid van Chinese, Nederlandse en Amerikaanse sprekers van het Engels

In de afgelopen eeuw heeft het Engels zich ontwikkeld tot de lingua franca van de wereld. Het is nu de taal van de internationale zakenwereld, handel, politiek en wetenschap. Deze ontwikkeling heeft geleid tot het ontstaan van een grote verscheidenheid aan 'non-native Englishes', d.w.z. variëteiten van het Engels die worden gesproken door personen wier moedertaal geen Engels is. Zulke variëteiten worden enigszins laatdunkend wel aangeduid als Chinglish (Chinees Engels), Dunglish (Nederlands Engels, Dutch-accented English), Spanglish (Spaans Engels), enzovoort. In deze non-native variëteiten wordt het Engels uitgesproken met een onmiskenbaar buitenlands (d.w.z. niet-Engels) accent. Aan de hand van zo'n accent kunnen luisteraars in het algemeen gemakkelijk vaststellen wat de moedertaal-achtergrond van de spreker is. Erger is dat het accent de verstaanbaarheid van de spreker kan aantasten. Ook heeft een non-native luisteraar meer moeite dan de moedertaalluisteraar met het verstaan en begrijpen van het Engels, als gevolg van onvolledige kennis van het Engelse klanksysteem, van de woordenschat en van de woord- en zinsgrammatica. Er is inmiddels veel onderzoek gedaan naar de productie en waarneming van het Engels door non-native leerders. Heel weinig is er nog maar bekend over de specifieke problemen die zich voordoen wanneer non-native sprekers met elkaar in het Engels moeten communiceren die ieder een andere moedertaal hebben. Die situatie doet zich b.v. voor als een Nederlandse piloot zich in het Engels moet verstaan met een Spaanse verkeersleider. In mijn onderzoek richt ik me op de problemen die ontstaan wanneer Chinese en Nederlandse sprekers met elkaar moeten communiceren in het Engels.

Mijn doel is de onderlinge verstaanbaarheid te bepalen van Chinese, Nederlandse en Amerikaanse sprekers in het Engels. Het Nederlands en het Engels zijn verwante West-Germaanse talen, die een groot deel van hun woordenschat delen, en waarvan de klanksystemen niet al te zeer verschillen. De structuur van het Standaard Chinees (of: Mandarijn), een Sino-Tibetaanse taal, verschilt sterk van die van het Nederlands of het Engels, en nagenoeg iedere overeenkomst in de woordenschat berust op toeval. In eerste benadering testen we de hypothese dat Chinese sprekers van het Engels moeilijker te verstaan zijn door Nederlandse (en Amerikaanse) luisteraars dan Nederlandse (en Amerikaanse) sprekers zijn voor Chinese luisteraars. In tweede instantie vragen we of non-native Engels gemakkelijker te verstaan is wanneer de spreker en de luisteraar dezelfde moedertaal hebben. Verstaan Chinese luisteraars Engels met een Chinees accent beter dan Engels met een Nederlands, en zelfs een Amerikaans accent? In dezelfde lijn

voortredenerend, hebben Nederlandse luisteraars er minder moeite mee het Engels van een landgenoot te verstaan dan dat van een Chinees of een Amerikaan? Dit onderzoek naar wat wel het voordeel van de tussentaal (interlanguage benefit) is genoemd, is nog maar kort geleden op gang gekomen. Mijn onderzoek is waarschijnlijk de eerste poging tot een grootschaliger bestudering van dit verschijnsel. Mijn meer specifieke onderzoeksvragen vermeld ik aan het einde van deze samenvatting, in samenhang met de conclusies die ik trek uit mijn experimenten, aan de hand waarvan ik de vragen beantwoord.

Verstaanbaarheid testen we door na te gaan hoe goed luisteraars in staat zijn de woorden die een spreker uit, te herkennen, in dezelfde volgorde waarin de spreker ze gezegd heeft. Verstaan is een voorwaarde om te komen tot begrip van het gesprokene, maar verschilt daarvan omdat verstaan niet expliciet een betekeniscomponent kent. In mijn onderzoek stel ik de verstaanbaarheid van woorden vast in betekenisloze en betekenisdragende zinnen. Om te kunnen begrijpen waarom de woordherkenning in non-native communicatie een probleem vormt, test ik ook het vermogen bij Chinese, Nederlandse en Amerikaanse luisteraars om individuele klinkers, medeklinkers en medeklinkerverbindingen (clusters) te identificeren in het Engels van Chinezen, Nederlanders en Amerikanen, in alle negen mogelijke combinaties van spreker- en hoordernationaliteit (of nog liever: moedertaalachtergrond). Amerikaanse in plaats van Britse sprekers van het Engels fungeerden als controlegroep omdat de uitspraaknorm voor het Engels in het Chinese onderwijssysteem de Amerikaanse is, terwijl Nederlands Engels min of meer het midden houdt tussen de Britse en de Amerikaanse uitspraak.

Na mijn inleidend hoofdstuk, waarin ik deze onderzoeksvragen formuleer, presenteer ik in Hoofdstuk twee de relevante literatuur met betrekking tot het testen van verstaanbaarheid, over vreemde-taalverwerving en over de effecten van niet-moedertaligheid (non-nativeness) op de productie en perceptie van spraak. Hoofdstuk drie bevat een gedetailleerde contrastieve analyse van he klanksysteem van het (Mandarijn) Chinees tegen over dat van het Engels, en een soortgelijke vergelijking van het Nederlandse met het Engelse klanksysteem. Potentiële moeilijkheden in de productie en perceptie van Engelse klanken door Chinese en Nederlandse leerders van het Engels worden geïdentificeerd en gedocumenteerd aan de hand van ervaringen die opgetekend zijn in de pedagogische literatuur.

In Hoofdstuk vier beschrijf ik de procedure die ik gevolgd heb om de materialen te verkrijgen die nodig waren voor het experimentele deel van mijn onderzoek. Ik heb me ingespannen om optimaal vergelijkbare sprekers van het Engels te vinden met een Chinees en met een Nederlands accent, één manlijke en één vrouwelijke spreker per groep. Deze optimale sprekers werden geselecteerd uit grotere groepen van tien manlijke en tien vrouwelijke sprekers in elk land, zodanig dat de optimale sprekers precies in het midden van hun ‘peer’-groep zaten. In beide landen heb ik mijn sprekers gezocht in de populatie van jonge academische gebruikers van het Engels die zich niet hadden gespecialiseerd in het Engels, en die nooit hadden gewoond in Engelssprekende omgevingen.

In Hoofdstuk vijf presenteer ik, bij wijze van intermezzo, een gedetailleerde akoestische analyse van de klinkers die zijn geproduceerd door de drie groepen van

20 sprekers (tien mannen, tien vrouwen per taalachtergrond). De resultaten laten zien dat Chinese en Nederlandse sprekers hun Engelse klinkers minder goed van elkaar onderscheiden maken dan de Amerikaanse moedertaalsprekers. Ondanks hun buitenlands accent kunnen Engelse klinkers van de Chinese en de Nederlandse sprekers heel succesvol geïdentificeerd worden met behulp van een Lineaire Discriminant Analyse (LDA). Deze automatische classificatieprocedure laat zien dat er behoorlijk wat akoestisch detail verscholen zit in de klinkers met buitenlands accent, dat goed gebruikt kan worden om de klinkers te identificeren maar dat de menselijke luisteraar ontgaat.

Hoofdstukken zes, zeven en acht presenteren respectievelijk de resultaten van de herkenning van klinkers, medeklinkers en clusters door 36 Chinese, 36 Nederlandse en 36 Amerikaanse luisteraars. De klinkers werden aangeboden in /hVd/-contexten en moesten worden geïdentificeerd met gedwongen keuze uit de 20 klinkers van het Engels. Alle beginmedeklinkers (24) werden aangeboden in invervocale positie (tussen twee klinkers in) /ɑ:Cɑ:/ context, hetgeen eveneens het geval was voor een selectie van 20 twee- (CC) en drieledige (CCC) clusters. De resultaten worden allereerst gepresenteerd in termen van percentage correct geïdentificeerde doelklanken. In het tweede deel van elk hoofdstuk presenter ik dan een foutenanalyse aan de hand van de verwarringsstructuur, onder verwijzing naar verwarringsmatrices (in de appendix) en naar verwarringsdiagrammen (in de tekst zelf), waarmee ik de belangrijkste klinker- en medeklinkerverwarringen in elk van de negen mogelijke combinaties van spreker- en luisteraar achtergrond kan belichten. Een verwarringsanalyse van de medeklinkerclusters kon achterwege blijven omdat die structuren tamelijk gemakkelijk te identificeren waren bij alle spreker-luisteraar-combinaties, waardoor er te weinig fouten waren om tot een zinvolle analyse te komen. *Grosso modo* wijzen de resultaten uit dat succes bij de communicatie van klinkers, medeklinkers en clusters van spreker naar luisteraar primair bepaald wordt door de moedertaalachtergrond van de luisteraar, en minder door die van de spreker. De Amerikaanse proefpersonen waren in het algemeen succesvoller als sprekers en als luisteraars van het Engels dan de Nederlanders, die op hun beurt weer beter waren dan de Chinese sprekers en luisteraars. Ondanks deze globale effecten vind ik echter een systematische interactie tussen de taalachtergrond van spreker en luisteraar, die wijst op een duidelijk voordeel op grond van gemeenschappelijke tussentaal (interlanguage benefit).

Hoofdstuk negen test de woordherkenning, eerst in zgn. Semantically Unpredictable Sentences (SUS), and daarna in een selectie van zinnen uit de Speech-in-Noise (SPIN) test. In SUS-zinnen, zijn woorden geplaatst in zes verschillende grammatische schema's zonder dat het geheel ooit een betekenisvolle zin oplevert, b.v. *The state sang by the long week* of *Why does the range watch the fine rest?* In zulke zinnen hebben de latere woorden geen voordeel bij correcte herkenning van woorden eerder in dezelfde zin. Luisteraars moesten in deze zinnen alle inhoudswoorden invullen, terwijl de functiewoorden al afgedrukt waren op de antwoordformulieren. In de SPIN-zinnen moesten de luisteraars alleen het laatste woord van elke zin opschrijven. Dit was in de helft van de gevallen onvoorspelbaar uit de eerdere zincontext (zoals in *We should consider the* ***map***) en in de andere helft

juist in hoge mater voorspelbaar (zoals in *Keep your broken arm in the **sling***). De resultaten laten zien dat de effecten van de moedertaalachtergrond van de sprekers en de luisteraars in het algemeen sterker naar voren komen bij deze woordherkenningstaken dan bij de eerdere klankidentificatietaken. Maar andermaal oefende de moedertaalachtergrond van de luisteraar een sterker effect uit dan die van de spreker, en opnieuw vinden we een sterk effect van interlanguage benefit.

In Hoofdstuk tien vat ik de belangrijkste uitkomsten van deze studie samen en probeer tevens systematisch antwoorden te geven op de onderzoeksvragen die ik in mijn inleidend hoofdstuk aan de orde heb gesteld. Deze vragen en hun antwoorden staan hieronder, in verkorte vorm.

1. Zijn sprekers/luisteraars met een moedertaal die verwant is aan de doeltaal in het voordeel ten opzichte van leerders met een moedertaal die verder af staat van de doeltaal? Mijn resultaten laten inderdaad zijn dat Nederlandse leerders meer succes hebben als sprekers en luisteraars in het Engels dan hun Chinese tegenhangers, zelfs als de groepen leerders geselecteerd zijn uit vergelijkbare groepen jonge academische gebruikers van het Engels als vreemde taal.
2. In hoeverre bevatten de verschillende tests op de lagere niveaus (klinkers, medeklinkers, clusters) en die op het hogere niveau (woordherkenning in nonsense-zinnen en laag-/hoog-voorspelbaar aan het eind van betekenisvolle zinnen) onafhankelijke informatie over onderlinge verstaanbaarheid van de sprekergroepen? Het blijkt dat klinker-, medeklinker- en clusteridentificatiescores slechts matig geïntercorreleerd zijn, zo dat elke lagere deelvaardigheid intercorrelated tamelijk onhankelijke informatie kan bijdragen aan de voorspelling van het succes bij de hogere-orde woordherkenningsvaardigheid. Voor de Chinese luisteraars zijn de intercorrelaties lager (r-waarden tussen .25 en .60) dan voor de Nederlandse of de Amerikaanse luisteraars (r-waarden tussen .51 en .72).
3. Kunnen we woordherkenning voorspellen uit de ate van success bij de identificatie van klinkers, medeklinkers en clusters op het lagere niveau? Meer in het algemeen, wat zijn de correlaties tussen de verschillende testtypen? In het algemeen kunnen we de resultaten van de hogere-orde vaardigheden niet erg accuraat voorspellen op grond van de lagere-orde foneem- en clusteridentificatietests. Multipele R komt nooit boven de .70, zodat maximaal 49 procent van de variantie in de woordherkenningsscores verklaard wordt door de lagere-orde vaardigheden. Interessant genoeg kunnen we woordherkenning beter voorspellen uit de foneemidentificatiescores als de luisteraars Amerikaans of Nederlands zijn (R-waarden tussen .25 en .70) dan wanneer zij Chinees zijn (R-waarden tussen – .27 en +.25).
4. Welke tests zijn het meest succesvol als we de geode van de slechte luisteraars willen scheiden? Over het geheel genomen discrimineren de hogere-orde vaardigheden (woordherkenning) beter tussen de drie luisteraargroepen dam de lagere-orde (foneemidentificatie) vaardigheden. De beste separatie tussen de drie groepen (Chinese, Nederlandse, Amerikaanse luisteraars) vinden we bij de hoog-voorspelbare SPIN-zinnen, waarin het zinsfinale woord met meer kans op success herkend wordt als de hoorder ook de eerdere woorden in de zin correct

herkend heeft. in Merk op dat dit type test het dichtst verstaanvaardigheidstaken in het echte leven benadert.

5. Kunnen we klinker- en medeklinker fouten/verwarringen voorspellen uit een contrastieve analyse van de klanksystemen in de bron- en in de doeltaal? Triviaal is dat klanken die in de leerder's moedertaal (brontaal) en in het Engels (doeltaal) (nagenoeg) hetzelfde zijn, met meer succes worden overgedragen van spreker naar hoorder dan klanken die tussen bron- en doeltaal verschillen. Dit is ook de voorspelling van Lado's klassieke transfermodel. Maar binnen de klasse van klanken die in bron- en doeltaal van elkaar verschillen, falen verdere voorspellingen. Zogenoemde nieuwe klanken (new sounds), d.w.z. doelklanken die aanzienlijk verschillen van enige klank in de brontaal, worden niet beter overgedragen dan zogenoemde gelijkende klanken (similar sounds), welke op meer subtiele wijze verschillen tussen bron- en doeltaal. Hier falen de voorspellingen van het recentere Speech Learning Model.
6. Kunnen we de waarneming en de verwarringsstructuur van klinkers voorspellen uit een akoestische analyse? Levert een LDA met F1, F2 (akoestische correlaten van respectievelijk kaak- en tongstand) en klinkerduur hetzelfde type fouten op als wat we vinden in menselijke herkenning? Onze resultaten laten zien dat de menselijke waarneming van klinkers met uiteenlopende mate van succes – maar altijd (veel) beter dan op grond van toeval verwacht mag worden – voorspeld kan worden uit de akoestische eigenschappen van de klinkers zoals die worden geprodiceerd door moedertaal- en vreemde-taalsprekers, op basis van Lineaire Discriminant Analyse. We kunnen de LDA-techniek ook goed gebruiken om (althans ten dele) de verwarringsstructuur bij (Engelse) klinkers te voorspellen in non-native communicatie waarin de spreker of hoorder (of beiden) een andere (eventueel ook onderling verschillende) moedertaal heeft dan het Engels.
7. Welke factoren dragen het meest bij tot onderlinge verstaanbaarheid? Is de kwaliteit van de spreker meer of minder van belang voor de effectiviteit van het communicatieproces dan de kwaliteit van de luisteraar? Mijn resultaten wijzen eenduidig uit dat het effect van de moedertaalachtergrond meer gewicht in de schaal legt dan dat van de spreker. Het doorslaggevend belang van de luisteraar komt naar voren elk elke van de zes tests in de testbatterij.
8. Is de moedertaalluisteraar altijd de beste taalgebruiker? Het blijkt dat, in termen van absolute scores, de Amerikaanse moedertaalsprekers in de regel, maar niet altijd, de beste resultaten behalen. In drie tests, waren de Nederlandse luisteraars meer succesvol dan de Amerikaanse controleluisteraars maar alleen als de sprekers ook Nederlands waren. Dit is dan een goed voorbeeld van wat we de absolute interlanguage benefit zouden kunnen noemen.
9. Steunen onze resultaten de hypothese dat er zoiets bestaat als moedertaal-/tussentaalvoordeel (native/interlanguage benefit)? Hoewel interlanguage benefit is aangetroffen in de testresultaten zelfs als we alleen de absolute scores als criterium nemen (zie punt 8 hierboven), betoog ik dat het verschijnsel van het tussentaalvoordeel inzichtelijker bestudeerd kan worden in relatieve termen. Daartoe dienen we eerst een verwachte verstaanvaardigheidsscore berekenen op basis van de gemiddelde prestaties van de luisteraar- en sprekergroep die in de

vergelijking zijn betrokken. Ten opzichte van deze verwachte score behalen combinaties van dezelfde spreker- en luisteraarnationaliteit hogere scores in 16 van de 18 testsituaties. De algehele conclusie is dan ook dat het tussentaal-voordeel zich bijna altijd doet gelden.

中文摘要

以英语作为通用语：测试母语分别为汉语,荷兰语和美国英语者之间的可懂度

自从上个世纪英语成为国际政治，经济贸易和科学研究的通用语言以来，它给国际间交流带来便利；但同时，由于母语背景的差异而产生的英语发音多样性与复杂性， 也而给国际间沟通带来困难。明显的带有各种口音的英语， 如中式英语，法式英语，荷式英语，西式英语等等，不仅能让人们很容易认出发音人的母语，同时，也降低了发音人的英语的可知性。由于对英语音系结构，词汇结构，以及语法结构知识的缺乏，听音人对这种带有口音的英语的感知程度会进一步降低。因而，有大量研究着眼于的非英语母语背景下英语学习者的发生与感知，但到目前为止，对于学习者在不同的母语背景下，用英语交流时的所遇到的具体问题与困难的研究还不多见。在我们的研究中，我们致力于荷兰人与中国人用英语交流时所遇到的困难和问题

具体地说，我们旨在发现中国人，荷兰人，美国人在用英语交流时的相互的沟通程度。英语与荷兰语作为与西日尔曼语系相联系两种语言，共享很多的词汇，语音结构的差别也不很大， 而汉语作为汉藏语系的一支，与英语和荷兰语有着完全不同的语言结构，而且没有任何共享的词汇。作为研究的一个假设，我们预测对于荷兰人和美国人来讲，中国的英语发音人要比荷兰的英语发音人更难于理解；作为研究的另一个假设，我们预测当发音人和听音人共享同一语言背景时，他们的英语会更容易理解一些；中国人理解中式英语是否要好于中国人理解荷式英语或者美式英语？同样，荷兰人是否在听带有自己口音的英语时就比听中式英语或者美式英语时少了许多的困难？这是最近才引起关注的所谓的 “过渡语具有优越性”的语言现象。本研究试图对此现象进行全方位的调查研究。下面我将对本研究着重的问题进行陈述，同时简要列出我们研究的结果与相关的结论。

可知性是由听音人对于发音者的识别率决定的，是听力理解的先决条件。它与听力理解的区别在于它并不涉及语音的意义。本研究试图建立词在有意句和无意句两种情况下的可知性。 为了理解为什么对于词的辨认是交流中的困难，我测试了中国人，荷兰人，和美国人对于英语元音，辅音和辅音群的辨别。这些音由中国人，荷兰人与美国人分别发生，以听音人和发音人组成九组以后分组辨别的。

第一章的简介中提出问题之后，在第二章中讨论了关于可知性测试，外语语言习得和非母语的学习者的发生与感知文献资料。第三章中包含了细致的汉语，荷兰语与英语的语音结构对比，根据教学文献资料对三种语音结构的分析，中国人，荷兰人的英语发生与感知的潜在问题得到了预测。

在第四章中，所有实验材料和试图找到理想的具有代表性的和可比性的荷兰口音和中国口音的男女发音人的过程得到描述。这些理想的发音人是各自国家大学里从非英语专业的大学生中选出的 20 个（男女各 10 名）发音人里恰好处于中间水平代表者。

在第五章对具体的 20 个发音人的元音发生细节进行分析，从而展示了和美国发音人相对比下的中国和荷兰发音人英语元音发生的不清晰与不准确之处。中国发音人与荷兰发音人依然能够被比较成功地线性分析辨别。这种自动的计算机辨认结果揭示了潜在的带有口音的元音发生细节。

在第六章， 第七章和第八章中，36 中国听音人，36 荷兰听音人和 36 美国听音人的对于元音，辅音，和辅音群的辨认结果得到展示。这些音是 20 个 /hVd/ 元音结构，24 个/ɑ:Cɑ:/ 辅音结构和 21 个/ɑ:CCɑ:/ 辅音群结构组成的。辨认结果首先是以辨认的正确率表现出来的，然后展示了具体的混淆结构图。这些章节中的主要内容是关于九种组合中听音人和发音人混淆分析。对于辅音群的辨认结果比较清洗，错误相对很少， 所以对于辅音群并没有过多分析。 关于元音，辅音和辅音群的总的结果表明交流的成功率取决于听音人胜过发音人。美国发音人和听音人总的来讲比好于荷兰人，而荷兰人则好于中国人。除了这些总的结果，我们发现了听音人与发音人的语言背景之间的相互间系统的反应，揭示了清晰的中介语具有优越性的效果。

在第九章词的辨认中，首先是所谓的“语义不可预知句”（SUS）的测试，其次是从 Speech-in-Noise (SPIN) 的测试中选择的句子的测试。在 SUS 句子中，词是在无意义的句子中的，以 6 种句子结构出现的。如：*The state sang by the long week* 或者 *Why does the range watch the fine rest?* 在这样的句子中，对于后面的词听音人不能从前面的词中得到暗示。听者按照顺序写出听到的实词。在 SPIN 句子测试中，听音人写下最后一个词，这个词可以/不可以从前面的句子中得知。如：不可预知的 *We should consider the **map*** 和可预知的 *Keep your broken arm in the **sling***. 这一部分的测试结果表明词的测试比以前的单音测试更能表现出听音人和发音人的语言背景。 同时，同一语言背景下潜在的过渡语优越性又一次表现出来。

第十章是关于所以结果的系统陈述与相关分析以更细致地回答据第一章提出的问题：

1. 是否源语与目的语关系近的听音人和发音人比源语与目的语关系远的听音人和发音人具有优越性？回答是肯定的。实验结果表明荷兰听音人和发音人都比中国的听音人和发音人成功率高。
2. 低层次的单音测试和较高层次的词的测试，哪一种测试对语音可知性的测试更可靠？实验结果表明对于元音，辅音和辅音群的测试只是适度相关，所以说，每一个底层次的测试都独立地为高一层次的测试提供信息，中国听音人的相关系数（r 值在 .25 和 .60 之间）就小于荷兰人和美国人（r 值在 .51 和 .72 之间）。
3. 能否从低的语言层次元音，辅音，辅音群的辨认中预测对于较高的语言层次词的辨认？不同种类测试中的结果是否相关？怎样相关？一般来讲，较高层次的词的测试结果没有能够在较低层次的测试中得到预测。 多项 R 值从来没超过 .70, 所以在词的辨认中最多有 49% 差异来自于较低层次的辨认结果。有趣的是，当听音人是荷兰人或者美国人时，词的辨认可以更好的从较低层次的测试中预测出来， R 值 在 .25 和 .70 之间， 而听音人是中国人时，R 值在 – .27 和 +.25 之间。
4. 那一种测试更成功地区别了较好的听音人与较差的听音人？一般来讲，要求较高技巧的词的辨认比要求较低的语音的辨认更好地区别了较好与较差的听音人。最好地区别了三组听音人的测试是高预知性的 SPIN 句子测试,

在这种句子中，句子的最后一个词可以成功地被听音人预测出来。有趣的是，这种测试的句子也是最真实的接近生活可知性测试句子。

5. 元音与辅音的辨认错误是否可以从源语与目的语的语音对比分析中预测出来？源语与目的语之间相似的语音被比较成功地辨认出来，而差别较大的语音则相对困难些。这是符合 Lado 的迁移理论模式的。但有一些语音的预测却出乎预料。所谓的新音(目的语中与源语中差别很大的音)并没有被转化为那些在源语与目的语之间有微妙差别的所谓的相似音，就这一点而言，最近提出的“言语学习模式”没有得到验证。
6. 元音的感知和混淆结构能否从发生的声学分析中预测出来？LDA 关于第一和第二共振峰的以及音长的分析而预测的错误是否能在真人的感知测试中得到证实？我们的结果表明在跨语言的真人元音感知测试是可以通过 LDA 进行预测的。这种技术也许也可以应用在非英语母语的听音和发音人用英语交流时，或者来自不同的语言背景的听音人和发音人用英语交流时的元音的混淆结构的预测。
7. 哪一因素更能为可知性测试提供信息？是发音人的水平还是听音人的水平对于有效的交流更重要？实验结果毫无疑问地表明听音人的语言背景比发音人的语言背景更强有力地表现了测试的结果。在六个测试中， 听音人在每一个测试中都表现了一贯的重要性。
8. 母语的听音人总是最好的吗？从总的结果来讲， 回答是肯定的，但并非总是如此。在三个测试中，荷兰听音人比美国听音人结果更佳， 但都是在发音人是本族荷兰人时发生的。这 一结果又一次证明了过渡语的优越性。
9. 我们的结果支持母语/过渡语具有优越性的假设吗？ 虽然过渡语具有优越性的现象在结果得到证实，（见 8）但这一现象应该得到更深刻的分析。 我们根据结果平均计算听音人组和发音人组的可知性的期待值，根据期待值，结合同一组的听音人和发音人的分值，从 18 种测试情形中得出 16 个高于平均值的结果。所以我们说母语/过渡语具有优越性现象极其广泛。

Summary

English as a lingua franca: Mutual intelligibility of Chinese, Dutch and American speakers of English

In the last century, English has developed into the lingua franca of the world. It is now the language of international business, trade, commerce, politics and science. This development has led to a large variety of non-native Englishes, i.e. varieties of English spoken by learners whose native language differs from English. Such varieties are sometimes disparagingly referred to as, for instance, Chinglish (Chinese-accented English), Dunglish (Dutch-accented English), Spanglish (Spanish-accented English), and so on. In these non-native varieties, English is spoken with a distinct foreign accent. Such accents not only allow listeners to identify the non-native speaker's mother tongue, they may also reduce the non-native speaker's intelligibility. Also, a non-native listener's perception of English may be less effective, due to imperfect knowledge of the English sound system, lexicon and morpho-syntax. There is a large body of research on the production and perception of English by non-native learners. Very little, however, is known at this time about the specific problems that arise when non-native speakers communicate in English, if these speakers do not share the same native language. Such situations are found, for instance, when a Dutch airline pilot has to communicate in English with the control tower at an airport in Spain. In our research, we address the problems that come up when Chinese and Dutch speakers communicate with each other in English.

Specifically, I aim to determine the mutual intelligibility of Chinese, Dutch and American speakers in English. Dutch and English are related West-Germanic languages, which share a large part of their vocabularies and whose sound systems do not differ greatly. Standard Chinese (Mandarin), being a Sino-Tibetan language, has a structure that is very different from either Dutch or English, and shares none of the vocabulary. As a first approximation, we test the hypothesis that Chinese speakers of English are more difficult to understand by Dutch (and American) listeners than Dutch (and American) speakers are for Chinese listeners. Secondly, we ask whether non-native English is easier to understand when the speaker and the listener have the same native language. Do Chinese listeners understand Chinese-accented English better that either Dutch-accented English or even American native English? Similarly, do Dutch listeners have less difficulty in understanding a fellow Dutch speaker of English than when listening to a Chinese (or American) speaker of English? This so-called inter-language benefit has only recently begun to receive attention. My study is probably the first to attempt a full-scale investigation of this phenomenon. An itemized list of specific research

questions is included at the end of this summary, together with the conclusions that can be drawn from the experiments, which serve as the answers to the questions.

Intelligibility is tested by determining how well listeners recognize the words a speaker utters, in the order intended by the speaker. Intelligibility is a prerequisite for comprehension (or speech understanding) but differs from the latter in that it does not explicitly involve meaning. In my research I establish the intelligibility of words in meaningless and in meaningful sentences. In order to understand why word recognition is problematic in non-native communication, I also test the ability of Chinese, Dutch and American listeners to identify individual vowels, consonants and consonant clusters in English spoken by Chinese, Dutch and American speakers of English, in all nine possible combinations of speaker and hearer nationalities (or rather: native language backgrounds). American, rather than British, speakers of English were used as controls as the norm of English teaching for my Chinese speakers is American, and Dutch-accented English does not seem to differ more from the American than from the British pronunciation of English.

After my introductory chapter, in which I formulate these research questions, Chapter two presents relevant literature on the topics of intelligibility testing, foreign-language acquisition and the effect of non-nativeness on the production and perception of a language. Chapter three contains a detailed contrastive analysis of the sound systems of Chinese (Mandarin) versus English and of Dutch versus English. Potential problems in the production and perception of English sounds by Chinese and by Dutch learners of English are identified in the analysis, and supported by claims made in the pedagogical literature.

In Chapter four I describe the procedures followed to obtain the materials needed for the experimental part of the research. I attempted to find optimally comparable speakers of Chinese-accented and of Dutch-accented English, one male and one female speaker for each group. These optimal speakers were selected from larger groups of ten male and ten female speakers in each country, such that the optimal speakers were right in the middle of their peer group. In both countries, the speakers targeted were young academic users of English, who had not specialized in English and had never lived in English-speaking environments.

Chapter five, as in intermezzo, presents a detailed acoustical analysis of the vowels produced by the three groups of 20 speakers (ten males, ten females per language background). The results show that Chinese and Dutch speakers keep the English vowels less distinct than the American native speakers do. Nevertheless, the Chinese and Dutch-accented vowels can be identified quite successfully by Linear Discriminant Analysis (LDA). This automatic classification procedure revealed that there is substantial acoustic detail in the foreign-accented vowel tokens that may serve to identify the vowel tokens but is not used by human listeners.

Chapters six, seven and eight present the results of the vowel, consonant and consonant cluster identification tests, respectively, by 36 Chinese, 36 Dutch and 36 American listeners. Vowels were presented in /hVd/ contexts and had to be identified with forced choice from the 20 vowels of English. All onset consonants (24) were presented intervocalically in /ɑ:Cɑ:/ contexts, as was a selection of 21 CC and CCC clusters. Results are first presented in terms of percent correctly identified

targets. In the second part of each chapter an error analysis is presented in terms of confusion structure, using confusion matrices (in appendices) and confusion graphs (in body of text) highlighting the most important vowel and consonant confusions for each of nine possible combinations of speaker and hearer backgrounds. No confusion analysis is given of the consonant clusters as these structures proved relatively easy to identify for all speaker-listener combinations, so that there were not enough errors to make a confusion analysis worthwhile. The overall results show that success in communicating vowels, consonants and clusters depends primarily on the language background of the listener rather than that of the speaker. American speakers/ listeners are generally more successful as speakers and as listeners than are the Dutch subjects, who in turn are more successful than the Chinese speakers and listeners. In spite of these overall effects, however, I find a systematic interaction between speaker and listener language background, revealing a clear effect of the inter-language benefit.

Chapter nine tests word recognition, first in so-called Semantically Unpredictable Sentences (SUS), and second in a selection of sentences taken from the Speech-in-Noise (SPIN) test. In SUS sentences, words appear in six grammatical frames but do not make up a meaningful sentence, e.g., *The state sang by the long week* or *Why does the range watch the fine rest?* In such sentences, later words do not benefit from correct recognition of earlier words. Listeners wrote down all the content words in these sentences, while function words were pre-given on the answer sheets. In the SPIN materials, the listeners wrote down the final word in each sentence, which was either unpredictable from the earlier words in the sentence (as in *We should consider the **map***) or highly predictable (as in *Keep your broken arm in the **sling***). The results show that effects of speaker and listener language background are generally stronger in these word-recognition tasks than in the earlier sound identification tests. But again, the native-language background of the listener exerted a stronger effect than that of the speaker, and again substantial interlanguage benefit could be shown.

Chapter ten presents a summary of findings and then systematically tries to answer the research questions that were identified in the introductory chapter. These questions and answers are summarized below.

1. Is it true that speaker/hearers with an L1 that is close to the target language have an edge over learners with a more distantly related L1? My results show that, indeed, Dutch learners are more successful as both listeners and speakers of English than Chinese learners, even with both groups are selected from young academic users of English as a foreign language.
2. To what extent do separate tests at the lower levels (vowels, consonants, clusters) and at the higher levels (word recognition in nonsense sentences, and in low/high predictability meaningful sentences) contribute independent information to the measurement of mutual intelligibility? It turns out that vowel, consonant, and cluster identification scores are only moderately intercorrelated so that each subskill may contribute independent information to the higher-order word-recognition skill. For Chinese listeners the intercorrelations

are smaller (r-values between .25 and .60) than for either Dutch or American listeners (r-values between .51 and .72).

3. Can word recognition be predicted from success in identification of vowels, consonants and clusters at the lower level? What, more generally, is the correlation between the various types of test results? Generally, the results on the higher-order word recognition tests cannot be predicted with great accuracy from the lower-order phoneme and cluster identification tests. Multiple R is never better than .70, so that maximally 49 percent of the variance in the word recognition scores is accounted for by the lower-order skills. Interestingly, word-recognition can be predicted better from phoneme identification scores when the listeners are either American or Dutch (R-values between .25 and .70) than when they are Chinese (R-values between – .27 and +.25).
4. Which tests are most successful in discriminating the better from the poorer listeners? Generally, higher-order skills (word recognition) discriminate better between the three listener groups than lower-order (phoneme identification) skills. The best separation of the three groups (Chinese, Dutch, American listeners) is obtained for the high-predictability SPIN sentences, in which the sentence-final word can be recognized more successfully if the listener has also recognized the earlier words in the sentence. Interestingly, this type of test is also closest to real-life intelligibility tasks.
5. Can vowel and consonant errors/confusions be predicted from a contrastive analysis of the sound systems of source and target language? Trivially, sounds that are (almost) the same in the learner's native language (source language) and in English (target language), were transmitted more successfully between speaker and listener than sounds that differ between source and target language. This is as predicted by Lado's classical transfer model. However, within the class of sounds that differ between source and target language further predictions fail. So-called new sounds (target sounds that differ substantially from any sounds in the source language) are not transmitted any better than so-called similar sounds, which differ more subtly between source and target language. Here, the predictions made by the more recent Speech Learning Model fail.
6. Can vowel perception and confusion structure be predicted from an acoustical analysis? Does an LDA on F1, F2 and duration measurements yield the same types of errors as in human perception? Our results indicate that cross-linguistic human perception of vowels can be predicted, with varying success but invariably (much) better than chance, from the acoustic properties of the vowel tokens as produced by native speakers and foreign learners, using Linear Discriminant Analysis. The technique may also be used to predict (part of) the confusion structure of (English) vowels in non-native communication with either or both speaker and hearer having a different language than English and even different native languages.
7. Which factors contribute most to mutual intelligibility? Is the quality of the speaker more or less important to the effectivity of the communication process than the quality of the listener? My results show unequivocally that the effect of listener nationality (or native-language background) is stronger than the effect

of speaker nationality. The overriding importance of the listener effect is found in each of the six tests administered in the test battery.

8. Is the native listener always the best performer? It turns out that, in terms of absolute scores, the American native listeners generally, but not always, obtain the best results. In three tests, Dutch listeners were more successful than the American control listeners but only if the speakers were also Dutch. This, then, is an example of what we may call absolute interlanguage benefit.
9. Do our results support the hypothesis that native/interlanguage benefit exists? Although interlanguage benefit was found in the test results even when absolute scores were used as the criterion (see 8 above), I argue that the phenomenon of interlanguage benefit is more insightfully studied in relative terms. We should first compute an expected intelligibility score based on the mean performance of the listener group and of the speaker group. Relative to this expected score, combinations of same speaker and listener nationality yield higher scores in 16 out of 18 test situations. The overall conclusion, then, is that the interlanguage benefit is pervasive.

Curriculum vitae

Wang Hongyan was born on April 5, 1967 in Tongliao, Inner Mongolia, in the People's Republic of China. She left Middle School in 1986 and went on to study English at the National Teachers' College of Nei Menggu, where she got her Bachelor's certificate in 1989. After having taught English for a number of years, she then enrolled as a graduate at Jilin University. She obtained her Master's degree in English Linguistics and Literature from the College of Foreign Languages of Jilin University in 1997. From then on she was employed as a lecturer in English at Jilin University. In the period between 2002 and 2006 she worked at the Leiden Centre for Linguistics (LUCL), first as a guest researcher with a grant from the China Scholarship Council (2002/03), then with a Delta scholarship from the Leiden University Fund (2003/04), and the remaining two years with an LUCL scholarship. The present dissertation is the report of the work done in this four-year period. Wang Hongyan is currently employed as a lecturer in the English Department of Shenzhen University in Guandong Province, China, where she lives with her daughter Ziru.