Downloaded from UvA-DARE, the institutional repository of the University of Amsterdam (UvA)
http://hdl.handle.net/11245/2.52668

File ID uvapub:52668
Filename Koet.pdf
Version unknown

SOURCE (OR PART OF THE FOLLOWING SOURCE):
Type PhD thesis
Title Polder English in Dutch ears : empirical studies on the evaluation of the pronunciation of English as a foreign language
Author(s) A.G.M. Koet
Faculty FMG: Research Institute Child Development and Education (CDE)
Year 2007

FULL BIBLIOGRAPHIC DETAILS:
http://hdl.handle.net/11245/1.283932

Copyright

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content licence (like Creative Commons).

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
(pagedate: 2014-11-17)

POLDER ENGLISH IN DUTCH EARS

EMPIRICAL STUDIES ON THE EVALUATION OF THE PRONUNCIATION OF ENGLISH AS A FOREIGN LANGUAGE

UNIVERSITEIT VAN AMSTERDAM

Cover design: Monique van Hootegem
Cover photograph: Thomas Koet
Printed by PrintPartners Ipskamp B.V., Enschede

ISBN 978-90-78087-11-3

This publication was partly supported by grants from the HBO-raad, the Educatieve Hogeschool van Amsterdam and the Graduate School of Teaching and Learning, University of Amsterdam

POLDER ENGLISH IN DUTCH EARS

Empirical studies on the evaluation of the pronunciation of English as a foreign language

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr J.W. Zwemmer
ten overstaan van een door het college van promoties ingestelde
commissie, in het openbaar te verdedigen in de Aula der Universiteit

op dinsdag 18 september 2007, te 14.00 uur

door

Antonius Gerardus Maria Koet

geboren te Amsterdam

Promotiecommissie

Promotoren:	Prof. dr G.C.W. Rijlaarsdam Prof. dr H.H. van den Bergh
Overige leden:	Prof. dr A.J. van Essen Prof. dr W. Herrlitz Prof. dr B.H.A.M. van Hout-Wolters Prof. dr F. Kuiken Prof. dr ir L.C.W. Pols
Faculteit	Maatschappij- en Gedragswetenschappen

Het promotieonderzoek is voorbereid aan het
Instituut voor de Lerarenopleiding van de Universiteit van Amsterdam

Table of contents

ACKNOWLEDGEMENTS

CHAPTER 1
Introduction 1

CHAPTER 2
Literature study 11

CHAPTER 3
Perceived pronunciation quality in mother tongue and foreign language 37

CHAPTER 4
Amsterdam English 47

CHAPTER 5
A choice of judges 59

CHAPTER 6
Judgements of Dutch and Dutch English 81

CHAPTER 7
Summary and conclusion 107

REFERENCES 121

AUTHOR INDEX 127

SAMENVATTING 131

CURRICULUM VITAE 143

.

ACKNOWLEDGEMENTS

Now that this dissertation has finally been completed I need to express my gratitude to those who have contributed to this work.

Without my two supervisors I might well still have been rudderless in an ocean of incoherent phrases and meaningless data, in which – without my two rescuers – I would probably have drowned. When, more than fifteen years ago, I approached Gert Rijlaarsdam with the first rough sketches, he told me that he was not allowed to supervise me; I was highly fortunate in that Gert was later authorized to be my supervisor; without his editorial acumen the separate parts would no doubt still have lacked cohesion; without his layout skills the appearance would have been a mess. When data analysis upon data analysis did not produce any results that made sense to me and I was desperate, Huub van den Bergh appeared on the scene; whatever the complaint was – tables that looked hopeless, data that could no longer be retrieved, results that were clearly impossible – Huub could be trusted to come up with a miraculous remedy. I have learnt a lot from these tutors.

The Institute for Phonetic Sciences of the University of Amsterdam has been more than helpful. Without Leo van Herpt I could never have designed the experiments; Leo was always there to help me with the data analysis; I am sorry that I cannot thank him in person. Also Ton Wempe always had time for me when I needed to use the recording facilities.

I am indebted to my own students, who inspired this study and whom I have always found willing to cooperate. I am most grateful to the schools in the Netherlands and the UK that allowed me to record their students, to the expert phoneticians who helped me in the selection of the speakers and to the students, teachers and lecturers, both in the Netherlands and the UK, who gave up part of their valuable spare time to listen to the recordings for little or no reward.

I must thank my colleagues in the department of English of d'Witte Leli – which later became Faculteit Onderwijs en Opvoeding of Algemene Hogeschool Amsterdam, later Hogeschool van Amsterdam, which faculty merged and became Educatieve Faculteit Amsterdam and after the failure of the merger resurfaced as Educatieve Hogeschool van Amsterdam, although the department retained a remarkable continuity – for their support. I realize that I have not always been able to contribute as much to the work of the department as I would have liked to. Of these colleagues I would like to mention Koos Meijers, who assisted me with the preliminary studies, and Leo Huisman, who was always most supportive.

Willy Weijdema taught me the necessary basic ICT skills and was always ready to help.

Theo van Els and Carlos Gussenhoven contributed a lot to an earlier version. I am sorry that Theo's encouragement did not produce the desired result in time for him to play a more active role in the completion of these studies. Of the many per-

sons who commented on earlier versions I must mention Cor Koster and the members of the commission, whose advice was most valuable.

In spite of all my supervisors' efforts, the work would probably still be unfinished if I had not been awarded a grant by the HBO raad; thanks to this grant I was privileged to work in the relative tranquillity of the premises of the Graduate School of Teaching and Learning of the University of Amsterdam, which, moreover, made a contribution towards the costs of printing this publication, as did the Amsterdam Institute of Education (EHvA).

I must thank my cousin, Bart Koet, for long and interesting discussions and my two sons, Thomas for producing a sand sculpture dedicated to Polder Dutch in Dutch ears and Bram for standing by me in the defence of this dissertation. I am profoundly indebted to my excellent wife, Nelly, who provided a domestic environment that could soothe me, even when I had been driven to distraction.

Finally, I have come to realize how much this work owes to my parents. I deplore the tardiness that has prevented them from witnessing its completion and dedicate this dissertation to their memory.

INTRODUCTION

1. PERSONAL REFLECTION

It is appropriate that I should start with a personal reflection on the work that I will report on in the following chapters. In this personal reflection I will first try to relate the work to my history as a native speaker of Dutch and as a non-native speaker of English. I will then attempt to relate it to my history as a teacher of English. Finally I will venture to relate it to my steps in research.

1.1 Language autobiography

In the following chapters I will describe work on judgements of pronunciation. I hope to show that judges' attitudes play an important part in the judging process. In judging pronunciation, judges do not seem be able to separate the speaker's identity as a member of a social group and his or her pronunciation. The anecdotal evidence of my language autobiography will bear this out.

I am a native of West Amsterdam[1] and a native speaker of a Western variety of Amsterdam Dutch. My parents both grew up in Amsterdam. My mother was from West Amsterdam and my father, who was born in East Amsterdam, spent his early years in North Amsterdam. My paternal grandparents differed from my maternal grandparents in that the former were natives of Amsterdam, whereas my maternal grandfather had been born in the East and my maternal grandmother in the South of the Netherlands and they only moved to Amsterdam later in their lives. As a result the men in my father's family spoke a variety of Dutch that had more Amsterdam characteristics than that of the men in my mother's family. As far as I remember none of the women in my father's or mother's family spoke a variety of Dutch that showed Amsterdam characteristics.

I became aware of the linguistic differences between my parents at an early age. Whereas my father prided himself on his Amsterdam identity, I never caught my mother asserting her Amsterdam identity. The difference was brought home to me when I was in primary school. The school was located in West Amsterdam, not far from the Central Amsterdam Jordaan area, part of which was in its catchment area. As a result two clearly different varieties of Dutch were spoken by my peers, a Western and a Central Amsterdam one, and the Central variety appealed to me as "tough" and "masculine". I distinctly remember the moment when I became conscious of the relative appropriateness of the two varieties. When I was sawing a piece of wood, the local medical general practitioner asked me what I was doing; to his rhetorical question I replied "sage" (Dutch *zagen*: to saw); my mother was not

[1] *The area was known as "Oud West" (Old West).*

amused; I had let her down in the presence of a "statusful" person. I realized that, apart from the inappropriately short one-word answer, I had perpetrated two other serious sociolinguistic crimes: to devoice and palatalise the sound represented by "s" and to retract and round the sound represented by "a" and decided not to sin in my mother's presence again. Although I must certainly have sinned in my father's presence, I do not remember him ever correcting my spoken Dutch.

In the last two years of primary school I was introduced to French and English – in that order. I remember that the French classes appealed to me more than the English ones; the (female) French teacher's pronunciation struck me as much better than the (male) English teacher's.

While in secondary school I found myself in an environment where non-standard speech was frowned upon. To my knowledge few teachers spoke Dutch with a trace of a non-standard accent. I well remember an outstanding Latin teacher with a distinct Amsterdam accent, which was remarked upon by the students. In the modern language curriculum little to no attention was paid to the spoken languages. As a result I had no idea whatsoever what native speakers of languages like English and French actually sounded like.

I spent my first undergraduate year in two universities in the Pacific Northwest of the USA[2]. For me, English had been a written language and the spoken English that I had heard most often was the British variant used by my Dutch English teacher. I was totally unprepared for the American English that I heard spoken around myself and – in retrospect – I must have displayed a serious lack of competences in nearly every field: sociocultural knowledge, intercultural awareness and intercultural skill and know-how; the most serious lack was probably in general phonetic awareness and skills.

Strangely enough my rudimentary phonological competence started to develop because of names, others' and my own[3]. One of the students I associated with was called *Hal*; I always addressed this person as *Hell*, which probably bothered nobody, until we had been invited to have dinner with my friend's family, where I caused him great distress; he accused me of using swear words in front of his younger sister. It was pointed out to me that the word *Hell* was inappropriate in mixed company, where I could have got away with *heck*. From this moment onwards I tried to distinguish the phonemes /e/ and /æ/. As to my own name: I was dismayed to find myself called what I understood as *Taan*; whenever I tried to correct this, the result was usually worse: *Daan.* Of course I did not realize that most English speakers in the Pacific Northwest pronounce words like *Tom* and *on* with unrounded vowels (cf. Wells, 1982, p. 130, p. 473), nor could I have known that an unaspirated pronunciation of the /t/ might strike an American listener as a realization of /d/. Also my last name caused great concern; the head of the English department insisted on calling

---

[2] *The universities were Gonzaga University in Spokane, Washington, and the University of Washington in Seattle, Washington.*

[3] *I believe that names, and in particular the learner's own name, are a useful topic in the teaching of pronunciation (see also Koet, 1977).*

me Mr *Co et*, which I resented[4], and others used a sound that was too dissimilar from the Dutch sound to my taste. It was only when I attended a course on the history of the English language[5] that I was made aware of the existence of the two English sounds found in the names *Susan* and *Woody* and realized that the Dutch sound was different from both. Partly thanks to this course I had probably acquired an acceptable pronunciation of the Northwestern variant of General American by the end of the academic year; Americans from other areas would on first acquaintance identify me as a person from the Pacific Northwest.

Another factor that may have contributed to my acquisition of an acceptable American English pronunciation was the accommodation policy of the first university that I attended. Exchange students like myself were lodged in one hall of residence; this was not a matter of segregation; each exchange student was made to share a room with an American student. This arrangement offered many advantages; the forced interaction promoted not only general but also communicative language competences; non-native speakers had to communicate with native speakers. As a result I had to make sense of various types of native and non-native English. Most of the American students were from the Northwest, but there were also speakers from the East, the New Yorkers amongst whom struck me as very different. I was privileged to listen to Danish, Finnish, French, German, Greek, Spanish, Italian, Pakistani, Brazilian Portuguese, Norwegian, Swedish and probably more non-native varieties of English. We unkindly made fun of speakers whose lack of phonological competence caused them to unwittingly produce obscene statements (the Greek speaker of English e.g. was unable to distinguish between the initial consonants in *Sam* and *Sean* and between the vowels in the stressed syllables of *Jean* and *Linda*, which caused great mirth in male but discomfort in mixed company). I found that there were considerable differences in the English produced by speakers of one and the same language; the Swiss German speakers' English sounded completely different from the North German speakers' English, but so did the English of speakers of Spanish from Mexico from that of Spanish speakers from Argentina and Uruguay. Non-native speakers were not the only victims of jokes; the friend's mother was a lady from the South, whose kind language and quaint phrases were continually made fun of by her husband and her son, and during Spanish classes, which several of the exchange students attended, they had a good laugh at the American students' inability to give vowels their correct "continental" value.

On my return from the USA, I enrolled in a course in Dutch and later English linguistics and literature in the University of Amsterdam. Although I do not recall being told so in so many words, it was brought home to me that in the English department an American pronunciation would not do. It was probably acceptable for native speakers of American English but definitely not for native speakers of Dutch. I much admired not only the contents of the lectures on phonetics and linguistics but also their presentation, which has always been an example to me[6]. I believe that

---

[4] *Later it became clear to me that he was right and I was wrong: his dissyllabic pronunciation of the originally Welsh name was historically correct.*
[5] *The textbook used was Robertson and Cassidy (1954).*
[6] *I am still grateful to the late Andries Vos.*

through my attempts to imitate the lecturer I managed to learn the pronunciation of British English fairly soon; but although it may have been acceptable to my tutors in the University of Amsterdam, it was certainly not liked by many of those American friends that I had not seen for a year; they accused me of not being myself.

When I was doing postgraduate work in Dublin[7], I was again confronted with the inability of listeners to separate pronunciation and social identity. Although I was always careful to tell people that I had to deal with that I was Dutch, most Irish people that I only spoke to briefly assumed that I was British – in fact I was employed by the English Language Institute in Dublin, which had a policy of only employing native speakers of English. When the troubles started in Northern Ireland and anti-British feelings were strong in Dublin, I encountered negative attitudes from time to time. I therefore adopted a modified Dublin accent, which provoked fewer hostile reactions.

The need to change varieties, both in the mother tongue (L1) and the second language (L2), was, I believe, a useful experience that made me more aware of the relation between accent and social identity. For me as a speaker, a change of accent was probably more an adoption of a different persona rather than an alteration of identity. After having studied the work on articulatory settings by Honikman (1964), I have consciously changed non-verbal behaviour, when switching between languages and accents, both in Dutch and in English. But in retrospect the new persona must have given some (naïve?) listeners the impression of a different identity, which fit one of their stereotypes.

1.2 History as a teacher and teacher educator

My personal history as a teacher of English started in Amsterdam. For two years I was privileged to work in an English department where the coordinator was totally dedicated to the use of the target language[8]. The children had to be kept busy in a sort of language laboratory with interminable, often meaningless, drills. I found that nearly all the pupils, regardless of their linguistic background – there were some native speakers of languages other than Dutch and several speakers of non-standard Dutch – managed to acquire an acceptable pronunciation, although it must be admitted that the dogmatic methodology did not always contribute to their motivation.

In the English Language Institute in Dublin I had to teach students from many different linguistic backgrounds; there were speakers of European languages such as French, German, Italian and Spanish but also of Arabic and Japanese. Although the students had been assigned to groups on the basis of their proficiency levels, there was no doubt in my mind that even within these groups there were considerable differences in the ability of speakers of the various L1s to improve their pronunciation. Where I had some basic knowledge of the phonology of the languages concerned, I tried to base my attempts to help the students improve their pronunciation on a contrastive analysis of their L1 and the intended L2, which for practical purposes was not Irish English or Dublin English but British English. I never questioned the con-

---

[7] *I followed the graduate course in Anglo-Irish studies at University College Dublin.*
[8] *For an in memoriam of Jan Mooijman, see Van Essen (2003).*

trastive approach but I soon discovered that there was no such thing as a single German L1; the differences between the difficulties experienced by speakers of Swiss German and those experienced by speakers from Northern Germany were probably just as great as the difference between the difficulties experienced by speakers of Italian and those experienced by speakers of Spanish. What I found as a teacher of English was reinforced by my work as a student in University College Dublin. Although the focus of the postgraduate work was on Anglo-Irish literature, there were supporting courses on history, culture and linguistics and in addition students were encouraged to study Gaelic. Here I was introduced to the development of the different varieties of spoken English in Ireland, both on the basis of a Gaelic substratum – the several dialects of Gaelic producing slightly different varieties of Hiberno-English – and on the basis of a dialect from the West of England in the case of Dublin English (cf. Bliss, 1979), which made me look at the various non-native Englishes from a different perspective.

When I became a teacher educator, first in the South of the Netherlands and later in Amsterdam, my first priority was always to make the students aware of the phonemes of English. It was my experience that the relation of English sounds and English spelling was the main obstacle to the acquisition of an acceptable pronunciation and that it was best to help learners to overcome this obstacle first (cf. Koet, 1977). Once the learners were aware of the sounds of English, the interference of their native language needed to be addressed. I did not hesitate to use a contrastive approach, encouraging students to study the sound systems of their L1 and the L2. At first I did not pay attention to the different varieties and restricted myself to the standard accent of Dutch (Algemeen Nederlands or AN) as the L1 and of English (Received Pronunciation or RP) as the L2, but as the student population became more heterogeneous – or perhaps as I became more alert to its heterogeneity – as far as the L1 and the intended L2 were concerned, I tried to discriminate between the varieties of L1 – Dutch, Spanish, Papiementu, Turkish, Moroccan Arabic, Berber, etc. – and L2 – British and American. The trainees did not only have to become aware of the sounds of their own L1 and L2 but also of those of their (future) pupils. For this purpose I designed a number of diagnostic exercises that would provide the trainees with a relatively simple instrument for dealing with pupils' pronunciation problems.

My practice in teaching the phonetics and pronunciation of English has always been based on contrastive linguistics and in that respect I would probably differ little from most other practitioners – perhaps only in that I try to pay more attention to the varieties of the L1 and the L2. Although I had no reason to doubt the great value of the contrastive approach, however, I began to suspect that it could not account for all the problems in pronunciation; it was not so that all my students with the same variety of L1 were equally successful in acquiring the pronunciation of English. Apart from general phonetic competence a positive or negative attitude towards the L2 and a strong or weak desire to become proficient speakers of that L2 must have played an important part.

As a teacher educator I had the privilege to visit many schools, where I was supposed to observe the trainees. I became acquainted with the views and the practices of many teachers of English and found that foreign language teaching was still as

strange as it had been in my own secondary school (cf. Rijlaarsdam, 2001). In the first school in Amsterdam I visited as a teacher educator I was not entirely surprised to hear the experienced teacher assert that the main aim of English classes was to improve the pupils' Dutch. He acted in accordance with his aims, since no English was spoken in his classes. Indeed, one unfortunate child, a boy from England, failed his English course miserably because of his spelling of Dutch, a fate that he shared with several other native speakers of English in schools that I visited later. Also during French and German classes native speakers of French and German could be seen or heard to be penalized for their lack of proficiency in Dutch, so that, paradoxically, some of the most proficient users of the foreign languages received the lowest marks. This made me suspect that Dutch teachers' judgements of English, French and German proficiency were often a more accurate reflection of their perceptions of the learners' Dutch language proficiency than of the learners' foreign language proficiency.

Attitudes towards varieties, both in the native and in the foreign language, seemed to me more and more important in judging pronunciation as I worked with more colleagues, native as well as non-native speakers of English, who were asked to assess fluency and pronunciation, not only of first-year but also of second-year and third-year students; in the first years this was no easy task as there were hardly any guidelines as to the distinction of the levels. During training sessions, in which teacher educators from many institutes in the Netherlands met to create some sort of consensus on standards, native-speaker judges' relative leniency and non-native speaker judges' relative strictness became apparent to me. It seemed to me that this was a serious matter; students who were judged by native speakers would have an advantage over the other students.

1.3 First steps in research

My first attempt at quantitative empirical – I must admit that the term would not have made any sense to me at the time – study of attitudes towards varieties concentrated on native speakers. Using the matched guise technique popular at the time I created a number of readings of the same text, each with a different pronunciation of one phoneme and a reading where all the words that would normally be shortened (weak forms) had their citation forms (strong forms), which I then collated on a tape (this was before audio cassettes had become popular). I managed to get a number of native speakers to listen to the recording and indicate their reactions on a number of seven-point scales. I found that these listeners were most negative about the two versions with suprasegmental deviations and one version with an allophonic deviation (i.e. a deviation that could not bring about a difference in meaning); curiously they were less negative about phonemic deviations (i.e. deviations that caused a different meaning, such as the substitution of the initial consonant in *Sean* for the one in *Sam*). This was a lot of fun and it is unfortunate that the data have disappeared, so that they cannot be analysed again[9]. I then embarked on a study of the pronunciation of Amsterdam speakers. By that time I had become convinced that great injustice

[9] *The report on this study (Koet, 1976) was never made available in English.*

was done to speakers of non-standard varieties of Dutch; as a result of the curious foreign language teaching practices prevailing in the Netherlands it was impossible for many learners whose Dutch language proficiency left something to be desired but whose English was perfectly all right to obtain good marks for their English. When I presented my first tentative results, they met with disbelief and I tried to think of better ways of supporting my hypotheses. In this I received generous and disinterested help from staff of the Institute of Phonetic Sciences of the University of Amsterdam.

Initially it was never my intention for these exercises in research into the relation of the mother tongue and foreign language to grow into a more serious effort. In a way the various institutes for teacher education for lower-secondary schools had been a world to themselves, with relatively few direct links with the world of higher education, which was perceived as a series of ivory towers, divorced from the grim reality of lower-secondary education. Although in the first few years lip service was paid to the notion that staff at these institutes should do research, time set aside for it soon fell to the axe of budget reductions. When the institutes became part of the higher professional education sector, the notion was abandoned. I well remember naïvely asking the director of one university research institute for assistance with the set up of my study. When it became clear to him that I had not brought a bag of funds, he told me in no uncertain terms that he had no intention of doing things for love and that teachers should not be involved in research[10]. I would probably have heeded his "terribly sensible advice"[11] if it had not been for the establishment of the Graduate School of Teaching and Learning of the University of Amsterdam, to which I was referred by my vice-principal.

In spite of the help and support so generously given by the staff of the Institute of Phonetic Sciences and the Graduate School of Teaching and Learning of the University of Amsterdam and the department of Applied Linguistics and Methodology of Nijmegen University it would be another fifteen years before the work could be presented in a more definitive form. I became involved in a number of international projects; these projects provided me with access to what I so badly needed: British listeners. At the same time the international cooperation projects consumed so much time and energy that not enough could be devoted to the completion of my research. I made some vain attempts to obtain funding but all my submissions were rejected. Perhaps more seriously, I had not found a way of analysing my data in such a way that my hypotheses could be supported and my plight seemed hopeless until I was introduced to Utrecht Institute of Linguistics, where many of my problems were miraculously solved.

---

[10] *The terms used were "we hebben geen tijd voor liefdewerk, oud papier" en "die leraren moeten onderzoek aan ons overlaten".*

[11] *I am quoting the phrase used by Thomas Hardy to describe the letter sent to Jude Fawley by the master of Biblioll College: "Sir, I have read your letter with interest; and, judging from your description of yourself as a working-man, I venture to think that you will have a much better chance of success in life by remaining in your own sphere and sticking to your trade than by adopting any other course". (Hardy, 1960, p. 138).*

Although I deplore the long period that has elapsed between the beginning of my work and its completion – and wish I had saved time by conducting the experiments properly from the beginning – I do not regret having done things this way. Working on this task has proved a valuable learning experience. On a trivial level I have been forced to learn to use tools or use tools better that stood me in good stead in other tasks. I have acquired skills that I would never have acquired otherwise. Starting from an almost complete lack of numeracy I have lost my worst awkwardness with numbers, which was of great benefit to my work in other endeavours. I hope to think that my writing skills have improved. Having seen my submissions for funding for my research rejected again and again, has taught me how to make project proposals that did get accepted. Although I am still lacking in accuracy I have learnt the hard way that I must be more accurate than I used to be. It would be vain for me to pretend that all my competences as a teacher educator have developed to an acceptable level, but I am quite confident that some have improved – although others may well have deteriorated. But perhaps the most important thing that I have learned is to mistrust human judgement, the judgement of others but most of all my own judgement, and always study the facts. I can only hope that I apply what I learnt – most of the time.

2. CONTENTS

After this introduction, some of the relevant literature on the judging of pronunciation is discussed. This literature study starts by examining the aims of the teaching and learning of pronunciation; it is shown that there is no aim that is generally accepted; although there seems to be a majority in favour of intelligibility, a substantial minority appears to be in favour of accent reduction. The aims are then related to several levels of language proficiency and it is shown that the aim of intelligibility is found more frequently at the lower levels, whereas the aim of accent reduction is found more frequently at the higher levels. In the second part of the literature study an attempt is made to identify factors that may influence the judging of pronunciation; most factors belong to three groups: factors having to do with the judges, factors having to do with the learners, and factors having to do with the tasks of both speakers and judges. Finally questions that are in need of further study are identified.

The following chapters are preliminary studies. In the first preliminary study the circumstances that occasioned this study are described: the opinions expressed by several Dutch teachers – that it was harder for speakers of certain varieties of Dutch to receive a positive judgements of their pronunciation of English than for speakers of other varieties – and the suspicion that there might be an element of truth in these opinions; an attempt is made to support these suspicions by some findings; at the end of the chapter three tentative hypotheses are formulated to account for the difficulties experienced by these speakers. In the next preliminary study the first of these tentative hypotheses is worked out. This study is on the pronunciation of English by speakers of Amsterdam Dutch and an attempt is made to demonstrate that it may

well be harder for these speakers to acquire the pronunciation of English than for speakers of other varieties of Dutch.

Each of the following two chapters is devoted to a report on one of two empirical studies. The first empirical study deals with judgements of the pronunciation of English by speakers of Dutch, which judgements were made by several groups of native and non-native listeners; an attempt is made to identify characteristics of judges and of speakers that may influence the judgements. The second empirical study concerns judgements of the pronunciation of English and of Dutch by speakers of Dutch; another attempt is made to identify characteristics of judges and of speakers; finally the relation between judgements on the mother tongue and the foreign language is addressed[12].

A final chapter is devoted to the limitations of the present work, the conclusions that may be drawn in spite of these limitations, suggestions for further research and recommendations for practitioners.

I gave this dissertation the title *Polder English in Dutch ears.* The term polder was intended as a synonym of Dutch, on the analogy of such recent Dutch coinages as *poldermodel* and *polderpop*[13]. I preferred this title with its suggestion of a recent variety spoken in the low areas of the Netherlands to the alternative "Dutch English in Dutch ears", which might have been taken to refer to the English spoken in the entire Dutch language area, including Belgium. The title *Polder English in Dutch ears* would not have been chosen if I had not been aware of the work in which the term "Polder Engels" (Polder English) was, to my knowledge, used for the first time (Stroop, 1998, p. 75). There it was used to refer to a variety of English spoken in London and in the southern part of England. "The speakers and the variety are called Cockney (….) but it is of course ordinary Polder English, brought about by the same factors as Polder Dutch. Phonologically, the English variety is one round ahead of the Dutch one. But there is another difference. Cockney is 'lower-class' English, whereas Polder Dutch was initially spoken by the upper middle classes." (my translation). It seems to me that it would be more appropriate to compare Polder Dutch with Estuary English, which is not confined to the lower classes, rather than with Cockney (see Rosewarne, 1994a and 1994b). But here I do not use the term Polder English as a synonym for Cockney or Estuary English, which are native varieties of English. Neither do I use it to refer to a clearly defined non-native variety of Eng-

---

[12] *As the literature study was written after the other chapters had been written, these chapters had to be revised. In particular chapter 3 and chapter 4 had to be rewritten in the light of more recent insights.*

[13] *Den Boon & Geeraerts (2005, p. 2730) describe these as "a consultation model aimed at consensus, as used in the Netherlands in the nineties of the twentieth century" and "Dutch pop music," respectively (my translations).*

lish. It is not restricted to the English produced by the upper middle class speakers of Polder Dutch; rather it refers to the English produced by speakers of Dutch in the Netherlands, whether these varieties are upper or lower middle class, or even working class.

LITERATURE STUDY

> Then said they unto him. Say now Shibboleth: and he said Sibboleth: for he could not frame to pronounce it right. Then they took him and slew him at the passages of Jordan: and there fell at that time of the Ephraimites forty and two thousand. [Judges 12-6]

The importance of pronunciation in language communication was already demonstrated in the Old Testament. Unfortunately the text does not provide the reader with full details of the situation but a number of facts can be inferred: the pronunciation of an Ephraimite speaker of Hebrew was judged by Gileadite speakers; the task was the production not of a sentence but of a word; what was judged was the pronunciation of a single sound, the sound represented by "sh"; the error was the substitution for this sound of a different sound, the one represented by "s". Other questions may be raised: was there one judge per Ephraimite or were there more, were the judges persons with experience in language teaching or were they naïve, had the judges been given specific training or were they untrained, were they male or female, were the judges strict or lenient, were the judges consistent? It is known that the consequences of an incorrect pronunciation for the unfortunate test-takers were disastrous and it may be concluded that in the period concerned learners of Hebrew did well to pay attention to their pronunciation. It looks as if there is some doubt if the same is true for learners of English in the Netherlands today.

Some ninety years ago Nolst Trénité (1920) encouraged Dutch learners of English to drop their foreign accents. Presumably learners were intended to drop their Dutch accents in favour of standard British ones. No doubt a lot has happened since this exhortation, but the pronunciation of English by speakers of Dutch has not received a great deal of attention in the last few years. Derwing, Munro and Wiebe (1998) were able to state that "interest in pronunciation has increased dramatically in the last decade" (p. 394). To make the same claim now about the Netherlands would be an overstatement. Indeed, in the introduction to his defence of the teaching of the pronunciation of English to Dutch secondary-school students Lowie (2004) states that one can rightly wonder if it makes sense to spend time and energy on this task in view of the fact that such students are generally fairly intelligible. But the Gileadite must also have been fairly intelligible; he pronounced most of the sounds correctly. Fortunately Lowie (2004) goes on to point out that it is the goal of the teaching of pronunciation that will determine whether the answer to his rhetorical question is affirmative or negative and the aims of teaching (and learning) pronunciation are what need to be examined first. After this examination of the aims, the factors that influence judgements on pronunciation will be examined.

1. AIMS OF TEACHING PRONUNCIATION

1.1 Intelligibility or native-likeness?

It seems that more authors of textbooks for language teachers now seem to have been converted to "intelligibility" as the aim of teaching pronunciation than in the past. If this is true of the majority of practitioners is doubtful. Levis (2005) compared the two "competing ideologies", claimed that "although an overwhelming amount of evidence argues against the nativeness principle, it still affects pronunciation teaching practice" (p. 370) and recommended that more attention should be paid to the communicative context of native – non-native speaker interaction as well as language identity and language attitudes (p. 374). The native speaker norm was shown to be dominant in Greek teachers' beliefs by Sifakis and Sougari (2005), who suggested that there is no reason to believe that the same is not true in other countries where English is a foreign language (p. 483). Adherents to the "nativeness principle" were castigated by Canagarajah (2005), who stated that "pedagogies for accent reduction have bordered on the pathological" (p. 365) and Levis (2005), who criticised "many teachers, especially those unfamiliar with pronunciation research," who "may see the rare learner who achieves a native-like accent as an achievable ideal, not an exception" (p. 370). Cook (1999) claimed that second language "users who cannot be distinguished from native speakers are as typical of human beings as are Olympic high jumpers or opera singers" (p. 191). These are strong words and one may wonder if such sweeping statements would have been made if the authors had been more familiar with the level of English proficiency in some European countries in general and the Netherlands in particular, where e.g. Bongaerts, Summeren, Planken and Schils (1997) demonstrated that it is possible even for persons who started learning English late to attain a native-like pronunciation.

If it is true that Dutch secondary-school students are generally fairly intelligible, it could not possibly be claimed that these students have also generally succeeded in dropping their foreign accents. And if they have not, would it not make sense to spend time and energy improving their pronunciation? Munro and Derwing (1995a) examined the interrelationship between accentedness, perceived comprehensibility and intelligibility and showed clearly that foreign accent on the one hand and intelligibility and comprehensibility on the other are correlated, but that "a strong foreign accent does not necessarily reduce the comprehensibility or intelligibility of L2 speech" (p. 74). Although they showed that the processing time of Mandarin accented English was longer than that of unaccented English they claimed that the Mandarin "accented productions were nearly always intelligible and were often rated as highly comprehensible" (Munro & Derwing, 1995b, p. 320). But listening conditions in the real world are not always as ideal as in laboratories. Munro (1998) studied the effect of artificial cafeteria noise on the intelligibility of Mandarin accented speech and found suggestions "that the addition of noise may result in a larger drop in intelligibility for accented speech than for native-produced speech" (p. 149). The addition of "multi talker babble noise" was shown to have a similar effect on the perception of Swiss-German English, which was perceived just as easily as

native English when it was not degraded (Bürki-Cohen, Miller & Eimas, 2001, p. 165).

Most authors of recent textbooks seem to be in favour of "intelligibility" as the aim of teaching and learning pronunciation. In his language teaching methodology textbook Nunan (1995, p. 115) affirmed that "in terms of teaching goals, the shift has been to focus on the development of communicative effectiveness and intelligibility, rather than on the development of native-like pronunciation". Pennington (1996, p. 220) held intelligibility to be "the most obvious, justifiable and pressing goal in the area of phonology"; to her native pronunciation seemed an unrealistic goal in the majority of cases, where fluency and accuracy would be possible goals in addition to intelligibility. Her position was modified when she admitted "that the nature of students, their levels and the goals of the language program will determine what other goals (if any) are appropriate" (p. 222). Celce-Murcia, Brinton and Goodwin (1996, p. 8) having distinguished six groups of learners, affirmed that "the goal of teaching pronunciation to such learners is not to make them sound like native speakers of English, but only to help them pass the threshold of intelligibility"; the fifth group of these learners consisted of "teachers of English as a foreign language who are not native speakers of English and who expect to serve as the major model and source of input in English for their students". Jenkins (2000) went so far as to recommend the Lingua Franca Core, a less ambitious goal for pronunciation teaching, to solve problems of "mutual phonological intelligibility and acceptability"; she herself, having studied in-depth interviews with eight non-native teachers of English, found that most of these teachers regarded deviations from the native norm – Received Pronunciation (RP) or General American (GA) – as incorrect and she had to admit that it cannot be taken for granted that non-native teachers of English as a foreign language "wish unequivocally to use their accented English to express their L1 identity…" (Jenkins, 2005, p. 541). In her review of recent research in teaching pronunciation Jenkins (2004, p. 113) distinguished the learners of a lingua franca from learners of a foreign language, who would "acquire not only the phonemic distinctions of the L2 but also near-nativelike realizations of individual phonemes according to the phonetic environment, along with many of the suprasegmental features of the foreign language".

Authors of research studies are not unanimous in claiming that intelligibility is the only goal. Wang and Van Heuven (2003), having described the practical solution of a much reduced version of English for aviation, stated that it would be more principled "to find out what it is that causes a foreign accent in the non-native forms of English, and to develop a successful pedagogy that would allow anyone in the world to learn to speak English without even the slightest trace of a foreign accent" (p. 213). Macdonald, Yule and Powers (1994) used the degree to which the pronunciation is targetlike as the criterium for improvement in English L2 pronunciation. In other studies the aims of intelligibility and native-likeness seem to be combined. Andersen-Hsieh, Johnson and Koehler (1992) indicated that speakers can reach three levels: a low level which is heavily accented and unintelligible, a mid level, which is accented but intelligible and a high level, which is near-native – apparently

at this high level they felt no need for intelligibility to be specified[14]. Albrechtsen, Henriksen and Færch (1980), who studied British listeners' reactions to Danish students' interlanguage, made it clear that an interlanguage "which is deviant from the target language is not necessarily difficult to understand" (p. 394).

As early as in 1978 Gimson allowed for the possibility that for some categories of learners it would not be necessary to "reach a level of pronunciation which will be not only easily intelligible but also readily acceptable to a native English speaker" and that such learners could resort to some sort of Rudimentary International Pronunciation or other (Gimson, 1978, p. 46). Although Gimson (1970) did not explicitly speak out in favour of or against "intelligibility" as a goal in learning - teaching pronunciation, he warned against "exaggerated articulation in order to achieve clarity", presumably because such an exaggerated articulation would not be perceived as native-like (p. 4). Although recommending pronunciation instruction guided by intelligibility Derwing (2003) showed that many of her respondents wished to sound more native-like because "they had been treated badly because of their accent". These respondents may have experienced that "in addition to non-standard pronunciation's potential for obstructing meaning, it has also been shown to set up social barriers between speaker and interlocutor," as Schairer (1992, p. 310) pointed out. It would therefore seem that the aims of intelligibility and "target-likeness" may be at odds [15].

Even if the only aim of the teaching and learning of pronunciation is for the learner to be understood, it is not always clear by whom. Most earlier authors seem to have assumed implicitly that the intended listener was a native speaker of English, to whom the most native-like pronunciation may well have been the most intelligible. Recently, however, several authors rejected the idea that the native speaker is necessarily the intended listener. One argument for the rejection of the native speaker is that there are now – or that there will soon be – more non-native speakers than native speakers of English. Crystal (1997) estimated that there were some 337 million L1 speakers of English versus 235 million speakers of English as an L2, as well as a larger number of speakers of English as a foreign language and Graddol (1999) showed that the proportion of native speakers versus non-native speakers of

---

[14] *In their study they asked their judges to score prosodic elements of the pronunciation of non-native speakers on a "least native-like … most native-like scale", which would not seem appropriate for their lowest level, where native-likeness need not be achieved.*

[15] *A most curious position seems to have been taken by Golombek and Jordan (2005), who asserted that non-native teachers of English should no longer strive to be intelligible: " ..the single-minded focus the profession has placed on developing intelligibility within speaking and pronunciation classrooms should be questioned. For L2 speakers, intelligibility is an illusion…" (p. 529); apart from the question if it is justified to draw such a far-reaching conclusion on the basis of a case study involving only two Taiwanese teacher trainees, the objection might be raised that, if these teachers did not have to pass the threshold of intelligibility, their students would probably be totally incomprehensible. On closer reading, however, it appears that the authors did not heed the advice by Munro and Derwing (1995a, p. 93) that "scales of accent, perceived comprehensibility and intelligibility ought not to be confused with one another" and recognized themselves that the distinction between the terms intelligibility and accent became blurred in their article.*

English is declining. It is doubtful if these arguments would necessarily hold for all learners. It would probably be hard to prove that significantly fewer learners of English in the Netherlands would expect to use English in their communication with native speakers, whether they are from the USA, the UK or other parts of the English-speaking world, than in their communication with non-native speakers, whether these are speakers of English as a second language or English as a foreign language.

Even if there were consensus about the intended listeners – native speakers or non-native speakers – that would not mean that there is agreement about the model to be chosen. But should it not be up to the language learners rather than the teachers or researchers to decide what the aims are? Until fairly recently most learners in the Netherlands would probably have been happy to accept a British English model. Van der Haagen (1998) studied Dutch secondary-school students' pronunciation preferences; she asked her respondents to choose between RP, GA and either as choices of the accent that they preferred to use themselves or that they preferred their teachers to use and found that slightly more secondary school students in the Netherlands would rather use GA than RP, although the vast majority wanted their teachers to use RP; she did not, however, offer a fourth possibility, an intelligible accent, but she mentioned that teachers should not "allow a kind of 'Dutch English', only intelligible to other Dutch (and possibly other Germanic) speakers" (p. 105). The speculation by Riney (1998) about "speakers from Singapore, India, or perhaps Holland, whose target language norm is, as time goes on less and less likely to be a native speaker one based in the UK or the USA" notwithstanding, it is probably unlikely that many of Van der Haagen's respondents would have chosen such a fourth option. It is of course possible that preferences have changed in the last few years and one could speculate that a native Caribbean English accent could now be attractive to some learners in the Netherlands.

It would not seem advisable for learners in the Netherlands to choose a non-standard model; Deterding (2005) found that native speakers of Singapore English did not only find fragments of Estuary English (a variety that is widely spoken in the South East of England) hard and sometimes impossible to understand, but – although recommending exposure of learners to such non-standard varieties – he also reported that the listeners strongly resented the speakers of this variety. This might be an argument for agreeing with Riney, Takagi and Inutsuka (2005), who concluded that for the English as a foreign language context either GA or RP would suffice as models. This conclusion combined with the distinction made by Roach (2000, p. 6) between the model and the goal – "the model chosen is BBC" (English) "(RP) but the goal is normally to develop the learner's pronunciation sufficiently to permit effective communication with native speakers" – would seem to make sense for the majority of learners in the Netherlands.

In view of the debate about the aims of teaching and learning pronunciation, it would seem advisable to examine the positions taken in official documents that are relevant for the situation in the Netherlands.

1.2 The Common European Framework of Reference on pronunciation

Probably the most influential document about the aims not only of teaching pronunciation but of language learning in general is now the Common European Framework of Reference (CEFR, Council of Europe, 2001). In setting out the concept of plurilingualism, it is outspoken in repudiating the notion of the "ideal native speaker" as the ultimate model (pp. 4-5). Although it does not explicitly reject the ordinary – as opposed to the ideal – native speaker as a model, it is obvious that here the intended listener is not always assumed to be a native speaker of the target language.

The document does not provide clear guidelines as to what can be required from language learners and users as far as pronunciation is concerned in the relevant communicative language activities, spoken interaction and spoken production. Here the document indicates merely that in order "to speak the leaner must be able to *articulate* the utterance (phonetic skills)" and that in order "to listen the learner must be able to *perceive* the utterance (auditory phonetic skills)" going on to state that "the more mechanical meaning-preserving activities (repetition, dictation, reading aloud, phonetic transcription) are currently out of favour in communication-oriented language teaching owing to their artificiality and what are seen as undesirable backwash effects". Nevertheless the document argues that many learners will profit from general phonetic awareness and skills as distinct from the ability to pronounce a particular language (Council of Europe, 2001, pp. 90, 99).

In the subsequent chapter about the user's / learner's competences, however, the document distinguishes a number of linguistic competences under the heading communicative language competences, among which phonological competence and orthoepic competence, thus giving new currency to the eighteenth century term "relating to correct or accepted pronunciation" (Murray, Bradley, Craigie, & Onions, 1933). The phonological competence is said to involve a knowledge of and skill in the perception and production of phonemes and allophones, the phonetic features, the phonetic composition of words, sentence phonetics and phonetic reduction.

The levels of phonological control as related to the CEFR levels of language proficiency are presented in table 1.

As control at each higher level entails control at the lower level, it can be observed that at the three lowest levels intelligibility is of the greatest importance. After level A2, however, intelligibility is no longer sufficient. At level B1 the foreign accent is evident only sometimes and mispronunciations can only be occasional. At level B2 pronunciation must not only be clear but also natural and intonation also starts playing a part. At level C1 intonation and sentence stress must be used correctly, level C2 being identical with C1 for this competence. There seems to be a progression from segmental to supra-segmental features as phonological control and language proficiency increase.

The orthoepic competence is said to be relevant for users who may be required to read aloud a prepared text, or to use words in speech that they had first encountered in their written form. This competence is said to involve – amongst other matters – the ability to consult a dictionary and knowledge of the conventions used there for the representation of pronunciation.

Table 1. Levels of phonological control and language proficiency
Adapted from Council of Europe, 2001, p. 117

Level of language proficiency		*Level of phonological control*
C1	Effective Operational Proficiency	Can vary intonation and place sentence stress correctly in order to express finer shades of meaning.
B2	Vantage	Has acquired a clear, natural, pronunciation and intonation.
B1	Threshold	Pronunciation is clearly intelligible even if a foreign accent is sometimes evident and occasional mispronunciations occur.
A2	Waystage	Pronunciation is generally clear enough to be understood despite a noticeable foreign accent, but conversational partners will need to ask for repetition from time to time.
A1	Breakthrough	Pronunciation of a very limited repertoire of learnt words and phrases can be understood with some effort by native speakers used to dealing with speakers of his/her language group.

Although the guidelines provided by the CEFR as to pronunciation are not as detailed as they might have been and allow for more interpretations – e.g. what is meant by a natural pronunciation; is it natural for the speaker or for the listener? – they nevertheless provide a basis from which others – learners, teachers – can start measuring progress in each of the two relevant competences from intelligible pronunciation at the lowest level (basic user), via some traces of an accent at the intermediate level (independent user) to control of intonation and sentence stress at the advanced level (proficient user) in each of the two relevant communicative language activities. Examples of situations in which language learners and users can demonstrate the two competences are given in table 2.

Table 2. Situations in which phonological and orthoepic competences can be demonstrated

Activity	*Competence*	
	Phonological	*Orthoepic*
Spoken production	Delivering a monologue	Reading aloud; delivering a paper
Spoken interaction	Taking part in conversation	Introducing unheard word in conversation

The CEFR does not attempt to provide illustrative scales for the various levels of the orthoepic competence; admittedly it would be hard to think of an activity of spoken interaction in which this competence could be demonstrated (the introduction into a conversation of words that the learner has only seen in writing but has never heard spoken) but communicative activities that demonstrate the orthoepic competence in

spoken production can be found easily (reading of stories, delivering a paper). Of course, authors of less recent studies could not have been aware of the framework at the moment of their studies but in retrospect it would seem that – if they had been aware of the two competences and the two communicative language activities – more attention would have been paid to pronunciation in spoken interaction – however problematic that would have been as Van der Haagen (1998, p. 19) rightly pointed out – than in spoken production and that the orthoepic competence could have been distinguished more clearly from the phonological competence. This distinction was made implicitly by Gussenhoven and Broeders (1997, p. 20), when they warned student teachers that the first obstacle encountered by Dutch students in their acquisition of the pronunciation of English is the spelling and that interference of the native language is only the second obstacle.

If the linguistic competences, communicative language activities and levels are combined, one would arrive at a hierarchy that would strongly resemble the three levels found in Lowie (2004, pp. 4-5) as presented in table 3.

Table 3. Possible correspondence between CEFR levels and levels found in Lowie (2004)

CEFR level	*Level in Lowie (2004)*
B1	Intelligible
B2	Pleasantly intelligible
C1	Native-language like

1.3 Cambridge ESOL examinations on pronunciation

The guidelines found in the CEFR are followed by providers of examinations such as Cambridge ESOL. The CAE Handbook for the Certificate in Advanced English e.g., which examination is claimed to be at level C1, requires the candidate "to produce comprehensible utterances...the production of individual sounds, the appropriate linking of words, and the use of stress and intonation to convey the intended meaning. First language accents are accepted, provided communication is not impeded." (Cambridge ESOL, 2003a, p. 51). At the higher C2 level the CPE Handbook for the Certificate of Proficiency in English requires

> the ability to produce easily comprehensible utterances. Articulation of sounds is not required to be native speaker-like but should be sufficiently clear for all words to be easily understood. An acceptable rhythm of connected speech should be achieved by the appropriate use of strong and weak syllables, the smooth linking of words and the effective highlighting of information-bearing words. Intonation, which includes the use of a sufficiently wide pitch range and the appropriate use of contours, should be used effectively to convey meaning. (Cambridge ESOL, 2003b, p. 75).

Although these guidelines are also not as detailed as they might have been, it appears that the requirements for Cambridge Advanced English are less strict than

those of CEFR level C1, whereas those for the Certificate of Proficiency in English would seem to be at that level. The communicative language activities are spoken production and spoken interaction with a native speaker (the examiner) and a non-native speaker (another candidate), who may be a native speaker of the same L1. It is to be observed that in neither of these two Cambridge ESOL examinations an attempt is made to distinguish between the phonological and the orthoepic competence.

1.4 Aims of teaching English pronunciation in the Netherlands

Unfortunately the implementation of the CEFR has not yet reached the stage at which one can safely relate the aims of English language teaching in the various types of schools in the Netherlands to the levels of the CEFR but it would not be too far fetched to speculate that the aims for English for the more academic types of secondary schools might well be B1, whereas they would be A2 for the more vocational types of secondary schools, B2 being required after one year of higher education and C1 at the end of the bachelor period (Koet & Weijdema, 2005).

As far as the goals of teaching pronunciation in the Netherlands are concerned, secondary-school teachers may be guided by a number of official documents: the attainment targets for the lower secondary school and the requirements for the several examinations in the upper secondary school.

The relevant attainment targets, which were valid from 1998, are mentioned under conversational skills; at the end of the lower secondary school students are required to express themselves in such a way that their intention can be understood without great effort; a certain degree of correctness of pronunciation is necessary (Examenblad, 2006). What is required is therefore not intelligibility of individual words but the perceived comprehensibility of the utterance. Staatsen (2004) added that a higher degree of correctness would be required from students in the more academic types of secondary school. In 2006 these attainment targets were replaced with new attainment targets, which are much less detailed (Onderbouw VO, 2006). The relevant target is number 11: the pupil learns how to become more familiar with the sound of English through frequent listening to spoken and sung texts. The attainment targets are related to the CEFR in the sense that the five aspects of language proficiency (listening, conversation, speaking, reading and writing) are mentioned but the required level is not specified.

In the specifications for the examinations Modern Foreign Languages for the more academic types of secondary school, conversational skills are found; candidates can reach one of four levels; at the lowest level (one) candidates are required to express themselves fairly comprehensibly and intelligibly, at the intermediate levels (two and three) candidates are required to express themselves comprehensibly and intelligibly, whereas at the highest level (four) no more is said about comprehensibility and intelligibility (Directie Voortgezet Onderwijs, 2005b). It is not clear if candidates at level four are expected to have progressed beyond the level at which comprehensibility and intelligibility are the only issues. Although an attempt was made to determine the relation of the NL examinations to the levels of the CEFR

and the relevant skill (speaking) was found to be at B1.2 for the more academic types of secondary school (Van Hest, De Jong & Stoks, 2001)[16], the actual requirements do not make this link. It is clear that these guidelines are of less help for the practitioner than those of the CEFR. Indeed, for the examinations for the more vocational types of secondary schools there are no guidelines at all as to the required level of the conversational skills, let alone the pronunciation (Directie voortgezet onderwijs, 2005a). Also for the primary schools there are no guidelines as to the required level yet, although it was proposed that level A1 should be adopted (Vedocep, 2004). In none of the specifications is there an attempt to distinguish between the two relevant competences and the two communicative language activities.

Fortunately the module guides for the bachelor's programmes in English language and literature or English language and culture of several universities in the Netherlands provide more information as to what is required from the students as far as their pronunciation is concerned.

The aims of the Language Proficiency module in Utrecht University e.g. stipulate that the student should improve his-her English fluency and pronunciation, that the student should acquire insight in articulatory and contrastive phonetics and that the student should learn how to make a phonetic transcription (Universiteit Utrecht, 2005). The aims of the Phonology course in the University of Amsterdam are "to improve students' pronunciation of English by increasing their analytic awareness of the sound patterns of English and sensitising them to social and geographical variations; and to equip them with skills for independent further work on their pronunciation of English" (Universiteit van Amsterdam, 2005). Vrije Universiteit Amsterdam requires knowledge of auditory-articulatory phonetics; knowledge of segmental and suprasegmental phonology of English (NRP: Non-Regional Pronunciation); knowledge of important English accents; improved pronunciation skills; the ability to make a phonemic transcription of spoken language (Vrije Universiteit Amsterdam, 2005). The aims of the Radboud University Nijmegen modules are most detailed; by the end of both the Oral Communication Skills – British/American English and the Phonetics British/American English modules, first-year students must have laid a basis for a near-native pronunciation of British or American English, which can be used as a model for others. In the course of the Phonetics British/American English modules second-year students must also acquire knowledge of the accentual and phonemic structure of the standard British English, c.q. the standard American English pronunciation, and of the variation between strong and weak forms of function words. Students must be able to transcribe a written text, including a list of words whose pronunciation is different from what would be expected on the basis of their spelling. Although the information provided in the on-line module guides cannot be detailed enough to provide a true picture of what is required from the students, it appears that the phonological competence is addressed as well as the orthoepic competence. Radboud University Nijmegen is unique in that it makes it clear that a near-native pronunciation will be expected; it is also unique in that it specifies that this must be a near-native pronunciation of either standard British or standard American

[16] *Each of the six levels of language proficiency of the CEFR can be subdivided as follows: A1.1, A1.2 etc. (Council of Europe, 2001, p. 32).*

English (Radboud Universiteit Nijmegen, 2005, pp. 16, 18). The other universities are much less strict in that they only require students to improve their pronunciation; they do not mention at what level this improvement should start and what level it should reach.

If there are considerable differences between the stated aims of the modules dealing with pronunciation in the universities, the differences between the aims of the modules offered by the teacher education courses that are provided by polytechnics in the Netherlands are even greater. Of only one of the modules offered by the Hogeschool Utrecht teacher education course e.g. does the description refer to pronunciation, one of the aims being: knowledge of the pronunciation of English and pronunciation problems (Hogeschool Utrecht, 2005). The English full-time course of the Noordelijke Hogeschool Leeuwarden, by contrast, offers a module Pronunciation and Phonetics, in which the aims for pronunciation are that the student can pronounce each sound of a list of three words correctly or almost correctly and that in a conversation and before a group the student has a natural and clear pronunciation and intonation without disturbing errors and without an annoying Dutch or Frisian accent. Although some readers may wonder if e.g. a German accent would be acceptable, it is clear that these aims address both the phonological and orthoepic competences in oral production and oral interaction. In the second-year pronunciation and phonetics course students must be able to read a text fluently and with correct intonation and read and speak in such a way that they can be understood even at the back of a large classroom and speak English with very little or no interference from their mother tongue. In these aims both intelligibility and lack of accent are addressed (Instituut Educatie & Communicatie, n.d., 2.1 p. 14, 2.2. p. 11).

In this far from exhaustive account of the attainment targets for the pronunciation of English in the Netherlands it is hard to find common denominators but it appears that an implicit or explicit distinction is made on the basis of the levels. For the two lower levels, which in the CEFR would be those of the basic user, intelligibility would be the sole goal. At the intermediate and higher levels of the independent user and the proficient user, however, learners would certainly be asked to "drop their foreign accent", although they would not often be required to get rid of all the traces of their accent; neither would they be required to replace their accents with the standard British one; the standard American accent would be equally acceptable; the same would probably be true of other standard accents.

2. JUDGING PRONUNCIATION

Whether the aims of learning and teaching pronunciation are to achieve intelligibility or some degree of nativelikeness, it is in the learners' interest that their achievements should at one time or other be judged, if only to inform them of the success or failure of their efforts. Although many informal judgements about pronunciation are no doubt made quickly, summarily and with little regard for the, at times serious, consequences, most of those who are professionally involved will probably have become aware of serious problems; these problems are associated with the judges,

the language learners, the tasks that must be elicited from these learners before their pronunciation can be judged and the judging process.

2.1 The choice of judges of pronunciation

In his research methodology textbook De Groot (1961, p. 173) warned that there is every reason to mistrust human observations and judgements and there is no reason to assume that his warning could be disregarded in the matter of pronunciation. One can only agree with Bongaerts (1999) when he identified the choice of judges as a central question in the study of the pronunciation skills of learners of a foreign language, the first choice being that between native and non-native judges, the second choice that between experienced and inexperienced judges and the last choice that between "competent" judges and those judges that are either too flexible or over-precise.

2.1.1 Native speaker judges or non-native speaker judges?

If there is no agreement on who the intended listeners will be, there must also be differences of opinion as to who are to judge the learners' achievements: native speakers or non-native speakers; if native speakers are to be chosen, should they be native speakers of the target variety or would native speakers of other varieties also be acceptable? There seems to be considerable divergence in practice; whereas e.g. students' oral performance in secondary-school examinations in the Netherlands is judged by the teachers, whether they are native speakers of the target language or not, the Netherlands Cambridge ESOL testing centre only employs native speakers of English, who are British, American, Australian, Scottish and Irish (personal communication British Council, 13 10 2005). Lynch and McNamara (1998) described a test for intending immigrants to Australia in which all the judges were native speakers. Kachru (1996) estimated that there were four non-native English speakers for each native English speaker. If this is true – and it might well be true in the Netherlands – it will not always be feasible to have native speaker judges in countries where English is a foreign language.

If non-native speaker judges are preferred, should they be speakers of the learner's L1 or speakers of another L1? On the one hand, Jenkins (2000) demonstrated that prolonged exposure to a certain type of interlanguage would improve understanding and that native speakers of the same L1 found each other's interlanguage much easier to understand than the interlanguages of speakers of another L1. Wang and Van Heuven (2003) conducted small-scale experiments in the mutual intelligibility of Chinese, Dutch and American speakers of English and found that Chinese listeners understood Chinese English best, whereas Dutch listeners understood Dutch English best and American listeners understood American English best. Wang and Van Heuven (2004) conducted an experiment on cross-linguistic confusion of vowels produced and perceived by Chinese, Dutch and American speakers of English and found that listeners tended "to identify those vowel tokens best that were produced by speakers of their own language background" (p. 214); interest-

ingly, they also found that the Chinese speakers, who had been exposed to Dutch accented English for some six months, had less difficulty identifying the vowels produced by the Dutch speakers than those produced by the American native speakers (p. 207). Might therefore the judgements by the non-native listener with the same L1 of the speaker's intelligibility not be too positive?

On the other hand, although such speakers' English might be more intelligible, that does not mean that the listeners with the same L1 were always more lenient than native speakers. On the contrary, most authors found non-native judges less positive than native judges. Fayer and Krasinski (1987) e.g. showed that Spanish speaking judges were harsher in their judgements of Spanish speakers of English than were native speakers of English; they warned that "a negative attitude towards the speaker of a particular variety will tend to decrease intelligibility in spite of the listener's familiarity with that variety" (p. 314) and felt that "non-natives are embarrassed by their compatriots' struggle with the non-native language" with pronunciation and hesitation as the most distracting features (p. 321). Similar findings were reported for Dutch versus English judges of English spoken by native speakers of Dutch by Koster and Koet (1993), who attributed the Dutch judges' negative reactions to their "undue fastidiousness" (p. 69) and by Brown (1995), who studied the judgements in a performance test of non-native Japanese speakers by both experienced and inexperienced non-native and native speakers of Japanese and found that "non-native speakers in addition to being less variable were substantially harsher on pronunciation". It looks as if increased understanding does not make for greater lenience and that familiarity breeds contempt.

Yoshida (2004), however, concluded on the basis of her data that native speaker judges and non-native speaker judges were equally capable of assessing Japanese students' intelligibility. But both her native speaker and non-native speaker judges were highly trained listeners who had had years of experience listening to the English produced by Japanese speakers. It is doubtful that her findings would apply to the majority of non-native teachers of English in the Netherlands. But even in a Japanese context her findings were not confirmed. Riney, Takagi and Inutsuka (2005) studied the perception of accent in native and non-native English by non-native Japanese listeners and native American listeners, whose judgements were subsequently checked by trained listeners; the differences were considerable; the non-native listeners appeared to be guided by prosodic features, amongst which fluency and rate of speech, whereas the native listeners appeared to be guided by segmental features.

The picture of the non-native speaker judges that emerged was one of not very competent, sometimes untrustworthy but also relatively harsh judges and it would not be surprising if learners were found to prefer native speaker judges; that learners did in fact prefer native speaker judges was shown by Lasagabaster and Sierra (2002), who asked students in the Basque region of Spain to indicate their preference for native or non-native teachers of English and found the clearest preference for native speaker teachers in the fields of the *assessment* of speaking and pronunciation, although they did not find such a preference in the field of the *teaching* of pronunciation (p. 137).

2.1.2 Experienced or inexperienced judges?

Is the implied listener an expert linguist or even a person who has undergone intensive phonetic training or might it not be more realistic for learners to be judged by non-expert listeners who have not studied linguistics and are not familiar with their interlanguages? Schoonen (1991) asked Dutch expert and lay judges to rate writing assignments of Dutch primary-school children and found that the lay judges were often stricter than the expert judges, who, moreover, were sometimes more reliable; Van As, Koopmans-van Beinum, Pols and Hilgers (2003) studied the perceptual evaluation of speech by naïve and experienced judges and found that the naïve judges were more negative about deviant speech than the experienced judges but believed their reliability to be high enough "to support the use of naïve listeners in perceptual studies, depending on the goal of the evaluations" (p. 957). However, these two studies concerned the L1.

Cunningham-Anderson (1996) reported that teachers of Swedish as a second language were more lenient as regards the importance of non-native speakers' errors in Swedish than were naïve secondary school students. Wesdorp (1978) cited a number of studies about the training of judges and called the results of such studies rather disappointing (p. 190). Lumley and McNamara (1995) studied the behaviour of raters during two training sessions aimed at removing great differences in rater severity and during the actual judging of an occupational English test and found that the training had not eliminated the differences, which returned after some time. Bonk and Ockey (2003) reported that experienced judges were more consistent but did not mention that they were stricter or more lenient. Brown (1995) found that experienced judges were more concerned about grammatical and lexical errors, whereas inexperienced judges took a dimmer view of pronunciation errors. Cheng and Warren (2005) asked engineering students, native speakers of Cantonese, to assess each others' English language proficiency; they found that their assessments of oral skills in seminars and writing skills in written reports differed significantly from the teachers' assessments but that there were no significant differences as far as the assessments of language proficiency in oral presentations were concerned and concluded that their students' assessments "did not reliably supplement their teachers' marks" (p. 110). Bongaerts, Van Summeren, Planken and Schils (1997), however, studied thirteen British judges, six of whom were or had been teachers of English as a foreign language and seven of whom had not been trained in languages or linguistics, and found no significant differences in strictness. Bongaerts (1999) studied three groups of judges: experienced and inexperienced native speakers of English and native speakers of Dutch with considerable experience in teaching the phonetics and pronunciation of English; he found no significant differences between the native speakers of English in their ability to identify passages spoken by non-native speakers, whereas the non-native listeners were less successful. It would therefore appear that the evidence in favour of experienced judges is not conclusive and that this matter requires further study.

2.1.3 Younger or older judges?

It is possible that for some speakers the intended listeners would be persons of their own age group; therefore it might be argued that it is preferable for adolescent speakers to be judged by adolescent listeners and for adults to be judged by adults. Albrechtsen, Henriksen and Færch (1980) asked native speakers of English from London, Leeds and Edinburgh, the majority of whom were students in secondary schools (sixth formers) and a minority of whom were adults, to judge interviews of Danish secondary-school students with a native speaker of British English; they found no differences in the way the judges from the different regions ranked the speakers but did find differences between the rankings by younger judges and those by older judges; unfortunately, they did not indicate if there were differences in strictness. Since they indicated that the adult listeners had been selected from groups such as hotel staff that would be likely to be in contact with Danish visitors and since these listeners may therefore well have had experience in dealing with Danish speakers' pronunciation of English, it is possible that the differences in ranking must be ascribed to greater experience rather than age [17].

2.1.4 Lenient or strict judges?

Some judges may be disposed to leniency and other judges may be disposed to strictness. Bongaerts (1999) found that some judges (experienced as well as inexperienced) were competent, in being able to successfully identify native pronunciation; other judges, however, were too lenient in accepting non-native speech as native, whereas yet other judges were too strict in identifying native speech as non-native speech; he therefore recommended careful screening of candidates so that only competent judges are selected.

2.1.5 Male or female judges?

Studies of language attitudes in the native language context have revealed substantial differences between men and women. Van Herpt and Fagel (1981) and Boves, Fagel and Van Herpt (1982) e.g. asked male and female speakers of Dutch to indicate on 35 scales what, in their opinion, ideal male and ideal female speech were; they found that male judges wanted female speakers to have more pleasant and lively voices than female judges did. It also appeared that male judges were less strict as far as the broadness of ideal speech was concerned. On nearly all the scales they found that female judges were more demanding in requiring a more extreme score for speech to be judged ideal, whereas male judges were less strict, in particu-

---

[17] *Van den Doel (2006) asked native speakers of English to listen to a number of sentences in which the speaker had introduced a pronunciation error and found that older listeners did not only detect fewer errors than younger listeners, which lower detection rate might be ascribed to reduced auditory sensitivity, but were also more lenient about the seriousness of the errors they did detect (p. 241). It would appear that learners who are judged by older judges enjoy an advantage over learners who are judged by younger judges.*

lar with male speakers, thus revealing an interaction effect between listeners' and speakers' sex. Van Bezooijen and Van den Berg (2001), however, found that young women were more positive than old women and young and old men towards a new variety of Dutch that the other groups did not regard as standard; the reason for the younger women's greater lenience might be that they were more sensitive to the emergence of a new standard than the other groups. Broeders (1981) asked male and female Dutch university students to listen to seven speakers, six of them with different native accents of English and one with an L1 (Dutch) accent, who all read the same text. They found that the female Dutch listeners were more positive about RP than were the male listeners (p. 130); both male and female listeners were more positive about RP than about GA and about GA than about L1 accented English; curiously, they were more positive about the L1 accented English than about the other native varieties, which they presumably felt to be less prestigious. Van der Haagen (1998), however, found no significant differences between her male and female listeners. Schairer (1992, p. 311) reported that in their judgements of non-native speech female native speakers of Spanish were "found to be more strict than their male counterparts in the evaluation of comprehensibility, particularly at the lower performance levels". In view of the different findings, it would appear that there is not enough evidence to recommend either female or male judges and that further study would be needed.

2.1.6 Regional origin and socio-economic background of judges

Both in the English speaking and the Dutch speaking world there are considerable differences related to regional origin, socio-economic background and a combination of the two. In his comprehensive survey of accents of English, Wells (1982) discussed some eighteen varieties of English found not only in the UK but also e.g. in the US, the West Indies and the Southern hemisphere. Donaldson (1983) distinguished as many varieties of Dutch found in the Netherlands and Belgium. In view of this great diversity, it would stand to reason that there are also differences between judges with different regional and/or socio-economic background, whether they listen to their native language as an L1 or as an L2. Curiously, such differences have not always been found; Albrechtsen, Henriksen and Færch (1980) found no differences between listeners from London, Leeds and Edinburgh, the local accents in which cities are radically different (Wells 1982, pp. 301-332, 349-374, 393-415); this may have been due to the fact that, although the listeners were residents of these cities, they had not lived there for long periods; no information about length of residence or socio-economic background was provided. But also judges who came from the West and the East of the Netherlands were shown not to be significantly different in their judgements of an advanced variety of standard Dutch by Van Bezooijen and Van den Berg (2001), who moreover found no differences between judges with different socio-economic backgrounds. Nevertheless the possibility that listeners from countries that are as far apart as the UK and the US would differ in their reac-

tions to the pronunciation of non-native speakers seems one that deserves further investigation[18].

2.2 Judges' complaints

After his warning about the untrustworthiness of human observations and judgements, De Groot (1961, pp. 239-244) identified five phenomena that may have an undesirable effect on judgement[19]. Judges may e.g. be influenced by the signification effect; they may not have a clear picture of what is required of them and may therefore change this picture in the course of the judging process. The halo effect may e.g. cause judges to arrive at a positive judgement of aspects of pronunciation because the speaker is highly fluent or at a negative judgement because the speaker is hesitant. Other effects that De Groot (1961) warned against are the sequence effect, norm shift and a personal tendency for some judges to be either strict or lenient, an effect that was clearly present in some of the judges in Bongaerts (1999). Finally, the contamination effect may e.g. cause judges to use the judgement of the learner's pronunciation for other purposes, such as the punishment of a recalcitrant pupil or the reward of a hard-working student.

Native speaker judges are known to be susceptible to the halo effect while judging native speakers' speech as well as non-native speakers' speech. There is considerable evidence that these judges use features of speech to arrive not only at judgements on the speakers' voice and pronunciation but also at judgements of the speakers' personality. Trudgill (1975) and Lippi-Green (1997) e.g. provided convincing evidence for the existence of this phenomenon in speakers of British resp. American English. Van Bezooijen (1988) studied the relative importance of pronunciation, prosody and voice quality and found suggestions that native speakers of Dutch needed only very short fragments of standard and regional varieties of Dutch to describe the speakers as more or as less statusful or attractive, using prosodic clues to decide on the strength of the speaker's personality and segmental clues to decide on the speaker's "intellectual qualities and socio-economic status" (p. 101). Van Bezooijen (1994) studied the aesthetic evaluation of Dutch language varieties and asked 7-year old, 10 year-old and adult native speakers of Dutch to listen to 20-second fragments of standard, non-standard urban and two varieties of non-standard regional Dutch; the 7-year olds were able to distinguish between standard and non-standard varieties, whereas the 10-year olds were able to locate most of the speakers

---

[18] *Van den Doel (2006) offered a GA version of the sentences with typically Dutch errors to North American listeners and an RP version to listeners from other English-speaking countries. He found that "there was no significant difference in severity between those judges taking the RP version of the experiment as against those taking the GA form" (p. 107), although listeners to the RP version were harsher on some errors, where their greater strictness could be attributed to a higher detection rate (p. 150).*

[19] *See also Wesdorp (1978), who discussed these phenomena as sources of unreliability in the evaluation of writing in the L1 (pp. 9-20) and Meuffels (1994), who distinguished no fewer than 8 factors that can affect the objectiveness of a judgement: signification effect, halo effect, contamination in the wider sense of the word, contamination in the narrower sense of the word, sequence effect, norm shift, personal comparison and competence effect (p. 5).*

and the adults all of the speakers geographically; the standard accent received the most positive evaluation, the regional accents the least positive, the non-standard urban accent occupying an intermediate position.

Native speakers can also arrive at judgements of non-native speakers' speech quickly. Lindemann (2003) found that undergraduates in the US rated Korean speakers of English negatively as far as the speakers' status was concerned; although they were seldom able to identify the accents as Korean, they did identify the speakers as belonging to "some stigmatised non-native group" (p. 358). Doeleman (1998) concluded from her study of native Dutch secondary-school children's reactions to short fragments of non-native speech that these students were led by both the perception of phonetic speech features and the identification of ethnic group membership but that "the respondents based their evaluation of nonnative speakers primarily on their stereotypical views on the nonnative groups from which the speakers originated, especially on the social distance they felt between the ethnic group and themselves" (p. 277); ironically, the identification of the non-native group was not always correct, so that judgements were at times based on misidentification.

Of the five undesirable effects mentioned the halo effect seems to be the one that is a particularly great risk to non-native speaker judges; listeners who find it hard to perceive the difference between incorrect and correct realisations of certain segments (e.g. /r/ and /l/) may well allow themselves to be guided by features that they can easily perceive (such as intonation or hesitation). If they are native speakers of the same L1 as the persons to be judged, non-native speaker judges are even more at risk. There is considerable evidence that language learners will at first perceive the sounds of the L2 in terms of the categories of the L1 (see e.g. Flege, 1996) and the ability to perceive the sounds of the L2 in terms of the categories of that language may well be what distinguishes the successful language learner from the less successful one. It is not unlikely that less successful language learners who are asked to judge the L2 pronunciation of speakers of their own L1 will perceive this pronunciation in terms of the categories of the L1 rather than those of the L2; it is not unthinkable that even successful learners may, from time to time, judge aspects of the L2 pronunciation as if they were aspects of the L1. Such judges might well allow the perceived quality of a speaker's L1 to influence their judgements of the speaker's L2. It seems that little attention has so far been paid to this effect.

2.2.1 Speakers' sex

Several studies did not specify if the speakers were male or female. Some studies such as Andersen-Hsieh, Johnson and Koehler (1992) concentrated on male speakers. Doeleman (1998, p. 80) justified her choice because "male voices are preferred to female voices in the manipulation of speech signals". Yoshida's (2004) study, by contrast, only used female speakers as did Sparks, Artzer, Ganschow, Siebenhar, Plageman and Patton (1998) in their first study; this may have contributed to the fact that the results of their second study, which used both male and female speakers, differed from those of the first study. Of Cheng and Warren's (2005, p. 97) Cantonese students 49 were male and only two female. The question may well be raised if

findings based on single sex or sexually unbalanced groups of speakers can be generalized to groups made up of approximately equal numbers of both sexes.

Several authors studied the differences between Dutch men's and women's pronunciation (e.g. Boves, et al., 1982). More recently Tielen (1992) found, perhaps not surprisingly, that women speak with a higher pitch and more melodiously than men. Sulter and Peters (1996), who could not find significant differences between the genders, found an interesting interaction between training and gender in that the speech of untrained women was felt to be shriller, uglier, breathier and broader than the speech of untrained men, but that the speech of trained women was felt to be less broad and less slovenly than that of trained men. Van Donzel and Koopmans-van Beinum (1997) mentioned that the prosodic characteristics of the speech of all four female Dutch speakers studied was perceived as better than those of the speech of all four male speakers. Van Bezooijen, Kroezen and Van den Berg (2002) studied the realisation of the consonant /r/ and Van Heuven, Edelman and Van Bezooijen (2002) studied the realisation of the diphthong /εi/ in standard Dutch and found a tendency for women to use the innovative front approximant variant more often than men resp. a significantly greater occurrence of a more open realisation of the diphthong, which they described as more "advanced". Van der Haagen (1998) found – to her own surprise – that the male Dutch speakers did not produce significantly more GA forms, which in the context of her study were less prestigious than RP forms, than the female Dutch speakers did. Major (2004), however, found that female Japanese speakers of English produced a greater number of prestigious forms than male Japanese speakers of English as did female Spanish speakers of English compared to male Spanish speakers, a similar difference being found between male and female native speakers (p. 180). It would therefore appear that the jury is still out on this issue and that it would be sensible to discriminate between male and female speakers.

2.2.2 Speakers' L1 and variety thereof

Many studies have addressed the relation between the speakers' ability to achieve an intelligible or native-like pronunciation and their L1, often through a contrastive analysis of the sound system of the speakers' L1 and the L2 or through a study of the interlanguage (cf. Corder, 1981, who related the linguistic distance between the L1 and the L2 to the magnitude of the learner's task; for a survey see Ioup & Weinberger, 1987)[20]. Several authors have been aware of the relation between the different varieties of German and the ability to acquire a native-like accent. While discussing the vowels of English, Jones (1960) e.g. did not only give advice to speakers of such languages as French and German but also distinguished speakers from certain regions within Germany, e.g. when he mentioned "Saxons, Bavarians

[20] *A possible measure of the "linguistic distance" between English and other languages was tentatively proposed in a discussion paper; interestingly, Dutch was estimated to be among the languages with the smallest "linguistic distance" from English (Chiswick & Miller, 2004); it would probably be hard to prove that the "linguistic distance" between English and standard Dutch is smaller than that between English and Amsterdam Dutch.*

and some from the extreme North (Hamburg, Lübeck, etc.) who often use" a vowel like in *far* "similar to the English one" (p. 76). Barry (1974) asked native speakers of English and German to pronounce the German-English pairs *Tip – tip*, *Paß – pus* and *Busch – bush* and found that speakers from Hamburg produced aspiration that was perceived as native-like by English listeners, whereas speakers from Cologne produced aspiration that was felt to be too short. Karpf, Kettemann and Viereck (1980) studied the phonological interference of the Styrian dialect of German in English language learning and James (1983) studied the English of speakers of the Swabian variety of German.

Textbooks about the pronunciation of English for speakers of Dutch have tended to address speakers of the standard variety of Dutch – ABN (general cultured Dutch) or AN (general Dutch) – although there are references to differences within the L1, leading to different pronunciations of some sounds of the L2. Kruisinga (1960, p. 35) believed that "many Dutch people especially in the Western part of Holland pronounce a uvular r … and that for these it is often more difficult to learn the English sound…". Gussenhoven and Broeders (1976) mentioned some difficulties experienced by speakers from the South of the Netherlands. Collins and Mees (1981) larded their descriptions of the vowels and consonants of English with observations not only for speakers of the standard accent of Dutch but also for speakers of the town dialects of the Hague, Rotterdam, Amsterdam and Antwerp, the dialects of Noord-Brabant and Limburg, as well as for speakers of affected varieties of ABN (e.g. pp. 86, 87, 97, 101, 102, 125). Gussenhoven and Broeders (1997) pointed to differences within standard Dutch but also to differences between standard Dutch and the Dutch found in e.g. the North, the East and the South of the Netherlands and in Belgium (pp. 69, 72-73) and warned that such differences may lead to problems in the pronunciation of English e.g. with speakers in the South-East of the Netherlands (e.g. pp. 96, 98, 100). The advice offered by these textbooks is no doubt useful, although advice targeted at speakers of one variety or a group of similar varieties may have been more useful. It is not clear, however, if the advice was based on systematic study of the type of English produced by speakers of particular varieties of Dutch or on anecdotal evidence.

2.3 Instrumentation

2.3.1 Speakers' tasks

The tasks that learners are asked to perform for their pronunciation to be judged – or to demonstrate their competences – can probably vary in the same way as the activities in learning pronunciation. Pennington (1996, p. 225) distinguished five dimensions of meaningfulness in the activities that teachers may ask their students to perform: mechanical (e.g. repetition of minimal pairs), contextualised (e.g. repetition of key words), meaningful (choice of correct word in a sentence), realistic (a role play) and real (e.g. discussion of the students real life situation).

There are studies that used mechanical tasks, but these tasks were often used to determine listeners' ability to perceive sounds rather than the speakers' ability to produce them. Heeren (2004) e.g. studied Dutch listeners' ability to discriminate

between the initial consonants in the minimal pair of nonsense words *thif – sif*; Wang and Van Heuven (2004) studied the ability to perceive the difference between the vowels in identical /hVd/ contexts[21].

It is understandable that many authors favoured more realistic tests such as group oral discussions, performance tests or even occupation specific performance tests for the testing of proficiency in general (e.g. Brown, 1995, Bonk & Ockey, 2003). In such tests pronunciation is often not assessed as a separate skill but as part of the more general speaking skill or of spoken production and spoken interaction. If the need is felt to test pronunciation as a separate skill, however, more controlled tasks seem to be preferred. Although truly spontaneous speech would be the activity in which speakers can show their competence in spoken interaction, it may be a serious drawback that in conversations with native speakers the non-native speakers would be likely to accommodate their pronunciation to that of their interlocutors; a similar convergence would be found in the conversation between non-native speakers, as was shown by Jenkins (2000). For Van der Haagen (1998) this was an argument for using controlled free speech (p. 19); even though it could be argued that the speaker's ability to accommodate the pronunciation to the interlocutor's is part of the learner's competence, her decision not to use spontaneous speech seemed sensible in a study of pronunciation preferences.

The tasks found most frequently are the word-list style, reading passage style and controlled free speech. A problem with these three tasks concerns non-native speakers' ability – or inability – to introduce the appropriate stylistic variation between these tasks, the word-style list requiring the most formal and careful style, free speech requiring the least formal style. Major (2004) studied gender and stylistic differences in native speakers' and Japanese and Spanish speakers' English and found that the Japanese speakers of English showed no stylistic differences and the Spanish speakers showed few differences, whereas the native speakers showed significant differences (p. 180); interestingly, some Japanese speakers used more informal features in the word-list style task than in the other tasks. Derwing, Munro and Wiebe (1998) asked native speakers of English to evaluate the effects of three types of pronunciation instruction for learners of English as a second language; the learners were first asked to read out sentences and then extemporaneously produce narratives; the researchers found that there were not only effects of instruction but also of task and suggested that the listeners might have been better able to notice differences in accentedness in the sentences than in the narratives (p. 406).

As far as the reading passage style test is concerned, more learners may see themselves performing this kind of task in real life than the word list. A lot, however, would depend on the type of reading passage; reading out a story or an argumentative text would seem to make more sense than for one person to read out the two parts of a dialogue (e.g. Yoshida, 2004). An advantage of the reading-style test might be that it is easier to judge; Doeleman (1998) asked native speakers of Dutch

---

[21] *Although their claim that the /h_d/ consonant frame is fully productive in English is no doubt correct, it is doubtful if the listeners had ever heard such words as hoyed, how'd, hered and hoored, which unfamiliarity may have influenced their ability to perceive them (Wang & Van Heuven, 2004. p. 206).*

to listen to texts read by and semi-spontaneous speech produced by native and non-native speakers of Dutch and found that the judges were better able to distinguish native and non-native speakers and were more successful in identifying the countries of origin on the basis of the read texts (pp. 192-193).

One would imagine that most learners would look forward to situations in which they would engage in conversations with speakers of English, fewer learners would probably see themselves delivering monologues and very few would see themselves reading out lists of words. Many studies addressed the relevant competences in spoken production rather than in spoken interaction and one may well wonder if the findings would have been similar if they had addressed both situations. Also, certain tasks may be more difficult for some learners than other tasks, so that "task effects" may occur as well as effects of interaction between learner and task, as Schoonen (1991, p. 27) pointed out in his study of the evaluation of assessment of writing. It would therefore appear worthwhile to have another look at the differences between the speakers' tasks.

2.3.2 Judges' tasks

Having identified five complaints that judges may be subject to and that may cloud their judgement, De Groot (1961, pp. 244-250) prescribed five remedies, two of which seem to be of particular importance in the judging of pronunciation. Reduction of the judges' task through clear definition of the aspects that are to be judged was recommended as a remedy against the signification effect as were elimination of irrelevant aspects and concentration on relevant aspects as remedies against the halo effect; ideally, judges should only judge one aspect of a speaker's task and then move on to the same aspect of other speakers' tasks. In informal and traditional judgement of pronunciation the two remedies are not applied, since such holistic judgements may not even distinguish between the segmental, suprasegmental and/or paralinguistic features. But some tasks invite judges to demonstrate a precision that is only on the surface, as Wesdorp (1981) pointed out about analytical schemes with many, overlapping subcategories, which as a result of the halo effect may lead to repetitions of previously made judgements.

Many authors reported on analytical studies in which particular features were singled out. Schoonen (1991), citing De Groot (1961) presented strong arguments against global or holistic judgements and in favour of analytical judgements of writing assignments. Several authors went further and developed atomistic pronunciation tests. Van Weeren and Theunissen (1987) developed atomistic tests of Dutch learners' pronunciation of French and German, containing some 40 features each that were known to be problematic for Dutch learners, which tests they claimed to be much more accurate than traditional holistic tests. Few objections can be raised against a test with so many distinct features but sometimes one may well wonder if extremely atomistic approaches can provide valid judgements of the speakers' pronunciation. Flege (1996) e.g. asked Dutch and British speakers to read the English words *heat*, *hit*, *hot*, *hoot*, *hut* and *hat*; the consonants were removed and British and American listeners were asked to identify the vowels; he himself admitted that this

procedure was "likely to decrease the overall rate of correct identifications" (p. 26). Edelenbos, Van der Schoot and Verstralen (2000) e.g. reported on the pronunciation of primary school children; the children had been asked to read a text consisting of seven sentences, in which two sounds were underlined that were relevant for the assessment of the pronunciation; two native speakers judged whether the phonemes were correct or incorrect; there were no standards for pronunciation.

Varonis and Gass (1982) merely asked their judges to choose between "good" and "bad" but most studies have preferred scales that do not force judges to make such binary choices. Anderson-Hsieh et al. (1992) combined what they termed impressionistic scales with scales that were produced by counting instances of deviance in segmentals, prosody and syllable structure; unfortunately the scale that was used for determining prosodic deviance was the "least native-like – most native-like" scale, which may well be regarded as an impressionistic scale, and the choice of this scale may have contributed to their conclusion that the relation between prosodic deviance and judgements of non-nativeness was stronger than that between segmental deviance and judgements of non-nativeness. This is not what was found more recently by Riney et al. (2005), who reported that non-native listeners' judgements were more strongly influenced by prosodic features, whereas native listeners' judgements were more strongly influenced by segmental features, notably the pronunciation of the consonants /r/ and /l/. Schairer (1992, p. 316) concluded that "the most accurate and economical way of predicting native speaker evaluation of non-native comprehensibility is to focus on the degree of nativeness or non-nativeness of vowels in stressed and unstressed position", but her listeners were native speakers of Spanish. Koster and Koet (1993) reported that native listeners of English were more concerned about the pronunciation of consonants, whereas Dutch non-native listeners paid more attention to the pronunciation of vowels, which may suggest that it is not a matter of segmental versus prosodic but rather one of consonants versus vowels that distinguishes native from non-native judges. Cucchiarini, Strik and Boves (2000) compared the way speech produced by native and non-native speakers of Dutch was rated automatically with the ratings of three groups of native speaker experts and found that "the highest correlations are found between overall pronunciation and segmental quality, which suggest that when the raters judge overall pronunciation, they are most influenced by the quality of the segments uttered by the speaker" (p. 115); they also mentioned that their findings "warrant the use of overall ratings of pronunciation" rather than specific – temporal and segmental – ratings "as sole reference for automatic scores" (p. 118). It would therefore appear that it is highly important that judges are given tasks that are neither too global nor too atomistic.

2.3.3 Perceived quality of pronunciation and personality

In view of the great variety of instruments used to find out how pronunciation and speech are perceived, it is not surprising that there are considerable differences in the dimensions found. Blom and Van Herpt (1976) studied no fewer than sixty seven-point scales used to describe the quality of voice and pronunciation of speak-

ers of Dutch; they extracted three dimensions – voice appreciation, articulation quality and abnormality – and suggested that just fourteen scales would be sufficient. Fagel (1981) made a selection of 38 scales and extracted one very important dimension – evaluation – and a second major dimension – articulation quality; in addition he extracted two minor dimensions: voice quality and pitch level. Fagel, Van Herpt and Boves (1983) analysed the perceptual qualities of Dutch speakers' voice and pronunciation and found that there were five dimensions, which they interpreted as melodiousness, articulation quality, voice quality, pitch level and speaking rate, to which they added two additional dimensions – evaluation and potency – on the basis of which they proposed 14 scales to be used in future research (p. 325). The proposed scales or a selection from the scales were subsequently used in studies of the voice and pronunciation of native speakers of Dutch several times (see e.g. Sulter and Peters, 1996, and Van Donzel and Koopmans-van Beinum, 1997, who found the dimensions voice appreciation, dynamic and articulation quality, and Van As et al., 2003) and also of speakers of English (e.g. Van der Haagen, 1998, who found the dimensions status, dynamism, affect and norm).

In several studies scales for the description of voice and pronunciation were combined with scales for the description of personality. Van Bezooijen (1988) e.g. used 23 voice scales and extracted eight factors: "voice quality, pitch variation, accent, loudness variation/tempo variation, pitch, loudness, articulation and tempo" (pp. 95-96), which she then related to 16 scales for the attribution of personality characteristics. Knops (1988) used seven speech scales and seven personality scales and found that "speech scales referring to norms for good or correct articulation correlate highly with personality scales referring to speakers' social and intellectual status" whereas "speech scales displaying raters' concern with the naturalness and aesthetic qualities of speech correlate highly with personality scales relating to the social attractiveness of speakers" (p. 119). Nine "sociolinguistic" scales were used to study reactions to non-native speakers' Dutch by Doeleman (1998), who added four extra "social-psychological" scales to arrive at three dimensions: status, attractiveness and social distance (p. 237).

Mulac, Hanley and Prigge (1974) asked American listeners, experienced as well as inexperienced, male as well as female, to rate the speech of native and non-native speakers of English on no fewer than fifty seven-point scales and found three dimensions: socio-intellectual status, aesthetic quality and dynamism (p. 413). Albrechtsen et al. (1980) identified four factors: personality, content, language and comprehensibility. In earlier work on linguistic attitudes Cunningham-Andersson (1995, p. 136) found the need for two dimensions "corresponding to the status of the speaker and solidarity felt by the listener with the speaker"; she herself only measured friendliness and intelligence.

Although some dimensions seem to emerge more frequently than others it is by no means so that there are clear common denominators. It would therefore seem unwise to make any a priori assumptions about the dimensions that might appear in future studies.

3. QUESTIONS IN NEED OF FURTHER STUDY

Previous studies of the pronunciation of English by non-native speakers and the judging of pronunciation have provided many answers to many questions, but unfortunately the answers to each question have seldom been identical. Also in the context of the Netherlands questions may still be raised about the judging of the pronunciation of English. These questions concern the judges, the speakers and their tasks.

As far as the judges are concerned, in view of the large number of non-native teachers in the Netherlands it would be in the Dutch learners' interest to know if non-native judges are stricter or more lenient than native speaker judges, and if these non-native judges arrive at the same ranking as the native judges. It would also be important for them to know if experienced judges are stricter than inexperienced judges. Finally, it would be in the learners' interest to know if female judges are stricter than male judges.

Questions concerning the speakers could focus on the different types of English that may be produced by speakers of other varieties of Dutch than the standard one. It would certainly benefit the Dutch learners if their teachers knew more about these different types and could adapt their teaching accordingly.

In addition to the questions about the listeners and the speakers questions may be raised about the interaction between judges and speakers. There are two questions to which the answers would profit the learners: the question if judges are more lenient towards speakers of the same sex than to speakers of the opposite sex and the question if judges are more lenient towards speakers of the standard variety of Dutch than towards speakers of other varieties.

As far as the speakers' tasks are concerned it is important for learners to know if judgements about language activities that demonstrate the phonological competence – e.g. taking part in a conversation and delivering a monologue – differ from those about language activities that demonstrate the orthoepic competence – e.g. reading aloud and delivering a paper.

These provisional questions will be worked out in the next chapter, Perceived pronunciation quality in mother tongue and foreign language. A question about the speakers will be dealt with in the chapter about Amsterdam English. Several questions about the judges as well as the interaction between speakers and judges will be addressed in the chapter A choice of judges. The chapter Judges on Dutch and Dutch English will discuss more aspects of the interaction between speakers and judges.

PERCEIVED PRONUNCIATION QUALITY IN MOTHER TONGUE AND FOREIGN LANGUAGE

1. INTRODUCTION

When I was an inexperienced teacher in a secondary school in Amsterdam, I was warned by a senior colleague that boys who spoke Dutch with a broad Amsterdam accent would never acquire a good pronunciation of English. When I was an inexperienced teacher in the South of the Netherlands I was told in forceful terms that "those Limburgers" were never going to learn how to pronounce English properly. When I had returned to Amsterdam I was confronted with similar statements about speakers of a regional North-Holland variety of Dutch.

Of course it never occurred to me to question the veracity of these statements, let alone to enquire how my elder colleagues had arrived at their insights. But on reflection it seemed to me that such statements were indicative of a view that might well be held by other foreign-language teachers in the Netherlands, namely that those learners who did not speak the standard variety of the Dutch language – whether this was the native language or not – were bound to experience more difficulties in speaking a foreign language than those who did and that learners who spoke Dutch with a broad accent were therefore less likely to acquire a native-like accent in a foreign language. Suggestions that speakers of other varieties than the standard variety of Dutch might experience greater difficulties in learning the vowels of English than speakers of the standard variety were found in Gussenhoven and Broeders (1976, pp. 71-73). If this were true there might well be a correlation between the ranking of Dutch speakers in terms of their pronunciation of English and their ranking in terms of their pronunciation of Dutch. On the other hand, if no such correlation were found there would be every reason for doubt. Therefore a preliminary research question was formulated: is there a correlation between judgements on the pronunciation of English of speakers of Dutch and judgements on their pronunciation of Dutch? To answer this question judgements on the pronunciation of Dutch were compared with judgements on the same speakers' pronunciation of English.

2. METHOD

2.1 The listeners

The listeners to English speech were two native speakers of Dutch, one of them male and one of them female, and one male native speaker of English. The listeners were experienced teachers of English as a foreign language, who moreover had

This chapter was based on a paper presented at the ANELA conference in November 1980. A Dutch version is found in Koet (1981a).

taken part in two training sessions of one day each for teachers of English in the Netherlands engaged in the assessment of fluency and pronunciation.

The listener to Dutch speech was a male native speaker of Dutch (the author), who was familiar with the varieties of Dutch but had received no training in the assessment of voice and pronunciation in Dutch.

2.2 The speakers

The speakers (n = 50) were first-year students of English who had started their full-time teacher-education course at an Amsterdam teacher-education institute[22]. They were working towards a degree that would allow them to teach in lower secondary education. At the time approximately one hundred full-time students had entered the institute to receive their training as teachers of English and one other subject. There was considerable diversity as far as geographical origin and educational background were concerned. Most of these students came to the institute after they had finished one of the more academic types of secondary school; others had had a career in industry or had studied at other institutes for tertiary education; they had had at least five years of formal instruction in English. Most first-year students were residents of the province of North Holland; they were frequently from Amsterdam or the areas in the northern part of the province. Because of its central location, however, the institute had also attracted several students from other provinces in the Netherlands such as Limburg. There were also a dozen non-native speakers of Dutch, mostly from Surinam, the Antilles and Portugal; all of them were proficient speakers of Dutch. The majority of these students was female. No distinction was made between male and female speakers and between native and non-native speakers of Dutch.

2.3 The speaking tasks

There were three tasks for the speakers: one Dutch task and two English tasks. As far as the Dutch speaking task was concerned, it was assumed that the students of English would find it a little peculiar to have their pronunciation of Dutch, of which most of them were native speakers, judged, a procedure that students of Dutch or of speech therapy would have been more familiar with (see e.g. Blom &Van Herpt, 1976, and Fagel 1980, 1981). Also care needed to be taken that the judgement of pronunciation would not – as in daily life – be too much obscured by a number of non-phonological matters, such as vocal quality, vocabulary, syntax, facial expression and outward appearance. An attempt was therefore made to exclude non-auditory factors as much as possible by asking the students to produce semi-spontaneous speech, a simple narrative suggested by twelve pictures (Land, 1981). This production took place during the regular oral-skills classes in the first half of the academic year. In order to obtain as natural a pronunciation of Dutch as possible, students were told that the Dutch version was merely an exercise to prepare them for a subsequent English version, which they were obliged to produce for their oral-

---

[22] *I am grateful to staff and students of the English department of d'Witte Leli.*

skills tutor. They were asked to record their renderings of the story on audiocassettes in the language laboratory, an Electron SLA C4; none of the recordings was longer than a few minutes. Students' names on the cassettes were replaced with numbers.

The first English speaking task consisted of semi-spontaneous speech, also a simple narrative suggested by twelve pictures. Like the Dutch fragments the English fragments were recorded on audiocassettes in the same language laboratory. This production took place during the regular oral skills classes in the second half of the academic year. Students' names on the cassettes were not replaced with numbers.

The second English speaking task consisted of a twenty-minute interview at the end of the academic year. For this interview first-year students had been asked to prepare two topics and read two books, to be discussed with two tutors, one of whom conducted the interview, the other taking notes.

2.4 Judgements on the pronunciation of Dutch and English

In judging the pronunciation of Dutch an impressionist rather than an analytical method was used. To prevent too much distraction by non-phonological factors a list of points to which attention should be paid was used, partly based on such works on the phonetics and phonology of Dutch as Cohen, Ebeling, Fokkema and Van Holk (1961), Hermkens (1969) and Van den Berg (1978). This list reflected the North Holland provenance of the majority of the participants and was based on an earlier informal study of students' Dutch; it pretended in no way to give criteria for the assessment of the pronunciation of Dutch speakers in general.

As far as the consonants were concerned, particular attention was paid to the realisation of /s/ – /z/, /f/ – /v/ as in *c - zee* (c – sea) and *fee - vee* (fairy – cattle) and of /s/ – /ʃ/ as in *Does - douche* (shower).

With the vowels notice was taken of the realisation of the diphthongs in words like *ei* and *ui* (egg – onion) and the monophthongs found in words like *zee* and *zo* (sea – so). The former had frequently been found to be pronounced with very little diphthongisation – almost as if they were monophthongs – and the latter had been found to be pronounced with strong diphthongisation. Another point was the realisation of the so-called "long" <a> as found in *haak* (hook), which had often been rendered as a lengthened version of the "short" <a> in *hak* (heel) and which, in addition, was sometimes pronounced with lip rounding.

Nasalisation was also attended to; in many cases there had been strong nasalization of vowels preceding nasals as in *mensen* (people), where the preconsonantal nasal sometimes disappeared.

The pronunciation of loanwords, like *stewardess*, was studied closely and in particular the pronunciation of unstressed syllables in such loanwords as *chauffeur*.

Besides these segmental matters attention was also paid to such prosodic matters as intonation in the widest sense of the word and durational aspects.

Also in judging the pronunciation of English, an impressionist method was used. No guidelines had been provided by the institute for the judgements on the pronunciation of English.

2.5 Procedure

The judgements on the pronunciation of Dutch were entered on a 1 to 10 scale, 10 representing the opinion that the pronunciation was no doubt Standard Dutch ("ABN" – short for "Algemeen Beschaafd Nederlands" i.e. General Civilised Dutch or "AN" – short for "Algemeen Nederlands" i.e. General Dutch) and 1 representing the opinion that the pronunciation was neither "general" nor "civilised" and perhaps not even Dutch. Of course, these marks were no attempts to give an absolute value but merely attempts to arrive at a ranking within the group.

The judgements on the pronunciation of English on each of the recordings were given by one of the three listeners, who entered his or her judgements on a 1 to 10 scale after having listened to the recordings at any time that was convenient.

The judgements on the pronunciation of English in the interview test were entered on a 1 to 10 scale by each of the two examiners independently, immediately after the test.

Although some months had elapsed between the recording of Dutch speech and the recording of English speech and one or two weeks between the recording of English speech and the interview tests, all the judgements took place at the end of the academic year.

The average of the two judgements on each speaker's pronunciation of English in the interview test was calculated; then the average of the judgement on the interview test and the judgement on the recording was calculated[23].

Results of speakers who had not performed all three speakers' tasks were not considered. The average judgements on the pronunciation of Dutch were compared with the judgements on the pronunciation of English.

3. RESULTS

When the judgements had been compared it appeared that there was a weak but significant correlation between the ranks for the pronunciation of English and those for the pronunciation of Dutch ($r = .63$; $n = 50$; $p = .046$). Therefore the question if there was a correlation between judgements on the pronunciation of English of speakers of Dutch and judgements on their pronunciation of Dutch was answered in the affirmative.

4. DISCUSSION

The inexperienced author had not heeded the sensible advice that was available to him; therefore there was a number of serious flaws in this preliminary study[24].

---

[23] *Only the final marks were recorded, so that no data are available about each individual listener's judgements. Therefore it is impossible to give indications of the listeners' reliability.*

[24] *Ironically, the author of a widely used research methodology textbook (De Groot, 1961) and the present author had their places of work in adjacent buildings for a number of years. It is doubtful if positive effects of this proximity could be demonstrated.*

Unlike the assessment of the pronunciation of Dutch the assessment of the pronunciation of English in the interview test was not anonymous; also the recordings of English speech were not anonymous. As a result the judgements, those of the native speaker judge as well as those of the non-native speaker judges, may have been subject to a halo effect and even a contamination effect. Moreover the interview and the English recording were also used for judgements on the students' fluency in English; judgements on pronunciation may therefore have been influenced by judgements on fluency.

The final marks given for the students' pronunciation of English had an "absolute" value in the sense that any mark lower than 5.5 would have meant that the student had failed the test and would have to do it again. Although in this study the mark was only used to rank the students, the fact that the students' future depended on the mark was likely to have had an influence on the judgements as a result of a contamination effect.

Several months had elapsed between the recording of the Dutch speech and the recording of the English speech; some weeks had elapsed between the recording of the English speech and the interview test. The three tasks were not judged in the same order, so that, if the judgements had been subject to a sequence effect, this effect would have affected them in different ways.

Also the use of the ten-point scale was less felicitous as had already been demonstrated by De Groot (1966). Many students received the lowest sufficient mark for the pronunciation of English (5.5), which meant that they were given the same rank.

Only one person judged the pronunciation of Dutch – which person was the author, who, as was admitted above, might have been subject to the contamination effect, in the know as he was of the aim of the study – and the pronunciation of English was judged by only three persons, one of whom was a native speaker of English and two of whom were native speakers of Dutch. Although the native speaker of English was a long time resident of the Netherlands, who spoke Dutch well and was familiar with the varieties of Dutch, it was not likely that his judgements of English would be affected by the quality of the speakers' Dutch, whereas such an effect was more likely to occur with the native speakers of Dutch[25].

Native and non-native speakers of Dutch were not kept apart. Although some influence of the pronunciation of an L2 (in this case Dutch) on an L3 (in this case English) would not have been impossible (cf. Koet, 1986), it would have been unlikely that the English of non-native speakers would have been influenced by their Dutch to the same extent as the English of native speakers of Dutch.

The presence of native as well as non-native speaker listeners, however, would have made it less likely for a correlation between the judgements on Dutch and the judgements on English to be found. The fact that a correlation, weak though it may have been, was found would therefore suggest that in study with only non-native listeners a stronger correlation might have emerged.

Although there were serious flaws, for which the author must be held responsible, he could not be blamed for the procedure used to arrive at the judgements on the

---

[25] *As was admitted above, it was a serious omission that no correlations between the individual judges' judgements on English and the judgements on Dutch were calculated.*

pronunciation of English, which was a "real life" testing procedure and not an experiment.

In spite of the objections that could be raised it appeared therefore that there were strong indications that first-year students of English at the teacher education institute concerned had a better chance of obtaining a good mark for their pronunciation of English if their pronunciation of Dutch was judged to be closer to AN.

A number of possible explanations were considered. A first hypothetical explanation would take into account the relative difficulty of the subject matter for the different groups of learners.

The phoneme system of the variety of English taught (RP) may have been more difficult for students with a non-AN Dutch background than for learners with an AN background. It might well be so that speakers of AN have more L1 phoneme oppositions at their disposal than speakers of non-standard varieties of Dutch. From the fragments of speech studied it appeared e.g. that in their production of Dutch speech not only many speakers from Amsterdam but also several speakers from some other parts of the province of North Holland did not at all distinguish between fortis and lenis fricatives as found in Dutch *c* and *zee* and *fee* and *vee*. In addition several speakers from Amsterdam did not distinguish between /s/ and /ʃ / as found in Dutch *Does* and *douche*. The ability to realize these phoneme oppositions would seem to be of the greatest importance in the acquisition of the pronunciation of English (see Gussenhoven & Broeders, 1997, p. 16). Therefore a first "linguistic" explanation was formulated.

> *For speakers of non-standard varieties of the mother tongue it might be more difficult to produce the sounds of the L2 than for speakers of the standard variety because their phoneme inventory lacks certain sounds that are present in that of the standard variety of the mother tongue.*

But not all speakers of a non-standard variety of Dutch lack these phoneme oppositions in their variety of the native language. Also, it might be possible that certain RP vowels would have been easier for speakers of a non-standard variety of Dutch than for speakers of the standard variety. It is doubtful, however, if this could have compensated for these speakers' deficiencies as far as the consonants are concerned, since, as Gimson (1970, p. 4) suggested, the consonants are more important for an understanding of spoken English than the vowels. (Whether this consideration would have prompted the judges to give more weight to the consonants in the assessment of the pronunciation, however, seems doubtful.)

A second, "sociolinguistic", explanation that was formulated would take into account characteristics of the judges. Judges of pronunciation, whether this is the pronunciation of the L1 or of the L2, may favour speakers of the standard variety.

> *Non-native speaker listeners who recognise features in the L2 of non-native speakers that are characteristic of certain non-standard varieties – not of the L2 but of the mother tongue – might therefore be inclined to give a lower rank to these speakers' L2.*

Several studies suggested that in Great Britain and the USA listeners have a more positive attitudes towards speakers of RP and GA, who are felt to be more intelligent and competent, though not necessarily friendlier, than speakers of other varieties; listeners were found to have negative attitudes towards speakers of such varieties as Southern and New York City English in the USA and of Birmingham English in Great Britain (e.g. Labov, 1966, Giles, 1970, Trudgill, 1972, Preston, 1989). There is no reason to suppose that a similar situation in the Netherlands with respect to AN would not exist (cf. Van Hout & Knops, 1988). It is therefore by no means unthinkable that a judge who was a speaker of AN himself or herself or had a positive attitude towards AN would as a result of the halo effect have been inclined to award more positive judgements to speech that he or she had perceived to have been produced by speakers of AN. Similarly one could imagine that a speaker of AN would have exhibited greater tolerance towards mispronunciations that were the results of the interference of AN phonology than towards mispronunciations that were the result of interference of non-standard phonology, if only because the latter would have been less familiar to him or her.

True, one of the assessors of the pronunciation of English was a native speaker of English who, although a long time resident of Amsterdam, would have been less vulnerable to such a halo effect. Also, it was departmental policy to conduct as many of the classes as possible in English. Nevertheless, it would seem that most judges could hardly have been unaware of the quality of the students' pronunciation of the native language. Although the judges concerned might have been inclined to reject this possible explanation, which would have reflected on their own judging behaviour, it appeared to be worth exploring.

A third explanation would take into account learners' general phonetic competence. Accuracy of pronunciation may not be language-specific and may manifest itself in the pronunciation of the native as well as the second or foreign language. This explanation agrees with what was found by Sparks et al. (1998), who demonstrated that there was a connection between foreign language proficiency and native language skills.

> *Speakers who have acquired a high level of general phonetic competence might therefore be able to pronounce both the mother tongue and the L2 well, whereas speakers who have not reached such a high level might pronounce both the mother tongue and L2 less well.*

Before a serious attempt was made to explore these three possible explanations a large number of questions might have been raised.

Were most speakers native speakers of another variety of Dutch than AN? There was an intuitive feeling that for the students studied this question could be answered in the affirmative. It was certainly true for several non-Dutch students and probably true for many natives of Amsterdam and of North Holland. There were, however, many residents of areas in North Holland that were traditionally supposed to harbour many speakers of AN (cf. Donaldson, 1983). It was therefore impossible to give a definite answer.

Did the speakers feel that in producing the recordings it would have been desirable to use AN? On the one hand the situation in the language laboratory was relatively formal: it was a monologue rather than a dialogue (cf. Fagel, 1980, p.8), there was the opportunity to go over the text several times and correct errors and the speakers knew that their products would be studied afterwards. On the other hand it was not a list of words or the reading of a text but semi-spontaneous speech and the learners had not been informed that the focus of interest would be on the pronunciation of Dutch rather than on non-phonological aspects. But even if the speakers had been able to produce what might be regarded as stylistically appropriate variants, one may well wonder if they had wanted to produce these. Again, there was a feeling that for many students studied this was not unequivocally true. Even assuming that for them AN was more "statusful", there was no reason to suppose that another variety of Dutch – e.g. a more "demotic", regionally coloured, variety not too dissimilar from AN – may not have had more covert prestige (cf. Trudgill, 1972).

Did the speakers feel it was desirable to produce the type of pronunciation of English that the assessors expected them to acquire, which in theory was one of the major national varieties of English but in practice RP? Or did they perhaps feel that a variety of Dutch English or even Amsterdam English was more natural to them? There was a suspicion that amongst certain groups of students there was a less positive attitude towards RP and a preference for other varieties of English, in particular General American. Broeders (1981, p. 130) found that his Amsterdam respondents, students from the same teacher-training institute as in the present study, were less positive towards RP than his respondents in Nijmegen in the East of the Netherlands. The fact that one of the oral skills tutors of the Amsterdam students was a speaker of Scottish English rather than RP may also have contributed to a less positive attitude to RP.

These were only some of the questions that might have been raised. Interesting though these may have been, under the circumstances it was felt that it would be more fruitful to focus further research on the first two possible explanations: firstly the phonological differences between the varieties of Dutch and the relative gravity of the phonological interference of these differences in the acquisition of the pronunciation of English and secondly the behaviour of native and non-native judges of the pronunciation of English in judging the English of speakers of different varieties of Dutch. Only if it turned out that these explanations could not account for the correlation that was found, would it be necessary to turn to the third "general phonetic competence" explanation.

5. RESEARCH QUESTIONS

A non-native accent of Dutch may well have an influence on the pronunciation of English, just like a native accent of Dutch and the perception of such a non-native accent may have a negative influence on judgements of the pronunciation of English – the serious effects of the perception of a non-native accent of Dutch were demonstrated by Doeleman (1998). Indeed, it is not unlikely that the perception of a non-native accent of Dutch may have a more negative influence on judgements on the

pronunciation of English than the perception of a native accent of Dutch. But the number of non-native speakers of Dutch among the first-year students concerned was much smaller than that of native speakers. This would have made it less practical to take into account non-native speakers of Dutch in the search for answers to the questions. Also Van Gelderen (1995) found no general differences in the learning of English between native and non-native speakers of Dutch. Therefore it was decided to focus on native speakers of Dutch.

As to the first – "linguistic" – explanation, it also seemed impractical to study more than one non-standard variety. In view of the location of the institute it was decided to concentrate on the Amsterdam variety. On the basis of these considerations the following research question was formulated.

> *Is the phonological interference of the Amsterdam variety of Dutch a more serious obstacle in the acquisition of the pronunciation of English than the phonological interference of the standard variety?*

As to the second possible explanation, the greater strictness that non-native speaker judges had been found to exhibit in their judgements of non-native speech compared to native speaker judges might have been the result of the halo effect that non-native speakers were subject to to a greater extent than native speakers. But as no conclusive evidence existed for such greater strictness in Dutch listeners, this needed to be studied first. Therefore the following research question was formulated.

> *Do Dutch and English listeners to English speech produced by speakers of Dutch arrive at the same judgements on pronunciation?*

Not before this question had been answered, would it be possible to study the effect of the perceived quality of the pronunciation of the native language on the listeners' judgements of the quality of the pronunciation of the L2.

> *Do Dutch listeners arrive at the same judgements on the English pronunciation of speakers of Dutch that they perceive to be of a lower quality as on the English pronunciation of speakers of Dutch that they perceive to be of a higher quality?*

How these three questions were answered will be discussed in each of the following three chapters, Amsterdam English, A Choice of Judges and Judgements on Dutch and Dutch English.

AMSTERDAM ENGLISH

1. INTRODUCTION

In the previous chapter it was seen that there was a correlation between the perceived quality of the pronunciation of English and the perceived quality of the pronunciation of Dutch. Although this correlation was weak, three possible explanations were suggested. One of these explanations, the "linguistic" one, was that the sound system of certain varieties of the L1 (e.g. the standard variety) could provide a better basis for the acquisition of the pronunciation of the L2 than the sound system of other varieties (e.g. non-standard varieties). As it would not have been feasible to study the sound systems of many native and non-native varieties – let alone all the varieties – of Dutch and their relation to the sound system of English, it was decided that the focus should be on an Amsterdam variety of Dutch.

The term "Amsterdam English" was coined for the interlanguage where the variety of Dutch spoken by many inhabitants of Amsterdam was the L1 and the standard British English taught by many teachers in the Netherlands was the L2. Of course it would have been wrong to treat Amsterdam English as a homogeneous non-native variety of English. Strictly speaking there would have been as many different varieties or idiolects of Amsterdam English as there were speakers. There were considerable differences caused by two sets of factors. The first set was also responsible for the nature of the L1 and consisted of such matters as social background and geographical origin. Daan (1969) distinguished three geographical varieties of Amsterdam Dutch, a Northwestern, a Northeastern and a central variety. The second set of factors referred to the speaker's progress in English and determined how close the interlanguage had come to the L2.

The variety of Amsterdam Dutch, the L1, was an accent that could be heard from native speakers of Dutch in West Amsterdam. It possessed many of the features of Amsterdam speech described by Daan (1948, pp. 18-19) and Schatz (1986, pp. 61-65)[26]. It lacked a number of the salient features of the more traditional Amsterdam pronunciation of Dutch that used to be found closer to the city centre (this *Jordaan* accent enjoyed great popularity as the basis for "stage Amsterdam", the accent used by actors portraying Amsterdam characters) and was closer to AN. The accent concerned showed some resemblance to a demotic pronunciation of Dutch heard in a wide area around Amsterdam from speakers for whom AN may have been too civi-

[26] *Schatz' data would have been a few years younger than the present data, although her speakers must have been older than the secondary-school children in this study. Her transcription of AN* lachen *as /lɑgə/ was most curious (Schatz, 1986, pp. 61). Even if it had been a slip of the pen for / lɑɣə/, /lɑχə/ would surely have been more usual (cf. Cohen et al., 1961).*

This chapter was based on a paper presented at the Second international conference on the teaching of Spoken English (Leeds, August 1979). A Dutch version was published as Koet, 1981b.

lised and the local accent not general enough. This accent may well have been more attractive for popular or populist speakers than AN, so that this modified Amsterdam Dutch could frequently be heard in the mass media (cf. Schatz, 1986, p. 23). Perhaps it shared some features with an "advanced" variety of AN that was described as "Polder Dutch" by Stroop (1998).

As far as the learner's progress towards the L2 was concerned, the "ideal naïve learner's" interlanguage was described. In this interlanguage phoneme substitution was used and every English phoneme was replaced with the sound in the L1 that the learner believed to be closest to it. This hypothetical "ideal learner" did not have very great difficulties with the distribution of the phonemes and did not make "transcriptional" errors, i.e errors that were not caused by phonological interference but by English orthography. As spelling appears to be the first obstacle encountered by Dutch learners of English and phonological interference only the second (Gussenhoven & Broeders, 1997, p. 20) not many students could have been such "ideal naïve learners". But there is no reason to assume that speakers of Amsterdam Dutch have greater problems with English spelling than speakers of the standard variety of Dutch, so that there was no need to consider errors caused by spelling.

The research question that was addressed in the study reported in this chapter is:

is the phonological interference of an Amsterdam variety of Dutch a more serious obstacle in the acquisition of the pronunciation of English than the phonological interference of the standard variety of Dutch?

2. METHOD

In order to answer the question the English pronunciation of speakers of "Amsterdam Dutch" was observed, both during English classes and during the reading of texts. The observations were compared with existing descriptions of Dutch English. The extent to which these two non-native varieties of English differed from the standard British English variety was considered as well as the relative seriousness of these differences.

2.1 The speakers and the listener

The speakers were a sample from a group of native speakers of Dutch, pupils in a lower-secondary school in West Amsterdam that catered for pupils with vocational rather than academic interests[27]. As there were hardly any female pupils in the school, the observations were based on the speech of male pupils. From these pupils five were selected to perform an extra speaking task. The selection was made by their teacher, who had been asked to choose five natives of West Amsterdam with a distinct Amsterdam accent.

The listener was a native speaker of Dutch (the author), a native of West Amsterdam ("Oud West").

---

[27] *The school was the Fourth LTS (technical school) on Postjesweg in Amsterdam, Oud West, whose staff and pupils were most cooperative.*

2.2 The speaking task

Five of the pupils were asked to read a Dutch and an English text. For the purpose the Dutch and English versions of "The North Wind and the Sun" were used (International Phonetic Association, 1949, pp. 20, 26)[28]. Their productions were recorded on audiocassettes with a portable Tandberg recorder. The recordings took place in a small room in the school. The pupils performed the task voluntarily and received no payment.

2.3 Procedure

The description of the interlanguage was based on the listener's notes, which he made during several English classes in the school, supplemented by notes based on his study of the extra speaking task. The description used traditional articulatory terminology. For the sake of convenience the standard accent of British English, RP, was referred to as "English" and the standard accent of Dutch, AN, as "Dutch".

3. RESULTS

3.1 The consonants of Amsterdam English

The most frequent realisations of the twenty-four English consonants (Wells, 2000) in Amsterdam English were indicated with the appropriate phonemic symbols, as well as the most frequent realisations in Dutch English (cf. Gussenhoven & Broeders, 1976, pp. 141-143) and examples (cf. Koet, 2005, p. 6). Where differences were found, these were presented in table 1.

From table 1 it will be clear that English has more consonants than Dutch English, which has twenty, and which, in its turn, has more consonants than Amsterdam English, which has fourteen. In word-final position the difference between Dutch and Amsterdam Dutch can be seen to have been reduced by the final tensing, which is found in both accents.

The opposition voiceless – voiced (for English often more correctly: fortis – lenis) is significant for the English obstruents: the plosives, fricatives and affricates. The opposition voiceless – voiced is found in the Dutch plosives and some of the fricatives but this opposition is lost in word-final position, which may lead to serious

---

[28] *Although short, these fragments contained realizations of most English and Dutch phonemes. The Dutch version contained realizations of 20 consonantal phonemes and of 14 vocalic phonemes; the vowels /y/ and /œ:/ were missing, as well as the consonants /ʃ/, /ʒ/ and /g/, the latter two being regarded as "marginal" by Gussenhoven and Broeders (1976, p. 41); therefore these omissions were not serious. The English version contained realizations of 23 consonantal phonemes and of 16 vocalic phonemes; the vowels /ɔɪ/, /ɪə/, /eə/ and /ʊə/ were missing as well as the consonant /ʒ/; the latter was characterized by Gimson as having "a particularly weak functional load" (Gimson, 1970, p. 190); therefore this omission was not regarded as serious.*

spelling problems[29]. Final tensing makes every word-final obstruent voiceless, unless the effect of final tensing is cancelled by assimilation of voice (cf. *dat is goed* (that is good): /χut/ but *dat is een goed boek* (that is a good book): /χud buk/). Dutch orthography ignores the effects of final tensing in the plosives (cf. *wij lijden – ik lijd*, *wij hebben – ik heb* (we suffer – I suffer, we have – I have)); it shows the effects of final tensing in two pairs of fricatives: /s/ – /z/ and /f/ – /v/ (cf. *twee wijven – een wijf*; *wij wijzen – ik wijs* (two females – one female; we point – I point))[30].

Table 1. Different realisations of consonants of English (En), Dutch English (initial & final (#) positions), and Amsterdam English (initial & final (#) positions)

En	*Example*	*Dutch En*	*Dutch En #*	*Amsterdam En*	*Amsterdam En #*
z	Zelda	z	s	s	s
ʃ	Sean	ʃ	ʃ	s	s
ʒ	Eugenie	ʃ	ʃ	s	s
tʃ	Charles	tʃ	tʃ	s	ts
dʒ	Jill	dj	tʃ	s	ts

In Amsterdam Dutch the opposition voiceless – voiced was only found in the labial and apical plosives: /p/ – /b/ and /t/ – /d/, unless these were in word-final position; it was not found in the fricatives, so that there was no opposition /s/ – /z/ or /f/ – /v/, let alone /x/ – /γ/. This does not imply that speakers of Amsterdam Dutch were incapable of producing voiced fricatives; these sounds were found between vowels but were just as indistinctive as the voiced [g] to be found in other varieties of Dutch in words like *zakdoek* (handkerchief) and combinations like *ik ben* (I am)[31].

---

[29] *Mees and Collins (1982) argued that it might be preferable to describe the contrast /s/ – /z/ and /f/ – /v/ as a fortis – lenis contrast; they found that the /f/ – /v/ contrast was unstable in the speech of undergraduates of Leiden University. The intervocalic contrast /χ/ – /γ/ as in lachen – vlaggen is called a "convenient fiction which appears to be maintained in virtually all recent treatments of the phonology of Dutch" (p. 5). Gussenhoven (1992, p. 46), describing Western, educated, middle-generation speech, stated that /v/ was only weakly voiced. The author has heard several school-children, not only in Amsterdam but also in other areas of North-Holland, refer to the letter <f> as the long <v>, which would seem to indicate that the contrast /f/ – /v/ did not exist for them.*

[30] *Final tensing is not shown in the spelling of the pair of velar fricatives /χ/ – /γ/ (cf. wij vlaggen – ik vlag (we flag – I flag)), where ik vlach would have been more consistent.*

[31] *There are AN speakers who produce "marginal" /g/ in loanwords. For these speakers goal and kool (cabbage) would be a minimal pair. Mees and Collins (1983, p. 73) referred to "the status of French as a second language for the higher social classes of the Netherlands over the period dating from the sixteenth to the mid-nineteenth centuries" to account for the large number of marginal phonemes in Standard Dutch and asserted that "the former linguistic position of French has left its mark on the phonological system (...) so that present-day Dutch*

A few examples of the consequences of the disappearance of the voiceless – voiced opposition in Amsterdam Dutch are given in table 2. In the first and second column there are some English minimal pairs that are homophones in Dutch English. In the third and fourth column there are some English minimal pairs that are minimal pairs in Dutch English but homophones in Amsterdam English.

Table 2. Homophones in Dutch English (1st & 2nd column); additional homophones in Amsterdam English (3rd & 4th column)

Dutch English		*Amsterdam English*	
lap	lab	figure	vigour
staff	starve	Sue	zoo
kit	kid	ferry	very
fuss	fuzz		
hack	hag		
could	good		

In Dutch as well as in Amsterdam Dutch there are two important places of articulation for the consonants, here referred to as labial, apical and dorsal (cf. Cohen et al., 1961); for the sake of convenience the /h/ has been left out of consideration. It is typical that in both accents one place of articulation is sufficient in the description, whereas for English several places of articulation need to be specified. This could certainly be called a problem area as may appear from table 3.

Table 3. Rendering of English apical consonants in Amsterdam English in initial and final positions

English	θ	δ	S	z	ʃ	ʒ	tʃ	dʒ
Amsterdam English initial	s	d	s	s	s	s	s	s
Amsterdam English final	s	s	s	s	s	s	ts	ts

From table 3 it appears that only two Amsterdam English apical consonants served as substitutes for no fewer than eight English sounds.

The problems in Dutch English were less serious than one might be inclined to believe at first. Speakers of Dutch would generally render the English palato-alveolar sound with the sound represented in Dutch orthography with <sj> in such

must be regarded as containing a French substratum". It would seem that this assertion would not hold for the non-standard variety described here.

words as *meisje* (girl), *lusje* (loop), *sjouwen* (drag), *sjorren* (lash). The same sound is used in loan words like *sherry* and *goulash*. Some authors regard it as a combination of /s/ and /j/ (cf. Cohen et al., 1961, p. 43, Booij, 1995, p. 7) and others give the sound phonemic status in its own right (cf. Gussenhoven & Broeders, 1976). With some speakers of AN the sound appeared to be subject to positional restrictions: they were unable or unwilling to produce it in word final positions replacing it with /ts/ (cf. *ik bridge*: /ɪk brɪts/) (cf. Mees & Collins, 1982, p. 3). Be that as it may, the presence of /sj/ in their native variety of Dutch would help such speakers in the pronunciation of English words like *shin*, *show* and *shoot*. This does not mean that the English phoneme can be substituted for Dutch /sj/: a pronunciation of the native word *meisje* (girl) or the loan word *sherry* with such a sound would strike most listeners as affected.

The minor redeeming feature in Dutch English referred to above was completely absent in Amsterdam English. In Amsterdam Dutch there was no distinction between /s/ and /sj/ within words, so that words like *douche* (shower) and *Does* (name given to dogs) and *John* and *zon* (sun) were homophones. The articulation of the Amsterdam apical fricative was closer to that of English /ʃ / than that of the Dutch /s/ and /z/, which because of their "dullness" may already have struck British listeners as similar to /ʃ /. Even those speakers of Amsterdam Dutch who had an /s/ that resembled the AN sound had great difficulty in distinguishing between /s/ and /sj/ within words, which was obvious from such hypercorrections as *sampanje* (champagne), *soosalist* (socialist) and *artisok* (artichoke)[32]. It should be pointed out that such speakers would have been perfectly capable of producing /sj/, as appeared from the combination *douche je* (do you shower; are you having a shower; have a shower): [dus jə]; cf. *ik douche*: [ɪk dus][33].

The typical realisation of the apical fricatives in Amsterdam Dutch was no isolated phenomenon. There was also a considerable degree of palatalisation in the apical plosives and nasals, which was particularly noticeable in word-final position. With these sounds, however, this would not usually have caused the merging of minimal pairs; it is doubtful if the plural *katten* (cats) would very often have been misunderstood as the diminutive *katje* (kitten), small though the phonetic difference may have been.

The next problem was the realisation of /r/, in which Amsterdam English differed from Dutch English both in a positive and in a negative sense. In Dutch there

---

[32] *The pronunciation of the names Sjors and Sjimmy, two well-known cartoon characters, would be a case in point. Sjors is a Dutch version of English George; Sjimmy was called Jimmy, when the cartoons first appeared in the Netherlands. The author remembers clearly that the three fricatives were identical and that no /r/ was present, so that the pair were called /sɔsənsɪmi/.*

[33] *It is possible that the degree to which the Amsterdam /s/ approximated English /ʃ / was a good measure of the strength of the accent. It would be comparable to the Biblical contrast Shibboleth – Sibboleth, although the incorrect realisation of the contrast was probably a social rather than the capital offence for which the perpetrator was slain at the passages of Jordan [Judges, 12, 6]. It is a curious coincidence that the name of the popular neighbourhood where the traditional Amsterdam accent was found is de Jordaan.*

is a large variety of /r/'s, such as the rolled, the flapped and the /j/-like sounds and the retroflex /r/, which seems to have gained great popularity with a younger generation of "civilised" speakers at the expense of uvular /r/ (cf. Van Bezooijen et al., 2002)[34]. Among speakers of Amsterdam Dutch the most common realisation of /r/ seemed to be an apical flap, resembling a very quickly articulated /d/. Possibly because of its resemblance to /d/ this /r/ would easily have disappeared before a following apical consonant, so that no /r/ was heard in such words as *hard* (hard), *vaart* (navigation), *vertalen* (translate), *vernieuwen* (renew), *karnemelk* (buttermilk), *Amsterdam* and *fortuin* (fortune). In this respect speakers of Amsterdam Dutch behaved like speakers of RP and other non-rhotic accents of English, where no /r/ occurs before consonants[35].

Before labial and dorsal consonants the reluctance to pronounce /r/ in preconsonantal positions led to the insertion of the so-called svarabhakti-vowel, so that words like *warm* (warm) and *werk* (work) did not sound as if they were written *wam* and *wek* but as if they were written *warrem* and *werrek*. The svarabhakti-vowel would be heard from many Dutch speakers who had an apical realisation of /r/; it seemed to be infrequent with those who had uvular, retroflex or /j/-like variants; although probably subject to some degree of stigmatisation, the svarabhakti-vowel manifested itself very clearly in Amsterdam Dutch, even in loan words, so that *workmate* might be realised as *wurrekmeet*. Some speakers of Amsterdam Dutch carried their fondness of this vowel so far that they inserted it before /r/, *broodje* (roll) becoming a homophone of *bureautje* (small desk) and *prijs* (price) with *Parijs* and *Patrick* rhyming with *vetterik* (fat person)[36]. Many speakers would delete /r/ altogether in unstressed syllables (*bijvoorbeeld* (for example) – *befobbeld*, *professor* – *pefesse*). This last phenomenon did not seem to have spread to Amsterdam English.

Of these two phenomena in the treatment of /r/, deletion and reinforcement, the former should be regarded as an advantage for Amsterdam English, provided of course that a non-rhotic accent like RP is considered more desirable than a rhotic accent like GA, which would depend on the learner's choice of model, the latter as a drawback. A discussion of the treatment of the vowels in Amsterdam English will reveal more substantial advantages.

---

[34] *This could well be a result of the increased prestige of American English and the diminished prestige of French.*

[35.] *Whether a trace of the /r/ was still present in a colouring of the vowel was indiscernible. There did not seem to be compensatory lengthening after r-deletion as in many native varieties of English nor was there a trace of prepausal deletion of /r/, except in unstressed syllables. In this context it may be interesting to note that, when an automobile plant in the Western part of Amsterdam was threatened with closure, the following slogan was carried during a demonstration: "Open de pot, Voor behoud van Ford". If this was intended as a rhyme that would indicate that the final parts of the words pot and Ford were felt to be identical.*

[36] *This phenomenon seemed to be still alive recently and to have made its way into the written language, witness a photograph taken by the author. On the roadsign "Prins Hendrikkade" was written "Prins Henderikkade".*

3.2 The Vowels in Amsterdam English

The most frequent realisations of the twenty RP vowels in Amsterdam English were indicated with the appropriate phonemic symbols, as well as the most frequent realisations in Dutch English (cf. Gussenhoven & Broeders, 1976, pp. 87-88) and examples (cf. Koet, 2005, p. 6). For the sake of convenience the weak vowels represented by /i/ and /u/ in Wells (2000) were left out of consideration.

From the tables it appears that there are twelve monophthongs and eight diphthongs in RP. As most authors arrive at fourteen Dutch monophthongs (cf. Booij, 1995, p. 4), one would be inclined to say that Dutch is the richer language as far as the monophthongs are concerned. A number of Dutch vowels, however, are produced with a combination of the features "front" and "rounded", e.g. those in *nu* /ny/ (now), *leun* /lø:n/ (lean) and *put* /pʏt/ (pit), which combination was lost in standard British English several hundred years ago (Gimson, 1970, p. 82). It would therefore be impracticable to use these front and rounded Dutch vowels as substitutes for the English phonemes. Also, many Dutch learners used the vowel in Dutch *put* to replace the English sound in *luck*. This would seem to happen under the influence of the orthography rather than of auditory similarity.

Where differences between Amsterdam English and Dutch English were found, these are presented in table 4.

Table 4 . English vowels and their counterparts in Dutch English and Amsterdam English

English	*Example*	*Dutch English*	*Amsterdam English*
		Monophthongs	
e	Fred	ɛ	e
æ	Ann	ɛ	e
ɑ:	Margaret	a:	ɑ:
		Diphthongs	
eɪ	Jane	e:	ɛ i
aʊ	Howard	ɔu	au
əʊ	Joan	o:	ɑu

From table 4 it appears that both in Dutch English and in Amsterdam English there are two monophthongs that serve as substitutes for three English monophthongs. From this table it also appears that in Dutch English two monophthongs serve as substitutes for two of the English diphthongs, whereas in Amsterdam English there are no monophthongs that replace English diphthongs.

One of the most serious mispronunciations of Dutch students concerned the opposition of /æ/ – /e/ as in *bad* and *bed*. Most Dutch speakers would use the Dutch

/ɛ/, which was acceptable as a substitute for neither /e/ nor /æ/. Unlike some southern Dutch speakers, whose / ɛ / was so open that it could well serve for /æ/, the Amsterdam speakers produced an / ɛ / that was so close that it could easily be used as a substitute for English /e/. Emphasis is laid on the /e/ - /æ/ contrast; of the vocalic phoneme oppositions lost in Dutch English it may well be the one with the greatest functional load. It scored high for error gravity in a hierarchy of errors of Dutch learners (Collins, 1979).

The close variant of /ɛ/ was not unique to Amsterdam Dutch, neither was it an isolated phenomenon. There was a close allophone of /ɛ/ in AN as well, viz. when a /j/ followed, as in the diminutives or the enclitic personal pronoun of the second person singular. With most AN speakers this palatalization manifested itself clearly in relatively short vowels, as in *Pietje*, *pitje*, *petje*, *ratje*, *rotje*, *doetje* (little Pete, pip, cap, rat, fire-cracker and wimp respectively), where the allophonic regressive assimilation could easily be recognised. In the Amsterdam English interlanguage the vowels in "Pete, pit, pet, pat, pot, putt" were closer than in Dutch English, which would definitely have been an advantage in "pet" but a disadvantage in "pat".

The so-called short vowels seemed to be shorter in Amsterdam Dutch than their Dutch counterparts; the opposite seemed to be the case in the "long" vowels and diphthongs. This lengthening could be very obvious in the long vowel found in *vader* (father). As the Amsterdam sound was a back vowel it was similar to the sound in the English "father", both in quantity and in quality. Many speakers of a more traditional Amsterdam accent would have used the lip-rounding characteristic of a *Jordaan* accent. It would seem that this lip rounding, which was probably stigmatised, was less frequent with the younger speakers and in the Western variant described here[37] (see figure 1).

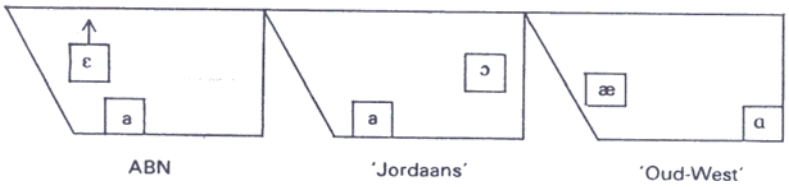
The other three "long" vowels (<ee>, <oo>, <eu>: /e:/, /o:/, /ø:/) showed some diphthongisation with most AN speakers[38]. As in other western urban accents of Dutch, this diphthongisation was so strong in Amsterdam Dutch that the sounds could no longer be classified as monophthongs. The Amsterdam English vowels in e.g. *name* and *gnome* therefore showed a greater resemblance to the corresponding English vowels than the Dutch English sounds, which were nearly always too "pure".

---

[37] *Cf. also Daan (1969), who indicated that these were rounded sounds in only two of the three neighbourhoods studied; she did not indicate that they were rounded in the Northwestern variety. Her data must have been collected about 1950 and concerned older speakers. Schatz (1986, p. 107) concluded that there were two distinct lower-status variants: the stigmatized rounded variant frequently found with male speakers and another non-standard unrounded nasalized variant found with female speakers; one could speculate that it was not so much an effect of sex as an effect of geographical origin.*

[38] *Mees and Collins (1983, p. 68) preferred to describe these vowels as "potential diphthongs", stating that "although conservative ABN still retains a very narrow glide, younger mainstream ABN speakers tend increasingly to use more diphthongal realizations".*

Figure 1. Schematic representation of the articulation of the vowels in "vader" and "rijk" in three varieties of Dutch: the standard variety, the traditional central Amsterdam variety and the Western variety (adapted from Koet, 1981b).



The opposite was the case with two of the three "diphthongs". The Amsterdam words *ei* (egg) and *ui* (onion) were often pronounced as monophthongs[39]. As the lip rounding in *ui* was very slight, confusion might have arisen between words like *rijk* (rich) and *ruik* (smell). Because of its monophthongal pronunciation a word like *bijt* (bite), where the vowel was quite long, may have resembled English *bad*.

The third diphthong <ou> was problematic in that there were two variants in AN, one variant with lip rounding in the first and second elements (/ɔu/) and a variant without lip rounding in the first element (/ɑu/). It would appear that the first element of the Amsterdam Dutch sound was usually further front than that of the AN sound, the lip rounding usually being very slight. This latter realisation seemed a very close approximation of the English diphthong in *out*.

3.3 Suprasegmental aspects

As far as the suprasegmental aspects were concerned few differences were found between Amsterdam Dutch and AN. The differences in intonation in the narrow sense of the word were probably negligible, slighter than the differences between AN and most Northern and Southern accents, slighter also than the differences between AN and some urban accents in the province of South Holland (e.g. Rotterdam, The Hague). As to the temporal aspects, speakers of Amsterdam Dutch seemed to make a clearer distinction between short and long vowels than most speakers of AN, the long vowel as in *vader* being particularly long, and as a result they could have

---

39. The different pronunciations of the <ij> and <ei> were no doubt very old. Halbertsma already made a distinction between the sounds as found in a Northwestern ("Haarlemerdijks") and a Southeastern ("Kattenburgs") variety of Amsterdam Dutch (Van Lennep, 1845). Alberdingk Thijm and Van Lennep (1885) distinguished two monophthongal variants of <ij> and <ei>, an open variant indicated by <aa> and a half-open variant indicated by <ae>. Kloeke (1934, pp. 3-4) stated that in the later 19th century no fewer than four different pronunciations were found, three of which were indicated as <èè > ("wèès voor wijs") <aa> and <ai> ("vaaf and vaif voor vijf"). He gave his personal impression that the sound was perceived as broader ("platter") if it was closer to the <aa>. The former was probably the variant described here, as was the West Amsterdam variant described in Daan (1948, p. 19). The author's impression was that the <ij> was somewhere between cardinal 3 and cardinal 4, whereas <ui> was in the lower part of cardinal 4.

been more similar to speakers of British English, if only the vowels /i/ and /u/ in *ziet* (sees) and *zoet* (sweet) had not been short in Dutch and even shorter in Amsterdam Dutch, unlike English /i:/ and /u:/.

3.4 Relative seriousness of differences

It was not easy to determine if Amsterdam English was superior or inferior to Dutch English. A serious deficiency as far as the consonants are concerned was diagnosed but in the vowels some redeeming features were found. As long as there is insufficient certainty about the relative gravity of errors (cf. Johansson, 1978) one would be inclined to take a dimmer view of mispronunciations of consonants than of vowels, considering the great social and regional variation within the English language and the probable tolerance of most native speakers in this respect (cf. Gimson, 1970). Gussenhoven and Broeders (1997, pp. 16-17) listed eleven points that teachers of English were advised to concentrate on; the first point concerning vowels was only number eight[40]. Most authors hold that intelligibility is of paramount importance and there seems little doubt that mispronunciations that cause loss of phoneme oppositions in the target language and may thus hinder communication are most serious where beginning students are concerned, although allophonic deviations, which may not cause confusion but only irritation, could be more serious where advanced students are concerned. The redeeming features of Amsterdam English vowels would therefore seem to be insufficient compensation for the serious deficiencies in the consonant system.

4. DISCUSSION

In this study there were a number of limitations regarding speakers, listeners and speaking task. The speakers were not representative of speakers of Amsterdam Dutch. They had been chosen from a limited area in West Amsterdam. Also, they were all lower-secondary school pupils, unlike the speakers studied in the previous chapter, who were students in higher education. But this was not a serious disadvantage, since these beginning learners could be expected to be more vulnerable to the difficulties caused by the L1 than intermediate or advanced learners. All the speakers were male and it would certainly have been preferable if both male and female speakers had been studied. As these male speakers were expected to produce a "broader" variety of the local accent than female speakers (cf. Schatz, 1986, and Brouwer, 1989), however, this was not considered a serious disadvantage in the context of this study. Moreover, all the speakers were native speakers of Dutch, whereas there were many non-native speakers of Amsterdam Dutch, not only in West Amsterdam but also in the school. But Van Gelderen (1995) found no relevant differences between native and non-native learners of foreign languages in the Netherlands, so, again, this could not be regarded as a serious disadvantage. Although

---

[40] *Van den Doel (2006, pp. 114-115) arrived at a hierarchy of 32 errors. In the top 13 seven errors involving consonants and only two errors involving vowels are found.*

some twenty-five speakers had been observed during the classes, only five speakers had been recorded for more careful study.

The recorded fragments were short and did not contain all the phonemes of English and Dutch (let alone all the marginal phonemes of Dutch), but the information that was missing in the recordings was supplied from the notes taken during the classes.

There was only one listener, the author, who knew the aims of the study and, moreover, used an impressionist approach. But the listener was thoroughly familiar both with Amsterdam Dutch and with the English produced by its speakers. Also, the findings about Amsterdam Dutch were in line with what had been found by other authors.

Another objection could be that the study did not take into account recent developments in the pronunciation of AN, which developments might cause the realization of some of the AN vowels to become more similar to the realization that was found in Amsterdam Dutch. But the result of these developments would be that AN would share some of the redeeming factors with Amsterdam English, so that the position of Amsterdam English versus the standard variety would be even worse.

In spite of the limitations, therefore, there would appear to be no reason for serious doubts about the findings of this study. The question if the phonological interference of the Amsterdam variety of Dutch was a more serious obstacle in the acquisition of the pronunciation of English than the phonological interference of the standard variety of Dutch was answered in the affirmative.

Therefore it would seem that the "linguistic" explanation would account for at least part of the more negative judgements on the pronunciation of the foreign language of speakers whose native language received more negative judgements.

In view of the limitations of this study, however, it would not have been wise to assume that this explanation could account for all of the negative judgements. The other "sociolinguistic" explanation will be looked into in the next chapter.

A CHOICE OF JUDGES

1. INTRODUCTION

In the chapter Perceived pronunciation quality it was seen that an informal pilot study had shown that there was a weak but significant correlation between non-native listeners' judgements on Dutch speakers' pronunciation of English and the perceived quality of their pronunciation of Dutch. Three possible explanations were offered for consideration:

1) a "linguistic" explanation: for speakers of non-standard varieties of the mother tongue it might be more difficult to produce the sounds of the L2 than for speakers of the standard variety because their phoneme inventory lacks certain sounds that are present in that of the standard variety of the mother tongue;
2) a "sociolinguistic" explanation: non-native speaker listeners who recognise features in the L2 of non-native speakers that are characteristic of certain non-standard varieties – not of the L2 but of the mother tongue – might therefore be inclined to give a lower rank to these speakers' L2;
3) a "general phonetic competence" explanation: speakers who have acquired a high level of general phonetic competence might therefore be able to pronounce both the mother tongue and the L2 well, whereas speakers who have not reached such a high level might pronounce both the mother tongue and L2 less well.

It was also suggested that the instrument that had been used was in need of improvement; the judgements on a 1 – 10 scale showed a restricted range.

In Amsterdam English the "linguistic" explanation was worked out. It appeared that the phoneme system of speakers of an Amsterdam variant of Dutch put these speakers in a less advantageous position for the acquisition of the pronunciation of English than speakers of "Algemeen (Beschaafd) Nederlands" ("General Dutch"), because the phoneme system of the standard variety contained more phonemes that could be used as substitutes for English phonemes than the phoneme system of the non-standard variety. Therefore the "linguistic" explanation might well account for part of the correlation found.

Before the second and third hypotheses could be explored, the choice of listeners needed to be addressed. The listeners in the preliminary experiment were non-native listeners. If these listeners tended to assign a lower ranking to the L2 speech produced by speakers whose variety of the native language was perceived to be non-standard than to the L2 speech produced by speakers whose variety of the native language was perceived to be standard, that tendency would not necessarily be objectionable. If the same tendency could be found in native-speaker listeners, it could

This chapter was based on two earlier reports on the experiment in Koet (1988) and Koet (1992), the latter a written version of a presentation at the New Sounds 92 conference in Amsterdam.

not be said that the non-native listeners were biased. Therefore it needed to be seen to what extent the judgements of native-speaker listeners and non-native speaker listeners corresponded. If the native and non-native speaker listeners were found to agree, features of L1 pronunciation that were present in the L2 of speakers of a non-standard variety would have provoked equally negative reactions from native and from non-native speaker listeners; it could then be argued that it was a fact that these speakers' English was inferior to the English of speakers of the standard variety of the L1. If they were found to disagree, however, and if the speakers of the non-standard variety were assigned lower rankings by the non-native listeners than by the native listeners, it could be argued that the non-native listeners were unfair to the speakers of a non-standard variety; this would put these speakers of a non-standard variety in a disadvantaged position.

Before this correspondence could be established, however, there was an issue of the choice of native-listener judges. In the literature study it was seen that two norms are generally accepted in the Netherlands: the standard British (Received Pronunciation: RP) and the standard American (General American: GA) ones, the first norm being preferred by most teachers, the second being preferred by many students in secondary education (cf. Van der Haagen, 1993). If it was found that teachers' judgements did not at all correspond with American listeners' judgements, that would not come as a surprise; unless the American listeners applied a British norm, which would be most unlikely, differences between the teachers and the American listeners were to be expected. It would stand to reason that British listeners' judgements should be compared with teachers' judgements. The first question then was if there were differences between the teachers' and the British listeners' judgements.

Although British listeners seemed the obvious choice to test the teachers' judgements, it was by no means certain that British and American listeners differed in their judgements of non-native speech. The literature had revealed considerable differences in British and American listeners' appreciation of native speech, it is true, but no such differences were found as regards non-native speech[41]. If there were no such differences, then it would not be necessary to choose either British or American listeners; Anglophone listeners would do. If there were differences, however, and British listeners were, for example, more positive towards speakers whose English showed traces of the non-standard variety of the L1, whereas American listeners were less positive to them, exposing such speakers to American listeners' judgements would put them at a disadvantage. The second question therefore was if there were differences between British and American listeners.

The listeners in the study described in Perceived pronunciation quality were experienced listeners. The literature study showed that there were considerable differences between experienced and inexperienced listeners. These differences, however, were not always consistent. Sometimes experienced listeners were found to be

---

[41] *Van den Doel (2006, p. 297) found that his North American listeners detected fewer errors but regarded the errors they did detect as more serious than other native-speaker listeners; the effects of the lower detection rate and the greater strictness seemed to neutralize each other, so that no significant differences between the mean judgements by North American listeners and those by native-speaker listeners from other countries were observed.*

stricter than inexperienced ones and sometimes they were found to be more lenient; sometimes experienced listeners were found to attach greater importance to certain features of pronunciation, e.g. prosodic features, but at other times they were found to attach greater importance to different, e.g. segmental, features. If there were no differences between experienced and inexperienced listeners, that would be most convenient: it would not matter if teachers or students judged pronunciation; it would not be necessary to distinguish between these two groups. If differences were found, however, and it appeared that experienced listeners arrived at lower rankings of the L2 of speakers of a non-standard variety of the L1 and inexperienced listeners did not, that would be a serious matter if the inexperienced listeners' judgements corresponded with the native-speaker listeners' judgements; in that case it could be argued that the experienced listeners were unfair and inexperienced listeners should be preferred. If the same was found but the experienced listeners' judgements corresponded with the native listeners' judgements, then it could be concluded that the experienced listeners' judgements were valid and that the inexperienced listeners had not learnt to give valid judgements. The third question, therefore, was if there were differences between experienced listeners' (teachers') judgements and inexperienced listeners' (students') judgements.

If differences were found between experienced and inexperienced listeners' judgements, then it would need to be established if the inexperienced listeners' judgements corresponded with those of American listeners. If they did not, the reason might be that the supposition that inexperienced listeners (students) used the American rather than the British norm was incorrect. It could be that, like the teachers, they used the British norm. It could also be that they had not yet learnt to apply any of the norms. The question therefore was if there were differences between inexperienced Dutch and American listeners' judgements.

Finally, the assumption that Dutch teachers used a British and Dutch students an American norm might be erroneous. If there were considerable differences between experienced Dutch and British listeners' judgements or between inexperienced Dutch and American listeners' judgements, that might point to another norm being used by Dutch listeners. It might be so that Dutch students used a British norm or, unlikely though this would seem, that Dutch teachers used an American norm. The fifth question that needed to be answered was if there were differences between experienced Dutch listeners' and American listeners' judgements and between inexperienced Dutch listeners' and British listeners' judgements and, if so, if these differences were greater than the differences between their judgements and those of the listeners that they were paired with for the previous questions. If this were not the case, then the assumptions about the norms used by these listeners would have to be revised.

Another issue was the definition of validity in this study. What criteria should be applied before deciding that a judgement is valid? In the chapter Perceived pronunciation quality the way speakers of English and speakers of Dutch were ranked was addressed. A high correlation would point to the same dimension being used by both groups of judges to assess the speakers. But using the same dimension would not have been sufficient for non-native judges to arrive at approximately the same ranking as native-speaker judges. In the literature study it was seen that judges differed

not only in the way they ranked speakers but also in strictness. Ideally non-native judges should not be stricter or more lenient than native-speaker judges. While searching for arguments for a valid choice of judges, therefore, both fairness (approximately the same ranking as native speaker judges) and strictness needed to be considered.

The literature study had also shown that female speakers' speech is often preferred to male speakers' speech but that this preference is by no means universal. It could well be that Dutch listeners preferred L2 speech produced by female speakers but that British listeners had no such preference. Provided that a British norm was accepted, that would mean that male speakers were put at a disadvantage if their speech was judged by Dutch listeners. The literature study had also shown that sometimes listeners are more positive towards speakers of the opposite sex and more negative towards speakers of their own sex; although this tendency seemed to be rarer than the preference for speech produced by female speakers, it was nevertheless felt advisable to see if this tendency could be found. If it were found that female Dutch listeners e.g. were more negative towards the speech produced by female Dutch speakers and female British listeners were not, that would put female Dutch speakers at an unfair disadvantage as opposed to female Dutch speakers who were judged by male Dutch listeners and male Dutch speakers, regardless of the listeners' sex. Another criterion therefore was if there were differences in the judgements of non-native speech between experienced Dutch and British listeners', British and American listeners', experienced and inexperienced Dutch listeners' and inexperienced Dutch and American listeners' judgements of male and female speech and if there were differences between male and female experienced Dutch and British listeners', British and American listeners', experienced and inexperienced Dutch listeners' and inexperienced Dutch and American listeners' judgements of male and female speech.

The five questions to be answered in this part of the study were:

1) Do experienced Dutch listeners' judgements differ from British listeners' judgements?
2) Do British listeners' judgements differ from American listeners' judgements?
3) Do experienced Dutch listeners' judgements differ from inexperienced Dutch listeners' judgements?
4) Do inexperienced Dutch listeners' judgements differ from American listeners' judgements?
5) Do experienced Dutch listeners' judgements differ from American listeners' judgements and do inexperienced Dutch listeners' judgements differ from British listeners' judgements and, if so, are the differences greater than those between their judgements and those of British and American listeners respectively.

In order to answer these questions differences in ranking and strictness were considered as well as differences in strictness towards male and female speakers and differences in strictness of male and female listeners.

2. METHOD

Four groups of listeners judged the pronunciation of nine Dutch speakers of English in two speaking tasks on five seven-point scales.

2.1 The listeners

The listeners were Anglophone and Dutch listeners. Each group consisted of two subgroups: the Anglophone listeners were either British or American listeners; the Dutch listeners were either "inexperienced" or "experienced" Dutch listeners.

Listeners	Description
British	First-year students of linguistics in the University of Edinburgh, 16 female and 4 male.
American	American and Canadian students at Afcent International School, Brunssum, the Netherlands. The students were in their final year in high school. No attempt had been made to discriminate between American and Canadian students, all of whom were native speakers of English. None of the students was said to be a native speaker of a distinct variety of American English other than General American (personal communication from their British English teacher[42]); there were 13 female and 7 male students.
Inexperienced Dutch	First-year students of English at d'Witte Leli teacher education institute; there were 14 female and 7 male students.
Experienced Dutch	3 female and 6 male tutors in English at d'Witte Leli teacher education institute.

2.2 The speakers

The speakers were first-year students of English at d'Witte Leli teacher education institute; there were five males and four females. They had been selected from the larger group of first-year students as follows: three students had scored a high mark for pronunciation on a test on their fluency and pronunciation in English, three an average mark and three a low mark.

2.3 The listening material

The listening material was a selection from the material described in Perceived pronunciation quality. The speakers had been asked to produce an "oral guided composition" on the basis of a so-called picture composition consisting of 12 pictures and to record their rendering of the story on their cassettes in the language laboratory; their performances were recorded in an Electron SLA C4 language laboratory; none of the recordings was longer than a few minutes. The fragments were collected on a tape in random order. Identical copies of the tape were produced.

---

[42] *Wells (1982, p. 491) stated that "the British usually take English-speaking Canadians for Americans".*

2.4 Procedure

The recordings were offered to the listeners with the fragments in the same order. The Dutch students were asked to listen to the recordings in the language laboratory described above. The British listeners and the Dutch teachers were given the recordings on cassettes and asked to listen to them at their own convenience. The American students were asked to listen to the recordings at the school at any time that was convenient for them. The listeners performed the listening tasks voluntarily and did not receive payment.

2.5 The scales

The instrument used for the pilot study reported in Perceived pronunciation quality had involved the use of 1 – 10 scales, which were found not to discriminate sufficiently between the speakers; as a result many speakers had received the same marks and had to be given the same rank. Another instrument had to be developed. In the literature positive reports about the usefulness of seven-point scales for the description of the pronunciation of Dutch had been found. It needed to be seen if these scales were suitable for judging the pronunciation of English as well as the pronunciation of Dutch. Listeners were asked to indicate their judgements of the pronunciation on five seven-point scales as found in table 1.

Table 1. The five English scales and the corresponding Dutch scales

English scales	*Dutch scales*
1. unpleasant – pleasant	onaangenaam – aangenaam
2. ugly – beautiful	lelijk – mooi
3. broad – cultured	plat – beschaafd
4. strong accent – no accent	sterk accent – geen accent
5. monotonous – melodious	monotoon – melodieus

The scales had been used in the ONU project on the pronunciation of Dutch (see Fagel, 1980, 1981) and the terms were translations of the original Dutch terms[43]. The procedure was based on Snider and Osgood (1969) (see Blom & Van Herpt, 1976). The first two terms were the original terms used by Osgood, Suci & Tannenbaum (1957). The terms "strong accent – no accent" had been adapted from those used in Fagel et al. (1983), where they appeared as "accentedness (high – low degree)".
The scales were presented to the listeners in such a way that the "positive" end was not always on the right, so as to avoid a left or right oriented rating bias (but cf. Van Bezooijen, 1988, p. 90).

[43] *Fagel (1981) used the term "eentonig", whereas the term "monotoon" was used here.*

3. RESULTS

3.1 Correlations between the scales

It would have been convenient if data reduction could be applied. But such a reduction would only have made sense if the scales loaded on the same factors for each of the four groups. In the literature study it had been found that a scale that loaded on one factor in one study loaded on a completely different factor in another study, depending on the groups of judges. Before data reduction could be attempted it was therefore necessary to see if the scales correlated in more or less the same way for each of the groups. Correlation coefficients between the judgements on the scales were calculated for each of the four groups. Results are shown in table 2.

Table 2. Correlation coefficients between the scales for British, American, inexperienced and experienced Dutch listeners (weak correlation (≤ .70) in italics)

	Pleasant	Beautiful	Cultured	No accent
British listeners				
Beautiful	.78			
Cultured	*.56*	*.60*		.
No accent	*.33*	*.42*	*.38*	
Melodious	*.66*	*.60*	*.51*	*.23*
American listeners				
Beautiful	.74			
Cultured	*.17*	*.20*		
No accent	*.20*	*.26*	*.13*	
Melodious	*.62*	*.69*	*.27*	*.16*
Inexperienced Dutch listeners				
Beautiful	*.62*			
Cultured	*.21*	*.39*		
No accent	*.21*	*.36*	*.63*	
Melodious	*.57*	*.47*	*.26*	*.23*
Experienced Dutch listeners				
Beautiful	.82			
Cultured	*.54*	*.67*		
No accent	*.45*	*.63*	*.60*	
Melodious	.79	*.68*	*.55*	*.42*

From this table it appears that the correlation between the scales was not strong, i.e., higher than .90 anywhere. Each of the scales therefore measured a different feature of pronunciation for each of the groups of listeners. The correlation coefficients

clearly differed between the groups of listeners; the hypothesis that the correlation matrices were equal had to be rejected (χ^2= 172.75; df = 30; p < .001).

The British and American listeners' judgements were similar in that there was a moderately strong correlation between judgements on two scales only, "beautiful" and "pleasant". The experienced and inexperienced Dutch listeners showed themselves different from each other and from the two groups of native listeners, in that there was a moderately strong correlation between the experienced Dutch listeners' judgements on the "pleasant" and "beautiful" as well as between their judgements on the "pleasant" and "melodious" scales, whereas the correlation between the inexperienced Dutch listeners' judgements on the five scales was weak.

In view of these differences between the groups of judges it was decided that no data reduction should be applied.

3.2 Reliability

One possible criterion for the choice of judges was their reliability. If one group proved to be less reliable than the other groups that might have been an argument for rejecting that group. The reliability (Cronbach's alpha) of the four groups of listeners was therefore calculated for each of the scales. Results are given in table 3.

Table 3. Reliability estimates of the British, American, inexperienced Dutch and experienced Dutch listeners for a jury with twenty members each (observed reliability in parentheses; high reliability (≥ .90) in bold, low reliability (≤ .70) in italics)

	Anglophone Listeners		Dutch Listeners			
	British	American	Inexperienced		Experienced	
	(n=180)	(n=176)	(n=174-180)		(n=72)	
Pleasant	**.91**	**.96**	.88	(.89)	**.93**	(.86)
Beautiful	.89	**.96**	.88	(.89)	**.97**	**(.93)**
Cultured	**.92**	*.66*	**.95**	**(.96)**	**.95**	**(.90)**
No accent	.78	*.66*	**.90**	**(.91)**	**.91**	(.82)
Melodious	.89	**.94**	**.94**	**(.95)**	**.95**	(.89)

It was not easy to interpret these reliability coefficients since the height of the coefficient is related to the number of judges. The number of judges varying from nine (experienced Dutch listeners) to twenty-one (inexperienced Dutch listeners) could exert a considerable influence on the height of the estimated jury reliability. The number of judges was therefore arbitrarily set at twenty so as to be able to compare estimated reliability with the help of Spearman-Brown's formula for test length (see Lord and Novick, 1968).

From this table it appears that the groups of listeners differed in reliability. The experienced Dutch listeners were highly reliable on each of the scales. The inexperienced Dutch listeners were slightly less reliable as far as their judgements on the "pleasant" and "beautiful" scales were concerned. The British listeners were less reliable than the experienced Dutch listeners on the "no accent" scale and the "beautiful" scale. The American listeners were unreliable on the "cultured" and "no accent" scales.

3.3 Raters using same dimensions: correlation between judgements of the four groups

One criterion for the selection of groups of judges was the ability of non-native speaker judges to arrive at more or less the same ranking as native speaker judges. To find out if a group arrived at more similar rankings than another group the correlation between the judgements of each of the four groups of listeners was calculated for the five scales. Results are presented in table 4.

Table 4. Correlation between the judgements of British (Br), American (Am), inexperienced Dutch (I Du) and experienced Dutch (E Du) listeners on the five scales (strong correlation (≥ .90) in bold, weak correlation (≤ .70) in italics)

	Pleasant	Beautiful	Cultured	No Accent	Melodious
Br-Am	*.66*	.86	**.90**	.78	.72
Br-E. Du	**.94**	**.94**	.81	.76	**.95**
Br-I Du	*.62*	.82	.85	*.48*	.74
Am-E Du	.77	.87	.71	*.70*	.84
Am-I Du	*.13*	*.56*	*.68*	*.48*	*.34*
E-I Du	*.52*	*.67*	**.90**	.83	.71

From table 4 it appears that the correlation between the British and experienced Dutch listeners' judgements was strong on three scales and moderately strong on the two scales relating to accuracy of pronunciation. The correlation between the British and American listeners' judgements was strong on only one scale, the "cultured" scale, and it was weak on only one scale. The correlation between the experienced and inexperienced Dutch listeners' judgements was also strong on the "cultured" scale but weak on the "pleasant" and "beautiful" scales. The correlation between the inexperienced Dutch and the American listeners' judgements was weak on all the scales and even negligible on the "pleasant" scale. From the above the following is clear:

1) The question if experienced Dutch listeners' judgements differed from British listeners' judgements could not be answered in the affirmative for any of the five scales.

2) The question whether British listeners' judgements differed from American listeners' judgements could be answered in the affirmative for only one scale ("pleasant").
3) The question whether experienced Dutch listeners' judgements differed from inexperienced Dutch listeners' judgements had to be answered in the affirmative on two scales ("pleasant" and "beautiful").
4) The question whether the inexperienced Dutch listeners' judgements differed from the American listeners' judgements had to be answered in the affirmative on all the scales.
5) Finally, the experienced Dutch listeners' judgements showed a weak correlation with those of the American listeners on one scale ("no accent"), the experienced Dutch listeners thus showing themselves more similar to the British listeners than to the American listeners; the inexperienced Dutch listeners' judgements also showed a weak correlation with the British listeners' judgements on two scales ("pleasant" and "no accent"), the inexperienced Dutch listeners thus showing themselves more similar to the British than to the American listeners.

3.2 Strictness: judgements of Anglophone and Dutch listeners

In order to determine if there were differences in strictness between the Anglophone and Dutch listeners' judgements, the mean judgements of the Anglophone and Dutch listeners on the speakers' English were calculated. Multilevel model analyses with speakers nested within listeners were carried out to determine the significance of the differences. Results are given in table 5.

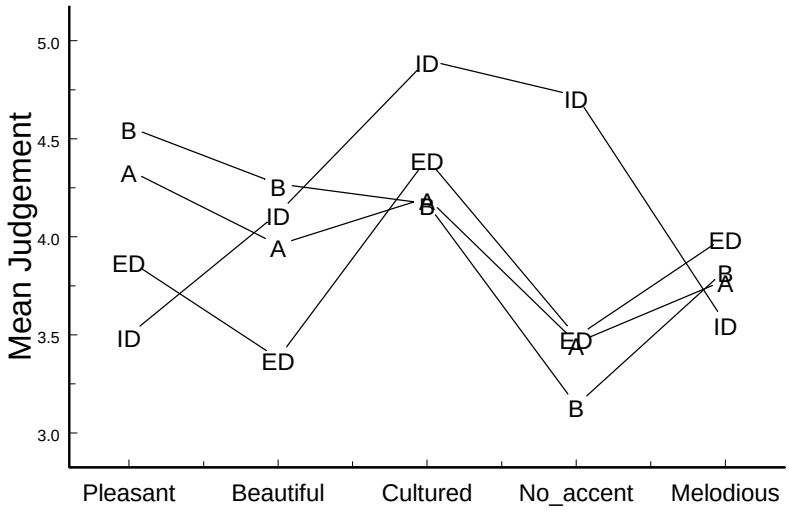
Table 5. British (Br), American (Am), inexperienced (I) Dutch (Du), and experienced (E) Dutch listeners' judgements on the speakers' English; number of cases, means and standard errors (se) are given

	British (n=180)		American (n=175-176)		I Dutch (n=174-180)		E Dutch (n=72)		
	Mean	(se)	Mean	(se)	Mean	(se)	Mean	(se)	More positive than
Pleasant	4.56	(.30)	4.34	(.50)	3.50	(.32)	3.88	(.43)	Br > E, I Du ***
Beautiful	4.27	(.24)	3.96	(.44)	4.12	(.26)	3.38	(.42)	Br, Am, I Du > E Du *
Cultured	4.17	(.32)	4.20	(.21)	4.90	(.44)	4.40	(.30)	I Du > Br, Am, E Du *
No accent	3.14	(.22)	3.46	(.23)	4.72	(.37)	3.49	(.38)	I Du >Br, Am, E Du ***
Melodious	3.83	(.32)	3.78	(.43)	3.56	(.43)	4.00	(.43)	

The mean judgements of British listeners varied between 3.83 for "melodious" and 4.56 for "pleasant". The mean judgements on the "pleasant" scale varied from 3.50

for inexperienced Dutch listeners to 4.56 for British listeners. The means of Table 5 are presented in Figure 1 as well. This figure gives an easier overview of the differences between scales and type of listeners.

Figure 1. Mean judgements (y-axis) by British (B), American (A), experienced Dutch (ED) and inexperienced Dutch (ID) listeners on each of the five scales (data taken from table 5).



These were the results for each of the scales:

- "Pleasant": the British listeners' judgements were more positive than those of experienced Dutch listeners and inexperienced Dutch listeners ($\chi^2 \geq 11.36$; df = 1; $p \leq .001$); no differences could be demonstrated between the judgements of British and American listeners ($\chi^2 = 0.34$; df = 1; p = .56); no differences were found between the judgements of inexperienced and experienced Dutch and American listeners ($\chi^2 \leq 2.07$; df = 1; $p \geq .15$).
- "Beautiful": British listeners' judgements were more positive than those of experienced Dutch listeners, as were the American listeners' judgements and the inexperienced Dutch listeners' judgements ($\chi^2 \geq 5.91$; df = 1; $p \leq .015$), which might lead to the conclusion that the experienced Dutch listeners were unduly strict; no differences were found between the judgements of inexperienced Dutch and British and American listeners and no differences could be demon-

strated between the judgements of British and American listeners ($\chi^2 \leq 0.92$; df = 1; $p \geq .337$).

- "Cultured": the mean judgements of all the groups of listeners were well above the theoretical mean; inexperienced Dutch listeners' judgements were more positive than those of British listeners, experienced Dutch listeners and American listeners ($\chi^2 \geq 4.33$; df = 1; $p \leq 0.037$), which might lead to the conclusion that the inexperienced Dutch listeners were unduly lenient; it would appear that these inexperienced listeners had not yet learnt to apply these terms to the English of non-native, or at least Dutch, speakers; no differences were found between the judgements of experienced Dutch and American listeners; no differences could be demonstrated between the judgements of British and American listeners ($\chi^2 \leq 1.52$; df = 1; $p \geq .217$).
- "No accent": the mean judgements of three groups of listeners were well below, whereas those of the inexperienced Dutch listeners were well above the theoretical mean; the inexperienced Dutch listeners' judgements were more positive than those of British listeners, experienced Dutch listeners and American listeners ($\chi^2 \geq 14.73$; df = 1; $p \leq .001$), which might lead to the conclusion that the inexperienced Dutch listeners were unduly lenient, possibly for the same reason as with the "cultured" scale. American listeners' judgements were more positive than those of British listeners ($\chi^2 = 4.35$; df = 1; $p = .037$); a reason for this greater lenience might be that American listeners are more tolerant of accent in non-native speech than British listeners; another possible explanation is that these American listeners did in fact live in the Netherlands - and would therefore be familiar with Dutch speakers' English - whereas the British listeners did not; no differences were found between the judgements of experienced Dutch and British and American listeners (for al comparisons: $\chi^2 \leq 2.24$; df = 1; $p > .134$).
- "Melodious": no differences were found ($\chi^2 \leq 1.80$; df = 1; $p \geq .129$).

The answers to each of the five research questions with respect to the mean judgements are given below.

- The question whether the experienced Dutch listeners' judgements differed from the British listeners' judgements had to be answered in the affirmative as far as the judgements on the "pleasant" and "beautiful" scales were concerned.
- The question whether the British listeners' judgements differed from the American listeners' judgements had to be answered in the negative.
- The question whether the experienced Dutch listeners' judgements differed from the inexperienced Dutch listeners' judgements had to be answered in the affirmative as far as the judgements on the "beautiful", "cultured" and "no accent" scales were concerned.
- The question whether the inexperienced Dutch listeners' judgements differed from those of the American listeners had to be answered in the affirmative as far as the judgements on the "cultured" and "no accent" scales were concerned.
- Finally, the question whether the experienced Dutch listeners' judgements differed from those of the American listeners differed as far as the judgements on the "beautiful" scale were concerned; therefore they showed themselves more

similar to the American than to the British listeners in the matter of mean judgements; the inexperienced Dutch listeners' judgements differed from those of the British listeners on the "pleasant", "cultured" and "no accent" scales, thus showing themselves more similar to the American than to the British listeners.

The differences between the experienced and the inexperienced Dutch listeners would suggest that the experienced listeners tended to be stricter in their judgements than the inexperienced listeners. Greater lenience on the part of the inexperienced Dutch listeners was found in the two scales associated with accuracy of pronunciation; it would appear that, unlike the experienced Dutch listeners, the inexperienced Dutch listeners had not yet learned how to judge English speech on this dimension. Alternatively it could be stated that these inexperienced listeners' judgements on voice quality were not yet subject to the halo effect of an accurate or inaccurate of pronunciation, between which they had not yet learned to discriminate. Greater strictness on the part of the experienced Dutch listeners was found on the "pleasant" and "beautiful" scales, which might suggest that non-native speech provoked a reaction of revulsion in these experienced teachers.

The fact that so few differences could be demonstrated between the British and American listeners would suggest that British and American listeners did not differ much as far as the strictness or lenience of their judgements of non-native speech were concerned.

The fact that no significant differences were found as far as the "melodious" scale was concerned may come as a surprise in view of the considerable differences between the English intonation produced by native speakers and by Dutch learners found by Willems (1982); since the judgements of three groups of listeners were below and the judgements of one group at the theoretical mean, however, it could be argued that most of the listeners felt that the Dutch speakers' intonation fell short of their standards.

3.3 Judgements of female and male listeners to female and male speakers

One of the characteristics of valid judgements is that these judgements should not be related to irrelevant characteristics of judges or judged persons. Female speakers may well receive more positive judgements than male speakers, but such more positive judgements should be found in all types of judges. If, for example, it were so that female speakers received more positive judgements from American listeners only or even from female American listeners only and not from other groups of listeners, then being judged by American listeners would give female speakers an unfair advantage over male speakers.

To determine if there were differences in strictness in the female and male listeners' judgements and in the judgements on female and male speakers, the mean judgements of the female and male listeners on the female and male speakers' English were calculated. Multilevel model analyses with speakers nested within listeners were carried out to determine the significance of the differences. Results are given in table 6.

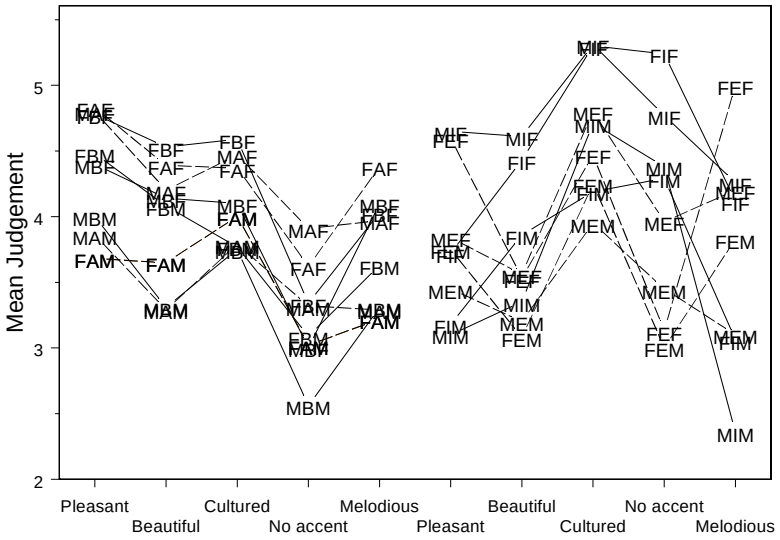
Table 6. Female and male British, American, inexperienced Dutch and experienced Dutch listeners' judgements on the female and male speakers' English; means (M) and standard errors (se) are given

	British listeners							
	Female				Male			
Speaker	Female		Male		Female		Male	
	M	*(se)*	*M*	*(se)*	*M*	*(se)*	*M*	*(se)*
Pleasant	4.78	(0.19)	4.48	(0.22)	4.40	(0.39)	4.00	(0.43)
Beautiful	4.53	(0.17)	4.08	(0.19)	4.15	(0.33)	3.31	(0.37)
Cultured	4.59	(0.19)	3.77	(0.21)	4.10	(0.38)	3.75	(0.42)
No accent	3.34	(0.23)	3.09	(0.25)	3.00	(0.47)	2.56	(0.50)
Melodious	4.03	(0.19)	3.63	(0.21)	4.10	(0.38)	3.31	(0.42)
	American listeners							
	Female				Male			
Speaker	Female		Male		Female		Male	
Pleasant	4.83	(0.22)	3.68	(0.24)	4.80	(0.29)	3.86	(0.33)
Beautiful	4.39	(0.19)	3.65	(0.21)	4.20	(0.25)	3.29	(0.28)
Cultured	4.37	(0.22)	4.00	(0.23)	4.47	(0.29)	3.79	(0.32)
No accent	3.62	(0.26)	3.02	(0.28)	3.91	(0.35)	3.32	(0.38)
Melodious	4.38	(0.22)	3.22	(0.24)	3.97	(0.29)	3.29	(0.32)
	Inexperienced Dutch listeners							
	Female				Male			
Speaker	Female		Male		Female		Male	
Pleasant	3.72	(0.22)	3.18	(0.25)	4.65	(0.28)	3.10	(0.31)
Beautiful	4.43	(0.19)	3.86	(0.21)	4.61	(0.25)	3.35	(0.28)
Cultured	5.30	(0.22)	4.19	(0.24)	5.31	(0.27)	4.71	(0.30)
No accent	5.24	(0.27)	4.29	(0.28)	4.77	(0.35)	4.38	(0.38)
Melodious	4.12	(0.22)	3.06	(0.24)	4.26	(0.27)	2.36	(0.31)
	Experienced Dutch listener							
	Female				Male			
Speaker	Female		Male		Female		Male	
Pleasant	4.60	(0.45)	3.75	(0.50)	3.84	(0.35)	3.45	(0.39)
Beautiful	3.53	(0.39)	3.08	(0.43)	3.56	(0.30)	3.20	(0.33)
Cultured	4.47	(0.44)	4.25	(0.48)	4.80	(0.34)	3.95	(0.37)
No accent	3.13	(0.54)	3.00	(0.58)	3.96	(0.42)	3.45	(0.45)
Melodious	5.00	(0.44)	3.83	(0.49)	4.20	(0.34)	3.10	(0.38)

From table 6 it appears that female British listeners' average judgement of female speakers' English on the "pleasant" scale is 4.78, whereas their average judgement of male speakers' English is 4.48. For each of the five scales the average judgements

have been separated for listeners' and speakers' sex. To facilitate comparison the average judgements have also been presented in a figure (see Figure 2).

Figure 2. Mean judgements (y-axis) by female British (FB), male British (MB), female American (FA), male American (MA) – left-hand side – and female experienced Dutch (FE), male experienced Dutch (ME), female inexperienced Dutch (FI) and male inexperienced Dutch (MI) listeners on female (FBF etc.) and male (FBM etc.) – right-hand side – speakers' English on each of the five scales. Data taken from table 6.



In Figure 2 the mean judgements of the various groups of listeners on the male and the female speakers' English have been presented. From this figure it appears that clear differences have been found, which differences are related to listeners' sex, nationality and experience as well as to speakers' sex.

These were the results on each of the scales:

- "Pleasant": both the American and the experienced Dutch listeners gave more positive judgements of the female than of the male speakers' English ($\chi^2 \geq 15.85$; df = 1; $p \leq .001$); no difference could be demonstrated between the British and the inexperienced Dutch listeners' judgements on the female speakers' English and their judgements on the male speakers' English ($\chi^2 \leq 1.23$; df = 1; $p \geq .267$).
- "Beautiful": both the British and the American and the experienced Dutch listeners gave more positive judgements of the female than of the male speakers' English ($\chi^2 \geq 5.64$; df = 1; $p \leq .018$); no differences could be demonstrated between the inexperienced Dutch listeners' judgements of the female speakers' English and those of the male speakers' English on this scale ($\chi^2 = 1.31$; df = 1; $p = .267$).

- "Cultured": both the British and the American and the experienced Dutch listeners gave more positive judgements of the female than of the male speakers' English ($\chi^2 \geq 4.05$; df = 1; $p < .044$); no differences could be demonstrated between the inexperienced Dutch listeners' judgements of the female speakers' English and those of the male speakers' English on this scale ($\chi^2 = 1.97$; df = 1; $p = .160$).
- "No accent": both the American and the experienced Dutch listeners gave more positive judgements of the female than of the male speakers' English ($\chi^2 \geq 5.48$; df = 1; $p \leq .019$); no differences could be demonstrated between the British and the inexperienced Dutch listeners' judgements of the female speakers' English and those of the male speakers' English on this scale ($\chi^2 \leq 0.67$; df = 1; $p \geq .413$).
- "Melodious": both the American and the inexperienced and the experienced Dutch listeners gave more positive judgements of the female than of the male speakers' English ($\chi^2 \geq 7.46$; df = 1; $p \leq .006$); no differences could be demonstrated between the British listeners' judgements of the female speakers' English and those of the male speakers' English on this scale ($\chi^2 = 3.50$; df = 1; $p = .061$).

In table 7 the situations in which more positive judgements of female speakers were given are summarized.

Table 7. More positive judgements about female than about male speakers' English (p. ≤ 0.05, ** p.≤ 0.01, *** p. ≤ 0.001)*

	More positive about female than about male speakers
Pleasant	American, experienced Dutch ***
Beautiful	British, American, experienced Dutch **
Cultured	British, American, experienced Dutch *
No accent	American, experienced Dutch *
Melodious	American, experienced Dutch, inexperienced Dutch **

From table 7, it is obvious that the inexperienced Dutch listeners were the only group that did not favour the female speakers on more than one scale. The answers to the five research questions appear from the above table.

1) The experienced Dutch listeners' and the British listeners' judgements differed on three scales ("pleasant", "no accent" and "melodious"), where the experienced Dutch listeners' judgements were more positive of the female than of the male speakers.
2) The British and American listeners' judgements also differed on the same three scales ("pleasant", "no accent" and "melodious").

3) The experienced Dutch and inexperienced Dutch listeners' judgements differed on no fewer than four scales ("pleasant", "beautiful", "cultured" and "no accent").
4) The inexperienced Dutch listeners' and the American listeners' judgements differed on no fewer than four scales ("pleasant", "beautiful", "cultured" and "no accent").
5) Finally, the experienced Dutch listeners' judgements differed from the American listeners' judgements on none of the scales, the experienced Dutch listeners thus showing themselves more similar to the American than to the British listeners, and the inexperienced Dutch listeners' judgements differed from the British listeners' judgements on two scales, the inexperienced Dutch listeners thus showing themselves more similar to the British than to the American listeners.

Both the American and the experienced Dutch listeners gave more positive judgements about the female speakers' English than about the male speakers' English on each of the five scales; the British listeners, however, gave more positive judgements on two scales and the inexperienced Dutch listeners on only one scale. The fact that more positive judgements were given about the female speakers' English is not at all surprising in view of earlier studies (e.g. Boves et al., 1982); the fact that it could not be demonstrated that the British and inexperienced Dutch listeners gave more positive judgements on the female speakers' English on some scales was probably due to the greater variance within these groups of listeners. It could, however, have provided an argument against the choice of the British and of the inexperienced Dutch listeners.

As far as the "beautiful" scale is concerned the British female listeners' judgements were more positive than the male listeners' judgements ($\chi^2 = 3.94$; df = 1; p = .047), which were particularly harsh on the male speakers, which would suggest that what Boves et al. (1982, p. 20) found for Dutch listeners was not true for these British listeners; like Dutch females these British males were more demanding of members of their own sex[44]. No difference could be demonstrated between the female American, inexperienced and experienced Dutch listeners' judgements and the male American, inexperienced and experienced Dutch listeners' judgements on this scale. No difference could be demonstrated between the female and the male listeners' judgements on any of the other scales. The fact that the male British listeners differed in this respect from all the other groups might have been argument against their selection as judges, if this difference had been found on more scales. As it only occurred on one scale, it would not have been a strong argument. What became clear from these findings, however, was that it would be inadvisable not to discriminate between male and female speakers and male and female listeners.

[44] *A greater overall lenience in female listeners was also found by Van den Doel (2006, p. 106), who added, however, that "as a rule, women judge the errors they detect more severely than do men" (p. 297).*

4. DISCUSSION

The instrument that was used in this experiment proved itself useful. Five more or less distinct dimensions were found. In table 8 the scales on which the various groups differed are summarized; the differences are: weak correlation of ranking, differences in mean judgements and differences in mean judgements of female speakers.

Table 8. Scales on which the British (Br), American (Am), experienced (E) and inexperienced (I) Dutch (Du) listeners differed as regards ranking of speakers, mean judgements of all speakers and mean judgements of female speakers

	Ranking of speakers	Mean judgements of all speakers	Mean judgements of female speakers
Br – E Du		pleasant, beautiful	pleasant, no accent, melodious
Br – Am	pleasant		pleasant, no accent, melodious
E – I Du	pleasant, beautiful,	beautiful, cultured, no accent	pleasant, beautiful, cultured, no accent,
Am – I Du	pleasant, beautiful, cultured, no accent, melodious	cultured, no accent	pleasant, beautiful, cultured, no accent
Am E – Du	no accent	beautiful	
Br – I Du	pleasant, beautiful	pleasant, cultured, no accent	pleasant, beautiful, cultured, no accent

From table 8 it appears that most differences (11) were found in the judgements on the "pleasant" and "no accent" scales, whereas few differences (3) were found on the "melodious" scale, the "beautiful" and "cultured" scales occupying a middle position. This would suggest that judges disagree most about pleasantness and accentedness of non-native speech but do not disagree too much about intonation.

It also appears that all the pairs of groups considered showed differences regarding ranking, strictness of judgements and lenience towards female speakers, except the British and the experienced Dutch listeners, who did not differ much in their ranking of the speakers and the British and American listeners, who did not differ much in mean judgements and differed in their ranking on only one of the scales.

There seems little doubt that the two pairs, the British and experienced Dutch listeners and the British and the American listeners, were more similar than the other pairs. Indeed, the differences in ranking between the British and American listeners' judgements concerned the judgements on one scale only. True, there were differences between the British and the American listeners as regards their mean judgements of female speakers and as regards the strictness of male British listeners towards male British speakers, but these differences were only found on the "pleasant" scale, it would therefore appear that these differences were much less serious than differences between the other pairs.

The differences between the British and the experienced Dutch listeners concerned the mean judgements on two scales only; as far as the differences between the mean judgements of female speakers were concerned, it might be so that these differences could be ascribed to a greater variation among the British speakers; if there had been less variation their judgements of female speakers would probably have been significantly more positive as well. Therefore the experienced Dutch listeners and the British listeners would appear to be well matched. Obviously, the Dutch teachers had acquired considerable experience in listening to non-native speech but the British listeners would – as students of linguistics – also have acquired some expertise in the matter.

The minor differences between the British and the American listeners might be ascribed to the latter group's more limited experience. It might be argued that the difference between the native speaker listeners was not so much a matter of nationality (British versus American) as a matter of experience; the British listeners were after all students of linguistics, who could be expected to possess a fair degree of linguistic sophistication, whereas the American listeners were high-school students, who could be expected to be linguistically naïve. It must be remembered that the American listeners were supposed to be a match for the Dutch students, who would be approximately the same age; moreover, the American listeners were in fact residents of the Netherlands and it is to be expected that they had been exposed to English speech produced by native speakers of Dutch.

The experienced and inexperienced Dutch listeners differed considerably. It was only in their judgements on the "melodious" scale that no differences were observed, either in ranking or in mean judgements. It is obvious that both in ranking of speakers and in strictness towards speakers, these teacher trainees had not learnt the norms applied by their teachers. The experienced Dutch listeners' greater strictness, which was found on three scales, might be accounted for by a greater sensitivity to the quality and/or standardness of the native language on the part of these experienced listeners, a sensitivity that the inexperienced listeners had not yet developed. These findings differed from what was found in studies such as those by Bongaerts et al (1997), Bongaerts (1999), Chang and Warren (2005), who found no differences between experienced and inexperienced listeners, and in studies such as those by

Schoonen (1991), Cunningham-Anderson (1996) and Van As et al. (2003), all of whom found inexperienced judges stricter rather than more lenient compared with experienced judges.

In spite of the differences between the Dutch listeners, it cannot be denied that both the experienced and the inexperienced Dutch showed themselves clearly different from the native-speaker listeners in several respects. A plausible explanation may be that their perception of the quality of the speakers' English may have been clouded by their perception of the quality of the speakers' Dutch, a quality that the Anglophone listeners cannot reasonably be assumed to have perceived.

Curiously enough the experienced Dutch listeners did not show themselves dissimilar from the American listeners in their mean judgements; where they were more similar to the British listeners as far as the ranking of the speakers was concerned, they were more similar to the American listeners as far as strictness was concerned. The inexperienced Dutch listeners, however, showed themselves as different from the American as from the British listeners. The assumption that the inexperienced Dutch listeners used an American rather than a British norm appeared to be in need of re-examination.

On balance there seemed to be an equal number of arguments in favour of either the British or the American listeners. In the preliminary experiment similarity of ranking was taken as a criterion; if this criterion were applied, then the evidence would definitely be in favour of the British listeners. Also their greater reliability would be an argument in their favour. Arguments against them would be their idiosyncratic behaviour towards female speakers and the male British listeners' harshness on male speakers. In view of the similarity of British and American listeners, however, it seemed safe to proceed with British listeners.

As to the inexperienced Dutch listeners, their greater lenience, which emerged from this experiment, might be the result of a relative lack of sensitivity towards the quality and the degree of standardness of the pronunciation of the mother tongue, a sensitivity that the experienced listeners might well have developed through their exposure to English speech produced by speakers of Dutch. If it were so that these inexperienced listeners were less susceptible to the influence of the speakers' Dutch than the experienced listeners, that would be a serious matter. Speakers who were judged by inexperienced listeners would have an advantage over speakers who were judged by experienced listeners and speakers who were judged by experienced listeners would be at a disadvantage against speakers who were judged by inexperienced listeners. Also in view of the differences between the results of this experiment and those of several other studies it seemed in order to proceed with separate groups of experienced and inexperienced Dutch listeners and see if a similar effect could be found again.

5. CONCLUSION

In the study reported in this chapter arguments were sought in favour of and against the selection of one of the two groups of native speaker listeners. Most of the evidence seemed to be in favour of the British listeners; they were more reliable than

the American listeners and agreed with the experienced Dutch listeners as far as the ranking of the speakers was concerned; but they did not differ much in strictness from the other groups, except from the experienced Dutch listeners, and only on some scales. True, the British listeners did not often show themselves more positive about female speakers than about male speakers, as the experienced Dutch and the American listeners did, and the British male listeners were less positive of speakers of their own sex, whereas the other groups were not. But this only occurred on one scale. On balance, therefore, there seemed to be slightly more arguments to find in favour of the British listeners.

During the search for the best non-native speaker judges, it emerged that there were considerable differences between inexperienced and experienced Dutch listeners. The magnitude of the differences between experienced and inexperienced Dutch listeners was such that it would have been inadvisable not to proceed with these two groups. The differences between these groups of Dutch listeners is one of the topics that were addressed in the final experiment.

JUDGEMENTS ON DUTCH AND DUTCH ENGLISH

1. INTRODUCTION

In the experiment reported in the previous chapter it had been found that the Dutch listeners to English speech produced by native speakers of Dutch differed often from the Anglophone listeners in the strictness of their judgements; it had also been found that these Dutch listeners could often be said to be unfair in the sense that there was no strong correlation between their judgements and the judgements of the Anglophone listeners. An explanation for these differences was suggested: the perceived quality of the pronunciation of the native language, to which quality the Dutch listeners, unlike the Anglophone listeners, could well be sensitive. In the study reported here an attempt was made to demonstrate the existence of such a sensitivity.

In the process of acquiring the pronunciation of a foreign language learners are confronted with a number of obstacles, such as the inherent difficulty of the foreign language and the spelling system. Another obstacle is the interference of the pronunciation of the native language; thus beginning learners will be inclined to interpret the sounds of the target language as the sounds of their variety of the mother tongue and to use the sounds of this variety of the mother tongue in the attempt to produce the sounds of the target language. It is a well-known fact that speakers of some languages have an advantage over speakers of some other languages in acquiring the sounds of certain target languages; the same applies to speakers of some varieties of languages.

In the assessment of the pronunciation of the foreign language assessors often refer to the extent to which learners have overcome the interference of the native language and approximated the pronunciation of the native speakers of the target language. It seems probable that in practice other factors play a part in the formation of the assessment. It is a well-known fact that listeners associate the varieties of a language with social and personal characteristics (cf. e.g. Labov, 1966, Giles, 1970). There is no reason to assume that assessors of the pronunciation of a foreign language do not engage in such an association. As the pronunciation of a foreign language of most beginning and intermediate learners shows perceptible interference by the native language, it stands to reason that listeners who are also native speakers of that language can perceive what variety of the native language is responsible for the interference, a variety that may be associated with social and personal characteristics. As a result the assessment of the pronunciation of the target language is not only determined by the extent to which interference of the native language takes place but also by what particular variety of the native language causes the interference. From the preliminary study of the situation in the Netherlands reported in the chapter Perceived pronunciation quality in mother tongue and foreign language it appeared that interference from other varieties of Dutch than standard Dutch might

A first version of this chapter was presented at the AILA conference in Jyväskylä in 1996.

well lead to a more negative assessment of the pronunciation of English. It was the aim of the present study to indicate in what way the Dutch listeners' assessment of the pronunciation of English by Dutch speakers was influenced by the interfering variety of Dutch.

Before this problem was addressed, however, the findings of the first experiment about the Dutch listeners' strictness and unfairness needed to be confirmed. In the present study an attempt was made to answer three questions, the first two of which were to confirm the earlier findings.

1) Are Dutch listeners stricter in their judgements of English speech produced by native speakers of Dutch than English listeners?
2) Are Dutch listeners unfair in their judgements of English speech produced by native speakers of Dutch in the sense that there is a weak correlation between their judgements and the judgements of Anglophone listeners?
3) Are Dutch listeners prejudiced in their judgements of English speech produced by native speakers of Dutch?

The main issue of the differences between Dutch and English listeners may well be obscured by other factors. These factors concern the differences between:

a) free speech and the reading of a text,
b) experienced and inexperienced listeners,
c) male and female speakers,
d) listeners to speakers of the same sex and listeners to speakers of the opposite sex.

It is not unthinkable that Dutch listeners are stricter in their judgements of free speech than of the reading of texts, whereas English listeners are stricter in their judgements of the reading of texts than of free speech and vice versa. If English listeners are less experienced than Dutch listeners, this may cause them to be more lenient than Dutch listeners. Dutch listeners may be more lenient towards female speakers than towards male speakers, whereas English listeners may be stricter with female speakers than with male speakers. Finally, male Dutch listeners may be more lenient in their judgements of female speakers than of male speakers and female Dutch listeners may be more lenient in their judgements of male speakers than of female speakers, whereas male and female English listeners are not and vice versa. The first question was therefore divided in four sub-questions.

1A. Are Dutch listeners stricter in their judgements of English free speech produced by native speakers of Dutch than English listeners and are Dutch listeners stricter in their judgements of the reading of an English text by native speakers of Dutch than English listeners?

1B. Are experienced Dutch listeners stricter in their judgements of English speech produced by native speakers of Dutch than experienced English listeners and are inexperienced Dutch listeners stricter in their judgements of English speech produced by native speakers of Dutch than inexperienced English listeners?

1C. Are Dutch listeners stricter in their judgements of English speech produced by male native speakers of Dutch than in their judgements of English speech produced by female native speakers of Dutch and are English listeners stricter in their judge-

ments of English speech produced by male native speakers of Dutch than in their judgements of English speech produced by female native speakers of Dutch?

1D. Are Dutch listeners stricter in their judgements of English speech produced by native speakers of Dutch of the same sex than in their judgements of English speech produced by native speakers of Dutch of the opposite sex and are English listeners stricter in their judgements of English speech produced by native speakers of Dutch of the same sex than in their judgements of English speech produced by native speakers of Dutch of the opposite sex?

2. METHOD

A listening experiment was carried out to seek answers to these three questions and four sub-questions. Two groups of listeners judged the pronunciation of twenty Dutch speakers of English in two speaking tasks on five seven-point scales and one of these group of listeners judged the pronunciation of these twenty Dutch speakers of Dutch in two speaking tasks on five seven-point scales.

2.1 The listeners

The findings presented here are based on the judgements of 126 listeners, 76 female and 50 male, who judged some 4700 fragments. There were 70 Dutch listeners, 57 of whom were students in tertiary education (most of them 1st-year students of English) and 13 of whom were teachers of English, most of them in higher education. There were 56 English listeners, 54 of whom were students (most of them in the sixth form) and two of whom were teachers, none of them teachers of languages, in three secondary schools in county Suffolk in England.

2.2 Listening situations

The listening situations involving the two nationalities are presented in table 1. There were three listening situations: Dutch listeners judging Dutch speech, Dutch listeners judging English speech and English listeners judging English speech.

Table 1. Listening situations

Listener	Speech	
	Dutch	English
Dutch	x	x
English		x

Note that in this table the lower left-hand cell has been left empty; it was felt that the English listeners would not be able to give meaningful judgements about Dutch speech, few English listeners being familiar with Dutch. This lack of reciprocity reflects the dominant position of the English language over the Dutch language. A case could, however, be made for asking the English listeners to judge Dutch speech, assuming that they would concentrate on non-linguistic aspects. It was felt that this would be outside the scope of this study.

2.3 The speakers

The speakers were twenty upper-intermediate to advanced learners of English, first-year full-time students of English in a teacher education institute in Amsterdam[45]. In the previous experiment it had been found that the Dutch and English listeners did not give similar judgements of male and female speakers; amongst the speakers in this experiment, therefore, equal representation of male (ten) students and female (ten) students was ensured. The students had been chosen in such a way that the variety of Dutch of half had been judged "standard" and the variety of the other half "non-standard" by six trained Dutch listeners (three female, three male).

2.4 The speaking tasks

The students had been asked to perform two types of task. First they were asked to read out two short texts, one Dutch text and one English text. The texts were "The king of the birds" (Gussenhoven & Broeders, 1976, p. 210) and "Koel, helder water" (Schilthuizen & Vink, 1977, p. 31). Then they were asked to tell a story, first in Dutch and then in English, on the basis of the "picture composition" that had been used in the previous experiments. A Sennheiser MKH 105 T microphone was used and the material was recorded on TDK HS cassettes with a Panasonic Hifi VHS recorder. The students performed these tasks voluntarily and did not receive payment.

2.5 Instrumentation

In the previous experiment as reported in A choice of judges five scales had been used, all of which had been studied by Fagel et al. (1983). In this experiment six more scales were used, three of which had also been used by Fagel et al. (1983): "expressionless – expressive", "slovenly – polished" and "inaccurate – precise". Three new scales were introduced "unintelligible – intelligible", "standard – non-standard" and "common – distinguished". The last scale had been adapted from the "distinct – indistinct" scale. It was hoped that the "standard – non-standard" scale would allow the listeners to indicate in how far they felt the pronunciation approximated the standard set by educational institutions (the secondary school, the college, cf. Van der Haagen, 1998, p. 73), whereas the scale "common – distinguished" would allow them to give another intuitive judgement on the "statusfulness" of the

---

[45] *The institute was the Faculteit Onderwijs en Opvoeding in the Hogeschool van Amsterdam.*

pronunciation in addition to the "broad – cultured" scale[46]. The scales used were the following:

1) unpleasant – pleasant,
2) broad – cultured,
3) ugly – beautiful,
4) slovenly – polished,
5) strong accent – no accent,
6) non-standard –standard,
7) monotonous – melodious,
8) expressionless – expressive,
9) unintelligible – intelligible,
10) inaccurate – precise,
11) common – distinguished[47].

2.6 Procedure

The fragments were presented to the listeners on two audiocassettes (one with Dutch speech and one with English speech) of some 30 minutes each, containing 44 fragments of 40 seconds; the 22 text fragments were presented first, the 22 fragments of free speech last. Some fragments were presented twice; the first fragments of each pair, intended to allow the listeners to familiarise themselves with the texts, were not considered here. The Dutch listeners had not been told that each speaker had produced Dutch and English speech; the listeners had not been told that some fragments occurred twice.

The English listeners listened only to one audiocassette containing the English fragments; they did so at any time that was convenient for them. The Dutch students listened to the audiocassettes in a Tandberg Educational IS 10 language laboratory; they first gave their judgements on the Dutch audiocassettes (on the assumption that it would be easier for them to judge listening material in their native language); they were asked to listen to the English audiocassettes at a later date. The Dutch teachers also listened to the Dutch audiocassette before they listened to the English audiocassette; they also listened at any time that was convenient for them.

The Dutch and the English students and the English teachers performed the task voluntarily and did not receive payment. The Dutch teachers received a small payment.

There were nearly as many judgements by Dutch listeners of Dutch as of English fragments (37% resp. 36%), with slightly fewer judgements by English listeners (27%). 37 % of the judgements were made on the fragments of free speech, the remainder being on the read texts. Nearly all the data on the read texts were complete;

---

[46] *Berns (2002, p. 65) indicates that in Amsterdam street language the Dutch word standard means: obvious, natural.*

[47] *The Dutch listeners used the Dutch translation: (1) onaangenaam – aangenaam, (2) plat – beschaafd, (3) lelijk – mooi, (4) slordig – verzorgd, (5) sterk accent – geen accent, (6) niet-standaard – standaard, (7) monotoon – melodieus, (8) expressieloos – expressief, (9) onbegrijpelijk – begrijpelijk, (10) onnauwkeurig – precies, (11) ordinair – gedistingeerd.*

unfortunately this was not the fact for the judgements on free speech, since several listeners had only performed the first task[48]. As a result there were a number of missing values. There were slightly more judgements on fragments produced by female speakers (53%) than by male speakers.

3. RESULTS

3.1 Data reduction

In order to find if it was possible to cluster the scales factor analyses were performed on the 11 scales for each listening situation. In order to obtain similar clustering for the three listening situations four factors were allowed. Results of the factor analyses are given in Appendix B, tables 2.

For the three listening situations the scales 1, 2, 3, and 5 loaded on a first factor, scales 7 and 8 on a second factor, scales 9 and 10 on a third factor; scale 6 loaded on a fourth factor. Scale 11, however, loaded on factor 1 for the Dutch listeners and on factor 2 for the English listeners. It would, therefore, appear that, for the English listeners, the question whether speech is common or distinguished related to accuracy and intelligibility; one might speculate that the term "distinguished" suggested to them that the speech sounds were kept distinct, as the original scale "indistinct – distinct" would have done. In view of the fact that the scale probably conveyed a different meaning to the two nationalities, it was decided not to consider this scale in the further analysis.

It appeared that scales 9 and 10 loaded on factor 3 for Dutch speech and on factor 2 for English speech; it also appeared that scales 7 and 8 loaded on factor 2 for English speech and on factor 3 for Dutch speech; scale 6 was alone in loading on factor 4. As far as scales 1, 2, 3, 4 and 5 were concerned, the situation was less clear. For Dutch speech they all clearly loaded on factor 1, as they did, although less clearly, for English speech as judged by English listeners; for Dutch listeners judging English, however, scale 4 appeared to load on factor 2, more so than on factor 1, but the difference was small. In spite of this uncertainty, it was decided to retain the "standard' scale and create three new composite scales as follows: aestheticity (composite of the "pleasant", "cultured", "beautiful", "polished" and "no accent" scales), intonation (composite of the "melodious" and "expressive" scales) and comprehensibility (composite of the "intelligible" and "precise" scales).

3.2 Reliability

Scale reliability coefficients (Cronbach's alpha) were calculated for the three intended composite scales. Results for the three listening conditions are summarised in table 2[49]. It is to be observed that the reliability of the aestheticity scales was slightly greater than that of the other scales. The reliability in the Dutch – Dutch listening

[48] *It appeared that several English listeners had forgotten to turn the cassettes, the fragments of free speech being on the B side.*
[49] *Differences in N between scales were due to missing values.*

situation tended to be greater than in the other situations. This may well have been so since in that situation the Dutch listeners judged native speech, which must have been easier than judging non-native speech. The reliability was least for the intonation scale in the English – English listening situation. The reason for this may well have been that the intonation was more difficult to judge for the English listeners than other aspects of non-native speech. The internal consistency of the composite scales appeared to be satisfactory and differences between the listening situations were not so substantial that a need was felt for other scales.

Table 2. Scale reliability coefficients composite scales; Dutch listeners to Dutch speech, Dutch listeners to English speech and English listeners to English speech

		Dutch listeners to Dutch speech		*Dutch listeners to English speech*		*English listeners to English speech*	
Composite scale	N items	N	Reliability	N	Reliability	N	Reliability
Aestheticity	5	1572	.89	1679	.87	1166	.85
Intonation	2	1604	.87	1730	.89	1175	.74
Comprehensibility	2	1549	.80	1725	.73	1184	.77

3.3 Correlations

To determine if each of the composite scales and the standardness scale contributed to the judgements, correlations between the composite scales were calculated for the three listening situations. Results are given in table 3[50].

*Table 3. Correlation between the four scales; listening situations Dutch – Dutch, Dutch – English and English – English (*p < .0001)*

	Du. listeners to Du. speech N of cases: 1444			*Du. listeners to Eng. speech* N of cases: 1669			*Eng. listeners to Eng. Speech* N of cases: 1151		
	Aesth.	Inton.	Compr.	Aesth.	Inton.	Compr.	Aest.	Inton.	Compr.
Intonation	.52*			.52*			.62*		
Comprehensibility	.60*	.51*		.60*	.45*		.56*	.47*	
Standardness	.28*	n.s.	.17*	.32*	.11*	.24*	.39*	.32*	.41*

[50] *Missing values were listwise deleted.*

It appears that there were significant correlations between all the scales; correlations between the standardness scale and the composite scales were weak as far as the Dutch listeners were concerned but less so as far as the English listeners were concerned; correlations between the composite scales were stronger, but not so strong that it would have been advisable to combine them.

3.4 Judgements of English and Dutch listeners on free speech and reading a text

To determine if there were differences in strictness in the Dutch and English listeners' judgements on free speech and on the reading of a text, the mean judgements of the Dutch and English listeners on the speakers' English were calculated for the free speech and the reading of the texts. Multilevel model analyses with speakers nested within listeners were carried out to estimate means and variance components and to determine the significance of the differences (see Appendix A for variance components). Results are given in table 4.

Table 4. Dutch and English listeners' judgements on the speakers' English in free speech and reading a text; number of cases, means and standard errors (se) are given

	Dutch Listeners						English Listeners					
	Free Speech			Reading Text			Free Speech			Reading Text		
	N	Mean	(se)	N	Mean	(se)	N	Mean	(se)	N	Mean	(se)
Aesthet.	756	3.66	(.07)	923	3.64	(.07)	93	3.91	(.13)	1073	3.91	(.05)
Inton.	773	3.30	(.09)	957	3.48	(.09)	95	3.87	(.15)	1087	3.98	(.05)
Compreh.	772	4.29	(.10)	953	4.54	(.09)	96	4.51	(.15)	1088	4.64	(.07)
Standard	772	3.90	(.10)	943	3.95	(.09)	96	4.07	(.15)	1072	4.10	(.06)

For Aesthetics and Intonation: Language effect ($\chi^2 \geq 6.87$; df = 1; $p \leq .01$)
For Comprehensibility: Type of task effect ($\chi^2 = 6.05$; df = 1; $p = .01$)

From this table it appears that average judgements on the aestheticity and intonation scales were below the theoretical scale mean (less than 4.00). Average judgements on the comprehensibility scale, however, were above the theoretical scale mean (more than 4.00). On the standardness scale average judgements by the Dutch listeners were negative, whereas average judgements by the English listeners were positive. From this table it also appears that the English listeners were more positive about the Dutch speakers' English than the Dutch listeners on the aestheticity and intonation scales; for these scales the effect was one of language. No interaction effect between language and type of task was found.

In addition both the Dutch and the English listeners were more positive about the speakers' English on the comprehensibility scale when these speakers had read a text than when they had produced free speech. The latter is not surprising; while reading the texts, the speakers were probably more careful with their pronunciation, which may well have caused them to be understood more easily.

In view of the fact that there were only significant differences between judgements on these tasks as far as this comprehensibility scale was concerned, there seemed to be little point in continuing to keep the reading of a text and free speech apart in the further analysis. Also the fact that there was no interaction between language and type of task indicated that it would not be necessary to distinguish the two types of task.

3.5 Inexperienced and experienced Dutch and English listeners' judgements of the speakers' English

In the previous experiment it had been found that the experienced Dutch listeners were stricter than the inexperienced Dutch listeners in their judgements on some of the scales. In order to validate this result for the speakers and listeners in this experiment, the inexperienced and experienced Dutch and English listeners' mean judgements on the speakers' English were calculated for each of the four scales. Multilevel model analyses with speakers nested within listeners were carried out to determine the significance of the differences. Results are presented in table 5.

Table 5. Inexperienced and experienced Dutch and English listeners' judgements on the speakers' English; number of cases, means and standard errors (se) are given

	Dutch Listeners						English Listeners					
	Inexperienced			Experienced			Inexperienced			Experienced		
	N	Mean	(se)	N	Mean	(se)	N	Mean	(se)	N	Mean	(se)
Aesthet.	1243	3.70	(.08)	436	3.50	(.13)	1128	3.97	(.05)	38	3.83	(.26)
Inton.	1281	3.32	(.10)	449	3.64	(.17)	1135	3.98	(.05)	40	4.17	(.27)
Compreh.	1279	4.48	(.11)	446	4.35	(.19)	1145	4.61	(.07)	39	5.10	(.35)
Standard	1257	4.01	(.10)	458	3.68	(.18)	1128	4.10	(.06)	40	4.02	(.31)

For Intonation and Comprehensibility: Language effect ($\chi^2 \geq 4.46$; df = 1; $p \leq .03$)

From this table it appears that there were no differences between the inexperienced and experienced Dutch and English listeners, as far as their judgements on the aestheticity and standardness scales were concerned. From this table it also appears that both the Dutch inexperienced and experienced listeners' judgements were less positive than the experienced and inexperienced English listeners' judgements on the intonation and the comprehensibility scales, but this was not a matter of experience or inexperience. The fact that there were so few experienced English listeners and relatively few experienced Dutch listeners may lead one to speculate that with larger numbers significant differences could have been found. Under the circumstances, however, these data were inconclusive.

In view of the fact that no significant differences between experienced and inexperienced listeners had been found, there seemed to be little point in continuing to keep experienced and inexperienced listeners apart in the further analysis.

3.6 Dutch and English listeners' judgements of male and female speakers' English

In the previous experiment it had been found that the British listeners gave more positive judgements about the female than about the male speakers on the "beautiful" and "cultured" scales, which scales became part of the composite aestheticity scale in this second experiment, whereas the experienced Dutch listeners gave more positive judgements about the female listeners' speech than about the male listeners' speech on all the scales.

In order to determine if the same was true for the speakers and listeners in this experiment, the Dutch and English listeners' mean judgements on the male and the female speakers' English were calculated for each of the four scales. Multilevel model analyses with speakers nested within listeners were carried out to determine the significance of the differences. Results are presented in table 6.

Table 6. Dutch and English listeners' judgements on female and male speakers' English; number of cases, means and standard errors (se) are given

	Dutch Listeners						English Listeners					
	Female Speakers			Male Speakers			Female Speakers			Male Speakers		
	N	Mean	(se)	N	Mean	(se)	N	Mean	(se)	N	Mean	(se)
Aesthet.	840	3.74	(.07)	839	3.54	(.07)	557	3.91	(.06)	609	3.92	(.06)
Inton.	865	3.56	(.09)	865	3.23	(.09)	564	4.09	(.06)	617	3.86	(.06)
Compreh.	865	4.42	(.10)	860	4.46	(.10)	566	4.43	(.08)	681	4.83	(.08)
Standard	856	4.02	(.09)	859	3.82	(.09)	562	3.94	(.05)	606	4.23	(.07)

For Aesthetics, Comprehensibility and Standard: Language x Sex effect ($\chi^2 \geq 7.21$; df = 1; $p \leq .01$)
For Intonation: Sex effect ($\chi^2 = 33.67$; df = 1; $p < .01$)
For Intonation: Language effect ($\chi^2 = 35.80$; df = 1; $p < .01$)

From this table it appears again that the English listeners gave more positive judgements about the speakers' English than the Dutch listeners on the intonation scale. In addition the female speakers received more positive judgements than the male speakers from the English and the Dutch listeners on the intonation scale. On three scales, aestheticity, comprehensibility and standardness there was an interaction between the speakers' sex and the listeners' language; this took a different form on the three scales, the Dutch listeners giving more positive judgements about the female speakers than about the male speakers on the aestheticity and the standard scale, and the English listeners giving more positive judgements about the male

speakers than about the female speakers on the comprehensibility scale and the standard scales.

3.7 Male and female Dutch and English listeners' judgements of male and female speakers' English

In the previous experiment it had been found that the female British listeners gave more positive judgements on the "pleasant" scale about the English of speakers of their own sex than about that of speakers of the other sex.

To find out if these findings could be replicated with the speakers and listeners in this experiment, the male and female Dutch and English listeners' mean judgements on the male and the female speakers' English were calculated for each of the four scales. Multilevel model analyses with speakers nested within listeners were carried out to determine the significance of the differences. Fixed effects of speakers' sex and listeners' sex as well as their interaction were established. Results are presented in tables 7 A and B.

Table 7 A. Female and male Dutch listeners' judgements on female and male speakers' English; number of cases, means and standard errors (se) are given

	Dutch Listeners											
	Female Speakers						Male Speakers					
	Female Listeners			Male Listeners			Female Listeners			Male Listeners		
	N	Mean	(se)	N	Mean	(se)	N	Mean	(se)	N	Mean	(se)
Aesthet.	554	3.77	(.09)	286	3.68	(.12)	551	3.61	(.08)	288	3.42	(.12)
Inton.	572	3.56	(.11)	293	3.56	(.15)	573	3.20	(.11)	292	3.28	(.15)
Compreh.	571	4.48	(.12)	294	4.30	(.16)	569	4.54	(.12)	291	4.30	(.16)
Standard	560	4.07	(.11)	296	3.91	(.16)	557	3.90	(.16)	302	3.66	(.16)

For Aesthetics, Intonation and Standard: Sex speakers effect (F>M) ($\chi^2 \geq 15.95$; df = 1; $p < .01$)

From table 7 A it appears again that the Dutch listeners gave more positive judgements about the female speakers' English than about the male speakers' English, except where the comprehensibility scale was concerned. It also appears that the Dutch listeners did not give more positive judgements about the English of speakers of the opposite sex than about that of speakers of their own sex.

From table 7 B it appears again that the English listeners gave more positive judgements about the male speakers than about the female speakers on the comprehensibility and standardness scales, whereas they gave more positive judgements about the female speakers than about the male speakers on the intonation scale. From table 7 B it also appears that the English listeners gave more positive judgements about speakers of the opposite sex on the aestheticity scale, whereas the Dutch listeners did not. This effect seems to be stronger with male listeners than

with female listeners. The lack of experience on the part of the English listeners may have caused them to be influenced by a greater attractiveness of the voices of the speakers of the opposite sex. If it had not been for the male English listeners' more positive judgements of the aestheticity of the female speakers' English, the English listeners' judgements of the male speakers' English would have been more positive than their judgements of the female speakers' English on all the scales, except the intonation scale.

Table 7 B. Female and male English listeners' judgements on female and male speakers' English; number of cases, means and standard errors (se) are given

	English Listeners											
	Female Speakers						Male Speakers					
	Female Listeners			Male Listeners			Female Listeners			Male Listeners		
	N	Mean	(se)	N	Mean	(se)	N	Mean	(se)	N	Mean	(se)
Aesthet.	302	3.86	(.08)	252	3.95	(.09)	354	4.02	(.07)	255	3.77	(.09)
Inton.	306	4.02	(.09)	258	4.17	(.10)	355	3.87	(.08)	256	3.86	(.10)
Compreh.	305	4.44	(.10)	261	4.41	(.12)	355	4.87	(.10)	263	4.77	(.12)
Standard	302	3.88	(.09)	260	4.02	(.10)	347	4.24	(.09)	259	4.26	(.11)

For Comprehensibility and Standard: Sex speakers effect (M>F) ($\chi^2 \geq 19.88$; df = 1; p< .01)
For Intonation: Sex speakers effect (F>M) ($\chi^2 = 9.25$; df = 1; p< .01)
For Aesthetics: Sex speakers x sex listeners effect ($\chi^2 = 7.46$; df = 1; p< .01)

3.8 Research question 1 and sub-questions a, b, c and d

On the basis of what appeared from tables 4 to 7 it was possible to answer research question 1 and sub-questions a, b, c and d.

1A: "Are Dutch listeners stricter in their judgements of English free speech produced by native speakers of Dutch than English listeners and are Dutch listeners stricter in their judgements of the reading of an English text by native speakers of Dutch than English listeners?" was answered in the affirmative as far as the aestheticity and intonation scales were concerned.

1B: "Are experienced Dutch listeners stricter in their judgements of English speech produced by native speakers of Dutch than experienced English listeners and are inexperienced Dutch listeners stricter in their judgements of English speech produced by native speakers of Dutch than inexperienced English listeners?" was answered in the affirmative as far as the intonation and comprehensibility scales were concerned.

1C: "Are Dutch listeners stricter in their judgements of English speech produced by male native speakers of Dutch than in their judgements of English speech produced by female native speakers of Dutch and are English listeners stricter in their judgements of English speech produced by male native speakers of Dutch than in their judgements of English speech produced by female native speakers of Dutch?" was only answered in the affirmative as far as the Dutch and English listeners' judgements on the intonation scales were concerned. Indeed, the opposite was found as far as the English listeners' judgements on the comprehensibility scale were concerned.

1D: "Are Dutch listeners stricter in their judgements of English speech produced by native speakers of Dutch of the same sex than in their judgements of English speech produced by native speakers of Dutch of the opposite sex and are English listeners stricter in their judgements of English speech produced by native speakers of Dutch of the same sex than in their judgements of English speech produced by native speakers of Dutch of the opposite sex?" was only answered in the affirmative as far as the English listeners' judgements on the aestheticity scale were concerned.

Question 1 "Are Dutch listeners stricter in their judgements of English speech produced by native speakers of Dutch than English listeners?" was therefore answered in the affirmative as far as the aestheticity, intonation and comprehensibility scales were concerned. It appeared that the main issue (the difference between Dutch and English listeners) was only obscured by the sex of the speakers and the interaction between the sex of the speakers and the sex of the listeners in a limited number of situations. Therefore it was felt to be unnecessary to take these factors into account in the further analysis.

3.9 Validity of Dutch listeners' judgements on the speakers' English

In order to determine if the Dutch listeners were fair in their judgements and if their judgements were valid in the sense that their ranking of the speakers was similar to that of the English listeners, correlations between the Dutch and the English listeners' judgements on the speakers' English were calculated. Results are presented in the middle column of table 8. In this table there are high correlations between judgements of English and Dutch listeners for intonation and comprehensibility (.91 and .90 resp.); speakers who were given a high ranking by Dutch listeners were also given a high ranking by English listeners. That these correlations are relatively high as far as comprehensibility is concerned may indicate that the ideas of the two group of listeners as to what intonation and comprehensibility are do not diverge too much. For standardness the correlation was .70, whereas for aestheticity a meagre correlation of only .53 was observed. There were considerable differences between Dutch and English listeners on these scales; speakers who were given a high ranking by one group of listeners were not necessarily given a high ranking by the other group. This may indicate that the ideas of the two groups of listeners as to what standardness and aestheticity are diverge considerably.

The second research question "Are Dutch listeners unfair in their judgements of English speech produced by native speakers of Dutch in the sense that there was a weak correlation between their judgements and the judgements of the Anglophone listeners?" was answered in the affirmative as far as the aestheticity and the standardness scales were concerned.

One plausible explanation for the differences between these groups of listeners might be the fact that only the Dutch listeners were influenced by factors that were present in the speakers' Dutch. In order to determine if this was so, correlations between the Dutch listeners' judgements on the speakers' English and the speakers' Dutch were calculated. Results are presented in the right-hand column of table 8.

Table 8. Correlation between Dutch and English listeners' judgements on the speakers' English and between Dutch listeners' judgements on the speakers' English and the speakers' Dutch

	Dutch and English listeners' judgements on the speakers' English	Dutch listeners' judgements on the speakers' English and the speakers' Dutch
Aestheticity	.53	.45
Intonation	.91	.46
Comprehensibility	.90	.34
Standardness	.70	.21

From the third column of table 8 it appears that for all the scales there was a significant correlation between the Dutch listeners' judgements on the speakers' Dutch and their judgements on the speakers' English, which correlation varied from .21 to .45. Although the height of these correlations did not invalidate the Dutch listeners' judgements, it might have influenced their judgements of the speakers' English considerably. In view of the significant correlations between Dutch listeners' judgements on the speakers' English and the speakers' Dutch it was decided to correct the Dutch listeners' judgements on the speakers' English for their judgements on the speakers' Dutch and to correlate these corrected judgements with the English listeners' judgements. In fact a multilevel model was estimated in which the Dutch listeners' judgements of English were corrected for their judgements of Dutch. The covariance coefficient between the English judges' judgements and the Dutch judges' judgements was used as a measure of "true" correlation. In fact this came down to calculating a partial correlation. Results are presented in table 9.

From table 9 it appears that there was a high correlation between the Dutch listeners' judgements that had been corrected for their judgements on Dutch and the English listeners' judgements for intonation and comprehensibility. Therefore there was hardly any difference between the judgements of these speakers by English listeners and Dutch listeners, after their judgements had been corrected. For the aestheticity and standardness scales, the corrected correlation was considerably higher

than the observed correlation. For aestheticity, however, the correlation was still rather low, which indicated that aestheticity as judged by English listeners is not the same as aestheticity as judged by Dutch listeners, although the two concepts have something in common.

Table 9. Corrected correlations between English listeners' judgements on the speakers' English and Dutch listeners' judgements on the speakers' English as corrected for their judgements on the speakers' Dutch

Scales	Corrected correlations
Aestheticity	.63
Intonation	1.00
Comprehensibility	.91
Standardness	.79

It is to be observed that the correlations between Dutch and English listeners' judgements on the speakers' English as corrected for the judgements on the speakers' Dutch were consistently higher than those that had not been corrected, although the difference was negligible as far as the judgements on the comprehensibility scale were concerned. This indicates that the Dutch listeners' judgements on the speakers' English had been influenced by factors that were present in the speakers' Dutch. Therefore the third research question "Are Dutch listeners prejudiced in their judgements of English speech produced by native speakers of Dutch?" was answered in the affirmative as far as the judgements on the aestheticity, intonation and standardness scale were concerned.

4. DISCUSSION

The first two research questions were addressed to confirm the findings of the previous experiment reported in the chapter A choice of judges. It will be seen that many of the results of this final experiment were similar to those of the previous experiment but that there were also a number of striking differences. As the differences and similarities may have been obscured by the clustering of a number of scales in the final experiment, which clustering was not felt to be advisable in the previous experiment, the correspondence of the scales in the two experiments is indicated in table 10.

The differences and similarities may also have been caused by the choice of the groups of judges, the correspondence of which is indicated in table 11. It can be seen that the composition of the groups of native-speaker judges in the final experiment was substantially different from that in the previous experiment because of the omission of the American judges. In addition it must be remembered that the native-speaker judges in the previous experiment were older than the majority of the na-

tive-speaker judges in the final experiment, and were, moreover, university students of linguistics in Scotland rather than secondary-school students in England, so that they may well have acquired a certain expertise in the judging of non-native speech. Although the composition of the groups of non-native listeners in the final experiment was less obviously different from that in the previous experiment, there were differences in the degree of experience.

Table 10. Correspondence of scales in previous (single scales) and final experiment (composite scales)

	Previous experiment	Final experiment
unpleasant – pleasant	x	
broad – cultured	x	
ugly – beautiful	x	aestheticity
slovenly – polished		
strong accent – no accent	x	
standard – non-standard		x
monotonous – melodious	x	intonation
expressionless – expressive		
unintelligible – intelligible		comprehensibility
inaccurate – precise		

Table 11. Correspondence of groups of judges in previous and final experiment

Groups of judges	Previous experiment	Final experiment
British	x	x
American	x	
Experienced Dutch	x	x
Inexperienced Dutch	x	x

4.1 Research question 1

As to the greater strictness of Dutch listeners, in the previous experiment it had been found that there were differences in strictness between British and Dutch listeners as far as the "pleasant" and the "beautiful" scales were concerned, the experienced native listeners being more positive than the non-native ones on both scales, the inex-

perienced Dutch listeners being more positive on the "pleasant" scale only. In the previous experiment greater lenience had been found as far as the inexperienced Dutch listeners' judgements on the "cultured" and "no accent" scales were concerned. For the sake of convenience the findings in the previous experiment on the relevant scales are presented in table 12.

Table 12. Presence of greater strictness and lenience, in experienced and inexperienced Dutch listeners' judgements than in British listeners' judgements as found in previous experiment

	Greater strictness	Greater lenience
Pleasant	Experienced Dutch, Inexperienced Dutch	
Beautiful	Experienced Dutch	
Cultured		Inexperienced Dutch
No accent		Inexperienced Dutch

In the final experiment greater strictness of Dutch listeners had been found on the aestheticity scale. Also on the intonation scale, where no significant differences between British and Dutch listeners had been found in the previous experiment, although the American listeners were much more positive than the inexperienced Dutch listeners, the British listeners were more positive than the Dutch listeners, which finding would appear to dispose of the notion that Dutch speakers' intonation of English strikes English listeners as unacceptable (cf. Willems, 1982). Therefore the results of the previous experiment can only be said to have been confirmed to a limited extent. Also on the comprehensibility scale, which was not present in the previous experiment, the British listeners were more positive than the Dutch listeners.

4.2 Research question 1a

As there had been no scale in the previous experiment that corresponded to the comprehensibility scale, the findings of the final experiment cannot be considered in the light of earlier findings. The result that the speakers were found more comprehensible when they read a text than in free speech cannot cause great surprise. Although this was not felt sufficient reason for keeping these tasks apart in the present experiment, it would have been advisable to do so in an experiment where the focus was on intelligibility; beginning learners might well have experienced much greater difficulty in making themselves understood without the support of the written text than the upper-intermediate to advanced learners in the present experiment.

4.3 Research question 1b

In the previous experiment substantial differences had been found between inexperienced and experienced listeners. In the final experiment, however, no such differences were found between inexperienced and experienced listeners and the findings of the previous experiment were not confirmed[51]. It is possible that real differences that would have been found if the results on the original scales had been analysed were obscured by the data reduction. A more plausible explanation could be found in the degree of experience of the non-native listeners in the two experiments. The experienced listeners in the previous experiment were teachers in higher education who, moreover, had taken part in training sessions together with native listeners and these sessions may well have resulted in greater similarity of their judgements to those of native listeners. Most experienced listeners in the second experiment were teachers in higher education but there were also teachers in secondary education. Moreover, these experienced listeners had not received specific training. As a result they may have been less experienced than the experienced listeners in the previous experiment. Also, the inexperienced listeners in the first experiment were students of English and another subject, whereas the inexperienced listeners in the second experiment were students of English only, who in theory had had twice as much time to devote to their English course as the first group. As a result they may well have been less inexperienced. Considerable differences in experience that existed between the experienced and inexperienced listeners in the previous experiment may have been much smaller in this final experiment.

4.4 Research question 1c

As far as the judgement on male and female speakers' English are concerned, the experienced Dutch listeners' judgements on female speakers' English were more positive on all the scales in the previous experiment as were the American listeners' judgements, whereas the British listeners' judgements of female speakers were more positive on the "beautiful" and "cultured" scales and the inexperienced Dutch listeners' judgement were more positive on the "melodious" scale only. In the final experiment all the listeners gave more positive judgements of the female speakers on the intonation scale. In this respect the findings of the previous experiment were confirmed. But the intonation scale was the only scale where the effect of speakers' sex was straightforward. On all the other scales there was an interaction effect of speakers' sex and listeners' nationality, the Dutch listeners giving more positive judgements about the female speakers than about the male speakers on the aesthetic-ity scale confirming the results of the previous experiment. As the comprehensibility and standard scale had not been used in the previous experiment, the findings in this experiment cannot be said to confirm earlier findings. The English listeners giving

---

[51] *This seems to be in line what was found by Cheng and Warren, who studied tutor and peer assessments of language proficiency and found few significant differences between mean tutor and peer judgements of oral production, although they found that the latter were in a narrower range (Cheng and Warren, 2005).*

more positive judgements about the male speakers than about the female speakers on the comprehensibility scale may well have been the result of the greater intelligibility of the male speakers; indeed, one male speaker happened to speak very clearly and it is surprising that the Dutch listeners did not register this; they may have been the victims of the halo effect that caused them to give higher judgements of female speech, whether this speech was in fact more comprehensible or not. A similar explanation could account for the differences in judgements on the standardness scale, where the English listeners were more positive about the male than about the female speakers.

4.5 Research question 1d

In the previous experiment the British female listeners had been found to be more positive about the English of speakers of their own sex than about that of speakers of the opposite sex as far as their judgements on the "pleasant" scale were concerned. Also in the final experiment an interaction effect between listeners' and speakers' sex was found with the English listeners. Unlike the previous experiment, the final experiment revealed that both the male and the female listeners were more positive about the English of speakers of the opposite sex, the male listeners even more so than the female. It would therefore appear that native speaker listeners are more susceptible to the interaction effect between listeners' and speakers' sex than non-native listeners. Why this effect affected both the male and the female listeners and why it took a different direction than in the previous experiment would be hard to explain. The obvious difference between the listeners in the first and those in the second experiment was the formers' greater experience, but it is hard to see why this greater experience could have caused this difference.

4.6 Research question 2

In the previous experiment a strong correlation between British and experienced Dutch listeners' judgements of the speakers' English had been found on the "pleasant", "beautiful" and "melodious" scales, whereas a weak correlation between British and inexperienced Dutch listeners' judgements had been found on the "pleasant" and "no accent" scales, which scales had become part of the aestheticity and intonation composite scales. In the final experiment, where experienced and inexperienced listeners were not kept apart, a weak correlation between English and Dutch listeners' judgements of the speakers' English was found on the aestheticity scale, which confirmed the findings as to the inexperienced Dutch listeners in the previous experiment; a strong correlation between English and Dutch listeners' judgements of the speakers' English was found on the intonation scale, which confirmed the findings as to the experienced Dutch listeners in the previous experiment.

That the correlation between the native and non-native listeners' judgements was strong on the intonation and comprehensibility scales need not give surprise. More surprising is the meagreness of the correlation between the native and non-native listeners' judgements on the aestheticity scale; in the previous experiment a correla-

tion as weak as these was only found between the British and the inexperienced Dutch listeners' judgements on one scale, the "no accent" scale, although an even weaker correlation was found between the American and the inexperienced Dutch listeners' judgements. An explanation for these findings might be that the experienced and inexperienced Dutch listeners were not kept apart for this analysis. A more plausible explanation, however, would appear to be that the native and non-native listeners in this final experiment used radically different dimensions to arrive at their judgements on aestheticity and to a lesser degree on standardness.

No unfairness was found as far as the comprehensibility scale was concerned.

4.7 Research question 3

Research question three did not address a problem that was raised in the previous experiment. The answer, therefore, cannot be said to confirm the findings reported in the previous chapter. Rather, it addressed the problem raised in the earlier chapter Perceived pronunciation quality in mother tongue and foreign language. In this chapter it was shown how judgements on the pronunciation of Dutch were compared with judgements on the pronunciation of English and how, in spite of the differences of the judges and of the speakers' tasks, a weak but significant correlation was found. Also in this final experiment judgements on the pronunciation of Dutch were compared with judgements on the pronunciation of English and a weak but significant correlation was found, which correlation was admittedly meagre on the aestheticity and intonation scales and almost negligible on the comprehensibility and standardness scales. In that sense the results of this final experiment could be said to confirm the earlier findings.

It appears that Dutch listeners are influenced to a highly limited extent by the quality of the speakers' Dutch in judging the comprehensibility of their English; this stands to reason; judgements on intelligibility can be fairly objective, and, although the Dutch listeners had shown themselves to be stricter than the English listeners, they had not shown themselves to be unfair in their judgements on this scale. This would suggest that these non-native listeners, if only they had not been too strict, would not have been less capable judges of comprehensibility than native listeners, which would be in line with the findings of Yoshida (2004), who found non-native listeners just as capable of assessing intelligibility as native listeners[52].

Dutch listeners appear to be influenced to a slightly less limited extent by the quality of the speakers' Dutch in judging the standardness of the speakers' English; this also stands to reason; apparently Dutch listeners have a fairly good notion of what a standard pronunciation of English is; although there may be considerable disagreement on the norm (British or American), apparently this norm has little – but certainly not nothing – to do with Dutch.

In the matter of aestheticity and intonation, however, Dutch listeners do not appear to be influenced by the quality of the speakers' Dutch to a negligible extent. Indeed, although the correlation is weak, there appears to be a perceptible influence.

---

[52] *Yoshida (2004, p. 96) stated that "significant differences in severity found were not peculiar to the raters' L1 background".*

If Dutch listeners are influenced by characteristics of the speakers' Dutch that they perceive in the speakers' English, it stands to reason that these are characteristics that they associate with varieties of Dutch, to the speakers of which they also attribute personal and social characteristics. If this is so, that would be a serious matter, which would suggest that Dutch listeners are not capable of assessing aestheticity, intonation and to a lesser extent standardness. If this is so, it would be unwise to ask Dutch listeners to indicate if the English pronunciation of speakers of Dutch is e.g. pleasant, beautiful or without accent, although they could be asked to indicate if it is e.g. intelligible.

5. CONCLUSION

The judgements of Dutch and English listeners of English speech produced by native speakers of Dutch were studied as well as the judgements of Dutch listeners of Dutch speech. It was shown that the Dutch listeners, whether they were experienced or inexperienced, male or female, were nearly always stricter than the English listeners in their judgements of the English speech, whether this speech was spontaneous or the reading of a text, of native speakers of Dutch, whether these speakers were male or female. This greater strictness concerned their judgements on the aestheticity, intonation and comprehensibility scales. It was also shown that the Dutch listeners could be called unfair. This unfairness concerned their judgements on the aestheticity and standardness scales. Finally it was shown that the Dutch listeners could be described as prejudiced. This prejudice concerned their judgements on the aestheticity, intonation and – to a lesser extent – standardness scales. No unfairness and prejudice were found as far as the comprehensibility scale was concerned. The answers to the three research questions are summarized in table 12.

Table 12. Presence of greater strictness, unfairness and prejudice in Dutch listeners' judgements of Dutch speakers' English

	Stricter	Unfair	Prejudiced
Aestheticity	x	x	x
Intonation	x		x
Comprehensibility	x		
Standardness		x	x

APPENDIX A

VARIANCE STRUCTURE

In order to determine if there were any differences between the Dutch and the English listeners as far as the variance between listeners and within listeners were concerned the variance structures were calculated. Results are given in table 13.

Table 13. Dutch and English listeners' judgements on the speakers' English in free speech and reading a text; variance and standard errors (se) are given

	Between Listeners				Within listeners			
	Dutch Listeners		English Listeners		Dutch Listeners		English Listeners	
	Free Speech S^2 (se)	Reading Text S^2 (se)	Free Speech S^2 (se)	Reading Text S^2 (se)	Free Speech S^2 (se)	Reading Text S^2 (se)	Free Speech S^2 (se)	Reading Text S^2 (se)
Aesthet.	0.18 (.05)	0.17 (.05)	0 (0)	0.09 (.03)	1.03 (.05)	1.17 (.06)	1.04 (.13)	1.12 (.05)
Inton.	0.41 (.11)	0.20 (.06)	0.07 (.10)	0.06 (.03)	1.49 (.08)	1.89 (.09)	1.61 (.21)	1.83 (.08)
Compreh.	0.29 (.08)	0.24 (.07)	0.21 (.18)	0.21 (.06)	1.47 (.08)	1.74 (.08)	1.75 (.23)	1.78 (.08)
Standard	0.49 (.12)	0.28 (.07)	0 (0)	0.17 (.05)	1.06 (.06)	1.29 (.06)	1.94 (.24)	1.24 (.06)

From this table it appears that the variance between listeners was smaller for the English listeners than for the Dutch listeners. This would indicate that there was greater agreement – sometimes even unanimity – between the native speakers than between the non-native speakers. This does not go so much for the comprehensibility scale. For the Dutch listeners variance tended to be greater where free speech was concerned, which might indicate that the non-native listeners disagreed even more on this task than on the reading of a text. There were no great differences in variance within listeners except where the standardness scale was concerned, the judgements of the English listeners on free speech showing greater variance.

The generalised stability coefficients were calculated for the four scales, for the Dutch and the English listeners, for free speech and for reading a text. Results are given in table 14. From this table it appears that the Dutch listeners were fairly reliable. The English listeners were fairly reliable as far as their judgements on the reading of texts were concerned, with the exception of the intonation scale, due to the relatively small variance between listeners (see table 4). As far as their judgements on free speech were concerned, the English listeners were not at all reliable, except for their judgements on the comprehensibility scale, which were fairly reliable and those on the intonation scale, which were somewhat reliable. The Dutch listeners were always more reliable than the English listeners.

Table 14. Generalised stability coefficients of Dutch and English listeners' judgements on the speakers' English in free speech and reading a text

	Dutch listeners		English listeners	
	Free speech	Reading a text	Free speech	Reading a text
Aestheticity	0.78	0.74	0	0.62
Intonation	0.85	0.68	0.47	0.40
Comprehensibility	0.80	0.73	0.71	0.70
Standardness	0.90	0.81	0	0.73

APPENDIX B

FACTOR ANALYSES

Table 15a. Rotated Factor Matrix for Dutch-Dutch (78% of variance explained);

	Aestheticity.	Intonation.	Comprehensibility.	Standardness
a1-pleasant	**.66**	.40	.19	.06
a2-cultured	**.85**	.14	.19	.11
a3-beautiful	**.79**	.35	.23	.02
a4-polished	**.71**	.21	.40	.01
s5-no accent	**.79**	.44	.08	.21
s6-standard	.18	.00	.06	**.98**
e7-melodious	.22	**.88**	.21	.02
e8-expressive	.23	**.88**	.21	.02
e9-intelligible	.22	.21	**.86**	.05
e10-precise	.36	.24	**.79**	.06
s11-distinguished	**.69**	.16	.27	.08

Table 15b. Rotated Factor Matrix for Dutch-English (75% of variance explained)

a1-pleasant	**.74**	.20	.36	.16
a2-cultured	**.77**	.24	.18	.02
a3-beautiful	**.78**	.24	.33	.16
a4-polished	.54	**.58**	.17	.19
s5-no accent	**.71**	.10	.02	.20
s6-standard	.16	.13	.016	**.94**
e7-melodious	.20	.22	**.89**	.01
e8-expressive	.25	.18	**.89**	.02
e9-intelligible	.16	**.81**	.20	.05
e10-precise	.29	**.81**	.16	.11
s11-distinguished	**.60**	.38	.16	.12

Table 15c. Rotated Factor Matrix for English-English (72% of variance explained)

a1-pleasant	**.75**	.10	.38	.11
a2-cultured	**.75**	.10	.30	.04
a3-beautiful	**.78**	.24	.16	.13
a4-polished	**.65**	.37	.26	.17
s5-no accent	**.62**	.38	.04	.24
s6-standard	.17	.18	.16	**.91**
e7-melodious	.34	.22	**.75**	.18
e8-expressive	.23	.23	**.81**	.06
e9-intelligible	.23	**.73**	.23	.25
e10-precise	.19	**.81**	.09	.24
s11-distinguished	.22	**.73**	.28	.15

The above tables reveal factor loadings that are similar for the three listening situations with the notable exception of scale 11; it is obvious that the English scale had a different meaning for the listeners than the Dutch scale. It is curious that scale 4 loaded on factor 2 in the Dutch – English listening situation, whereas it loaded on factor 1 in the other situations; as the Dutch listeners used the same scales in both situations, it cannot have been a matter of language. The difference, however, was so small as to be almost negligible.

The tables do not reveal two factors that one might expect: there are no factors that could be associated with status and affect or solidarity. Apparently the listeners were more concerned with matters like appreciation of voice and pronunciation, comprehensibility and intonation than with the attribution of characteristics of personality.

As many of the scales used in the second experiment were identical to or had been adapted from the scales used by Fagel et al. (1983), factor loadings obtained were compared. In table 16 loadings on the first two factors are shown (five resp. four factors had been found).

Table 16. Varimax rotated factor loadings > 45 on the first two factors as found by Fagel et al. and varimax rotated loadings on the first two factors as found in the second experiment

	Fagel et al.		2nd experiment	
	F1	F2	F1	F2
unpleasant – pleasant	x		x	
broad – cultured		x		
ugly – beautiful	x		x	
slovenly – polished		x	x	
strong accent – no accent		x	x	
monotonous – melodious	x			x
expressionless – expressive	x			x
inaccurate – precise.		x		
common – distinguished	x		x	

From the above table it is obvious that the listeners in the present study arrived at different dimensions. Fagel et al. (1983) described the first factor as "melodiousness" and the second as "articulation quality" (p. 325). Such a description would have been inappropriate here. It must be remembered that the conclusions in Fagel et al. (1983) were based on Dutch texts read out by five male and five female speakers. The ratings were done by students of speech therapy. There can be little doubt that these listeners applied different dimensions than the listeners in the present study, most of whom were aspiring or experienced foreign-language teachers who, moreover, listened to L1 and L2 speech.

SUMMARY AND CONCLUSION

The studies presented here were prompted by my concern that no justice was done to speakers of a non-standard variety of Dutch when their pronunciation of English was judged by Dutch listeners. While addressing this problem I raised a large number of questions – perhaps too large a number – which may have obscured the line of my argument. This final chapter therefore starts with the briefest possible summary, in which I tried to reduce the argument to its barest essentials. This is followed by a fuller summary of the argument, which for the sake of brevity is restricted to the answers to the research questions, which answers were simplified, where necessary. This is followed by a discussion of the limitations of these studies, which discussion, naturally, leads to a consideration of the desirability of further research. Recommendations for the teaching of English in the Netherlands are followed by a brief epilogue.

1. SUMMARY

I suspected that Dutch listeners gave more negative judgements of the English pronunciation of native speakers of a non-standard variety of Dutch than of the English pronunciation of native speakers of a standard variety of Dutch. This may have been caused by the greater difficulties that native speakers of a non-standard variety of Dutch may have in acquiring the pronunciation of English than speakers of the standard variety of Dutch. If that is so – and one of the findings of the preliminary studies was that there were indications that this was so – no fault could be found with the Dutch listeners' judgements. But another reason may have been that the Dutch listeners' judgements were influenced by characteristics of the speakers' Dutch, which the listeners perceived in the speakers' English. If this is so – and the most salient finding of the empirical studies was that there were indications that this influence may have affected their judgements of the aestheticity, intonation and standardness of the Dutch speakers' English – fault could be found with the Dutch listeners' judgements. I also found that the Dutch listeners were often stricter than native speaker listeners when judging the aestheticity, intonation and comprehensibility of the Dutch speakers' English. I therefore expressed serious doubts about Dutch listeners' ability to produce trustworthy judgements about the pronunciation of English of other Dutch persons.

1.1 Summary of Perceived pronunciation quality in mother tongue and foreign language

In the chapter Perceived pronunciation quality in mother tongue and foreign language I introduced the problem that fragments of English speech produced by

speakers of a non-standard variety of Dutch received lower rankings than those produced by speakers of the standard variety. First I tried to establish if this problem really existed or was a mere figment of my imagination. Having shown evidence suggesting that the problem did in fact exist I came up with three possible explanations.

- For speakers of non-standard varieties of the mother tongue it might be more difficult to produce the sounds of the L2 than for speakers of the standard variety because the phoneme inventory of their variety lacks certain sounds that are present in that of the standard variety.
- Non-native speaker listeners who recognize features in the L2 of non-native speakers that are characteristic of certain non-standard varieties – not of the L2 but of the mother tongue – might therefore be inclined to assign a lower ranking to these speakers' L2.
- Speakers who have acquired a high level of general phonetic competence might therefore be able to pronounce both the mother tongue and the L2 well, whereas speakers who have not reached such a high level might pronounce both the mother tongue and the L2 less well.

In the following studies I tried to present evidence for the first two explanations.

1.2 Summary of Amsterdam English

In the chapter Amsterdam English I reformulated the first possible explanation as the first research question.

> *Is the phonological interference of an Amsterdam variety of Dutch a more serious obstacle in the acquisition of the pronunciation of English than the phonological interference of the standard variety of Dutch?*

The pronunciation of lower-secondary school pupils was observed and five male pupils were asked to read an English text, which was recorded. The observations and the study of the recording suggested that speakers of Amsterdam Dutch would have more serious problems with the English consonants than speakers of the standard variety of Dutch but that they might have fewer problems with the vowels. As problems with consonants have been found to form greater obstacles to intelligibility than problems with vowels, the research question could therefore be answered in the affirmative: it might well be a fact that speakers of an Amsterdam variety of Dutch were less likely to pronounce English well than speakers of the standard variety of Dutch; therefore no blame could be laid on the Dutch listeners, when they judged these Amsterdam students to pronounce English less well than speakers of the standard variety.

1.3 Summary of A Choice of judges

In spite of the answer to my first research question I was still not convinced that the Dutch listeners' judgements were to be trusted; to test their trustworthiness I needed

judgments by native speakers. I addressed the second explanation in the chapter Judgements on Dutch and Dutch English. Before I could address this second explanation, however, I needed to solve the problem of the choice of judges; what native speakers would provide trustworthy judgements? How this choice was made was reported in the chapter A Choice of judges.

I asked four groups of listeners – two Anglophone groups, British and American, and two Dutch groups, experienced and inexperienced – to judge the English produced by nine Dutch speakers on five seven-point scales, four scales related to the aestheticity of the pronunciation and one scale relating to the intonation. I considered differences in ranking and strictness as well as differences in strictness towards male and female speakers and differences in strictness of male and female listeners. These were the results.

Experienced Dutch listeners vs. British listeners. The experienced Dutch listeners did not differ from the British listeners in the way they ranked the speakers but they were stricter on two of the aestheticity scales; also they were more positive of female speakers than of male speakers on two aestheticity scales and on the intonation scale, whereas the British listeners' judgements of male and female speakers did not differ.

British vs. American listeners. The British listeners differed from the American listeners in their ranking of the speakers on one aestheticity scale. They were not stricter but they differed from the American listeners, who were more positive of the female speakers on two aestheticity scales and on the intonation scale

Experienced vs. inexperienced Dutch listeners. The experienced Dutch listeners differed from the inexperienced Dutch listeners in their ranking of the speakers on two aesthetictiy scales. They were stricter than the inexperienced Dutch listeners on three aesthetictiy scales. They differed from the inexperienced Dutch listeners in their judgements of female speakers on all four aestheticity scales.

Inexperienced Dutch vs. American listeners. The inexperienced Dutch listeners differed from the American listeners in their ranking of the speakers on all the scales. (This was the only instance of a difference in ranking on the intonation scale.) They differed from the American listeners in strictness on two aestheticity scales. They differed from the American listeners in their judgements of female speakers on all four aestheticity scales.

The experienced Dutch vs. American listeners. The experienced Dutch listeners differed from the American listener in their ranking of the speakers on only one asetheticity scale, which was also the case for their strictness. They differed from the American listeners in their judgements of female speakers on none of the scales.

The inexperienced Dutch vs. British listeners. The inexperienced Dutch listeners differed from the British listeners in their ranking of the speakers on two aestheticity scales. They differed from the British in strictness on three aestheticity scales. They differed from the British listeners in their judgements of female speakers on one aestheticity scale.

Most of the differences were found on aestheticity scales, with very few differences on the intonation scale.

Most of the evidence seemed to be in favour of the British listeners; they were more reliable than the American listeners and agreed with the experienced Dutch listeners as far as the ranking of the speakers was concerned; but they did not differ much in strictness from the other groups, except from the experienced Dutch listeners, and only on some scales. True, the British listeners did not often show themselves more positive about the female speakers than about the male speakers, as the experienced Dutch and the American listeners did, and the British male listeners were less positive of the speakers of their own sex, whereas the other groups were not. But this only occurred on one scale. On balance, therefore, there seemed to be slightly more arguments to find in favour of the British listeners.

During the search for the best non-native speaker judges, however, it emerged that there were considerable differences between the inexperienced and the experienced Dutch listeners. The magnitude of the differences between experienced and inexperienced Dutch listeners was such that I thought it inadvisable not to proceed with these two groups. The differences between these groups of Dutch listeners was one of the topics that I had to address in the final experiment.

1.4 Summary of Judgements on Dutch and Dutch English

In the chapter Judgements on Dutch and Dutch English I reformulated the second explanation in the form of three research questions as follows.

1) Are Dutch listeners stricter in their judgements of English speech produced by native speakers of Dutch than English listeners?
2) Are Dutch listeners unfair in their judgements of English speech produced by native speakers of Dutch in the sense that there is a weak correlation between their judgements and the judgements of the Anglophone listeners?
3) Are Dutch listeners prejudiced in their judgements of English speech produced by native speakers of Dutch?

Although I had solved some of the problems concerning the choice of native-speaker judges, namely the choice between British and American judges, I had not solved all the problems; the problem of the choice between male and female listeners remained. Moreover there were problems concerning the Dutch judges, concerning the speakers and concerning their tasks. To confirm the findings of the first experiment and to answer the first new research question I had to divide the first research question in four sub questions.

- In the first sub question I distinguished the two tasks: the reading of a text and the production of free speech.

- In the second sub question I segregated experienced and inexperienced listeners.
- In the third sub question I separated male and female speakers.
- In the fourth sub question I discriminated between listeners to speakers of the same sex and listeners to speakers of the opposite sex.

Twenty Dutch speakers produced two fragments of free speech and read out two texts, both in Dutch and in English. I asked some 140 Dutch and English listeners, both experienced and inexperienced, to judge these fragments on eleven seven-point scales, relating to aestheticity, intonation, comprehensibility and standardness.

These were the answers to the four sub questions.

Free speech and reading a text. The Dutch listeners were stricter than the English listeners in their judgements of free speech and the reading of a text as far as the aestheticity and intonation scales were concerned.

Experienced and inexperienced listeners. The experienced Dutch listeners were stricter in their judgements of English speech produced by native speakers of Dutch than the experienced English listeners and the inexperienced Dutch listeners were stricter in their judgements of English speech produced by native speakers of Dutch than the inexperienced English listeners as far as the intonation and comprehensibility scales were concerned.

Male and female speakers. The Dutch listeners were stricter in their judgements of English speech produced by the male native speakers of Dutch than in their judgements of English speech produced by the female native speakers of Dutch and the English listeners were stricter in their judgements of English speech produced by the male native speakers of Dutch than in their judgements of English speech produced by the female native speakers of Dutch only as far as their judgements on the aestheticity and intonation scales were concerned. Indeed, the opposite was found as far as the judgements on the comprehensibility scale were concerned.

Listeners to speakers of the same sex and listeners to speakers of the opposite sex. The English listeners were stricter in their judgements of English speech produced by native speakers of Dutch of the same sex than in their judgements of English speech produced by the native speakers of Dutch of the opposite sex as far as the English listeners' judgements on the aestheticity scale were concerned.

The greatest number of differences were found on the aestheticity scale and the smallest number of the comprehensibility scale. The answers to the three main research question were as follows.

The Dutch listeners were

- stricter in their judgements of English speech produced by the native speakers of Dutch than the English as far as the aestheticity, intonation and comprehensibility scales were concerned.
- unfair in their judgements of English speech produced by the native speakers of Dutch in the sense that there was a weak correlation between their judgements and the judgements of the Anglophone listeners as far as the aestheticity and the standardness scales were concerned.
- prejudiced in their judgements of English speech produced by the native speakers of Dutch as far as the judgements on the aestheticity, intonation and standardness scale were concerned.

The research questions relating to the second explanation had been answered in the affirmative on most of the scales, with the exception of the comprehensibility scale. Therefore Dutch listeners could be said to be too strict, but not to be unfair or prejudiced in their judgements of Dutch speakers' comprehensibility.

The research questions relating to the first and the second possible explanations having been answered and serious doubt having been cast on Dutch listeners' trustworthiness, I felt no need for a further exploration of the third possible explanation: the higher or lower level of general phonetic competence.

2. LIMITATIONS

Before considering the implications of the results of the three studies reported here, their limitations need to be recognized. These limitations were the result of the choice of speakers, that of listeners, that of tasks and the explanations that were addressed.

2.1 Limitations resulting from the choice of speakers

The limitations concerning the choice of speakers appear from table 1.

Table 1. Speakers in the three studies

Speakers	*Number*	*Educational level*
Amsterdam English	(n=5; 5 male)	Lower secondary
First experiment	(n=9; 5 male, 4 female)	1st-year higher
Final experiment	(n=20; 10 male, 10 female)	1st-year higher

The first study concerning the disadvantages of speakers of Amsterdam Dutch in the acquisition of the pronunciation of English was based on speech produced by no more than five male students in a non-academic type of lower-secondary school. It could be argued that the results are not valid for female students, students in lower-

secondary schools of a more academic type, let alone for students in upper-secondary schools or the students in higher education, whose problems prompted this study.

The second and third studies were based on speech produced by 1st-year students of English in higher education; also in view of the fact that differences were found between the results of these studies – which differences were partly explained by reference to the different curricula followed by the students – it could be argued that these students were in no way representative for other students in higher education. Moreover, the speakers in the first experiment had been selected on the basis of the marks that they had been given, possibly by teachers who were amongst the judges, whereas the speakers in the final experiment had been selected by external experts, so that it could be argued that no objections could be raised against the choice of speakers in this final experiment.

The differences between the speakers did not only concern the levels. Also the geographical origins differed. Whereas the speakers in Amsterdam English were all Amsterdam persons, little is known about the geographical origin of the speakers in the second study; although it was assumed that most of them were residents of Amsterdam or North Holland, this had not been ascertained; in the final experiment, where the speakers' pronunciation had been judged by expert listeners as either standard or non-standard, the speakers were from Amsterdam, North Holland but also from South Holland and Utrecht. Unfortunately, little could be done with the information that was available about their geographical origin[53].

Another serious objection concerns the time elapsed between the three studies. Whereas the Amsterdam speakers produced their speech in the late seventies, the speakers in the first experiment produced theirs in the early eighties and those in the final experiment in the nineties. In how far could what was found concerning these speakers have any relevance in the twenty-first century? Is it not a fact that the Dutch language has undergone changes in the last decades, some of which changes could have made the standard accent of Dutch more similar to Amsterdam Dutch and would attitudes towards Amsterdam Dutch not have become more positive as a result? These are questions that were not answered.

Even more seriously, it might be objected that radical changes in the composition of the population of West Amsterdam – influx of speakers of other languages than Dutch as a result of external immigration, of speakers of other non-standard varieties of Dutch than Amsterdam Dutch as a result of internal immigration and of speakers of standard Dutch as a result of gentrification – may have caused speakers of West Amsterdam Dutch to disappear, so that the problem solved itself. But such objections were not borne out by the data on the ethnic composition of the West Central Amsterdam ("Oud West") population provided by the city of Amsterdam (Dienst Onderzoek en Statistiek, 2006), which data, admittedly, did not include information on linguistic background. Also, it was the author's experience, anecdotal though this evidence may be, that a broader variety of Amsterdam Dutch could often be heard

---

[53] *No significant differences were found between judgements on the speech produced by speakers from the three areas: Amsterdam, North Holland and South Holland plus Utrecht.*

from speakers whose first language was not Dutch than from native speakers of Dutch.

2.2 Limitations resulting from the choice of listeners

The limitations resulting from the choice of listeners can become clear from table 2.

Table 2. Listeners in the three studies

Study	*Non-native speaker*	*Native speaker*
Amsterdam English	Experienced Dutch (n=1)	
First experiment	Experienced Dutch (n=9), Inexperienced Dutch (n=21)	British undergraduate (n=20), American secondary-school student (n=20)
Final experiment	Experienced Dutch (n=13), Inexperienced Dutch (n=70)	British secondary-school student (n=54), British teacher (n=2)

As can be seen from table 2, a serious omission in the first study was the absence of native-speaker listeners, whose judgements must have been more trustworthy than those of non-native speaker listeners, if the findings of the second and third study reported here are to be believed. Moreover, in the first experiment the native speaker-listeners were perhaps not well matched. The British undergraduate students of linguistics must have been different from the American secondary-school students in other respects than nationality. No such objections could be raised against the choice of native-speaker listeners in the final experiment; indeed British secondary-school students as listeners to trainee secondary-school teachers would seem to be a natural choice; an objection could, however, be made against the small number of British teachers.

As with the speakers, it might be objected that changes in the composition of the population of the Netherlands may have altered listeners' attitudes towards varieties of Dutch; it is not unthinkable that the influx of large numbers of speakers of other languages has made Dutch listeners more tolerant or less tolerant of deviations from the standard variety.

2.3 Limitations resulting from the choice of speakers' tasks

Limitations resulting from the choice of tasks may be clear from table 3. As appears from table 3, it was only in the final experiment that the speakers were asked to perform two tasks. Although the findings of the final experiment suggested that there were no differences in judgements between the two tasks, it must be borne in mind that this concerned tasks at CEFR level B2. At level A2, at which level the speakers in the first study might with hindsight have been supposed to aim, although several speakers had certainly not reached that level, there would probably have been much greater differences. But even if the two tasks were equal, it is doubtful if what was

found in these contextualised activities, would also hold for realistic or even real tasks.

Table 3. Speakers' tasks in the three studies

Study	*Reading text*	*Semi-spontaneous speech*
Amsterdam English	x	
First experiment		x
Final experiment	x	x

The tasks related to spoken production and did not relate to spoken interaction and listening, let alone to reading and writing. But it is not impossible that speakers of a non-standard variety who have difficulties in producing the sounds of English would also experience problems in perceiving these sounds. This potential problem in perception was not touched upon.

The speaking tasks only concerned English and Dutch. But there is no reason to assume that what was found for learners of English could not possibly be true for learners of other language such as French, German and Spanish. There was no attempt to even consider this possibility.

2.4 Limitations resulting from the number of explanations addressed

In the three studies only two of the three possible explanations were worked out; after the research questions based on the two explanations had been answered, it was felt that there was no need to study the third explanation. It may be objected that the answers only accounted for part of the problem and that the differences in general phonetic competence could also contribute to problems experienced by speakers of non-standard varieties of Dutch in their acquisition of the pronunciation of English. Also it could be objected that no attempt has been made to estimate the relative importance of the linguistic and sociolinguistic factors, let alone the relative importance of the differences in general phonetic competence.

As a result of the limitations doubts may be entertained about the findings of the first two studies, in particular where they contradict those of the final study. As the objections to the final experiment must be less serious, greater credence should be attached to the findings of the final experiment, also where these findings differed from those of the first experiment.

This account of the limitations of the studies presented here points to the need of further study to replicate and to supplement what had been done. The limitations of these studies, however, do not seem to be so serious that it would be totally presumptuous to base recommendations for language teaching on the findings, in particular where these findings confirm those of other studies.

3. SUGGESTIONS FOR FURTHER STUDY

The limitations of the studies suggest topics that could be fruitfully explored by students with similar and related interests.

As far as the speakers were concerned, samples could be selected of speakers with different geographical origins and with different levels of expected proficiency. The focus of the first study e.g. was very narrow. The speakers in the first study were male speakers of a western variety of Amsterdam Dutch in a lower-secondary school. It would be of great interest to know if this variety of Dutch is still found in lower-secondary school children. If it has not disappeared, female speakers could also be taken into account. But the choice in favour of Amsterdam speakers was relatively arbitrary in that it was motivated by the location of the teacher education institute. Other varieties of Dutch, not only in the Netherlands but also in Flanders, not only native but also non-native, might provide interesting material. The focus of the second and the final study was less narrow geographically but just as narrow as far as educational level was concerned (only 1st-year higher education students). It would be worthwhile to replicate these studies with speakers that are representative of each of the five relevant CEFR proficiency levels, A1, A2, B1, B2 and C1.

As far as the native listeners were concerned, it had been decided to abandon the American listeners, although clear differences had been found between these American listeners and the British listeners. These differences could be explored in much greater detail, and other major groups of native speakers could be taken into account[54]. Also, the selection of the native-speaker listeners could be related more clearly to the learners' future audience. For future secondary-school teachers, e.g. British secondary-school students would appear to be able to yield more valid judgements than students of linguistics.

As far as the non-native listeners were concerned, the choice of Dutch listeners seemed to be the obvious one but there is no reason why non-native listeners who do not speak Dutch, e.g. French or German listeners, should not be chosen; in fact non-native listeners who do not share the speakers' language are likely to be less liable to prejudice and thus preferable as judges. The inexperienced Dutch listeners in these studies were 1st-year students in higher education. But would it not have been more sensible to ask secondary-school students to judge their future teachers' pronunciation?

Of these suggested topics for further study, the most urgent topic would seem to be that of Dutch listeners' judgements of the English of speakers of non-native varieties of Dutch. In earlier studies no differences had been found between native and non-native speakers of Dutch as far as their ability to acquire foreign languages was concerned (see Van Gelderen, 1995). But this study used interviews with teachers and analyses of protocols of reading comprehension rather than spoken production and spoken interaction and it cannot be assumed that what was found for the receptive written skill would necessarily apply to the productive oral skill. Also it is hard to believe that what was found by Doeleman (1998) about the reactions of secon-

---

[54] *Van den Doel (2006) found Australians, New Zealanders and South Africans more lenient than respondents from the Northern hemisphere (p. 107).*

dary-school students to non-native Dutch could be entirely untrue of reactions to the foreign language speech produced by such speakers.

As far as the tasks were concerned, the findings of the final experiment suggested that there was little to choose between semispontaneous speech and the reading of a text. But even the semispontaneous speech could hardly be said to constitute spoken interaction. Therefore it would seem worthwhile to find out if tests with more realistic or even real tasks (e.g. performance tests like the one described in Lynch & McNamara, 1998) would yield results similar to the reading of a text. It would also seem worth exploring the possibility that problems in production could have parallels in problems in perception.

4. RECOMMENDATIONS FOR THE TEACHING OF ENGLISH IN THE NETHERLANDS

At first sight the findings of these studies seem to cast serious doubt on the ability of Dutch listeners, whether these listeners are teachers or teacher trainees, to judge Dutch speakers' pronunciation of English. But it would be a mistake to conclude from the findings that Dutch listeners should not be allowed to judge Dutch speakers' English at all. If the remedies suggested by De Groot (1961) – reduction of the judges' tasks and concentration on the relevant aspects – are applied consistently, Dutch listeners might well be capable of producing valid and reliable judgements. Consideration of the educational levels of the speakers in these studies and their expected proficiency levels would assist in such a reduction and concentration. In table 4 an attempt is made to relate the levels that the speakers might have been expected to reach to those of the CEFR.

Table 4. Educational and expected proficiency levels of speakers

Speakers	*Educational level*	*CEFR*
Amsterdam English	Lower secondary	A2
First experiment	1st-year higher	B2
Final experiment	1st-year higher	B2

For the lower-secondary school students, intelligibility would have been all that could be required. For the 1st-year students in higher education, however, a clear and natural pronunciation and intonation would have been required in addition to intelligibility.

There seems to be little doubt that the lower-secondary school students studied would have had considerable difficulty in making themselves understood, both to native and non-native listeners. The two groups of 1st-year students in higher education by contrast would most certainly have been intelligible, the last group almost certainly highly intelligible. Native and non-native listeners to English speech pro-

duced by Dutch learners agreed on this point. But there seemed to be very little agreement on aspects beyond intelligibility. True, in the first experiment native and non-native listeners agreed on the speakers' intonation. In the second experiment, however, they did not.

It would appear that Dutch listeners, whether they are experienced or inexperienced, would be capable of judging intelligibility, if only they would be more lenient. As intelligibility appears to be all that can be required from learners aiming for the basic proficiency levels and as these learners are found in primary and lower-secondary schools, it would seem the safest policy to restrict judging by Dutch listeners to learners whose expected proficiency level is not expected to go beyond A2.

Although learners aiming for level B1, which learners might be found in upper-secondary school and in further education, are allowed to show only occasional evidence of a foreign accent and to make only occasional mispronunciations, Dutch listeners could still judge their pronunciation, provided that they do not attempt to judge other aspects than intelligibility and are cautioned not to be too strict. They should definitely not be asked to form aesthetic judgements or to pronounce on the presence of an accent.

It would not appear to be a good policy to ask Dutch listeners to judge the pronunciation of learners aiming at the higher proficiency levels, which learners would presumably be found in higher education. One can only agree with the Spanish higher education students studied by Lasagabaster and Sierra (2002) when they expressed a strong preference for native-speaker listeners as judges of their pronunciation. It might be argued that after a number of training sessions, similar to the ones attended – apparently with some success – by the experienced listeners in the first experiment, non-native listeners might become better judges, but the fear that they would revert to their old ways would always exist (cf. Lumley & McNamara, 1995). In their students' interest institutes of higher education in the Netherlands ought, therefore, to be strongly urged to follow the example of such professional language testers as Cambridge ESOL and to employ native speakers only as examiners of spoken English.

Learners who are speakers of a non-standard variety of Dutch may suffer from a double disadvantage: because of their accent in Dutch they may have to overcome greater obstacles in acquiring the pronunciation of English and because of their accent they may be treated more harshly by their Dutch teachers than speakers of the standard variety. There is every reason to exercise extra caution when judging the pronunciation of such learners.

5. EPILOGUE

When discussing the limitations of these studies, I alluded to the changes in the composition of the Amsterdam population, the extent of which might lead some readers to dismiss my first study as irrelevant. It is perhaps appropriate that I should end with some remarks about these changes, which I could not pretend to have no bearing on my work.

While reviewing my study on Amsterdam English, I tried to visit the school where I made the recordings. I found that the 4th LTS has disappeared; the school building, a striking example of Amsterdam School architecture, had first been used as a school for silversmiths but had then been turned into a multi-cultural centre. The building of the teacher-education institute where I made the recordings used in the first experiment faced the Central Amsterdam Jordaan area, from which it was separated by the water of one canal only. It was less than one hundred metres from the Westertoren, a landmark overlooking the Jordaan. Through its location the institute seemed to assert its Amsterdam identity. The building where the students listened to the recordings used in the final experiment was located in Central Amsterdam. The institute had, by that time, been merged with other institutes of higher education and its original name must have long been forgotten.

At the moment of writing the institute was located in East Amsterdam but it was to be moved to an anonymous industrial estate in South East Amsterdam. It still bore the name of the city but, when it asserted its Amsterdam identity, it was necessarily a very different identity.

A similar tale could be told about my primary and secondary schools. My primary school disappeared completely. The building, together with those of the neighbouring schools for girls and infants and the church, was demolished to make way for new housing. The secondary school, which was only separated by a canal from a traditional Amsterdam working-class neighbourhood, was moved to a new location, far away from any working-class areas.

Although the Amsterdam speech that surrounded me when I grew up could still be found in some areas, mainly from older people, it had been replaced with radically different speech in other areas. Of course I knew that, as society changes, its language must necessarily change. But I could not help regretting the gradual disappearance of a vernacular variety that for me had so many associations with such positive qualities as friendliness and humour.

REFERENCES

Alberdingk Thijm, J.A., & Van Lennep, W.W. (1885). Het tegenwoordig Amsterdamsch, volgens nasporingen. *Onze Volkstaal, 2*, pp. 121-136.

Albrechtsen, D., Henriksen, B., & Færch, C. (1980). Native speaker reactions to learners' spoken interlanguage. *Language Learning*, *30*, pp. 365-396.

Andersen-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, *42, 4*, pp. 529-555.

Barry, W. (1974). Language background and the perception of foreign accent. *Journal of phonetics*, *2*, pp. 65-89.

Berns, J. (2002). *Amsterdams*. Den Haag: Sdu Uitgevers.

Binnenpoorte, D., Van Bael, C., Den Os, E., & Boves, L. (2005). Gender in everyday speech and language; a corpus-based study. *Interspeech 2005*, pp. 1-4.

Bliss, A.J. (1979). *Spoken English in Ireland 1600 – 1740*. Dublin: Dolmen Press.

Blom, J. & Van Herpt, L. (1976). The evaluation of jury judgements on pronunciation quality. *Proceedings Inst. Phon. Amsterdam*, *4*, pp. 31-47.

Bonk, W. J., & Ockey, G. J. (2003). A many-facet Rasch analysis of the second language group oral discussion. *Language Testing*, *20*, pp. 89-110.

Bongaerts, T. (1999). De keuze van beoordelaars in onderzoek naar de uitspraakvaardigheid in een vreemde taal. In M. Gerritsen, & D. Springorum (eds), *Een bundel bedrijfscommunicatie voor Ger Peerbooms bij gelegenheid van zijn 65e verjaardag*. Nijmegen: Nijmegen University Press.

Bongaerts, T., Van Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in second language acquisition*, *19*, pp. 447-465.

Booij, G. (1995). *The Phonology of Dutch*. Oxford: Clarendon Press.

Boves, L., Fagel, W., & Van Herpt, L. (1982). Opvattingen van vrouwen en mannen over de spraak van mannen en vrouwen. *De Nieuwe Taalgids, 75*, pp. 1-23.

Broeders, A. (1981). Engels in Nederlandse oren: uitspraakvoorkeur bij Nederlandse studenten Engels. *Toegepaste taalwetenschap in artikelen*, 9, pp. 127-138.

Brown, A. (1995). The effect of rater variables in the development of an occupation-specific language performance test. *Language Testing*, *12*, pp.1-15.

Brouwer, D. (1989). *Gender variation in Dutch: a sociolinguistic study of Amsterdam speech*. Dordrecht: Foris.

Bürki-Cohen, J., Miller, J., & Eimas, P. (2001). Perceiving non-native speech. *Language and speech, 44*, *2*, pp.149-169.

Cambridge ESOL (2003a). *CAE Handbook for the Certificate in Advanced English*. Cambridge: Cambridge ESOL.

Cambridge ESOL (2003b). *CPE Handbook for the Certificate of Proficiency in English*. English. Cambridge: Cambridge ESOL.

Canagarajah, S. (2005). Editor's note. *TESOL Quarterly, 39, 3*, pp. 365-366.

Celce-Murcia, M., Brinton, D.M., & Goodwin, J.M. (1996). *Teaching pronunciation: a reference for teachers of English to speakers of other languages*. Cambridge: Cambridge University Press.

Cheng, W., & Warren, M. (2005). Peer assessment of language proficiency. *Language testing*, *22, 1*, pp. 93-121.

Chiswick, B., & Miller, P. (2004). Linguistic distance: a quantitative measure of the distance between English and other languages. *ICA DP*, 1246.

Cohen, A., Ebeling, C., Fokkema, A., & Van Holk, A. (1961). *Fonologie van het Nederlands en het Fries*. (2nd ed.). Den Haag: Nijhoff.

Collins, B. (1979). Hierarchy of Errors of Dutch Learners. Paper presented at the Second International Conference on the Teaching of Spoken English, Leeds.

Collins, B., & Mees, I. (1981). *The Sounds of English and Dutch*. The Hague: Leiden University Press.
Collins, B., & Mees, I. (2003). *The Phonetics of English and Dutch.* (5th ed.). Leiden: Brill.
Cook, V. (1999). Going beyond the native speaker in language teaching. *TESOL Quarterly, 33*, pp. 185-209.
Corder, S. P. (1981). *Error Analysis and Interlanguage*. Oxford: Oxford University Press.
Council of Europe (2001). *Common European framework of reference for languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.
Cunningham-Andersson, U. (1996). Native speaker reactions to non-native speech. In A. James, & J. Leather (eds), *Second-language speech. Structure and process*, pp. 133-144. Berlin, New York: Mouton de Gruyter.
Crystal, D. (1997). *English as a Global Language*. Cambridge: Cambridge University Press.
Cucchiarini, C., Strik, H., & Boves, L. (2000). Different aspects of expert pronunciation quality ratings and their relation to scores produced by speech recognition algorithms. *Speech Communication, 30*, pp. 109-119.
Daan, J. (1948). *Hij zeit wat. Grepen uit de Amsterdamse volkstaal*. Amsterdam: Jacob van Campen.
Daan, J. (1969). *Dialektatlas van Noord-Holland*. Antwerpen: de Sikkel.
De Groot, A. (1961). *Methodologie. Grondslagen van onderzoek en denken in de gedragswetenschappen*. 's-Gravenhage: Mouton & Co.
De Groot, A. (1966). *Vijven en zessen: cijfers en beslissingen: het selectieproces in ons onderwijs.* Groningen: Wolters-Noordhoff.
Den Boon, T. & Geeraerts, D. (2005). *Van Dale groot woordenboek van de Nederlandse taal*. (14th ed.) Utrecht/Antwerpen: Van Dale lexicografie.
Deterding, D. (2005). Listening to Estuary English in Singapore. *TESOL Quarterly, 39, 3*, pp. 425-440.
Derwing, T. (2003). What do ESL students say about their accents? *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, *59, 4*, pp. 547-566.
Derwing, T., Munro, M., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning, 48, 3*, pp. 393-410.
Dienst Onderzoek en Statistiek (2006). Bevolking in Oud West naar leeftijdsgroepen en etnische groepen, 1 januari 2005. Gemeente Amsterdam.
Retrieved 25/10/06 from http://www.os.amsterdam.nl/tabel/3308/.
Directie Voortgezet Onderwijs. (2005a). Examenprogramma VMBO Moderne vreemde talen. Retrieved 25 09 2005 from http://www.eindexamen.nl/.
Directie Voortgezet Onderwijs (2005b). Examenprogramma Engelse taal en letterkunde v.w.o., examenprogramma Engelse taal en letterkunde h.a.v.o.
Retrieved 25 09 2005 from http://www.eindexamen.nl/.
Doeleman, R. (1998). *Native reactions to nonnative speech*. Tilburg: Tilburg University Press.
Donaldson, B. (1983). *Dutch. A linguistic history of Holland and Belgium*. Leiden: Martinus Nijhoff.
Edelenbos, P., Van der Schoot, F., & Verstralen, H. (2000). *Balans van het Engels na de Basisschool 2*. Arnhem: Citogroep.
Examenblad (2006). Retrieved 25/10/06 from http://www.eindexamen.nl/.
Fagel, W. (1980). *Literatuuronderzoek uitspraakbeoordeling*. Amsterdam: Instituut voor fonetische wetenschappen Universiteit van Amsterdam.
Fagel, W. (1981). *De beoordeling van stem en uitspraak*. Amsterdam: Instituut voor fonetische wetenschappen Universiteit van Amsterdam.
Fagel, W., Van Herpt, L., & Boves, L. (1983). Analysis of the perceptual qualities of Dutch speakers' voice and pronunciation. *Speech Communication, 2*, pp. 315-326.
Fayer, J., & Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning*, *37*, pp. 313-326.
Flege, J. (1996). English vowel production by Dutch talkers. A. James, & J. Leather (eds), *Second-language speech. Structure and process*. Berlin, New York: Mouton de Gruyter, pp. 11-52.
Gatbonton, E., Trofimovich, P., & Magid, M. (2005). Learners' ethnic group affiliation and L2 pronunciation accuracy: a sociolinguistic investigation. *TESOL Quarterly*, *39, 3*, pp. 489-511.
Giles, H. (1970). Evaluative reactions to accents. *Educational Review*, *22*, pp. 211-227.
Gimson, A. (1970). *An introduction to the pronunciation of English.* (2nd ed.). London: Edward Arnold.
Gimson, A. (1978). Towards an international pronunciation of English. In P. Strevens (ed.), *In honour of A. S. Hornby*. Oxford: Oxford University Press, pp. 45-54.

Golombek, P., & Jordan, S. (2005). Becoming "black lambs" not "parrots": a poststructuralist orientation to intelligibility and identity. *TESOL Quarterly, 39, 3*, pp. 369-377.
Graddol, D. (1999). The decline of the native speaker. *AILA Review*, *13*, pp. 57-68.
Gussenhoven, C. (1992). Illustrations of the IPA, Dutch. *Journal of the international phonetic association*, *22, 2*, pp. 45-47.
Gussenhoven, C., & Broeders, A. (1976). *The pronunciation of English. A course for Dutch learners*. Groningen: Wolters-Noordhoff-Longman.
Gussenhoven, C., & Broeders, A. (1997). *English pronunciation for student teachers.* (2nd ed.*)*. Groningen: Wolters-Noordhoff.
Hardy, T. (1960). *Jude the obscure*. London: Macmillan & Co Ltd.
Heeren, W. (2004). The perceptual development of a British-English phoneme contrast in Dutch adults. *Linguistics in the Netherlands 2004*, pp. 91-101.
Hermkens, H. (1969). *Fonetiek en fonologie*. (2nd ed.). 's-Hertogenbosch: Malmberg.
Hogeschool Utrecht (2005). Play your part.
Retrieved 25 09 2005 from http://www.hu.nl/home/homepage.htm.
Honikman, B. (1964). Articulatory settings. In D. Abercrombie, D. Fry, P. MacCarthy, N.C. Scott, & J. Trim (eds), *In Honour of Daniel Jones*. London: Longman, pp. 73-84.
Instituut Educatie & Communicatie (n. d.). *Studiegids 2005-2006 Engels voltijd*. Leeuwarden: NHL.
International Phonetic Association (1949). *The Principles of the International Phonetic Association.* London: University College.
Ioup, G., & Weinberger, S. (eds) (1987). *Interlanguage phonology. The acquisition of a second language sound system*. New York: Newbury House.
James, A. (1977). Dialect and English language teaching. *English Language Teaching Journal*, pp. 146-149.
James, A. (1983). Transferability and dialect phonology: Swabian English. In A. James, & B. Kettemann. (eds), *Dialektphonologie und Fremdsprachenerwerb*. Tübingen: Narr, pp. 173-197.
Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.
Jenkins, J. (2004). Research in teaching pronunciation and intonation. *Annual review of applied linguistics, 24*, pp. 109-125.
Jenkins, J. (2005). Implementing an international approach to English pronunciation: the role of teacher attitudes and identity. *TESOL Quarterly, 39, 3*, pp. 535-543.
Johansson, S. (1978). *Studies of error gravity. Native reactions to errors produced by Swedish learners of English*. Goteborg: Acta Universitatis Gothenburgensis.
Jones, D. (1960). *An Outline of English Phonetics.* (9th ed.). Cambridge: W. Heffer & Sons, Ltd.
Kachru, B. (1996). The paradigms of marginality. *World Englishes*, *15, 3*, pp. 241-255.
Karpf, A., Kettemann, B., & Viereck, W. (1980). Phonology of Dialect Interference in Second Language Learning. Styrian dialect and English language learning. *International Review of Applied Linguistics in Language Teaching*, *18*, pp. 193-208.
Kloeke, G.G. (1934). De Amsterdamsche volkstaal voorheen en thans. *Mededeelingen K. N. A. W. A 77*, pp. 1-14.
Knops, U. (1988). Attitudes towards regional variation. In R. van Hout, & U. Knops (eds), *Language attitudes in the Dutch language area*. Dordrecht: Foris Publications, pp. 104-120.
Koet, T. (1976). Storende fouten. Prioriteiten bij het onderwijs van de uitspraak van het Engels. *Levende talen, 320*, pp. 409-417.
Koet, T. (1977). Namen als hulp bij het uitspraakonderwijs. *Engels, 12*, pp. 206-209.
Koet, T. (1981a). Plat Nederlands is slecht Engels*. Fonetiek en vreemdetalenonderwijs. Toegepaste taalwetenschap in artikelen, 9*, pp. 44-51.
Koet, T. (1981b). Engels op zijn Amsterdams. *Levende talen, 365*, pp. 789-797.
Koet, T. (1986). Het Engels van Niet-Nederlandstaligen. *Toegepaste Taalwetenschap in Artikelen*, *25*, pp. 20-27.
Koet, T. (1988). De Beoordeling van de uitspraak van het Engels van nederlandstaligen. *Toegepaste Taalwetenschap in Artikelen, 32*, pp. 123-137.
Koet, T. (1992). The assessment of Dutch speakers' pronunciation of English by American listeners. In J. Leather, & A. James (eds), *New sounds 92*, pp. 122-127. Amsterdam: University of Amsterdam.
Koet, T. (2005). *Contrastive sound awareness. A course for students of English in the Netherlands*. Amsterdam: Amsterdam faculty of education.

Koet, T., & Weijdema, W. (2005). International standards in the assessment of teacher trainees' language proficiency in the Netherlands. Proceedings ATEE conference. Retrieved from http://www.atee2005.nl.

Koster, C., & Koet, T. (1993). The evaluation of accent in the English of Dutchmen. *Language Learning*, *43, 1*, pp. 69-92.

Kruisinga, E. (1960). *An introduction to the study of English Sounds*. (12th ed.). Groningen: P. Noordhoff N.V.

Labov, W. (1966). *The Social Stratification of English in New York City*. Arlington: the Center for Applied Linguistics.

Land, G. (1981). *Picture stories for composition*. London: Evans.

Lasagabaster, D., & Sierra, J. (2002). University students' perceptions of native and non-native speaker teachers of English. *Language Awareness*, *11, 2,* pp. 132-142.

Levis, J. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly, 39, 3*, pp. 369-377.

Lindemann, S. (2003). Koreans, Chinese or Indians? Attitudes and ideologies about non-native English speakers in the United States. *Journal of sociolinguistics, 7, 3*, pp. 348-364.

Lippi-Green, R. (1997). *English with an accent*. London and New York: Routledge.

Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading: Addison -Wesley.

Lowie, W. (2004). De zin en onzin van het uitspraakonderwijs. *Levende Talen Tijdschrift*, *5*, pp. 3-11.

Lumley, T., & McNamara, T. (1995). Rater characteristics and rater bias: Implications for training. *Language Testing*, *12,* pp. 54-71.

Lynch, B., & McNamara, T. (1998). Using G-theory and Many-facet Rasch measurement in the development of performance assessments of the ESL speaking skills of immigrants. *Language Testing*, *15*, pp. 158-180.

Macdonald, D., Yule, G., & Powers, M. (1994). Attempts to improve English L2 pronunciation: the variable effects of different types of instruction. *Language Learning, 1*, pp. 75-100.

Major, R. (2004). Gender and stylistic variation in second language phonology. *Language variation and change*, *16*, pp. 169-188.

Mees, I., & Collins, B. (1982). A phonetic description of the consonant system of standard Dutch (ABN). *Journal of the international phonetic association*, *12, 1,* pp. 2-12.

Mees, I., & Collins, B. (1983). A phonetic description of the vowel system of standard Dutch (ABN). *Journal of the international phonetic association*, *13, 2*, pp. 64-75.

Meuffels, B. (1994). *De verguisde beoordelaar. Opstellen over opstelbeoordeling*. Amsterdam: Thesis Publishers.

Mulac, A., Hanley, T., & Prigge, D. (1974). Effects of phonological speech foreignness upon three dimensions of attitude of selected American listeners. *The Quarterly Journal of Speech*, *60,* pp. 411-420.

Munro, M. (1998). The effects of noise on the intelligibility of foreign-accented speech. *Studies in second language acquisition*, *20*, pp. 139-154.

Munro, M., & Derwing, T. (1995a). Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning*, *45, 1*, pp. 73-97.

Munro, M., & Derwing, T. (1995b). Processing time, accent and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38, 3*, pp. 289-308.

Murray, J., Bradley, H., Craigie, W., & Onions, C. (1933). *The Oxford English Dictionary.* Oxford: Clarendon Press.

Nolst Trénité, G. (1920). *Drop your foreign accent. Engelsche uitspraakoefeningen.* Haarlem: Tjeenk Willink.

Nunan, D. (1995). *Language teaching methodology. A textbook for teachers*. Hemel Hempstead: Phoenix ELT.

Onderbouw VO. (2005). Retrieved 25 09 2005 from http://www.onderbouw-vo.nl/ventura/engine.php?Cmd=see&P_site=694&P_self=45&PMax=0&PSkip=0.

Onderbouw VO. (2006). Retrieved 31 08 2006 from http://www.onderbouw-vo.nl/ventura/?720_d1347,overzicht_kerndoelen.pdf.

Osgood, C., Suci, G., & Tannenbaum, P. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.

Pennington, M. (1996). *Phonology in English language teaching*. Harlow: Longman.

Preston, D. (1989). *Perceptual dialectology: nonlinguists' views of areal linguistics*. Dordrecht: Foris Publications.

Radboud Universiteit Nijmegen (2005). Opleiding Engels taal en cultuur. Studiegids 2005-2006. Retrieved 25 09 2005 from http://www.ru.nl/contents/pages/140841/engels_boek_scherm.pdf.

Rijlaarsdam, G. (2001). *Vreemd talenonderwijs. Over vernieuwing, samenhang, didactiek en onderzoek*. Utrecht: Universiteit Utrecht.

Riney, T. (1998). Review of M. Yavas (ed.): *First and Second Language Phonology*. *Applied Linguistics*, *18, 2*, pp. 237-239.

Riney, T., Takagi, N., & Inutsuka, K. (2005). Phonetic parameters and perceptual judgments of accent in English by American and Japanese listeners. *TESOL Quarterly*, *39, 3*, pp. 441-465.

Roach, P. (2000). *English phonetics and phonology*. (3rd edition). Cambridge: Cambridge University press.

Robertson, S., & Cassidy, F. (1954). *The development of modern English*. Englewood Cliff: Prentice-Hall, Inc.

Rosewarne, D. (1994a). Estuary English: tomorrow's RP? *English Today, 37, 10, 1,* pp. 3-8.

Rosewarne, D. (1994b). Pronouncing Estuary English. *English Today, 40, 10, 4,* pp. 3-8.

Schairer, K. (1992). Native speaker reaction to non-native speech. *The modern language journal*, *76, 3*, pp. 309-319.

Schatz, H. (1986). *Plat Amsterdams in its social context: a sociolinguistic study of the dialect of Amsterdam*. Amsterdam: P. J. Meertens-Instituut voor dialectologie, volkskunde en naamkunde.

Schilthuizen, A., & Vink, A. (1977). *Easy does it*. Groningen: Wolters-Noordhoff-Longman.

Schoonen, R. (1991). *De evaluatie van schrijfvaardigheidsmetingen*. Amsterdam: Universiteit van Amsterdam.

Sifakis, N., & Sougari, A. (2005). Pronunciation issues and EIL pedagogy in the periphery: a survey of Greek state school teachers' beliefs. *TESOL Quarterly*, *39, 3*, pp. 467-488.

Snider, J., & Osgood, C. (1969). *Semantic differential technique, a sourcebook*. Chicago: Aldine Pub.

Sparks, R., Artzer, M., Ganschow, L., Siebenhar, D., Plageman, M., & Patton, J. (1998). Differences in native-language skills, foreign-language aptitude, and foreign language grades among high-, average- and low-proficiency foreign-language learners: two studies. *Language Testing*, *15, 2*, pp. 181-216.

Staatsen, F. (2004). *Moderne vreemde talen in de onderbouw.* (Derde herziene druk). Bussum: Coutinho.

Stroop, J. (1981). Diffuse Diftongering. *De nieuwe Taalgids*, *74, 1,* pp. 1-16.

Stroop, J. (1998). *Poldernederlands: waardoor het ABN verdwijnt*. Amsterdam: Bakker.

Sulter, A., & Peters, H. (1996). Perceptive characteristics of speech of untrained and trained subjects, and influences of gender and age. In: A. Sulter (ed.), *Variation of Voice Quality Features and Effects of Voice Training in Males and Females,* pp. 73-94. Groningen: Groningen University.

Tielen, M. (1992). *Male and female speech. An experimental study of sex-related voice and pronunciation characteristics*. Amsterdam: Universiteit van Amsterdam.

Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society, 1*, pp. 179-195.

Trudgill, P. (1975). *Accent, dialect and the school*. London: Edward Arnold.

Universiteit Utrecht (2005). Spoken English, Engelse taalvaardigheid 2. Retrieved 25 09 2005 from http://www2.let.uu.nl/Solis/engels/studieschema.htm.

Universiteit van Amsterdam (2005). Phonology of English. Retrieved 25 09 2005 from http://studiegids.uva.nl/web/sgs/nl/c/5186.html/

Van As, C., Koopmans-van Beinum, F., Pols, L., & Hilgers, F. (2003). Perceptual evaluation of tracheosophageal speech by naive and experienced judges through the use of semantic differential scales. *Journal of speech, language and hearing research, 46*, pp. 947-959.

Van Bezooijen, R. (1988). Relative importance of pronunciation, prosody and voice quality. In R. van Hout, & U. Knops (eds), *Language attitudes in the Dutch language area*. Dordrecht: Foris Publications, pp. 85-104.

Van Bezooijen, R. (1994). Aesthetic evaluation of Dutch language varieties. *Language & communication, 14, 3*, pp. 253-263.

Van Bezooijen, R., & Van den Berg, R. (2001). Who power Polder Dutch? *Linguistics in the Netherlands, 2001*. Amsterdam, Philadelphia: John Benjamins, pp. 1-12.

Van Bezooijen, R., Kroezen, S., & Van den Berg, R. (2002). Front approximant /r/: a new and vigorous change in Dutch. *Linguistics in the Netherlands 2002*. Amsterdam, Philadelphia: John Benjamins, pp. 1-12.

Van den Berg, B. (1978). *Foniek van het Nederlands*. (8th ed.). Amsterdam: Van Goor Zonen.
Van den Doel, R. (2006). *How friendly are the natives?* Utrecht: LOT.
Van Donzel, M.E., & Koopmans-van Beinum, F.J. (1997). Evaluation of Prosodic Characteristics in retold stories in Dutch by means of semantic scales. Proceedings Eurospeech '97. Retrieved 31-10-2005 from http://uvafon.hum.uva.nl/monique/Eurospeech.ps.
Van der Haagen, M. (1998). *Caught between norms. The English pronunciation of Dutch Learners*. The Hague: Holland academic graphics.
Van Essen, A. J. (2003). In memoriam Jan Mooijman SJ. *Levende talen magazine*, *90*, *4*, pp. 15-16.
Van Gelderen, A. (1995). *Allochtone leerlingen in het vreemde-talenonderwijs*. Amsterdam: SCO-Kohnstamm Instituut.
Van Gelderen, A. (1996). Zijn er relevante verschillen. Allochtone en autochtone leerlingen in het vreemde-talenonderwijs. *Levende Talen*, *506*, pp. 10-16.
Van Herpt, L., & Fagel, W. (1981). Sexe-invloeden bij de beoordeling van stem en uitspraak. *Toegepaste taalwetenschap in artikelen, 9*, pp. 146-155.
Van Hest, E., De Jong, J. H. A. L., & Stoks, G. (2001). *Nederlandse taalkwalificaties in Europees verband*. Enschede: SLO.
Van Heuven, V., Edelman, L., & Van Bezooijen, R. (2002). The pronunciation of /εi/ by male and female speakers of avant-garde Dutch. *Linguistics in the Netherlands 2002*. Amsterdam, Philadelphia: John Benjamins, pp. 61-72.
Van Hout, R., & Knops, U. (eds) (1988). *Language attitudes in the Dutch language area*. Dordrecht: Foris Publications.
Van Lennep, J. (1845). *Proeve van Platamsterdamsch met ophelderende aanmerkingen van Dr J. Halbertsma.* Deventer: J. de Lange.
Van Weeren, J., & Theunissen, T. (1987). Testing pronunciation: an application of generalizability theory. *Language Learning, 37*, pp. 109-122
Varonis, E.M., & Gass, S. (1982). The comprehensibility of non-native speech. *Studies in Second Language Acquisition*, *4*, pp. 114-136.
Vedocep (2004). Goed nieuws voor Engels in het basisonderwijs. Retrieved 25 09 2005 from http://home.planet.nl/~bodde009/pages/oudnieuw/nieuws28.htm.
Vrije Universiteit Amsterdam (2005). Fonetiek en fonologie. Retrieved 25 09 2005 from http://www.studiegids.vu.nl/index.html?home_file.cfm? objectnaam=OPLEIDINGEN&ruid=35870463&vak_ruid=0&datum=01-Sep-2005&taal=NL.
Wang, H., & Van Heuven, V. (2003). Mutual intelligibility of Chinese, Dutch and American speakers of English. *Linguistics in the Netherlands 2003*, pp. 213-224.
Wang, H., & Van Heuven, V. (2004). Cross-linguistic confusion of vowels produced and perceived by Chinese, Dutch and American speakers of English. *Linguistics in the Netherlands 2004*, pp. 205-216.
Wells, J.C. (1982). *Accents of English*. Cambridge: Cambridge University Press.
Wells, J.C. (2000). *Longman pronunciation dictionary.* (2nd ed.). Harlow: Pearson education limited.
Wesdorp, H. (1981). *Evaluatietechnieken voor het moedertaalonderwijs*. Den Haag: Staatsuitgeverij, RITP, SVO.
Willems, N. (1982). *English intonation from a Dutch point of view.* The Hague: Mouton, de Gruyter.
Yoshida, H. (2004). *An analytic instrument for assessing EFL pronunciation*. Unpublished doctoral dissertation. Temple University, Philadelphia.

AUTHOR INDEX

Alberdingk Thijm, J.A., 56
Albrechtsen, D., 14, 25, 26, 34
Andersen-Hsieh, J., 13, 28
Artzer, M., 28

Barry, W., 30
Berns, J., 85
Bliss, A. J., 5
Blom, J., 33, 38, 64
Bongaerts, T., 12, 22, 24, 25, 27, 77
Bonk, W. J., 24, 31
Booij, G., 52, 54
Boves, L., 25, 29, 33, 34, 75
Bradley, H., 16
Brinton, D. M., 13
Broeders, A., 18, 26, 30, 37, 42, 44, 48, 49, 52, 54, 57, 84
Brouwer, D., 57
Brown, A., 23, 24, 31
Bürki-Cohen, J., 13

Cambridge ESOL, 18, 19, 22, 118, 133
Canagarajah, S., 12
Cassidy, F., 3
Celce-Murcia, M., 13
Cheng, W., 24, 28, 98
Chiswick, B., 29
Cohen, A., 39, 47, 51, 52
Collins, B., 30, 50, 52, 55
Cook, V., 12
Corder, S. P., 29
Council of Europe, 16
Council of Europe., 16, 17, 20
Craigie, W., 16
Crystal, D., 14
Cucchiarini, C., 33
Cunningham-Andersson, U., 34

Daan, J., 2, 47, 55, 56
De Groot, A., 22, 27, 32, 40, 41, 117
De Jong, J. H. A. L., 20
Den Boon, T., 9
Derwing, T., 11, 12, 14, 31
Deterding, D., 15
Dienst Onderzoek en Statistiek, 113
Directie Voortgezet Onderwijs, 19, 20
Doeleman, R., 28, 31, 34, 44, 116
Donaldson, B., 26, 43

Ebeling, C., 39
Edelenbos, P., 33
Edelman, L., 29
Eimas, P., 13
Examenblad, 19

Færch, C., 14, 25, 26
Fagel, W., 25, 34, 38, 44, 64, 84, 105
Fayer, J., 23
Flege, J., 28, 32
Fokkema, A., 39

Ganschow, L., 28
Gass, S., 33
Geeraerts, D., 9
Giles, H., 43, 81
Gimson, A., 14, 42, 49, 54, 57
Golombek, P., 14
Goodwin, J. M., 13
Graddol, D., 14
Gussenhoven, C., 1, 18, 30, 37, 42, 48, 49, 50, 52, 54, 57, 84

Hanley, T., 34
Hardy, T., 7
Heeren, W., 30
Henriksen, B., 14, 25, 26
Hermkens, H., 39
Hilgers, F., 24
Hogeschool Utrecht, 21
Honikman, B., 4

Instituut Educatie & Communicatie, 21
International Phonetic Association, 49
Inutsuka, K., 15, 23
Ioup, G., 29

James, A., 30
Jenkins, J., 13, 22, 31
Johansson, S., 57
Johnson, R., 13, 28
Jones, D., 29
Jordan, S., 11, 14, 52

Kachru, B., 22
Karpf, A., 30
Kettemann, B., 30
Kloeke, G. G., 56
Knops, U., 34, 43
Koehler, K., 13, 28
Koet, T., ii, 2, 5, 6, 19, 23, 33, 41, 49, 54, 56, 143
Koopmans-van Beinum, F.J., 29
Koster, C., 2, 23, 33
Krasinski, E., 23
Kroezen, S., 29
Kruisinga, E., 30

Labov, W., 43, 81
Land, G., 38
Lasagabaster, D., 23, 118
Levis, J., 12
Lindemann, S., 28
Lippi-Green, R., 27
Lord, F.M., 66
Lowie, W., 11, 18
Lumley, T., 24, 118
Lynch, B., 22, 117

Macdonald, D., 13
Major, R., 29, 31
McNamara, T., 22, 24, 117, 118
Mees, I., 30, 50, 52, 55
Meuffels, B., 27
Miller, J., 13, 29
Miller, P., 13, 29
Mulac, A., 34
Munro, M., 11, 12, 14, 31
Murray, J., 16

Nolst Trénité, G., 11
Novick, M.R., 66
Nunan, D., 13

Ockey, G. J., 24, 31
Onderbouw VO, 19
Onions, C., 16
Osgood, C., 64

Patton, J., 28
Pennington, M., 13, 30
Peters, H., 29, 34
Plageman, M., 28
Planken, B., 12, 24
Pols, L., 24
Powers, M., 13
Preston, D., 43
Prigge, D., 34

Radboud Universiteit Nijmegen, 21
Rijlaarsdam, G., 1, 6
Riney, T., 15, 23, 33
Roach, P., 15
Robertson, S., 3
Rosewarne, D., 9

Schairer, K., 14, 26, 33
Schatz, H., 47, 48, 55, 57
Schils, E., 12, 24
Schilthuizen, A., 84
Schoonen, R., 24, 32, 78
Siebenhar, D., 28
Sierra, J., 23, 118

Sifakis, N., 12
Snider, J., 64
Sougari, A., 12
Sparks, R., 28, 43
Staatsen, F., 19
Stoks, G., 20
Strik, H., 33
Stroop, J., 9, 48
Suci, G., 64
Sulter, A., 29, 34

Takagi, N., 15, 23
Tannenbaum, P., 64
Theunissen, T., 32
Tielen, M., 29
Trudgill, P., 27, 43, 44

Universiteit van Amsterdam, 20

Van As, C., 24, 34, 78
Van Bezooijen, R., 26, 27, 29, 34, 53, 64
Van den Berg, B., 26, 29, 39
Van den Doel, R., 25, 27, 57, 60, 75, 116
Van der Haagen, M., 15, 18, 26, 29, 31, 34, 60, 84
Van der Schoot, F., 33
Van Donzel, M.E., 29, 34
Van Essen, A. J., 4
Van Gelderen, A., 45, 57, 116
Van Herpt, L., 25, 33, 34, 38, 64
Van Hest, E., 20
Van Heuven, V., 13, 22, 29, 31
Van Holk, A., 39
Van Hout, R., 43
Van Lennep, J., 56
Van Lennep, W.W., 56
Van Summeren, C., 24
Van Weeren, J., 32
Varonis, E. M., 33
Vedocep, 20
Verstralen, H., 33
Viereck, W., 30
Vink, A., 84
Vrije Universiteit Amsterdam, 20

Wang, H., 13, 22, 31
Warren, M., 24, 77, 98
Weijdema, W., 1, 19
Weinberger, S., 29
Wells, J. C., 2, 26, 49, 54, 63
Wesdorp, H., 24, 27, 32
Wiebe, G., 11, 31
Willems, N., 71, 97

Yoshida, H., 23, 31, 100
Yule, G., 13

SAMENVATTING

De hier gepresenteerde studies gaan over de beoordeling van de uitspraak van het Engels van Nederlandstaligen door Nederlandstaligen. De aanleiding voor deze studies was mijn vermoeden dat het voor sprekers van een andere variant van het Nederlands dan de standaardvariant moeilijker zou kunnen zijn van andere Nederlandstaligen een positief oordeel over de uitspraak van het Engels te krijgen dan voor sprekers van de standaardvariant. Indien het inderdaad het geval zou zijn dat deze sprekers een lagere beoordeling zouden krijgen, dan zijn daar drie mogelijke verklaringen voor.

Een eerste verklaring zou zijn dat het voor sprekers van de standaardvariant van het Nederlands makkelijker is de uitspraak van het Engels te verwerven dan voor sprekers van niet-standaardvarianten aangezien het klanksysteem van de standaardvariant een beter uitgangspunt is dan de klanksystemen van niet-standaardvarianten.

Een tweede verklaring zou zijn dat Nederlandstalige luisteraars in het Engels van sommige Nederlandstaligen uitspraakverschijnselen kunnen horen die zij in verband brengen met niet-standaardvarianten van het Nederlands; de uitstraling van deze verschijnselen is zodanig dat dit het oordeel over het Engels in ongunstige zin beïnvloedt.

Een derde verklaring zou kunnen zijn dat sprekers een hogere of geringere mate van algemene fonetische competentie hebben, die zich zowel in de moedertaal als de vreemde taal manifesteert; een spreker die voldoende competent is om de standaardvariant van de moedertaal te verwerven zal waarschijnlijk ook competent genoeg zijn om zich de gewenste uitspraak van de vreemde taal eigen te maken. De eerste twee verklaringen heb ik onderzocht.

Deze dissertatie bestaat uit een inleidend gedeelte van twee hoofdstukken, het centrale deel van vier hoofdstukken en een afsluitend hoofdstuk. In het eerste hoofdstuk schetste ik de achtergrond die mij ertoe bracht mij in dit onderwerp te verdiepen. Het betreft mijn jeugd in Amsterdam Oud-West, mijn studie in de Verenigde Staten en Ierland, mijn ervaringen als beginnend leraar Engels in het voortgezet onderwijs in Amsterdam, als docent aan een taleninstituut in Dublin en als opleider in Tilburg en later in Amsterdam. In het tweede hoofdstuk besprak ik relevante literatuur en vragen die het m.i. de moeite waard waren te beantwoorden. In het derde hoofdstuk beschreef ik een eerste studie over oordelen over de moedertaal en de doeltaal en hun mogelijke samenhang; deze studie had tot doel na te gaan of er indicaties waren dat het geschetste probleem inderdaad bestond. In het vierde hoofdstuk over de mogelijke invloed van de variant van de moedertaal op de doeltaal gaf ik een beschrijving van een onderzoek naar de uitspraak van het Engels van leerlingen op een school voor technisch onderwijs in Amsterdam. In het vijfde hoofdstuk over de keu-

ze van de beoordelaars deed ik verslag van een eerste experiment dat tot doel had argumenten te vinden voor een verantwoorde keuze van Nederlandstalige en Engelstalige beoordelaars. In het zesde hoofdstuk over oordelen over de doeltaal en de moedertaal deed ik verslag van het laatste experiment dat tot doel had vast te stellen of het oordeel van Nederlandstalige beoordelaars van het Engels van Nederlandstalige sprekers werd beïnvloed door eigenschappen van het Nederlands van deze sprekers. In het zevende hoofdstuk presenteerde ik mijn conclusies, ging ik in op de beperkingen van de studies en deed ik aanbevelingen voor eventueel vervolgonderzoek en voor de onderwijspraktijk.

1. INLEIDING

Ik ben me al vroeg bewust geworden van de verschillende reacties die taalvarianten in luisteraars oproepen. De Amsterdamse wijk waar ik ben opgegroeid moet dicht bij een soort taalgrens hebben gelegen, die tussen het gebied waar het "Jordaans" werd gesproken en dat waar een westelijke variant van het Amsterdams gebruikelijk was. De mannelijke, niet de vrouwelijke, leden van de familie van mijn vader spraken een Nederlands met een Amsterdamse kleur, die afwezig was in het Nederlands van de familie van mijn moeder. De vroege kennismaking met taalverschillen in de moedertaal werd aangevuld door mijn kennismaking met de verschillen in het Engels, zowel dat van Engelstaligen als van niet-Engelstaligen, tijdens mijn studie in de Verenigde Staten en in Ierland. Als beginnend leraar Engels in het voortgezet onderwijs in Amsterdam merkte ik hoe de (variant van de) moedertaal het Engels van mijn leerlingen beïnvloedde. Oudere collega's waarschuwden mij geen al te hoge verwachtingen te hebben van het Engels van leerlingen met een "plat" Amsterdams accent. Toch merkte ik dat ook deze leerlingen wel degelijk een acceptabel niveau in het Engels konden bereiken. Als beginnend opleider in Tilburg en later in Amsterdam werd ik door collega's gewaarschuwd voor de problemen van studenten met een Limburgs resp. Noord-Hollands accent. Bij bezoeken aan stagescholen merkte ik dat leerlingen (niet alleen Nederlandstalige maar ook moedertaalsprekers van het Engels, Frans en Duits) niet zelden lage cijfers kregen voor hun Engels, Frans en Duits wegens hun problemen met het Nederlands. In dit "vreemde" talenonderwijs op deze scholen, gekenmerkt door vermijding van de doeltaal en grote nadruk op de vertaling uit het Engels in het Nederlands, werd het cijfer voor het Engels in belangrijke mate bepaald door de kwaliteit van het Nederlands.

2. LITERATUURSTUDIE

Het tweede hoofdstuk bevat een studie van de relevante literatuur. In deze studie begon ik met de vraag wat het doel van het uitspraakonderwijs is. Daarna keek ik naar de beoordeling van de uitspraak. Ik eindigde met een aantal vragen die mijns inziens verder onderzoek behoefden.

Wat het doel van het onderwijs van de uitspraak betreft vond ik een onderscheid tussen enerzijds auteurs die stelden dat verstaanbaarheid het doel zou moeten zijn waar de lerende naar moet streven en anderzijds diegenen die meenden dat een uit-

spraak die vrij is van een buitenlands accent het doel zou moeten zijn. De standpunten van de twee groepen stonden soms diametraal tegenover elkaar – vooral vertegenwoordigers van de groep die verstaanbaarheid als enig doel bepleitte, drukten zich sterk uit – zodat het moeilijk leek de standpunten te verzoenen. Bij beide groepen deed zich de vraag voor welk model de leerlingen dienden na te streven of voor wie de leerling verstaanbaar moest zijn. Moest de leerling de moedertaalspreker als voorbeeld nemen en verstaanbaar zijn voor bijv. Britten of Amerikanen of mocht de leerling een ander model kiezen en verstaanbaar willen zijn voor andere niet-Engelstaligen? Indien de keuze viel op de moedertaalspreker bleef de vraag welke variant men moest kiezen: Brits of Amerikaans of eventueel een andere variant. Indien de keuze viel op een niet-moedertaalmodel was het nog moeilijker een keuze te maken: moest men kiezen voor een kunstmatig model of zou bijv. een Nederlands Engels acceptabel zijn?

Vervolgens besprak ik het Gemeenschappelijk Europees Referentiekader (GERK) voorzover dat aanwijzingen bevatte die tot een verzoening van de standpunten zouden kunnen leiden. Deze aanwijzingen bleken aanwezig te zijn. Het GERK onderscheidde twee competenties waarin de uitspraak een rol speelt: de fonologische en de orthoëpische. Bij de eerste competentie ging het om kennis van en vaardigheid in de perceptie en productie van fonemen en allofonen, bij de tweede om het vermogen de uitspraak van een woord uit de spelling af te leiden (al of niet met behulp van een woordenboek). Voor elk van de eerste vijf niveaus van het GERK – A1, A2, B1, B2 en C1 – stelde het document doelstellingen voor; wat betreft de twee laagste niveaus ging het uitsluitend om verstaanbaarheid; bij het middenniveau en zeker bij het hoogste niveau was alleen verstaanbaarheid onvoldoende: het vreemde accent mocht op niveau B1 niet meer dan enkele malen blijken; de uitspraak moest evenals de intonatie duidelijk en natuurlijk zijn op niveau B2; op niveau C1 moest de lerende intonatie en zinsaccent zo beheersen dat hij/zij subtiele verschillen in betekenis kon uitdrukken. Hoewel het GERK dus geen uitspraak vereiste als die van de moedertaalspreker, diende de uitspraak op de hoogste niveaus niet veel daarvan te verschillen.

De door de Universiteit van Cambridge aangeboden examens in het Engels voor sprekers van andere talen (English for speakers of other languages: ESOL) leidden tot internationaal erkende certificaten. De door Cambridge ESOL voor de verschillende examens voor de Engelse taalvaardigheid geformuleerde eisen waren ook voor wat betreft de uitspraak gebaseerd op de niveaus van het GERK, met dien verstande dat Cambridge ESOL een onderscheid maakte tussen de twee hoogste niveaus, C1 (Certificate in Advanced English) en C2 (Certificate for Proficiency in English). Zowel op niveau C1 als op niveau C2 was een vreemd accent toegestaan, mits dit de verstaanbaarheid niet belemmerde.

De kerndoelen en eindtermen van het voortgezet onderwijs in Nederland gaven weinig houvast voor de in het uitspraakonderwijs na te streven doelen. Bij de eindexamens HAVO en VWO bijv. dienden de kandidaten op het laagste niveau verstaanbaar te zijn, op twee hogere niveaus zowel verstaanbaar als begrijpelijk en op het hoogste niveau was bij de exameneisen van verstaanbaarheid en begrijpelijkheid geen sprake. In het hoger onderwijs hadden sommige universiteiten en hogescholen wel duidelijke doelstellingen geformuleerd, welke doelstellingen onderling echter

sterk verschilden. Bij de ene universiteit ging het om verbetering van de uitspraak en het verwerven van vaardigheden, bij een andere om het verwerven van kennis en het vermogen een fonetische transcriptie te maken; slechts één universiteit eiste een uitspraak die dicht lag bij die van moedertaalsprekers van hetzij het Britse of het Amerikaanse standaardaccent. Bij de lerarenopleidingen aan de hogescholen varieerde het doel van kennis van de uitspraak tot het vermogen helder en duidelijk Engels te spreken zonder een storend Nederlands of Fries accent. Hoewel het moeilijk was te zien wat de verschillende doelstellingen gemeen hadden, leek het erop dat op de laagste niveaus verstaanbaarheid het enige doel was terwijl op de hogere niveaus wel degelijk eisen aan het accent gesteld werden; slechts bij één instelling werd echter een uitspraak vereist die de uitspraak van moedertaalsprekers dicht benaderde.

Evenals bij de doelstellingen van het uitspraakonderwijs moest men bij de beoordeling van de uitspraak een groot aantal keuzes maken. Ik besprak de verschillende keuzes die een aantal auteurs hebben gemaakt met betrekking tot de beoordelaars, de sprekers en de taken die de sprekers moesten verrichten. Ook besteedde ik aandacht aan een aantal problemen die zich bij het beoordelingsproces voordeden.

Moest men voor beoordelaars kiezen die moedertaalsprekers van de doeltaal waren of voor niet-moedertaalsprekers? En als men niet moedertaalsprekers koos, moesten het dan sprekers van dezelfde taal zijn als de te beoordelen persoon? De keuze voor niet-moedertaalsprekers was zeker niet onomstreden; sommigen meenden dat deze te streng zouden zijn in hun oordeel; luisteraars die naar de verrichtingen van hun landgenoten luisterden, zouden kieskeurig zijn, in verlegenheid worden gebracht, geïrriteerd raken en/of vervuld worden van afkeer; anderen daarentegen stelden dat er geen enkele reden was om aan te nemen dat niet-moedertaalsprekers de verstaanbaarheid slechter zouden kunnen beoordelen dan moedertaalsprekers.

Een vergelijkbare situatie deed zich voor bij de keuze voor ervaren dan wel onervaren of jonge dan wel oude beoordelaars. Sommige auteurs kwamen tot de bevinding dat ervaren luisteraars de voorkeur verdienden, andere auteurs daarentegen zagen geen verschil. Sommige auteurs merkten dat oudere luisteraars minder streng waren, wellicht doordat hun gehoor minder gevoelig was.

Over de keuze tussen mannelijke en vrouwelijke beoordelaars bestond evenmin unanimiteit. De meeste auteurs gaven aan dat vrouwelijke beoordelaars strenger waren maar anderen hadden geen verschillen gevonden en weer anderen hadden gemerkt dat mannen strenger waren.

De regionale herkomst en sociaal-economische achtergrond van beoordelaars maakten volgens de meeste auteurs weinig verschil, hoewel anderen wel verschillen hadden aangetroffen; zo zouden Noord-Amerikaanse beoordelaars strenger kunnen zijn in hun oordeel over de fouten die zij opmerkten dan andere Engelstalige beoordelaars.

Beoordelaars, niet alleen van de uitspraak maar ook van andere aspecten van de taalvaardigheid, bleken aan veel kwalen te kunnen lijden. Het halo-effect, het verschijnsel dat de uitstraling van één aspect van de te beoordelen prestatie het oordeel over andere aspecten beïnvloedt, was één van die kwalen. Zowel moedertaalsprekers als niet-moedertaalsprekers leden hieraan; zo zou een uitspraak die de beoordelaar als “plat” of “beschaafd” voorkwam, kunnen uitstralen naar het oordeel over de per-

soonlijkheid van de spreker; eveneens zou een oordeel over segmentale aspecten van de uitspraak het oordeel over suprasegmentale aspecten kunnen beïnvloeden.

Ook over de sprekers kon men zich een aantal vragen stellen. De eerste vraag betrof het geslacht van de sprekers, de tweede de regionale herkomst. De meeste auteurs, maar zeker niet alle, hadden geconstateerd dat de spraak van vrouwen gemiddeld gunstiger werd beoordeeld dan die van mannen en dat meer vrouwen een beschaafde uitspraak hadden dan mannen. Een aantal studies was gewijd aan de verschillen in de uitspraak van het Engels tussen sprekers van diverse varianten van een aantal moedertalen, waaronder het Duits en het Nederlands.

Welke taken laat men de sprekers en de luisteraars verrichten? Deze vraag is zeer uiteenlopend beantwoord. Bij de sprekers ging het meestal om een monoloog: het voorlezen van een lijst van woorden of een tekst of het produceren van een "vrije" gesproken tekst; slechts in een aantal gevallen ging het om een dialoog en zelden om een polyloog. Gesproken productie kwam vaker aan de orde dan gesproken interactie. Bij de luisteraars zag men een verschil tussen enerzijds analytische taken en anderzijds holistische. Sommige luisteraars dienden een onderscheid te maken tussen segmentale en suprasegmentale aspecten. De vraag deed zich voor of voorstanders van sterk analytische en soms zelfs atomistische taken het halo-effect niet onderschatten.

Veel auteurs hebben getracht de vele oordelen van de luisteraars op een groot aantal schalen terug te brengen tot een hanteerbaarder aantal dimensies. Ook hier trof men een breed scala aan dimensies aan.

Bij bijna alle keuzes die men kon maken heb ik moeten constateren dat er geen eenstemmig antwoord was te vinden. Daarom formuleerde ik vervolgens een aantal vragen die het mijns inziens verdienden beantwoord te worden, namelijk over de keuze van de te bestuderen beoordelaars, de keuze van de te bestuderen sprekers en de keuze van de taken van de sprekers.

3. WAARGENOMEN KWALITEIT VAN UITSPRAAK IN MOEDERTAAL EN VREEMDE TAAL

In het derde hoofdstuk introduceerde ik het probleem dat fragmenten van Engelse spraak die waren geproduceerd door sprekers van een niet-standaard variant van het Nederlands minder positief werden beoordeeld dan fragmenten die door sprekers van een standaardvariant waren geproduceerd. Ik probeerde eerst na te gaan of dit inderdaad het geval was. Hiertoe vroeg ik vijftig 1^{e}-jaarsstudenten in het hoger onderwijs semi-spontane spraak in het Nederlands te produceren. Mijn oordeel over de kwaliteit van de uitspraak van het Nederlands gaf ik aan op een tienpuntsschaal. Het Engels van de studenten werd beoordeeld door drie beoordelaars, twee Nederlandse en een Engelse, eveneens op een tienpuntsschaal. De oordelen van de drie beoordelaars werden gemiddeld. Vervolgend bepaalde ik de rangorde van de sprekers voor de gemiddelde oordelen over het Engels en de oordelen over het Nederlands. Er bleek een weliswaar zwakke maar significante correlatie tussen beide oordelen te bestaan. Aangezien er aanwijzingen waren dat het probleem inderdaad bestond formuleerde ik drie mogelijke verklaringen.

- Voor sprekers van een niet-standaard variant van de moedertaal zou het moeilijker kunnen zijn om de klanken van een vreemde taal te produceren dan voor sprekers van de standaardvariant aangezien het foneemsysteem van hun variant een aantal klanken mist die wel aanwezig zijn in het foneemsysteem van de standaardvariant.
- Luisteraars die geen moedertaalsprekers zijn van de vreemde taal en in de uitspraak van de vreemde taal kenmerken herkennen van niet-standaardvarianten – niet van de vreemde taal maar van de moedertaal – zouden daardoor geneigd kunnen zijn deze sprekers een lagere plaats in de rangorde te geven
- Sprekers die een hoog niveau van algemene fonetische competentie hebben bereikt zouden daardoor zowel de moedertaal als de vreemde taal goed kunnen uitspreken, terwijl sprekers die dat hoge niveau niet hebben bereikt zowel de moedertaal als de vreemde taal minder goed zouden kunnen uitspreken. Deze laatste mogelijke verklaring heb ik niet onderzocht.

4. AMSTERDAMS ENGELS

In het hoofdstuk over Amsterdams Engels formuleerde ik de eerste verklaring als een onderzoeksvraag.

Is de fonologische interferentie van een Amsterdamse dialect in het Nederlands een ernstiger hindernis bij de verwerving van de uitspraak van het Engels dan de fonologische interferentie van het standaarddialect van het Nederlands?

Ik bestudeerde de uitspraak van leerlingen in de laagste klassen van een Amsterdamse school voor voortgezet onderwijs en vroeg daarnaast vijf mannelijke leerlingen een Engelse tekst voor te lezen. Mijn observaties en de bestudering van de voorgelezen teksten duidden erop dat sprekers van een Amsterdamse variant van het Nederlands grotere moeite zouden hebben met de medeklinkers van het Engels maar minder grote moeite met de klinkers dan sprekers van de standaardvariant. Aangezien een aantal auteurs aannam dat problemen met de medeklinkers een grotere barrière voor de verstaanbaarheid vormden dan problemen met de klinkers stelde ik dat de onderzoeksvraag bevestigend kon worden beantwoord. Daarom zou het heel goed zo kunnen zijn dat sprekers van een Amsterdams dialect het Engels waarschijnlijk minder goed uitspraken dan sprekers van het standaarddialect van het Nederlands. Dus zou men Nederlandse beoordelaars geen verwijten kunnen maken als ze het Engels van sprekers van het Amsterdams ongunstiger beoordeelden dan het Engels van sprekers van het ABN.

5. DE KEUZE VAN BEOORDELAARS

Ondanks het bevestigende antwoord op mijn eerste onderzoeksvraag was ik er nog niet van overtuigd dat het oordeel van de Nederlandse luisteraars geheel geloofwaardig was en besloot ik de tweede mogelijke verklaring te onderzoeken. Om de betrouwbaarheid van de Nederlandse beoordelaars op de proef te stellen moest ik de oordelen van Engelstalige luisteraars gebruiken. Hoe dit gebeurde, beschreef ik in hoofdstuk 6. Voor ik echter de tweede verklaring kon onderzoeken, moest ik het

probleem van de keuze van beoordelaars aanpakken; welke Engelstalige beoordelaars zouden mij oordelen verschaffen waar ik op kon vertrouwen? Hoe ik tot een keuze kwam beschreef ik in hoofdstuk 5.

Ik vroeg vier groepen luisteraars – twee Engelstalige groepen, Britten en Amerikanen, en twee Nederlandstalige groepen, ervaren en onervaren – om het door negen Nederlandse sprekers geproduceerde Engels op vijf zevenpuntsschalen te beoordelen, waarvan vier schalen gerelateerd waren aan de estheticiteit van de uitspraak en één schaal aan de intonatie. Ik nam daarbij verschillen in rangorde in overweging en verschillen in strengheid maar ook verschillen in strengheid tegenover mannelijke en vrouwelijke sprekers en verschillen in strengheid van mannelijke en vrouwelijke luisteraar en verschillen in strengheid van luisteraars tegenover sprekers van hetzelfde en het andere geslacht. Dit waren de uitkomsten.

Ervaren Nederlandse en Britse luisteraars. De ervaren Nederlandse luisteraars verschilden niet van de Britse luisteraars in de rangorde van de sprekers die zij aanbrachten maar waren wel strenger in hun oordeel op twee estheticiteitsschalen; ook waren zij positiever over vrouwelijke sprekers dan over mannelijke sprekers in hun oordeel op twee estheticiteitsschalen en op de intonatieschaal, terwijl de oordelen van de Britse luisteraars over mannelijke en vrouwelijke sprekers niet verschilden.

Britse en Amerikaanse luisteraars. De Britse en Amerikaanse luisteraars verschilden in de rangorde die zij tussen de sprekers aanbrachten op één estheticiteitsschaal. Zij verschilden niet in strengheid maar de Britse luisteraars verschilden van de Amerikaanse luisteraar in die zin dat de laatsten op twee estheticiteitsschalen en op de intonatieschaal positiever waren over de vrouwelijke sprekers.

Ervaren en onervaren Nederlandse luisteraars. De ervaren Nederlandse luisteraars verschilden van de onervaren Nederlandse luisteraars in de rangorde van de sprekers op twee estheticiteitsschalen. Ze waren strenger dan de onervaren luisteraars op drie estheticiteitsschalen. Ze verschilden in hun oordeel over vrouwelijke sprekers op alle vier de estheticiteitsschalen.

Onervaren Nederlandse en Amerikaanse luisteraars. De onervaren Nederlandse en Amerikaanse luisteraars brachten op alle schalen een verschillende rangorde tussen de sprekers aan. De onervaren Nederlandse luisteraars verschilden op twee estheticiteitsschalen van de Amerikaanse luisteraars. Ze verschilden in hun oordeel over vrouwelijke sprekers op alle vier de estheticiteitsschalen,

Ervaren Nederlandse en Amerikaanse luisteraars. De ervaren Nederlandse en de Amerikaanse luisteraars verschilden in de rangorde van de sprekers, in strengheid op slechts één schaal en in strengheid tegenover vrouwelijke sprekers op geen enkele schaal.

Onervaren Nederlandse en Britse luisteraars. De onervaren Nederlandse luisteraars verschilden qua aangebrachte rangorde op twee estheticiteitsschalen van de Britse luisteraars en qua strengheid op drie. In hun strengheid over vrouwelijke luisteraars verschilden de onervaren Nederlandse en Britse luisteraars op één estheticiteitsschaal.

De luisteraars verschilden het meest in hun oordelen over de estheticiteit en slechts zelden in hun oordeel over de intonatie.

Het was niet zo dat Engelstalige beoordelaars op alle schalen milder waren dan Nederlandstalige of dat ervaren beoordelaars op alle schalen strenger waren dan onervaren beoordelaars.

Ik concludeerde dat de ervaren Britse luisteraars het meest geschikt waren; ze waren betrouwbaarder dan de Amerikaanse luisteraars en kwamen overeen met de ervaren Nederlandse luisteraars in de aangebrachte rangorde; in strengheid verschilden ze niet van de andere groepen behalve op enkele schalen, waar ze van de ervaren Nederlanders verschilden. Daar stond tegenover dat de Britse luisteraars niet positiever waren over vrouwelijke sprekers, zoals de ervaren Nederlandse en Amerikaanse luisteraars, en dat ze ook minder positief waren over sprekers van hetzelfde geslacht; dit deed zich slechts op één schaal voor. Daarom koos ik ervoor met de Britse luisteraars verder te gaan.

Bij de zoektocht naar de meest geschikte Engelstalige luisteraars waren aanzienlijke verschillen tussen ervaren en onervaren Nederlandse luisteraars aan het licht gekomen. Deze verschillen vond ik dermate groot dat het mij onverstandig leek niet met deze twee groepen verder te gaan. Het verschil tussen ervaren en onervaren Nederlandse luisteraars moest dus in het volgende hoofdstuk terugkomen.

6. OORDELEN OVER HET NEDERLANDS EN HET NEDERLANDS ENGELS

Aan het begin van het hoofdstuk over oordelen over het Nederlands en het Nederlands Engels had ik de tweede verklaring in de vorm van drie onderzoeksvragen geformuleerd.

1) Zijn Nederlands luisteraars strenger in hun oordeel over de door Nederlanders geproduceerde Engelse spraak dan Engelse luisteraars?
2) Zijn Nederlandse luisteraars naar door Nederlanders geproduceerde Engels onrechtvaardig in de zin dat er een zwakke samenhang is tussen hun oordelen en de oordelen van Engelse luisteraars?
3) Zijn Nederlandse luisteraars bevooroordeeld in hun oordelen over de door Nederlanders geproduceerde Engelse spraak?

Hoewel ik één probleem bij de keuze van Engelstalige beoordelaars had opgelost, namelijk de keuze tussen Britse en Amerikaanse beoordelaars, had ik een ander probleem nog niet opgelost, namelijk dat van de keuze tussen mannelijke en vrouwelijke luisteraars. Bovendien waren er problemen betreffende de Nederlandse beoordelaars, de sprekers en hun taken. Om de resultaten van het eerste experiment te bevestigen en om de nieuwe onderzoeksvraag te beantwoorden, deelde ik de eerste nieuwe onderzoeksvraag op in vier subvragen. In de eerste subvraag maakte ik een onderscheid tussen de taken: het voorlezen van een tekst en het produceren van semi-spontane spraak. In de tweede subvraag scheidde ik ervaren en onervaren luisteraars. In de derde subvraag haalde ik mannelijke en vrouwelijke luisteraars uit elkaar. In de vierde subvraag maakte ik onderscheiden groepen van luisteraars naar sprekers van hun eigen geslacht en luisteraars naar sprekers van het andere geslacht.

Ik vroeg twintig Nederlandse sprekers om twee stukjes semi-spontane spraak te produceren en twee tekstjes voor te lezen, beide zowel in het Nederlands als in het

Engels. Ik vond ongeveer 140 Nederlandse en Engelse luisteraars, zowel ervaren als onervaren, bereid om deze stukjes te beoordelen op elf zevenpuntsschalen die betrekking hadden op de estheticiteit, de intonatie, de begrijpelijkheid en de standaardheid. Dit waren de antwoorden op de vier subvragen.

Vrije spraak en het voorlezen van een tekst. De Nederlandse luisteraars waren strenger dan de Engelse luisteraars in hun oordeel over estheticiteit en intonatie van de door Nederlandse sprekers geproduceerde vrije spraak en het voorlezen van een tekst.

Ervaren en onervaren luisteraars. De ervaren Nederlandse luisteraars waren strenger in hun oordeel dan de ervaren Engelse luisteraars en de onervaren Nederlandse luisteraars waren strenger in hun oordeel dan de onervaren Engelse luisteraars voorzover het hun oordeel op de intonatie en begrijpelijkheidsschalen betrof.

Mannelijke en vrouwelijke sprekers. De Nederlandse luisteraars waren strenger in hun oordeel over het door mannelijke sprekers geproduceerde Engels dan over het door vrouwelijke sprekers geproduceerde Engels en de Engelse luisteraars waren strenger in hun oordeel over het door mannelijke sprekers geproduceerde Engels dan over het door vrouwelijke sprekers geproduceerde Engels alleen in zoverre het hun oordeel op de estheticiteits- en intonatieschalen betrof. Het omgekeerde was het geval bij de oordelen op de begrijpelijkheidsschaal.

Luisteraars naar sprekers van hetzelfde geslacht en luisteraars naar sprekers van het andere geslacht. De Engelse luisteraars waren strenger in hun oordeel over het door sprekers van hetzelfde geslacht geproduceerde Engels dan in hun oordeel over het door sprekers van het andere geslacht geproduceerde Engels bij de estheticiteitsschaal.

Het grootste aantal verschillen trof ik aan op de estheticiteitsschaal en het kleinste op de begrijpelijkheidsschaal. De antwoorden op de drie hoofdonderzoeksvragen waren als volgt.

De Nederlandse luisteraars waren strenger dan de Engelse luisteraars waar het hun oordeel over het door de Nederlandse sprekers geproduceerde Engels op de estheticiteit-, intonatie- en begrijpelijkheidsschalen betrof.

De Nederlandse luisteraars waren onrechtvaardig in hun oordeel over het door de Nederlanders geproduceerde Engels in die zin dat er een zwak verband bestond tussen hun oordelen en de oordelen van de Engelse luisteraars op de estheticiteit-, de intonatie- en de standaardschalen.

De Nederlandse luisteraars waren bevooroordeeld in hun oordeel over het door de Nederlanders geproduceerde Engels voor de estheticiteit-, intonatie- en standaardschalen.

De aan de tweede verklaring gerelateerde onderzoeksvragen waren voor alle schalen meestal bevestigend beantwoord behalve voor de begrijpelijkheidsschaal. Daarom zou men kunnen stellen dat Nederlandse luisteraars weliswaar te streng maar niet onrechtvaardig of bevooroordeeld waren in hun oordeel over de begrijpelijkheid.

Nadat ik de aan de eerste twee verklaringen gerelateerde onderzoeksvragen had beantwoord en ernstige twijfel had uitgesproken over de geloofwaardigheid van Nederlandse beoordelaars, zag ik geen noodzaak om de derde verklaring – de geringere of hogere mate van algemene fonetische competentie – te onderzoeken.

7. CONCLUSIES

7.1 Beperkingen

Aan de hier gepresenteerde studies kleven een aantal beperkingen. Deze beperkingen waren het gevolg van de keuze van de sprekers, de keuze van de luisteraars, de keuze van de taken die de sprekers moesten verrichten en van de nader onderzochte verklaringen.

Bij de studie over het Engels van sprekers van een Amsterdams Nederlands ging het om een gering aantal mannelijke sprekers, leerlingen in de onderbouw van wat toen het lager beroepsonderwijs was. Bij het eerste experiment waren de sprekers negen eerstejaarsstudenten in het hoger onderwijs (vijf mannelijke en vier vrouwelijke). Alleen bij het tweede experiment was door externe deskundigen een selectie gemaakt van sprekers; deze selectie bestond uit twintig sprekers, allen eerstejaarsstudenten in het hoger onderwijs, tien mannelijke en tien vrouwelijke. Alle sprekers in de eerste studie waren Amsterdams; in het eerste experiment was de geografische herkomst van de sprekers niet bekend; in het tweede experiment ging het om sprekers uit Noord-Holland (en Amsterdam), Zuid-Holland en Utrecht. Ook in de tijd verschilden de sprekers. De Amsterdamse sprekers produceerden hun spraak in de zeventiger jaren, de sprekers in het eerste experiment in de tachtiger jaren en die in het laatste experiment in de jaren negentig. Men zou de vraag kunnen stellen of de uitkomsten nog enige relevantie hebben in de eenentwintigste eeuw. De veranderingen in de samenstelling van de populatie van Amsterdam zouden de relevantie nog twijfelachtiger hebben kunnen maken, ware het niet zo dat in het beschreven deel van Amsterdam deze veranderingen niet zo groot schijnen te zijn als elders.

Ook door de keuze van luisteraars was de relevantie van een deel van de studies beperkter dan wenselijk was geweest. In de studie naar het Amsterdams Engels ontbrak het oordeel van een Engelstalige luisteraars: een ernstige omissie, zeker voor wie enige waarde aan de bevindingen van de andere studies hecht. Het geringe aantal ervaren Engelstalige luisteraars in het laatste experiment was eveneens te betreuren. Wat gold voor de sprekers gold ook voor de luisteraars; door de veranderingen in de samenstelling van de bevolking zou de houding van luisteraars tegenover verschillend varianten van het Nederlands gewijzigd kunnen zijn; men zou zich kunnen voorstellen dat de komst van veel niet-Nederlandstaligen tot grotere tolerantie tegenover afwijkingen van de standaardvariant heeft geleid; eveneens is het mogelijk dat de tolerantie tegenover een afwijkende uitspraak juist is afgenomen.

Alleen in het laatste experiment was de sprekers gevraagd twee taken te verrichten: het voorlezen van een tekst en het produceren van halfspontane spraak. Als ik de Amsterdamse sprekers had gevraagd twee verschillende taken uit te voeren, waren er vrijwel zeker verschillen geweest in de beoordeling van het voorlezen en van de semi-spontane spraak. Ook ging het uitsluitend om de productieve mondelinge vaardigheid en niet om andere vaardigheden, en om het Engels en niet om andere vreemde talen.

7.2 Aanbevelingen voor verder onderzoek

De aangegeven beperkingen leidden tot aanbevelingen voor verder onderzoek. De jonge Amsterdamse mannelijke sprekers waren vertegenwoordigers van een beperkte groep; het was zeer wel mogelijk dat sprekers van andere varianten van het Nederlands, niet alleen in Nederland maar ook in België, niet alleen moedertaalsprekers maar ook tweedetaalsprekers, vergelijkbare problemen met het Engels hadden. Ook zou men naar sprekers met een andere vaardigheidsniveau – bijv. B1 – kunnen kijken.

Wat de Engelstalige luisteraars betreft, heb ik mij in het laatste experiment beperkt tot Engelsen (in de nauwe zin van het woord), hoewel ik in het eerste experiment toch verschillen tussen Britse en Amerikaanse luisteraars had gevonden. Daarom zou het interessant zijn als er nader onderzoek naar de oordelen van luisteraars uit andere delen van de Engelstalige wereld kon worden gedaan. Wat de niet-Engelstalige luisteraars betreft, heb ik alleen naar Nederlanders, landgenoten van de sprekers gekeken. Het zou alleszins de moeite waard zijn naar andere niet-Engelstalige luisteraars, bijv. francofone, te kijken. Mijn Engelstalige luisteraars waren weliswaar voor een groot deel leerlingen in het voortgezet onderwijs, maar dit was niet het geval bij de Nederlandstalige luisteraars. De oordelen van Nederlandse scholieren zouden voor toekomstige docenten zeker interessant materiaal kunnen bevatten.

Van de gesuggereerde onderwerpen voor nadere studie, leek de beoordeling van het Engels van niet-Nederlandstalige sprekers in Nederlands mij het meest urgent. Het leek onwaarschijnlijk dat de negatieve reacties op met een buitenlands accent gesproken Nederlands die bij Nederlandstalige luisteraars zijn aangetroffen afwezig zouden zijn bij met een buitenlands accent gesproken Engels.

Bij de taken van de sprekers, tenslotte, heb ik mij grotendeels beperkt tot gesproken productie en heb ik slechts weinig aandacht besteed aan gesproken interactie; ondanks de moeilijkheden die dit met zich mee zou brengen, zou de beoordeling van het Engels in gesproken interactie zeker een onderwerp zijn dat het bestuderen waard is.

7.3 Aanbevelingen voor de onderwijspraktijk

In mijn conclusies heb ik ernstige twijfels geuit over het vermogen van Nederlandstalige luisteraars om te oordelen over de uitspraak van het Engels van andere Nederlanders. Toch zou het niet juist zijn om te stellen dat Nederlandstalige luisteraars uitgesloten dienen te worden van deze taak bij sprekers van alle niveaus van taalvaardigheid. Op de niveaus waar het uitsluitend gaat om de begrijpelijkheid en verstaanbaarheid (A1 en A2) kunnen Nederlandse luisteraars wel degelijk een bijdrage leveren, mits zij zich bewust zijn van hun neiging te streng te oordelen. Ook op niveau B1 zouden Nederlandstalige luisteraars een rol kunnen spelen, mits zij zich in hun oordeel beperken tot de begrijpelijkheid en de verstaanbaarheid en zich onthouden van esthetische oordelen en oordelen over de sterkte van het accent (m.a.w. men zou Nederlandse luisteraars niet moeten vragen of de uitspraak bijvoorbeeld aangenaam, beschaafd, mooi, verzorgd of vrij van accent is maar wel of deze begrijpelijk

of nauwkeurig is). Het lijkt niet verstandig Nederlandstalige luisteraars te betrekken bij het beoordelen van sprekers op niveau B2 en hoger, tenzij deze beoordelaars goed getraind zijn; zelfs in dat geval bestaat echter het gevaar dat het effect van de training na enige tijd verdwijnt. Het verdient daarom aanbeveling om op deze hogere niveaus bij voorkeur met Engelstalige beoordelaars te werken.

CURRICULUM VITAE

Ton Koet (1945) was born in Amsterdam, where he attended primary and secondary school. He studied at Gonzaga University in Spokane, and the University of Washington in Seattle (USA). He studied English Language and Literature and Educational Sciences in the University of Amsterdam, where he passed his doctoral examination in 1970. He followed the post-graduate course in Anglo-Irish studies in University College Dublin and received his master's degree from the National University of Ireland in 1971. He started teaching English in 1968; he taught in several secondary schools in Amsterdam and in the English Language Institute, Dublin, where he received the TESOL diploma.

In 1971 he became a teacher educator, first in Tilburg and after two years in Amsterdam. He now works for the Amsterdam Institute of Education and the Graduate School of Teaching and Learning of the University of Amsterdam. He has taught a variety of courses in initial teacher education, ranging from English phonetics and phonology to the English speaking world and from 20th century literature to English for classroom management. He has been active in the professional development of teachers, organising courses on new technologies in language teaching.

.

Other dissertations in this series

Braaksma, M.A.H. (2002). *Observational learning in argumentative writing*. Amsterdam: Graduate School of Teaching and Learning, University of Amsterdam (120 pg.). ISBN: 90-9015630-5

Broekkamp, H.H. (2003). *Task demands and test expectations. Theory and empirical research on students' preparation for a teacher-made test*. Amsterdam: Graduate School of Teaching and Learning, University of Amsterdam (152 pg.). ISBN: 90-9016618-1

Trinh Quoc, L. (2005). *Stimulating learner autonomy in English language education. A curriculum innovation study in a Vietnamese context*. Amsterdam: Graduate School of Teaching and Learning, University of Amsterdam (222 pg.). ISBN: 90-7808-701-3

Saab, N. (2005). *Chat and Explore. The role of support and motivation in collaborative discovery learning*. Amsterdam: Graduate School for Teaching of Learning, University of Amsterdam (127 pg.). ISBN: 978-90-7808-702-1

Löhner, S. (2005). *Computer based modeling tasks. The role of external representation*. Amsterdam: Graduate School of Teaching and Learning, University of Amsterdam (123 pg.). ISBN: 978-90-7808-703-8

Sins, P.H.M. (2006). *Students' reasoning during scientific computer-based modeling. The impact of epistemology, motivation and communication mode*. Amsterdam: Graduate School of Teaching and Learning, University of Amsterdam (126 pg.). ISBN: 978-90-7808-705-2

Kieft, M.H. (2006). *The effects of adapting writing instruction to students' writing strategies*. Amsterdam: Graduate School of Teaching and Learning, University of Amsterdam (118 pg.). ISBN: 90-78087-07-2

Kneppers, H.C. (2007). *Leren voor transfer. Een empirisch onderzoek naar de concept- en contextbenadering in het economieonderwijs*. Amsterdam: Graduate School of Teaching and Learning, University of Amsterdam (166 pg.). ISBN: 978-90-7808-708-3

Pijls, M.H.J. (2007). *Collaborative mathematical investigations with the computer: learning materials and teacher help*. Amsterdam: Graduate School of Teaching and Learning, University of Amsterdam (120 pg.). ISBN 978-90-78087-10-6

.