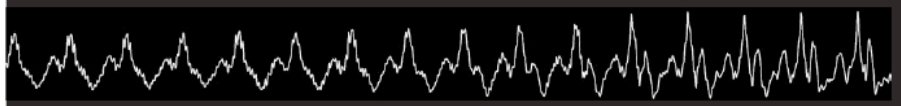


On Variation and Change in Diphthongs and Long Vowels of Spoken Dutch

Irene Jacobi

On Variation and Change in Diphthongs and Long Vowels of Spoken Dutch

Irene Jacobi

Uitnodiging

Voor het bijwonen van de verdediging van mijn proefschrift

"On Variation and Change in Diphthongs and Long Vowels of Spoken Dutch"

Op vrijdag 13 februari 2009
om 14.00 uur
Agnietenkapel
Oudezijds Voorburgwal 231
Amsterdam

Aansluitend receptie

Irene Jacobi

On Variation and Change in Diphthongs and Long Vowels of Spoken Dutch

ISBN/EAN 978-90-9023884-5

Typeset by the author with the LaTeX Documentation System
Cover: *Ei* and *ui* by the author
Printed in the Netherlands by PrintPartners Ipskamp, Enschede (www.ppi.nl)

On Variation and Change in Diphthongs and Long Vowels of Spoken Dutch

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college voor promoties ingestelde
commissie, in het openbaar te verdedigen in de Agnietenkapel

op vrijdag 13 februari 2009, te 14.00 uur

door Irene Jacobi

geboren te Bad Cannstatt, Duitsland

Promotiecommissie

Promotores:	Prof. dr. ir. L.C.W. Pols
	Prof. dr. F.P. Weerman
Co-promotor:	Dr. J.P.A. Stroop
Overige leden:	Prof. dr. P.P.G. Boersma
	Prof. dr. J. Harrington
	Prof. dr. V.J.J.P. van Heuven
	Prof. dr. R. van Hout
	Prof. dr. M. van Oostendorp
	Dr. M.E.H. Schouten

Faculteit der Geesteswetenschappen

CONTENTS

1. General Introduction 1
1.1 Introduction 2
1.2 Present Research Objective and Outline 4
1.3 Literature on Dutch Long Vowels and Diphthongs 7
1.3.1 Auditory-Articulatory Description of Vowel Quality 8
1.3.2 Diphthong Quality 11
1.3.3 Long Vowel Quality 14
1.4 Summary 16

2. On Measuring and Analyzing Vowels and Diphthongs 17
2.1 Introduction 18
2.2 Aspects of Duration, Speech Mode, Speech Rate, and Context 18
2.3 Acoustic Cues to Vowel Quality 19
2.3.1 Formants 20
2.3.2 Whole-Spectrum Representations 23
2.3.3 Concluding Remarks 24
2.4 Normalization Procedures 25
2.5 Conclusion 26

3. Preliminary Study on /ɛi/ 29
3.1 Introduction 30
3.2 Data 31
3.3 Analysis 32
3.3.1 Formant Analysis 33

3.3.2 Durational Aspects . . . 35
3.3.3 Bandfilter Analysis . . . 36
3.3.4 Comparing Formants and Principal Components . . . 38
3.3.5 The Position of /ɛi/ in Relation to /a/, /i/, and /ɛ/ . . . 39
3.3.6 Temporal Diphthong Structure . . . 41
3.4 Summary . . . 42

4. 70 Speakers - An Acoustic Analysis Considering Speaker Backgrounds . . . 47
4.1 Introduction . . . 48
4.2 Corpus . . . 49
4.2.1 Speaker Distribution and Social Encoding . . . 50
4.2.2 Regional Encoding . . . 50
4.2.3 Recording Situation . . . 52
4.2.4 Segmentation and Choice of Vowels . . . 52
4.3 Analysis . . . 55
4.3.1 New Dimensions: PC's . . . 55
Vowel Space . . . 57
Effect of Noise . . . 59
4.3.2 Normalization . . . 63
Relative Onset . . . 64
Relative Degree of Diphthongization . . . 65
4.4 Results . . . 66
4.4.1 Males versus Females . . . 70
4.4.2 Higher versus Lower Socio-Economic Status . . . 72
4.4.3 Testing Interactions of 'Level of Education' and 'Sex' . . . 73
4.4.4 The Effect of Speaker Age . . . 74
'Age'/'Age^2' Effects within the Low Educated . . . 75
'Age'/'Age^2' Effects within the High Educated . . . 76
4.4.5 Differences between Age Groups . . . 77
Old Generation . . . 78
Mid Generation . . . 78
Young Generation . . . 79
Effects within the Low Educated Group . . . 80
Effects within the High Educated Group . . . 80
4.4.6 Effects of Region . . . 84
Effects of Region on the High Educated Speakers . . . 84
Effects of Region on the Low Educated Speakers . . . 85
4.5 Summary . . . 85

5. Perceptual Dissimilarity of Acoustic Differences . . . 91
5.1 Introduction . . . 92
5.2 Method and Material . . . 93
5.2.1 Stimuli . . . 93

5.2.2 Procedure 96
5.3 Results 98
5.3.1 Overall Response Behavior and Acoustic Distances 100
5.3.2 Age Dependent Response Behavior and Acoustic Distances . . . 102
5.4 Summary 105

6. On Speech Variation and Social Behaviour 109
6.1 Introduction 110
6.2 The Structure of Variation and Change 110
6.2.1 Sociolinguistic Approach 111
6.2.2 On the Origin of Sound Change 111
6.2.3 Social Relations, Identity and Social Cognition 113
6.2.4 Summary 117
6.3 Interpretation of the Results of the Present Study 118

7. General Summary, Limitations and Prospects 123
7.1 Hypotheses Reconsidered 124
7.2 Limitations and Future Prospects 126

English Summary 129

Nederlandse Samenvatting 131

References 134

1. GENERAL INTRODUCTION

Abstract In this dissertation, long vowels and diphthongs of contemporary casual Dutch as spoken in the Netherlands are investigated. The main concern will be to discover structures in variation and interconnected changes within the Standard Dutch vowel system of the last decades. This first chapter provides an introduction and motivation for the underlying variation research; it outlines the background and summarizes literature related to the topic. After the introduction and the subject specification, a short description of the articulatory-auditory vowel space is given, followed by a discussion of the literature on the phonetic quality of the long vowels and diphthongs of the last decades.

1.1 Introduction

Speech is most commonly and naturally used as an interaction medium in social settings. Along with communicating meaning, the acoustic signal is a product of physical properties and changes, as well as of more generally all those factors that form the identity of the speaker, such as social affiliation or family origin. The choice of words but also the way they are realized differs from speaker to speaker, as well as within a speaker. Even more, from an acoustic point of view, each utterance is unique.

In this study, we will concentrate on variation in the realizations of Dutch vowel phonemes. Next to variation caused by anatomical differences, the articulatory-acoustic variation between speakers often turns out to be regionally and socially structured. The objective of our investigation will be the phonetic variation between speakers that is caused by other factors than speaker-specific biological attributes of the vocal tract. In addition to this inter-speaker variation, present at a certain point in time, the sound system of a language is in a state of flux, and sounds that once were contrasted get merged and vice versa. In this research, we will consider both kinds of vowel variation: synchronic and diachronic.

The present research was triggered by the appearance of a new diphthong variant: In the beginning of the 1990's an ear-catching pronunciation in Dutch was noticed, a lowered variant of the diphthong /ɛi/. Stroop (1998 [140]) documented the phenomenon, claimed that it was primarily produced by young and highly educated progressive females, e.g. from the world of art, research, and politics, and predicted that men of the same status would soon apply the pronunciation pattern as well.

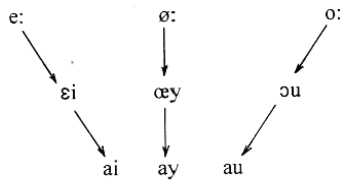


Figure 1.1: *Shift of long vowels and dipthongs in Dutch according to Stroop (1998). The first two rows show the long vowels and diphthongs of the Dutch vowel system. Here, the arrows indicate the recent movements within this vowel system, and, in the bottom row, the resulting new forms.*

Stroop stated that the standard pronunciation of the three Dutch diphthongs /ɛi, œy, ɔu/ (also referred to as /ɛi, ʌy, au/) had been lowered in the late 20th century, and the phenomenon had become widespread since then. Figure 1.1 shows his perceived changes. The diphthongs /ɛi, œy, ɔu/ are lowered to /ai, ay, au/. The lowering of these diphthongs in the articulatory-auditory space drags along the long vowels /eː, øː, oː/, which, by being lowered as well, fill in the empty space previously occupied by the diphthongs. Stroop named this pronunciation variant *Polder Dutch*[1], and suggested that the lowering phe-

[1] see Stroop's website http://www.hum.uva.nl/poldernederlands

nomenon could be the Dutch counterpart of British 'Estuary English'. 'Estuary English', named after the banks of the Thames and its estuary, is expected to become the future RP (Received Pronunciation).

With the term 'Polder Dutch' Stroop wanted to refer to the Dutch political term *poldermodel*, a model of political consensus in the seventies that brought economic growth. According to Stroop, the changes in society induced by the model supported individualism and informality, and the end of authorities such as the norm ABN (Stroop, 1998 [140]). ABN, the abbreviation of *Algemeen Beschaafd Nederlands*, is the term for the Dutch speech standard, meaning 'general cultivated or civilized Dutch'. Stroop argued that women made the most of the new possibilities and Polder Dutch started in the seventies, with women's emancipation bringing along a looser attitude towards language norms. He also stated that, following Labov's findings (Labov, 1994 [84]), the tendency to lower diphthongs seems to be a rather natural language change[2], once the prestige of a narrow articulation in the 16th century was lost. In the neighbouring languages English and German the cognates of Dutch [ɛi] or [œy] are fully open diphthongs, starting with a low vowel (compare Dutch <ijs> [ɛi] vs. English or German <ice> or <Eis> [aɪ], and <huis> [œy] vs. <house>, <Haus> [ɑu]). A look at the Middle Ages reveals that these developed from long /i/ and long /u/ respectively (for Dutch see Janssen & Marynissen, 2003 [62], for English Fennell, 2001 [37]). To others than Stroop, the new Polder Dutch variety simply showed that a few patterns from rural varieties, the so-called "plat Nederlands", found their way into (informal) Standard Dutch (Janssen & Marynissen, 2003 [62]).

Whatever the source, research confirmed the perceived change in quality of the diphthong /ɛi/ to almost [aɪ]. To investigate whether the lowered realization of /ɛi/ predominated in females as opposed to males, Edelman (1999 [33]) and van Heuven et al. (2002 [156]) used recordings of 'Het blauwe licht' ("The blue light"). The latter was then a regular TV-show where two presenters discussed a recent event, or relevant issues with invited guests belonging to the Dutch avant-garde. Having measured the magnitude of formant change between onset and offset of the guests' diphthongs, the investigators concluded that within this homogenous group of 'avant-garde' speakers, the women's diphthongs were lowered more than the men's. For the females, their data also show longer diphthong durations, together with lowered onsets and stronger movement. The onset of the reported female variety of /ɛi/ was therefore close or even identical to the Dutch monophthong /a/ (Edelman, 1999 [33]).

Since this change in pronunciation was first noticed amongst younger well-educated women from the upper middle classes, including women working in universities, left-wing politicians, artists or authors, van Heuven et al. (2002 [156]) suggested calling it *Avant-garde Dutch* rather than Polder Dutch, as the latter might lead to the wrong conclusion of

[2] One of Labov's principles of linguistic change is that "... in chain shifts, the nuclei of upgliding diphthongs fall ..." (Labov, 1994 [84], p.116).

a geographic epicentre, whereas it truly qualifies as a sociolect.

Research by van Bezooijen et al. (2001 [151, 153]) investigated how people value and differentiate speech assigned to ABN, Polder Dutch, and speech strongly affected by dialect. To the younger subjects taking part in her experiment, Polder Dutch was as highly appreciated as ABN, or even more appreciated, though it was thought of as not as "beschaafd" (cultivated) as ABN. In contrast, Polder Dutch was less appreciated among elderly people. Furthermore, young females identified themselves more with the new variety than young males.

All of these investigations suggested that for /ɛi/ indeed a new pronunciation pattern had arisen, and that its appearance is, or at least was, sex-dependent. However, most of the above-mentioned studies on the ear-catching lowering included only speakers of the avant-garde, and investigations have been restricted to the diphthong /ɛi/. Testing a phenomenon only where it is expected or predicted by a theory is a common procedure in linguistics, yet, one that might produce biased results. Reliable conclusions can only result from testing other assumptions included in the hypothesis as well - namely the implicit predictions for the non-target group; there is a need to find out whether the new pronunciation pattern is indeed only apparent within the avant-garde, as assumed by the previous investigators.

Investigations have been restriced to /ɛi/, but the existence of chain shifts suggest that more vowels changed in interdependence with /ɛi/. As an example, a vowel shift of crucial importance during the 16th century that marked the end of Middle Dutch was the diphthongization of <ij> and <uu>. Spreading from the dialect of southern Brabant and from within the lower classes of Holland, [iː] became [ɛi], and [yː] became [œy], and the new patterns became part of later Standard Dutch. Both diphthongs had already been part of the Middle Dutch phoneme inventory. These days, <ei> and <ij> are homophonous.

Following this, other vowels might differ as well within speakers whose /ɛi/ is lowered. The previous investigations on /ɛi/, and the hypotheses they were based on, led to the subject of the present research; analyzing the variation – and its presumed social structure – in the Dutch long vowels and diphthongs.

1.2 Present Research Objective and Outline

With respect to Dutch vowels, the most recent realization that has been documented to diverge from a previous standard, is the pronunciation of /ɛi/. Here, a social markedness was attributed to its lowered and more strongly diphthongized realization, as well as a sex-specific occurrence. Our first, general hypothesis will therefore be:

- The realizations of the Standard Dutch vowel phonemes show sub-phonemic variation that is socially marked.

Testing the general relation between the pronunciation pattern on the one hand, and sex, education and age on the other will clarify whether the pronunciation of 'well-educated' speakers indeed differs from that of other speakers, and whether the term 'avant-garde Dutch' matches the appearance of the pronunciation variant. Previous studies were mainly limited to the speech of avant-garde speakers. An analysis of vowel variation in larger corpora of speakers is lacking for Dutch, which gives rise to the second hypothesis of our research:

- While the well-educated (the avant-garde) have lowered /ɛi/, led by the females, the phenomenon is not apparent in other speakers.

One hypothesis in the previously mentioned studies on /ɛi/ was that highly educated women lead in the lowering process. The studies of Edelman (1999), and van Heuven et al. (2002, 2003) seemed to have proved this. To test these findings, the emphasis will have to lie on both an adequate method for the analysis and comparison of vowel qualities between various speakers, including males and females. For variation research, special attention has to be paid to gender (the cultural attribute) differences contrary to sex (the biological attribute) differences. The third hypothesis is:

- Vowel space sizes (to be defined later) differ, and gender differences may be caused by anatomical differences between the sexes: When comparing realizations across speakers and sexes, a speaker's realized vowel quality needs to be defined in relation to the size of the individual's vowel space.

A more detailed research question is whether onset lowering, longer duration and stronger diphthongization of /ɛi/ are entangled as reported in Edelman (1999) and van Heuven et al. (2002, 2003). Lowering and diphthongization are entangled in the method of measurement, whereas duration is known to be affected by sex (for Dutch vowels see Koopmans-van Beinum, 1980 [77]).

Measuring and comparing various speakers' realizations will thus require the application of procedures for inter-speaker normalization. The latter should make different speakers' data comparable by reducing speaker-dependent physical attributes while keeping variation without getting artifacts. A principal component analysis on bandfiltered spectra as described by Plomp et al. (1967 [117]), could be applied to variation analysis as a more objective method than formant analysis for measuring and comparing the quality of vowels. Our fourth hypothesis is:

- Principal component analysis on barkfiltered spectra are a more objective method of measurement in vowel variation research than formant analysis.

To take into account the occurrence of systematic vowel shifts, next to the analysis of /ɛi/, the pronunciation of the other Dutch diphthongs and the diphthongized long vowels will

yield a more complete picture of the pronunciation variation and vowel changes. The fifth hypothesis is:

- The long vowels and diphthongs of Dutch vary interdependently. If the pronunciation of /ɛi/ is changing, the diphthongs /œy/ and /ɑu/, and the long vowels /eː/, /øː/, and /oː/ are, too.

A recent schematic articulatory-acoustic description of these vowels is given in figure 1.2. To find out more about speech changes in motion in general, these contemporary vowel variants will be measured and compared under aspects of speaker sex, age, and (social) background. Thus, next to investigating to what extent the pronunciation of the vowels of Standard Dutch speakers varies, factors that possibly form a speaker's pronunciation pattern will be analyzed. Hereby we expect to get further insight into the interrelation of social alteration and spoken language as one of the routes to the emergence of variety and language change. We will argue that acoustic differences in realization between speakers that are related to the speakers' background data are caused by differences in their acoustic input.

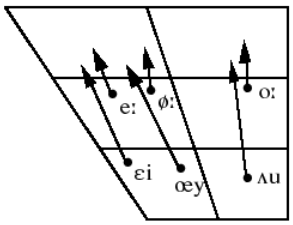


Figure 1.2: *Articulatory-acoustic schema of Dutch long vowels and diphthongs taken from Gussenhoven (1999 [43]): The beginning of each long vowel and diphthong is marked by a dot; the arrows direct to their endpoints.*

The following chapter 2 will focus on general aspects that we need to consider when acoustically measuring and comparing vowels and diphthongs. In chapter 3 a preliminary analysis of /ɛi/ is provided to determine how a perceived vowel lowering can be captured acoustically. Also, it is tested to what extent vowels in the 'spontaneous speech' mode can be analyzed reliably, and whether this can be automated by a principal component analysis on barkfilters in contrast to a formant analysis. The central chapter 4 develops the method of chapter 3 to compare vowel realizations between speakers, and describes the acoustic analysis of the diphthongs and long vowels of 70 Dutch Standard speakers in relation to the speakers' background. The perception experiment presented in chapter 5 was set up to verify that the sub-phonemic acoustic differences found in chapter 4 can be perceived as well. In chapter 6 we investigate social behavior and how it is related to the appearance of variation. In that chapter we will also present literature that underlines the connection of perception and production. Chapter 7 will summarize our research on variation within the Dutch long vowels and diphthongs and its limitations, and will identify prospects.

However, first, research on variation holds that there is something to diverge from, and considering spoken Dutch, presumably, this would be a pronunciation standard. To consider (changes in) variation, in the next section, we will first gather general attributes

of the Dutch speech standard, followed by a more detailed documentation of the vowel quality during the last decades.

1.3 Literature on Dutch Long Vowels and Diphthongs

Lacking methodical acoustic analyses, descriptions of spoken Standard Dutch have usually been phonological-descriptive[3], and thus strongly related to the writing system.

Considering Modern Dutch, the reciprocal relationship of standard speech and the writing system goes back to the beginning of the 19th century, when for the first time spelling rules were officially published. To keep pace with the development of spoken Dutch, spelling rules are officially changed from time to time; the last spelling reforms took place in 1954 and 1994. Since the 19th century, the main principle for written Dutch has been the striving for a phonological spelling, based on 'educated' speech (Janssens & Marynissen, 2003 [62]). A popular and still common definition of the Dutch spoken standard has been that this proper spoken Dutch has no traces of a speaker's area of origin (see e.g. Jespersen, 1929 [63]).

Apart from variation in realization due to the speaker's regional background, there is still a large variety of possible pronunciations that lie within the boundaries of this definition. Accordingly, the *Nederlandse Taalunie*[4] defines the spoken standard as "...the varieties of Dutch spoken all over the Netherlands excluding dialects...". Due to the inclusion of variation in the definition of the standard, it is difficult to assign articulatory-acoustic categories to the phonemes of what is called the spoken standard, and any grapheme-phoneme alignment or phonological-phonetic boundaries will be abstract rather than physically clearly defined. This vagueness is central to another common definition: interpreting a standard language not as something uniform but rather as an abstraction of a usage description, an abstract norm (compare Kloeke, 1951 [74]).

In 1895, ABN, the acronym of *Algemeen Beschaafd Nederlands* was introduced as a term for the Dutch speech standard. Algemeen Beschaafd Nederlands means 'general cultivated or civilized Dutch', suggesting that people who do not speak ABN are not civilized. Politically correct or not, the term ABN implies that there is a social attribute attached to this pronunciation of Dutch, namely 'well-bred'. The number of speakers who use ABN is reported to have grown in the 20th century (Janssens & Marynissen, 2003 [62]). If this is the case, either the number of 'well-bred' people increased, or the ABN pronunciation has been adopted more generally, or the criteria have become more lenient. Also, it is

[3] Cf. Smakman (2006 [133]) for an extensive review on the history and definition of Standard Dutch.

[4] These days, the Dutch, Flemish and Surinamese governments coordinate their language activities in a language union called *De Nederlandse Taalunie* (abbreviated to NTU). This Dutch union is an association established by the Dutch government and the government of Flanders. Within this policy organization, Dutch and Belgians work together on various Dutch language fields, including the standard language to be used by authorities, language education and humanities. In 2004, Suriname joined the union.

often reported that what is referred to as ABN (in terms of the Dutch speech standard) has changed over the years. With respect to the actual Dutch speech standard, an audible change has been asserted compared to the middle-class ABN of the 50's and 60's [62].

In conclusion, the following attributes were assigned to the Dutch Standard: It is the speech of the educated and 'well-bred', it has no traces of a speaker's origin, and it is in a state of flux. To account for existing variants and documented changes within the long vowels and diphthongs of Modern Standard Dutch, in the next section, previous descriptions of the Dutch vowel qualities are considered for reference.

1.3.1 Auditory-Articulatory Description of Vowel Quality

Vowel realizations are usually perceived in categories that match the phonemic system of the language one uses, and have therefore more often been described in a normative way or on the phoneme level than in terms of phonetic variation. Since many definitions of vowel quality in the Dutch phonetic literature are transcriptions based on auditory-articulatory categories, we will start with a short outline of the application of phonetic symbols and their interpretation.

Across and within languages, vowels and the symbols they were transcribed with had often been used inaccurately. More congruent transcriptions came with articulatory models of vowel production, firstly resting on x-rays. Seminal for the objectivity of vowel quality descriptions and the associated articulations were Jones's *cardinal* vowels (see Jones, 1967 [65]), a vowel system with reference to the most peripheral tongue positions as anchors of vowel articulation. After its introduction at the beginning of the 20th century, the vowel system was later implemented by the IPA[5], where, based on a phonemic principle, a separate symbol is provided for each distinctive sound (see fig. 1.3 for the monophthongs). For a phonemic or broad transcription, the symbols are written within oblique lines; when placed between square brackets they represent a narrow phonetic transcription, encoding phonetic variation and allophones[6] (IPA, 1999 [56]).

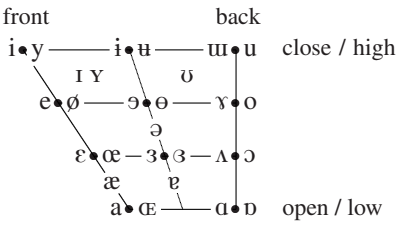


Figure 1.3: IPA vowels: Vertically, the schematic vowel space is based on openness and tongue height. Horizontally it is based on tongue position (fronted or backed), with the unrounded vowels to the left side of the dots, and their rounded counterparts to the right.

[5] IPA, the acronym of the *International Phonetic Alphabet*, a notational standard for the phonetic representation of all languages, provided to the academic community world-wide by the International Phonetic Association (also IPA). The latter was established in 1886 in Paris and is the major as well as the oldest representative organization for phoneticians (with the journal JIPA and the conference ICPhS).

[6] e.g. /ɑxt/ versus [ʔɑxtʰ] for a German pronunciation of the digit 8

Although transcriptions are useful in many ways, studies are manifold that show the influence of the listener's speech background such as the size of his/her vowel inventory, and expectations on the perception of vowel categories or the assignment of phonetic symbols (e.g. Liberman et al., 1957, [89], Cohen et al., 1963 [18], Terbeek, 1977 [145], Dioubina & Pfitzinger, 2002 [29], Iverson et al., 2003 [58], Magnuson & Nusbaum, 2007 [93]). Mees & Collins (1983 [96], 2003 [21]) reported how the Dutch perceive and interpret German or French diphthongs, and their errors in pronouncing English. Cross-language discrimination studies indicate a shift from a language-general to a language-specific pattern of vowel- and consonant-contrast perception already during the first year of life (Polka et al., 1994, 1996 [119, 118]). And when trained adults' transcripts of speech (segments) across or within linguistic boundaries are compared, significant interrater differences are found, especially for narrow transcriptions (as in Shriberg & Lof, 1991 [131], or Cucchiarini, 1993 [22]).

Yet, most descriptions of vowel quality are based on articulatory-auditory transcriptions, also for Dutch. This has changed only recently as methods of signal analysis became generally practicable and more easily accessible. With acoustic correlates to the transcriptions, rendition and interpretation can be objectified. (A further possibility would be articulatory measurements, but due to the sparseness of articulatory data, and the difficulties in accessing articulation, acoustic measurements of vowels have generally been preferred.)

Besides the vowels used in loanwords, and a schwa occurring in unstressed positions, Dutch is said to have twelve vowel phonemes, traditionally divided into short and long ones, plus three diphthongs (Moulton, 1962 [103], Booij, 1995 [13]). Measurements in a corpus of sentences from Dutch public news broadcasts showed the shortest mean duration for schwa, followed by short vowels, then the vowels /i/, /u/, /y/ (before /r/ they are lengthened), followed by the long vowels, and finally the diphthongs with the longest duration (Klabbers & van Santen, 2000 [71]). Though duration generally adds to the classification of Dutch vowels, it is heavily influenced by context and speech condition (cf. the following chapter). The Dutch long vowels /oː/, /eː/, /aː/, /øː/ are said to have the corresponding short vowel phonemes /ɔ/, /ɪ/, /ɑ/, /œ/. However, the long vowels and their short counterparts differ not only in duration but also in spectral composition, at least for /aː/: Shortening a phonemically long /aː/ resulted in the perception of the short vowel phoneme, but the opposite effect of perceiving a long /aː/ was not found when lengthening the short vowel phoneme (Nooteboom, 1980 [107]). Thus the cues for a short versus long vowel distinction must include more than duration.

Most vowel research has been carried out on the analysis of monophthongs, with diphthongs being comparatively neglected, partly as a result of traditional phonological theory (Zonneveld & Trommelen, 1980 [167]). Phonologically, diphthongs behave as monophthongs, the presence of the glide being phonemically irrelevant (Moulton, 1962 [103]).

In phonetic terms, monophthongs are often referred to as 'single target' vowels, since they aim at only one articulatory target gesture, whereas diphthongs require two target specifications to represent their changing nature (Lehiste & Peterson, 1961 [88]). In conventional transcription following the IPA chart, diphthongs are described by a sequence of two phonetic symbols, representing the two articulatory gestures, often with a bottom tie bar to show the phonological unity of the segments (e.g. [ɑu] or [ɑ͜u]). These days, though a monophonematic transcription of diphthongs is accepted, there is a consensus that a diphthong is not a sequence of two monophthongs (Lehiste & Peterson, 1961 [88], Holbrook & Fairbanks, 1962 [52], Gay, 1968 [39], Ladefoged, 1972 [85]).

Next to the *genuine* diphthongs /ɛi/, /ɑu/ (also referred to as /ɔu/) and /œy/ (also referred to as /ʌy/), and the aim of our investigation, Dutch is said to have some so-called *pseudo*-diphthongs /aj, oj, uj, iw, ew, yw/. Cohen (1971 [17]) proposed treating only genuine diphthongs as unitary segments. For the Belgian variant of Standard Dutch, Collier et al. (1982 [19]) found that the main difference between genuine and pseudo diphthongs, both articulatorily and acoustically, lies in the dynamics of movement. Though in an auditory-acoustic study with synthesized dipthhongs, the perceptual distinctiveness between genuine and pseudo-diphthongs turned out to be less clear to listeners in Dutch, pseudo-diphthongs can be distinguished from genuine diphthongs in terms of articulatory dimensions, speech errors, and phonological rules (Collier & 't Hart, 1983 [20]). Also, the production of pseudo-diphthongs shows a greater rate of formant change. Regardless of the distinctions between genuine and pseudo-diphthongs, in the present study we will concentrate on the genuine diphthongs only, and on the long vowels.

Regarding diphthongs, Modern Standard Dutch begins with no longer contrasting the diphthongizations of <au> and <ou>. The stagnating pronunciation difference between these diphthongs is situated around the turn of the 19/20th century, with the last grammar in 1911 to distinguish the pronunciations of <au> from <ou>[7] (Den Hertog, 1911 [28]). Our literature research on the long vowels and diphthongs of Modern Standard Dutch will start in the first half of the 20th century, after the pronunciation of <au> and <ou> had merged.

Though there is the Dutch *Uitspraakwoordenboek* by Heemskerk and Zonneveld (2000 [48]), unlike the German *Ausspracheduden* (Dudenredaktion, 1990 [31]) or the *English Pronouncing Dictionary* (Jones, 1997 [66]), there is no tradition of Dutch pronunciation codification or generalization of the IPA (see 1.3.1, page 8) within Dutch dictionaries. The English standard pronunciation RP (Received Pronunciation) for example, was first defined by Jones at the beginning of the 20th century by a quantitative-qualitative transcription (Jones, 1997 [66]), and a redefinition in 1990 by Wells shows the changes that

[7] Modern Standard Dutch still encodes the two spelling variants of homophonous <au>/<ou> and <ij>/<ei> respectively, as besides the main principle of a phonological spelling, another important principle for spelling is the rule of derivation that takes into account etymological differences.

RP underwent in the course of almost a century (Wells, 1990 [165]).

To get a grip on possible changes in the Dutch standard pronunciation, we will have a look at descriptions of the qualities of the Dutch diphthongs and long vowels by several phoneticians and phonologists throughout the last century, starting with descriptions of /ɛi/, the object of the most recent investigations. Most of the literature on vowel pronunciation, however, does not consider aspects of social pronunciation differences, and the authors usually refer to the pronunciation of a Standard speaker without further explanation of his background.

1.3.2 Diphthong Quality

In 1928, the fronted Dutch diphthong /ɛi/ was described by Zwaardemaker and Eijkman as being articulated with a smaller mouth-opening and a higher tongue than the vowel [ɛ] (Zwaardemaker & Eijkman, 1928 [168]), thus starting with a different vowel quality than the transcription would suggest[8]. A decade later, Eijkman (1937 [34]) wrote of a tendency to widen the first part of the diphthong, [ɛ], to strengthen the contrast between the diphthong parts. Around the same time, the first component of the diphthong /ɛi/ was like the vowel sound of <bek> to Kaiser (1941 [69]).

In 1969 ‘t Hart reported trying to find those formant combinations of vowel segments that were most suitable to represent the diphthongs. He presented fragments of diphthongs of increasing length to listeners (‘t Hart, 1969 [143]) and concluded that /ɛi/ started with [ɛ], and was followed by a movement into the direction of [i], the usual endpoint being [ɪ]. The same can be taken from a short and speaker-specific acoustic description of Dutch diphthongal qualities of three speakers by Pols (1977 [121]). Based on a PCA on spectral bands the starting point for the diphthong /ɛi/ is in the close area of the speaker’s /ɛ/.

To Nooteboom (1976 [106]), /ɛi/ moved from a position before [ɛ] in the direction of [ɪ], thus presumably starting a little lower than [ɛ]. Thus, compared to Zwaardemaker and Eijkman’s description of an articulation closer than [ɛ] in the 1920’s, a lowering of the first part of /ɛi/ is described. Yet, for want of reference, all descriptions are difficult to compare and should be interpreted with caution.

The variation (over time) in the descriptions concerning the components of the diphthong /ɛi/ is also apparent for the other two diphthongs /ɑu/ (also referred to as /ɔu/), and /œy/ (also referred to as /ʌy/). Similar to his description of /ɛi/, Eijkman (1937 [34]) stated a tendency to articulate [ɔ] more open ([34]). In 1949, Kaiser put the first components of the diphthongs of <ei>, <ou>, <ui> (/ɛi/, /ɑu/, /œy/) on a par with Dutch <bek>, <hok>, and English <up>. Also, she assigned the first part of the diphthong of Dutch

[8] Zwaardemaker and Eijkman (1928 [168], p.155): " [ɛɪ] – Het eerste deel van dezen tweeklank heeft gewoonlijk een eenigzins hoogeren vóórtoon dan de enkele klinker [ɛ]. Dit wordt het gemakkelijkst verkregen door den mond ietwat minder open te doen en daarbij de vóórtong wat meer op te heffen." ... "Onbeschaafd klinkt [ɛː] voor [ɛɪ], b.v. in het Leidsche dialect: [kɛːk sɛːn] (kijk hem eens)."

<bruin> and French <brun> the same vowel quality, both cases showing an identical vowel that undergoes a resonatory change (Kaiser, 1949 [69]). The same is said about the diphthong of Dutch <jasmijn> and French <jasmin>. Again, with little reference, especially for the French/English counterparts, the descriptions are difficult to interpret.

Moulton (1962 [103]) commented on differences in the degree of diphthongization of the diphthongs. He described the three diphthongs /ɛi, œy, ɑu/ as 'strongly' diphthongal; all second vowels being allophonically non-syllabic. He mentioned that the pronunciation of the diphthongs is considered 'substandard' when /ɛi, œy, ɑu/ are diphthongized too weakly.

Besides the diphthong /ɛi/, previously mentioned, 't Hart analyzed the diphthongs /ɑu/ and /œy/, and stated in 1969 that /ɑu/ started with [ɑ] and was followed by a movement in the direction of [u], the usual endpoint being [ʊ]. His analyses showed that the first part of /œy/, often referred to as /ʌy/, was normally unrounded: /œy/ started with [ʌ] and was followed by a movement in the direction of [y]; the usual endpoint was [ø]. Unlike the long vowels /eː, øː, oː/, the diphthongs /ɛi, œy, ɑu/ could not be synthesized satisfactorily by a single homogenous spectral composition ('t Hart, 1969 [143]). Several years later, to Nooteboom (1976 [106]), /œy/ moved from a position before /ʌ/ (as in English <but>[9]) to /y/, and /ɑu/ from a position before /ɑ/ in the direction of /u/.

As mentioned before, all descriptions are difficult to interpret for want of an (acoustic) reference. Yet, several Dutch vowel descriptions come up with acoustic data, such as the early detailed acoustic investigations of Dutch vowels by Pols et al. (1973 [122]), and Van Nierop et al. (1973 [157]). Most of these, however, focus on monophthongs (see e.g. Koopmans-van Beinum, 1980 [77], van Son, 1993 [158], Weenink, 2006 [164]).

A recent official transcription of the vowel system of Dutch can be found in the Handbook of the IPA from 1999 (Gussenhoven, 1999 [43]), of which the long vowels and diphthongs are displayed in figure 1.2, page 6, all with a closing movement (moving towards a closer tongue position). Except for /eː/, the description is consistent with the description of Mees & Collins in 1983 [96], where /eː/ is located slightly higher. Also, the diphthong movements (arrows) go a little less far in the description from 1983.

There are two articulatory-acoustic descriptions of the long vowels and diphthongs, 20 years apart, which we will compare in the following. Figure 1.4, p. 13 shows a formant plot of the Dutch vowels of an ABN-speaker[10] after a graph by Koopmans-van Beinum, 1969 (p. 250 [75]). Her representation is comparable to a graph by Mol (1969 [101]), pub-

[9] When describing the quality of [ʌ], Dutch authors often referred to the vowel of English <but>. The interpretation of the symbol [ʌ] following IPA is a lower back vowel. It is doubtful whether the vowel of English <but> was still a lower back vowel [ʌ] in the 1970's. In Daniel Jones' (1967) English vowel space for example, [ʌ] is placed more central in the vowel space than its assigned place in the IPA-vowel chart. The symbol [ʌ] was probably used inappropriately as a transcription for the corresponding Dutch sound to the English vowel of <but>.

[10] refer to subsection 1.3, p. 7 for 'ABN'

lished in the same collection. Additionally, it includes data on the formant movement of the long vowels /e:, o:, ø:/ (in the figure, 'ö' represents /ø:/). The original figure furthermore included arrows to indicate the movement of the vowels when appearing before [r], where all move towards [ə]. Effects of coarticulation on vowel quality and further factors (e.g. suprasegmentalia) will be considered in the next chapter. In figure 1.4, the formant values of the same Dutch vowel phonemes taken from the sound files of the 1999 IPA-handbook[11] have been added in grey. With some caution since the two male speakers' vowel space sizes differ, it can be seen that diphthongization has increased for the diphthongs, and the onset of /au/ ('ou') has become centralized.

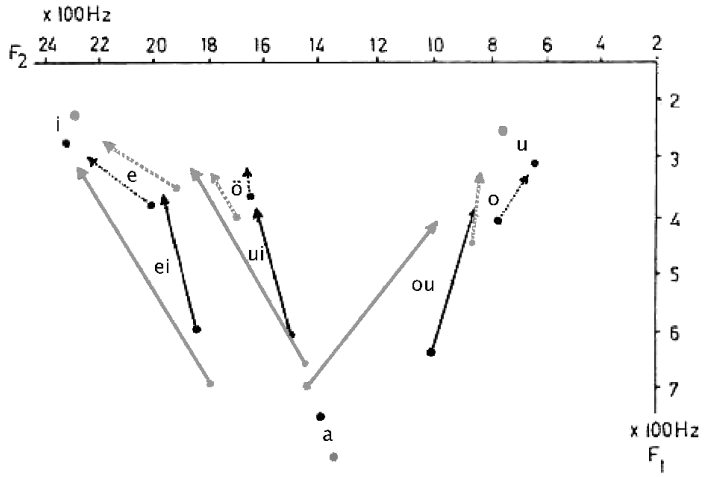


Figure 1.4: *Dutch vowel system after the figure of Koopmans-van Beinum: In black the vowels of a male speaker measured by her in 1969, in grey the measured values of the vowels uttered by Gussenhoven for the IPA-handbook 30 years later. The thick arrows show the formant movements of the three diphthongs /ɛi/, /œy/, /au/ ('ei', 'ui', 'ou'); the dotted arrows show the movements of the three long vowels /e:/, /ø:/, /o:/ ('e', 'ö', 'o').*

In 1999, as mentioned in section 1.1, Edelman (1999 [33]), and van Heuven et al. (2002 [156]) found a lowered variety of /ɛi/ with stronger diphthongization to predominate within avant-garde females. The most recent acoustic descriptions including Dutch long vowels and dipthongs are by Adank (2003, 2004 [1, 2]) and Smakman (2006 [133]). Smakman stated that in his corpus of male news readers, recorded from 1950-1990, the degree of diphthongization of /ɛi, œy, ɑu/ has been stable, though some of his speakers showed slightly lowered first elements. In Adank's research on normalization procedures for variation research, she investigated the vowel qualities of read speech, secondarily considering diphthongs. Contrary to Edelman (1999 [33]), van Heuven et al. (2002 [156]), and Smakman (2006 [133]), her male speakers diphthongized more than the females. However,

[11] see http://web.uvic.ca/ling/resources/ipa/handbook for corresponding sound files

the investigators used different methods of formant normalization, which compromizes a comparison of the results (in chapter 2 we will consider methods of normalization).

In sum, the phonetic descriptions of /ɛi/, /œy/, and /ɑu/ differ in terms of the diphthongal quality and the diphthongs' starting positions in the vowel space. Similar to the Dutch diphthongs, the next section will show that variation is apparent in the descriptions and transcriptions in the phonetic literature on the Dutch long vowels /e:/, /ø:/, and /o:/ as well.

1.3.3 Long Vowel Quality

The Dutch long vowels /e/, /ø/, and /o/ are traditionally transcribed as steady-state vowels, though being realized slightly diphthongized (compare fig. 1.2, p. 6, and figure 1.4, p. 13). To differentiate them from the short vowels, they are usually noted with a length attribute, /e:/, /ø:/, /o:/.

In 1937, Eijkman described a tendency to make 'unreal' diphthongs of long and tense /o:/, /e:/, /ø:/: The diphthongal character showed up as a slight front-upwards movement for [e], and as a little more rounding for /o:/ and /ø:/ (Eijkman, 1937 [34]). According to van de Velde (1996 [154]), the slight diphthongization of /e:, ø:, o:/ started in the 1920's.

Mees and Collins (1983 [96]) mentioned that in 1877, /e:/, /ø:/, /o:/ were noted as being unusual with diphthongal glides, whereas in 1962 a moderately diphthongal quality was mentioned (Moulton, 1962 [103]). However, the pronunciation was considered 'substandard' when diphthongized too much. Accordingly, in 1962, Blancquaert transcribed [ɛɪ] for /e:/, and [ɔu] for /o:/ with some dialects of Brabant (Blancquaert, 1962 [11]).

A few years later, Koopmans-van Beinum (1969 [75]) described the measured diphthongal quality of separately uttered /o:, ø:, e:/ as follows: after a constant beginning, [o] moves in the direction of [u], [ø] in the direction of [y], and [e] in the direction of [ɪ] or [i] (compare figure 1.4, p. 13).

Also in 1969, 't Hart concluded after his speech perception experiment that the first part of /e:/ was identified as [ɪ], followed by a change towards [i]. /ø:/ started as [œ], then changing to [y]. /o:/ started as [ɔ] then changing to [u] ('t Hart, 1969 [143]). However, he also stated that to listeners, the long vowels /e:, ø:, o:/ could be synthesized satisfactorily by a single homogenous spectral composition.

In a phonetic description from 1983 (Mees & Collins [96]), the long vowels /e:/, /ø:/, and /o:/, referred to as 'potential' diphthongs, were alluded to as retaining a narrow glide within conservative Standard Dutch (ABN), but being realized increasingly wider (diphthongized more strongly) by younger mainstream speakers of Standard Dutch (see also Collins & Mees, 2003 [21]). To Mees and Collins, the strength of diphthongization is regionally and socially marked: While steady-state realizations were said to be restricted to areas outside the central conurbation of the Netherlands, the popular speech in the central

conurbation (the 'Randstad' speech) was mentioned to be socially marked by wide diphthongs [96]. This view has recently been adopted by Stroop (1998, 2003 [140, 141]), who said both the diphthongs and long vowels have been diphthongized and lowered, and that the lowering is socially marked. On the assumption that news readers reflect the standard speech, Van de Velde (1996, 2001 [154, 155]) investigated the variation and change in the spoken Dutch of male presenters of the years 1935 to 1993. Within these speakers, he found the pronunciation of /oː/ and /eː/ to change from monophthongal to diphthongal from 1935 to 1993.

Peeters (1991 [113]) referred to the long vowel phonemes /eː/ and /oː/ as "/e(ɪ)/" and "/o(ʊ)/", suggesting they are diphthongized to a certain amount but not lowered.

The slightly diphthongized quality has thus been perceived for decades (Moulton, 1962 [103], Mees & Collins, 1983 [96], Booij, 1995 [13]). Since the diphthongization of the mid vowels has been mentioned for so long, Smakman (2006 [133]) argued that a change in progress in the degree of diphthongization is exaggerated. In his corpus of seven news readers as representatives of Standard Dutch, all except one recorded in the 1990's, the females showed stronger diphthongizations for /eː, øː, oː/ than the males. In Adank's read corpus (2003 [1]), the females did not differ from the males in terms of diphthongization, though the females' long vowels started at lower onset positions.

So, today, the status of the long vowels is still uncertain, and the reported reluctance in attributing a certain diphthongal quality to /eː/, /øː/, /oː/ indicates on the one hand coexisting variation, such as regional or social variants (e.g. suggested by Mees & Collins, 1983 [96]), whereas on the other hand a change in quality over the decades is reflected (following van de Velde (1996 [154]). The different modes of speech that have been used for the studies hinder a clear definition of recent and previously existing vowel qualities (see Koopmans-van Beinum, 1980 [77] for effects of speech mode on vowels). Except for Edelman (1999) and van Heuven et al. (2002) who used spontaneous speech, the speech mode was read or semi-spontaneously carefully pronounced syllables. For American English, early studies already showed that vowels which are used to describe diphthongs do not necessarily reflect the measured formants of the vowel targets (Lehiste, 1961 [88], Potter & Peterson, 1964 [125]). Concerning Dutch vowels, the findings were similar: In spontaneous speech, variability is very large, and the vowel positions indicated by the phonetic transcriptions are not reached by the three long vowels (compare Pols, 1977 [121]). However, the variation in the literature seems to be limited to differences in the degree of diphthongization, whereas there is not much variation in the descriptions of the long vowel onsets. In the previous section, variation had also been assumed when the phonetic descriptions of /ɛi/, /ɑu/ and /œy/ were compared in terms of their diphthongal quality. Contrary to the long vowels, variation was found both in the diphthongs' starting positions in the vowel space as well as in terms of diphthongization.

1.4 Summary

Triggered by recent findings of socially structured variation in the pronunciation of the Dutch diphthong /ɛi/ ('Polder Dutch'), our objective is the analysis of variation in the long vowels and diphthongs of Standard Dutch. Even when excluding regional accents, variation is still included in the Dutch Standard pronunciation. The previously indicated effect of social background (section 1.3) on the pronunciation pattern has not yet led to a consistent speaker control when the pronunciation of Dutch is reported or measured. The understanding of the necessity to control the speaker background is a rather recent development in phonetics, and for Dutch, the social markedness of diphthongs as described by Mees and Collins (1983 [96]) has only been revived some years ago.

The transcriptions and descriptions of the vowel qualities of /eː/, /oː/, /øː/, /ɛi/, /ɑu/, and /œy/ indicate changes in realization through the years as well as synchronous vowel variation. Whereas for /eː/, /oː/, /øː/ the differences in the various transcriptions seemed to be limited to the degree of diphthongization, with little variation in the descriptions of the long vowel onsets, the phonetic descriptions of /ɛi/, /ɑu/ and /œy/ vary in terms of their diphthongal quality and as well in their starting positions in the vowel space.

However, many studies show that transcriptions are affected by the transcriber's own background, and accordingly, variation research that is based on transcriptions of vowels is probably not very reliable, especially when it comes to phonetic detail and differences within phoneme categories. Besides inter-speaker differences in the perception of vowel quality, unknown speaker backgrounds or homogenous speaker data, the usual falling back on traditional ways of transcribing the vowel categories (the strong relation to the writing system), and thereby neglecting potential changes, make an interpretation more complex. To disentangle all effects, different speaker groups should be formed and the vowel realizations of /ɛi/, /ɑu/, /œy/, /oː/, /eː/, and /øː/ should be compared within and between these groups.

As a more objective approach to vowel quality, we prefer an analysis of spontaneous speech within the (articulatory-)acoustic domain. However, though probably being the more objective method in vowel variation research, assessing the vowel quality in acoustic terms brings difficulties as well, even more when analyzing spontaneous speech.

The next chapter describes how vowel quality is measured in acoustic terms, problems that occur in measuring and comparing spontaneous speech data, and the complexity of matching the acoustics of a vowel with the perceived vowel quality.

2. ON MEASURING AND ANALYZING VOWELS AND DIPHTHONGS

Abstract This chapter provides an overview of various aspects that need to be taken into account when measuring and comparing vowel qualities in spontaneous speech. The goal was to find an efficient and reliable method for measuring and comparing vowel qualities across speakers and sexes. In an optimal procedure for our vowel variation analysis, the variance caused by the speakers' differing vocal tract properties should be reduced, whilst linguistic trends rather than artifacts should be maintained. First, acoustic vowel properties and the most important sources of variation within and between speakers are given. Second, two different methods to measure vowel quality acoustically, formant analysis and principal component analysis based on spectral filter output, are reconsidered, as well as procedures to normalize for unwanted speaker-effects in linguistic vowel research. Principal components derived from a principal component analysis built on the vowels /a/, /i/, /u/, which have been unaffected by linguistic trends and delimit the acoustic vowel space, are expected to yield the most objective results when measuring acoustic variation within other vowel phonemes.

2.1 Introduction

In casual speech, the speaker can neglect the articulatory-acoustic quality of a vowel realization up to a certain degree without being misunderstood. This is tolerated by the listener due to the complex speech processing, which weighs the various layers of speech depending on meaning, context, predictability and redundancy, and supports a quick accommodation to the interlocutor's sound inventory. So generally, variation hardly hinders communication, and perceptually, the accommodation to speech variability is an automatic process (Magnuson & Nusbaum, 2007 [93]).

When analyzing the quality of vowels and (social) variation, aspects of duration, speech mode, speech rate and context need to be considered. The following section will outline these effects, followed by sections on two methods of spectral analysis, common methods to normalize for unwanted speaker effects on vowel quality, and conclusions on how to objectively analyze the diphthongs and diphthongized long vowels for our research.

2.2 Aspects of Duration, Speech Mode, Speech Rate, and Context

In this section we will dwell on durational aspects and aspects of context that need to be considered when analyzing vowel quality in acoustics and perception.

Though for example duration generally adds to the identification of Dutch vowels, in spontaneous speech, vowel duration can be heavily influenced by speech style, speech rate, and context. Generally, stressed vowels are longer and they are articulated more accurately (more peripheral in the articulatory-acoustic vowel space) compared to unstressed vowels (for studies on Dutch see e.g. Koopmans-van Beinum, 1973 [76], van Bergem, 1993 [150]). This implies that they are less affected by coarticulation and more reliable in terms of acoustic regularity.

Differences in vowel quality are also apparent in isolated tokens versus read speech, versus spontaneous speech (see Koopmans-van Beinum, 1980 [77]). The (static) spectral values of many studies are taken from accurately read speech recorded in noiseless environments, or from synthesized stimuli, and thus are based on a vowel quality that is rarely reached in spontaneous speech. As an example, vowels taken from casual speech are often not identified as belonging to the phoneme category intended by the speaker when presented out of context. Contrary to vowel realizations of isolated tokens or read speech, vowel realizations from spontaneous speech are often more centralized in the articulatory-acoustic vowel space (Joos, 1948 [68]), and phoneme categories are more diffuse and overlap considerably. Next to effects of coarticulation, following Lindblom's (1971 [91]) argumentation, the main reason for the quality differences is the varying speech rate, causing an 'undershoot' in reaching the articulatory-acoustic target position, with increases in tempo and decreases in vowel duration respectively. Other studies could not find effects of un-

dershoot with increasing tempo, and suggest that compensatory articulations such as an increase in articulatory velocity or an increase in coarticulatory overlap let the articulators nevertheless reach their target positions (van Son 1993 [158], see Harrington & Cassidy, 1999 on this topic [45]).

Nonetheless, besides the possibility of undershoot due to speech rate differences, the speaker's awareness of his production, attention, and communicative intention probably differ for each speech condition. Differences in speech conditions can then result in differences in e.g. prosodic realization, or, more generally, differences in suprasegmental realization.

To control for durational and coarticulatory effects on vowel realizations, only stressed vowels will be considered in our analysis. This will also reduce the strongest effects of coarticulation (influence of neighboring sounds) on the target sounds.

Considering diphthongs, results of experiments with mostly synthesized stimuli showed that the temporal pattern of diphthong movement varies depending on dialect or language, and it was not duration itself that differentiated the quality of the vowels. In a cross-language perception test, Peeters (1991 [113, 114]) studied subjects' preferences for synthesized possible productions of diphthongs. The stimuli were continua of long vowels and diphthongs (/e/, /o/, /ai/, and /au/) with manipulated onsets, offsets and transitions, and subjects were asked to choose the best match of two. The languages included were Dutch, English, and German. The results suggested that not duration but spectral transitions were relevant for the perceived quality.

Gay (1968 [39]) investigated American English diphthongs in three conditions of speaking rate. The results indicated that the offset target positions were variable across different diphthong durations, whereas the onset target position and the rate of change of the second formant were constant. Bladon (1985 [9]), studying fast speech, stated that the integrity of diphthongs is not compromised by offset undershoot since the second targets have little competition.

Following the latter studies, the most promising acoustic values for the analysis of our diphthongal vowels seem to be the spectral composition at the vowel onset position and spectral change. The following section will focus on how spectral information reflects the articulatory-auditory quality of a vowel and in what ways this acoustic information is commonly represented.

2.3 Acoustic Cues to Vowel Quality

The usual acoustic parameter to characterize and differentiate between and within vowels is the spectral energy distribution. When measuring and representing the distribution of energy of vowel spectra, the question arises what forms a better sketch of the acoustic and perceptual vowel cues: whole spectrum representations or formant representations. In the

following sections, these two different ways to acoustically cover vowel quality, and their advantages and disadvantages with respect to the present variation research, are described.

2.3.1 Formants

The importance of the first two formants for vowel quality distinctions has been reported by Helmholtz and others from as early as the middle of the 19th century on (Helmholtz, 1862 [161]). Formants represent concentrations of energy around particular frequencies in the vocal tract. During vowel production, these resonance frequencies occur while the vocal tract is excited by the chain of air pulses that passes through the folds during the open phases of the vibratory cycle. Here, the vocal tract resembles a one-side-closed tube with minimum pressure and maximum velocity at the open end (mouth opening), and maximum pressure and minimum velocity at the closed end (the glottis). Based on Webster's horn equation to describe pressure waves in a duct, natural frequencies that correspond to the vocal tract area functions can be calculated (compare the early investigations on vowels by Chiba and Kajiyama, 1942 [15]).

Formants are defined by the frequency at which air vibration is maximal, the center frequency, and by the bandwidth, which is defined as the range of neighbouring frequencies falling within 3 dB below the peak amplitude. Corresponding to the characteristics of the vocal tract filter, the resonance frequencies differ with the tract size, and shift with its shape (compare fig. 2.1, p. 21). Accordingly, they indicate the articulatory pattern. A detailed description can be found in Stevens (1998 [137]).

The most commonly used acoustic cues for vowels are the first two formants. From the fourth formant on, the formant attributes carry mainly speaker-specific and little vowel-specific information; they contribute to natural sound perception and are of importance for speaker recognition. The first formant (F1) is associated with the degree of constriction, and indicates the vertical tongue position and mouth opening. A narrowing of the cross-sectional front part of the vocal tract is accompanied by a widening in the back part, and results in a decrease of F1 (Stevens, 1998 [137]). The more the jaw is lowered, and the more open the vocal tract, the higher F1 (Lindblom, 1971 [91]). The second formant (F2) is dependent upon the length of the front cavity (Fant, 1970 [36]), and indicates the articulatory front-back dimension. Lip rounding, which increases vocal tract length, was found to lower all formants (Lindblom, 1971 [91]).

An aggravating factor when trying to detect articulatory patterns based on formant tracking is the variety of human vocal tract sizes and shapes, causing their characteristic tract resonances to differ accordingly. Peterson & Barney (1952 [115]) tried to relate formant patterns to vowel qualities in a seminal experiment. A plotted F1-F2 plane, created from the vowel values of uttered /h/_*vowel*_ /d/ -words could be divided into vowel areas. However, the areas overlapped considerably for males, females and children, and absolute

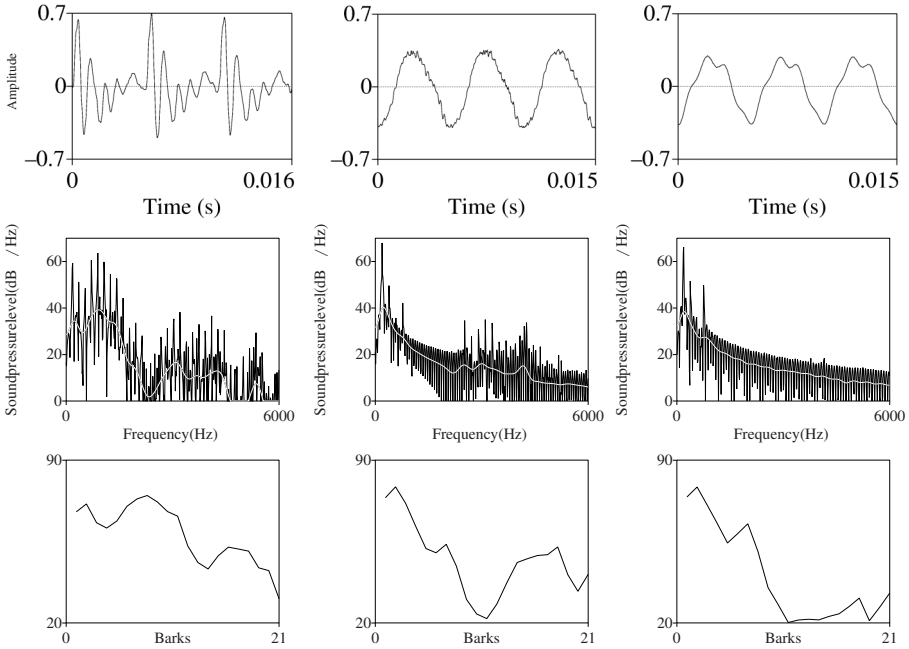


Figure 2.1: Top row from left to right: oscillograms of the vowels /a/, /i/, and /u/, three periods each. Middle row from left to right: spectra of /a/, /i/, /u/, with superimposed smoothed spectra (grey lines). Bottom row from left to right: barkfiltered spectra of /a/, /i/, /u/.

F1 and F2 values as the only parameters were not sufficient to define the produced and perceived vowel categories.

Usually, transformations based on the cochlea's tonotopic bank of filters are applied on formant values[1] (see e.g. Traunmüller, 1981 [147] and Syrdal & Gopal, 1986 [142]). Frequency scales that have been used for vowel analysis such as the Koenig scale, the cochlear position scale, the mel scale, or the Bark scale are all based on auditory findings. Generally, all these scales share a linear Hertz-scale in the low-frequency region and a logarithmic scaling in the high frequencies (see Miller, 1989 [97]). By mapping the differences in Hertz-values onto the perceptual vowel quality or timbre domain, it is taken into account that the same distance (of say 100 Hz) between low tones is experienced as greater than the same distance between high tones.

The perceived vowel quality is also affected by several other acoustic properties. Proposed on the basis of psychophysical considerations, the relevance of spectral relations (feature interaction) as opposed to formant peak extraction was suggested (Miller, 1989 [97]). Crucial for auditory perception are thus not only the positions of the formant peaks but also the distance between them. Several studies have pointed out the role of f0 in rela-

[1] formula by Traunmüller (1990) [148]: $z = \frac{26.81 f}{(1960+f)}$ - 0.53, inverse $f = 1960 \frac{z+0.58}{26.28-z}$

tion to F1 (for instance Miller, 1953 [98], Traunmüller, 1981 [147], Hoemke & Diehl, 1994 [51]). Syrdal and Gopal (1986 [142]) found the F3-F2 difference to be a more accurate cue for the perceived front-back dimension than the F2-F1 difference. An improvement of judgements by taking F3 into account had been reported earlier, as well as a combined influence of the higher formants on vowel perception, as in Delattre et al. (1952 [26]), or Bladon (1983 [8]). In variation studies, however, F3 is hardly ever mentioned.

Although formants are important acoustic vowel features, extracting reliable formant values from a harmonic spectrum is notoriously difficult. To draw any conclusions from the speech signal about the momentary vocal tract characteristics, the spectral envelope has to be separated from the source signal. Measuring formant frequencies by spectrographic analysis was one of the first attempts to accomplish this, but this method is not very accurate (c.f. Monsen & Engebretson, 1983 [102]). Usually, formants are calculated from the mean frequency spectrum of the target segment. For vowels of monophthongal quality, the mean of its mid section is taken, whereas for vowels of diphthongal quality the mean of a section in the beginning of the vowel is taken, as well as the mean of the end section of the vowel.

Linear Predictive Coding (LPC) is used to separate the effects of source and filter. With Linear Predictive Coding (Markel & Gray, 1976 [94]), the signal is broken down into a signal source and LPC coefficients. Knowing that neighbouring consecutive samples are statistically not independent, a subsequent value can be approximately predicted by the weighted sum of preceding values. The signal is encoded as a set of coefficients (usually N=10), representing the vocal tract filter, plus an error signal that represents the difference between predicted and real value (Tempelaars, 1991 [144]). With a flat spectrum the minimized error signal looks like an impulse train (a glottis signal) or noise, and when using only the LPC coefficients a spectrum can be calculated without the interference of the source/error signal. Formants are computed from the LPC coefficients. However, the model is not perfect and the coefficients also contain information about less variable unwanted filters of the vocal tract. These can be factored out by pre-emphasis.

The Praat program (Boersma & Weenink, 1992 [12]) is frequently used in vowel variation studies, and with it its standard procedure to analyze formants using LPC. In the standard setting for formant analysis, after resampling and pre-emphasis, the coefficients are computed with the Burg algorithm and five formants are assigned to candidates in the frequency range from 0 to 5500Hz (for an adult female speaker).

Though formants are clearly indispensable in human vowel perception, formant tracking algorithms produce errors which are not comparable to listeners' errors (Bladon, 1982 [7]). Using LPC it is assumed that there are no prominent antiformants, which could cause problems e.g. when vowels are nasalized and spectral valleys are of significance (c.f. Johnson, 2003 [64]). Also, for formant tracking, the specification of the number of formants is important to anticipate the right peaks: For e.g. Praat [12], the given standard value for

analyzing a human female voice assumes five formants within the first 5500Hz, for a male voice five formants within 5000Hz. Difficulties occur when f0 and F1 interact, formant peaks lie close, or when formants are low (e.g. in high back vowels).

The spectral interaction of frequency bands lays open one of the problems of depicting and comparing formants. Especially with high fundamental frequencies formants are hard to define. The splitting and merging of formants and antiformants entails the assignment of spurious formant peaks or 'missing' peaks, which causes errors in the serial numbering of the formants, and thereby problems in the further processing of the results. There is hardly a formant-based study that does not mention a hand correction of data.

The mentioned integration or adjoining effects on the acoustic level are comparable to some perceived formant integration effects. Following Chistovich et al. (1979 [16]), perceived formant averaging or integration occurs within a critical distance of 3 to 3.5 Bark for two-formant signals. Also, spectral amplitude relations and spectral density seem to determine vowel quality (Chistovich et al. 1979, Ito et al., 2001[16, 57]). Similarly to the results of Bladon & Lindblom (1981 [10]) and Bedder & Hawkins (1990 [6]), Kiefte & Kluender (2005 [70]) conclude that gross spectral properties (tilt) at least contribute to more detailed spectral cues (formants peaks) in vowel perception. However, after their experiments with synthesized monophthongs and diphthongs, they also found that the role of spectral tilt is less important in signals with changing spectral characteristics, and they argue that in this case, it is change over time that dominates the perception of speech.

2.3.2 Whole-Spectrum Representations

Alternatively to formant tracking, from the 1960's on, Dutch researchers used principal component analyses (PCA's) on bandfiltered spectra to analyze and compare vowels (starting with the studies of Plomp et al., 1967 [117], Pols et al., 1969 [123], followed by others [73, 157, 122, 138, 129, 160, 124, 150]). Using this method, it is assumed that there is a finite amount of independent variation that appears in the spectral data. Instead of using raw spectral data, the ensemble of spectral variation is used to arrange the data: The original n-dimensional feature space is rotated in such a way that in the new n-dimensional space most of the variability is placed in the first dimension; the smallest (noisy) variance will end up in the highest dimension.

By analyzing the whole spectrum in frequency bands, one band represents one dimension and the different levels of energy within the band can be described as a coordinate in the single dimension. Changes in the spectrum then reflect different concentrations of constituents, and each spectral sample becomes a point in a multidimensional Euclidean space (compare Pols, 1971 [120]). Based on the principle of combining two or more (correlated) variables into a single factor, the principal component analysis breaks down the information in the bandfilters into its most basic variations. Physical or psychophysical

properties of the human listener can be included by choosing filters of the same properties as the bandfilters of the human ear, frequency-dependent excitation levels, or other neurophysiological or psychophysical scales.

Numerous studies have shown that using multi-dimensional scaling, the first two principal components (pc) dimensions correspond to the frequencies of F1 and F2, though the acoustic properties they have been calculated from do differ from formant tracking algorithms: In the 70's, Klein, Plomp, Pols and Tromp compared a principal component representation of 12 different bandfiltered /h/_*vowel*_/t/-vowels, produced by 50 Dutch male speakers, to the frequency and level data of the first three formants from the same vowel segments (Klein et al., 1970 [73], Pols et al., 1973 [73, 122]). At the time, the formants were derived by drawing the spectral envelope by eye. The results were then compared to the results of the PCA. Corresponding to the ear's critical bandwidths, the sound spectra had been filtered in 21 $\frac{1}{3}$-octave bands from 10 to 10000 Hz with the sound levels as dimensions. To reduce the influence of the fundamental frequency, the energy in the first three filters was added, and the energy in the fourth and fifth filter, resulting in 18 filters altogether. The overall sound pressure output levels were normalized. An analysis of the principal components yielded a reduction to four factors explaining 77% of the total variance in the 600 vowel spectra. The average vowel configuration resembled the logarithmically plotted F1-F2 formant plane. The largest part of variance caused by different vowel classes could be explained by logF1 and logF2, confirming F1 and F2 as the most characteristic vowel features. Adding logF3 further improved the identification score (Klein et al., 1970 [73], Pols et al., 1973 [122]).

2.3.3 Concluding Remarks

As reported, formants are the most important cues to vowel quality, and many experiments with (synthetic) speech have proven that changes in formant frequencies affect perceived vowel categories more than formant bandwidth or spectral bends do (e.g. Klatt, 1982 [72]).

However, finding formant peaks remains a problem when energy is distributed over a range of frequencies, or when formants come close together. Whenever formant tracking is used for vowel analysis, a considerable amount of hand correction of the formant data is reported. This problem does not occur when measuring the spectral distribution by principal components derived from spectral filters.

Next to the finding that no errors need to be corrected by hand, an important advantage using PCA on the whole spectra is that, contrary to formant analysis, no previous knowledge about vowel categories is needed. Thus, the analysis can be reliably automated. In vowel variation research with large amounts of vowel data, a reliable automation of the measuring procedure is highly desirable. Least influenced by expectations, this is likely to be the more objective way to analyze vowel variation, and hence we will apply this method

in the present research. Tracing back the articulatory patterns from the PCA coefficients' could be accomplished by building the PCA on only certain vowels, and by including and referring to vowels with clear or steady articulatory properties. Ideally, the measured vowel quality differences should correspond with perceived differences as well. Research showed that the first pc's are comparable to the first formants, the most important cues to perceived vowel quality. Direct evidence for the correspondence of pc's and perceived differences was given in Klein et al. (1970 [73]).

The basis of our study will be vowels in spontaneous speech. Though we will only use stressed vowels, considering the artificial nature of the measured sounds in the cited studies on formants or PCA on barkfilters, and taking into account the reported effects of speech condition, we expect the vowels of spontaneous speech to be more centralized and/or coarticulated than vowels of semi-spontaneous or read speech. Also, with various speakers, we will have to deal with differences in vocal tracts shaping the acoustic output. The following section will consider a common problem in vowel studies: dealing with inter-speaker differences. An overview will be given of popular methods used in vowel formant analyses to normalize for speaker-specific effects, especially dealing with speaker-effects due to sex, in order to make a speaker-independent comparison of vowel quality possible.

2.4 Normalization Procedures

Numerous methods have been developed to reduce the impact of specific speaker effects and make a representation of e.g. vowels of a speaker community possible. Some procedures use only vowel-intrinsic information and categorize the vowel e.g. by transforming f0 and the formant patterns. Other extrinsic normalization procedures take into account information distributed over several vowels. A more detailed discussion of the different approaches and classification procedures can be found in Nearey, 1989 [105]. Joos (1948 [68]) was one of the first to suggest a speaker-specific normalization procedure. He suggested that listeners relate the phonetic quality to the speaker's point vowels /i, a, u/. A decade later in 1957, Ladefoged and Broadbent confirmed his theory [86]. They shifted the complete vowel system in a synthesized sentence except for the test vowel. If, as a result, the vowel was placed within the acoustic category of another phonologically distinct vowel, the listeners reliably normalized: The same vowel was perceived as belonging to different phonological classes, dependent on the context (embedded in a sentence or separately)[2].

Considering vowels, most research has been focused on methods to generalize vari-

[2] By changing the formants' range of the carrier sentence, e.g. the vowel of <head> could be made heard as the vowel of <hid>. A partial reproduction (by Malcolm Slaney) of the original experiment can be found on the web: http://cobweb.ecn.purdue.edu/~malcolm/interval/1997-056/VowelQuality3.html

ation within vowel classes, for example to enhance robust speech recognition (see e.g. Weenink, 2006 [164]). Contrary to the more general aim of these procedures to minimize variation, for phonetic variation research, the organic variation (variation caused by physical attributes of the individual vocal tract) has to be disentangled from learned and/or acquired variation, the latter being the object of our interest. This implies that the rather complex or abstract relationship of the acoustic regularities within and between speakers versus the irregularities within and between speakers have to be defined and categorized. On the basis of these findings, a normalization procedure can be built. Hence, for the present variation study, the phonemic speaker characteristics had to be further divided into physiological (anatomical) variation versus intra-phonemic variation, and the physiological/anatomical variation will have to be factored out.

Disner (1980 [30]) evaluated several formant normalization procedures to find out about their overall ability to reduce variance while yielding truly linguistic trends and not artifacts. She compared Gerstman's (1968 [40]), Lobanov's (1971 [92]), Nearey's (1977 [104]) and the PARAFAC procedure (Harshman, 1970 [46]). All procedures use a mean or standard deviation of the speaker's whole vowel system. She came to the conclusion that for cross-linguistic studies, or for comparisons across dialects, the application of normalization procedures which use the mean or standard deviation of the vowel system are too effective in reducing interspeaker variance, and might result in procedural artifacts (Disner, 1980 [30]). In 2003, for the analysis of vowel variation in Dutch read speech, Adank favored the Lobanov procedure after evaluating formant normalization procedures (Adank, 2003 [1]).

A method of normalizing vowel data for variation analysis derived from Gerstman's method was introduced by van Heuven et al. (2002, 2003 [156]). They compared formant values of the onset and the offset of Dutch /ɛi/ diphthongs, and concentrated on the height of the diphthong onset and the extension of the glide of the diphthongs. As reference they took a speaker's most extreme high front vowel /i/ (in terms of F2), and the most extreme open front vowel /a/ (in terms of F1), to which the /ɛi/ onset was related. After calculating the Euclidean distance of F1 and F2 in Bark between diphthong onset and /a/, they related it to the distance between /a/ and /i/. The resulting values showed differences between the pronunciation of /ɛi/ by males and females in terms of relative onsets and diphthongization (both related to each speaker's /a/ and /i/).

2.5 Conclusion

For our vowel analysis, we are searching for a reliable method to (automatically) measure variation in vowel realizations across speakers and sexes. As already mentioned, every sound is acoustically unique, even if uttered by the same speaker in a sequence. Considering vowel dispersion, there are considerable effects in terms of speech mode, accent,

stress, and speaker sex. Our speech mode will be spontaneous speech, and the conclusion is to consider only stressed vowels in our variation analysis.

The spectral composition at the target onset position and the rate of change, rather than duration, were found to determine the quality of diphthongal sounds, and when measuring our spontaneous vowel data, we will start our investigations with these acoustic properties.

For (socio-)phonetic variation research and between-speaker comparison, speaker-specific physical variation and 'externally' (environmentally) caused variation need to be separated. The variance caused by the physical differences between males and females needs special attention: Females are supposed to have more dispersed vowels, and for variation analysis, a normalization procedure should account for these sex differences, so that linguistic effects in terms of gender differences can still be differentiated from biological sex.

Our preferred method of analysis is a PCA on bandfilters. The vowels that are used for a PCA should also mark the size of the speaker's vowel space (limited by the individual's most extreme articulatory-acoustic realizations) to be able to capture all of the speakers' vowel qualities, as for example the anchor vowels /a/, /i/, and /u/. Ideally, these anchor vowels should be stable vowels with a minimal probability of carrying gender effects. Then, the variance within these vowel classes is merely due to sex differences and presumably smaller than the variation between the vowel classes, so that, when running a PCA, the variation of the total size of the vowel space due to speaker sex could be filtered out, whereas gender differences of paths within the vowel space should not be affected.

In the following chapter we will apply, and try to test this in more detail in a small variation study on the Dutch vowel /ɛi/, to prepare for the automatic analysis of vowel phoneme realizations in a larger corpus.

3. PRELIMINARY STUDY ON /ɛi/

Abstract This chapter describes a pilot study on the acoustic cues to a perceived lowering of /ɛi/. The aim was to affirm auditory-acoustic properties of lowered versus non-lowered variants that have been found for formants, also in our alternative representation of vowel quality: principal components (pc's) built on bandfilter output. From the spontaneous speech of a dozen speakers, realizations of /ɛi/ were measured in terms of formants and pc's. A principal component analysis on the barkfilter output of all speakers' /a/, /i/, /u/ spectra, as a basis for the calculation of other vowels' acoustics, was a valuable approach to reliably analyze vowel quality. The first two principal components correlated with the first two formant values. Differences between the sexes were smaller for the pc's than for the formants. However, there were sex-independent acoustic speaker-differences whose source has yet to be detected. Following the conclusion of the previous chapter, next to the spectral composition of the onset values, the offset values, duration, and spectral change were considered as possible characterizing variant attributes. To make the /ɛi/ realizations comparable between the speakers, the onset values of /ɛi/ were related to the individual speaker's /a/, /i/ and /ɛ/ values. When related to /a/, /i/ and /ɛ/, the onset values of /ɛi/ indicated the perceived vowel quality.

This chapter is an extended version of Jacobi et al., 2005 [59].

3.1 Introduction

A pilot study on twelve Dutch speakers' vowels was performed for a first measurement of realizations of the diphthong /ɛi/. Recent studies of spoken Standard Dutch have supported an ongoing change in the phonetic quality of this diphthong (see van Heuven et al., 2002, 2003 [156]). Before investigating a larger Dutch spontaneous speech corpus, we wanted to test on a small corpus to what extent our considerations on spontaneous speech given in the previous chapters will play a role in the analysis of vowels taken from spontaneous speech, and to what extent onset and offset, or rate of change are convenient for measurement and comparison of diphthongal vowel quality. And first of all, we needed to see whether existing Dutch speech corpora are appropriate for our needs in terms of objective investigations on the lowered /ɛi/ phenomenon in Standard Dutch. The purpose of the preliminary study on /ɛi/ was to find an automatable method to reliably analyze and define vowel variation in a large corpus of spontaneous speech produced by various speakers. In the previous chapter we described two methods of analysis: formant analysis and bandfilter measurements followed by data reduction such as principal components analysis (PCA). Our preference was towards a PCA on bandfiltered output, and by this preliminary study we want to test to what extent the pc's are comparable to formants and whether our assumptions on the preferred method can be verified.

For an acoustic definition of the /ɛi/-variants, features are gathered that were shared only by speakers who were assigned by listeners to use the more open variant of /ɛi/. The speakers' diphthong variants were analyzed by measuring formants and, additionally, by bandfiltering their spectral energy distributions to find out to what extent the preferred acoustic definition by principal components of a PCA on bandfilter output is as meaningful as formants. Given the content-focused attention and the variable articulatory-acoustic realizations across and within speakers, assigning acoustic cues to auditory effects is a difficult task. Moreover, physical properties can cause major spectral differences in vowels uttered by males versus those uttered by females, and habits, dialects or accents add to the acoustic diversity of the speech (segments). For the present variation research we need to distinguish between acoustic differences that originate in each speaker's individual anatomical properties or speaking condition, and those that were acquired and are of linguistic interest. With variation in pronunciation as a social construct being the aim of the study, we need a normalization procedure that would enable us to compare various speakers' realizations by normalizing for effects of speaker sex, while keeping possible gender effects. As suggested in the previous chapter, the speakers' point vowels /a/, /i/, /u/ could act as references in this normalization procedure.

In the following, the analysis of the vowel qualities of a dozen speakers' /ɛɪ/ realizations are presented. Next to verifying the lowered /ɛɪ/ variant, we were looking for an objective method to reliably analyze vowel quality speaker-independently.

3.2 Data

Six females' and six males' realizations of /a/, /i/, /u/, /ɛ/, and /ɛi/ were taken from the IFA Speech Corpus[1] and a prerelease of the Spoken Dutch Corpus [2] (CGN), both recorded around the year 2000. The IFA Speech Corpus contains recordings of eight Dutch speakers in a variety of speaking styles. From this corpus, only the informally uttered and spontaneously retold speech of the seven adult speakers, four females and three males, was considered. To form a dozen, and for an equal amount of females versus males, two female and three male speakers were added from the spontaneous data pool of the then available prerelease (Release1)[3] of the Spoken Dutch Corpus. The data were approached in the order given in the database, and the additional speakers were selected in such a way that the age distribution was roughly equal between males and females. Of the twelve speakers, the six female (F) speakers were aged 20, 28, 36, 40, 46 and 60, the six male (M) speakers 32, 36, 40, 54, 56 and 66.

The segmentation and labeling of both corpora is comparable. The IFA Speech Corpus is hand-labeled and segmented at the phoneme level (van Son et al., 2001 [159]). The CGN is segmented at the phoneme level as well, and the orthographic transcription was used as a starting point for a lemmatization and part-of-speech tagging of the corpus (Oostdijk et al., 2002 [111]). A broad phonetic transcription has been added for a selection of one million words, and the alignment of the transcripts and the speech files has been verified at the word level.

As mentioned, both corpora were recorded around 2000 and neither of the two has been built with any regard for the aspects and appearance of so-called 'Polder Dutch', or new pronunciation styles in general. Also, the standard transcriptions in the corpora are too broad to carry information below the phoneme level of Standard Dutch, and hence the variation we are looking for. An impartial acoustic variation analysis could thus be based on the speech segments that had been aligned to the Dutch homophones <ij> and <ei> (/ɛi/).

For speaker comparison, additionally, the twelve speakers' realizations of <aa> (/a/), <ie> (/i/), and <oe> (/u/) were selected. At a later point in the study, the Dutch short vowel /ɛ/ was added to see to what extent its acoustic value coincides with the onset of /ɛi/. We wanted to use /a/ and /i/ as references for the relative position of /ɛi/. For the later PCA on bandfiltered output, we planned to build the principal components on the point vowels /a/, /i/, and /u/ which define the articulatory-acoustic space (see section 3.3.3, p. 36). Considering these vowels and their quantal articulatory-acoustic relation (see Stevens, 1972 [135]), we expect less linguistic speaker variation than within other vowel phonemes. Generally,

[1] http://www.fon.hum.uva.nl/Service/IFAcorpus/

[2] http://tst.inl.nl/cgndocs/doc_English/start.htm

[3] The final corpus should later form the basis for a larger variation analysis, but at the time of this pilot study, the final version was not finished yet and there was merely access to a limited part of the data.

and cross-linguistically, extremities of the vowel space, such as /a/ and /i/, show more stability and are produced with less variation than vowels of the space within. In a recent study on Dutch speech where the point vowels /a/, /i/, /u/ from speakers of the Northern and Southern Standard Dutch (Flanders) variants have roughly the same formant values (Adank, 2004 [2]), it is confirmed that /a/, /i/, /u/, have been left untouched by language changes. With the help of these anchor vowels, we might be able to normalize unwanted speaker-effects and identify acoustically the perceived quality.

Our selected vowel phonemes /a/, /i/, /u/, and /ɛi/ appear in a diversity of contexts. Restrictions for the extraction of vowels from the spontaneous speech for the analysis were minimized to capture a preferably large number of realizations. The only criterion for selection was their occurrence in a stressed syllable, as generally, stressed vowels are longer and they are articulated more accurately, which implies that they are less affected by coarticulation and more reliable in terms of acoustic regularity (see Koopmans-van Beinum, 1973 [76], van Bergem, 1993 [150], van Son, 1993 [158]).

In our data pool, the occurance and frequency of words and vowels in the spontaneous speech of our selected twelve speakers differed between speakers and topics. Segments of /a/ were most frequent in the selected speech data (953), followed by /i/ (543), /ɛi/ (428), and /u/ (293). Per speaker, at least ten realizations of each vowel phoneme were included.

The following section describes how the selected vowels and diphthongs were analyzed in terms of formants and principal components on barkfiltered spectra, to get an acoustic definition of lowered versus non-lowered variants of /ɛi/.

3.3 Analysis

Four experienced listeners evaluated the twelve speakers auditorily and put them in two categories: speakers who lower their diphthong and speakers who do not. Eight of the twelve speakers (the females aged 20, 28, 36, 40, and the males aged 32, 36, 56, 66, compare table 3.1, p. 34) were categorized as speakers of the rather openly articulated [aɪ]-like variant of /ɛi/. We will refer to this group of speakers as the 'PL'-group; 'PL' for 'perceived lowering'. The most obvious speakers within this group were the female speakers aged 20, 36, and 40, and the male speakers aged 32 and 36. The remaining (two females aged 46, and 60, and two males aged 40, and 55) will be referred to as the 'noPL'-group (for 'no perceived lowering').

The sample rate for all selected vowels was 16000 Hertz. Sound segments shorter than 0.027 seconds were not considered for analysis. The time step for the spectral analysis was set to 1 millisecond. The window size for the bandfilter calculations was 13 ms; for the formant calculations, the window size was related to the vowel's mean pitch to fit a duration of three periods.

All vowel sounds were formant-tracked and bandfiltered at comparable points in time

with the Praat program (Boersma & Weenink 1992-2006 [12]). For realizations of /ɛi/, the spectral calculation at one tenth of the segment duration was then used as the diphthong onset value, and the spectrum calculated at nine tenths of the segment duration was taken to represent the diphthong offset. Frames at the very beginning and end were thus ignored. This was to exclude major coarticulatory effects at onset and offset, and to avoid measurement artifacts or miscalculations which can occur at segment borders. It left the major diphthong phase with rather unidirectional spectral transitions for measurement.

For monophthongs, the spectrum calculated at the temporal midpoint, presumably at a rather steady-state phase of the vowel, was used for further analysis. The fundamental frequency was measured using the Praat standard analysis. However, f0 yielded no systematic differences between the lowering vs. the non-lowering variants, nor concerning their relation to the anchor vowels, and so f0 was ignored from hereon.

In the next section, we will start analyzing and comparing the speakers' variants in terms of formants, followed by durational aspects that might affect the measurements. Next, the same speech data are measured in terms of principal components derived from barkfiltered spectra. Formants and pc's will be compared and their correspondence and explanatory power in view of the perceived lowering will be discussed. The last subsection 3.3.6 will check on analogies of the dynamic pattern within the lowered versus non-lowered variants.

3.3.1 Formant Analysis

The sound was resampled to 2 x 5500 Hz for female, and 2 x 5000 Hz for male speakers for the extraction of five formants. These differing frequency ranges are usually applied to account for the physical differences in the female and male vocal tube size, and, especially considering back vowels, the larger range covering five formants yields more stable results for the first formants than a smaller range defining fewer formants. The formants were computed using the standard settings in Praat [12]: After pre-emphasis, the LPC coefficients were computed, applying the Burg algorithm on Gaussian-like windows, with a time step of 1ms, not encompassing frequencies below 50 Hz. The first three calculated formants were scaled to Bark and used for further analysis. No hand corrections were applied.

To gain a first insight into the acoustic formation of our data, the mean diphthong onset values of the twelve speakers were compared and related to their mean values of the vowels /a/ and /i/. Table 3.1, p. 34, shows the speakers' mean values of F1 in Bark and F2 in Bark.

For a homogeneous group of the Dutch 'avant-garde', van Heuven et al. (2002 [156]) had found that the F1/F2 in Bark taken from /ɛi/ onsets of the female speakers were lower and closer to /a/ on the /a/-/i/ line than for the male speakers. It appeared that their acous-

Table 3.1: *Table of mean F1 and F2 in Bark of the twelve speakers' vowels (/a/, /i/, /u/ measured at the mid point, /ɛi/ measured at the onset). In grey the speakers who were perceived to lower the diphthong /ɛi/, in white those who were not. Within both groups, the speakers are sorted according to age.*

	F20	F28	M32	F36	M36	F40	M56	M66	M40	F46	M54	F60
F1 /a/	7.2	7.1	6.1	7.8	6.0	6.8	5.5	5.7	5.2	6.3	6.0	6.7
/ɛi/ $_{onset}$	7.1	6.8	5.8	7.0	6.0	7.3	4.5	5.6	5.1	5.0	5.2	6.2
/i/	3.6	3.4	3.4	3.3	4.1	3.3	2.9	3.1	3.3	2.6	3.1	3.1
/u/	4.0	3.8	3.7	3.7	4.1	3.4	3.4	3.5	3.3	2.5	3.1	3.4
F2 /a/	11.7	11.4	10.1	11.2	10.1	11.2	9.8	9.3	10.7	10.7	10.9	10.8
/ɛi/$_{onset}$	11.9	12.3	10.8	11.9	10.8	12.2	10.9	11.4	12.0	11.6	12.2	12.0
/i/	14.0	14.1	12.4	14.1	12.5	14.3	12.8	13.6	12.5	13.8	13.8	13.7
/u/	10.2	8.7	8.6	8.5	8.5	9.0	9.0	8.6	8.7	8.8	8.3	8.1

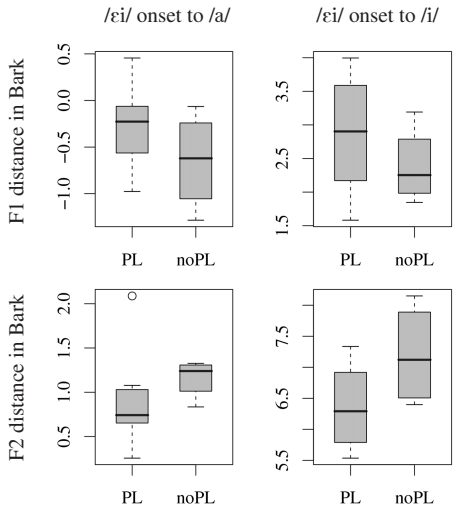


Figure 3.1: *Top row from left to right: boxplots on the distance of the /ɛi/ onset to /a/ (/ɛi/–/a/) and the /ɛi/ onset to /i/ (/ɛi/–/i/) in F1Bark. Bottom row from left to right: boxplots on the distance of /ɛi/ onset to /a/, and /ɛi/ onset to /i/ in F2Bark. 'PL' for the group of 8 speakers who were perceived to lower /ɛi/, and 'noPL' for the other speaker group of 4. The boxplots show the median, the minimum, the maximum, the first & last quartile, and outliers in the data.*

tic data matched their perceptual impression of lowering. An analysis of the first three formants in our data showed the following tendencies: Generally, for those speakers who were perceived to lower /ɛi/, F1 and F2 of the diphthong onsets were closer to the corresponding values taken from /a/ than for the other speakers, and further away from /i/ (compare fig. 3.1). This indicates less articulatory space between their articulation of /a/ and the /ɛi/ onset than for speakers who were not perceived to lower their /ɛi/ (c.f. fig. 3.7, p. 39). So far, the auditory impression thus goes together with the acoustics. All but one (the male aged 66) of the PL speakers showed even a large overlap of both F1 and F2 for the /a/ values and the start of the diphthong /ɛi/. In other words, the acoustics of their /ɛi/-onsets in terms of F1/F2 coincided with the values measured for their /a/ realizations (compare left panels of figure 3.7, p. 39). However, this was not found for all lowering speakers (compare the F2 means of /ɛi/ and /a/ of M56 in Table 3.1, or the F1 means of non-PL speaker M40). The acoustic distance of /ɛi/ to /a/ measured in formants was thus not always a safe indication for the perception of lowering.

In the next subsection, we check whether duration is another cue in defining the perceived variants, or whether it is intertwined in the onset or offset values we measured.

3.3.2 Durational Aspects

To describe the quality of diphthongs, the spectral composition at the beginning and end of the vowel are the most commonly used measurements. Investigations on American English diphthongs in three conditions of speaking rate showed that the offset target positions are variable across different diphthong durations, while the onset target position and the rate of change of the second formant are constant (Gay, 1968 [39], see page 19). These results are comparable to a study on German diphthongs, where the duration of a diphthong influenced the extent of formant transitions within a diphthong (Wrede et al., 2000 [166]). Experiments on American English and German thus indicate that durational aspects influence the overall extent of movement but do not influence the rate of change itself.

The durations of the present diphthong data varied from speaker to speaker and within speakers. To see if the mean of a speaker's onset and offset values is a reliable value for further time-independent comparison, we checked the data on a systematic influence of overall diphthong duration on the onset and offset values. The onset formant values (F1, F2, F3 in Bark), measured at one tenth of the vowel duration, showed rather fixed speaker values and did not correlate with the overall diphthong length (c.p. example in fig. 3.2). Where the diphthong onset showed no systematic correlation with length, the offset values of F1Bark and F2Bark, measured at nine tenths of the total vowel duration, correlated speaker-dependently and only slightly with duration in getting more extreme with an increase of overall diphthong duration. 'Extreme' to their accordant articulatory [i]-like production, which causes F1 to decrease and F2 to increase with increasing duration (all speakers' mean r_{F1}= -.18, r_{F2}=+.44, compare fig. 3.2 of a speaker with a comparably

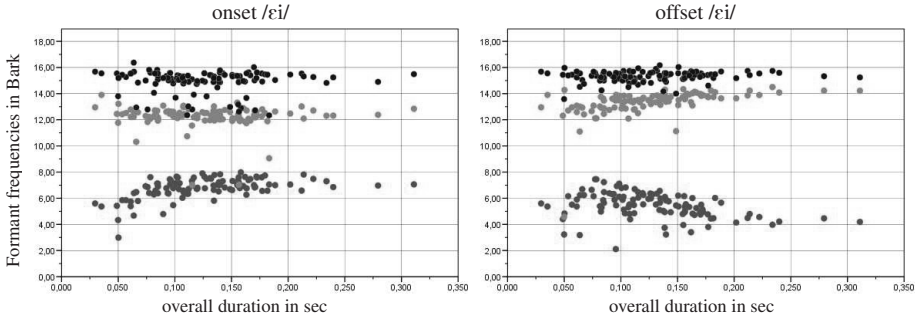


Figure 3.2: *Example of one of the speakers' measured formants in /ɛi/ realizations: Onset (left) and offset (right). Overall diphthong duration on the x-axis with formant frequencies F1, F2, F3 in Bark on the y-axis.*

strong correlation of F1 and F2, r_{F1}=-.37, r_{F2}=+.45). This suggests that further examination can reliably refer to the diphthong onset formant values. The offset values could carry small effects of speaking rates.

As mentioned in section 2.5, an analysis by means of formants entails problems, not only in terms of objectivity, and we argued that a bandfilter analysis of the spectrum may be more consistent and appropriate than a formant analysis. One of these problems is reflected in the left plot of fig. 3.2, p. 35: Some of the points that were assigned to the third formant (black dots) by the algorithm clearly lie in the area of the second formant (light grey dots). The same goes for some of the points that were assigned to the second formant; they are in the area of the first formant. The cause could be e.g. a relatively high pitch. As a result, a harmonic might have been picked as F1, and F2 was wrongly picked as first formant, and the numbering of all following formants will be shifted as well. Usually, in formant studies, these miscalculations are corrected by hand, which we decided not to do. In the next section we therefore additionally defined the acoustic quality by means of a PCA on bandfilter output.

3.3.3 Bandfilter Analysis

The spectra of the same segments that were previously analyzed by means of formant peaks were bandfiltered in order to calculate a PCA on the filter output. For the analysis the barkfiltered spectra were level-normalized to 80 dB. Though the resulting principal components (pc's) are not directly related to vocal tract attributes, articulatory attributes can nonetheless be gathered from the data with the help of point vowels and the relative position of the vowels in question.

The range and width of the set of bandfilters was adapted to psycho-physical findings. Literature shows that up to 500 Hz the critical bandwidth of the human inner ear is consistently 100 Hz, and from thereon, the critical bandwidth grows progressively (section 2.3.1, p. 21). Simulating the physical characteristics of the auditory filters in the human ear, we set our bandfilters with progressively increasing bandwidths, each over an area of one Bark. Frequencies higher than the fourth formant are not specific in their impact on vowel categories. Up to now we have only considered the area of the first three to four formants for the vowel quality analysis, a frequency range up to 5500 Hz. For the principal components analysis on the bandfiltered spectra we used the same frequency range, resulting in 20 filters, overlapping at -3dB (fig. 3.3, p. 37).

A problem for bandpass filtering can be the fundamental frequency, sometimes resulting in empty filter outputs and thus high variance in the lowest filters. Pols et al. (1973 [122]) decided to combine the first three and next two one-third octave filters to make sure that all speakers' fundamental frequencies were represented within the same filter. To get rid of the unwanted influence on the PCA of our one-Bark filter set, the first two filters

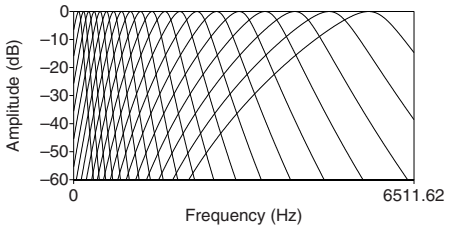


Figure 3.3: *Barkfilters 1 to 20 on a linear frequency scale. In the actual application, the first two filters are replaced by their mean to reduce the impact of f0.*

(with center frequencies 93 Hertz and 188 Hertz) were combined and represented by their mean intensity. The total number of dimensions thus decreased from 20 to 19.

To compare the speakers' vowel structures, the calculated pc-dimensions had to include as little variance caused by individual speaking style variants as possible. Additionally, they should reflect the major articulatory-auditory vowel quality dimensions to facilitate an interpretation of the results. To calculate the pc-dimensions, we decided to use only the speakers' realizations of the three rather stable anchor vowels /a/, /i/, /u/. The large articulatory-acoustic variance in the vowel space between /a/, /i/, and /u/ accounts for the possible differences in vowel quality, in contrast, the speaker variance within the classes of /a/, /i/, and /u/ should be non-linguistic, i.e. non-cultural, and small in comparison. Then, the resulting principal components should be ruled by the acoustic differences between the three reference vowels, with differences within each reference vowel being of minor influence. The dimensions that result from this calculation can then be used to represent the acoustic quality of all other vowels and diphthongs within the vowel space.

The number of realizations of /a/, /i/, /u/ differed between speakers, and in order to include all /a/, /i/, /u/ data and give each speaker and each vowel equal influence in the analysis, the twelve speakers' mean vowel values, 36 in total, were used for the PCA (the eigenvectors were calulcated of the covariance matrix.) Since the number of speakers was rather small and the eigenvectors, which show the weighing of the dimensions, only differed slightly between the sexes (figure 3.4, with reservations given the small amount of data the PCA was calculated on), female and male speakers were analyzed together. The first three dimensions together accounted for 95% of the total variance in the data (fig. 3.5, p. 38). The first two new dimensions pc1 and pc2 each explained far more of the total variance than any of the original dimensions (see fig. 3.6, p. 38).

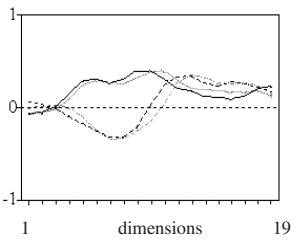


Figure 3.4: *Eigenvectors 1 (continuous line) and 2 (dotted line). PCA on male (black) and PCA on female (gray) speakers' /a/, /i/, /u/.*

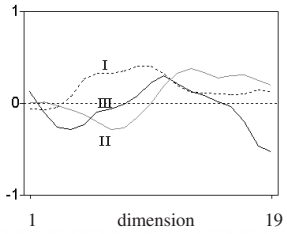


Figure 3.5: *Eigenvectors 1 (I), 2 (II) and 3 (III) of the 19 dimensions of the PCA on all speakers' means of bark-filtered /a/, /i/, /u/.*

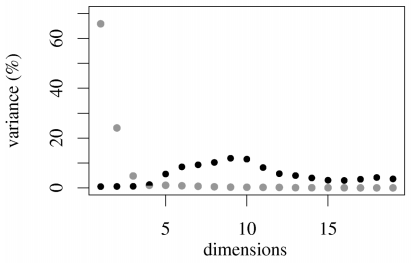


Figure 3.6: *The explained variance in the 19 old and new dimensions in percentage. Black dots: explained variance in the original dimensions (barkfilters). Grey dots: explained variance in the new dimensions (pc's).*

3.3.4 Comparing Formants and Principal Components

As can be seen in table 3.2, pc1 correlated positively with F1Bark, and pc2 with F2Bark. A rotation of the pc1-pc2 plane might bring about even stronger correlations. The inter-speaker variance considering the point vowels /a/, /i/, /u/ was percentage-wise smaller for the pc1-pc2 plane than for the F1-F2 Bark plane (compare table 3.3).

Table 3.2: *Correlations of F1/2/3 with pc's 1/2/3, 2767 speech segments (/a/, /i/, /u/, /ɛ/, /ɛi/ onsets and offsets). On the very right the percentage of total variance explained by the first three dimensions of the PCA on Bark filtered /a/, /i/, /u/ of 12 speakers.*

	$F1_{Bark}$	$F2_{Bark}$	$F3_{Bark}$	expl. var.
pc1	+.81	-.12	+.26	65%
pc2	-.08	+.70	+.10	25%
pc3	-.19	+.05	-.15	5%

Table 3.3: *Mean and SD of F1 (Bark) vs. pc1 for anchor vowels /a/, /i/, /u/ of all twelve speakers.*

	F1 (SD)	pc1 (SD)
/a/	6.36 (12%)	191 (6%)
/i/	3.26 (11%)	131 (8%)
/u/	3.50 (12%)	135 (9%)

As argued before, when it comes to errors and automation of the measurement procedure, barkfiltering the spectral energy distribution reduced by a PCA prevails over a formant analysis. Also, with the pc1-pc2 plane being comparable to the F1Bark-F2Bark plane (compare examples in figure 3.7, p. 39), articulatory and perceived attributes can be traced back from the principal components as well. Hence, the further analysis of the vowel realizations will be carried out in terms of principal components.

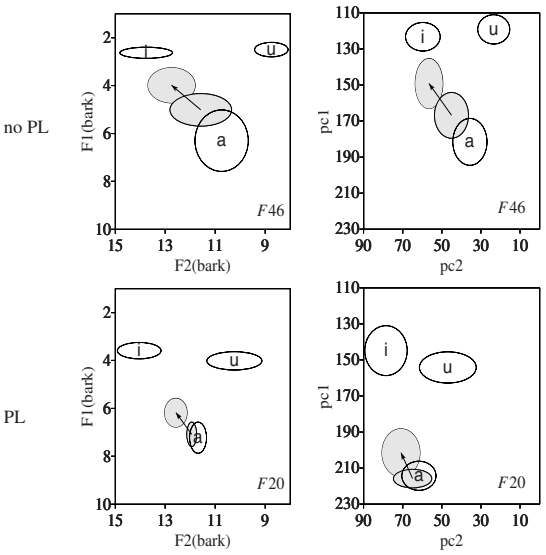


Figure 3.7: *F1-F2 Bark planes (left) vs. pc1-pc2 planes (right) of non-lowering speaker F46 (top) and perceived lowering (PL) speaker F20 (bottom). Mean anchor vowel values with one-sigma ellipses. In grey the diphthong /ɛi/ on- and offsets, the means connected by an arrow.*

3.3.5 The Position of /ɛi/ in Relation to /a/, /i/, and /ɛ/

The aim was to see how the behavior of /ɛi/ could best be described in acoustic terms and measured automatically. We started by comparing the absolute values of onset and offset positions, and then the relative positions in terms of the values of /ɛi/ in relation to values of /a/ or /i/ of the same speaker. The amount of data of this preliminary study was too small for a reliable statistic analysis, and so we can only talk of indications when the results of the measurements are described in the following.

The proximity of the /ɛi/-onset to /a/ and its distance to /i/ in the pc-plane was an indicator for the perceived lowering of /ɛi/, as can be seen in figure 3.7 and in figure 3.8 on page 40. Compared to the measured mean of the non-lowering speakers' values, the lowering speakers' /ɛi/-onsets were closer to /a/ and further away from /i/.

Traditionally, the first articulatory-acoustic goal for the Dutch standard pronunciation of <ei> or <ij> is said to be [ɛ]. For further relativization of the starting position of the diphthong in the articulatory-acoustic /a/-/i/ space, we decided to add measurements of the speakers' /ɛ/ realizations in lexically stressed syllables for comparison. Unlike the stressed vowels /a/, /i/, /u/, or /ɛi/, Dutch /ɛ/ is significantly shorter and influenced more strongly by coarticulation. We therefore opted for fewer realizations but similar phonetic environments, and merely considered /ɛ/ realizations taken from the words <hebben> and <heb>. Table 3.4, p. 40 shows the mean individual pc1 and pc2 values for all measured vowels. Pc1 was the dimension that explained most of the variance in the data, for all vowel classes and for each speaker.

The position of /ɛ/ in the vowel space turned out to be very speaker-specific. The /ɛ/

Table 3.4: *Table of mean pc1 and pc2 of the speakers' vowels (/a/, /i/, /u/, /ɛ/ measured at the mid point, /ɛi/ measured at the onset). In grey the eight speakers who were perceived to lower the diphthong /ɛi/, in white the four who were not. The speakers are sorted according to age.*

	F20	F28	M32	F36	M36	F40	M56	M66	M40	F46	M54	F60
pc1 /a/	214	192	202	192	193	192	182	182	183	181	170	204
/ɛi/$_{onset}$	216	196	197	186	185	203	177	196	191	167	142	203
/ɛ/	181	181	185	178	165	184	175	163	176	172	141	189
/i/	145	132	151	132	141	116	124	127	129	123	119	134
/u/	154	135	154	137	143	125	132	134	131	119	116	134
pc2 /a/	62	45	44	36	35	45	34	36	51	36	36	50
/ɛi/$_{onset}$	65	55	49	41	36	57	46	60	70	45	36	65
/ɛ/	67	60	55	42	53	54	47	55	56	41	44	60
/i/	79	69	68	58	56	59	65	75	60	60	55	72
/u/	47	22	39	13	35	11	36	33	26	23	13	26

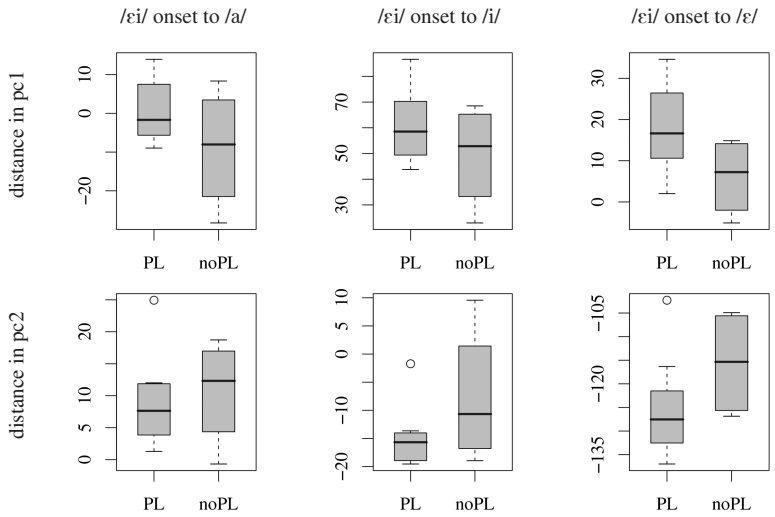


Figure 3.8: *Top row from left to right: boxplots on the distance of /ɛi/ to /a/ (/ɛi/-/a/), /i/ (/ɛi/-/i/), and /ɛ/ (/ɛi/-/ɛ/) in pc1. Bottom row from left to right: boxplots on the distance of /ɛi/ to /a/, /i/, and /ɛ/ in pc2. 'PL' for the speaker group that was perceived to lower /ɛi/; 'noPL' for the group of speakers that was not.*

of five of the eight speakers who lowered their diphthong surfaced around the middle of an imaginery /a/-/i/ line, whereas for the other group /ɛ/ was close to /a/ (compare fig. 3.9, p. 41, and table 3.4), except for one speaker M54. Regarding the acoustic meaningfulness in relation to the perceived lowering, the acoustic results were more interesting when /ɛ/ was considered in relation to the /ɛi/-onset: The diphthong onset of all non-lowering speakers was within or just outside the one sigma ellipse of their /ɛ/ realizations, never in between /ɛ/ and /a/. This was not the case for the PL speakers. Compared to the measured mean of the non-lowering speakers' values, the lowering speakers' /ɛi/ onsets are closer to /a/, further away from /i/, and also further away from /ɛ/.

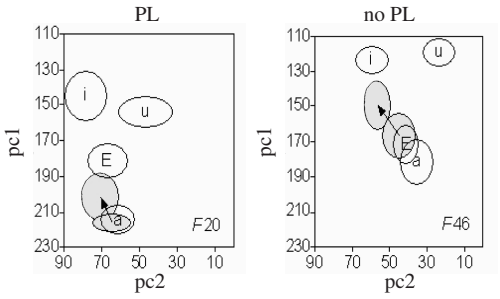


Figure 3.9: *Pc1-pc2 planes with one-sigma ellipses for the vowel phoneme realizations of a female (left) who was perceived to lower /ɛi/, and a female (right) who was not perceived to lower. In grey the diphthong /ɛi/ on- and offsets, their means connected by an arrow.*

These results suggest that relative values that take into account the distance of /ɛi/ to more than one vowel might be the most successful when the perceived difference between lowering and non-lowering has to be defined acoustically. As a short vowel, /ɛ/ is less stable and more affected by coarticulation, and a normalization procedure that includes relative distances to other vowels should then preferably include the longer and more stable anchor vowels /a/, /i/, and perhaps /u/.

Before we summarize the results of the whole analysis with suggestions for the analysis of a larger corpus, we will first check whether the temporal diphthong structure plays an important role in differentiating the /ɛi/ variants.

3.3.6 Temporal Diphthong Structure

A lowering of the Dutch diphthongs could be seen as a movement towards the diphthongs of the surrounding Germanic languages, which begin with a lower - more open articulated - sound. Nonetheless, Peeters (1991 [113]) states that temporal dynamics are the markers for language-specific diphthong differences. Neither onset and offset formant-frequency positions nor formant-frequency glide directions could unambiguously reveal language-specific diphthong properties. The diphthongs of the Germanic languages generally showed overlap in onset and offset formant frequencies, and according to Peeters, his data point to a temporally-based articulatory pattern. If a spectral overlap is assigned for cross-language diphthong formant frequencies, variants within a language probably overlap at least as much. However, acoustic results based on principal components might lead to differing results.

Investigating the temporal structure of the Dutch varieties might reveal more classification cues. Besides further classification of the stage of diphthong change within Standard Dutch, the temporal structure could add to explaining the auditory judgment of the variants, which could not be explained definitely by relative beginning and end diphthong values related to the anchor vowels or /ɛ/.

The mean duration of the diphthong segments was 130 ms, and so the temporal structure was analyzed by measuring in 13 equidistant steps along the diphthongs. The dynamic diphthong patterns varied within and between speakers. Usually, the further one gets in the

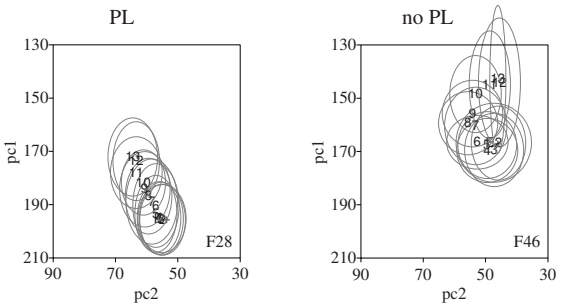


Figure 3.10: *Pc1-pc2 planes of mean /ɛi/ diphthong dynamics and one-sigma ellipses of the 13 measured points in time. Lowering speaker F28 (left) and non-lowering speaker F46.*

duration of the diphthong, the greater the standard deviation of the measured mean points in time. While the lowering speaker's movement in the pc1-pc2 plane was more linear, with increasing pc2 values and decreasing pc1 values (fig. 3.10, left plot), the movement of the non-lowering speakers was less steady, showing a decrease in pc2 values with decreasing pc1 values from the middle of the movement on (fig. 3.10, right plot). However, given the small number of speakers and the diverse temporal movements, correspondences between the diverse temporal movements patterns were difficult to define.

3.4 Summary

Measuring formants is still the most common and preferred method in the literature to visualize the vowel space, and when it comes to articulatory patterns, there is a direct relation between the vocal tract properties and the formants. In this study, pc1 and pc2 of a PCA on barkfiltered /a/, /i/, /u/ yielded comparable results to F1Bark and F2Bark, and are thus easily interpretable in terms of articulation. Moreover, the bandfilter method could be automated without hand corrections and its results revealed less unwanted variance compared to the formants.

Table 3.5, p. 43 summarizes the results of the perceived lowering and its acoustic cues in the measured pc-dimensions of the bandfilter analysis as described in the previous section. Within our sample, the variants were predictable from relative distances in the pc1-pc2 vowel space. The vowel analyses in the first two principal component dimensions revealed that the onset is the most stable attribute of the diphthong. Speakers who were perceived to lower showed acoustic values for their onset of /ɛi/ that were closer to /a/, and further away from /i/. A closer look at /ɛ/ taken from realizations of ⟨hebben⟩ and ⟨heb⟩ revealed that its distance to the /ɛi/-onset contributes to the classification of the two perceived variants. These findings clearly show that other vowels, and the speaker-specific acoustic distances between the vowels need to be considered when defining a single vowel quality.

Besides the relative vowel positions, the acoustic cue that turned out to be most meaningful regarding the perceived categorization of the twelve speakers' /ɛi/ was the differing

Table 3.5: *Table of presence (+) or absence (-) of attributes for female (F) and male (M) speakers of different ages.*

	PL	F20	F28	M32	F36	M36	F40	M56	M66
	perceived lowering of /ɛi/	+	+	+	+	+	+	+	+
1.	/ɛi/ onset overlaps /a/ in pc1/pc2	+	+	+	+	+	+	+	-
2.	pc2 is only increasing in /ɛi/ dynamics	-	+	+	+	+	+	+	+
3.	pc2 /ɛi/ onset ≤ pc2 /ɛ/	+	+	+	+	+	-	+	-
4.	high /ɛ/ in pc1	+	+	+	-	+	-	-	+

	no PL	M40	F46	M54	F60
	perceived lowering of /ɛi/	-	-	-	-
1.	/ɛi/ onset overlaps /a/ pc1/pc2	-	+	-	-
2.	pc2 is only increasing in /ɛi/ dynamics	-	-	-	-
3.	pc2 /ɛi/ onset ≤ pc2 /ɛ/	-	-	+	-
4.	high /ɛ/ in pc1	-	-	+	-

diphthong-dynamics of pc2 in the temporal diphthong movement. Yet, our calculations were based on means, with increasing standard deviations the further one got in the duration of the diphthong, and we had not enough speakers to interpret the movements reliably. We therefore consider the position of the /ɛi/-onset in relation to /a/, /i/, or /ɛ/ a better approach to map the perceived vowel quality. As mentioned, /ɛ/ as a short vowel is less reliable as an acoustic reference; the long anchor vowels /a/ and /i/ are therefore considered better references when defining the quality of other vowels, in this case /ɛi/. The higher orientation of /ɛ/ in the vowel space for the group of lowering speakers might indicate a further change within the vowel inventory of the speaker group, and underlines its instability when considered as a reference.

Generally, the observations in this study call for a closer acoustic analysis of the whole vowel system in future research. More detailed analyses on more vowels by more speakers might reveal a more complete pattern. As indicated in the first chapter, the lowering of /ɛi/ might go together with the disposition or dynamic changes of other vowels. The results also show that, contrary to long vowels, investigations on short vowels such as /ɛ/ are more restricted, especially when considering spontaneous speech. Investigations on positional or dynamic changes of the other Dutch diphthongs /ɑu/ and /œy/ (from words with <au/ou> and <ui>) and the long vowels /eː/, /oː/, and /øː/ (from words with <ee>, <oo>, <eu>) might thus be more fruitful. It might also reveal whether there are pronunciation attributes which come back in all or some vowel phonemes of a speaker, or whether the vowel phonemes are independent considering quality aspects such as lowering or dipthongization. A larger sample of speakers from various age groups could then reveal the temporal order of change within the whole vowel system over the last decades, and whether there is a regular relationship between lowering and an increase of diphthongization.

The present analysis affirms that the relation of the vowels to each other, contrary to absolute measurements, plays a major role when the acoustics are mapped to perceived

vowel quality differences. For the further study, the speaker-individual vowel dispersion in relation to point vowels such as /a/, /i/, /u/ will be elaborated. No clear tendencies for female speakers as opposed to male speakers were found in our sample, but more data are needed to confirm all indications.

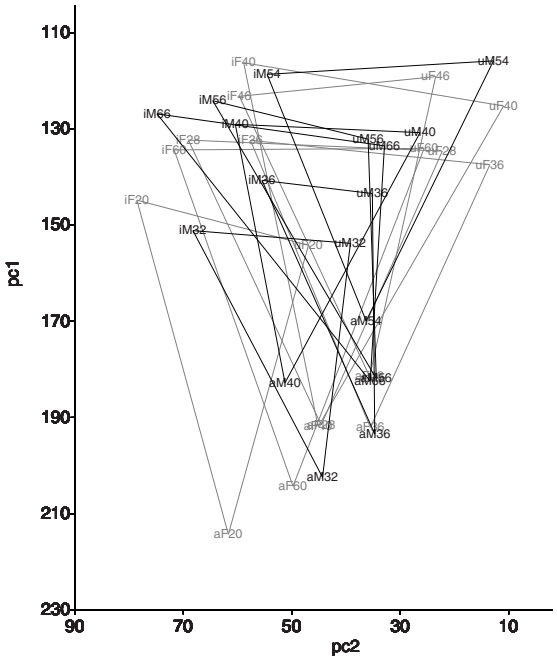


Figure 3.11: *Pc1-pc2 plane with /a/-/i/-/u/ triangles of the six female (gray) and six male (black) speakers.*

In figure 3.11 we plotted the pc1-pc2 plane with each speakers' /a/-/i/-/u/ triangle. The females' triangles are plotted in gray, the males' in black. Considering relative as opposed to absolute values, differences in /a/-/i/-/u/ triangle-size and the point vowel positions in the pc-space showed a larger variability within the females and within the males than a consistent variability between the sexes. A speaker representation entailing both females and males thus seems to be possible in the pc-dimensions. The source of the sex-independent differences has yet to be figured out. Differing signal-to-noise ratios could be a probable reason: Contrary to picking only spectral peaks (i.e. formants), the pc's were built on sequential filters and therefore are as sensitive to areas with little energy, as they are to areas with higher energy. The spectral areas of the vowels with little energy on the other hand are more sensitive to noise. Additional calculations to normalize the influence of these unwanted quality differences on the measured pc-values could improve the speaker mapping. If the differences are indeed mainly due to differences in the signal-to-noise ratio, an automated rotation and linear transformation of the speakers' /a/-/i/-/u/ triangles in the pc1-pc2 plane could be applied. This would further simplify a speaker-independent comparison of vowel qualities.

All in all, we consider the principal component analysis an adequate way to analyze the vowel quality of spontaneous speech automatically and rather sex-independently. Problems of a formant analysis such as the need for hand correction were not encountered by the pc-analysis. The next chapter will focus on the acoustic analysis of vowel quality based on a PCA on the point vowels, and on an appropriate manner to normalize for unwanted speaker variation. In this chapter, the social background of the speakers was not yet considered. Contrary to the IFA speech corpus, of which most of our small corpus was derived, the CGN, meanwhile available as a whole, includes data on the speaker background that are relevant (see chapter 1) when analyzing the structure of social vowel variation. The CGN thus seems appropriate for further vowel research in a larger speaker sample of spontaneous speech, including other diphthongs and the long vowels, with controlled speaker backgrounds. A larger sample of speakers would allow for a reliable statistical analysis taking into consideration the speakers' ages and social backgrounds. Taking into account various factors that could influence a speaker's pronunciation pattern, variation and changes over the last decades can then hopefully be identified for the diverse speech groups.

In the following chapter, a larger corpus of 70 speakers and their vowel realizations in spontaneous speech will be analyzed. Taking into account the results of the present chapter, we will apply a normalization procedure that delimits speaker-differences in the anchor vowel dispersion, and improves the comparability of the speakers' acoustic vowel qualities. The selection of speakers will be controlled in view of their social background to investigate the relationship between social speaker attributes, such as age or education, and vowel variation.

4. 70 SPEAKERS - AN ACOUSTIC ANALYSIS OF DIPHTHONGS AND LONG VOWELS CONSIDERING SPEAKER BACKGROUNDS

Abstract In this chapter, we analyze the vowel realizations in the spontaneous speech of 70 speakers with different social backgrounds. Our aim was to find out how the pronunciation variants of vowels in a representative sample of Dutch speakers coincide with attributes of the speakers' backgrounds. Presumably, the speakers' socio-economic affiliations go together with diverse speech communities. Hence, we expect a classification of pronunciation variants according to the diverse speech communities that are reflected in the speaker background data. For the analysis of vowel pronunciation, a speaker sample was built of 35 males and 35 females, taken from the CGN, half of them categorized as speakers of a higher social stratum, the other half as speakers of a lower social stratum. Here, social status was defined by the level of education and occupation. The speakers' vowel acoustics were calculated by means of a Principal Component Analysis (PCA) on bark-filtered spectra. A normalization procedure was applied to account for influences of noise and unwanted non-linguistic speaker-specific attributes. When related to the speakers' individual /a/, /i/, /u/ acoustics, measurements on the realizations of the Dutch diphthongs and long vowels /ɛi/, /ɑu/, /œy/, /eː/, and /oː/ showed significant differences between social groups and ages. /øː/ was omitted due to its infrequent occurrence. Significant spectral differences were found in terms of vowel onset values and degrees of diphthongization. Most salient were the findings for the vowel phonemes /eː/ and /oː/. For speaker generations with an assigned higher social status changes in the vowel pronunciation patterns were found, in contrast to speaker generations with an assigned lower social status. Given our PCA analysis, no significant differences between the vowel phoneme realizations of females and males were found.

Parts of this chapter have been published in Jacobi et al., 2006 [60], and Jacobi et al., 2007 [61].

4.1 Introduction

In our preliminary study (see chapter 3), the acoustic properties of the Dutch diphthong /ɛi/ from the spontaneous speech of twelve speakers were compared by means of formants and a principal component analysis (PCA) on Bark-filtered spectra (Jacobi et al., 2005 [59]). To be able to interpret the individual variation, the speakers' vowels /a/, /i/, and /u/ were used as reference points on which the PCA was calculated. The resulting first two components pc1 and pc2 (of the PCA on the Bark-filtered spectra of the sound segments) were comparable to F1 and F2 in Bark of the same sound segments.

In the small corpus of twelve speakers, the vowel realizations of speakers who were perceived to lower their diphthongs differed in several acoustic properties from the realizations of speakers for whom no lowering was perceived. Differences between speakers of the lowered vs. non-lowered variant of /ɛi/ showed up in the diphthongs' onset values relative to the speakers' /a/ and /i/. Pc1 was the strongest acoustic cue to the classification of the long vowel variants, and explained 65% of the variance in the data. As a robust and meaningful account that can be automated without any need for manual correction, the PCA on barkfiltered /a/, /i/, /u/ will be employed for the analysis of the following broader vowel corpus.

Our initial results called for further investigations on more data. This includes data on the speakers' backgrounds, as well as realizations of other Dutch vowels, since a change in one vowel phoneme is often accompanied, or usually affects, the quality of other vowel phonemes.

The following corpus analysis will focus on the effects of speaker background data on the vowel acoustics, in addition to age. These effects could not be tested in the previous chapter given the small number of speakers. According to chapters 1 and 2, the occurrence of variants can be classified by the speaker's sex, age, and education. Consequently, we predict for the vowel realizations of the sample of the 70 Dutch speakers that their pronunciations can be differentiated by means of their background data. The results of studies on realizations of the Dutch diphthong /ɛi/ (see Stroop 1998, and van Heuven et al. 2002, 2003 [140, 156]) let us expect our more highly educated females to lower the onset of the diphthong more, and to diphthongize to a larger extent than male speakers of higher social status, or speakers of lower social status. This pronunciation variant was called 'Polder Dutch'. This chapter will examine to what extent the findings for /ɛi/ can be confirmed by the present speaker sample, and to what extent the other diphthongs /ɑu/ and /œy/, or the long vowels /eː/ and /oː/ match this pattern. In line with the previous study, we will focus on the onset and the diphthongization as characteristic vowel attributes. Moreover, effects need to be considered that are the result of differences in the applied acoustic measurement and normalization procedures. Based on general articulatory-acoustic findings that mouth opening and tongue lowering bring about certain acoustic and perceived differences, we

will focus on acoustic similarities and dissimilarities in the vowel positions between the speakers. To normalize for unwanted non-linguistic speaker-attributes, each speaker's individual /a/, /i/, /u/ acoustics will be used as relative measures.

After having normalized the unwanted speaker-specific signal attributes, we expect to find differences in vowel realizations that can be explained by the speakers' sex, social class, and age. These three variables have repeatedly been cited as significant speaker attributes that account for differences in speech behavior between speakers of the same (standard) language (see chapter 1).

To be able to reliably analyze the variables 'sex', 'age', and 'social class' in a representative sample, a corpus of speakers was composed that included an even spread of speakers within these (sub)groups.

In the following, the results of 70 speakers' acoustic realizations of the diphthongs /ɛi/, /ɑu/, /œy/ from words with <ij/ei>, <au/ou>, and <ui>, and their realizations of the long vowels /eː/, /oː/ and /øː/ from words with <ee>, <oo> and <eu> are investigated. Compared to the other vowels, /øː/ was less frequent in the data. Due to the small amount of data for /øː/, it will be omitted from this study. Changes in pronunciation probably affect more phoneme realizations in the Dutch vowel system as well, but here, we will concentrate on the long vowel and diphthong classes.

4.2 Corpus

To measure speaker group differences in vowel realizations of /ɛi/, /ɑu/, /œy/, /eː/, and /oː/ (initially also /øː/), a sample of 70 speakers was taken from the Dutch part of the *Corpus Gesproken Nederlands*. The CGN contains nearly 9 million spoken words from adult speakers, of which over 5.6 million were collected from the Netherlands and about 3.3 million from Flanders (Oostdijk et al., 2002 [111]). The corpus was built to serve the interests of different user groups by a plausible sample of contemporary Standard Dutch from speakers in the Netherlands and Flanders[1]. It was recorded around the year 2000 and includes several subcorpora, characterized in terms of socio-situational settings, communicative goal, interlocutors and medium. For the selection of the 70 speakers, the speaker database was approached in the given order of the CGN, and speakers were chosen or skipped according to the attributes that were essential for a representative sample in this variation analysis. The six CGN speakers that had been selected for the preliminary analysis in the previous chapter were part of the present sample as well, but none of the speakers of the IFA corpus.

Only speech and speakers of the spontaneous sub-corpus were considered (for more

[1] The project was funded by the Flemish and Dutch governments and the Dutch Organization for Scientific Research (NWO). The Dutch Language Union (Nederlandse Taalunie) holds all rights. http://tst.inl.nl/cgndocs/doc_English/start.htm

details see section 4.2.3). Telephone recordings which differ from all other recordings in their recorded frequency range, were excluded. The choice of speakers for the present study was designed by trying to achieve an equal spread in two speaker attributes that have repeatedly been cited to affect phoneme pronunciation over generations: sex and social class. For the study of changes the speakers' ages were ranged as equally as possible.

4.2.1 Speaker Distribution and Social Encoding

Of the available speaker background information in the CGN, the level of education and occupation were the most plausible ones to represent the speaker's social class. Though in the CGN the level of education is split into six ranks, here, to increase statistical power, the ranks were merged into two distinct levels: Speakers who were enrolled in or who had completed university/academy or a college of higher education (Dutch 'hogeschool') were assigned to the class 'high educated', and all others to the class 'low educated'.

Considering the ranking of occupations, the CGN defines nine levels. Again, we merged some levels to form two distinct classes, 'high occupied' versus 'low occupied'. We relied on the occupation ranks of the CGN and merged occupations requiring a higher level of education (such as doctor, lawyer, etc.) and occupations requiring a middle level of education (such as journalist, teacher, etc.) to 'high occupied'. All occupations requiring a lower level of education (nursery school teacher, bank employee, mechanic) or not any level of education (cleaning lady, taxi driver, garbage collector, housewife, unemployed, unfit) were assigned to the class 'low occupied'[2]. In this two-class system, the level of education and the level of occupation turned out to be the same for all but one speaker, and so we decided that the level of education was sufficient to reflect the speaker's social class.

The 70 speakers consisted of 35 females and 35 males. Of the 35 females, 18 were labeled as 'high educated', and 17 as 'low educated; of the males, 17 were labeled as 'high educated', and 18 were labeled as 'low educated' (table 4.1, p. 51). At the time of recording, the 70 speakers were between 19 and 76 years old (compare the plotted distribution of age in fig. 4.18, p. 74). Contrary to the speakers from the previous chapter 3, these speakers were not judged and labeled according to a 'preceived lowering'. In this respect, the data of the 70 speakers have to speak for themselves.

4.2.2 Regional Encoding

All speakers were acknowledged speakers of Standard Dutch in terms of their first, home, and work language, and so the regional background was seen as being of minor importance in the choice of speakers. Still, as part of the speakers' meta data, the CGN encodes

[2] More detailed information on the encoding of all available meta-data can be found on the following site: http://tst.inl.nl/cgndocs/doc_English/topics/metadata/speakers.htm

the place and region of birth, education, and residence. Of these, the region of education and the region of residence were considered as a possibly relevant and evaluable influence on the pronunciation pattern.

The CGN-coding of the four regions is displayed in table 4.2. Speakers were assigned to the central region 1 when their education or residence was in one of the cities of the 'Randstad' or the area within. This central region includes the provinces Noord-Holland, (excluding West Friesland), Zuid-Holland (excluding Goeree Overflakkee) and West Utrecht. Region 2 comprises the transitional areas Oost Utrecht (excluding the city of Utrecht, a 'Randstad' city), the Gelders river area, Zeeland (including Zeeuws-Vlaanderen and Goeree Overflakkee), the Polders, the Veluwe up to the river IJssel, and West Friesland. Region 3, the north-east peripheral region, includes the Achterhoek, Overijssel, Drenthe, Groningen and Friesland. Region 4, the south-peripheral region, comprises Noord-Brabant and Limburg.

With an accumulation of centers for higher education and jobs in the big cities of the 'Randstad', there was an inevitable overlap between a subject's level of education, region of education, and residence region within the CGN, and hence, in the sample of the 70 speakers (compare tables 4.4 and 4.3).

Table 4.1: *Between-subjects factors.*

	levels	N
sex	f	35
	m	35
level of education	h	35
	l	35
region of education	1	15
	2	26
	3	12
	4	17
residence region	1	11
	2	39
	3	7
	4	13

Table 4.2: *Regional coding*

Region 1:	central region
Region 2:	transitional region
Region 3:	north-east peripheral region
Region 4:	south peripheral region

Table 4.3: *The high (h) and low (l) educated speakers' region of residence and education*

		h	l
region of education	1	11	4
	2	9	17
	3	7	5
	4	8	9
residence region	1	8	3
	2	20	19
	3	3	4
	4	4	9

Table 4.4: *Overlap of the speakers' region of residence and region of education*

region of education		1	2	3	4
residence region	1	9	1	1	0
	2	5	24	5	5
	3	0	0	6	1
	4	1	1	0	11

4.2.3 Recording Situation

As already mentioned, the CGN is a spontaneous speech corpus that was built irrespective of the aspects of vowel change. The data in the spontaneous speech part come from diverse recordings: private conversations, interviews, broadcasts, lectures, discussions or meetings. In view of the various recording circumstances, we have to consider that *code switching* might play a role in, for example, a private conversation vs. an interview situation. The private conversations were recorded in circles and situations familiar to the interlocutors, whereas the interview situation might have included an unfamiliar interlocutor or situation. As a result, code switching could have affected the pronunciation patterns. Yet, in the interview situation, the interviewer merely talked to keep conversation going, so that the situation might as well be defined as a spontaneous monologue of the interviewee.

Of the 70 speakers, two speakers were recorded while commenting on the radio, five were recorded during discussions and debates, 16 were recorded during interviews, and the rest during private conversations. These four recording situations were labeled to allow later analysis of code switching effects or level of background noise. However, all speaker data that were not recorded during private conversations were from high educated speakers, whereas the speech of the low educated speakers was never recorded during an interview, but during private conversations. Within the scope of the spontaneous speech of the CGN, effects of code switching can therefore not be analyzed reliably. Hence, we excluded code switching from our list of factors for statistical analyses.

Another effect that is partly dependent on the recording situation is the signal quality of the speech recording. Generally, the speech recorded in interview situations is of much better quality than speech recorded during private conversations. Noises accompanying the speech of the private conversations suggest, for example that the speakers were having dinner during the recording. And considering the broadcast recordings, there are cases of background music. As mentioned in the previous chapter (section 3.4), the recording quality can show up as an effect in pc-dimensions that are based on bandfilters. In sections 4.3.1 and 4.3.2, we will therefore dwell on the effect of noise on the calculated spectra and the pc-dimensions.

4.2.4 Segmentation and Choice of Vowels

For our vowel study, we wanted to consider as many vowel segments from spontaneous speech as possible. For the segment boundaries and vowel classes we relied on the existing segmentations and annotations of the CGN that fitted the research in terms of a broad transcription: a phonemic representation that was based on the orthographic transcriptions of the corpus, and had been generated fully automatically by TreeTalk (Daelemans & van den Bosch, 2001 [24]). The symbols used were derived from SAMPA in such a way that the produced sounds were related to the phonemes of Dutch (Gillis, 2001 [41]), hence,

giving the same symbol to all variants of a phoneme: "E+" to all /ɛi/, "A+" to /ɑu/, "Y+" to /œy/, "e" to /eː/, and "o" to /oː/.

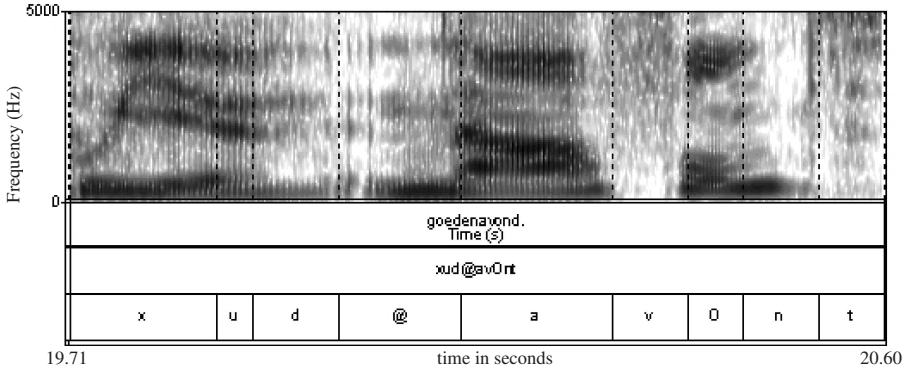


Figure 4.1: *Example of an incorrect phonetic annotation in the CGN: Right after the initial fricative /x/ of <goedenavond>, the spectrogram shows a clear triphthong during the first syllable indicating [xuɪənafɔnt], whereas the annotation sticks to the Dutch Standard [xudəavɔnt]. As a result, the segmentation and labeling up to segment /a/ failed. The CGN labels represent broad phonetic categories based on phoneme classes, so that the fricative labeled with /v/ can as well include voiceless realization such as [f], which is the case in the example above.*

For one million words, the automatic transcriptions of the corpus are reported to have been checked manually (Oostdijk et al., 2002 [111]), but not for all of our chosen data. Disregarding which data had been checked or not, we re-checked all our data manually for errors generated by the automatic labeling and segmentation process. Figure 4.1 shows an example of such a labeling error. (The vowel segment of /a/ in the given example matched our corpus criteria). The realization of <goedenavond> is transcribed as [xudəavɔnt], the Dutch standard, whereas the spectrogram indicates a different realization, namely [xuɪənafɔnt]. Enforcing the standard labeling and annotation resulted in misplaced segment boundaries in the first half of the utterance. A minute number of such suspect transcriptions and segmentations was excluded, as were the more frequent segments with overlapping speakers and distortion noises, altogether approximately 5% of the data. File and speaker names were checked as well for incidentally occurring switches in tier and speaker identities.

Of the spontaneous utterances, almost all (see below) realizations of the vowels /a/, /i/, /u/, /ɛi/, /ɑu/, /œy/, /eː/ and /oː/ in stressed syllables were selected (/øː/ was omitted), in a variety of phonetic contexts. The extraction criterion was based on the presence of lexical stress, as well as on a minimum duration of the vowel (30ms).

To avoid strong coarticulatory influences, vowels from specific environments were excluded. Our aim was, however, to include as many vowels as possible and so only vowels

were omitted that were followed or preceded by /l/, and those that were followed by /r/. Not only does their semi-vocalic character bring about stronger coarticulatory effects on adjacent sounds, Dutch /l/ is usually realized with a secondary approximation in the back (velar/uvular). The [ɬ] realization of /l/ – in contrast to e.g. the German /l/, [l] – can cause a lowered F3 in the more open vowels. Figure 4.2 below shows an example of the influence of such a secondary approximation on the spectrum and formant position of [a] in words like <maar> or <maal>. The same effect can be reached by strongly retracting the tongue root.

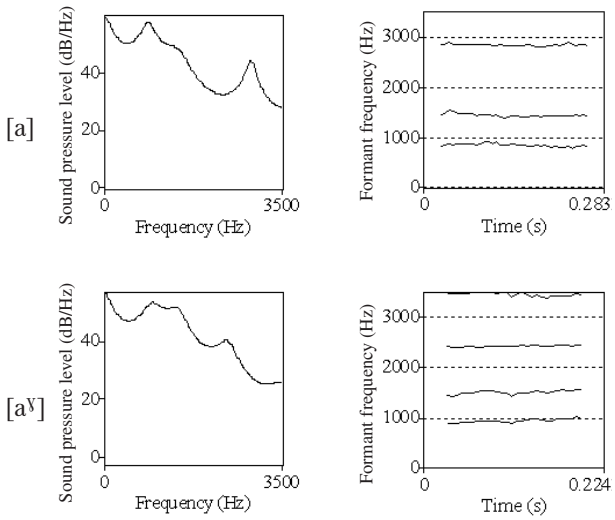


Figure 4.2: *Two example vowels spoken by a trained phonetician: Vowel /a/ realized with and without velar approximation. Top: mean spectrum and first three formants of unvelarized [a]. Bottom: mean spectrum of [aˠ] (with velar approximation) and a low third formant, with the fourth formant appearing at around 3500 Hz.*

The fact that /l/ and especially /r/ strongly influences preceding vowels has often been reported; for Dutch e.g. thirty years ago by Koopmans-van Beinum (1969 [75]) and Pols (1977 [121]). An articulatory characteristic that has been proliferating in the last decade in final or pre-final position is a slightly retroflexal /r/. This popular realization of /r/ shares the characteristics of a vowel and has a strong coarticulatory influence on preceding vowels. Vowels that matched these /l/ and /r/ contexts were therefore omitted. Other coarticulatory influences were accepted, being less severe, and expected to level out due to the variety of phonetic contexts and the large amount of data.

The following section describes how the speech data were analyzed for an inter-speaker comparison of vowel quality. For interpretable and reliable results, various influences on

the acoustic vowel measurements were analyzed. As a result of the investigations, a procedure was applied to normalize unwanted speaker-specific attributes and minimize effects of variable recording qualities.

4.3 Analysis

The common values for an acoustic definition of diphthong quality are beginning and end values, usually taken at one quarter and three quarters respectively of the diphthong duration. The beginning value, and the difference between the beginning and end are then reported as characterizing and differentiating the vowel qualities.

Most Dutch diphthong studies have been conducted on segments of read speech or words or syllables spoken in isolation (e.g. van de Velde, 2001 [155], Adank, 2004 [2], Smakman, 2006 [133]). This results in speech segments that are of longer duration and clearer articulation than is the case for spontaneous speech (refer to section 2.2, p. 18). To optimize onset and offset and get a larger vowel segment, for our spontaneous data we decided to measure at more extreme points of the diphthongs and long vowel segments: at one tenth and nine tenths of their total segment duration. This left out the very first and very last frames, and thereby the strongest coarticulatory effects. The two points in time represent our onset and offset values for the diphthongs and long vowels.

Since more speakers and more variable recording qualities were included, the resulting dimensions of a PCA on the speakers' point vowels were likely to differ from the dimensions calculated in our preliminary analysis (see chapter 3, p. 38). Though in the present chapter the focus was on the analysis by means of pc's, the automatic uncorrected calculation of formants for comparison was pursued alongside. The congruence of the first pc-dimensions with the first formants appeared to change slightly as compared to the preliminary calculations.

4.3.1 New Dimensions: PC's

The acoustic analysis is highly comparable to the one used in our preliminary study. Sample rate, window sizes, time step and measured points in time were the same (see section 3.3), and again, all measurements were done with Praat (Boersma & Weenink, 1992-2007 [12]).

The three corner vowels /a/, /i/, /u/ were analyzed at the temporal mid point of the segment to capture the steady state phase. The spectrum of each sound segment was filtered up to ca. 4200 Hz by using 18 filters with progressively increasing bandwidths, according to the Bark-scale. The mid-frequencies of the 18 filters are listed in table 4.5, p. 56. The overall bandwidth covered the important information concerning vowel quality, including

Table 4.5: *Mid-frequencies of the 18 barkfilters (filters 1 and 2 were combined).*

barkfilter	1	2	3	4	5	6	7	8	9
mid-frequency in Hz	93	188	287	392	505	628	764	915	1086
barkfilter	10	11	12	13	14	15	16	17	18
mid-frequency in Hz	1278	1497	1746	2031	2357	2732	3163	3657	4228

the area of the first, second and third formants. Higher formants include mainly speaker-specific information.

For the analysis by means of a principal component analysis (PCA), the barkfiltered spectra were level-normalized to 80 dB. Each filter covered a frequency band of one Bark and adjacent filters overlapped at -3 dB (compare figure 3.3, p. 37 in the previous chapter). To prevent possible strong variance caused by the speakers' varying fundamental frequency, the first two filter outputs were added up and represented by one mean intensity, resulting in 17 filters and dimensions.

As the stressed anchor vowels are hardly influenced by sound changes or by the speakers' individual speech style (see previous chapter), we took all speakers' /a/, /i/, and /u/ means to calculate the PCA dimensions. The resulting principal components were then used to analyze all other vowel tokens. For the 70 speakers, altogether 11381 /a/, /i/, and /u/ tokens were analyzed, and the new dimensions were calculated on altogether 210 (3 x 70) /a, i, u/ means of the 70 speakers. No hand corrections were applied. The first dimension then explained 65.4% of the variance, the second added another 25.4%, and by the third dimension no more than 3.8% was added. For the data set of more than 12400 measured long vowels (/oː/, /eː/) and diphthongs (/ɛi/, /ɑu/, /œy/), the resulting values of the first dimension correlated positively with F1 in Bark (r=.70), the second with F2 in Bark (r=.72). Figure 4.3 below shows the eigenvectors of the first three dimensions.

As argued in the previous chapter (p. 44), the dimensions could be sensitive to noise.

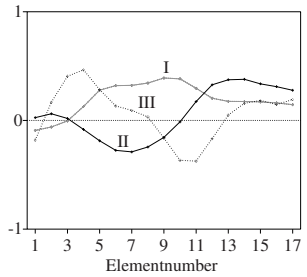


Figure 4.3: *Considering the 210 means of the 70 speakers' Bark-filtered /a/, /i/, /u/, the first eigenvector I accounts for 65.4%, II for 25.4%, III for 3.8% of the fraction variance.*

To what extent the coordinate values had been affected by variable noise levels is investigated in the following sections.

Vowel Space

In figure 4.4, p. 58, /a/-/i/-/u/-vowel triangles, based on each of the 35 female and 35 male speaker's mean /a/, /i/, and /u/, are plotted in the pc1-pc2 plane for a comparison of the vowel space sizes between the sexes. For males and females, the covered areas of the pc-vowel triangles are comparable. Differences in the sizes of the /a/-/i/-/u/-triangles are more salient within the sexes than differences between the sexes.

For comparison, the 70 speakers' /a/-/i/-/u/-triangles based on the averaged first two formants in Bark are plotted in figure 4.6, p. 58. In view of the differences in the resulting triangle plots, we calculated the vowel triangle areas following Heron's formula[3], to test for sex differences in the pc-dimensions as well as in the formant-dimensions in Bark. A t-test on the 35 male areas versus the 35 female areas revealed that the females' vowel sizes were highly significantly larger than the males' sizes in the formant dimensions in Bark ($t(68) = 4.848$, $p<.0001$, the 95% confidence interval (C.I.) ranges from 1.586 to 3.806). When the vowel space sizes were compared in the pc-dimensions, however, the sex difference was not significant ($t(68) = 1.3854$, $p=.0852$, the 95% C.I. ranges from -42.186 to 233.783). Thus, even after applying a logarithmic transformation by means of the Bark scale, the formant vowel space sizes differ significantly between the males and females. The pc-dimensions are comparable between males and females. Considering normalization procedures in general that take into account the speaker-specific vowel space sizes and anchors, and in view of the common finding of gender effects in vowel variation research, this is an important finding for our following acoustic analysis of Standard Dutch speakers, in which we strive to separate sex effects from gender effects. An analysis by means of principal components seems to enable this separation.

For both females and males, the /a/-/i/-/u/ pc-triangles are dislocated more or less in one direction (compare fig. 4.4, p. 58). To make the speaker data better comparable, a primary normalization could be obtained by a linear transformation in the pc-dimensions. Following this, the point of gravity of each speaker's /a/-/i/-/u/ triangle was calculated, and set to zero in each pc-dimension to normalize for the diverse positioning of the triangles in the space. Figure 4.5, p. 58 shows the speakers' vowel triangle spaces after the linear transformation.

While investigating the influences of noise and its adherent spectral attributes, the next section explicates what filtered speech attributes the linear transformation of the triangle midpoints actually normalized.

[3] Following the formula by Heron, the area A of a triangle with the sides a, b, c can be calculated by $A = \frac{\sqrt{(a+b+c)(a+b-c)(-a+b+c)(a-b+c)}}{4} = \frac{\sqrt{(a^2+b^2+c^2)^2-2(a^4+b^4+c^4)}}{4}$

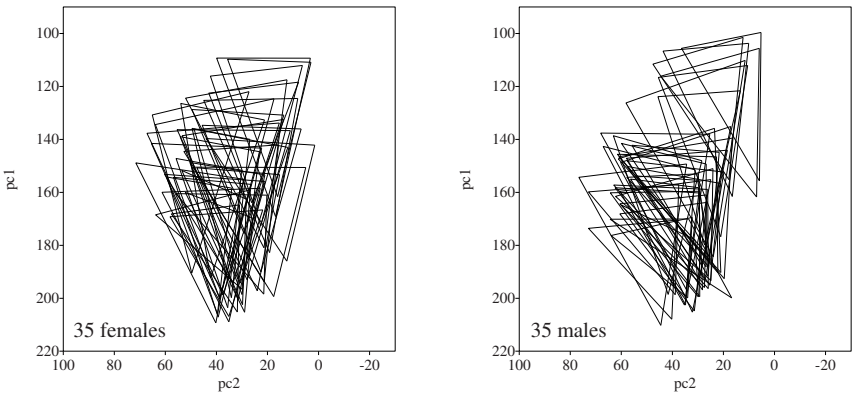


Figure 4.4: */a/-/i/-/u/ triangles in the pc1/pc2 dimensions of a PCA on the 210 mean /a/, /i/, /u/ filter output.*

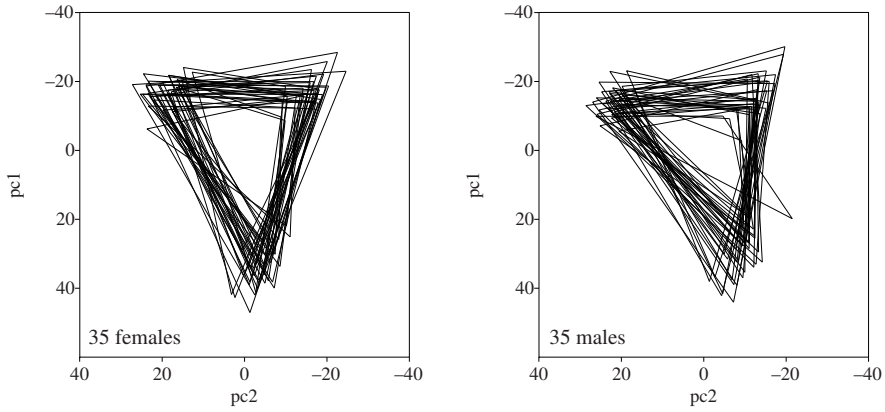


Figure 4.5: */a/-/i/-/u/ triangles in pc1 and pc2 after normalizing the focus to 0 in all dimensions.*

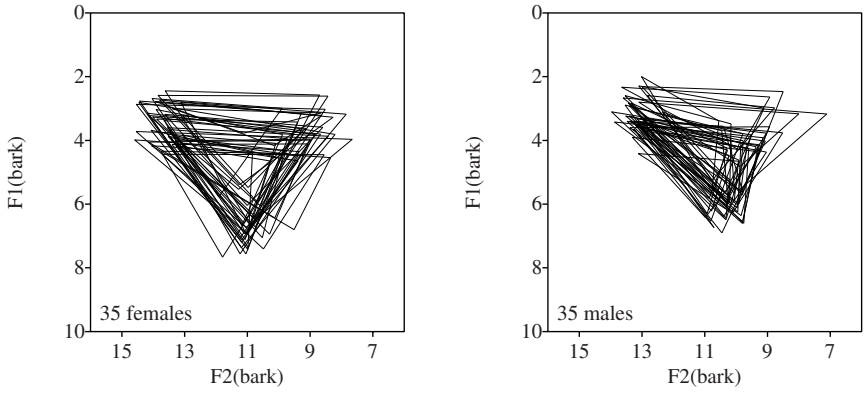


Figure 4.6: */a/-/i/-/u/ triangles of the females' and males' first and second formant in Bark.*

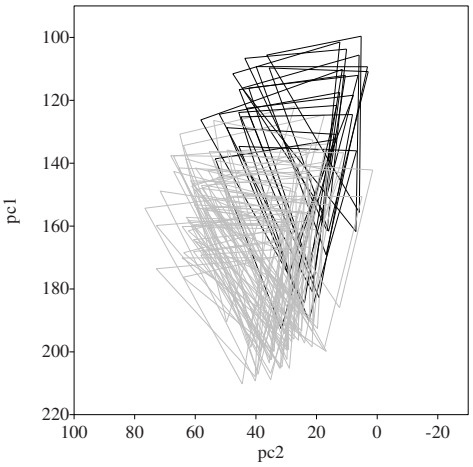


Figure 4.7: */a/-/i/-/u/ triangles in pc1 and pc2 before normalization. The black triangles are from female and male speakers whose speech was recorded in relatively good quality during an interview situation; the grey triangles are from female and male speakers whose speech was recorded in more noisy environments such as private conversations, debates or broadcasts.*

Effect of Noise

When comparing the vowel triangles and the positioning of each speaker's vowel set in the two-dimensional pc-space, inter-speaker differences were obvious that were beyond an attribution to the two sexes, and the variance within the sexes was more striking than the variance between the sexes (fig. 4.4, p. 58). As mentioned before, this variance could coincide with differences in the recording quality and the recording situation (e.g. distance to the microphone and background noise). Various recording qualities are a characteristic of the spontaneous speech part of the CGN, and so we investigated the implications of this variability on our vowel analyses.

Figure 4.7 shows the triangles marked according to good recording situation (interviews, in black) versus relatively bad recording recording situations (debates, private conversations, broadcasts, all in grey). As can be seen, the recordings of good quality from interview situations are generally located more in the upper right part of the two-dimensional pc-plot, whereas the other recordings appear more to the lower left. To get a better view on the effects of recording quality, the specific influence of noise was tested by degrading the quality of recorded speech samples.

Two contrastive speakers, in as far as their vowel space size and location in the pc1-pc2 plane was concerned, were compared in more detail; speakers Z and Y (both males). The speakers' vowel triangles are plotted in figure 4.8, page 60. The speech of one of the speakers (speaker Y) was of rather poor quality: a broadcast recording with music in the background which resulted in a poorer signal-to-noise ratio. The other speaker's speech (speaker Z) was recorded during an interview in a quiet environment, resulting in a good speech-to-noise ratio.

Both speakers' vowel spectra were manipulated to see to what extent noise changes

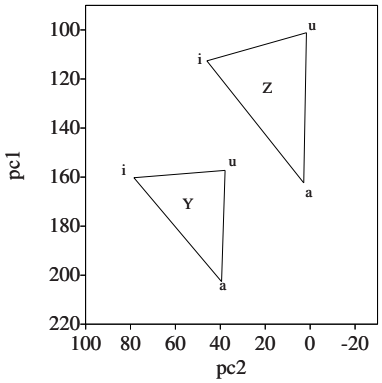


Figure 4.8: */a/-/i/-/u/ triangles in pc1 and pc2 of two speakers (Y and Z). Y's speech was recorded during a broadcast in rather bad quality, and Z's speech was recorded during an interview in rather good quality.*

the positioning and dispersion of their /a/, /i/, and /u/ in the first pc-dimensions. This was tested by gradually deteriorating the quality of their speech: All of the speaker's original filter-output that was below 20 (30/40/60) dB was set to 20 (30/40/60) dB, while the remaining filter values were kept as they were. By this, the minimum filter value was increased step-wise from 20dB to 60dB, so that the dB-range in the filters decreased step-wise. As can be seen in the plots of figure 4.9, on page 61, the worse the noise-level, the more the vowel positions shifted, and the more the distances between /a/, /i/, and /u/ decreased in the pc1pc2-plane.

The higher the minimum dB in the filters was set, the more the vowels shifted to the left and downwards in the pc1-pc2 plane. The increase of the minimum dB in the filters resulted ultimately in a mere point in the plot for more than 60 dB noise, which is an extreme condition. The bottom plots in figure 4.9, p. 61 show that whereas the vowels of speaker Y are not (yet) affected by a minimum noise level of 40 dB, most probably because the noise in those spectra was already 40dB, the vowels of speaker Z shift remarkably when the dB-level is set to 40dB.

The area of the back vowel /u/ is the first to change its position by added noise. Generally, /u/-like vowels are characterized by a spectrum with energy in the lower part of the spectrum and the absence of energy in higher parts (compare the /u/-spectrum in fig. 2.1, page 21). For a better view of the implication of spectral noise on the vowels in the pc-dimensions, the pc1- and pc2-values were recalculated to their according bandfilter output. The recalculated spectra were built on the eigenvectors of pc1 and pc2 alone while the other 15 pc-dimensions were set to 0 during the recalculation.

Figure 4.10, p. 62 shows the recalculation of spectra of corner points in the pc1-pc2 vowel space (points A, B, D, E), and of the centre (C). The figure reveals that two of the recalculated spectra (A and B), coincide with typical (common) spectral compositions: When the pc1- and pc2-eigenvectors are recalculated, the pc1 and pc2 value at the position of 'A' shows an /i/-like spectrum. 'B' displays an /u/-like vowel spectrum, as mentioned,

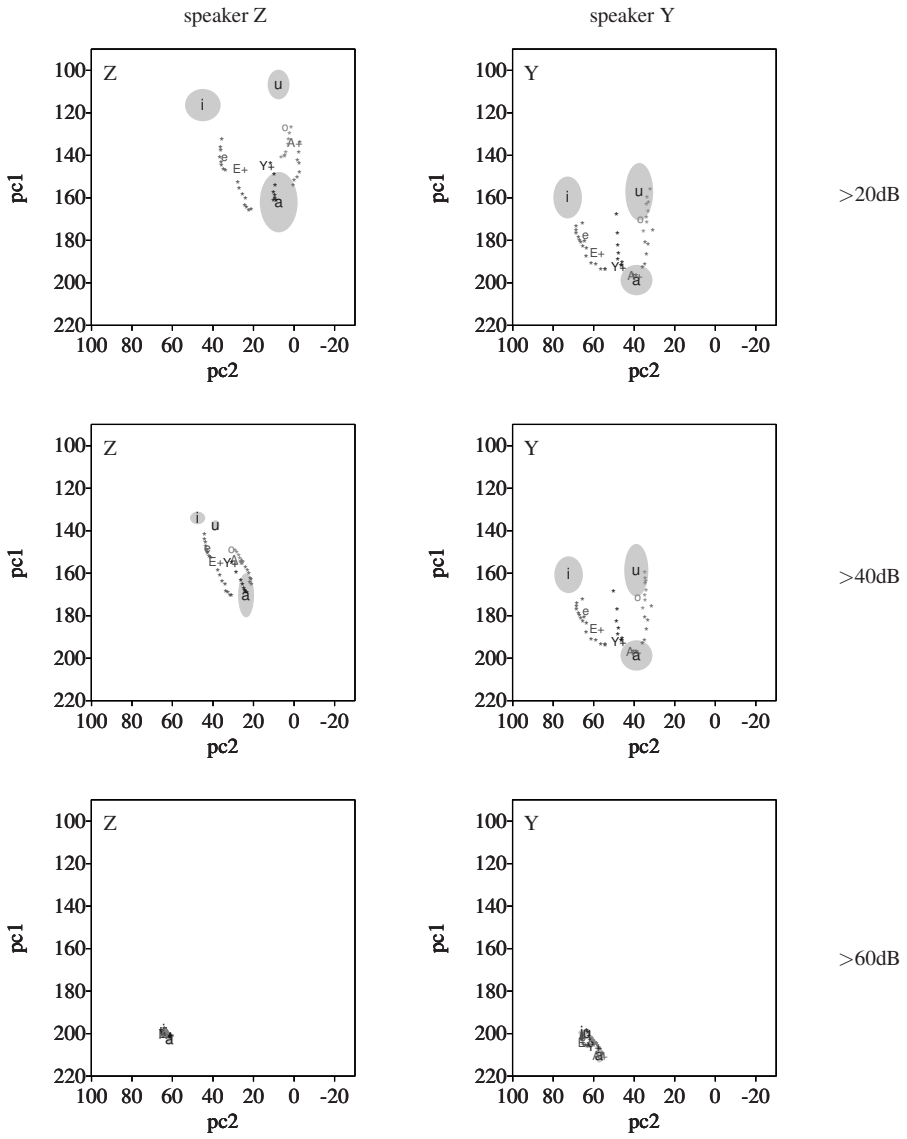


Figure 4.9: *The influence of the dB minimum-level (and thus of the noise level) in the filters on the vowel space size of male speaker Z (good recording quality) and male speaker Y (radio recording, often music in the background): from top to bottom, the dB level for the minimum filter value was increased from 20, to 40, to 60.*

characterized by the absence of energy in higher parts. As can be seen, by moving through the pc1-pc2 dimension from the right upper corner ‘B’ via ‘C’ to the left lower corner ‘D’, the minimum dB-level in the spectra increases. It is thus no coincidence that the B-D-line points out the direction in which the /a/-/i/-/u/ vowel triangles of recordings with bad noise ratios are dispositioned compared to recordings of good quality (figures 4.7, p. 59, and 4.9,

p. 61).

The recalculated spectra of corner points in figure 4.10 below, and the pc1 and pc2 values for speaker Y in figure 4.8, page 60 suggest that for speaker Y's vowel spectra, the intensities in the original (unmanipulated) filter output never decreased by 40 dB, whereas speaker Z's original filter output must have been below 40 dB in some filters. The higher minimum intensity in speaker Y's original filter output, and the late disposition (in terms of increasing noise levels) of speaker Y's vowels, indicates a poorer signal-to-noise ratio, compared to the earlier affected vowels of speaker Z. Different locations of the speakers' vowel spaces in the pc1-pc2 plane can at least partially be led back to differences in noise ratios.

We rechecked the results indicated by the recalculated spectra: Normalizing each

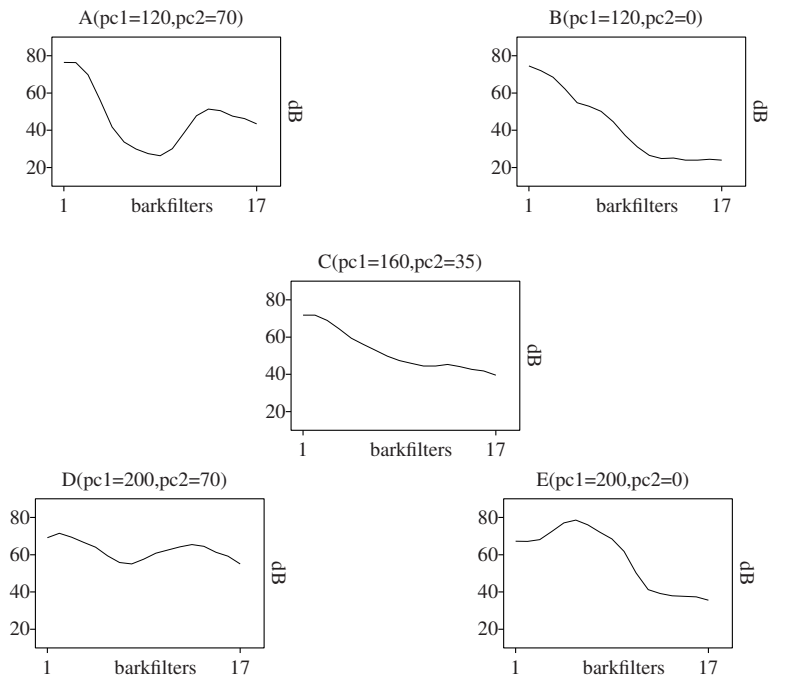
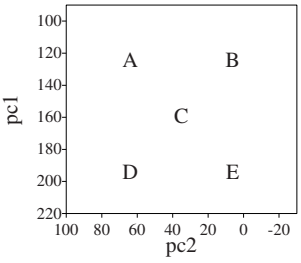


Figure 4.10: *Recalculation of the original spectra of the pc1-pc2 dimensions. Recalculated spectra (above) from the corners and from the center of the pc1-pc2 plane, indicated by A, B, C, D, E (as positioned in the pc1-pc2 plane to the right). For all pc3 to pc19 values, the 44 speakers' /a/-/i/-/u/ focal point pc3 to pc19 values were taken.*



speaker's barkfilter output according to his/her individual /a/, /i/, /u/ mean filter output before running a PCA on all 210 /a/, /i/, /u/ filter outputs resulted in almost the same first two pc-dimensions.

In the next section, we propose another normalization procedure that accounts for the variable recording qualities that affected the vowels' position and the dispersion in the pc-dimensions.

4.3.2 Normalization

The speaker-specific spectral vowel compositions, as discussed in the previous section, show that an interpretation of vowel quality must be based on relative rather than absolute pc values. Only in relation to each speaker's individual vowel space size and other (anchor) vowels do the speaker-specific spectra of the phoneme classes make sense.

Increasing the minimum dB in the filters resulted in decreasing sizes and shifting positions of the vowel space. Considering the various triangle positions, we could normalize the spectral data by setting the speakers' /a/-/i/-/u/ focal points to 0, and therewith eliminated part of the signal-to-noise ratio differences. However, this does not normalize for differences in vowel space size.

Though for the pc-dimensions there were no significant sex-differences in the vowel space sizes, we wanted to check remaining differences in the vowel space sizes. To see to what extent the remaining variance mirrors non-linguistic attributes such as speaker anatomy in terms of speaker sex, we ran analyses on the pc's of all speakers' anchor vowels /a/, /i/, /u/. A multivariate analysis of variance with the three vowels /a/, /i/, /u/ and their pc1 and pc2 values as dependent variables, and with 'sex' (female or male) as fixed factors, showed a significant effect for 'vowel' (Pillai's Trace, $F_{2,67}$=550.798, p<.001), for the interaction of 'vowel' with 'sex' ($F_{2,67}$=12.402, p<.001), and for 'vowel' with 'pc dimension' ($F_{2,67}$=2588.967, p<.001).

Which vowels were affected in what dimension was investigated by univariate analyses of variance on each anchor vowel phoneme separately. To find effects of recording quality, we also added the four-level fixed factor 'recording situation' as defined in section 4.2.3, p. 52. With 'sex' and 'recording situation' as fixed factors, there was a main effect of the recording situation on pc1, but no main effect of 'sex': Considering pc1 with 'sex' and 'recording situation' as fixed factors, for the vowel /u/, the recording situation (/u/ $F_{3,63}$=12.32, p<.001), and the interaction of 'sex' with 'recording situation' (/u/ $F_{2,63}$=5.18, p=.008) was highly significant. An analysis of variance on /a/ with 'sex' and 'recording situation' as fixed factors was only significant for the recording situation (/a/ $F_{3,63}$=11.26, p<.001). An analysis of variance on /i/ with 'sex' and 'recording situation' as fixed factors was just significant for the recording situation (/i/ $F_{3,63}$=2.81, p=.046).

So, all vowels were affected by the main effect 'recording situations', and for /u/ there

was an interaction effect; the effect of 'recording situation' was sex-dependent. This coincides with the findings on the influence of noise as described in the previous section. As mentioned there, /u/ was the most vulnerable vowel, and the worst recordings were those of two males who were recorded during broadcasts. A closer look at the distribution of the recording situations within the 70 speakers revealed that the recordings of the males included not only the worst recording qualities (broadcasts) but also fewer interview recordings (good quality) compared to the females' recordings. In conclusion, and of high importance for a variation analysis, the pc1 values of the anchor vowels do not carry significant main effects of 'sex'.

The same analyses of variance on pc2, again with 'sex' and 'recording situation' as fixed factors, showed that /a/ was significantly affected by 'sex' (/a/ $F_{1,63}$=17.10, p<.001), and by 'recording situation' (/a/ $F_{3,63}$=7.93, p<.001). The vowels /i/ and /u/ were not significantly influenced by 'sex' or by 'recording situation', nor was there an effect of interaction.

Thus, female and male data can be pooled for pc1. Pc2 shows an influence of speaker sex on the values of /a/. Pc1 explains most of the variance within the data, and it is the indicator for the phenomenon of 'lowering' (high correlation with $F1_{Bark}$). Therefore, we will concentrate on pc1 as the most important dimension from here on. To account for the various vowel space sizes and the remaining influences of recording quality and speaker sex, we will use the anchor vowels to relate the speaker's individual spectral phoneme realization to. We decided to put the long vowel and diphthong positions in the pc1-pc2 plane in relation to each other by measuring the relative distances within each speaker's set of anchor vowels. As a result the speaker normalization introduced earlier by equalizing all speakers' pc focal points to 0 will become superfluous. Nonetheless, we will continue to use it to make the vowel plots more comparable. With /u/ being more vulnerable than the other point vowels, we took each speaker's individual /a/ and /i/ values, and put all other measured values in relation to their positioning and distance. The following paragraphs describe how the onset and offset of each realized long vowel or diphthong was related to the speaker-specific /a/-/i/ distance or position.

Relative Onset

For the between-speaker comparison, the onset of a speaker's long vowel or diphthong was defined relative to his or her /a/ and /i/ values (compare van Heuven et al., 2002, 2003 [156]). First, the onset pc-values of a speaker's long vowel or diphthong were subtracted from his or her mean /a/ pc-values. The resulting pc-distance was then divided by the speaker's distance from /a/ to /i/. The relative onset as a percentage was thus calculated as follows:

$$rel.\ onset = \frac{pc_a - pc_{onset}}{pc_a - pc_i} * 100$$

Figure 4.11 below gives an example of how the relative onset was calculated, in this case for the onset of the long vowel /e:/. The vowel's movement in the pc1-pc2 plane is represented by one little star every 10 ms. For the relative value in pc1, the distance between the onset of /e:/ and the position of /a/ ($pc1_{/e:/}$ in the figure), is divided by the distance between /a/ and /i/ in pc1 ($pc1_{/a-i/}$ in the figure):
rel. onset /e:/ = $\frac{pc1_{e:}}{pc1_{a-i}} * 100 = \frac{19}{57} * 100 = 33$. The relative onset of /e:/ in pc1 is thus one third of the /a-i/ distance away from /a/, and two thirds away from /i/. For the relative value within the pc2-dimension, the calculation is equivalent. The distance between the onset of /e:/ and the position of /a/ ($pc2_{/e:/}$ in figure 4.11) is divided by the pc2-distance between /a/ and /i/ ($pc2_{/a-i/}$ in the same figure):
rel. onset /e:/ = $\frac{pc2_{e:}}{pc2_{a-i}} * 100 = \frac{20}{26} * 100 = 77$. The relative onset of /e:/ in pc2 is thus more than three quarters of the /a-i/ distance away from /a/. The speaker-specific vowel onset of /e:/ in the example of figure 4.11 would thus be represented by a pc1 value of 33 and a pc2 value of 77. By this normalization, every speaker's mean pc1 and pc2 value for /a/ will be represented by the value 0, and every speaker's mean pc1 and pc2 value for /i/ will be represented by 100, and so the speakers' vowels can be compared. Negative values are also possible, in case the vowel onset starts below /a/.

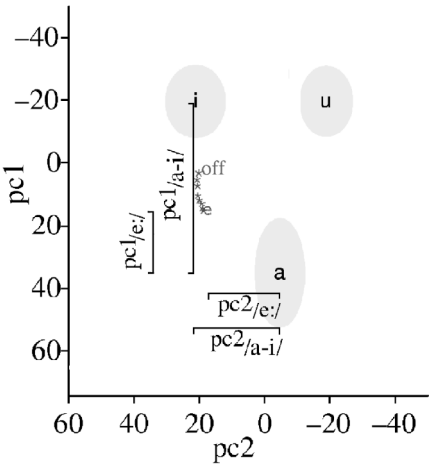


Figure 4.11: *Calculating the relative onset of /e:/ in pc1 and pc2: Every 10 ms, the movement of /e:/ in the pc1-pc2 plane is displayed by a star; the end of the movement marked by "off". The relative onset of /e:/ is calculated by* $\frac{pc1_{e:}}{pc1_{a-i}} * 100$. *By this equation, /a/ is always represented by an onset value of 0 in pc1 and pc2, whereas /i/ is represented by a onset value of 100. The relative pc1 onset of /e:/ in this example corresponds to* $\frac{19}{57} * 100 = 33$. *Related to the /a/-/i/ distance, the onset of /e:/ lies thus one third away from /a/ and two thirds away from /i/.*

Relative Degree of Diphthongization

For the calculation of the relative degree of diphthongization, one further value is added in each dimension: the distance from /a/ to the offset of the measured long vowel or diphthong. To calculate the degree of diphthongization of a speaker's vowel phoneme, the offset value is subtracted from the onset value. The resulting distance is then divided by the speaker-specific /a/-/i/ distance. The relative degree of diphthongization as a percentage along pc1 or pc2 was thus calculated as follows:

$$\textit{rel. degree of diphthongization} = \frac{pc_{onset} - pc_{offset}}{pc_a - pc_i} * 100$$

In figure 4.11, p. 65, the movement of /eː/ starts at a pc1 value of 19 ($pc1_{/e:/}$, see previous paragraph) and ends at a pc1 value of 3. The distance in pc1 between on- and offset of /eː/ is thus 16. /a/ is positioned at a pc1 value of 37 and /i/ at -20. The distance between /a/ and /i/ ($pc1_{a-i}$) is 57. The relative degree of diphthongization for /eː/ in figure 4.11 would be
rel. degree of diphthongization /eː/ = $\frac{pc1_{e:} - pc1_{offset}}{pc1_{a-i}} * 100 = \frac{19-3}{57} * 100 = 28$. The diphthongization of /eː/ in pc1 corresponds thus to 28% of the /a/-/i/ distance in pc1.

This normalization procedure was applied to each vowel and each pc-dimension separately. We also could have taken the Euclidean distance in an n-dimensional space, by for example representing the distance within the 2-dimensional pc1-pc2 plane by a single number (distance). As a result, though, the distance between two points *a* and *b* would be calculated by D(*a*,*b*)=$\sqrt{(pc1a - pc1b)^2 + (pc2a - pc2b)^2}$, and information would be lost considering the orientation of the measured points in the space: pc1 and pc2 values of *a* and *b* could be switched without changing the resulting distance D, and various *a* and *b* values would be represented by the same D value.

In our case, this would include the interchangeability in the front/back or high/low orientation of the vowels in relation to the /a/-/i/ line. As a result, the main focus of our investigation, 'lowering' (the more open pronunciation that matches the pc1-dimension) would get intertwined with the front-back dimension that matches pc2. Furthermore, pc1 is the value that explains most of the variation in the vowel space, and we have seen earlier in this section that for /a/, pc2 shows not only effects of the signal-to-noise ratio (recording situation), but also a main effect for speaker sex. To avoid unwanted effects of speaker sex on the results, and to keep the interpretability of the calculated distances, we decided to measure the distances separately for each dimension, and not to represent distances in pc1 and pc2 by a single value.

In the following sections, the long vowels' and diphthongs' onsets and diphthongizations will be represented by their relative onset and the relative degree of diphthongization.

4.4 Results

In this section, the results of the acoustic analysis of the 70 speakers' vowel realizations are presented. Next to the more general acoustic results including all speakers, the results of various speaker subgroups that were formed according to the speaker's sex, age and education, were compared.

In the corpus of 70 speakers, not only the quality, but also the quantity of speech data per person differed, such as the frequency of occurrence for words and vowels. In general, realizations of the vowel phonemes /a/ were most frequent, followed by /eː/, /ɛi/, /i/, and

/oː/ (table 4.6 below). Realizations of /u/, /ɑu/, /œy/ (and /øː/, the latter will be ignored) were less frequent. For each speaker and phoneme, at least four vowel realizations were analyzed. Altogether, the onsets and offsets of 12482 long vowels and diphthongs and the mid spectra of 11381 anchor vowels (/a/, /i/, /u/) were measured for analysis.

The mean duration of all vowel phonemes was generally longer for vowels uttered by females than for those uttered by males, as can be seen in table 4.6. This observation has also been made by Koopmans-van Beinum for Dutch vowels in 1980 [77]. Diverse biophysical factors, such as the greater pitch range of females, articulatory-dynamic properties, as well as gender differences are seen as the cause of this rather stable sex pattern. Simpson (2001 [132]) argues, though, that a socio-phonetic explanation is rather implausible given that the phenomenon is found cross-linguistically. Investigations on the cause of this phenomenon are beyond the extent of the present research, and we considered only possible interactions with other acoustically measured attributes.

Table 4.6: *Mean number of vowels and mean duration (broken down by sex) for the 70 speakers.*

mean	/ɛi/	/ɑu/	/œy/	/oː/	/eː/	(/øː/)	/a/	/i/	/u/
number of vowels	44	16	12	43	59	(4)	93	43	26
duration (ms) females	112	119	118	110	102	(90)	104	87	75
duration (ms) males	106	113	111	105	96	(79)	102	83	70

Interestingly, the durational order of the vowel phonemes - from longest to shortest vowel phoneme - has changed compared to previous literature. In our data, for the long vowels and diphthongs averaged over all 70 speakers (table 4.6), /ɑu/ showed the longest duration, followed by /œy/, /ɛi/, /oː/, /eː/, and /øː/ as the shortest vowel. In Nooteboom & Slis (1972 [108]) analysis of three-syllabic /pVpVpVp/ nonsense words (with V for the same vowel), with the second syllable stressed, /œy/ was the longest, followed by /ɛi/, /øː/, /oː/, /ɑu/, and the shortest vowel /eː/. Twenty years later, /eː/ and /oː/ have switched their durational positions: In Strik et al.'s analysis of isolated Dutch words from 1992 [139], uttered by one untrained male speaker, /œy/ was the longest, followed by /ɛi/, /øː/, /eː/, /ɑu/, and the shortest vowel /oː/.

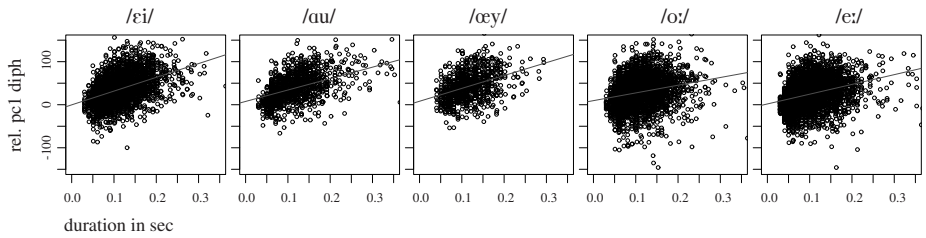


Figure 4.12: *Relative degree of diphthongization in pc1 (y-axis) and durations (x-axis) of the vowel phonemes /ɛi/, /ɑu/, /œy/, /oː/, and /eː/ with regression lines.*

These changes in the durational relations between the vowel phonemes might go together with changes in other attributes of vowel pronunciation. As described in section 2.2, p. 19, differences in the duration of diphthongs correlate with the diphthong offset values. Figure 4.12, p. 67 shows that the degree of diphthongization increases slightly with increasing vowel length. A steady correlation with the onset values was not found. However, there was a rather stable relation between the onset and the distance between onset and offset, displayed in figure 4.13. Generally, and for all five vowel phonemes, the onset values and the degrees of diphthongization show a reciprocal pattern: the lower the onset, the stronger the diphthongization, and vice versa. The following sections will display to what extent onsets and degrees of diphthongization are reciprocal within specific speaker subgroups.

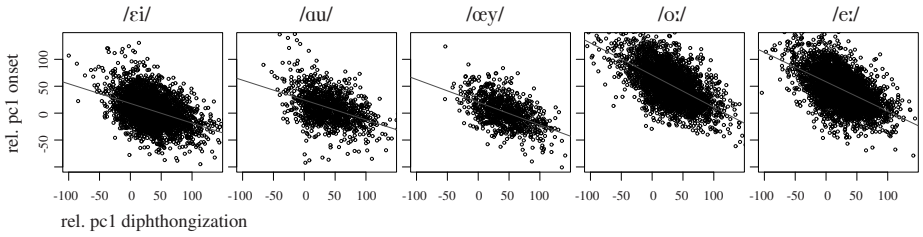


Figure 4.13: *Relative onsets (y-axis) and degrees of diphthongization (x-axis) in pc1 with regression lines for the vowel phonemes /ɛi/, /ɑu/, /œy/, /o:/, and /e:/.*

Each speaker's mean onset and mean degree of diphthongization in pc1 was taken for the inter-speaker comparison of the vowel phonemes /o:/, /e:/, /ɛi/, /œy/, /ɑu/. Table 4.7 on page 69 summarizes the average results per vowel phoneme and for some of the speaker subgroups (high vs. low socio-economic status; males vs. females). The measured values of their relative onsets and relative degrees of diphthongization, and the differences between the various subgroups will be analyzed in more detail in the following sections.

In section 4.1 we predicted that the vowel values of our 70 speakers are most likely to be classified by the socio-economic status (here reduced to level of education), sex, and age. To investigate the effects in our data, a multivariate analysis of variance was run with the five vowels' relative pc1 onsets as dependent variables, with sex (f/m) and level of education (high/low) as independent variables, and with the speaker's age (continuous scale) as a covariate. There was a significant effect for the within-subject factor 'vowel' ($F_{4,59}$=23.640, p<.001), and for the interaction of 'vowel' and 'age' ($F_{4,59}$=3.371, p=.015). Pairwise comparisons of the vowel phonemes' onsets showed they all differed significantly, except for /œy/ and /ɛi/. There was no main or interaction effect of sex (we will see this again in section 4.4.1). A significant between-subject effect was found for the level of education ($F_{1,62}$=6.085, p=.016). For the onsets, the mean difference between high and low level of education was significant at the .05 level (13.68, p<.001, the confidence inter-

Table 4.7: *Means and standard deviations of the relative onsets (on = $\frac{pc1_a - pc1_{onset}}{pc1_a - pc1_i} * 100$) and the relative degrees of diphongization (diph = $\frac{pc1_{onset} - pc1_{offset}}{pc1_a - pc1_i} * 100$) in pc1. The results for the 5 vowel phonemes /oː/, /eː/, /ɛi/, /ɑu/, and /œy/ are broken down by sex ('f' for females and 'm' for males) and level of education ('h' for high educated and 'l' for low educated).*

		rel.pc1	/oː/		/eː/		/ɛi/		/ɑu/		/œy/	
sex	eduL		on	diph	on	diph	on	diph	on	diph	on	diph
f	h	N	18	18	18	18	18	18	18	18	18	18
		Mean	41.5	36.2	38.3	30.6	5.4	40.2	3.9	44.8	-4.0	55.1
		SD	23.0	16.1	17.1	15.6	11.2	14.1	11.2	20.6	14.4	17.8
	l	N	17	17	17	17	17	17	17	17	17	17
		Mean	64.1	14.0	59.3	12.4	10.1	26.3	14.2	31.9	10.9	31.1
		SD	18.9	12.8	11.4	9.8	11.8	10.5	13.1	13.7	13.6	12.2
	Total	N	35	35	35	35	35	35	35	35	35	35
		Mean	52.5	25.4	48.5	21.8	7.7	33.5	8.9	38.6	3.2	43.4
		SD	23.7	18.3	17.9	15.8	11.6	14.2	13.1	18.5	15.8	19.4
m	h	N	17	17	17	17	17	17	17	17	17	17
		Mean	47.5	37.1	34.7	31.1	0.8	36.9	6.3	40.0	-4.9	46.2
		SD	12.8	12.4	11.4	10.0	9.7	11.3	8.8	12.0	7.8	11.2
	l	N	18	18	18	18	18	18	18	18	18	18
		Mean	73.5	11.5	50.9	15.1	5.2	25.0	19.4	35.6	8.0	25.7
		SD	20.1	14.3	20.2	11.2	17.3	11.5	20.3	16.4	22.7	19.0
	Total	N	35	35	35	35	35	35	35	35	35	35
		Mean	60.9	23.9	43.0	22.8	3.0	30.8	13.0	37.8	1.8	35.7
		SD	21.3	18.5	18.2	13.2	14.1	12.7	16.9	14.4	18.1	18.6
Total	N		70	70	70	70	70	70	70	70	70	70
	Mean		56.7	24.7	45.8	22.3	5.4	32.1	11.0	38.2	2.5	39.6
	SD		22.8	18.3	18.1	14.5	13.0	13.4	15.1	16.5	16.9	19.3

val (C.I.) ranges from 7.308 to 20.049). We will return to this in section 4.4.2.

The same multivariate analysis of variance was run with the five vowels' relative degrees of diphthongization in pc1 as dependent variables. With sex (f/m) and level of education (high/low) as independent variables, and with speaker age as a covariate, there was no significant within-subject-effect of 'vowel'. Pairwise comparison revealed that the mean differences between the vowel phonemes in terms of diphthongizations were significant at the .05 level, except for the mean difference between /eː/ and /o/, and the difference between /ɑu/ and /œy/. There was a significant interaction effect for the within-subject factor 'vowel' with 'age' ($F_{4,59}$=3.207, p=.019). There was no main or interaction effect of sex (see section 4.4.1). A significant between-subject effect was found for the level of education ($F_{1,62}$=16.208, p<.001). We will return to this in section 4.4.2. For the degrees of diphthongization, the mean difference between high and low level of education was significant at the .05 level (16.8, p<.001, the C.I. ranges from 11.615 to 21.987). The interaction of the level of education and age had a significant effect as well ($F_{1,62}$=4.422, p=.040).

In the study of van Heuven et al. (2002 [156]), a lowering of /ɛi/ was found for young female speakers of the 'avant-garde', and the lowering of the diphthong was reported to have spread since then. These female avant-garde speakers are presumably part of our speaker group that was assigned to the background category 'high educated/occupied'. If the statement of the mentioned study is true for our data as well, we should see within the high educated group of females that the oldest speakers do not lower. Furthermore, we expect increasingly lowered /ɛi/'s with decreasing female speaker age, and we also expect all of the younger more highly educated speakers to lower /ɛi/. For the males, we expect a lowered /ɛi/ at a later stage. The expectations for the low educated speakers are less clear. Following the assumptions of the literature on /ɛi/, the low educated speakers should not show a lowering of /ɛi/ (see Stroop, 1998 [140], Edelman, 1999 [33], van Heuven et al., 2002 [156]). However, previous studies did not include analyses of non-avant-garde, or speakers of lower socio-economic status, and so, these expectations are based merely on reported subjective perceptions. With respect to the other vowel phonemes /oː/, /eː/, /ɑu/, and /œy/, we expect similar patterns as those found for /ɛi/, or, if the vowel phonemes vary rather due to a chain shift, we should find changing phoneme patterns shifted in time according to the speaker ages.

To test the claims of the earlier studies and to investigate in more detail the variation in the onset values or degrees of diphthongization, and speaker attributes such as sex, age, and level of education, the results of analyses of variance (followed by post-hoc tests, with Bonferroni correction when the equality of variance assumption holds, and Dunnett t3 otherwise) are reported separately for each vowel phoneme. All analyses were calculated with the statistical software SPSS [134] and R [127].

4.4.1 Males versus Females

The females' pronunciation was thus said to differ from the males' pronunciation pattern [140, 33, 156]. In our data, the multivariate analysis (p. 68) revealed no main or interaction effect of sex. Analyses of variance on each vowel phoneme with either the relative pc1 onsets, or the degrees of diphthongization as dependent variable, and level of education and speaker sex as independent variables, and age as covariate yielded no significant main or interaction effect of sex. For all vowel phonemes, the mean onsets of the females did not differ significantly from the onsets of the male speakers (females-males: for /oː/ p=.092, mean difference -8.191, the 95% C.I. ranges from -17.768 to 1.387; for /eː/ p=.141, mean diff 5.856, C.I. -1.987 to 13.699; for /ɛi/ p=.133, mean diff 4.789, C.I. -1.493 to 11.072; for /ɑu/ p=.552, mean diff -2.042, C.I. -8.863 to 4.780; for /œy/ p=.337, mean diff 3.699, C.I. -3.950 to 11.348).

As can be seen in figure 4.14, p. 71 the range of the relative onset values of /oː/ and /eː/ was noticeably larger than the range of the diphthongs' onsets. The large range, and

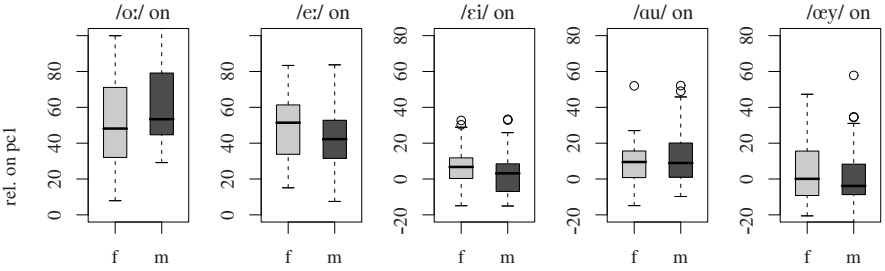


Figure 4.14: *Box plots of the 35 females' (f) and 35 males' (m) relative pc1 onset values (y-axis) for the five vowel phonemes.*

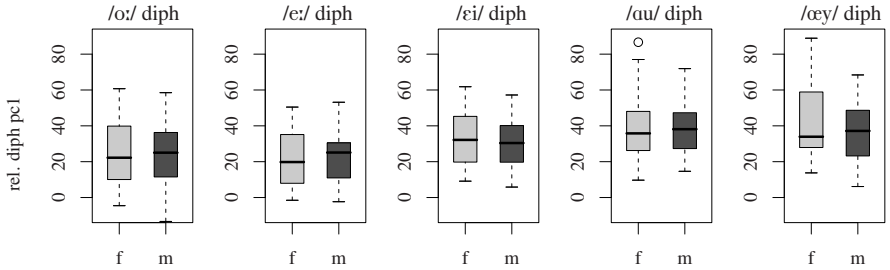


Figure 4.15: *Box plots of the 35 females' (f) and 35 males' (m) relative degrees of diphthongization in pc1 (y-axis) for the five vowel phonemes.*

thereby the large variation within these two vowel phonemes makes the existence of distinguishable vowel variants within these phoneme classes more probable than within the less varying diphthong phonemes. We would thus now predict that /oː/ and /eː/ are vowel phonemes whose onsets are pronounced differently, independently of the speaker's sex, and thus dependent on yet to be defined speaker group attributes. In terms of diphthongization (fig. 4.15), the range of variation was rather large for /oː/, /ɑu/ and /œy/.

The same analyses of variance on the relative degrees of diphthongization showed also no significant effect of sex for any of the five vowel phonemes (females-males: for /oː/ p=.743, mean difference 1.113, the 95% C.I. ranges from -5.652 to 7.878; for /eː/ p=.737, mean diff -0.995, C.I. -6.898 to 4.907; for /ɛi/ p=.576, mean diff 1.633, C.I. -4.177 to 7.444; for /ɑu/ p=.638, mean diff -1.818, C.I. -9.513 to 5.878; for /œy/ p=.158, mean diff 5.454, C.I. -2.177 to 13.085).

Against our expectations, the male speakers' and the female speakers' pronunciations in terms of onsets and diphthongization did not really differ, neither in variance nor in mean. Nonetheless, in general, the large range of the onsets and the diphthongization values for some of the phoneme pronunciations suggest the existence of variants for both groups.

4.4.2 Higher versus Lower Socio-Economic Status

As described in the previous section, there was no significant main effect of speaker sex on the relative onsets and the relative degrees of diphthongization. However, at least for the onsets of /eː/ and /oː/, pc1 showed a rather large range in the measured values both within the females and the males. The range in the degrees of diphthongization was considerable as well, suggesting variation due to other factors than sex. Following previous reports considering the realization of /ɛi/ we expected the avant-garde, arguably included in the high educated speaker group, to show lower onsets than the other speakers (here, the low educated speakers). Contrary to the insignificant effect of sex, for the onsets and the degrees of diphthongization in our data, the multivariate analysis (p. 68) yielded a significant main effect for the level of education.

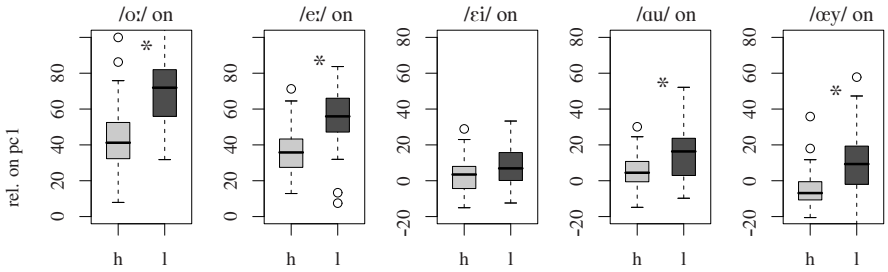


Figure 4.16: *Boxplots of the 35 high (h) and 35 low (l) educated speakers' relative pc1 onset values (y-axis) for the five vowel phonemes. Both females and males. Stars indicate significance.*

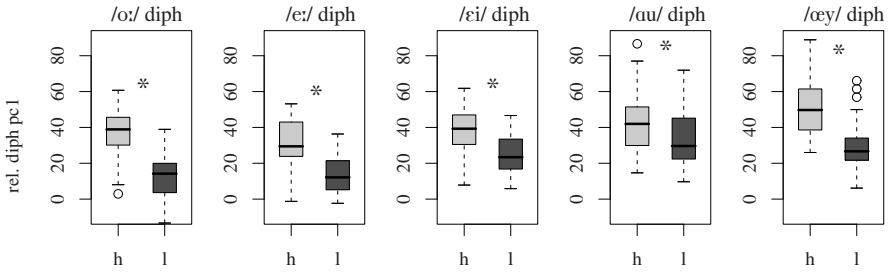


Figure 4.17: *Boxplots of the 35 high (h) and 35 low (l) educated speakers' relative degrees of diphthongization in pc1 (y-axis) for the five vowel phonemes. Stars indicate significance.*

As can be seen in figure 4.16 and figure 4.17, there was a considerable difference between the two educational groups for almost each vowel phoneme. Generally, the relative onsets in pc1 (fig. 4.16) of the high educated speakers were lower than the low educated speakers' onsets. Analyses of variance on each vowel phoneme (again with age as covariate and with sex and the level of education as factors) revealed that, except for /ɛi/, these differences between the onset values of the two educational groups were highly significant at the .05 level (high-low: for /oː/ $p<.001$, with a mean difference of -23.420,

the C.I. ranges from -32.997 to -13.843; for /eː/ p<.001, mean diff -17.919, C.I. -25.762 to -10.075; for /ɑu/ p=.003, mean diff -10.751, C.I. -17.572 to -3.929; for /œy/ p=.001, mean diff -13.199, C.I. -20.848 to -5.550; but for /ɛi/ p=.327, mean diff -3.106, C.I. -9.388 to 3.176).

As can be seen in figure 4.17, p. 72, the high educated speakers diphthongized the vowels to a stronger extent than the low educated speakers. These differences between the higher versus low educated speakers appeared to be (highly) significant for all vowel phonemes (high-low: for /oː/ p<.001, with a mean difference of 23.704, the 95% C.I. ranges from 16.939 to 30.470; for /eː/ p<.001, mean diff 16.782, C.I. 10.879 to 22.685; for /ɛi/ p<.001, mean diff 13.209, C.I. 7.399 to 19.020; for /ɑu/ p=.023, mean diff 8.998, C.I. 1.303 to 16.694; for /œy/ p<.001, mean diff 21.311, C.I. 13.681 to 28.942).

The low educated speakers showed considerably less diphthongization for the long vowels /oː/ and /eː/ than for the diphthongs /ɛi/, /ɑu/, and /œy/ (see figure 4.17, p. 72, compare table 4.7, p. 69). This coincides with the traditional contrast of long vowels and diphthongs. This traditional contrast is less obvious for the degree diphthongization within the more highly educated: Here, the degree of diphthongization of /oː/, /ɛi/, and /ɑu/ is alike (see figure 4.17, p. 72).

In conclusion, there was a main effect of 'level of education', and the expected differences considering the realization of /ɛi/ of the avant-garde (presumably included in the high educated speaker group) versus the other speakers (here, the low educated speakers) was confirmed in terms of diphthongization. However, the onsets of /ɛi/, contrary to the expectations, did not differ significantly between the levels of education. In general, the behavior of the vowel phonemes in terms of onset and diphthongization was comparable (except for /ɛi/), and suggests so far an overall pronunciation pattern.

4.4.3 Testing Interactions of 'Level of Education' and 'Sex'

The previous sections showed that there were no significant differences of onsets and degrees of diphthongization between the females and males (see figures 4.14/4.15, p. 71), whereas the level of education affected the pronunciation patterns significantly (see figures 4.16/4.17, p. 72), and independently of the vowel phoneme (the onset of /ɛi/ being an exception).

Following previous studies, we expected the /ɛi/ of high educated females to differ from high educated males' pronunciations. However, ANOVA on the speakers' onset values, or the speakers' degrees of diphthongization, with the speakers' level of education and the speakers' sex as fixed factors yielded no significant interaction of 'sex' and 'level of education' for any of the vowel phonemes. Against the expectations from the literature on the pronunciation of /ɛi/, the pronunciation of all high educated females did not differ significantly from the pronunciation of all high educated males; neither did low educated

females differ significantly in their pronunciation from low educated males.

4.4.4 The Effect of Speaker Age

Next to the effects of level of education and speaker sex, which we described in the previous sections, we expected an effect of speaker age on the measured vowel values. In the multivariate analysis (p. 68) there were no main effects of 'age', but significant interaction effects were found for 'age' and the vowel phonemes, and for 'age' and 'level of education'.

ANOVA on the onsets of each vowel phoneme separately, with 'sex' and 'level of education' as fixed factors, and with the speakers' age as covariate revealed that there was no main effect of 'age', and the interaction of 'age' and 'level of education' was not significant for the vowel phonemes either (for /oː/ $F_{1,62}$=.343, p=.560; for /eː/ $F_{1,62}$=.269, p=.606; for /ɑu/ $F_{1,62}$=2.187, p=.144; for /œy/ $F_{1,62}$=2.208, p=.142; for /ɛi/ $F_{1,62}$=1.059, p=.307).

Considering the degrees of diphthongization, there was no main effect of age, but age interacted with the level of education. This effect was significant for the vowel phonemes /ɛi/ and /ɑu/ (for /eː/ ($F_{1,62}$=.731, p=.396), for /oː/ ($F_{1,62}$=1.214, p=.275), for /ɛi/ ($F_{1,62}$=4.270, p=.043), for /œy/ ($F_{1,62}$=1.729, p=.193); for /ɑu/ ($F_{1,62}$=6.746, p=.012)). As an example, figure 4.18 below shows the relative degree of diphthongization for /ɛi/ in dependence of speaker age for both levels of education; figure 4.19, p. 75 shows the diphthongization for both sexes.

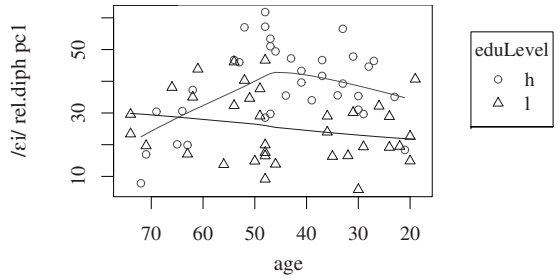


Figure 4.18: *Degrees of diphthongization of the vowel phoneme /ɛi/ in pc1 (y-axis) according to speaker age (x-axis). Circles for the high educated speakers, triangles for the low educated speakers, with lowess curves.*

In conclusion, for two vowel phonemes the pronunciation differences between the high and low educated speakers differed in dependence of the speakers' age. Yet, considering 'age', up to now, only linear dependencies were tested. To see whether the results would change (especially considering the previously insignificant interaction effect of 'age' on the onsets and the diphthongization of /œy/, /eː/, and /oː/) if 'age' is linear or rather a curvilinear variable, we replaced 'age' by the squared age 'age^2'. For /ɛi/ ($F_{1,62}$=6.231, p=.015)

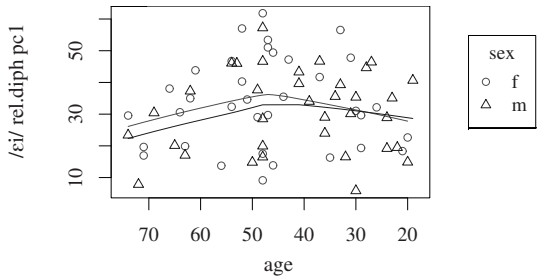


Figure 4.19: *Degrees of diphthongization of the vowel phoneme /ɛi/ in pc1 (y-axis) according to speaker age (x-axis). Circles for the female speakers, triangles for the male speakers, with lowess curve.*

and /ɑu/ ($F_{1,62}$=8.886, p=.004), the interaction effect was slightly higher significant, but the results did not change in general. A curvilinear factor (which would show exponential, parabolic, or logarithmic behavior) did not capture the changing effect of age any better.

Whether speaker age affects the vowel phoneme pronunciation of the low educated in the same way as the high educated, and whether this effect is linear or not, will be considered next. The literature on /ɛi/ (Stroop, 1998 [140], van Heuven et al., 2002 [156]) would suggest that the low educated do not increasingly lower their onsets, whereas the high educated do, with the females leading. For the low educated we would therefore expect no effects, whereas for the high educated, the reported increase of lowering /ɛi/ should result in an at least linear, or even exponential effect of 'age'. Given the previous general accordance of the five vowel phonemes' pronunciation patterns, we expect the other vowels to show the same pattern as /ɛi/.

Figure 4.18, page 74, displays the individual degrees of diphthongization for /ɛi/ according to the speaker's age and the level of education. As can be seen, the diphthongization pattern of the high educated speakers is changing, whereas the pattern of the low educated speakers is rather diffuse and almost horizontal, indicating its independence of the speakers' age. This indicates differing structures of change over time for the higher versus low educated speaker group:

'Age'/'Age^2' Effects within the Low Educated

Analyses of variance with 'sex', and either 'age^2', or 'age' as covariate on the low educated speakers' relative onsets and the degrees of diphthongization revealed significant effects only for the diphthongization of /ɑu/. For the low educated speakers' degrees of diphthongization, the effect of 'age' ($F_{1,35}$=8.09, p=.008) turned out to be slightly higher significant than the effect of 'age^2' ($F_{1,35}$=7.50, p=.010). Thus, both, 'age' and 'age^2', were significant, suggesting that the speakers' age is a linear factor when it comes to the diphthongization of /ɑu/. 'Age' was almost of significant influence in terms of the low

educated speakers' /ɑu/ onsets with $F_{1,35}$=3.87, p=.058 (for 'age^2' $F_{1,35}$=3.52, p=.070). A comparable (but also not significant) pattern was found for the low educated speakers' diphthongization of /œy/ (for 'age' $F_{1,35}$=4.07, p=.052, for 'age^2' $F_{1,35}$=3.72, p=.063). No significances for 'sex' as a main effect were found, and no interaction effects either. Against the expectations, the low educated speakers' pronunciation thus did show some (linearly progressing) pronunciation changes for /ɑu/ and /œy/ over time.

'Age'/'Age^2' Effects within the High Educated

For the onsets of each vowel phoneme, analyses of variance with 'sex', and either 'age', or 'age^2' yielded no significant differences between the sexes, and no interaction effects of 'sex' with 'age', or 'sex' with 'age^2'. Therefore, again, females did not differ significantly from the males in their pronunciation. Though there was no linear relation between the speakers' ages and their relative onsets, or their relative degrees of diphthongization, 'age^2' turned out to significantly affect the degrees of diphthongization of /eː/ ($F_{1,35}$=5.12, p=.031), /oː/ ($F_{1,35}$=5.37, p=.027), and /ɛi/ ($F_{1,35}$=5.83, p=.022), as well as the onsets of /ɛi/ ($F_{1,35}$=4.87, p=.034).

The age pattern of the high educated speakers was found to differ from the low educated speakers' pattern for each vowel phoneme. Within the high educated, other vowel phonemes were affected by the factor age than the ones that were affected within the lowered educated speakers. Whereas the low educated speakers showed a linear effect of speaker age for the vowel phoneme /ɑu/, and the same tendency for /œy/, the high educated speakers turned out to be affected by a non-linear effect of speaker age. This curvilinear effect was manifested in the degrees of diphthongization of /eː/, /oː/, and /ɛi/, as well as in the onsets of /ɛi/, and suggests that pronunciation habits changed in a non-gradual way. The significant results for the factor 'age^2' suggest for the high educated speakers that some pronunciation habits changed from generation to generation rather than gradually over the years, as was tested with the factor 'age'. Figure 4.18, p. 74 also suggests that in general, the expectations for the high educated speakers in terms of a changing pronunciation can be confirmed. However, as the next sections will show, there is no gradual lowering of /ɛi/ as expected by the results of previous /ɛi/-studies.

In the next section, the breakdown of speakers into age groups will give a better view on how the non-linear effect of speaker age manifests in the measured values. Given the speakers' age range of 57 years (adults aged 19 years up to 76 years of age), the speakers covered more than two generations. To discover these age effects more globally, next, we split the speakers into three age groups.

4.4.5 Differences between Age Groups

At the time of recording, the 70 speakers were between 19 and 76 years old. To calculate generation-dependent effects of speaker age, we decided to split the coviariate 'age' into three levels and therewith get a fixed factor 'age group'.

In terms of speaker age, the CGN had assigned six levels. Given the number of 70 speakers, their diverse group affiliations (see table 4.1, p. 51), and the purpose to distinguish generations, each two consecutive age levels were merged. This resulted in three generations: a 'young' group aged 18 years to 35 years, a 'mid' group ranging from 36 to 54 years of age, and an 'old' group of 55 years of age and older (fig. 4.20).

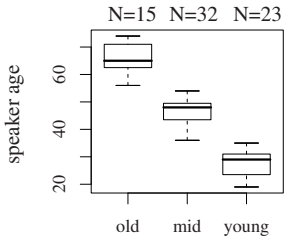


Figure 4.20: *Boxplots of speaker age split into the age groups 'old', 'mid', 'young'.*

Since these three age groups had not been considered in the initial design of our corpus, the number of speakers within each age group varied: 15 were assigned to the old generation, 32 to the mid generation, and 23 to the young generation. Following the results from the previous section, we expect some linear changes from the 'old' via 'mid' to the 'young' generation, as well as some changes from 'old' to 'mid', that are reversed again in the following generation step, thus from the 'mid' to the 'young' generation.

ANOVA on the relative onsets or degrees of dipthongization of all speakers, with 'level of education', 'age group', and 'sex' as fixed factors showed no significant differences between the generations ('age group') in terms of the relative vowel onsets. The degrees of diphthongization, however, showed significant differences between some of the age groups for the vowel realizations of /eː/, /oː/, and /ɛi/. Considering the degrees of diphthongization, the old generation differed significantly from the mid generation for /eː/ (mid-old: p=.005, mean difference 11.66, the 95% C.I. ranges from 2.870 to 20.453), and /oː/ (p=.020, mean difference 11.866, C.I. 1.446 to 22.286), and /ɛi/ (p=.004, mean difference 11.886, C.I. 3.145 to 20.627). For /ɑu/ and /œy/, there were no significant differences between generations. The boxplots in figure 4.21, p. 78 display the relative degrees of diphthongization of /ɛi/ for the three groups. As can be seen the diphthongization range of the mid and young group is considerable.

Except for /ɛi/, there was a significant interaction of 'level of education' (high or low) with 'age group' for all vowel phonemes' relative onsets. For the degrees of diphthongization, the interaction of 'level of education' with 'age group' (old, mid, young) was significant for all vowel phonemes. The subgroups' onset- and diphthongization values

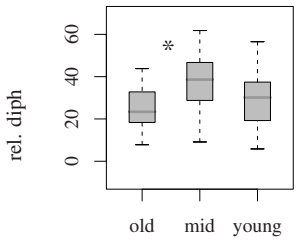


Figure 4.21: *Boxplots of the relative degrees of diphthongization (y-axis) of /ɛi/ for the age groups 'old', 'mid', 'young'.*

	old	mid	young
high	7	17	11
low	8	15	12
total	15	32	23

Table 4.8: *Number of speakers per level of education when split into the age groups 'old', 'mid', 'young'.*

are plotted per vowel in figure 4.24 on page 81. The number of speakers per subgroup is displayed in table 4.8.

Old Generation

ANOVA on the vowel onsets or degrees of diphthongization with 'level of education' and 'sex' as independent factors yielded no significant effects, neither in terms of the onsets, nor considering the degrees of diphthonization. Though some difference seems to be apparent for the diphthongization of /ɑu/, the vowel phoneme realization of the old generation did not differ significantly between the higher and low educated speakers (see the very left part of all plots of fig. 4.24, page 81), nor between the sexes.

Mid Generation

ANOVA on the vowel onsets or degrees of diphthongization with 'level of education' and 'sex' as independent factors yielded some significant effects: Considering the vowel onsets, the level of education was significant for /o:/ ($F_{1,32}$=23.53, p<.001), /e:/ ($F_{1,32}$=31.03, p<.001), /ɑu/ ($F_{1,32}$=9.96, p=.004), and /œy/ ($F_{1,32}$=5.07, p=.032), not for /ɛi/ ($F_{1,32}$=1.99, p=.168). Considering the degrees of diphthongizations, the level of education was significant for all vowel phonemes (for /o:/ ($F_{1,32}$=31.93, p<.001), for /e:/ ($F_{1,32}$=36.77, p<.001), for /ɛi/ ($F_{1,32}$=20.11, p<.001), for /ɑu/ ($F_{F1,32}$=7.72, p=.010), and for /œy/ ($F_{1,32}$=36.42, p<.001), fig. 4.22, p. 79).

High educated speakers of the mid generation lowered their vowel onsets more, and diphthongized the vowel phonemes to a stronger degree than low educated speakers of the same generation (see middle part of all plots of figure 4.24, page 81).

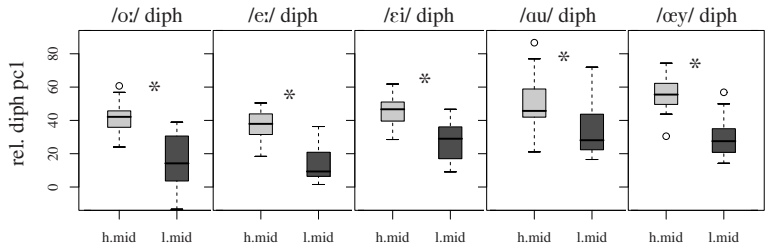


Figure 4.22: *Boxplots of the relative pc1 degrees of diphthongization (diph) of all speakers of the mid age group, split into high (h: light grey boxes) and low educated (l: dark grey boxes) speaker groups. Stars indicate significances between the levels of education within this generation.*

Young Generation

ANOVA's on the vowel's onsets or degrees of diphthongization with 'level of education' and 'sex' as independent factors yielded some effects for the young generation as well: As can be seen in figure 4.24, page 81 on the right side of all plots, the low educated show higher onsets than the high educated. However, the pronunciation differences between lower and high educated speakers in terms of the vowel onsets was not of significance, except for /oː/ ($F_{1,23}$=5.92, p=.025). Comparable to the mid generation, the high educated speakers of the young generation diphthongized the vowel phonemes to a stronger extent than the low educated speakers (fig. 4.23). The differences in diphthongization between the levels of education were significant for all vowel phonemes (/oː/ $F_{1,23}$=21.46 p<.001; /eː/ $F_{1,23}$=8.77, p=.008; /ɛi/ $F_{1,23}$=15.58, p=.002; /ɑu/ $F_{1,23}$=5.98, p=.024; /œy/ $F_{1,23}$=11.32, p=.003); compare figure 4.23 (and see right part of all plots of figure 4.24, page 81).

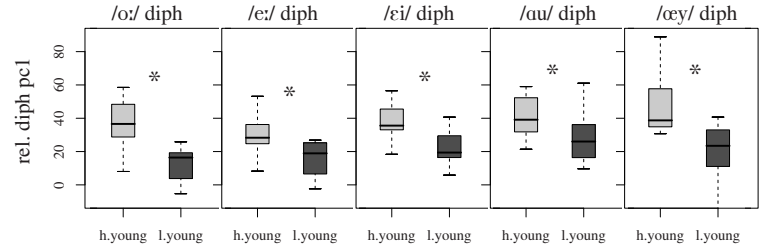


Figure 4.23: *Boxplots of the relative pc1 degrees of diphthongization (diph) of all speakers of the young age group, split into high (h: light grey boxes) and low educated (l: dark grey boxes) speaker groups. Stars indicate significant differences between the levels of education within this generation.*

Since we knew of the significant effect of level of education on the results, we investigated to what extent the main effect of 'age group' was apparent in both or merely one of the levels of education. Further analyses were run on the educational levels separately.

The data and subgroups are visualized in figure 4.25 on page 82. The speaker subgroups are rather small and the results should therefore be seen as indications.

Effects within the Low Educated Group

There were no significant generational effects within the low educated group (see right-hand side of fig. 4.25, p. 82). Analyses of variance with the onsets, and degrees of diphthongization respectively as dependent variable, and 'age group' and 'sex' as fixed factors showed no effects within the group of low educated. Neither 'sex', nor 'age group', nor their interaction affected any of the vowel phonemes' onsets or degrees of diphthongization significantly.

Following this, for the low educated group, there were no significant differences between the pronunciation behavior of the three generations. Also, the vowel phoneme realizations of the females did not differ from that of the males, independent of the speakers' ages.

Effects within the High Educated Group

The same analyses that were run on the group of low educated speakers were performed on the group of high educated speakers. Boxplots of the relative onsets and degrees of diphthongization of the high educated speakers for each age group are displayed on the left-hand side in figure 4.25, p. 82. Within the high educated, there were some significant main effects of the factor 'age group':

For the onsets, no significant main effects of 'sex', and no significant interactions of 'sex' and 'age group' were found. Post-hoc tests on the relative onset values revealed that the observed means of the mid generation differed significantly from the observed means of the old generation for the vowel phonemes /oː/, /ɛi/, and /ɑu/. Speakers of the old group had higher vowel onsets than the speakers of the mid group. Table 4.9 on page 83 shows all significant (and almost significant) effects. As can be seen, there were no significant differences between the young and the mid generation.

Post-hoc tests on the relative degree of diphthongizations displayed also no significant main effects of 'sex', and no significant interactions of 'sex' and 'age group', for either of the vowel phonemes. Yet, there was a significant main effect of 'age group' for all vowel phonemes but /œy/ (table 4.9, p. 83). For /oː/, /eː/, /ɛi/, and /ɑu/, the diphthongization of the old and mid generation differed significantly. The old generation diphthongized the vowel phonemes significantly less than the speakers of the mid generation. Considering their diphthongization of /oː/, /eː/, and /ɛi/, the old speakers differed also significantly from the young speakers.

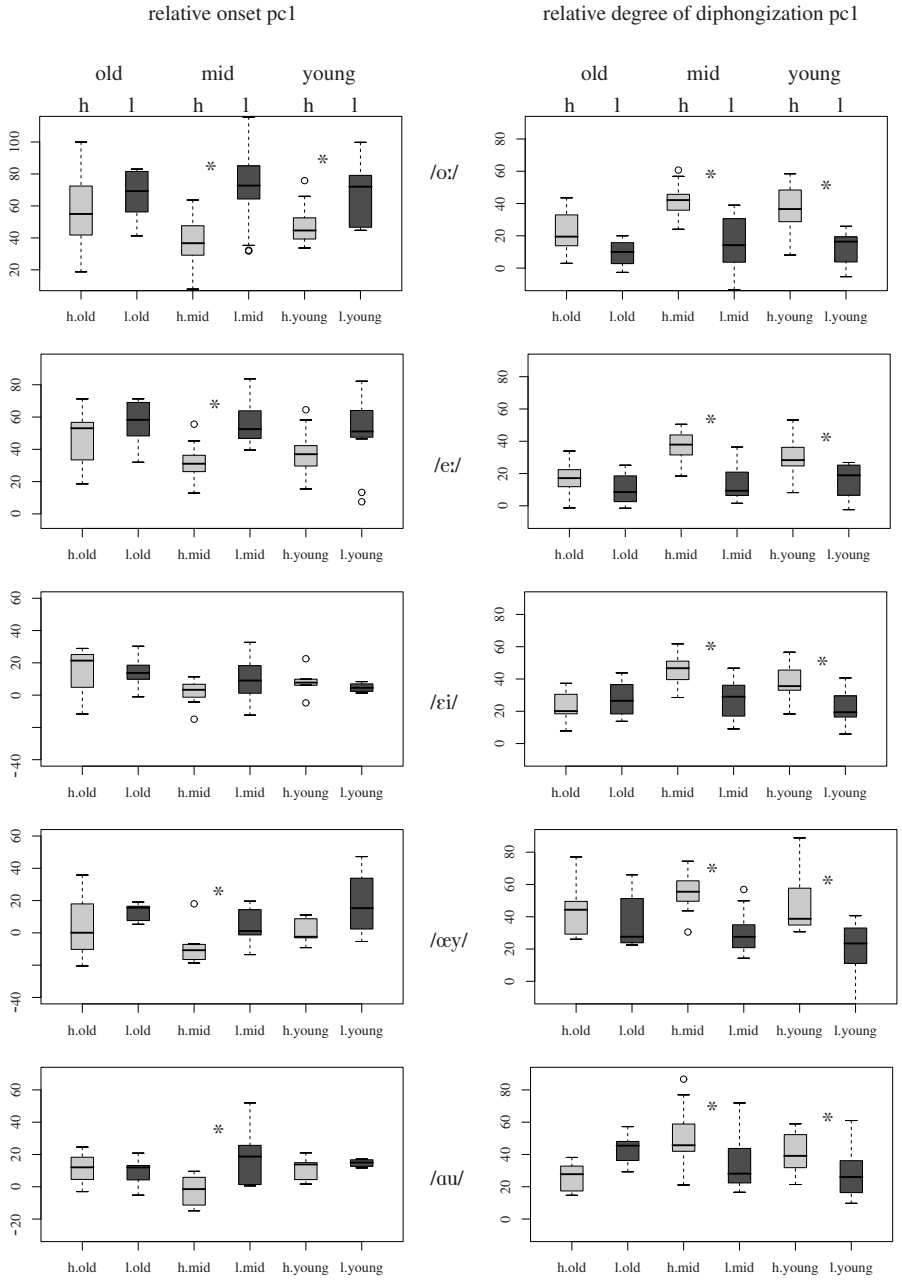


Figure 4.24: *Boxplots of the pc1 of all speakers' vowel phonemes, split into age groups 'old', 'mid', 'young', and split into high (h: light grey boxes) and low educated (l: dark grey boxes) speaker groups. Relative onsets of the vowel phonemes on the left, relative degrees of diphthongization to the right. Stars indicate significant differences between the levels of education within one generation.*

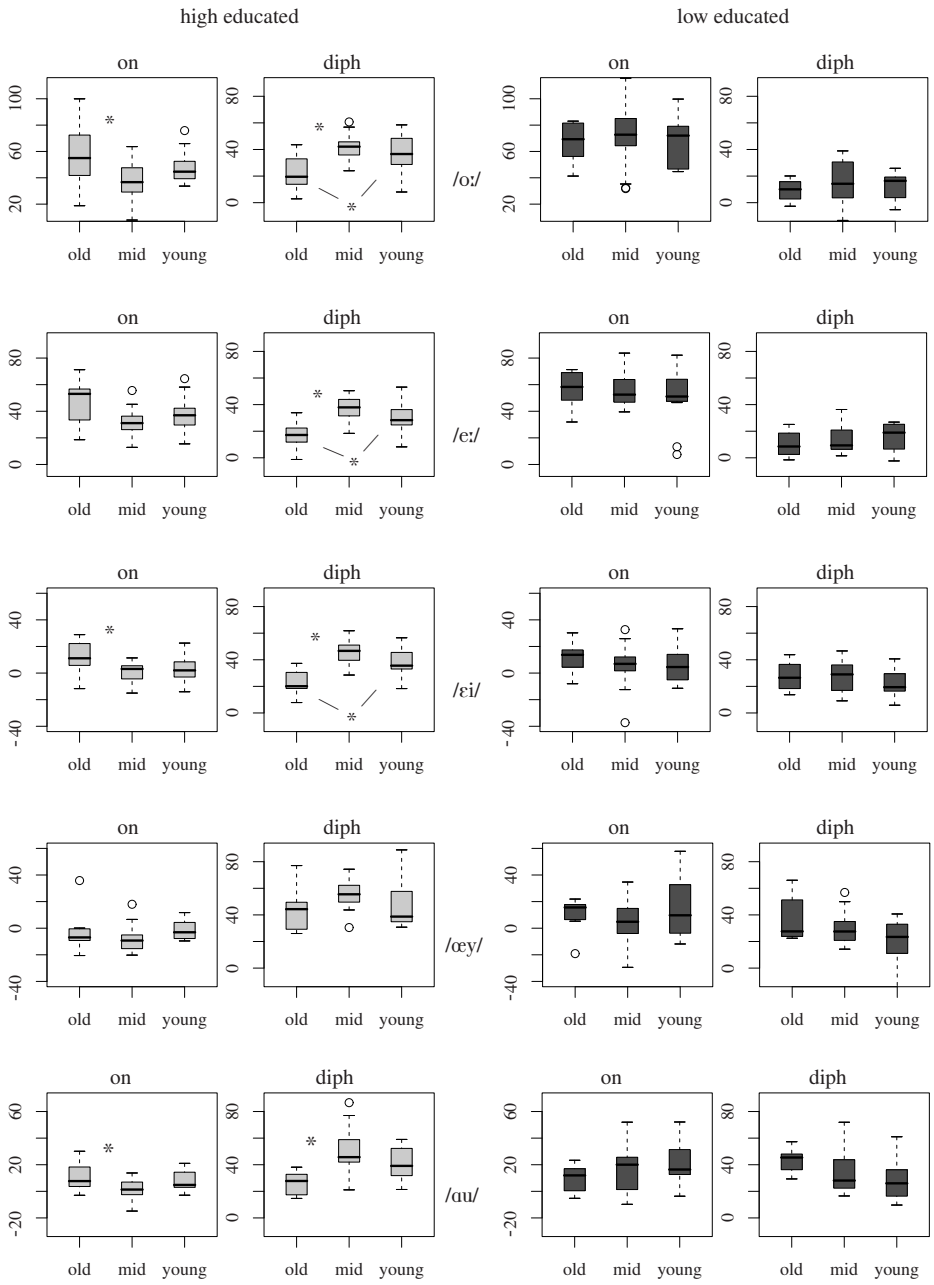


Figure 4.25: *Boxplots of the relative onset (on) and relative degree of diphthongization (diph) in pc1 of all high educated (light grey) speakers' vowel phonemes (to the left) and all low educated (dark grey) speakers vowel phonemes (to the right), split into age groups 'old', 'mid', 'young'. Relative onsets of the vowel phonemes on the left, relative degrees of diphthongization to the right. Stars indicate significant differences between the generations within the same level of education.*

Table 4.9: *Significant (and almost significant) age group differences within the high educated speakers' relative onset values (on) and degrees of diphthongization (diph). No significant, or almost significant, group differences were found for /œy/.*

		Multiple comparisons 'age group'						
	rel.pc1			MeanDiff	Std. Error	Sig.	95% C.I. Lower	UpperBound
/oː/	on	old	mid	20.6	7.9	.043(*)	.497	40.616
	diph	old	mid	-19.6	5.3	.003(*)	-33.069	-6.084
		old	young	-13.8	5.7	.067	-28.338	.714
/eː/	diph	old	mid	-19.9	4.7	.001(*)	-31.820	-8.120
		old	young	-13.4	5.0	.037(*)	-26.176	-.660
/ɛi/	on	old	mid	12.2	4.4	.027(*)	1.149	23.338
	diph	old	mid	-21.9	4.6	.000(*)	-33.581	-10.285
		old	young	-14.9	4.9	.016(*)	-27.401	-2.319
/ɑu/	on	old	mid	10.6	4.2	.050	-.008	21.190
	diph	old	mid	-24.0	6.6	.003(*)	-40.905	-7.186

Based on observed means. * The mean difference is significant at the .05 level.

All in all, the results show that 'level of education' is the most regular main effect, followed by the factor 'age group'. Significant patterns of change in pronunciation between generations were found for the high educated generations. In contrast to this, the low educated speakers showed hardly any change in their pronunciation pattern from generation to generation. Within the low educated speaker group, there was an effect of speaker age as covariate for /ɑu/ and /œy/. This suggests that the assigned age groups do not capture the changing behavior of the low educated speakers, or that the linear changes are too small to be of significance. Figure 4.24 on page 81 displays that for the low educated, the (almost) significant linear effect of 'age' for /ɑu/ and /œy/ is a decrease in diphthongization with increasing speaker age. The same figure also shows that the non-linear significant effect of 'age^2' for the high educated's diphthongization of /eː/, /oː/, and /ɛi/, and the onset of /ɛi/ is caused by a lowering of the onset and increase in diphthongization from the first (oldest) generation to the second generation, and that this increase is reversed again from the second to the third generation.

In the old generation, no significant differences between high and low educated are apparent (compare left part of the panels in figure 4.24, p. 81). In the mid generation, the pronunciation of the high and low educated speakers differs considerably: All vowels but /ɛi/ show significantly different onsets and degrees of diphthongization; for /ɛi/, only the diphthongization pattern differs between high and low educated. The patterns of the young speakers resemble those of the mid generation. However, their pattern is shifted a bit towards the old generation's pronunciation pattern, and the difference between the young and the old generation is less extreme than the difference between the mid and the old generation (compare table 4.9).

Although the level of significance is not the same for the vowel phonemes, the chan-

ging behavior of the long vowels and diphthongs within the high educated speakers is very much alike (see the plots on the left-hand side of figure 4.25, p. 82, and table 4.9, p. 83). Within a generation, the relative degree of diphthongization generally seems to be shared by all five vowel phonemes, so that it is difficult to assign a pull or push chain to the (high educated) vowel data. The visualization of the data might indicate that /ɑu/ and /œy/ were the first in the process of change. Their onsets are rather stable and already low in the oldest and the mid generation. Also, the diphthongization of /œy/ is rather strong with a median above 40 % for the high educated of the oldest generation. The change from the oldest to the mid generation is rather small compared to the changes in the realization of the other vowel phonemes. However, these are only speculations and in general, the accordance between the behavior of the vowel phonemes' relative onsets and degrees of diphthongization suggest that the vowel phonemes changed relatively simultaneously.

Having analyzed the factors that account for the largest part of variance in the data, we still want to make sure that the vowel phonemes of the assigned 'Standard Dutch' speakers carry no significant effects of speaker region. Keeping in mind that the regional background is not equally spread in our data pool, in the next section, we nonetheless tried to check on regional influences on the vowel realizations.

4.4.6 Effects of Region

As mentioned in the first section of this chapter, educational or residential regions of the speakers were not represented evenly (refer to tables 4.3 and 4.4 on page 51). The equal spreading of regions had not been a primary necessity in our corpus design since all speakers had been assigned as Standard Dutch speakers, so that we expected little regional influence.

Yet, we still wanted to see to what extent regional effects might coincide with factors that had been found to have a statistical effect on the measured vowels. And, if there were regional influences, if these effects would be salient within the two levels of education. ANOVA's and post-hocs on all speakers' onsets or degrees of diphthongization with 'region of residence' and 'region of education' as fixed factors showed significant differences for only the onsets of /eː/ (residence region p=.020, educational region p.=025), and /oː/ (residence region p=.004, educational region p.=003) between the regions 1 and 4; the central region and the south peripheral region.

Effects of Region on the High Educated Speakers

Analyses of variance were run on all high educated speakers' degrees of diphthongization and onsets with region of residence (4 levels) and region of education (4 levels) as fixed factors. For none of the vowel phonemes did any effect or interaction reach significance, neither for the onsets, nor for the degrees of diphthongization.

Effects of Region on the Low Educated Speakers

The same analyses on all low educated speakers' degrees of diphthongization and onset, with region of residence and region of education as fixed factors, did show some effect, though not significant. A post-hoc test on the effect of 'region of education' on the onsets of /oː/ reached almost significance (p=.065) for the central region 1 (mean 50.66) vs. the south peripheral region 4 (mean 80.99). The same was the case for the onsets of /eː/. Here, speakers of the central region 1 (mean 47.35) differed almost significantly (p=.064) in their relative onset from speakers of the south peripheral region 4 (mean 63.9). Speakers of the central region lowered their onsets more than speakers of the south-peripheral region.

Following these results the low educated speakers seem to be more affected by a regional pronunciation than the high educated speakers. However, none of the effects of region was significant, and it has to be kept in mind that the regions were represented very unequally, and so the results are not very representative or reliable.

4.5 Summary

Dutch vowel variants of 70 speakers were taken from a spoken Dutch speech corpus, the CGN (Oostdijk et al., 2002 [111]). The purpose was to analyze changes in long vowel and diphthong quality dependent on the speakers' sociological backgrounds and ages, and to deal with the variable recording qualities of the corpus. Realizations of the vowel phonemes /ɛi/, /ɑu/, /œy/, /oː/, and /eː/, as well as /a/, /i/, /u/, were measured and compared on the basis of more than 22000 vowel tokens. All vowels were taken from spontaneously uttered sentences and were analyzed automatically. They were presented in a space, based on a principal component analysis (PCA) on the anchor vowels /a/, /i/, /u/ Bark-filtered spectra. For comparison, automatic formant analyses were performed as well.

Recalculating spectral positions in the principal components (pc's) plane displayed the spectral interaction in the pc1-pc2 plane, and explained the high correlation of the first two formants with pc1 and pc2. The first pc's turned out to be rather insensitive to sex-differences, but they were sensitive to the background noise accompanying the speech data. Variable recording qualities manifested themselves in speaker-specific locations and sizes of the vowel spaces. For a detailed analysis of the effects of noise, vowel spectra of good quality were transformed to poorer signals by increasing the lowest possible dB values per filter. With increasing noise, the positions of the vowels shifted. Having analyzed the influence of noise on our data, we decided to normalize the data by taking each speaker's /a/ and /i/ positions as references for an inter-speaker comparison. This resulted in a new definition of the acoustic attributes of the long vowels and diphthongs in terms of relative onset values and relative degrees of diphthongization. These acoustic measures were potentially powerful to express a lowering of the onsets and the amount of diphthong-

ization for both long vowels and diphthongs.

To detect certain patterns within the vowel data, we concentrated on the pc1 values of the speakers, which explained most of the variance in the data, carried no effects of sex, and had also been the most efficient cue to indicate the perceived lowering in the preliminary study. Main purpose was to find out if a speaker's vowel set would highlight his or her socio-economic status in terms of educational or occupational level. The level of occupation (high or low) and the level of education (high or low) turned out to be the same for all, except one speaker, and so we concentrated on one level, the level of education.

The onset positions appeared to be more or less linearly correlated with the degree of diphthongization (compare fig. 4.13, p. 68); the lower the onset, the stronger the diphthongization. But, generally, the degree of diphthongization was computed as the more reliable cue to the speakers' background than was the onset position. For some speakers, the educational and residence regions affected the onsets of the vowel phonemes /oː/ and /eː/. No effects of region were found for the degrees of diphthongization. Speakers who were educated or resided in the Randstad-cities showed lower onsets for these vowel phonemes than speakers of the south-peripheral region. When split into high and low educated, contrary to the high educated, the low educated turned out to be affected significantly. However, the data did not equally cover the residence regions (compare tables 4.4 and 4.3, page 51).

The results of all measurements clearly showed different vowel quality patterns dependent on the speakers' educational level (fig. 4.26) and age, and indicate a progress of quality changes, with as parameters the degree of diphthongization as most meaningful parameter, and second, the lowering of the long vowels and diphthongs. The high educated speakers showed varying directions of change the younger the speakers' age was (fig. 4.27, p. 87), whereas for the low educated speakers, the few changes were gradual. The low educated speakers' onsets of /ɑu/ and /œy/ were lower the younger the speakers' ages were, whereas for the high educated speakers, from older to middle aged speakers,

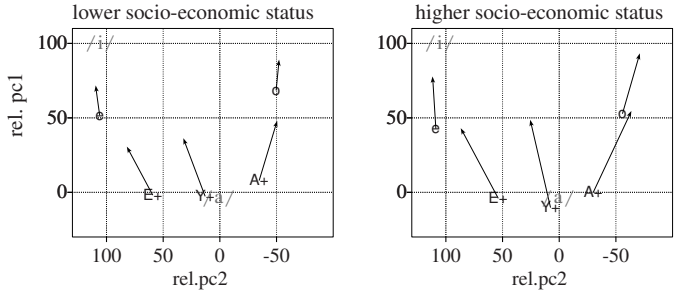


Figure 4.26: *Relative pc-values of the mean vowel pronunciation patterns of the 35 low (left) and the 35 high (right) educated speakers. /a/ has a relative pc1/pc2 value of 0, and /i/ has a relative pc1/pc2 value of 100 (compare section 4.3.2). The vowels' onsets are represented by 'e' for /eː/, 'o' for /oː/, 'E+' for /ɛi/, 'Y+' for /œy/, 'A+' for /ɑu/. The arrow represent the degree of diphthongization.*

the onsets got lower and the degrees of diphthongization got stronger, only to slightly reverse again within the youngest speakers (fig. 4.27).

All in all, the results indicate sound changes for all measured vowel phonemes of the high educated speakers. The most salient sound changes were found from the old to the mid generation, with as conspicuous parameters the changing degrees of diphthongization of all vowel phonemes, especially /oː/ and /eː/, and secondarily, the changing onsets (fig. 4.27). In general, the direction of the vowel changes and the social markedness confirmed what was reported by Mees & Collins (1983, 2003 [96, 21]), and by Stroop (1998, 2003 [140, 141], see section 1.3.2). Mees & Collins had assigned a popular stronger diphthongization to the cities of the Randstad and their younger non-conservative speakers, whereas Stroop assigned the phenomenon to 30 to 40 year old avant-garde females (under the name 'Polder Dutch'). In our data, for both females and males, from the old to the mid generation, by becoming [ɔu]-like and [ɛɪ]-like, the long vowels /oː/ and /eː/ take a position close to the area of the former realizations of the diphthongs /ɑu/ and /ɛi/ (compare figure 4.27[4], and figure 1.1, page 2). The pronunciation changes from the old to the mid generation, which were the most significant in our data, can be described as follows: For the more highly educated, /oː/ changed from [oʊ̈] to [ɔʊ̯], /eː/ from [eᴵ] to [ɛɪ], /ɛi/ from [ɛɪ] to [æɪ], /ɑu/ from [ʌu] to [ɑu], and /œy/ from [ɜʏ] to [ɐə].

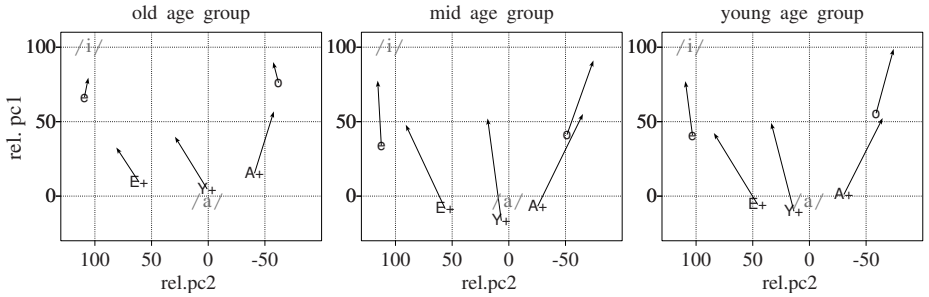


Figure 4.27: *Relative pc-values of the mean vowel pronunciation patterns of the higher educated speakers in the old (left), mid (center), and young (right) age group. /a/ has a relative pc1/pc2 value of 0, and /i/ has a relative pc1/pc2 value of 100 (compare section 4.3.2). The vowels' onsets are represented by 'e' for /eː/, 'o' for /oː/, 'E+' for /ɛi/, 'Y+' for /œy/, 'A+' for /au/. The arrow represent the degree of diphthongization.*

In our corpus of 70 speakers, there were no main or interaction effects of 'sex'. Many sociolinguistic studies reveal differing behaviour according to sex (the biological attribute), or gender (the social construct). Both factors overlap, and disentangling the biological differences in phonetic variation from the socially constructed variation can be difficult. Usually, no distinction is made between the two, as one assumes that unwanted anatomical effects have been normalized by e.g. applying a logarithmic scale to formant

[4] We can now assign the speaker plotted in grey in figure 1.4, p. 13, and recorded in 1999, to the categories 'high educated' and probably 'mid age group'. Whereas for the vowel pattern of the speaker plotted in black we can only suggest that he is rather not a high educated speaker of the mid or young age group.

values, and referring to anchor vowels or the vowel space size. The fact that differences between female and male behavior are a general finding in the formants of vowel variation studies, and given the significant sex-differences in our calculation of the formant vowel space in Bark (section 4.3.1), arises the question to what extent the normalization procedures applied actually normalized for biological vocal tube attributes, and thus, to what extent reported behavior differences in vowel realization are indeed due to gender and not due to sex. Following Heffernan (2007 [49]), the relationship between findings that for both sexes, speakers with less dispersed vowel spaces tend to lead merging changes, could indicate that sex-significances in vowel changes are determined by the differences in vowel dispersion, which are significant between sexes in terms of vowel formants. Due to differences in the vocal tract sizes, generally, female vowel formants are more dispersed than their male counterparts. Usually, logarithmic scales are applied to normalize sex differences, however, there is a lack of thorough research on possible remaining effects of sex as opposed to gender in these logarithmic Hertz-values. In our formant data, we found sex-significances in the vowel dispersion as well, even after applying a logarithmic transformation. Also, females are reported to produce longer vowels than males, and longer vowels, in turn, are articulated more clearly (i.e. less centralized) than shorter vowels. In our corpus, females showed longer duration for the vowel segments (table 4.6, p. 67).

This would suggest that reported sex-significances in vowel changes might not always be due to a social pronunciation construct. Instead, significances could be due to the fact that, statistically, female vowels are longer and more dispersed than male vowels, resulting in artifacts in the formant values by the unsolved problem of normalizing differences in the speakers' vowel dispersions, i.e., vocal tract sizes. On the other hand, the acoustic differences in vowel dispersions probably affect auditory perceptibility, and vowel qualities in dispersed vowel spaces should be easier to distinguish by listeners than the equivalents in speakers with less dispersed vowel spaces. Detailed investigations are needed to decide whether the reported sex differences in vowel changes in terms of formants are really social constructs and not attributes of the biological sex differences, and whether they are indeed audible to listeners. If the acoustically larger formant dispersions for the females were auditorily also more salient than the equivalent but less dispersed vowel spaces of the males, it would explain why females are often seen as the leaders of sound changes, and it would disqualify 'gender' as factor.

In view of the absence of a main effect of either gender or sex in our data, by building pc's on the stable anchor vowels /a/, /i/, /u/, we filtered out large parts of these male/female differences (the biological attribute) that are apparent in formant measurements. As can be seen in the /a-i-u/ vowel plots of figure 4.6 on page 58, and supported by statistics, the female and male vowel triangle areas in formants differ considerably, even after applying the quasi-logarithmic Bark scale (as reported in section 4.3.1, from p. 57 on). The differences in the vowel triangle areas between the sexes were not significant in the pc-dimensions,

before and after normalization. By relating each vowel to speaker-specific /a/ and /i/ values, we could normalize the remaining vowel space size differences. The argumentation in chapter 6 on variation and social behavior will also support an absence of gender differences in active speech communities that include both females and males.

Generally, more strongly diphthongized vowels brought about longer durations and lower onsets (fig. 4.12/4.13, p. 67/68). The relative degrees of diphthongization of the five vowel phonemes turned out to correspond highly within the more highly educated speakers (compare fig. 4.17, p. 72). Speakers strongly diphthongizing the genuine diphthongs also showed stronger degrees for /oː/ and /eː/. For the low educated speakers the degrees differed. On one hand /oː/ and /eː/ corresponded with each other, and on the other hand /ɛi/, /œy/ and /ɑu/ (compare fig. 4.17, p. 72), reflecting the traditional separation of long vowels and genuine diphthongs which was not apparent any more for the high educated.

The present data of the high educated speakers and their pronunciation changes might suggest that /ɑu/ and /œy/ were the first in the changing process: Compared to the other vowel phonemes and changes, their measured onsets are rather stable and already low in the oldest generation, with no further lowering in the mid generation. The diphthongization of /œy/, with a median above 40% of the /a/-/i/ distance, was rather strong for the speakers of the oldest generation, and changes (in the subgroups) from the oldest to the mid generation are insignificant, contrary to all other vowel phoneme changes. Nonetheless, generally, the behavior of the vowel phonemes' relative onsets and degrees of diphthongization change in accordance in the formed subgroups. In view of this phenomenon, and considering a chain reaction of sound changes, we would suggests that the 'degree of diphthongization' is an attribute that the speakers applied to all vowel phonemes rather equally.

Altogether, the results support the importance of the social background of the speaker when describing the acoustic quality of the long vowel and diphthong phonemes of Standard Dutch. For a reliable variation analysis, pc dimensions based on barkfilter output turned out to be more reliable than formant measurements. In our data, socially structured variation was apparent in the degree of diphthongization and onset of not only /ɛi/, but also in the other analyzed long vowel and diphthong phonemes. Unexpectedly, there were no significant differences in the realizations of males and females, contradicting the hypothesis of female precursors in the lowering process. The most important social factors that could be related to vowel variation were the level of education and occupation, and the speakers' age. In chapter 6, after having run a perception experiment in the following chapter, we will discuss some explanations for the structure of the differences in realization found within our speaker data. First, in the next chapter, we try to find out if at least the largest group differences in the acoustic vowel realizations that we found in the present chapter, are audible to listeners, and if the listener background has an effect on how the differences are perceived.

5. PERCEPTUAL DISSIMILARITY OF ACOUSTIC DIFFERENCES

Abstract In the previous chapter the variation in the acoustic vowel data in terms of onset and degree of diphthongization turned out to correlate with aspects of the speaker background. In this chapter we investigated in a small perception experiment to what extent listeners differentiate these sub-phonemic acoustic differences. Listeners had to judge whether vowel realizations of various speaker pairs were of the same or of a different quality. The response behavior revealed that realizations that were found to differ significantly in the previous chapter on acoustics, were differentiated by all listeners as well. Across all realizations, the larger the acoustic distances between the realizations, the more listeners perceived the realizations as differing. However, which acoustic difference matched the listeners' responses best was phoneme dependent: only for /eː/ did the acoustic distances explain a substantial amount of variance in the data. Effects of listener age on the perceived realization differences were tested as well. Including the listeners' ages in the response analysis increased the predictability of the listeners' responses behavior, especially for /œy/. When the listeners were split into age groups (old, mid, young), the response behavior of the young and the old age groups was comparable, whereas listeners of the mid age group appeared to differ, indicating that the listener age effects are comparable to the speaker age effects in the previous chapter.

5.1 Introduction

In the previous chapter, the speakers' articulatory-acoustic realizations of vowel phonemes were found to differ according to their social background. We conducted a perception experiment on the discriminability of sub-phonemic vowel variants in their original word environment by using speech from the spontaneous corpus that was analyzed acoustically in the previous chapter. The purpose was to verify that phoneme realizations which were found to differ significantly in the acoustic analysis in the previous chapter, are also differentiated by listeners. The core assumption was that at least part of the measured acoustic variation is perceived.

In the perception task described in the following sections, listeners had to give 'different' or 'similar' judgments on realizations of the same vowel phonemes in words taken from the spontaneous speech of six speakers. If the listeners' perception somehow corresponds to the significant acoustics described in the previous chapter, they should at least be able to differentiate rather extreme realizations of the acoustically defined vowel variants: within each phoneme, a strongly diphthongized vowel with a low onset versus a slightly diphthongized vowel with a high onset would represent the two extremes of the measured variation continuum of the data analyzed in the previous chapter. We expected all listeners to judge these two realizations as differing in quality. Though in general, vowel onset and degree of diphthongization correlated positively in the acoustic data of the speakers in the previous chapter (the lower the onset, the stronger the diphthongization), one of the two might be more speaker-dependent and/or auditorily more salient. Also, other values such as duration and f0 might be relevant, even if they were insignificant in the acoustic analysis of the previous chapter.

Experiments on the perception of sub-phonemic vowel variation are rare. One experiment dealing with sub-phonemic vowel categories investigated the identification of phoneme categories in the presence of a merger-in-process in the vowel pronunciations of New Zealand English (Hay et al., 2006 [47], and Warren et al., 2007 [163]). When listeners had to decide which of two words with merging pronunciations (e.g. <cheer> or <chair>) was uttered, the perceptually favored word was biased when social speaker attributes were available to the listeners in terms of speaker photos. In our case, no effort is made to evoke social stereotypes. Also, we will not be dealing with merging vowel phonemes, and there will be no doubt about which word was uttered by the speaker. The available information on the speaker will be the same for all listeners, and, except for the speech stimuli, the listeners got no social information on the speaker's background.

In addition to social information that is attached to the speech, the listener's background might have an effect on the task, as implied by a study on Dutch speech variants by van Bezooijen (2001 [151]): She had a group of younger females, and a group of older females judge speech samples of the Polder Dutch variety, of Standard Dutch, and of two

dialects. For each variety, she used random sequences of speech fragments of representative speakers: listeners marked on a seven-point scale whether they considered the speech 'normal' or rather 'deviant', or rather 'modern' than 'oldfashioned'. The results show that younger females had a more positive attitude towards Polder Dutch than older females. When groups of young and old males were included (van Bezooijen et al., 2001 [153]), they turned out to agree with the older females in their evaluation of the Polder Dutch variety. However, in their judgement to what extent the variants are 'normal', the listener generations differed in their answers independent of sex, with the young listeners being more habituated to the Polder Dutch variant. Our goal was simply to confirm that the acoustically salient variation in the vowel is perceived as well by all listeners. Listeners had to put their attention on the target vowel and thus concentrate on fine phonetic detail. Nonetheless, we gathered information on the listener's background, in case there might still be some listener effects.

Before presenting the results, in the following section, we will describe the stimuli and the design of our perception experiment, followed by the instructions the listeners were given when proceeding through the task.

5.2 Method and Material

A small same-different experiment using the AX-paradigm was carried out to investigate how well listeners can differentiate vowel variants of the same phoneme. For this purpose, the listeners had to compare words uttered by various speakers pairwise in terms of their similarity or difference in the realization of a target vowel phoneme.

5.2.1 Stimuli

Two males (A, F) and four females (B, C, D, E), and their realizations of the 5 vowel phonemes /eː/, /oː/, /ɛi/, /œy/, and /ɑu/ in their original word context were taken from the spontaneous speech data that was described in section 4.2. Our primary interest was whether the realization differences that were found to be significant in the previous acoustic chapter (e.g. the high versus the low educated speakers of the mid age group) can be (consciously) differentiated by normal listeners. For this purpose, two of the six speakers (B, D, see fig. 5.1, p. 94) had been selected as clear representatives each of one end of the measured acoustic variation continuum. As can be seen in figure 5.1, p. 94, speaker D shows noticably higher vowel onset positions and less diphthongization for the vowel phonemes than speaker B. For a perceptual validation of our significant acoustic categories from the last chapter, the listeners should be able to differentiate the variation in realization between these contrasting speakers. The other speakers A, C, E, F, were chosen randomly. All speakers' relative pc1 values are displayed in table 5.1 on the following page.

To keep the stimuli as natural as possible, the vowels were presented in their original

Table 5.1: *Relative pc1 mean values of the six speakers' realizations of the vowel phonemes /eː/, /oː/, /ɛi/, /œy/ that were chosen for the perception experiment. (f) for female and (m) for male.*

	speaker	A(m)	B(f)	C(f)	D(f)	E(f)	F(m)
/eː/ e	rel.onset	55	-4	57	52	42	37
	rel.diph	60	81	49	13	6	13
/oː/ o	rel.onset	20	-22	49	82	19	-23
	rel.diph	55	92	18	6	54	36
/ɛi/ E+	rel.onset	-22	-22	1	27	16	-11
	rel.diph	89	65	19	39	46	58
/œy/ Y+	rel.onset	-24	-50	-7	26	-2	3
	rel.diph	79	77	97	26	33	50

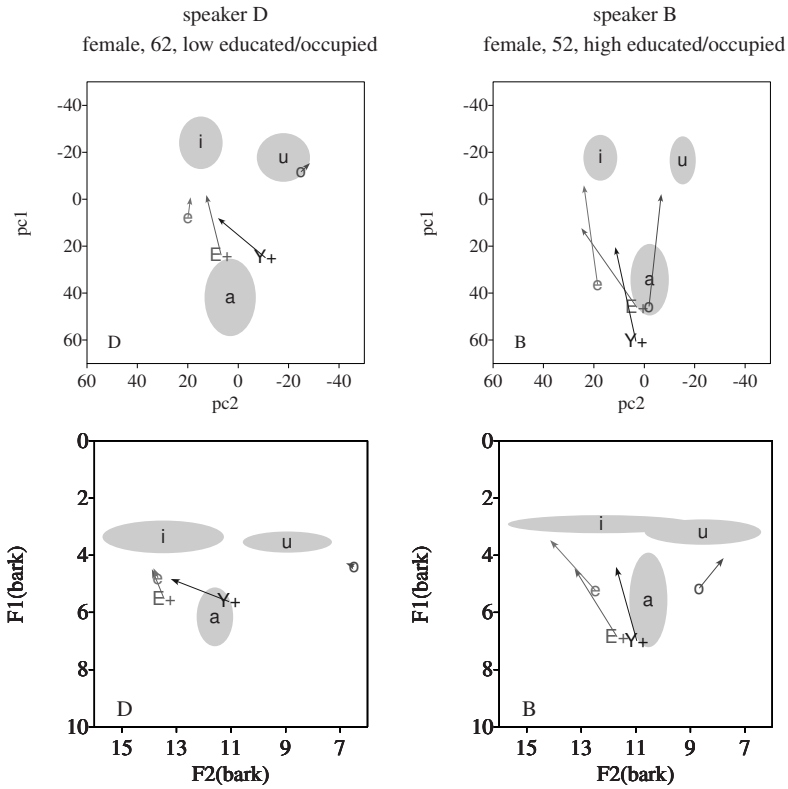


Figure 5.1: *Pc1/pc2 (top) and F1/F2 (bottom) dimensions of the vowels of the two in acoustic terms most contrastive speakers: The plotted anchor vowels /a,i,u/ with sigma elipses were based on all available sound segments in the corpus of speakers that was used in the previous chapter. The displayed vowels /eː/, /oː/ (e, o) and /ɛi/, /œy/ (E+, Y+) are based on the vowel means of the (two or three) stimulus words used for this experiment. Onset and offset of the vowels are connected by an arrow, representing the degree of diphthongization. The words they belong to are annotated in Table 5.2, p. 95.*

environment within single words. Due to the various speech topics and idiolects of the spontaneous speech, it was not possible to choose the same environment (words) for the vowel phonemes across all speakers, and so we decided to represent each vowel phoneme of a speaker by preferably three words. For each speaker, preferably, the target vowel was embedded in different consonantal contexts. To reduce unwanted influence on the judgements by other levels of speech processing, the words that were taken as stimuli had not been uttered sequentially.

Table 5.2: *Stimulus words per speaker and vowel phoneme*

phoneme category:	/ɛi/	/œy/	/eː/	/oː/
words speaker A:	wij actualiteit krijgen	buitenlandse huidige	daaromheen lezen	grote gekozen
words speaker B:	bekijken nationaliteiten vijf	buitenlanders Duits luisteren	collegezaal lezen	diploma gesloten nodig
words speaker C:	uiteindelijk zij krijgt	geruis duiken	uitgaansleven tegelen weet	grootste monument afgoden
words speaker D:	cichoreikoffie kijken	d'ruit huisgezinnen tuin	privileges tweede	bovenop gekookt hoogste
words speaker E:	bijgebouwtje zijkant	uitstappen buiten uit	bezem afgegeven stenen	geloof overstappen ook
words speaker F:	rijst kleine	kruiden uien	meestal varkensvlees	stoofpot ook hardgekookte

We selected, where possible, polysyllabic words that included vowels from other areas of the acoustic space, as it is known that listeners normalize a speaker's vowel by taking into account his or her vowel space. Also, in the previous chapter, the acoustic quality had been measured in relation to each speaker's individual vowel space size, by basing the PCA on all speaker's /a/, /i/, /u/, and relating acoustic distances to each speaker's /a/ and /i/ position or distance. Within a speaker, we tried to choose words where the acoustic realizations of the vowel phoneme closely matched; thus, realized with as much agreement as possible in terms of the onsets and/or degree of diphthongization. On account of these acoustic demands, and to control for other unwanted influences, we had to exclude some of the words initially chosen of a speaker with e.g. conspicuous differences in f0 or other acoustic dimensions, words with swallowed endings or other salient forms of reduction,

and words where the recording or prosodic quality differed too much from those of the other chosen words. As a consequence of meeting these various requirements, each vowel phoneme of a speaker was presented either by three or two single unconnected words. These words were separated by 500 ms of silence. Due to shortcomings in quality, /ɑu/ was finally left out of the task, leaving realizations of /eː/, /oː/, /ɛi/, and /œy/ of each of the six speakers for comparison. The words used are displayed in table 5.2, p. 95. All words were cut, equalized in intensity, and the signals were faded in and out using the Praat [12] software.

5.2.2 Procedure

Regarding previous positive experiences by phoneticians of our institute (most recently see Jongmans, 2008 [67]), we preferred to run the experiment online. To reach more listeners, the online format should demand as low an expenditure of time as possible from participants. After a pre-trial, the stimuli were presented via a web-interface[1] (compare figures 5.2 and 5.3, p. 97). First, the listeners were asked to state their age (*'leeftijd'*), sex (*'vrouw' / 'man'*), highest education (*'Wat is uw hoogst genoten opleiding'*), and whether they resided in one of the cities of the 'randstad' (*'Woont u in een van de randsteden'*). Next, they were asked about hearing loss (*'Is gehoorverlies bij u bekend'*), whether Dutch was their mother tongue (*'Is Nederlands uw moedertaal'*), and whether they had ever been phonetically trained (*'Heeft u een opleiding gehad in de fonetiek'*).

Before beginning the task, the listeners were instructed how to proceed during the experiment and how to use the sound buttons with the stimuli, which they were allowed to push and listen to repeatedly. The listeners were also instructed to choose a silent place, to adjust the volume, and to wear headphones when listening to the stimulus words (see fig. 5.2, p. 97). Allowing the listeners to participate in this online experiment at a place of their own choice implies that we could not check whether the listeners followed these instructions.

The listeners had to judge in total 66 stimulus pairings (4 vowel conditions in 15 speaker pairings, plus 6 repetitions), which on average took about 20 minutes. The first six stimulus pairings were the same for all listeners to familiarize them with the task and stimulus mode; they were repeated at the end. All other stimulus pairings were presented randomly, preventing only same stimuli from appearing in a row. Each pairing was presented on a separate webpage in the same standard form. An example is given in figure 5.3, p. 97. As can be seen, the words belonging to one speaker were represented by one clickable sound button. Per stimulus pairing, the participants saw two sound buttons, one for each speaker, each with the speakers' words written on it orthographically, and the target

[1] Our experimental design was supported by van Son's freely accessible web-form to construct online listening experiments: http://www.fon.hum.uva.nl/Service/Experiment/ConstructExperiment.html.

vowels visually marked. In figure 5.3, the words of the first speaker are *'hoogste, bovenop, gekookt'*, the second speaker's words are *'nodig, gesloten, diploma'*. In the lines above the buttons, the listeners were asked to attend to the pronunciation of the marked vowels, in the case of figure 5.3 it was the vowel phoneme /o:/, and the letters that represent it were marked in each word.

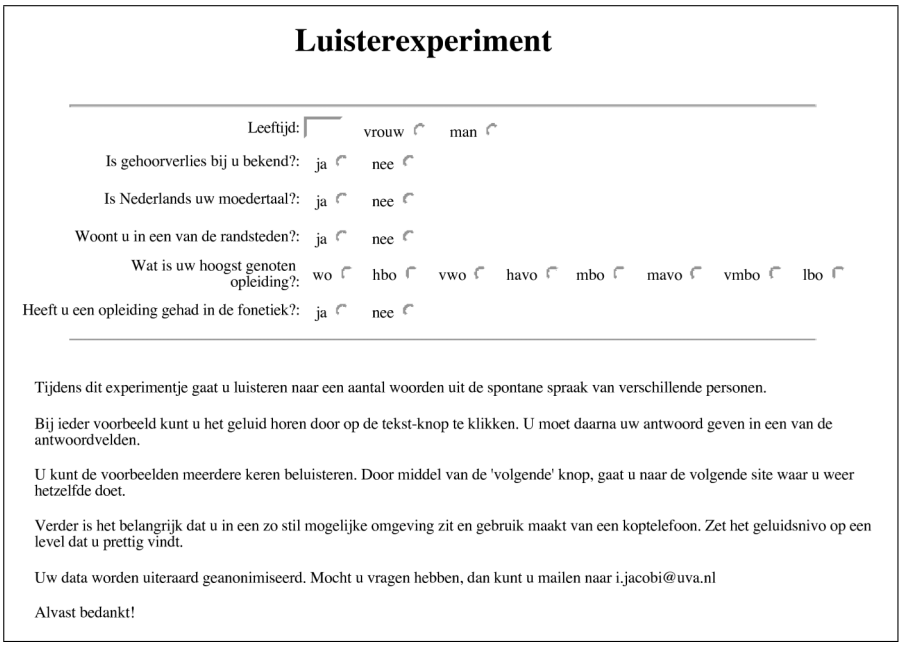


Figure 5.2: *Questionnaire prior to the participation in the online listening experiment*

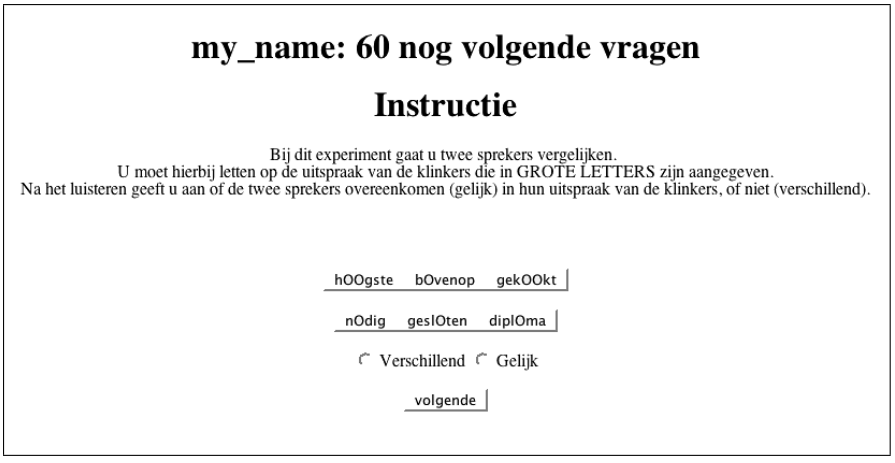


Figure 5.3: *An example page of the web-based listening task*

The listeners could listen repeatedly to the words of each of the paired speakers. They then had to make a forced decision on whether the two speakers pronounced the marked vowels in the same way or not, by marking the box 'same' (*'gelijk'*), or by marking the box 'different' (*'verschillend'*). By clicking on the button 'next' (*'volgende'* in figure 5.3, p. 97), they could proceed to the next stimulus pair.

The null hypothesis of the experiment was that the response behavior is the same for all stimulus pairings, and that the vowel realizations of the various speakers are not differentiated by the listeners.

5.3 Results

Thirty listeners, 18 females and 12 males, all with Dutch as their mother tongue, participated. 26 of them were inhabitants of one of the cities of the 'randstad'. Their mean age was 43.3 years (range 24-68 years of age). According to our categorization from the previous chapter (4.2.1, p. 50), 27 of the listeners were high educated (wo/hbo) and 3 low educated (mbo, vwo, havo). Three of the listeners indicated they were phonetically trained.

The proportions of the listeners' 'same' versus 'different' responses to each speaker pairing are displayed in table 5.3 below. Since the response variable is dichotomous, the percentage of the respective responses reflects as well the variance in the data. A stimulus response of 50% 'same' and 50% 'different' judgments thus holds the largest variance and would be interpreted as a random decision. As can be seen in table 5.3, speaker pairing BD was found to differ the most in terms of their vowel realizations. So far, this matched our expectations, as these speakers had been chosen as representatives of significantly differing acoustic groups of the previous chapter. Since the table is ordered according to the proportions of 'different' versus 'same' responses, the spreading of differing speaker sex (fm/mf) and same speaker sex (ff/mm) already indicates that speaker sex was not a decisive factor in the comparison task (compare figure 5.7, page 102).

Table 5.3: *Percentage of 'same' vs. 'different' responses to each speaker pairing with speaker sexes (f/m)*

	BD	CD	AD	DE	BF	BE	AB	BC	CF	DF	EF	AF	AC	CE	AE
	ff	ff	mf	ff	fm	ff	mf	ff	fm	fm	fm	mm	mf	ff	mf
SAME	10	19	19	19	23	24	35	43	43	46	48	51	58	61	62
DIFF	90	81	81	81	77	76	65	57	57	54	52	49	42	39	38

A closer look at the data showed that, next to being speaker-pairing-specific, the results were vowel phoneme-specific, with even significant differences in the amount of 'different' vs. 'same' responses between the vowel phonemes of the same speaker pairing (compare figure 5.4, p. 99 and table 5.4, p. 99). Sign tests on the number of 'same' versus 'different' responses split into vowel classes revealed non-random decisions on more than 50% (34 out of 60) of the paired vowel stimuli (compare table 5.4, p. 99, '*' for p<=0.03,

'**' for p<=0.01). Most of the non-random decisions on the stimulus pairs were 'different', but there were also some significant 'same' judgments. The realizations of a phoneme were thus not all judged in the same way.

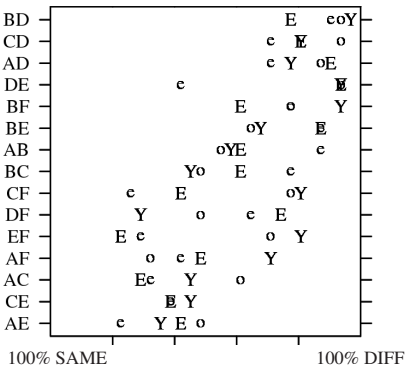


Figure 5.4: *30 listeners' mean response to vowel phoneme realizations of the 15 speaker pairings. 'e' represents /eː/, 'o' /oː/, 'E' /ɛi/, and 'Y' /œy/.*

The most inconsistent pattern in terms of the four vowel phonemes was found for the speaker pairing EF: Whereas the speakers' realizations of /ɛi/ and /eː/ were judged as sounding quite similar, /œy/ and /oː/ were found to sound significantly different. As apparent from figure 5.4 and table 5.4, the speaker pairing BD was significantly different to the listeners in all four vowel phoneme classes ('e' represents /eː/, 'o' /oː/, 'E' /ɛi/, and 'Y' /œy/). The quality of all vowel phonemes of this speaker pairing were judged as differing significantly, as were the vowel realizations of speaker pairings CD, AD, BF and BE. No speaker pairing was perceived as similar in all vowel phonemes.

Table 5.4: *Listener responses per vowel and speaker pairing in percentage. The stars indicate (highly) significant differences between the number of 'same' (SAME) and 'different' (DIFF) responses.*

/ɛi/	DIFF	SAME
DE	** 94	06
AD	** 90	10
BE	** 87	13
CD	** 81	19
BD	** 77	23
DF	* 74	26
AB	61	39
BC	61	39
BF	61	39
AF	48	52
CF	42	58
AE	42	58
CE	39	61
AC	29	* 71
EF	23	** 77

/œy/	DIFF	SAME
BD	** 97	03
DE	** 94	06
BF	** 94	06
EF	** 81	19
CD	** 81	19
CF	** 81	19
AD	** 77	23
AF	* 71	29
BE	68	32
AB	58	42
AC	45	55
BC	45	55
CE	45	55
AE	35	65
DF	29	* 71

/eː/	DIFF	SAME
BD	** 90	10
BE	** 87	13
AB	** 87	13
BC	** 77	23
BF	** 77	23
CD	* 71	29
AD	* 71	29
DF	65	35
DE	42	58
AF	42	58
CE	39	61
AC	32	68
EF	29	* 71
CF	26	* 74
AE	23	** 77

/oː/	DIFF	SAME
BD	** 94	06
DE	** 94	06
CD	** 94	06
AD	** 87	13
BF	** 77	23
CF	** 77	23
EF	* 71	29
BE	65	35
AC	61	39
AB	55	45
AE	48	52
BC	48	52
DF	48	52
CE	39	61
AF	32	68

To relate the listeners' response behavior in context to the acoustic vowel qualities, next, the acoustic distances between the vowel realizations of the six speakers were calculated.

5.3.1 Overall Response Behavior and Acoustic Distances

For an interpretation of the response outcome, between-speaker differences were calculated in terms of the realized vowel onset positions, and degrees of diphthongization. For each vowel phoneme, the speaker's mean acoustic value of his or her two or three realizations was taken. As in the previous chapter, the main focus was on the pc1 dimension.

Since we instructed the listener to put his or her attention on comparing different realizations of the same phoneme, and thus to concentrate on finer phonetic detail, we expected the participants' judgments to correlate with one or some of the measured acoustic dimensions. The more the speakers' realizations differed acoustically, the more we expected them to be judged as 'different', so that the extent of acoustic distance between the stimuli should ideally be reflected in the ratio of 'same' versus 'different' judgments.

However, when the acoustic distances were compared with the listeners' response behavior in general, i.e. phoneme-independently, the concordance between the acoustic values and the mean response was not very strong. Figure 5.5 displays the distances in the pc1 onsets and in the pc1 degrees of diphthongization in the realizations of all vowel phonemes. Roughly speaking, there was a tendency of increasing 'different' responses for increasing acoustic differences in the pc1 onsets (r=-.455, p<.001, thus roughly 21% of the variance is attributed to the distance in the pc1 onset). As can be seen in the left hand plot of figure 5.5, as the distance increases, the 'different' responses increase. The degree of diphthongization (compare the right hand plot of the figure) matched the response behavior of the listeners less.

To what extent acoustic dimensions in terms of distances in onsets or diphthongizations in pc1 matched the response behavior of the listeners is plotted in figure 5.6, p. 101

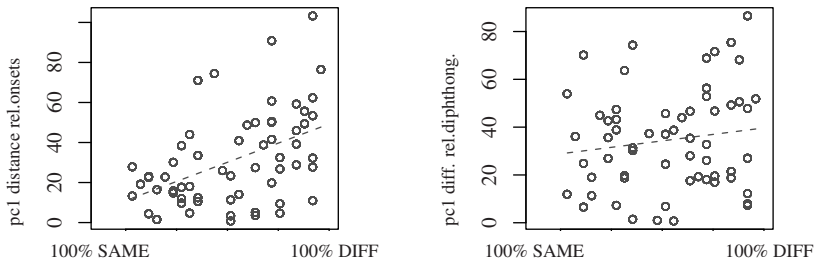


Figure 5.5: *The distance of the relative onsets (left) and the relative degrees of diphthongization (right) in the pc1 dimension of all 60 stimuli versus the mean same-different ratio of the 30 listeners.*

per vowel phoneme. As can be seen, the correspondence of the various acoustic values and the response behavior differed for the vowel phonemes. A clearly categorical behavior is seen in the response behavior to distances in /eː/ onsets (top row, third panel in figure 5.6).

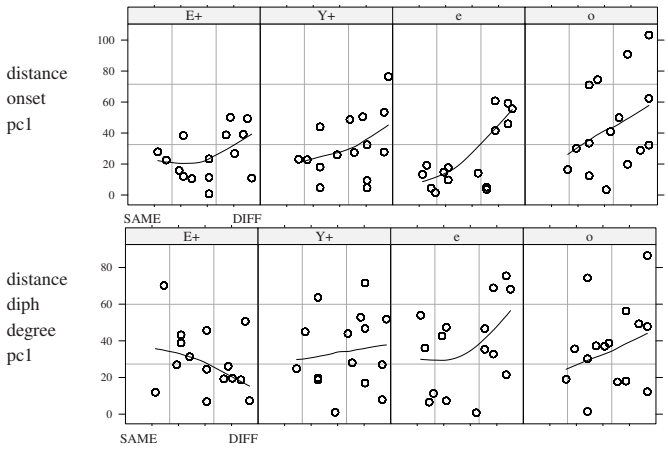


Figure 5.6: *Listener mean response (x-axis) to the phoneme realizations of the 15 speaker-pairings versus acoustic distances (y-axis). 'E+' for /ɛi/, 'Y+' for /œy/, 'e' for /eː/, 'o' for /oː/.*

Which of the acoustic dimensions correlated most with the listeners' response behavior, i.e. whether the distances between the onsets predicted the responses, rather than the differences between the degrees of diphthonization, or their combination, was calculated by logistic (also called binary or binomial) regressions (Hosmer & Lemeshow, 2000 [54]). This kind of regression is used when the probability of the occurrence of a dichotomous dependent (here, the response 'same'(0)/'different'(1)) has to be predicted. In our case, we tested the extent to which the speaker distances in terms of pc1 onsets and degrees of diphthongization predicted the response distribution. For all regression models and for all vowel phonemes, the distances in the relative pc1 onset were useful to predict the response; the differences in the degrees of diphthongization were only of use in predicting the response behavior towards the phonemes /eː/ and /ɛi/. 'Speaker sex' did not improve any of the models' prediction of the response behavior (compare fig. 5.7, p. 102), whereas for /oː/ and /œy/, f0 added significantly to the models' predictions, and for /eː/ and /œy/ it was vowel phoneme duration. However, the predictability of the best fitting regression model for each vowel phoneme was only acceptable for the response behavior towards /eː/ (as already indicated by fig. 5.6). Table 5.5, p. 102 shows the logistic regression coefficients of the best fitting model for /eː/. A relationship between the measured acoustic differences and their perceived (dis)similarity could thus not be generalized for all vowel phonemes.

The primary aim of our perception experiment had been to confirm the auditory dif-

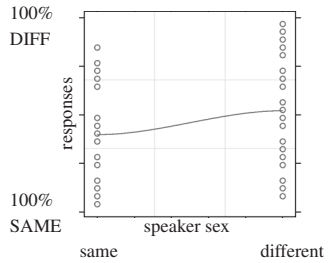


Figure 5.7: *Plot of the mean responses (y-axis) related to differences in speaker sex (x-axis). There was an insignificant tendency of more 'different' responses towards speaker pairs who differed in sex.*

Table 5.5: *Logistic regression coefficients with standard errors of the best fitting models for the response behavior to acoustic distances for /eː/.*

variable	coefficient	S.E.	Sig.
pc1 diph	.028	.007	0.000
pc1 onset	-.027	.006	0.000
dur ms	-.042	.008	0.000
-2 Log likelihood	519.227		
Nagelkerke R^2	.247		
% correct predict resp	71.6		

ferentiability of the acoustic categories that had been found to differ significantly in the previous acoustic chapter, represented by the realizations of speaker B and D. So far, we could confirm that their acoustic realizations were perceived as differing significantly by normal listeners. The concordance of acoustic differences in speaker realizations and the listeners' response behavior was highly phoneme-dependent, and the variance in the response behavior that could be explained by the acoustic distances was only considerable for /eː/. Thus, next to the acoustic distances, other attributes must have affected the listener's response. Next, we tested to what extent variance in the response behavior could have been affected by attributes of the listener.

5.3.2 Age Dependent Response Behavior and Acoustic Distances

Originally, the present experiment was set up to confirm the perceptual reality of the significantly differing acoustics we found in the previous chapter. That these acoustic categories are perceptually significant as well was confirmed in the previous section. However, logistic analyses with the acoustic distances as predictors yielded an acceptable model only for the responses to /eː/. For the other vowel phonemes, there was still a considerable amount of unexplained variance in the response behavior.

As mentioned in the beginning of the chapter, it is quite possible that 'normal' untrained listeners base their decisions on more than acoustic quality differences alone. In addition to speaker-dependent acoustic factors that correlated with response behavior, or other acoustic effects we did not test, listener-dependent factors might have had an effect on the judgments, therefore our data on the listener's background might help to explain

some variation in response behavior.

Before starting the perception experiment, the listeners had been asked to state their highest education, their age, mother tongue, and place of residence (see fig. 5.2, p. 97). Factors such as education and residence region were spread unevenly among the listeners, as almost all listeners were high educated and residents of one of the cities of the'randstad', and all marked Dutch as their mother tongue. Due to their uneven spread, these three factors had to be ignored in the further background data analysis, and we concentrated on the listeners' age.

In the previous chapter, sub-phonemic social-acoustic vowel categories seemed to crystalize from the old to the mid generation (compare section 4.4.5, p. 78), merging again from the mid to the young generation. Following the socio-economic categorization in the acoustic chapter, speaker D, a female aged 52 at the time of recording, belonged to the category 'high educated, high occupied', and 'age group: mid' (compare figure 5.1, p. 94). Speaker B, a female aged 62 at the time of recording, belonged to the category 'low educated, low occupied', and 'age group: old'. These socio-economic groups, and the age groups, were found to differ most significantly in the acoustic dimensions measured previously. If production and perception are as closely connected as the literature suggests (compare section 1.3.1 and section 6.2.3), we might find some listener-dependent effects in the response outcome. Given the results of the previous acoustic chapter, middle-aged listeners then might judge stimulus distances in another way than elderly listeners.

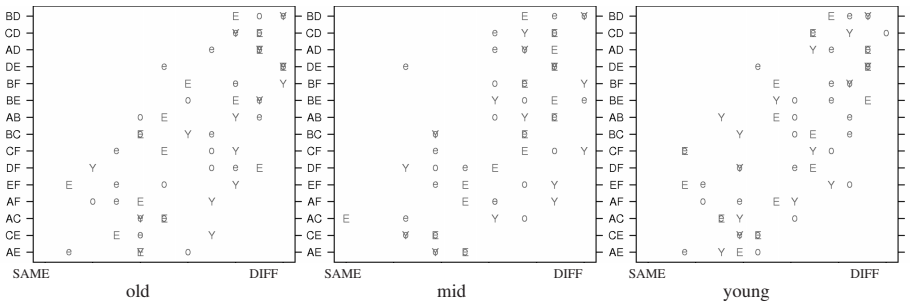
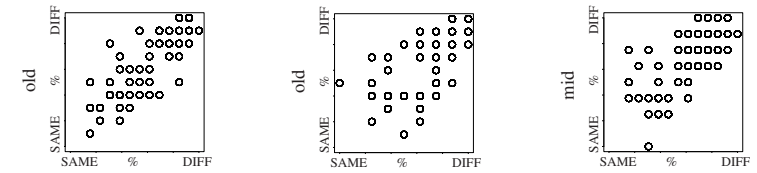


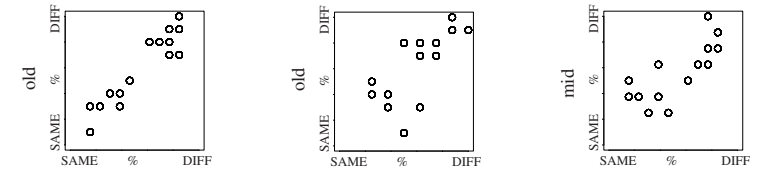
Figure 5.8: *Mean response (x-axis) per listener age group (old, mid, young) to each vowel phoneme of the 15 speaker pairs. The speaker pairs on the y-axis are ordered according to the listeners' overall mean response, with BD being judged as most different and AE as most similar.*

To see whether the effect of 'age group' is also reflected in auditory perception, our listeners were split into the same age groups as the speakers in the previous chapter: 'Old' for listeners of 55 years and older (N=12), 'mid' for listeners above 35 and below 55 years (N=8) of age, and 'young' for listeners below 35 years of age (N=10). Figure 5.8 shows the mean responses of each listener age group to the vowel realizations of the various speaker-pairings. The speaker-pairings on the y-axis are ordered according to the mean



r= .80 young r= .65 mid r= .71 young

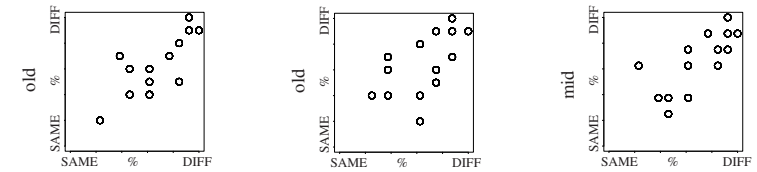
Figure 5.9: *Mean responses and correlation of the three listener age groups (old, mid, young) to all 60 stimulus pairings.*



/eː/

r= .94 young r= .73 mid r= .79 young

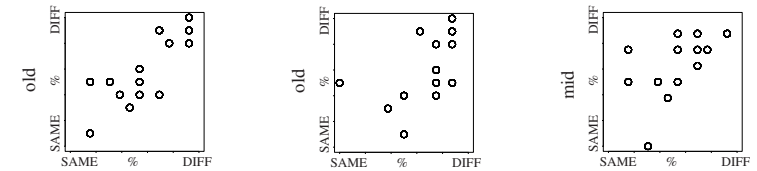
Figure 5.10: *Mean responsess and correlation of the three listener age groups to the 15 /eː/ stimulus pairings.*



/oː/

r= .78 young r= .55 mid r= .71 young

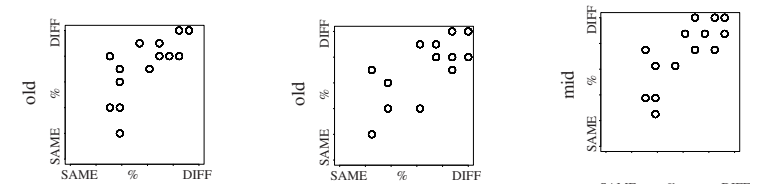
Figure 5.11: *Mean responsess and correlation of the three listener age groups to the 15 /oː/ stimulus pairings.*



/ɛi/

r= .79 young r= .55 mid r= .60 young

Figure 5.12: *Mean responses and correlation of the three listener age groups to the 15 /ɛi/ stimulus pairings.*



/œy/

r= .75 young r= .74 mid r= .78 young

Figure 5.13: *Mean responses and correlation of the three listener age groups to the 15 /œy/ stimulus pairings.*

response of all 30 listeners, with the speaker-pairing that was differentiated most clearly, the one getting the most 'different' responses, at the head (BD). The responses of the age groups are correlated on page 104. All plots on the left compare the response behavior of the old age group on the y-axis with that of the young age group on the x-axis. The plots in the middle show the responses of the old age group (y-axis) versus the responses of the mid age group (x-axis), and the plots on the right show the responses of the mid age group (y-axis) versus those of the young age group.

Generally, the plots in the left column of page 104, i.e. those that matched the responses of the old generation with the young, show a rather linear array of the data points. The correlations are high, and for /eː/ (see left panel in fig. 5.10, p. 104), the old and young listeners' response to the stimuli is most similar (r=.94). The least agreement on the other hand can be found in the mid column plots of page 104. For /oː/ (fig. 5.11) and /ɛi/ (fig. 5.12), the mid generation judged realization differences between more speaker pairings as 'different' than the younger or older speakers (both r=.55), whereas the response behavior of the old listener group matched that of the young group much better for /oː/ (r=.78) and /ɛi/ (r=.79).

Though our sample was rather small, we tested listener age as a predictor for the responses to each vowel phoneme separately in the logistic models. It contributed only significantly in the logistic models of /ɛi/ and /œy/. Yet, the models yielded no more than 66% correctly predicted responses; therefore the models do not yield an acceptable prediction of the response behavior. More data are needed to test the effect of listener age on the judgments of acoustic distances in vowel phoneme realizations, as except for /eː/, our data suggest that the listener's age group does play a role in the perception of sub-phonemic acoustic differences.

5.4 Summary

By means of a perception experiment, we tested whether listeners differentiate sub-phonemic acoustic vowel variants that had been found to significantly coincide with the background data of 70 speakers in the preceding acoustic chapter. By giving 'same' or 'different' responses, 30 listeners had to judge whether vowel phoneme realizations of various speaker pairings differed in phonetic quality or not.

Having analyzed the response behavior, it appeared that the significant acoustic distances described in the preceding acoustic chapter 4, and represented by the vowel realizations of the speakers B and D, were indeed differentiated by all listeners. Roughly speaking, the larger the acoustic distance between two stimuli, the higher the probability that the quality of the vowels was perceived to differ. Which acoustic distance measurement (onset or diphthongization in pc1) predicted the listeners' response behavior best, was phoneme dependent. Given the phoneme-dependent responses to acoustic speaker

distances in vowel realizations, the predictability of the response behavior had to be investigated separately for each phoneme. /eː/ was the only phoneme for which listener responses could be predicted by the stimulus distance in the pc1 onsets. As for the other vowel phonemes, /oː/, /ɛi/, and /œy/, even though the acoustic distances in pc1 onsets and degrees of diphthongization between the speakers added significantly to the predictability of the response behavior in the regression models, they could explain only little of the variance in the outcome.

Having checked the speaker-dependent attributes, for the remaining unexplained variance in the responses, listener-dependent effects on the response behavior were analyzed. When the listeners were split into the same age groups as applied to the speakers in the preceding acoustic chapter, some differences in the response behavior in dependence of the listeners' age group became apparent. Including the age level in regression models to predict the listeners' response behavior improved the predictability of the response behavior significantly. Yet, since the predictive power of the logistic models was still weak, we will rather talk of indications that the listeners' age had an effect on the acoustic discrimination behavior. Nevertheless, it is remarkable that effects of listener age on the responses coincided with some of the speaker behavior described in the previous acoustic chapter. There, the mid age group of the high educated speakers had been found to differ significantly in their vowel realizations from the old speaker group, whereas the old and young speaker group differed the least. Similarly, in our data, the poorest agreement was found between the response behavior of the mid and old age group, and the strongest agreement was found in the response behavior of the old and the young listeners.

Since we assume that listeners base their decisions on their daily perception of sub-phonemic acoustic categories, we presume that both the sub-phonemic realization of vowels (as described in the previous chapter) and the perception of sub-phonemic differences in vowel realization (as described in the present chapter) are socially tuned in the same way. Research showed that listeners can associate well-defined social patterns with pronunciation when asked to (see e.g. van Bezooijen, 1999 [152]). Yet, it is clear that more data are needed to prove the indications of listener age effects in the present perception experiment. Also, an equal amount of high versus low educated listeners should be included (90% of our listeners were high educated). Then it could be tested whether the socio-economic status of the speaker (high versus low educated) that significantly affected the realizations in the previous chapter, does not have the same effect in (high or low educated) listeners, as implied e.g. by Hay et al. (2006 [47]).

With (social) acoustic sub-phonemic differences not being perceived in the same way by all age groups, speakers of different age seem to internalize different sub-phonemic categories. In as far as (social) information is coded in acoustic variation it seemed to have been of different importance to the three listener age groups.

As a concluding remark on our perception experiment, the acoustic dimensions seemed

to carry different (social) information for each vowel phoneme, and probably for listeners of different age levels. How differences in phoneme realization between various (social) speaker groups emerge will be investigated in the following chapter. Literature on the finding that a human's articulatory production and auditory perception are interconnected will be discussed, and might explain findings of our experiment, and the results of our acoustic analysis. If the results indeed reflect some basic dependencies in human perception, effects of listener age on the perception of sub-phonemic categories should be replicable.

6. ON SPEECH VARIATION AND SOCIAL BEHAVIOUR

Abstract The previous chapters showed that the social background of speakers in terms of age and educational/occupational level had an effect on the sub-phonemic realization of the vowel phonemes /eː/, /oː/, /ɛi/, /œy/, and /ɑu/. From the small perception task that we described in chapter 5 we can infer that the listener's age had a comparable impact on the perceptual categorization of vowel variants. This chapter will offer a literature overview on how and why phonetic variation is socially intertwined. With the objective of defining its structure, and to explain origin and change in variation, linguistic approaches as well as processes studied in psychology will be considered. We will show that the effects found in literature on the articulatory-auditory interaction in human beings coincide with the effects found in our acoustic and perception data.

6.1 Introduction

Chapter 4 showed that socially structured variation can be found in fine-grained phonetic detail such as sub-phonemic differences in vowel realization. Our cohort of 70 speakers differing in social background, differed also in their pronunciation patterns of /eː/, /oː/, /ɛi/, /œy/, and /ɑu/. The speakers' level of education (or occupation) could be related to their vowel realizations. Generally, the low educated speakers showed higher onsets and less diphthongization than the high educated speakers.

Moreover, the speakers' vowel realization could be grouped according to the speakers' age group at the time of recording. Realization differences between the two educational levels varied characteristically between speaker generations: the largest difference between the two educational levels were found for the mid generation (aged 36 to 54 at the time of recording), which included the speakers who were born between 1945 and 1965. The pronunciation of the high educated changed remarkably from the old to younger speaker generations, whereas the pattern of the low educated hardly changed with the generations. Considering the role of age in listener behavior, the results of the perception experiment in chapter 5 coincide with the age effects that were found in the speakers' realizations in chapter 4. When split into the young, mid and old generations as the speakers in chapter 4, the listeners in chapter 5 differ in their perception of acoustic differences age-group dependently.

Together, the results of the acoustic speaker analysis and the perception experiment suggest that sub-phonemic vowel perception is affected by the social background as much as vowel production. Before interpreting the results of our present study, in the following we will try to explore what causes the variation in articulatory and perceptual behavior and why both dimensions seem to be connected. We will review relevant literature from linguistics and psychology on the topic.

6.2 The Structure of Variation and Change

To determine to what extent variation is perceived and can be imitated or accommodated, the role and the processing of phonetic and social features need to be considered. Though social research in the psychological-cognitive area that goes beyond the pragmatic-semantic level is comparably scarce and recent, (and often based on new techniques of brain-imaging, which might not yet be totally reliable in terms of mapping precision), the basic concept of social recognition and processing of inter-human contact will give further explanation to variation and changing speech behavior.

6.2.1 Sociolinguistic Approach

Starting with a study on the sociolinguistic structures and social tension in the community of Martha's Vineyard (Labov, 1963 [80]), foremost research on the connection of sound patterns and a speaker's social background or social awareness was carried out on American English by William Labov from the 1950's on. In his seminal study from 1966 on the pronunciation of the postvocalic /r/ in New York City, he found the production variants to correlate with socio-economic class (Labov, 1966 [81]). Labov's research resulted in a class stratification pattern, where a variant is used most frequently by the highest-status class and least frequently by the lowest-status class, and where alternatives of saying the same will have social significance.

With the findings on correlations of social factors and linguistic variants, the linguistic tradition to focus on competence (or internal language) in distinction to performance (or external language) became problematic (Milroy & Gordon, 2003 [100]). Rather than treating language structure as invariant and variation as asocial, the variationist approach is based on the assumption that language variation is intrinsic and structured. Behavior variables and their social embeddings are thereby seen as essential in understanding the dynamics of language change. By comparing the existence of variants and their relative frequency at different points in time, the quantitative research paradigm enabled linguists to propose social explanations for changing frequencies and emergence of varieties in time, space, and social space (Milroy & Milroy, 1997 [99]). Without data on usage and attitudes, and without interlocking the collected linguistic forms with ordinary verbal interaction, linguistic changes can hardly be explained (Labov, 1989 [83]).

According to Labov, research on change should focus on the following points: Firstly, find the continuous matrix of social and linguistic behaviour in which the linguistic change is embedded (embedding problem). Secondly, find the trigger of the linguistic change (actuation problem). And thirdly, find out if the change from below (below the level of conscious awareness) is dependent upon high status and will become a prestige model, or if it is dependent upon low status and will be stigmatized. There has been some debate about the role of prestige and other tacit generalizations, and by now it is assumed that the crucial indicators of language change are rather locally determined social categories (Milroy & Gordon, 2003 [100]). Since these have different meanings in different communities, various interactions need to be considered, which complicates any proposal of generalization.

6.2.2 On the Origin of Sound Change

The cultural and psychological forces that were found to structure variation and change are only accounts of the spread of a variant, the 'maxi-sound change' (Ohala, 1993 [110]). The origin of a sound change, i.e. the fine-grained phonetic detail which selectively becomes

spread or not, referred to as ‘mini-sound change’, cannot be explained by these factors.

Often, a principle of least effort has been put forward as motivator or mechanism for sound change. While ease of effort might fit other levels of speech (e.g. grammar), least effort in terms of articulation is hardly defensible when it comes to phonetic structures. The structures of the existing languages differ too much and are too diverse to support articulatory ease as the leading mechanism. Also, languages with similar structures fail to show similar sound changes under comparable conditions. From a sociolinguistic point of view, the principle of least effort cannot explain the underlying process of sound change, since it regularly appeared that speakers of advanced social positions within a local community who use language effectively and vigorously, are the innovators of sound changes (Labov, 1980 [82]).

Where previously the speaker and his striving for ease or intelligibility was seen as the origin of a sound change, to Ohala, it is the listener rather than the speaker who is assigned the leading role in the emergence of a new phonetic variant. He assumes that for the sake of communication, the interlocutors will pronounce and use words the way they (think they have) heard them (Ohala, 1981 [109]). Assimilatory and dissimilatory sound changes are due to misperception, and listeners develop new forms as a cause of failure in normalizing or correcting perceived speech variations. Inherent to his approach is also that sound changes are phonetically abrupt, however, he assumes that the phonetic changes might be easier to detect by outsiders than by speakers of the affected speaker community (Ohala, 1993 [110]).

If indeed based on misperception, the (abrupt) ‘mini-sound change’ should happen within phoneme classes, since the sound actually produced is misperceived in such a way that the listener is not aware of the misperception and his following ‘misproduction’ (in terms of deviating from the norm or mean): With multiple sources of information, including phonetic examples of various speakers, as well as knowledge of spelling or grammar, perception errors would be discovered. Thus, an ‘accepted’ misperception can only appear regularly, if the misperception and the new production do not break phonological rules.

A model of sound change based on inappropriate normalization or correction, however, is hardly able to explain more complex sound variation or change, i.e. the phenomenon of speech convergence during conversations, or chain shifts. The assumption that "speaker and hearer are interested in communicating and will pronounce words only as they have heard them (or think they have heard them) pronounced by others" (Ohala, 1981, p.197 [109]), could also hold for arising variants without misperception: As the perception experiences (input) differ for each listener, pronunciations following the individual input information will do, too. Instead of ‘misperception’, the speaker variation in pronunciation could simply be due to the individual auditory input. With strong connections between hearing and articulation no abnormal processes need to be included to explain variation.

The strong connections between the auditory and articulatory system have been sup-

ported by various findings, ranging from speech experiments with delayed feedback, and pseudo-word repetition tasks to neuro-anatomy. With respect to the latter, Hickok & Poeppel (2000 [50]) describe an interfacing network between auditory and articulatory representations of speech, where a sound-based representation is linked to a motor-articulatory system, as well as to an auditory-motor interface, and the auditory-conceptual interface. The motor-articulatory system is furthermore directly connected with the auditory-motor interface.[1] The network might establish when the child tunes his articulatory productions to the sounds of the target language. Next to playing a key-role during this critical period, it forms the basis of the phonological working memory in adults, providing access to sub-lexical speech segments (Hickok & Poeppel, 2000 [50]). The network's continuing activity, beyond the so-called 'critical period' of language development, can account for the phonetic tuning and therewith for changes in the productions of adults. Examples of tuning activity in adults are various, reflected by e.g. the phenomenon of (temporary) speech convergence during dialogs. In Pardo (2006 [112]), phonetic change and vowel variation could be linked to social interaction patterns of the interlocutors. A less temporary example are the gestural drifts that were found for the productions of bilingual speakers after a long stay in either of the two countries where one of their native languages is spoken (Sancier & Fowler, 1997 [130]). Furthermore, a longitudinal study over 50 years on the British Queen's realizations of vowels during her broadcast annual Christmas messages revealed a considerable shift over time (Harrington, 2006 [44]).

As mentioned at the beginning of this chapter, many changes in the phonetic repertoire are found to have social significance. Moreover, sound changes usually spread from groups of speakers with advanced social positions (Labov, 1980 [82]). Since there is no plausible physical explanation (i.e. misperception or individual perception) why some communities show salient changes over time, whereas others hardly do (excluding the unequal existence of hearing impairments), the reason for the different sound dynamics will be due to the unequal characteristics of the speaker groups. Previously it was concluded that the 'mini-sound change' will hardly be perceptible to speakers inside the affected group, and so a conscious adaptation of a pronunciation variant within a social group is rather unlikely. In the following section, we will consider to what extent the social structure and connections of a group can carry, adapt, or slow down innovations, and how findings from social psychology explain the impact of social relations on speech realizations.

6.2.3 Social Relations, Identity and Social Cognition

A speaker's social network is often described as a web of strong and weak ties, which interpersonal relations are defined in terms of strength, structure and density (Milroy &

[1] This dual-route sensory-motor interface is supported by neural research on the auditory switching from non-speech to speech modes (Dehaene-Lambertz et al., 2005 [25]).

Gordon, 2003 [100]). Weak interconnections within a network are seen as being more sensitive to external influences, and hence favourable to changes. Conversly, a linguistic system might be successfully supported within a network that consists of dense and multiplex interconnections.

According to theories from social psychology, intra-group differences are minimized within the social network, whereas inter-group differences are maximized (Tjafel, 1982 [146]). This accentuation of differences protects the group's value system and helps maintaining or enhancing it. Next to the value function, the accentuation of differences and the forming of stereotypes has a cognitive function: With the utilization of category membership, the complex network of social groups the individual has to deal with can be simplified.

Similar to the 'community of practice' within a social theory of learning (see Lave & Wenger, 1991 [87]), Eckert (1999) related linguistic variation to social practice and defined such communities of practice within the structure of school, where collections of people meet through common endeavours as "common goals, dreams, desires, jobs, necessities, and/ or problems, finding joint responses and strategies for dealing" (Eckert, 1999, p.40 [32]). Entering these multiple communities of practice, each scholar finds a personal path in juggling the benefits of the various communities. This principle of personal development or social conformation can easily be mapped on situations outside school.

Another theory originating from the area of social learning that explains the impact of social relations on speech is the 'Social Cognitive Theory'. It holds that environment, behavior and cognitive factors are interacting in a reciprocal relationship, and thereby are causing each other (Bandura, 1989 [5]). In being selective in their environments, people can get control over the happenings, and the social support helps in managing daily life. In anthropology studies the possibility is discussed that language evolved primarily to subserve social behavior (Adolphs, 2003 [3]). These theories recall the work of Vygotsky, who was one of the first researchers from the language field to emphasize the role of social learning. Earlier, speech was seen as the expression of thoughts, the latter being an inner process. Vygotsky argued that social interaction precedes development: Using tools such as speech to mediate the social environment, consciousness and cognition are formed through this socialization process and social behavior (Vygotsky, 1986 [162]). In a recent article, early speech learning is tied to social factors (Kuhl, 2007 [78]): In natural linguistic settings meaningful social cues like referential information (e.g. objects of reference or eye gazes) cause significantly higher attention and arousal, as well as overall increases in remembered and coded speech quality and quantity, both in perception and production. Social interaction, or even the simple presence of a human being (as opposed to a virtual human being on the tv-screen) significantly affects early speech learning. As inherent features of natural social settings, contingency and interactivity seem to be key components of speech learning. Furthermore, findings from studies with children with

autism spectrum disorder couple social deficits with early language disabilities (Kuhl et al., 2005 [79]).

Social interaction is not only a major factor during the acquisition of speech. Investigations on the 'Chameleon' effect, the non-conscious mimicry of various aspects of one's interaction partner, show, that mere perception triggers mimicry, also in adults (Chartrand & Bargh, 1999 [14]). Mimicry seems to smoothen and increase the linking between the interaction partners. For interaction partners who are well-disposed towards each other, the effect is found to be even greater. A similar effect is found in speech communication, where interlocutors converge their speech patterns during conversations (Vallabha & Tuller, 2004 [149], Pardo, 2006 [112], Magnus & Nusbaum, 2007 [93], Delvaux & Soquet, 2007 [27]). Pickering and Garrod (2004 [116]) assume that a largely automatic process causes the interlocutors' linguistic representations to become aligned at many levels in dialogue. In their model of 'interactive alignment account', the channels are bidirectional, and are assumed to be similar to the perception-behavior link that plays a central role in imitiation according to Chartrand & Bargh (1999 [14]).

Considering the reciprocal relationship of social interaction, today, two ways of research are persued: the representation of other minds, and the experiencing of other states of mind. The 'Theory of Mind' is used as a general term for research that investigates how we reason about others' mental states, mediated by our own social rules and norms, hence on the basis of our own theories of minds (Lieberman, 2007 [90]). The Theory of Mind as domain-specific format in terms of its claimed independence from general intellectual capabilities, or as a purely theoretical model, has been challenged by an alternative approach: Experiencing others' states of mind is associated with empathy and internally-focused processes, rather than interpretation based on theoretical concepts. The suggestion of a reciprocal relationship of environment, behaviour and cognition is supported by results from social and cognitive psychology. Most observations suggest that the performance of social actions and the processing of social stimuli are cognitively not different from the performance of other actions, or the processing of other stimuli (Hommel, 2006 [53]). Human behavior in general seems to be constantly affected and even conditioned by social interaction. Neural processes have been revealed, which mediate perception and the planning of action: Studies on the neuron system using brain-imaging techniques[2] suggest a neural mechanism that mediates own self-experienced multilevel knowledge and the implicit certainties we hold about others. Research on neural links between oneself and others, the *mirror*-neurons, display their role in social understanding and imitation (cf. Kuhl, 2007 [78]): During the third-person experience of action or emotion, the same structures are active as in self-experience. Besides a cognitive interpretation of what is perceived when for example recognizing emotions, experiential knowledge is generated

[2] such as C(A)Tscan (Computed (Axial) Tomography), PET (Positron Emission Tomography), fMRI (functional Magnetic Resonance Imaging), EEG (Electroencephologram), and ERPs (Event-related Potentials).

by a functional mechanism (Gallese et al., 2004 [38]): Observed social stimuli are mapped directly onto motor neural structures, that were generated by the experiential knowledge of the concerning social stimulus. Thus, in dichotomy with the sensory description of the observed stimuli, also the associated internal representation of the state which is evoked during self-experience of similar states is activated. In other words, the structures that are normally involved in personal experience take a part in how we perceive and understand the behaviour or states of third persons. The impact of social relations on speech is reflected by findings concerning Broca's area, a brain area involved in both speech and the adult's mirror system (Rizzolatti & Craighero, 2004 [128], Pulvermüller, 2005 [126]).

The implications of socially constrained variation and its effect on speech perception and language acquisition led to adjustments of several (phonological) models. Featural models of speech processing that directly map phonetic features to lexical representations (with the possibility of a post-lexical phonemic level) account for the representation and processing of speech variation. Not only socially constrained variation seems to be part of the mental representations. Following e.g. McMurray et al. (2002 [95]), fine-grained subphonemic differences are preserved and of use for higher levels of speech processing. The information might be important for the perceptual system, where the fine-grained acoustic/phonetic information is correlated with information of its phonetic environment. Other studies support abstract underspecified representations of phonological features in the mental lexicon (Eulitz, 2007 [35]). Based on both findings, the newest models of speech processing are hybrid, assuming a coexistence of abstract and episodic representations, with one type of representation being dominant, dependent on individual experience: During the perception of speech, at the same time that knowledge of linguistic meaning is added, detailed episodic memory traces are created of the words that were spoken. Suggesting that accumulated episodic traces represent the mental lexicon, not all phonemic features are stored in the mental lexicon, and top-down processes influence language-specifically the perception of phonetic contrast. Following this, a complementary system is suggested, consisting of both episodic and abstract perceptions and memories that work combined (Goldinger, 2007 [42]), and, depending on factors, with one of the representations being dominant.

By stating that detailed episodic memory traces are created of perceived words, which in accumulation form the mental lexicon, the hybrid model would account for an automatic and unconsciously acquired pronunciation pattern, and for a certain degree of flexibility and changes in phonetic realizations. The latter is needed to explain speech accommodation over time. It would also explain the increasing influence of phonological knowledge on the categorization of speech sounds during language development. The different weighing possibilities of the complementary system can account for the experience-based differences found in L1 and L2 learners (see Cutler & Wagner, 2007 [23]). Furthermore, the assumption that the patterns used in production are more or less dependent on perceived

patterns, goes together very well with the previous results from neuroanatomy considering a dual-route sensory-motor interface (see Hickok & Poeppel, 2000 [50]). However, though this hybrid model seems promising, more detail and challenges have yet to be investigated.

6.2.4 Summary

Theories holding that the speakers' representations are based or conditioned on the (frequencies of occurrence of) input, are supported by neuro-anatomical findings: With linked auditory- and articulatory systems, third-person experiences being mapped on the same areas as own experience, and brain areas that are involved in both speech and the mirror system, there is a clear interdependence of the speech input, social relations, and one's own behavior.

With speakers usually being unaware of using a particular pronunciation, it is not surprising that the acquisition process of pronunciation patterns is largely automatic. In view of code switching, as well as the stigmatizing of interlocutors, which leads to less convergence in speech, a volitional process seems to have an influence on the pattern adaptation process. Considering the processes that mediate perception and one's own behavior, stigmatization could as well be partly automatic; due to no or little contact with the rejected group, there will be no or little ability to mirror its behavior. Research must show, whether for example speakers who dislike their interlocutors, converge speech segments anyway when communicating over a long period of time. Nonetheless, the processes included in selective pattern accommodation are heavily based on social relations: Convergence goes with social understanding and the will to communicate in an optimal way, whereas non-convergence goes with social distance or reluctance to communicate. For an explanation of speech variation and change, an analysis of the embedding of the speaker within each community and the attributes of these social networks, as well as with whom the speaker identifies the most, is indispensable.

The temporary adjustment found in the speech of interlocutors in the cause of their dialogue, and the long-term speech adjustment such as for example in the longitudinal study of Queen Elizabeth's vowels are connected but have to be considered separately. (Phonetic) speech convergence during dialogues in general can be explained with the temporary storage of (acoustic) information, which, in being activated, will be the information that is primarily accessed. Following the general assumptions on memory, the temporarily converged speech patterns will have an effect on long-term memory if they are regularly reinforced. To change a speaker's speech pattern in the long run, first, the speaker's input will have to undergo a long-term change. These long-term effects of (temporary) speech accommodation have been reported with reference to the tuning of bilinguals, before and after they spent a considerable time in either of their two mother-tongue countries (Sancier & Fowler, 1997 [130]).

The fact that more or less stable interacting communities share a certain pronunciation pattern, and that convergence that arises during a conversation disappears later on, indicates that frequency of occurrence (probably in various communication partners) plays a major role in the longer adaptation and maintenance of pronunciation patterns. As speakers will choose to spend most of their time in communities they identify with, their speech pattern is very likely to have lineaments of the very community. Since there is no clear ending or critical period in the flexibility or adaptation of phonetic speech patterns, a stable pronunciation pattern will be more or less the result of a stable environment. To what extent these findings help interpreting the variation in the present data will be discussed in the following section.

6.3 Interpretation of the Results of the Present Study

In chapter 4, except for the oldest generation of speakers, the variation found in the acoustic analysis of the vowel realizations of 70 speakers was most significantly affected by the speakers' either high or low educational and occupational level (compare figures 4.22 and 4.23, p. 79).

Following the first sections of the present chapter, the two educational groups (high and low) can be seen as two different speech communities. Since the acoustic categories have been built individually to facilitate conversation, the social communities in the given data seem to be based on interactions and strong ties between more highly educated/occupied speakers, and on the other hand between the low educated/occupied speakers, with less strong contact between speakers of different educational levels. In the oldest generation, however, educational groups were not apparent in the acoustics, and hence, the contact between speakers of both levels might have been stronger and speech communities less differentiable than within the age generations thereafter.

In general, we assume that the results of our acoustic analysis, based on data all taken at the same given moment in time, reflect long-term speech patterns. However, considering the results of chapter 4, we assume that the speakers who belong to the middle and old age group are more settled in terms of their social communities than the speakers of the youngest age group. Thus, were all speakers measured again for a longitudinal research, we would suggest that our youngest speakers are the most probable age group to show a change in pronunciation.

The effect of age was of different relevance when the higher and low educated groups were surveyed separately. Unlike the high educated speaker group of chapter 4, the low educated speaker group hardly showed changes in its pronunciation pattern over time (compare the plots on the right-hand side of figure 4.25, page 82, with the plots on the left-hand side). For the low educated there was only a small linear effect of age and changes in the pronunciation of /œy/ and /ɑu/, which was significant only for the diphthongization

of /ɑu/ (section 4.4.4, p. 75).

Various suggestions are imaginable to explain the structure and spread of the pronunciation patterns within our limited data. Sociological research will have to show to what extent politics and economy or other factors affected the social structures in the society during the last decades, and thereby the social groupings and their pronunciation patterns. Then we could conclude if the pronunciation patterns of our sample of speakers indeed reflect the networks within Dutch society, and their changes over time.

The stable pronunciation pattern within the low educated group would for example suggest dense interpersonal relations, little sensitivity towards external influences, and hence the repression of changes over time. In view of the shared pronunciation pattern, there should have been pronounced ties between speakers of different age groups. Contrarily, within the high educated, the age differences in pronunciation might point to weak ties across age groups, suggesting less steady contact over time between speakers of consecutive generations. This would suggest the allowance of new contacts, thereby external influences, and hence changes in pronunciation.

If this is the case, we could hypothesize that the population of the high educated fluctuated much more than the population of the low educated. One could speculate that students who enroll for higher education might for example have come from various parts of the country, as well as from abroad, bringing various pronunciation patterns into the more highly educated/occupied community. With the origin of the more highly educated speakers being a rather unstable factor, the proportions of various vowel variants, and thus the pronunciation pattern of the speech community, would have been in a permanent state of flux. By contrast, lower education (and occupation) might have been available regionally, and the probability that students of various parts of the country mixed during lower education would have been comparatively small. Then, the low educated would have been much less affected by external influences and new pronunciation patterns than the high educated.

The finding of some almost significant regional traces in the pronunciation patterns of the low educated, as described in section 4.4.6, p. 84, would also suggest that the low educated speakers are a less firm speech community as a whole, but consist of several regional sub-communities. When this argumentation is transferred to the high educated speech group, where no significant regional traces were found but significant differences between the age groups, it would implicate sub-communities according to age. So, within the low educated group social sub-groups would have been structured by the factor ‘region’, whereas for the high educated, the factor ‘age’ would have been liable. However, more data and sociological research is needed to show which structures (and which changes of structures) of the Dutch society are indeed mirrored in our pronunciation data.

Within our high educated speakers (see table 4.9, p. 83), the pronunciations of the mid age group differed considerably from the pronunciations of the old age group. The pronunciation behavior of the youngest generation approached the pattern of the oldest age

group again though not differing significantly from the mid age group's pattern. Whereas the young age group might socialize with the old and the mid age group, in view of the outstanding of the high educated mid age group (36-54 years of age), thus the high educated speakers born between 1945 and 1965, we could hypothesize that these speakers were more dissociated from the previous generation than the young age group. Then, given the pronunciation differences in our data, similarly, they would have been dissociated from speakers of lower educational or occupational level. According to Stroop (1998 [140]), the 'poldermodel' in the Netherlands (see section 1.1, p. 3) triggered this behavior of separation, especially in women who profited by the growing equality. Whether the distinctiveness of this high educated speaker group has its roots in the 'poldermodel', or other societal movements, such as for example the well-known '68ers'-movement, is difficult to disentangle, and more complex investigations are needed. If we follow Stroop's argumentation of the 'poldermodel' as a supporter of women's emancipation, though in our data there were no gender effects or female precursors apparent, the women's new social strength in the early seventies might as well be reflected by the fact that they show the same behavior as male speakers. Nonetheless, in the generation before (the old generation) we found no gender effects either, nor in the generation thereafter (the young age group), and so we would suggest that the females' social strength could already have grown earlier, presumably within the war-generation, reflected by those raised before 1945. However, it could as well be the case that the females' vowel behavior has never differed significantly from the males'.

When considering our listening experiment, 90% of the listeners turned out to be high educated/occupied, and the effect of level of education thus could not be analyzed. The factor 'age group' that had been significant for the high educated speakers in the acoustic analysis had a (significant) influence on the listeners' categorization of some of the data as well. Despite the inclusion of rather few listeners per age group, the mid aged listeners judged some stimulus pairs significantly different from speakers of other age groups. The mid generation thus differed remarkably from the other generations, in acoustic terms as well as in the perception of these acoustics. The findings of the preceding sections on the auditory-acoustic linkage offer an explanation for our age-dependent speaker- and (presumably) listener behavior.

The acoustic distinctiveness of the mid age speakers' productions and the perceptual distinctiveness of the mid age listeners' behavior suggests that this age group was shaped by an acoustic input that differed considerably from the input that shaped the old or young age group. It could have been the case that social information got attached to acoustic variants in the time the mid aged group's speech was tuned. In view of the other age groups, we would then hypothesize that this social information had been irrelevant in the time before, and, given the smaller distinctiveness of the young and the old age group, and given as well the smaller distinctiveness of the young and mid age group, this social information

presumably became less important again in the tuning of the young generation.

Experiments on the perception of non-native speech sounds already showed that phonetic boundaries differ according to the listener's language background (see e.g. Ingram & Park, 2002 [55]). The perception of sub-phonemic boundaries of listeners with the same language background has not yet been a popular study objective. An investigation by van Bezooijen (2001 [151]) on what attributes younger females (aged 18-29) versus older females (aged 38-58) associate with speech samples of speakers of 'ABN', 'Polder Dutch' (see chapter 1), and two dialects, has some interesting results considering the effect of the listeners' age. To the younger, the difference between the two categories is less clear than to the older females. The younger judged ABN and Polder Dutch both as being 'normal', but ABN as less modern. The older females judged Polder Dutch as less normal and less cultivated. So, to the younger, the difference between both variants is less salient than to the older females. This suggests that the younger females are used to both variants equally, and that the only difference they perceive is the period in time that they associate with the variants. The older females, however, associate an additional factor, namely 'less cultivated' with the new variant, and thus a social attribute. In another study of van Bezooijen et al. younger and older males were included in the perception task, and the results showed that the listener generations differed in their answers towards the normalcy of the Polder Dutch variant independent of sex; contrary to the young, the old females and males judged it less normal than the Standard Dutch (ABN) variant (van Bezooijen et al., 2001 [153]). The categories of van Bezooijen's 'younger' versus 'older' listeners roughly correspond to the categories of the young versus the mid age group of our present studies. In our perception experiment, accordingly to the results of van Bezooijen, the mid generation showed categorization patterns that differed from the younger generation, suggesting that the generations are not equally sensitive to sub-phonemic differences. Though van Bezooijen's task was very different from our perception task, both results indicate the same: attributes attached to sub-phonemic variants and their perception vary at different points in time.

An interesting experiment underlines the finding that the listeners' ability to judge and categorize sub-phonemic speech sounds is influenced by the phoneme categories they experienced themselves. Recently, the effect of social information on the speech processing of merging diphthongs in New Zealand was investigated (Hay et al., 2006 [47]). In New Zealand English, the diphthongs of <near> and <square> words are merging, with the diphthong [eə] moving towards [iə]. Within the realizations of the younger, for example <air> and <ear> became homophonic [iə], and <chair> and <cheer> became [tʃiə]. Participants in this perception experiment were presented speech stimuli with and without varying (visual) social information on the speakers. When social speaker attributes were available to the listeners, the perceptually favored word of two words with merging pronunciations turned out to be biased. The results showed that – next to the influence of context- and word-specific characteristics – the accuracy in the perception task was not

only influenced by (visually) perceived speaker characteristics, but also by participant-specific factors. Though the authors note that the results are quite complex, apparently, the speaker's speech, including the merging vowel realizations, was processed not only dependent on the perceived social speaker characteristics, and participants with distinct representations of the merging vowels evoked speaker-age specific vowel distributions. To these 'unmerged' participants, perceiving a younger speaker activated less distinct vowel distributions, whereas perceiving an older speaker evoked more distinct distributions. Contrary, participants with merged representations appeared to activate less age-dependent distributions (Hay et al., 2006 [47]). Additionally, no effect of listener education on the perceptual categorization were found in contrast to the effect of listener age. Following the previous argumentation, this would imply that the different age groups were surrounded by differently equipped social-acoustic communities.

In general, where a certain sub-phonemic range of variation is socially unimportant to one generation, variation within this range can be crucial and socially structured in another generation. Given our acoustic and perception analysis and the findings in literature, we can conclude that in the same way the acoustic structure of sub-phonemic vowel realizations in Dutch differs during the course of time due to varying social constellations and influences, the sensitivity towards sub-phonemic variation differs between age groups due to own experienced social attributes.

However, interpreting the measured outcome of our speaker data remains difficult. Given the various speaker groupings, a generalization of the effects found in our sample of speakers is delimited, and many more (social) factors in addition to the ones we considered might have played a role in the distinctiveness of sub-phonemic pronunciation patterns. More research and still broader analysis is needed to identify all factors that affect the sub-phonemic tuning of production and perception. The strength of our corpus was seen in its objectivity towards the appearance of the Polder Dutch phenomenon. Now it would be interesting to see whether the same background effects that were related to the acoustic behavior of our speakers will be found in larger corpora that are specifically designed to investigate these or other social background effects in sub-phonemic speech production (with e.g. equal spreads in speaker background data and recording situation). Correspondingly, a perception experiment should be performed on the distinctiveness of the sub-phonemic variation with larger and more diverse listener groups.

7. GENERAL SUMMARY, LIMITATIONS AND PROSPECTS

Abstract This final chapter summarizes the main findings of our research on vowel variation in Standard Dutch. The hypotheses as given in the introduction are reconsidered in view of the main findings of the present research, and the limitations. Finally, suggestions for future research are given.

7.1 Hypotheses Reconsidered

In the following we will reconsider the hypotheses of the first chapter (cf. section 1.2).

The general hypothesis of the present research was that in Standard Dutch *"... the realizations of vowel phonemes show sub-phonemic variation that is socially marked."* The acoustic results of chapter 4 yielded patterns of vowel realization that coincided with the social background of the speakers in terms of 'level of education' and 'age group'. Our general hypothesis on the social structure in sub-phonemic vowel realization was thus supported. In view of our significant results, and unlike previous analyses of larger Dutch vowel corpora, future analyses of Dutch vowels should control the speakers' social background to minimize unwanted variation in the data, and to allow for a better interpretation of the measured acoustics, even more when speaker data are pooled.

The second hypothesis in section 1.2 focused on gender effects in terms of female precursors within the avant-garde speakers: *"While the well-educated (the avant-garde) have lowered /ɛi/, led by the females, the phenomenon is not apparent in other speakers."* In chapter 4, the results showed that, starting with the mid age generation, /ɛi/ was significantly lowered, longer, and more strongly diphthongized within the group of high educated speakers. Our high or low social classes were defined in terms of 'level of education and occupation', and we did not define avant-garde speakers versus the non-avant-garde. This was due to limitations in the available speaker data and the background attributes gathered within the spontaneous speech part of the CGN. However, we presume that avant-garde speakers are part of the speaker group of high educated and occupied speakers, and that the speaker group that was labeled 'low educated and occupied' does not include speakers of the avant-garde. From this angle, the hypothesis can be supported, though, in view of our results, we would rather name the appearance of the lowered and more strongly diphthongized variant 'higher Dutch' (the Dutch of the high educated and high occupied) than 'avant-garde Dutch' or 'Polder Dutch'. When it comes to the leading role of the females, however, we cannot support the hypothesis, as in our corpus of 35 males and 35 females there were no effects of gender. Due to the latter findings, we have to reject the hypothesis of women leading the change.

Yet, as mentioned in chapter 4, the hypothesis was based on earlier studies of the pronunciation of /ɛi/ which were based on formant values. Though including a logarithmic scale to prevent the unwanted effect of speaker sex contrary to gender, the previously applied normalization procedures probably still carried effects of speaker sex in the formant values which then got entangled in the research results, which led us to our third hypothesis: *"Vowel space sizes differ, and gender differences might be caused by anatomical differences between the sexes. When comparing realizations of various speakers and sexes, a speaker's realized vowel quality needs to be defined in relation to the size of his or her individual vowel space."* In chapter 4, the acoustic analysis of all speakers' /a/-/i/-/u/

vowel triangle spaces yielded significant effects of speaker sex when the acoustics were measured in terms of formants in Bark. For the triangle space size in pc's based on a PCA on barkfilter output, the differences between the sexes were not significant and speakers could be pooled. For formants, our data support the first part of the hypothesis considering the disentanglement of gender and sex effects in vowel variation research. This underlines the second part of the hypothesis; the need for a definition of vowel quality in relation to the speakers' individual vowel space size, which was indicated by the results of the preliminary study in chapter 3. A sophisticated procedure to compare vowel data of various speakers and independent of speaker-sex was developed in chapter 4. As shown in chapter 4, the sex-specific vowel space attributes could be normalized when our normalization procedure was applied to the outcome of pc's derived from a PCA on barkfiltered /a/, /i/, and /u/ spectra. In contrast, sex-specific vowel space attributes could not be normalized when the normalization procedure was applied to formant measurements in Bark.

Our next, more methodologically oriented hypothesis stated the following: *"Principal component analysis on barkfiltered spectra are a more objective method of measurement in vowel variation research than formant analysis."* Given the results of the PCA on all speakers' barkfiltered /a/, /i/, /u/ mean spectra, and the lack of significant sex effects in the analysis of the resulting various speakers' vowel space sizes in the pc dimensions, we assume that this method is reliable in vowel variation research. Furthermore, and contrary to formant analysis, this method can be reliably automated and needs no hand correction. The effect of noise in the vowel space sizes could be normalized by relating all vowels to the speaker-specific /a/-/i/ values. Contrary to our pc's, the vowel space size in formants in Bark did yield significant sex differences. Since we were analyzing degrees of lowering and diphthongization under aspects of social pronunciation constructs, effects of speaker sex need to be disentangled from effects of gender in this variation research. We thus assume that principal components on barkfiltered spectra are more objective and reliable in vowel variation research.

Having proven the socially marked diphthongization and lowering of /ɛi/, we now consider the last hypothesis stated in the introduction that deals with the interdependence of the Dutch diphthongs next to /ɛi/, and the long vowels: *"The long vowels and diphthongs of Dutch vary interdependently. If the pronunciation of /ɛi/ is changing, the diphthongs /œy/ and /ɑu/, and the long vowels /eː/, /øː/, and /oː/ are, too."* Due to its low frequency of occurrence we omitted the vowel phoneme /øː/ from our analysis. The results of chapter 4 indicate indeed a lowering and diphthongization pattern. This pattern was found for all of the vowels mentioned, and not merely for /ɛi/. Moreover, the lowering and stronger diphthongization were most apparent in the phonemes /eː/ and /oː/. The limited amount of data do not allow us to determine which of the vowel phonemes was first in the process of lowering and stronger diphthongization, but we speculated that /ɔu/ and /œy/ were the first to have moved. In general the outcome of this study suggests that, unlike previous Dutch

vowel analyses in larger corpora, future vowel analyses should control the social speaker background before generalizing pronunciation patterns.

7.2 Limitations and Future Prospects

Though most parts of our initial hypotheses could be clarified or even supported by our research, the results of our corpus analysis can only be generalized under reserve. Given the number of attributes in the 70 speakers' meta data that could be related to the realization behavior, there is a need for the analysis of even larger corpora with a more even spread of these attributes.

Although we are satisfied with the reliability of our pc's in our variation research compared to the analysis by formants, the fact that the pc's were sensitive to background noise, whereas the formant values were affected by sex differences, would suggest the benefit when both could be combined. Vowel variation analysis by a PCA on barkfilters could be improved by a combination with e.g. the stronger weighing of spectral peaks in the barkfiltered spectra, reducing the effects of noise that were apparent in the principal components.

Our study should represent diachronic changes in pronunciation from generation to generation. However, only a longitudinal study could confirm that the effects found in our apparent-time study are not the effects of age grading. In the summary of chapter 6 we speculated that our young speaker group is probably the least settled generation of our speakers, and therewith they are the speakers who are most likely to show changes in pronunciation, were they measured again at a later point in time. Longitudinal studies on pronunciation are lacking for Dutch, and measuring speakers and reporting on their social background at different points in their lives would help to disentangle the various background effects.

Considering social background effects we would also suggest more sophisticated investigations on the listeners' backgrounds. Although the interpretability of our perception experiment is limited, given the literature on social behavior and perception, we nonetheless assume that, parallel to the social effects found in sub-phonemic vowel realization, comparable effects can be found in sub-phonemic vowel perception. This is due to the individual acoustic input a speaker is confronted with, and which tunes his or her perception of sub-phonemic (social) variation and the production behavior as well. Referring to the strong ties between perception and production, we could have suggested another hypothesis: *"If the speakers' productions can be related to their social background, and thus to different speech communities, their perception might show traces of these effects as well."* We would thus claim that not only a speaker's production but also his or her perception is socially marked. The results of our perception experiment suggest that both are affected or formed by social behavior in the same way. Chapter 6 underlined the de-

pendence of acoustics and articulation and the large impact of social behavior on both. However, though it showed some effects of listener age in the same-different judgments of pairs of sub-phonemic vowel realizations, our perception experiment in chapter 5 was originally set up to prove the perceptibility of the acoustic variation as found in chapter 4. Given the indications of participant-related effects on the perception of acoustic distances, the listeners should have been chosen more carefully with an even spread of attributes such as the listener's age that turned out to have affected the judgments. Only then the suggested hypothesis could have been tested properly. Next to controlling the listener background, we would have chosen other stimuli. Major difficulties in mapping articulation, acoustics and the perception of vowels arise from the circumstance that different sounds or articulations can evoke the same perception, and thus equidistant articulatory or acoustic steps do not lead to equidistant perceptual categories (see for instance Peterson & Barney, 1952 [115], Miller, 1953 [98], Traunmüller, 1981 [147], Assmann et al., 1982[4], Miller, 1989 [97], Hoemke & Diehl, 1994 [51], Stevens, 1996 [136]). Restricting ourselves to the spontaneous speech data that have been analyzed in our study of 70 speakers made the choice of stimuli and controlling various stimulus attributes difficult. Including psychoacoustic, syntactic, semantic, lexical and pragmatic effects, human speech processing is very complex, and due to the interaction of the various layers of speech, the relation between perceived features and the acoustic signal is not biunique. Many more aspects of the chosen stimuli probably affected the listeners judgments than the ones we could consider, and synthesized stimuli might help in controlling these aspects.

To test the social effects in listeners we would suggest to use stimuli with synthesized vowel qualities from various parts of the vowel continuum, representing the sub-phonemic vowel variation. A perception task including synthetic stimuli with variously diphthongized and lowered vowel phonemes for comparison could clarify to what extent listeners can be differentiated in their judgment of acoustic differences, and whether the differences can be related to social attributes of the listener background. Were the vowel realizations of the same participants' acoustically analyzed as a function of their social background, a direct link could be established between effects of social background on production and perception behavior.

ENGLISH SUMMARY

Speech is most commonly and naturally used as an interaction medium in social settings. Along with communicating meaning, the speech signal is a product of physical properties and changes, as well as of generally all factors that form the identity of the speaker, such as social affiliation or family origin. The choice of words but also the way they are realized differs from speaker to speaker, and also within a speaker. Various observations of the lowering of the diphthong /ɛi/ (Polder Dutch) led to the start of this project.

In this study, the phonetic variation in the realizations of the Dutch vowel phonemes /ɛi/, /ɑu/, /œy/, /e:/, and /o:/ (as in words like <tijd>, <kous>, <huis>, <zeep> and <boot>) is analyzed in a representative sample of Dutch speakers taken from the Corpus Gesproken Nederlands (CGN). The aim was to find out whether the distribution of sub-phonemic pronunciation variants coincides with attributes of the speakers' background, and whether it changed over time as a function of age. To discover socio-phonetic variation and change, we investigated the apparent distribution of pronunciation variants in the spontaneous speech of 70 speakers, 35 females and 35 males, of different ages and with different socio-economic backgrounds. Presumably, the speakers' socio-economic affiliations go together with diverse speech patterns, and hence, pronunciation variants can be classified according to the speakers' background data.

In addition to the acoustic variation that we were looking for, there is acoustic variation between speakers that is caused by biological attributes, such as the difference between the vocal tracts of females and males. To be able to compare vowel qualities across speakers and sexes we needed an efficient and reliable method that minimizes unwanted variation but keeps the linguistic variation. Two different methods were compared to measure the vowel quality acoustically in our sample of speakers: formant analysis and principal component analysis (PCA) on spectral bandfilters. Differences in vowel quality between the speakers could be captured successfully by the PCA dimensions, and thus this method was used for all vowel analyses. Physiological differences (such as speaker sex) were further factored out by relating vowel differences speaker-individually to the point vowels /a/, /i/, and /u/.

When related to each speakers' individual /a/, /i/, and /u/ vowels, the realizations of the diphthongs /ɛi/, /ɑu/, /œy/, and the long diphthongized vowels /e:/, and /o:/ revealed significant differences between socio-economic groups and ages in terms of vowel onset

and degree of diphthongization. Given our analysis we found no significant differences between the vowel phoneme realizations of females and males. Speakers with a higher level of education and occupation showed lower onsets and stronger degrees of diphthongization. Contrary to speakers with an assigned lower socio-economic status, we also found remarkable changes in the vowel pronunciation patterns between speaker generations with an assigned higher socio-economic status.

A perception experiment was run to verify the perceptibility of these acoustic differences. 30 listeners had to judge whether the vowel realizations of several pairs of speakers were similar or different. The results were phoneme-dependent, and indicated that listener age affected the decision. The listener age effects in perception were compatible to the speaker age effects in the acoustics, indicating parallels of social factors in production and perception. The effects in the acoustic and perception data coincide with reported effects in literature on social interaction and imitation, and with literature on the results of investigations on the articulatory-auditory interaction in human beings.

The present research reveals a mutual sound change in the long vowels and the diphthongs of Standard Dutch. The results indicate that social information that is attached to sub-phonemic variation changes over time and affects the pronunciation and perception of the vowel phonemes studied.

NEDERLANDSE SAMENVATTING

Spreken is de meest gebruikelijke vorm van communicatie in een sociale omgeving. Naast de funktie van het overbrengen van inhoudelijke aspecten is het spraaksignaal ook het product van fysische eigenschappen en veranderingen. Tevens bevat het signaal alle factoren die de identiteit van de spreker vormen, zoals de sociale contacten en de afkomst van de spreker. De keuze van woorden, maar ook de manier waarop ze worden uitgesproken, verschilt van spreker tot spreker en zelfs binnen een spreker. De gesignaleerde uitspraakverandering van de tweeklank <ei> (Poldernederlands) was de aanleiding voor dit onderzoek.

In deze studie wordt de fonetische variatie in de realisatie van de Nederlandse klinkerfonemen /ɛi/, /ɑu/, /œy/, /eː/, en /oː/ (in woorden als <tijd>, <kous>, <huis>, <zeep>, en <boot>) binnen een representatieve steekproef van Nederlandse sprekers geanalyseerd. Het doel was te onderzoeken in hoeverre de distributie van bepaalde subfonemische uitspraakvarianten overeenkomt met de sociaal-economische achtergrond van de sprekers en of de distributie in de loop der tijd is veranderd. De klinkers uit de spontane spraak van een groep van 70 sprekers, afkomstig uit het Corpus Gesproken Nederlands (CGN) werden onderzocht. Het corpus bestond uit 35 vrouwen en 35 mannen van verschillende leeftijden en met verschillende sociaal-economische achtergronden. Over het algemeen gaat men er vanuit dat de sociale groepen waartoe sprekers behoren, overeenkomen met verschillende spreekpatronen. De achtergrond van een spreker zou deze spreekpatronen moeten weerspiegelen en uitspraakvarianten kunnen hierdoor geclassificeerd worden.

Naast de variatie door omgevingsfactoren wordt een deel van de variatie in het akoestische spraaksignaal tussen sprekers ook veroorzaakt door biologische eigenschappen, zoals het lengteverschil tussen de spraakkanalen van mannen en vrouwen. Om de akoestische kwaliteit van klinkers tussen verschillende sprekers en tussen de seksen te kunnen vergelijken, is een betrouwbare methode nodig. Deze methode moet de ongewenste variatie efficiënt minimaliseren en de linguïstische variatie behouden. Twee verschillende methodes om de klinkerkwaliteit akoestisch te kunnen meten werden vergeleken: formantanalyse en principale componenten analyse (PCA) op bandfilterdata werd toegepast.

Met de uit de PCA resulterende dimensies konden de verschillen in klinkerkwaliteit tussen de sprekers het meest succesvol worden gemeten. Deze methode werd vervolgens gebruikt voor alle verdere analyses. Om de fysiologische verschillen verder te minimali-

seren werden de klinkers per spreker aan de (hoek)klinkers /a/, /i/, en /u/ gerelateerd. Na het toepassen van deze methode werden verschillen tussen sociaal-economische groepen en leeftijden zichtbaar in de tweeklanken /ɛi/, /ɑu/, /œy/, en de lange klinkers /eː/, en /oː/. Gegeven onze analysemethode vonden we echter geen significante verschillen tussen de uitspraak van mannen en vrouwen. Sprekers met een hoger opleidingsniveau of beroep vertoonden een lagere onset en een sterkere diftongering bij alle onderzochte lange klinkers en tweeklanken. Anders dan bij sprekers met een lagere sociaal-economische status werden er in de groep van sprekers met een hogere sociaal-economische status ook significante verschillen tussen de generaties gevonden.

Om de gevonden akoestische verschillen te verifiëren is er een perceptie-experiment uitgevoerd. 30 luisteraars hebben beoordeeld of de klinkerrealisaties van telkens twee sprekers overeenkwamen of niet. De akoestische verschillen bleken inderdaad waarneembaar, maar de resultaten verschilden per foneem en waren afhankelijk van de leeftijd van de luisteraar. De invloed van de leeftijd van de luisteraars in het perceptie-experiment kwam overeen met de leeftijdseffecten in de uitspraak. Dit duidt op parallellen tussen sociale factoren in productie en perceptie. Deze effecten in de akoestische en de perceptieve data komen overeen met de in de literatuur beschreven effecten van sociale interactie en imitatie en met de resultaten van studies over de articulatorisch-perceptieve interactie.

Het hier beschreven onderzoek laat gekoppelde veranderingen zien in de lange klinkers en de tweeklanken van het gesproken Standaard Nederlands. De resultaten geven aan dat sociale informatie die met subfonemische variatie verbonden is, in de loop der tijd verandert en uitspraak en perceptie van meerdere klinkers beïnvloedt.

References

[1] ADANK, P. M. *Vowel normalization: a perceptual-acoustic study of Dutch vowels*. PhD thesis, Katholieke Universiteit Nijmegen, Nijmegen, 2003.

[2] ADANK, P. M., VAN HOUT, R., AND SMITS, R. An acoustic description of the vowels of Northern and Southern Standard Dutch. *Journal of the Acoustical Society of America 116(3)* (2004), 1729–1738.

[3] ADOLPHS, R. Cognitive neurosciences of human social behavior. *Nature Reviews Neuroscience 4* (2003), 165–178.

[4] ASSMANN, P., NEAREY, T., AND HOGAN, J. Vowel identification: orthographic, perceptual, and acoustic aspects. *Journal of the Acoustical Society of America 71* (1982), 975–989.

[5] BANDURA, A. Human agency in social cognitive theory. *The American Psychologist 44(9)* (1989), 1175–1184.

[6] BEDDOR, P. S., AND HAWKINS, S. The influence of spectral prominence on perceived vowel quality. *Journal of the Acoustical Society of America 87 (6)* (1990), 2684–2704.

[7] BLADON, A. Arguments against formants in the auditory representation of speech. In *The Representation of Speech in the Peripheral Auditory System*, R. Carlson and B. Granström, Eds. Elsevier, Amsterdam, 1982, pp. 95–102.

[8] BLADON, A. Two-formant models of vowel perception: shortcomings and enhancements. *Speech Communication 2* (1983), 305–313.

[9] BLADON, A. Diphthongs: a case study of dynamic auditory processing. *Speech Communication 4* (1985), 145–154.

[10] BLADON, R. A. W., AND LINDBLOM, B. J. Modeling the judgement of vowel quality differences. *Journal of the Acoustical Society of America 69* (1981), 1414–1422.

[11] BLANCQUAERT, E. *Praktische uitspraakleer van de Nederlandse taal (zesde uitgave)*. De Sikkel N. V., Antwerpen, 1962.

[12] BOERSMA, P. P. G., AND WEENINK, D. J. M. *Praat, a system for doing phonetics by computer*. http://www.praat.org, 1992-2006.

[13] BOOIJ, G. E. *The Phonology of Dutch*. Oxford University Press, Oxford, 1995.

[14] CHARTRAND, T. L., AND BARGH, J. A. The chameleon effect: the perception-behavior link and social interaction. *Journal of Personality and Social Psycholology 76* (1999), 893–910.

[15] CHIBA, T., AND KAJIYAMA, M. *The vowel: its nature and structure (1942)*. Setagaya: Phonetic Society of Japan, Tokyo, 1958.

[16] CHISTOVICH, L. A., AND LUBLINSKAYA, V. V. The 'center of gravity' effect in vowel spectra and critical distance between the formants: psychoacoustical study

of the perception of vowel-like stimuli. *Hearing Research 1* (1979), 185–195.

[17] COHEN, A. Diphthongs, mainly Dutch. In *Form and Substance*, L. Hammerich, R. Jakobson, and E. Zwirner, Eds. Akademisk Forlag, Copenhagen, 1971, pp. 277–289.

[18] COHEN, A., SLIS, I. H., AND 'T HART, J. Perceptual tolerances of isolated Dutch vowels. *Phonetica 9* (1963), 65–78.

[19] COLLIER, R., BELL-BERTI, F., AND RAPHAEL, L. J. Some acoustic and physiological observations on diphthongs. *Language and Speech 4* (1982), 59–69.

[20] COLLIER, R., AND 'T HART, J. The perceptual relevance of the formant trajectories in Dutch diphthongs. In *Sound Structures. Studies for Antonie Cohen*, M. van den Broecke, V. J. van Heuven, and V. Zonneveld, Eds. Foris Publications in Language Science, Dordrecht, Cinnaminson, 1983, pp. 31–45.

[21] COLLINS, B., AND MEES, I. *The phonetics of English and Dutch, revised edition*. Brill, Leiden, New York, Köln, 2003.

[22] CUCCHIARINI, C. *Phonetic transcriptions: a methodological and empirical study*. PhD thesis, Universiteit van Nijmegen, Nijmegen, 1993.

[23] CUTLER, A., AND WAGNER, A. Listening experience and phonetic-to-lexical mapping in L2. *Proceedings ICPhS* (2007), 43–48.

[24] DAELEMANS, W., AND VAN DEN BOSCH, A. Treetalk: Memory-based word phonemisation. In *Data-Driven Techniques in Speech Synthesis*, R. I. Damper, Ed. Klewer Academic Publishers, 2001, pp. 149–172.

[25] DEHAENE-LAMBERTZ, G., PALLIER, C., SERNICLAES, W., SPRENGER-CHAROLLES, L., JOBERT, A., AND DEHAENE, S. Neural correlates of switching from auditory to speech perception. *NeuroImage 25* (2005), 21–33.

[26] DELATTRE, P., LIBERMAN, A. M., COOPER, F. S., AND GERSTMAN, L. J. An experimental study of the acoustic determinants of vowel color; observations on one- and two-formant vowels synthesized from spectrographic patterns. *Word 8(3)* (1952), 195–210.

[27] DELVAUX, V., AND SOQUET, A. Inducing imitative phonetic variation in the laboratory. *Proceedings ICPhS* (2007), 369–372.

[28] DEN HERTOG, C. H. *De Nederlandsche taal: practische spraakkunst van het hedendaagsche Nederlandsch, 3e dr*. Versluys, Amsterdam, 1909-1911.

[29] DIOUBINA, O. I., AND PFITZINGER, H. R. An IPA vowel diagram approach to analysing L1 effects on vowel production and perception. *Proceedings ICSLP* (2002), 2265–2268.

[30] DISNER, S. F. Evaluation of vowel normalisation procedures. *Journal of the Acoustical Society of America 67* (1980), 253–261.

[31] DUDENREDAKTION WITH MAGOLD, M. *Duden. Aussprachewörterbuch. Wörterbuch der Deutschen Standardaussprache. Band 6, 3. Auflage*. Dudenredaktion,

Mannheim/Wien/Zürich, 1990.

[32] ECKERT, P. *Linguistic variation as social practice: the linguistic construction of identity in Belten High*. Blackwell Publishing, Oxford, 1999.

[33] EDELMAN, L. Het Poldernederlands: een vrouwentaal, een sociolinguistisch onderzoek. Unpublished report, Linguistics Program, 1999.

[34] EIJKMAN, L. P. H. *Phonetiek van het Nederlands*. De Erven F. Bohn N.V., Haarlem, 1937.

[35] EULITZ, C. Representation of phonological features in the brain: Evidence from mismatch negativity. *Proceedings ICPhS* (2007), 113–116.

[36] FANT, C. G. M. *Acoustic Theory of Speech Production, 2nd ed*. Mouton, The Hague, 1970.

[37] FENNELL, B. A. *A history of English*. Blackwell Publishers, Oxford, 2001.

[38] GALLESE, V., KEYSERS, C., AND RIZZOLATTI, G. A unifying view of the basis of social cognition. *TRENDS in Cognitive Sciences 8* (2004), 396–403.

[39] GAY, T. Effect of speaking rate on diphthong formant movements. *Journal of the Acoustical Society of America 44* (1968), 1570–1573.

[40] GERSTMAN, L. H. Classification of self-normalized vowels. *Proceedings IEEE Trans. Audio Electroacoust AU-16* (1968), 78–80.

[41] GILLIS, S. Protocol voor de brede fonetische transcriptie. Tech. rep., CGN, http://lands.let.kun.nl/CGN/home.htm, 2001.

[42] GOLDINGER, S. D. A complementary-systems approach to abstract and episodic speech perception. *Proceedings ICPhS* (2007), 49–54.

[43] GUSSENHOVEN, C. Dutch. In *Handbook of the International Phonetic Association*, IPA, Ed. Cambridge University Press, Cambridge, 1999, pp. 74–77.

[44] HARRINGTON, J. An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. *Journal of Phonetics 34* (2006), 439–457.

[45] HARRINGTON, J., AND CASSIDY, S. *Techniques in speech acoustics*. Kluwer Academic Publishers, Dordrecht, London, Boston, 1999.

[46] HARSHMAN, R. Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multi-modal factor analysis. *UCLA Working Papers in Phonetics 16* (1970), 1–84.

[47] HAY, J., WARREN, P., AND DRAGER, K. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics 34* (2006), 458–484.

[48] HEEMSKERK, J., AND ZONNEVELD, W. *Uitspraakwoordenboek*. Het Spectrum, Utrecht, 2000.

[49] HEFFERNAN, K. Vowel dispersion as a determinant of which sex leads a vowel change. *Proceedings ICPhS* (2007), 1485–1488.

[50] HICKOK, G., AND POEPPEL, D. Towards a functional neuroanatomy of speech perception. *Trends in Congnitive Science 4* (2000), 131–137.

[51] HOEMKE, K. A., AND DIEHL, R. L. Perception of vowel height: The role of F1-F0 distance. *Journal of the Acoustical Society of America 96* (1994), 661–674.

[52] HOLBROOK, A., AND FAIRBANKS, G. Diphthong formants and their movements. *Journal of Speech and Hearing Research 5* (1962), 33–58.

[53] HOMMEL, B. Bridging social and cognitive psychology? In *Bridging social psychology: Benefits of transdisciplinary approaches*, P. van Lange, Ed. Erlbaum, Mahwah, NJ, 2006, pp. 167–172.

[54] HOSMER, D. W., AND LEMESHOW, S. *Applied logistic regression, 2nd ed.* Wiley, New York, NY, 2000.

[55] INGRAM, J. C., AND PARK, S.-G. Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics 25(3)* (2002), 343–370.

[56] IPA. *Handbook of the International Phonetic Association*. Cambridge University Press, Cambridge, 1999.

[57] ITO, M., TSUCHIDA, J., AND YANO, M. On the effectiveness of whole spectral shape for vowel perception. *Journal of the Acoustical Society of America 110* (2001), 1141–1149.

[58] IVERSON, P., KUHL, P. K., AKAHANE-YAMADA, R., DIESCH, E., TOHKURA, Y., KETTERMANN, A., AND SIEBERT, C. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition 87* (2003), B47–B57.

[59] JACOBI, I., POLS, L. C. W., AND STROOP, J. Polder Dutch: Aspects of the /ɛi/-lowering in Standard Dutch. *Proceedings Interspeech* (2005), 2877–2800.

[60] JACOBI, I., POLS, L. C. W., AND STROOP, J. Measuring and comparing vowel qualities in a Dutch spontaneous speech corpus. *Proceedings Interspeech* (2006), 701–704.

[61] JACOBI, I., POLS, L. C. W., AND STROOP, J. Dutch diphthong and long vowel realizations as changing socio-economic markers. *Proceedings ICPhS* (2007), 1481–1484.

[62] JANSSENS, G., AND MARYNISSEN, A. *Het Nederlands van vroeger en nu*. Acco, Leuven, 2003.

[63] JESPERSEN, O. *Lehrbuch der Phonetik*. Teubner, Leipzig, 1926.

[64] JOHNSON, K. *Acoustic and auditory phonetics*. Blackwell Publishing, Malden, Oxford, Melbourne, Berlin, 2003.

[65] JONES, D. *An outline of English phonetics, ninth edition*. Heffer, Cambridge, 1967.

[66] JONES, D. *English pronouncing dictionary, 15th ed*. edited by Roach, P. and Hartman, J. , Cambridge University Press, Cambridge, 1997.

[67] JONGMANS, P. *The intelligibility of tracheoesophageal speech. An analytic and rehabilitation study.* PhD thesis, Universiteit van Amsterdam, Amsterdam, 2008.

[68] JOOS, M. Language monograph No.23: Acoustic phonetics. *Language 24(2)*

(1948), 1–136.

[69] KAISER, L. Diphthongs in Dutch. *Lingua 1* (1949), 303–305.

[70] KIEFTE, M., AND KLUENDER, K. R. The relative importance of spectral tilt in monophthongs and diphthongs. *Journal of the Acoustical Society of America 117(3)* (2005), 1395–1404.

[71] KLABBERS, E., AND VAN SANTEN, J. Predicting segmental durations for Dutch using the sums-of-products approach. *Proceedings ICSLP* (2000), 670–673.

[72] KLATT, D. Prediction of perceived phonetic distance from critical-band spectra: A first step. *Proceedings IEEE ICASSP* (1982), 1278–1281.

[73] KLEIN, W., PLOMP, R., AND POLS, L. C. W. Vowel spectra, vowel spaces and vowel identification. *Journal of the Acoustical Society of America 48* (1970), 999–1009.

[74] KLOEKE, G. G. *Gezag en norm bij het gebruik van verzorgd Nederlands*. J. M. Meulenhoff, Amsterdam, 1951.

[75] KOOPMANS-VAN BEINUM, F. J. Nog meer fonetische zekerheden. *Nieuwe Taalgids 62* (1969), 245–250.

[76] KOOPMANS-VAN BEINUM, F. J. Comparative phonetic vowel analysis. *Journal of Phonetics 1* (1973), 249–261.

[77] KOOPMANS-VAN BEINUM, F. J. *Vowel contrast reduction. An acoustic and perceptual study of Dutch vowels in various speech conditions*. PhD thesis, Universiteit van Amsterdam, Amsterdam, 1980.

[78] KUHL, P. K. Is speech learning 'gated' by the social brain? *Developmental Science 10:1* (2007), 110–120.

[79] KUHL, P. K., COFFEY-CORINA, S., PADDEN, D., AND DAWSON, G. Links between social and linguistic processing of speech in preschool children with autism: behavioral and electrophysiological measures. *Developmental Science 8:1* (2005), F1–F12.

[80] LABOV, W. The social motivation of a sound change. *Word 19* (1963), 273–309.

[81] LABOV, W. The social stratification of English in New York City. Tech. rep., Center for Applied Linguistics, Washington D.C., 1966.

[82] LABOV, W. The social origins of sound change. In *Quantitative analyses of linguistic structure, 2. Linguistic change*, W. Labov, Ed. Basil Blackwell, Oxford, New York, 1980.

[83] LABOV, W. On the mechanisms of linguistic change. In *Quantitative analyses of linguistic structure, 2. Linguistic change*, J. J. Gumperz and D. Hymes, Eds. Basil Blackwell, Oxford, New York, 1989.

[84] LABOV, W. *Principles of linguistic change; Volume 1. Internal Factors*. Malden, Oxford, 1994.

[85] LADEFOGED, P. *A course in phonetics*. Harcourt Brace Jovanovich, 1982.

[86] LADEFOGED, P., AND BROADBENT, D. E. Information conveyed by vowels. *Journal of the Acoustical Society of America 29* (1957), 98–104.

[87] LAVE, J., AND WENGER, E. *Situated learning: Legitimate peripheral participation*. Cambridge University Press, Camebridge, 1991.

[88] LEHISTE, I., AND PETERSON, G. Transitions, glides and diphthongs. *Journal of the Acoustical Society of America 33* (1961), 268–277.

[89] LIBERMAN, A., HARRIS, K. S., HOFFMAN, H. S., AND GRIFFITH, B. C. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology 54* (1957), 358–368.

[90] LIEBERMAN, M. D. Social cognitive neuroscience: A review of core processes. *Annual Review of Psychology 58* (2007), 259–289.

[91] LINDBLOM, B. E. F., AND SUNDBERG, J. E. F. Acoustical consequences of lip, tongue, jaw, and larynx movement. *Journal of the Acoustical Society of America 4* (1971), 1166–1179.

[92] LOBANOV, B. M. Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America 49* (1971), 606–608.

[93] MAGNUSON, J. S., AND NUSBAUM, H. C. Acoustic differences, listener expectations, and the perceptual accomodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance 33* (2007), 391–409.

[94] MARKEL, J. D., AND GRAY, A. H. *Linear prediction of speech*. Springer-Verlag, New York, 1976.

[95] MCMURRAY, B., TANENHAUS, M. K., AND ASLIN, R. N. Gradient effects of within-category phonetic variation on lexical access. *Cognition 86* (2002), B33–B42.

[96] MEES, I., AND COLLINS, B. A phonetic description of the vowel system of Standard Dutch (ABN). *Journal of the International Phonetic Association 13* (1983), 64–75.

[97] MILLER, J. D. Auditory-perceptual interpretation of the vowel. *Journal of the Acoustical Society of America 85* (1989), 2114–2134.

[98] MILLER, R. L. Auditory tests with synthetic vowels. *Journal of the Acoustical Society of America 25* (1953), 114–121.

[99] MILROY, J., AND MILROY, L. Varieties and variation. In *The Handbook of Sociolinguistics*, F. Coulmas, Ed. Blackwell Publishing, Oxford, Cambridge, 1997, pp. 47–64.

[100] MILROY, L., AND GORDON, M. *Sociolinguistics: Method and Interpretation*. Blackwell, Malden, Oxford, Melbourne, Berlin, 2003.

[101] MOL, H. Fonetische zekerheden. *Nieuwe Taalgids 62* (1969), 161–167.

[102] MONSEN, R. B., AND ENGEBRETSON, A. M. The accuracy of formant frequency measurements: a comparison of spectrographic analysis and linear predic-

tion. *Journal of Speech and Hearing Research 26* (1983), 89–97.

[103] MOULTON, W. G. The vowels of Dutch: phonetic and distributional classes. *Lingua 11* (1962), 294–312.

[104] NEAREY, T. M. *Phonetic feature systems for vowels*. PhD thesis, University of Connecticut, Storrs, CT, 1977.

[105] NEAREY, T. M. Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America 85* (1989), 2088–2113.

[106] NOOTEBOOM, S. G., AND COHEN, A. *Spreken en verstaan. Een inleiding tot de experimentele fonetiek*. Van Gorcum, Assen/Amsterdam, 1976.

[107] NOOTEBOOM, S. G., AND DOODEMAN, G. J. N. Production and perception of vowel length in spoken sentences. *Journal of the Acoustical Society of America 67(1)* (1980), 276–287.

[108] NOOTEBOOM, S. G., AND SLIS, I. H. The phonetic feature of vowel length in Dutch. *Language & Speech 15* (1972), 301–316.

[109] OHALA, J. J. The listener as a source of sound change. In *Papers from the parasession on language and behavior*, C. S. Masek, R. A. Hendrick, and M. F. Miller, Eds. Chicago Linguistic Society, Unversity of Chicago, Chicago, Illinois, 1981, pp. 178–203.

[110] OHALA, J. J. The phonetics of sound change. In *Historical linguistics: problems and perspectives*, C. Jones, Ed. Longman, London, 1993, pp. 237–278.

[111] OOSTDIJK, N., GOEDERTIER, W., VAN EYNDE, F., BOVES, L., MARTENS, J. P., MOORTGAT, M., AND BAAYEN, H. Experiences form the Spoken Dutch Corpus project. *Proceedings 3rd LREC* (2002), 340–347.

[112] PARDO, J. S. On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America 119(4)* (2006), 2382–2393.

[113] PEETERS, W. J. M. *Diphthong dynamics. A cross-linguistic perceptual analysis of temporal patterns in Dutch, English, and German*. PhD thesis, Rijksuniversiteit Utrecht, Kampen, 1991.

[114] PEETERS, W. J. M. Diphthong dynamics. A cross-linguistic perceptual analysis of temporal patterns in Dutch, English, and German. *Language 69* (1993), 632–633.

[115] PETERSON, G. E., AND BARNEY, H. L. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America 24* (1952), 175–184.

[116] PICKERING, M. J., AND GARROD, S. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences 27* (2004), 169–226.

[117] PLOMP, R., POLS, L. C. W., AND VAN DER GEER, J. O. Dimensional analysis of vowel spectra. *Journal of the Acoustical Society of America 41* (1967), 707–712.

[118] POLKA, L., AND BOHN, O.-S. A cross-language comparison of vowel perception in English-learning and German-learning infants. *Journal of the Acoustical Society of America 100* (1996), 577–592.

[119] POLKA, L., AND WERKER, J. F. Developmental changes in the perception of non-native vowel contrasts. *Journal of Experimental Psychology: Hum. Percept. Perform. 20* (1994), 421–435.

[120] POLS, L. C. W. Real-time recognition of spoken words. *Proceedings IEEE Transactions on Computers C-20 (9)* (1971), 972–978.

[121] POLS, L. C. W. *Spectral analysis and identification of Dutch vowels in monosyllabic words*. PhD thesis, Vrije Universiteit Amsterdam, Amsterdam, 1977.

[122] POLS, L. C. W., TROMP, H. R. C., AND PLOMP, R. Frequency analysis of Dutch vowels from 50 male speakers. *Journal of the Acoustical Society of America 53* (1973), 1093–1101.

[123] POLS, L. C. W., VAN DER KAMP, L. J. T., AND PLOMP, R. Perceptual and physical space of vowel sounds. *Journal of the Acoustical Society of America 46* (1969), 458–467.

[124] POLS, L. C. W., AND VAN SON, R. J. J. H. Acoustics and perception of dynamic vowel segments. *Speech Communication 13* (1993), 135–147.

[125] POTTER, R. K., AND PETERSON, G. E. The representation of vowels and their movements. *Journal of the Acoustical Society of America 20* (1948), 528–535.

[126] PULVERMÜLLER, F. Brain mechanisms linking language to action. *Nature Reviews Neuroscience 7* (2005), 574–582.

[127] R DEVELOPMENT CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006.

[128] RIZZOLATTI, G., AND CRAIGHERO, L. The mirror-neuron system. *Annual Review of Neuroscience 27* (2004), 169–192.

[129] ROSEN, S., AND FOURCIN, A. J. Frequency selectivity and the perception of speech. In *Frequency selectivity in hearing*, B. C. J. Moore, Ed. Academic Press, London, 1986, pp. 373–487.

[130] SANCIER, M. L., AND FOWLER, C. A. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics 25* (1997), 421–436.

[131] SHRIBERG, L. D., AND LOF, L. Reliability studies in broad and narrow transcriptions. *Clinical Linguistics and Phonetics 5* (1991), 225–279.

[132] SIMPSON, A. P. Dynamic consequences of differences in male and female vocal tract dimensions. *Journal of the Acoustical Society of America 109* (2001), 2153–2164.

[133] SMAKMAN, D. *Standard Dutch in the Netherlands, a sociolinguistic and phonetic description*. PhD thesis, Radboud Universiteit Nijmegen, LOT Utrecht, 2006.

[134] SPSS. *Rel. 14.1*. Chicago: SPSS Inc., 2006.

[135] STEVENS, K. N. The quantal nature of speech: Evidence from articulatory-acoustic data. In *Human Communication: A Unified View*, P. D. Denes and E. E. David, Eds. Jr. McGraw-Hill, New York, 1971, pp. 51–66.

[136] STEVENS, K. N. Critique: Articulatory-acoustic relations and their role in speech perception. *Journal of the Acoustical Society of America 99* (1996), 1693–1694.

[137] STEVENS, K. N. *Acoustic phonetics*. MIT Press, Cambridge, MA, 1998.

[138] STRANGE, W. Dynamic specification of coarticulated vowels. *Journal of the Acoustical Society of America 74* (1983), 695–705.

[139] STRIK, H., AND KONST, E. A duration model for phonetic units in isolated dutch words. *AFN-Proceedings Vol.15* (1992), 71–78.

[140] STROOP, J. *Poldernederlands, waardoor het ABN verdwijnt*. Bert Bakker, Amsterdam, 1998.

[141] STROOP, J. Van delta naar tweestromenland. Over het divergerende Nederlands. In *Waar gaat het Nederlands naartoe? Panorama van een taal*, J. Stroop, Ed. Bert Bakker, Amsterdam, 2003, pp. 14–24.

[142] SYRDAL, A. K., AND GOPAL, H. S. A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America 79* (1986), 1086–1100.

[143] 'T HART, J. Fonetische steunpunten. *Nieuwe Taalgids 62* (1969), 168–174.

[144] TEMPELAARS, S. *Signaalverwerking in spraak en muziek*. Koninklijk Conservatorium, Den Haag, 1991.

[145] TERBEEK, D. A cross-language multidimensional scaling study of vowel perception. *UCLA Working papers in phonetics 37* (1977), 1–271.

[146] TJAFEL, H. Social psychology of intergroup relations. *Annual Review of Psychology 33* (1982), 1–19.

[147] TRAUNMÜLLER, H. Perceptual dimension of openness in vowels. *Journal of the Acoustical Society of America 69* (1981), 1465–1475.

[148] TRAUNMÜLLER, H. Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America 88* (1990), 97–100.

[149] VALLABHA, G. K., AND TULLER, B. Perceptuomotor bias in the imitation of steady-state vowels. *Journal of the Acoustical Society of America 116(2)* (2004), 1184–1197.

[150] VAN BERGEM, D. R. Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech Communication 12,1* (1993), 1–23.

[151] VAN BEZOOIJEN, R. Poldernederlands: Hoe kijken vrouwen ertegenaan? *Nederlandse Taalkunde 6 (4)* (2001), 257–271.

[152] VAN BEZOOIJEN, R., AND GOOSKENS, C. Identification of language varieties: The contribution of different linguistic levels. *Journal of Language and Social Psychology 18* (1999), 31–48.

[153] VAN BEZOOIJEN, R., AND VAN DEN BERG, R. Who power Polder Dutch? A perceptual-sociolinguistic study of a new variety of Dutch. *Linguistics in the Netherlands* (2001), 1–12.

[154] VAN DE VELDE, H. *Variatie en verandering in het gesproken Standaard-Nederlands (1935-1993)*. PhD thesis, Katholieke Universiteit Nijmegen, Nijmegen, 1996.

[155] VAN DE VELDE, H. Watching Dutch change: a real time study of variation and change in Standard Dutch pronunciation. *Journal of Sociolinguistics 1* (2001), 361–391.

[156] VAN HEUVEN, V. J., EDELMAN, L., AND VAN BEZOOIJEN, R. The pronunciation of /ɛi/ by male and female speakers of avant-garde Dutch. *Linguistics in the Netherlands* (2002), 61–72.

[157] VAN NIEROP, D. J. P. J., POLS, L. C. W., AND PLOMP, R. Frequency analysis of Dutch vowels from 25 female speakers. *Acustica 29* (1973), 110–118.

[158] VAN SON, R. J. J. H. *Spectro-temporal features of vowel segments*. PhD thesis, Universiteit van Amsterdam, Amsterdam, 1993.

[159] VAN SON, R. J. J. H., BINNENPOORTE, D., VAN DEN HEUVEL, H., AND POLS, L. C. W. The IFA corpus: a phonemically segmented Dutch Open Source speech database. *Proceedings Eurospeech 3* (2001), 2051–2054.

[160] VAN SON, R. J. J. H., AND POLS, L. C. W. Formant frequencies of Dutch vowels in a text, read at normal and fast rate. *Journal of the Acoustical Society of America 88* (1990), 1683–1693.

[161] VON HELMHOLTZ, H. *Die Lehre von den Tonempfindungen, als physiologische Grundlage für die Theorie der Musik*. Vieweg, Braunschweig, 1862.

[162] VYGOTSKY, L. S. *Thought and Language*. MIT Press, Cambridge, MA, 1986.

[163] WARREN, P., HAY, J., AND THOMAS, B. The loci of sound change effects in recognition and perception. In *Laboratory Phonology 9th*, J. Cole and J. I. Hualde, Eds. Mouton de Gruyter, New York, 2007, pp. 87–112.

[164] WEENINK, D. J. M. *Speaker-adaptive vowel identification*. PhD thesis, Universiteit van Amsterdam, Amsterdam, 2006.

[165] WELLS, J. C. *A phonetic update on RP*. Moderna språk, 82(1), 1990.

[166] WREDE, B., FINK, G., AND G. SAGERER, G. Influence of duration on static and dynamic properties of German vowels in spontaneous speech. *Proceedings ICSLP* (2000), 82–85.

[167] ZONNEVELD, W., AND TROMMELEN, M. Egg, onion, ouch! On the representation of Dutch diphthongs. In *Studies in Dutch Phonology*, W. Zonneveld, F. van Coetsem, and O. W. Robinson, Eds. Martinus Nijhoff, Den Haag, 1980, pp. 265–292.

[168] ZWAARDEMAKER, H., AND EIJKMAN, L. P. H. *Leerboek der phonetiek, inzonderheid met betrekking tot het Standaard-Nederlandsch*. De Erven F. Bohn, Haarlem, 1928.

CURRICULUM VITAE

Irene Jacobi (1975, Bad Cannstatt, Germany) received her Abitur at the Otto-Hahn-Gymnasium, Ostfildern-Nellingen in 1994. After a short period of studying German and Anglo-American law at the Westfälische Wilhelms-Universität Münster, accompanied by internships at lawyer offices, she enrolled for Phonetics at the Ludwig-Maximilians-Universität in Munich in 1998. During her study of Phonetics, Psycholinguistics and Psychology she worked as a student assistant at the Phonetics Department of the LMU. In 2000 she took part in the Erasmus program and studied a half year at the University of Utrecht, the Netherlands. In 2003 she received her MA in Phonetics, with Psychology and Psycholinguistics as minor subjects, and started working as an assistant researcher at the Phonetics Department in Munich. From 2004 to 2008 she worked as "assistente in opleiding" within the research institute ACLC at the Dutch Department of the University of Amsterdam where she carried out the research that resulted in the present dissertation. Since 2008 she has been working as a researcher at the Netherlands Cancer Institute / Antoni van Leeuwenhoek hospital in Amsterdam.

ACKNOWLEDGEMENTS

Thank you all:

Prof.dr.ir. Louis Pols en Dr. Jan Stroop – beste Louis en Jan, met veel genoegen kijk ik terug op de afgelopen jaren van samenwerking. Jullie supervisie heb ik zeer gewaardeerd. Nu kunnen jullie eindelijk van je vrije tijd gaan genieten!

Prof.dr. Fred Weerman for his contribution to this dissertation.

The reading commission Prof.dr. P.P.G. Boersma, Prof.dr. J. Harrington, Prof.dr. V.J.J.P. van Heuven, Prof.dr. R. van Hout, Prof.dr. M. van Oostendorp, Dr. M.E.H. Schouten.

Being separated from fellow phoneticians was not always easy but it was surely gezellig: Thanks to my roommates during four years, Marian and Robert. Also everyone at the Neerlandistiek, the secretaries, all who joined the lunch-breaks at the PCH, Loulou, Maren, Suzanne, Hedde, Elma, Daniela, Antje, Catherine, Marije, Margarita, Nivja, Nel, Therese, the ACLC, and all ACLC-borrelaars, especially the aio's.

Mijn paranimfen Diana en Rafael.

Alle Amsterdamse foneten – altijd een open oor als ik langs de Herengracht fietste voor hulp, motivatie of raad. In het bijzonder David, Dirk-Jan, Rob, Ton en Paul.

Die Münchener (mittlerweile verstreut in aller Herren Länder), besonders die 'Trinker'. Und alle vom Phonetik Institut in München – gemeinsame Konferenzen waren nicht nur wissenschaftliche sondern auch soziale Höhepunkte.

Die Deutsch-Amsterdamse Clique Martin, Roli, Maren, Rachel, Berit, Sebastian, Raffi, die Kulturgang Nina & Marije, de gratiën Petra, Diana, Wieneke, the Maprevs, de roeiers, de crea's, de wandelaars, and all who shared time with me at conferences, in arts, sports, and more..

Prof.dr. Frans Hilgers en de HH-afdeling van het NKI. Anne en Eeke, Annemieke, Marion. De oio's Ewoud, Ingrid, Iris, Marieke, Maurits, Niels, Renske, Ronald, Tjeerd. En uiteraard kamergenootje Lisette.

Meine Mädels und Jungs aus dem Schwabenländle für Jahrzehnte Freundschaft.

Und natürlich meine Familie: Ro und die Mädels, Tante Wies, Tante Wila en Oom Jos, en – lest best – Ma & Pa!