



UvA-DARE (Digital Academic Repository)

On variation and change in diphthongs and long vowels of spoken Dutch

Jacobi, I.

Publication date

2009

Document Version

Final published version

[Link to publication](#)

Citation for published version (APA):

Jacobi, I. (2009). *On variation and change in diphthongs and long vowels of spoken Dutch*. [Thesis, fully internal, Universiteit van Amsterdam].

General rights

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

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

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/, /uu/, /œy/, /e:/, and /o:/ showed significant differences between social groups and ages. /ø:/ 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 /e:/ and /o:/. 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 / ϵi / 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 / ϵi / 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 / ϵi / (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 / ϵi / can be confirmed by the present speaker sample, and to what extent the other diphthongs / au / and / αy /, or the long vowels / e / and / o / 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 /ei/, /au/, /œy/ from words with <ij/ei>, <au/ou>, and <ui>, and their realizations of the long vowels /e:/, /o:/ and /ø:/ from words with <ee>, <oo> and <eu> are investigated. Compared to the other vowels, /ø:/ was less frequent in the data. Due to the small amount of data for /ø:/, 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 /ei/, /au/, /œy/, /e:/, and /o:/ (initially also /ø:/), 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 Flanders¹. 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 'perceived 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 /ei/, "A+" to /au/, "Y+" to /œy/, "e" to /e:/, and "o" to /o:/.

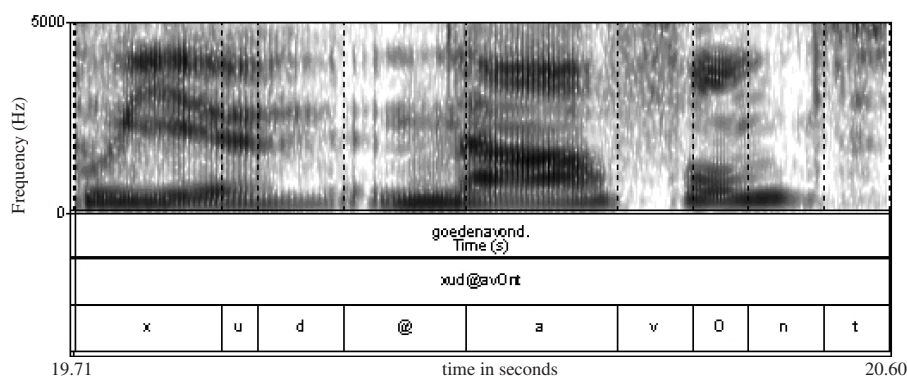


Figure 4.1: Example of an incorrect phonetic annotation in the CGN: Right after the initial fricative /x/ of <goedenavond>, the spectrogram shows a clear triphthong during the first syllable indicating [xuiənafɔnt], whereas the annotation sticks to the Dutch Standard [xudəvɔnt]. As a result, the segmentation 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 [xudəvɔnt], the Dutch standard, whereas the spectrogram indicates a different realization, namely [xuiənafɔnt]. 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/, /ei/, /au/, /œy/, /e:/ and /o:/ in stressed syllables were selected (/ø:/ 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 [ɫ] 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.

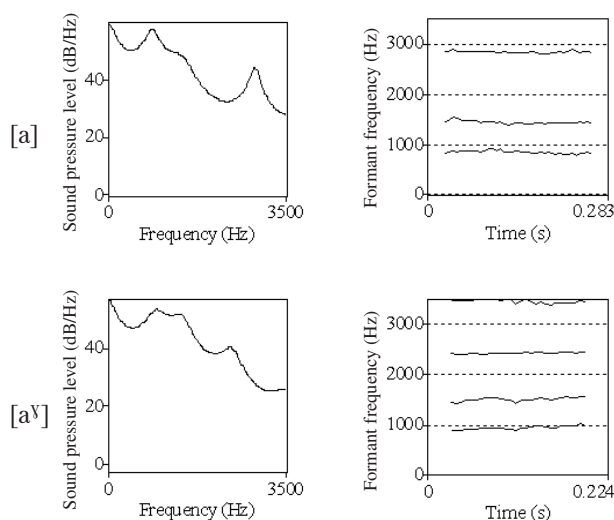


Figure 4.2: Two example vowels spoken by a trained phonetician: Vowel /a/ realized with and without velar approximation. Top: mean spectrum and first three formants of unvelarized [a]. Bottom: mean spectrum of [a^V] (with velar approximation) and a low third formant, with the fourth formant appearing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

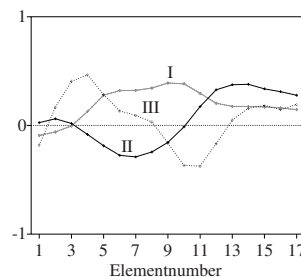
the area of the first, second and third formants. Higher formants include mainly speaker-specific information.

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 (/o:/, /e:/) and diphthongs (/ɛi/, /au/, /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 formula³, 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 $A = \frac{\sqrt{(a+b+c)(a+b-c)(-a+b+c)(a-b+c)}}{4} = \frac{\sqrt{(a^2+b^2+c^2)^2 - 2(a^4+b^4+c^4)}}{4}$

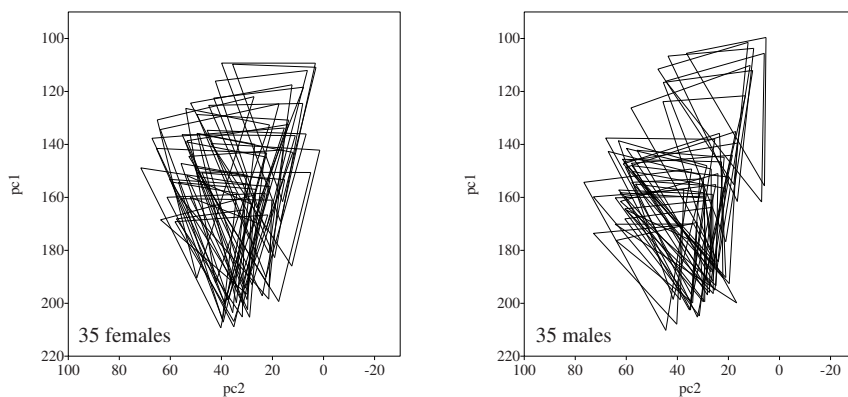


Figure 4.4: /a/-/i/-/u/ triangles in the $pc1/pc2$ dimensions of a PCA on the 210 mean /a/, /i/, /u/ filter output.

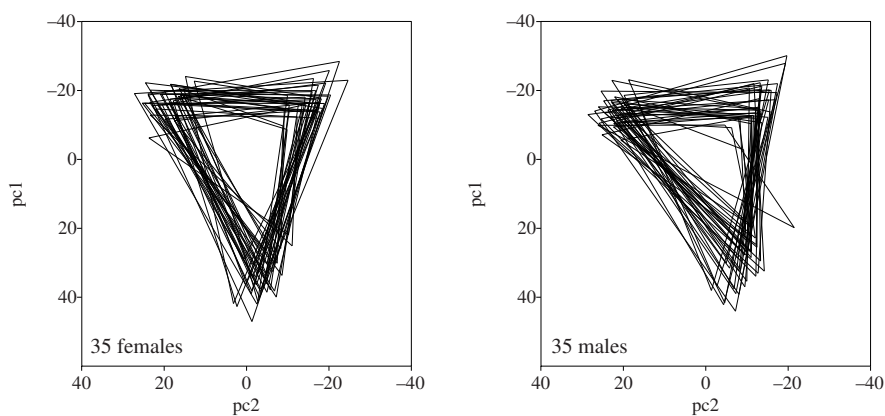


Figure 4.5: /a/-/i/-/u/ triangles in $pc1$ and $pc2$ after normalizing the focus to 0 in all dimensions.

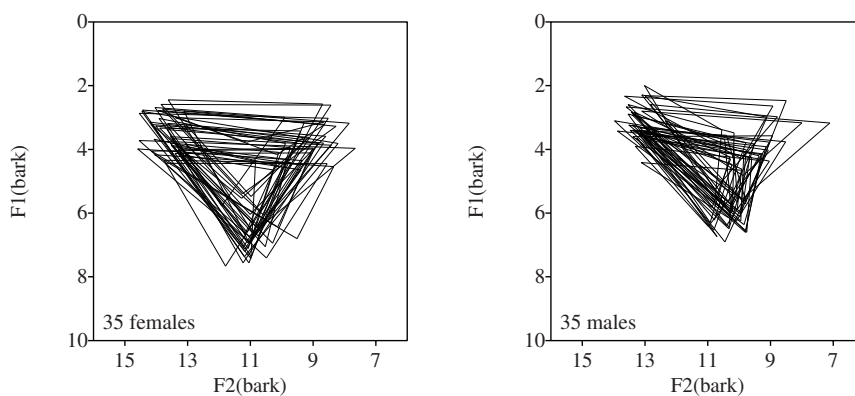
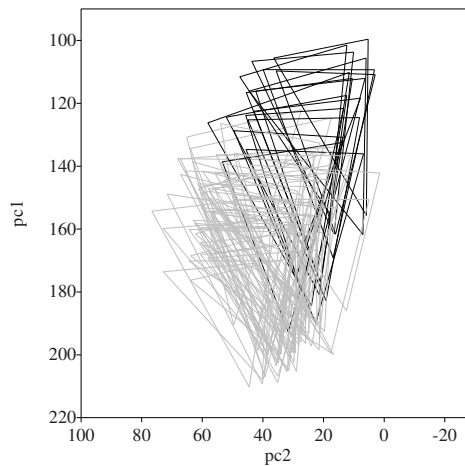


Figure 4.6: /a/-/i/-/u/ triangles of the females' and males' first and second formant in Bark.

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

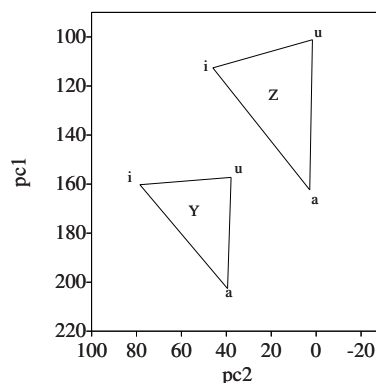
When comparing the vowel triangles and the positioning of each speaker's vowel set in the two-dimensional pc-space, inter-speaker differences were obvious that were beyond an attribution to the two sexes, and the variance within the sexes was more striking than the variance between the sexes (fig. 4.4, 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 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,

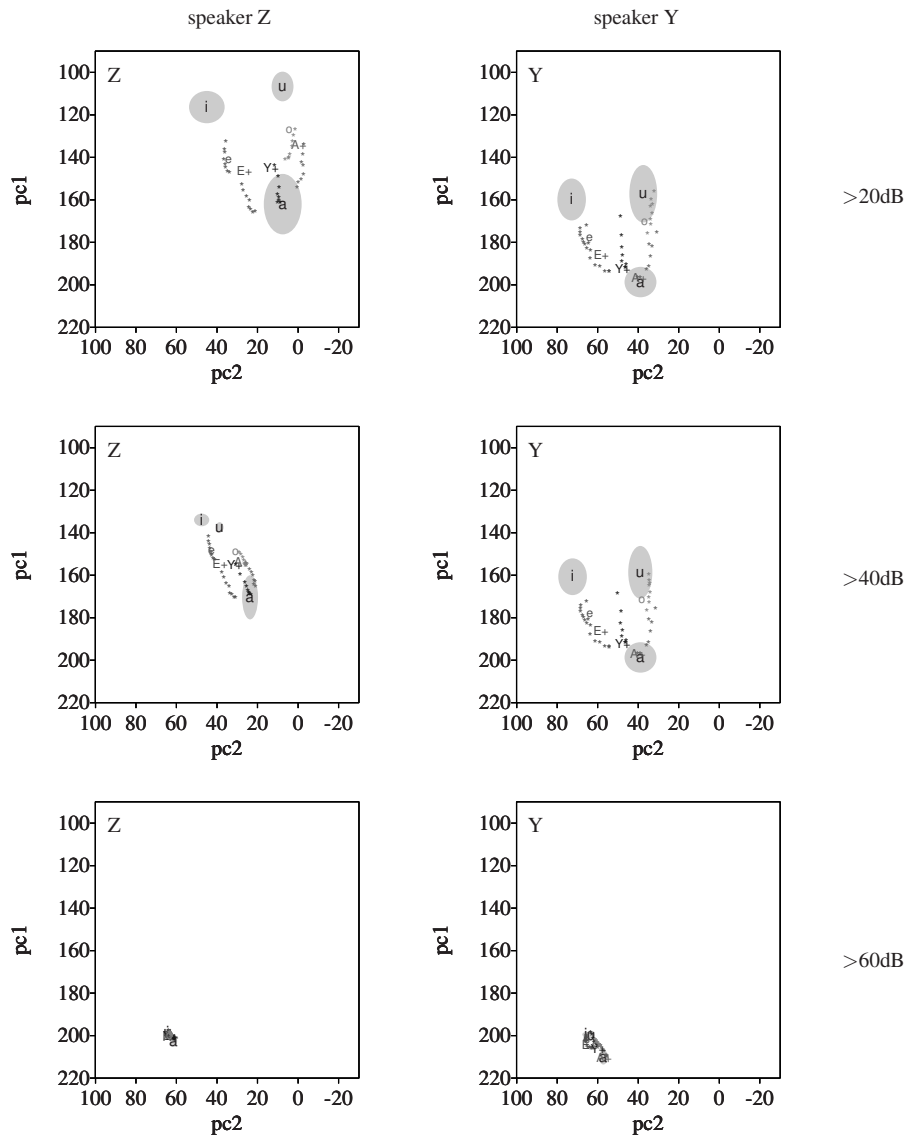


Figure 4.9: The influence of the dB minimum-level (and thus of the noise level) in the filters on the vowel space size of male speaker Z (good recording quality) and male speaker Y (radio recording, often music in the background): from top to bottom, the dB level for the minimum filter value was increased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

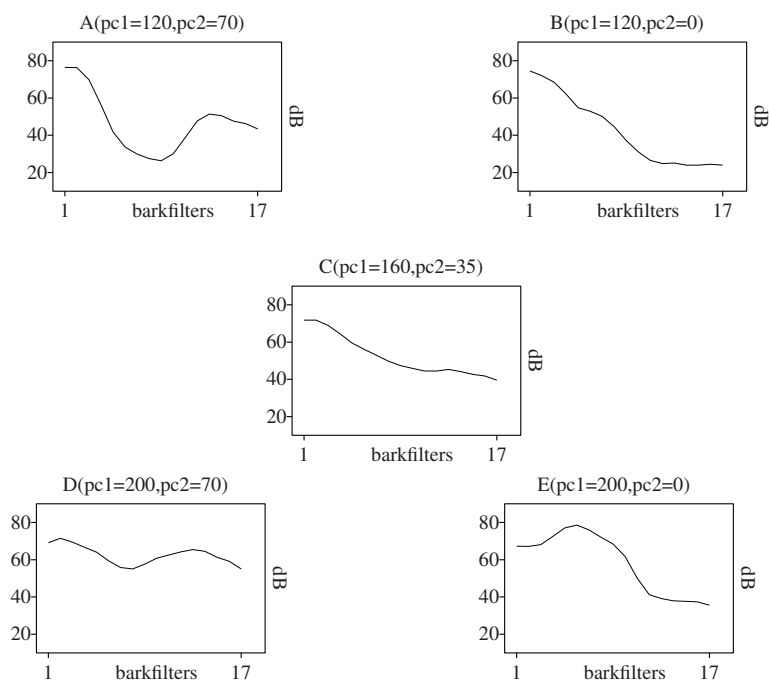
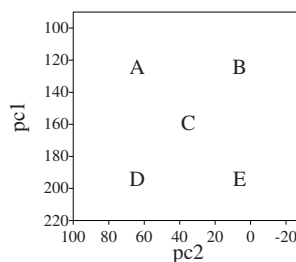


Figure 4.10: Recalculation of the original spectra of the pc1-pc2 dimensions. Recalculated spectra (above) from the corners and from the center of the pc1-pc2 plane, indicated by A, B, C, D, E (as positioned in the pc1-pc2 plane to the right). For all pc3 to pc19 values, the 44 speakers' /a/-/i/-/u/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

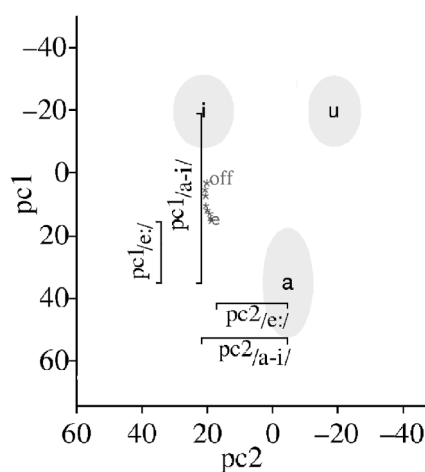
$$rel. \text{ onset} = \frac{pc_a - pc_{onset}}{pc_a - pc_i} * 100$$

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):

$rel. \text{ onset } /e:/ = \frac{pc1_{e:}}{pc1_{a-i}} * 100 = \frac{19}{57} * 100 = 33$. 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):

$rel. \text{ onset } /e:/ = \frac{pc2_{e:}}{pc2_{a-i}} * 100 = \frac{20}{26} * 100 = 77$. 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:

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

In figure 4.11, p. 65, the movement of /e:/ 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 /e:/ 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 /e:/ in figure 4.11 would be

rel. degree of diphthongization /e:/ = $\frac{pc1_{e:} - pc1_{offset} * 100}{pc1_{a-i}} = \frac{19-3}{57} * 100 = 28$. The diphthongization of /e:/ 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 /e:/, /ɛi/, /i/, and

/o:/ (table 4.6 below). Realizations of /u/, /au/, /œy/ (and /ø:/, 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/	/au/	/œy/	/o:/	/e:/	(/ø:/)	/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), /au/ showed the longest duration, followed by /œy/, /ɛi/, /o:/, /e:/, and /ø:/ as the shortest vowel. In Nootboom & 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/, /ø:/, /o:/, /au/, and the shortest vowel /e:/. Twenty years later, /e:/ and /o:/ 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/, /ø:/, /e:/, /au/, and the shortest vowel /o:/.

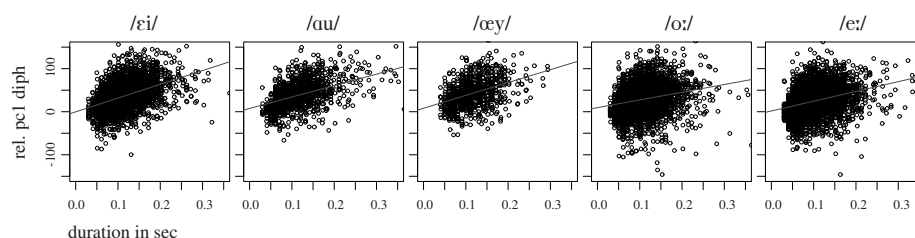


Figure 4.12: Relative degree of diphthongization in pc1 (y-axis) and durations (x-axis) of the vowel phonemes /ɛi/, /au/, /œy/, /o:/, and /e:/ with regression lines.

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.

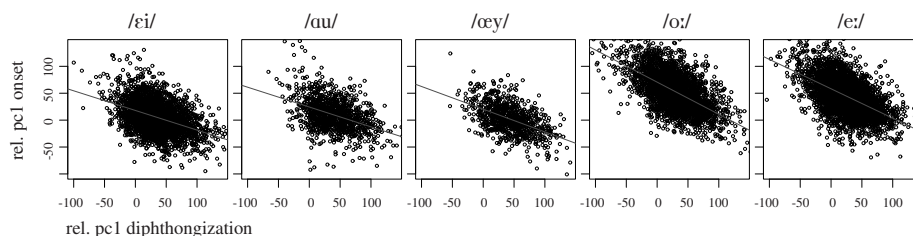


Figure 4.13: Relative onsets (y-axis) and degrees of diphthongization (x-axis) in pc1 with regression lines for the vowel phonemes /ɛi/, /au/, /œy/, /o:/, and /e:/.

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/, /au/. 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 diphthongization ($diph = \frac{pc1_{onset} - pc1_{offset}}{pc1_a - pc1_i} * 100$) in pc1. The results for the 5 vowel phonemes /o:/, /e:/, /ɛi/, /au/, 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).

sex	eduL	rel.pc1	/o:/		/e:/		/ɛi/		/au/		/œy/	
			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	
	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 /e:/ and /o/, and the difference between /au/ 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 /*ei*/ 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 /*ei*/s with decreasing female speaker age, and we also expect all of the younger more highly educated speakers to lower /*ei*/. For the males, we expect a lowered /*ei*/ at a later stage. The expectations for the low educated speakers are less clear. Following the assumptions of the literature on /*ei*/, the low educated speakers should not show a lowering of /*ei*/ (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 /*o*/, /*e*/, /*au*/, and /*æy*/, we expect similar patterns as those found for /*ei*/, 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 /*o*/ $p=.092$, mean difference -8.191, the 95% C.I. ranges from -17.768 to 1.387; for /*e*/ $p=.141$, mean diff 5.856, C.I. -1.987 to 13.699; for /*ei*/ $p=.133$, mean diff 4.789, C.I. -1.493 to 11.072; for /*au*/ $p=.552$, mean diff -2.042, C.I. -8.863 to 4.780; for /*æy*/ $p=.337$, 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 /*o*/ and /*e*/ was noticeably larger than the range of the diphthongs' onsets. The large range, and

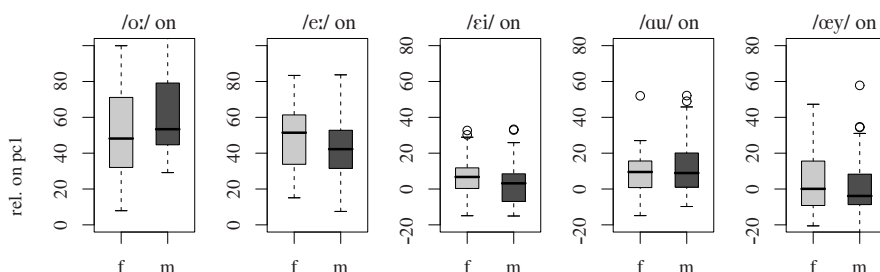


Figure 4.14: Box plots of the 35 females' (f) and 35 males' (m) relative pc1 onset values (y-axis) for the five vowel phonemes.

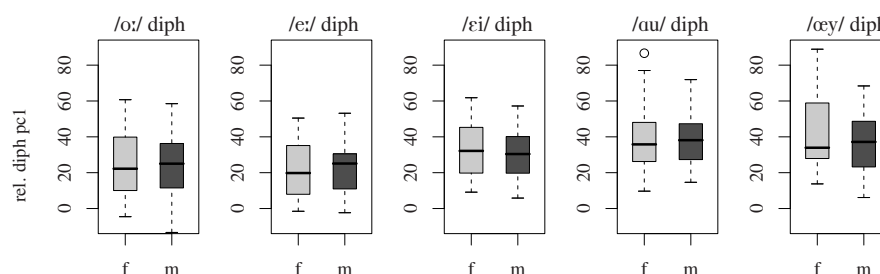


Figure 4.15: Box plots of the 35 females' (f) and 35 males' (m) relative degrees of diphthongization in pc1 (y-axis) for the five vowel phonemes.

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 /o:/ and /e:/ 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 /o:/, /au/ 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 /o:/ $p=.743$, mean difference 1.113, the 95% C.I. ranges from -5.652 to 7.878; for /e:/ $p=.737$, mean diff -0.995, C.I. -6.898 to 4.907; for /ɛ:/ $p=.576$, mean diff 1.633, C.I. -4.177 to 7.444; for /au/ $p=.638$, 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 /e:/ and /o:/, 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.

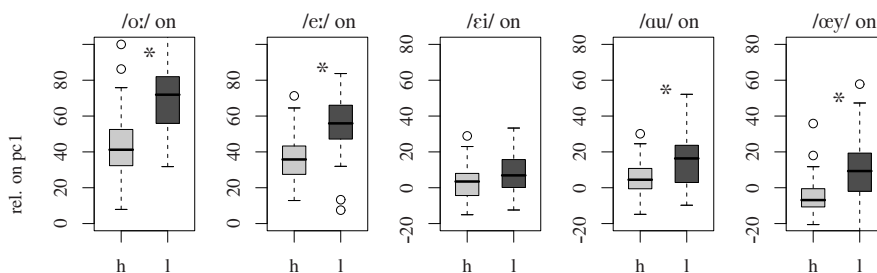


Figure 4.16: Boxplots of the 35 high (h) and 35 low (l) educated speakers' relative pc1 onset values (y-axis) for the five vowel phonemes. Both females and males. Stars indicate significance.

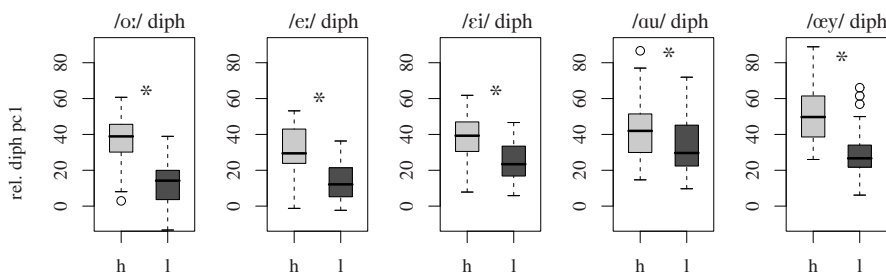


Figure 4.17: Boxplots of the 35 high (h) and 35 low (l) educated speakers' relative degrees of diphthongization in pc1 (y-axis) for the five vowel phonemes. Stars indicate significance.

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 /o:/ $p < .001$, with a mean difference of -23.420,

the C.I. ranges from -32.997 to -13.843; for /e:/ $p < .001$, mean diff -17.919, C.I. -25.762 to -10.075; for /au/ $p = .003$, mean diff -10.751, C.I. -17.572 to -3.929; for /æy/ $p = .001$, 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 /o:/ $p < .001$, with a mean difference of 23.704, the 95% C.I. ranges from 16.939 to 30.470; for /e:/ $p < .001$, 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 /au/ $p = .023$, mean diff 8.998, C.I. 1.303 to 16.694; for /æy/ $p < .001$, mean diff 21.311, C.I. 13.681 to 28.942).

The low educated speakers showed considerably less diphthongization for the long vowels /o:/ and /e:/ than for the diphthongs /ɛi/, /au/, and /æy/ (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 /o:/, /ɛi/, and /au/ 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 /o:/ $F_{1,62}=0.343$, $p=.560$; for /e:/ $F_{1,62}=0.269$, $p=.606$; for /au/ $F_{1,62}=2.187$, $p=.144$; for /æy/ $F_{1,62}=2.208$, $p=.142$; for /ei/ $F_{1,62}=1.059$, $p=.307$).

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 /ei/ and /au/ (for /e:/ ($F_{1,62}=0.731$, $p=.396$), for /o:/ ($F_{1,62}=1.214$, $p=.275$), for /ei/ ($F_{1,62}=4.270$, $p=.043$), for /æy/ ($F_{1,62}=1.729$, $p=.193$); for /au/ ($F_{1,62}=6.746$, $p=.012$)). As an example, figure 4.18 below shows the relative degree of diphthongization for /ei/ in dependence of speaker age for both levels of education; figure 4.19, p. 75 shows the diphthongization for both sexes.

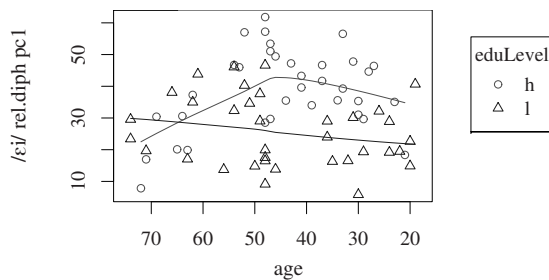


Figure 4.18: Degrees of diphthongization of the vowel phoneme /ei/ in pc1 (y-axis) according to speaker age (x-axis). Circles for the high educated speakers, triangles for the low educated speakers, with lowess curves.

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 /æy/, /e:/, and /o:/) if 'age' is linear or rather a curvilinear variable, we replaced 'age' by the squared age 'age²'. For /ei/ ($F_{1,62}=6.231$, $p=.015$)

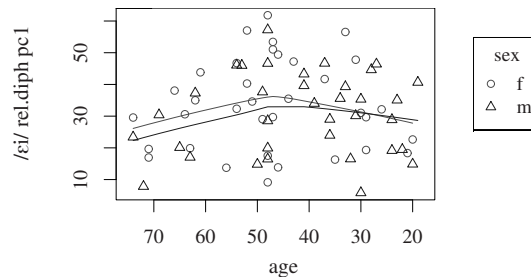


Figure 4.19: Degrees of diphthongization of the vowel phoneme /εi/ in pc1 (y-axis) according to speaker age (x-axis). Circles for the female speakers, triangles for the male speakers, with lowess curve.

and /au/ ($F_{1,62}=8.886$, $p=.004$), 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 /εi/ (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 /εi/ 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 /εi/.

Figure 4.18, page 74, displays the individual degrees of diphthongization for /εi/ 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'/'Age²' Effects within the Low Educated

Analyses of variance with 'sex', and either 'age²', or 'age' as covariate on the low educated speakers' relative onsets and the degrees of diphthongization revealed significant effects only for the diphthongization of /au/. 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 'age²' ($F_{1,35}=7.50$, $p=.010$). Thus, both, 'age' and 'age²', were significant, suggesting that the speakers' age is a linear factor when it comes to the diphthongization of /au/. 'Age' was almost of significant influence in terms of the low

educated speakers' /au/ onsets with $F_{1,35}=3.87$, $p=.058$ (for 'age²' $F_{1,35}=3.52$, $p=.070$). 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 'age²' $F_{1,35}=3.72$, $p=.063$). 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 /au/ and /œy/ over time.

'Age'/'Age²' Effects within the High Educated

For the onsets of each vowel phoneme, analyses of variance with 'sex', and either 'age', or 'age²' yielded no significant differences between the sexes, and no interaction effects of 'sex' with 'age', or 'sex' with 'age²'. 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, 'age²' turned out to significantly affect the degrees of diphthongization of /e:/ ($F_{1,35}=5.12$, $p=.031$), /o:/ ($F_{1,35}=5.37$, $p=.027$), and /ei/ ($F_{1,35}=5.83$, $p=.022$), as well as the onsets of /ei/ ($F_{1,35}=4.87$, $p=.034$).

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 /au/, 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 /e:/, /o:/, and /ei/, as well as in the onsets of /ei/, and suggests that pronunciation habits changed in a non-gradual way. The significant results for the factor 'age²' 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 /ei/ as expected by the results of previous /ei/-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 covariate ‘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).

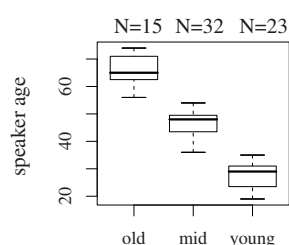


Figure 4.20: Boxplots of speaker age split into the age groups ‘old’, ‘mid’, ‘young’.

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 diphthongization 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 /e:/, /o:/, and /ei/. Considering the degrees of diphthongization, the old generation differed significantly from the mid generation for /e:/ (mid-old: $p=.005$, mean difference 11.66, the 95% C.I. ranges from 2.870 to 20.453), and /o:/ ($p=.020$, mean difference 11.866, C.I. 1.446 to 22.286), and /ei/ ($p=.004$, mean difference 11.886, C.I. 3.145 to 20.627). For /au/ and /æy/, there were no significant differences between generations. The boxplots in figure 4.21, p. 78 display the relative degrees of diphthongization of /ei/ for the three groups. As can be seen the diphthongization range of the mid and young group is considerable.

Except for /ei/, 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'.

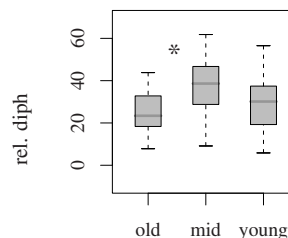


Table 4.8: Number of speakers per level of education when split into the age groups 'old', 'mid', 'young'.

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

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 diphthongization. Though some difference seems to be apparent for the diphthongization of /au/, 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:/ ($F_{1,32}=31.03$, $p<.001$), /au/ ($F_{1,32}=9.96$, $p=.004$), and /æy/ ($F_{1,32}=5.07$, $p=.032$), 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 /au/ ($F_{1,32}=7.72$, $p=.010$), and for /æy/ ($F_{1,32}=36.42$, $p<.001$), 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).

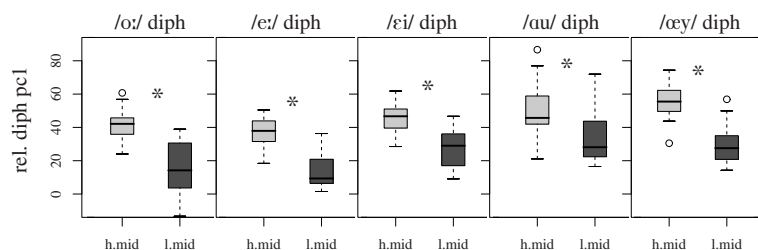


Figure 4.22: Boxplots of the relative *pc1* degrees of diphthongization (*diph*) of all speakers of the *mid* age group, split into high (*h*: light grey boxes) and low educated (*l*: dark grey boxes) speaker groups. Stars indicate significances between the levels of education within this generation.

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 /o:/ ($F_{1,23}=5.92$, $p=.025$). 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 (/o:/ $F_{1,23}=21.46$, $p<.001$; /e:/ $F_{1,23}=8.77$, $p=.008$; /ɛi/ $F_{1,23}=15.58$, $p=.002$; /au/ $F_{1,23}=5.98$, $p=.024$; /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).

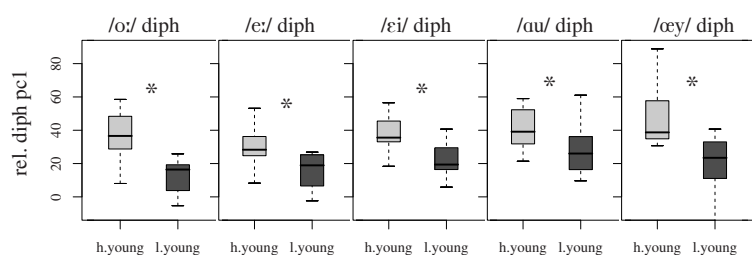


Figure 4.23: Boxplots of the relative *pc1* degrees of diphthongization (*diph*) of all speakers of the *young* age group, split into high (*h*: light grey boxes) and low educated (*l*: dark grey boxes) speaker groups. Stars indicate significant differences between the levels of education within this 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 /o:/, /ɛi/, and /au/. 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 /æy/ (table 4.9, p. 83). For /o:/, /e:/, /ɛi/, and /au/, 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 /o:/, /e:/, and /ɛi/, the old speakers differed also significantly from the young speakers.

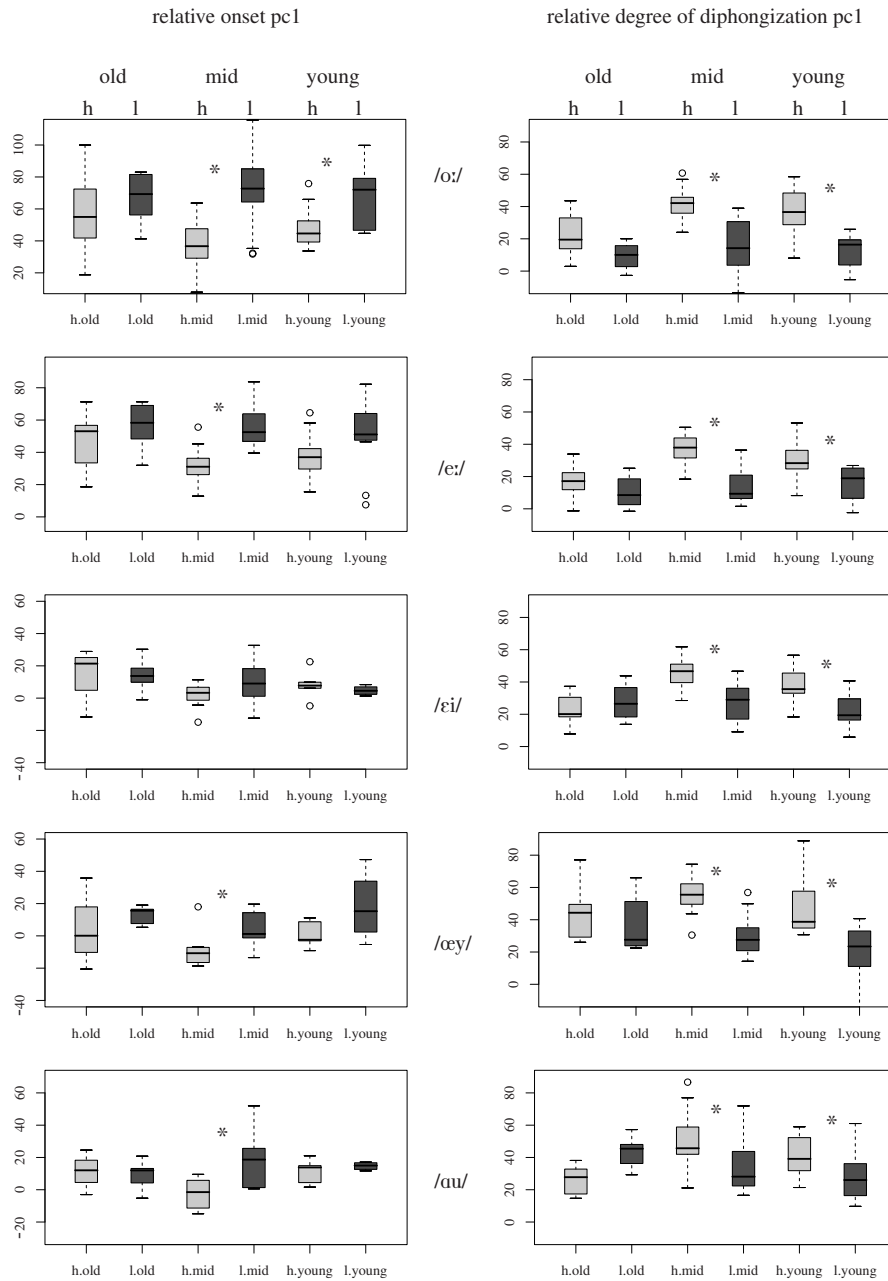


Figure 4.24: Boxplots of the *pc1* of all speakers' vowel phonemes, split into age groups 'old', 'mid', 'young', and split into high (h: light grey boxes) and low educated (l: dark grey boxes) speaker groups. Relative onsets of the vowel phonemes on the left, relative degrees of diphthongization to the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

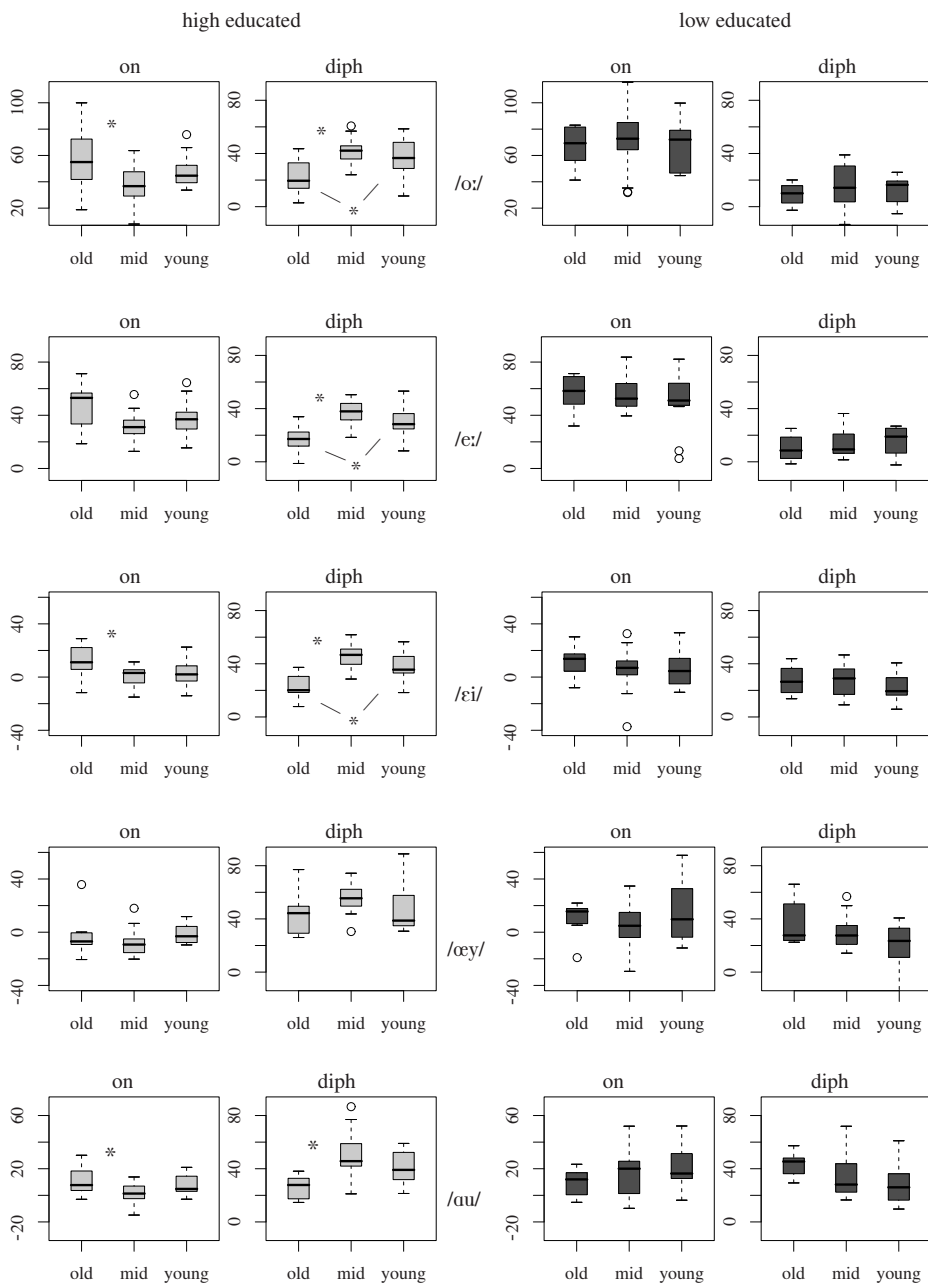


Figure 4.25: Boxplots of the relative onset (on) and relative degree of diphthongization (diph) in *pC1* of all high educated (light grey) speakers' vowel phonemes (to the left) and all low educated (dark grey) speakers' vowel phonemes (to the right), split into age groups 'old', 'mid', 'young'. Relative 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 /æy/.

rel.pc1			Multiple comparisons 'age group'				
			MeanDiff	Std. Error	Sig.	95% C.I. Lower	UpperBound
/o:/	on	old mid	20.6	7.9	.043(*)	.497	40.616
		diph	-19.6	5.3	.003(*)	-33.069	-6.084
	diph	old young	-13.8	5.7	.067	-28.338	.714
/e:/	diph	old mid	-19.9	4.7	.001(*)	-31.820	-8.120
		old young	-13.4	5.0	.037(*)	-26.176	-.660
/ei/	on	old mid	12.2	4.4	.027(*)	1.149	23.338
		diph	-21.9	4.6	.000(*)	-33.581	-10.285
	diph	old young	-14.9	4.9	.016(*)	-27.401	-2.319
/au/	on	old mid	10.6	4.2	.050	-.008	21.190
		diph	-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 /au/ and /æy/. 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 /au/ and /æy/ is a decrease in diphthongization with increasing speaker age. The same figure also shows that the non-linear significant effect of 'age²' for the high educated's diphthongization of /e:/, /o:/, and /ei/, and the onset of /ei/ 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 /ei/ show significantly different onsets and degrees of diphthongization; for /ei/, 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 /au/ 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 /e:/ (residence region $p=.020$, educational region $p=.025$), and /o:/ (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 /o:/ 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 /e:/. 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 /*ei*/, /*au*/, /*œy*/, /*o:*/, and /*e:*/, 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 /o:/ and /e:/. 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 /au/ and /æy/ were lower the younger the speakers' ages were, whereas for the high educated speakers, from older to middle aged speakers,

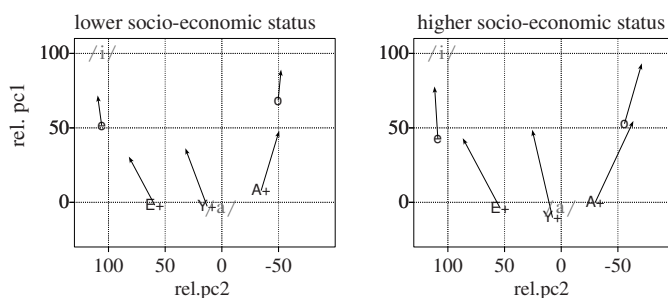


Figure 4.26: Relative pc-values of the mean vowel pronunciation patterns of the 35 low (left) and the 35 high (right) educated speakers. /a/ has a relative pc1/pc2 value of 0, and /i/ has a relative pc1/pc2 value of 100 (compare section 4.3.2). The vowels' onsets are represented by 'e' for /e:/, 'o' for /o:/, 'E+' for /ɛi/, 'Y+' for /æy/, 'A+' for /au/. 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 /o:/ and /e:/, 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 /o:/ and /e:/ take a position close to the area of the former realizations of the diphthongs /au/ and /ɛi/ (compare figure 4.27⁴, 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, /o:/ changed from [o^u] to [ɔu], /e:/ from [eⁱ] to [ɛɪ], /ɛi/ from [ɛɪ] to [æɪ], /au/ from [ʌu] to [au], and /œy/ from [ʒy] to [ɐə].

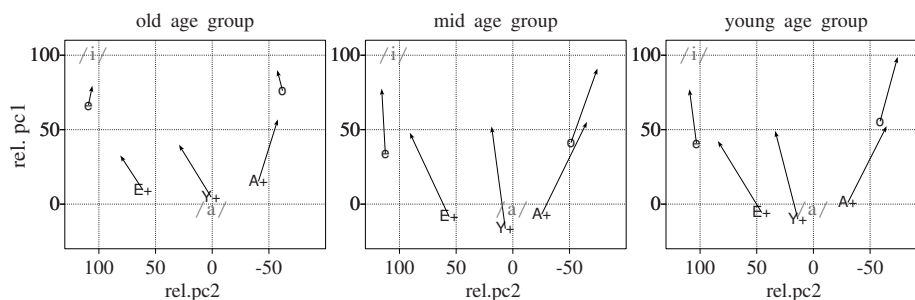


Figure 4.27: Relative *pc*-values of the mean vowel pronunciation patterns of the higher educated speakers in the old (left), mid (center), and young (right) age group. /a/ has a relative *pc1/pc2* value of 0, and /i/ has a relative *pc1/pc2* value of 100 (compare section 4.3.2). The vowels' onsets are represented by 'e' for /e:/, 'o' for /o:/, '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 /o:/ and /e:/. For the low educated speakers the degrees differed. On one hand /o:/ and /e:/ corresponded with each other, and on the other hand /ɛi/, /œy/ and /au/ (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 /au/ 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 suggest 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.