



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.

3. PRELIMINARY STUDY ON /ɛi/

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

3.1 Introduction

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

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

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

3.2 Data

Six females' and six males' realizations of /a/, /i/, /u/, /ɛ/, and /ɛi/ were taken from the IFA Speech Corpus¹ and a prerelease of the Spoken Dutch Corpus² (CGN), both recorded around the year 2000. The IFA Speech Corpus contains recordings of eight Dutch speakers in a variety of speaking styles. From this corpus, only the informally uttered and spontaneously retold speech of the seven adult speakers, four females and three males, was considered. To form a dozen, and for an equal amount of females versus males, two female and three male speakers were added from the spontaneous data pool of the then available prerelease (Release1)³ of the Spoken Dutch Corpus. The data were approached in the order given in the database, and the additional speakers were selected in such a way that the age distribution was roughly equal between males and females. Of the twelve speakers, the six female (F) speakers were aged 20, 28, 36, 40, 46 and 60, the six male (M) speakers 32, 36, 40, 54, 56 and 66.

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

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

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

¹ <http://www.fon.hum.uva.nl/Service/IFAcopus/>

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

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

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

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

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

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

3.3 Analysis

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

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

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

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

For monophthongs, the spectrum calculated at the temporal midpoint, presumably at a rather steady-state phase of the vowel, was used for further analysis. The fundamental frequency was measured using the Praat standard analysis. However, f_0 yielded no systematic differences between the lowering vs. the non-lowering variants, nor concerning their relation to the anchor vowels, and so f_0 was ignored from hereon.

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

3.3.1 Formant Analysis

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

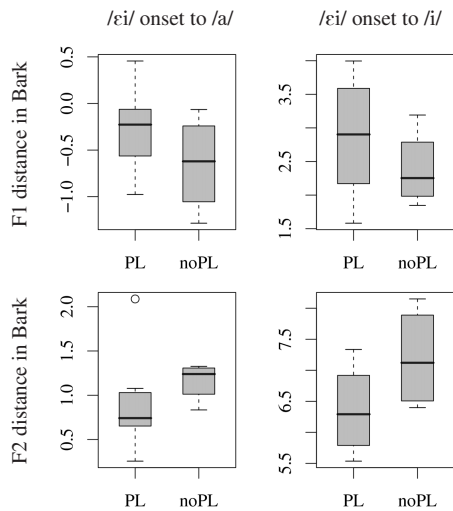
To gain a first insight into the acoustic formation of our data, the mean diphthong onset values of the twelve speakers were compared and related to their mean values of the vowels / a / and / i /. Table 3.1, p. 34, shows the speakers' mean values of F1 in Bark and F2 in Bark.

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

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

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

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



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

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

3.3.2 Durational Aspects

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

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

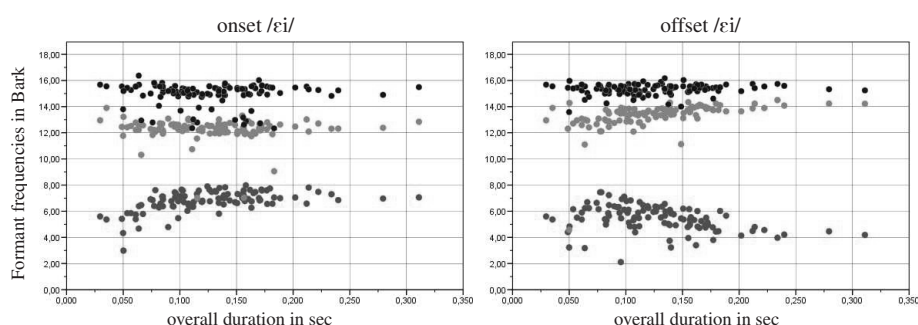


Figure 3.2: Example of one of the speakers' measured formants in /εi/ realizations: Onset (left) and offset (right). Overall diphthong duration on the x-axis with formant frequencies F1, F2, F3 in Bark on the y-axis.

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

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

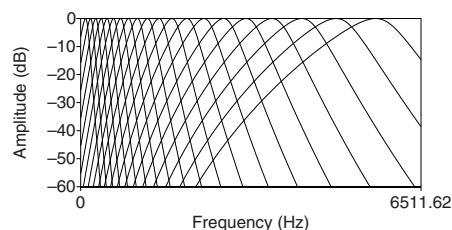
3.3.3 Bandfilter Analysis

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

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

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

Figure 3.3: Barkfilters 1 to 20 on a linear frequency scale. In the actual application, the first two filters are replaced by their mean to reduce the impact of f_0 .



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

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

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

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

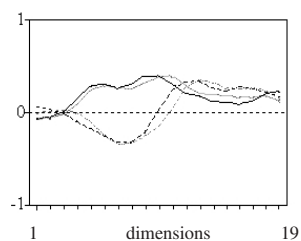


Figure 3.5: Eigenvectors 1 (I), 2 (II) and 3 (III) of the 19 dimensions of the PCA on all speakers' means of bark-filtered /a/, /i/, /u/.

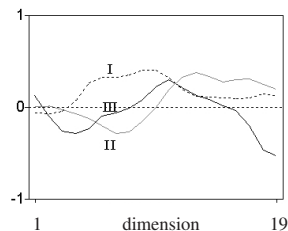
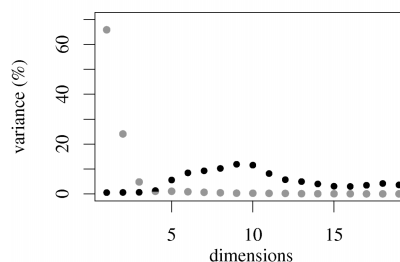


Figure 3.6: The explained variance in the 19 old and new dimensions in percentage. Black dots: explained variance in the original dimensions (barkfilters). Grey dots: explained variance in the new dimensions (pc's).



3.3.4 Comparing Formants and Principal Components

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

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

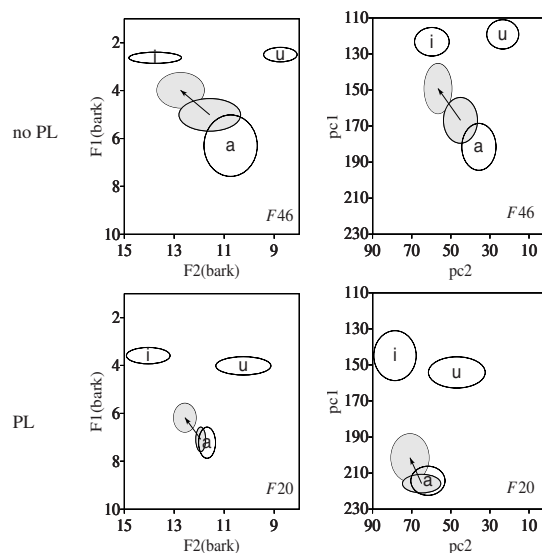
	F1 _{Bark}	F2 _{Bark}	F3 _{Bark}	expl. var.
pc1	+0.81	-0.12	+0.26	65%
pc2	-0.08	+0.70	+0.10	25%
pc3	-0.19	+0.05	-0.15	5%

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

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

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

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



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

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

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

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

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

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

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

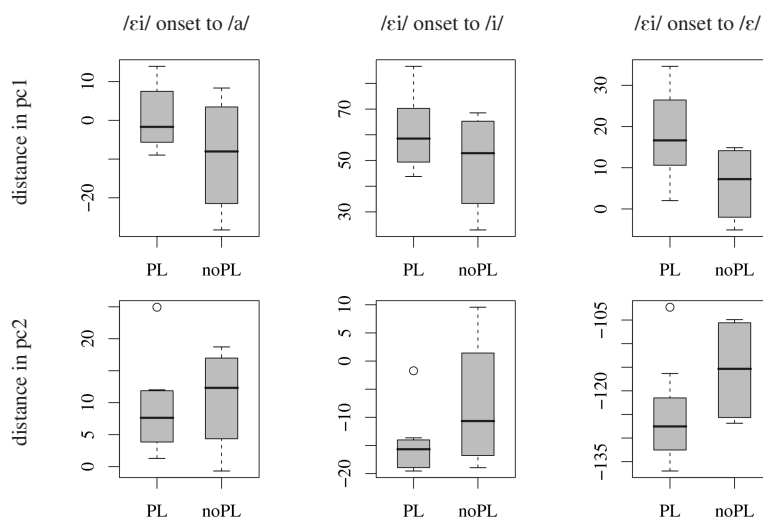
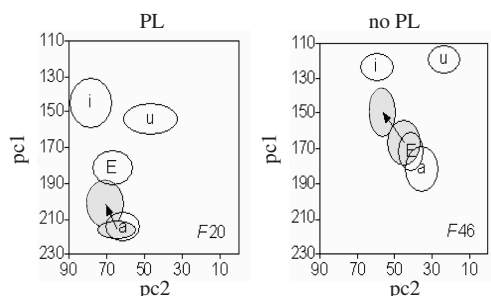


Figure 3.8: Top row from left to right: boxplots on the distance of /ɛi/ to /a/ (/ɛi/-/a/), /i/ (/ɛi/-/i/), and /ɛ/ (/ɛi/-/ɛ/) in pc1. Bottom row from left to right: boxplots on the distance of /ɛi/ to /a/, /i/, and /ɛ/ in pc2. 'PL' for the speaker group that was perceived to lower /ɛi/; 'noPL' for the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

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

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



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

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

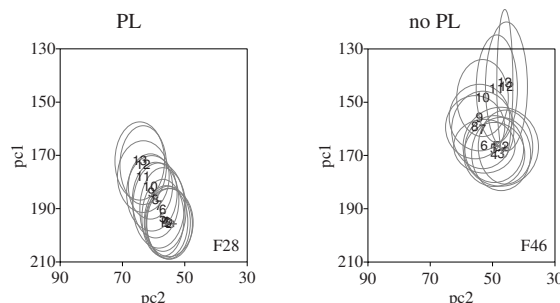
3.3.6 Temporal Diphthong Structure

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

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

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

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



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

3.4 Summary

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

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

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

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

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

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

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

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

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

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

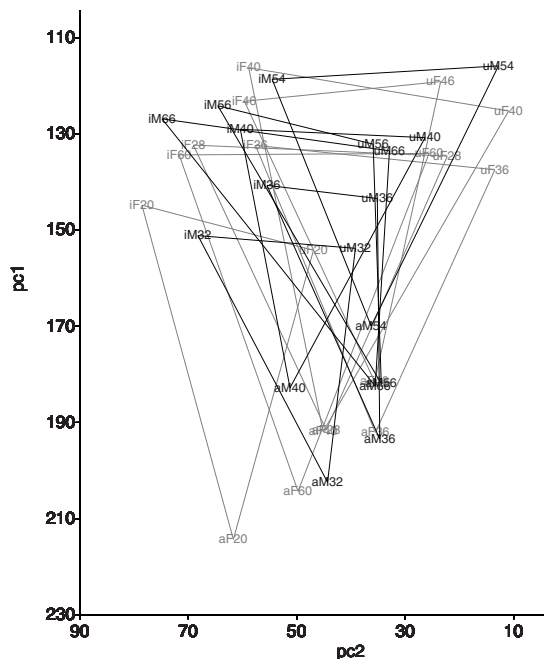


Figure 3.11: $PC1$ - $PC2$ plane with / a /-/ i /-/ u / triangles of the six female (gray) and six male (black) speakers.

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

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

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