



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.

5. PERCEPTUAL DISSIMILARITY OF ACOUSTIC DIFFERENCES

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

5.1 Introduction

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

In the perception task described in the following sections, listeners had to give 'different' or 'similar' judgments on realizations of the same vowel phonemes in words taken from the spontaneous speech of six speakers. If the listeners' perception somehow corresponds to the significant acoustics described in the previous chapter, they should at least be able to differentiate rather extreme realizations of the acoustically defined vowel variants: within each phoneme, a strongly diphthongized vowel with a low onset versus a slightly diphthongized vowel with a high onset would represent the two extremes of the measured variation continuum of the data analyzed in the previous chapter. We expected all listeners to judge these two realizations as differing in quality. Though in general, vowel onset and degree of diphthongization correlated positively in the acoustic data of the speakers in the previous chapter (the lower the onset, the stronger the diphthongization), one of the two might be more speaker-dependent and/or auditorily more salient. Also, other values such as duration and f_0 might be relevant, even if they were insignificant in the acoustic analysis of the previous chapter.

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

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

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

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

5.2 Method and Material

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

5.2.1 Stimuli

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

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

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

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

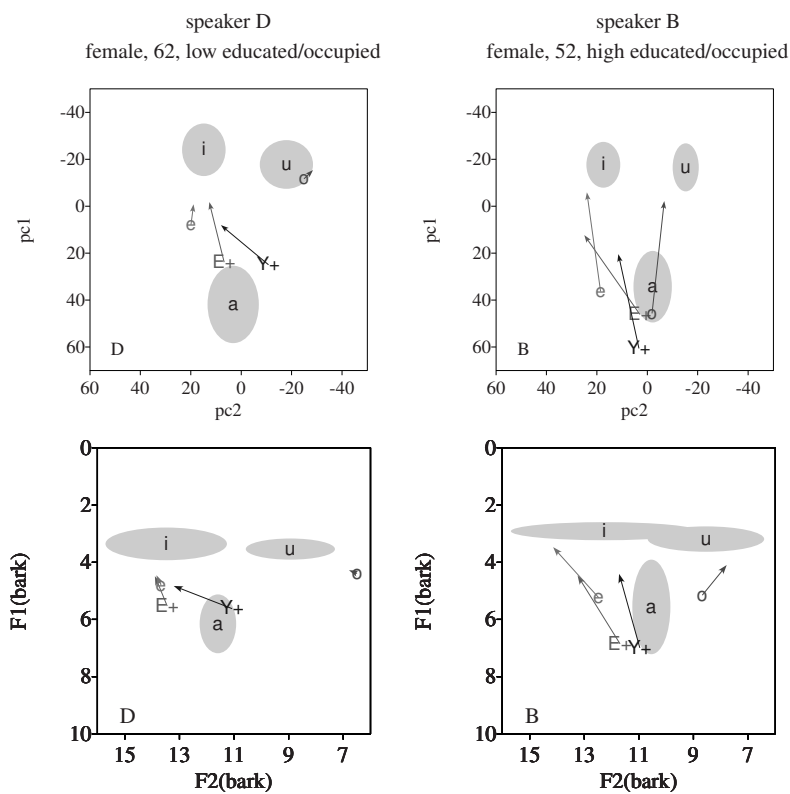


Figure 5.1: *Pc1/pc2* (top) and *F1/F2* (bottom) dimensions of the vowels of the two in acoustic terms most contrastive speakers: The plotted anchor vowels /a,i,u/ with sigma ellipses were based on all available sound segments in the corpus of speakers that was used in the previous chapter. The displayed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

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

Table 5.2: Stimulus words per speaker and vowel phoneme

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

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

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

5.2.2 Procedure

Regarding previous positive experiences by phoneticians of our institute (most recently see Jongmans, 2008 [67]), we preferred to run the experiment online. To reach more listeners, the online format should demand as low an expenditure of time as possible from participants. After a pre-trial, the stimuli were presented via a web-interface¹ (compare figures 5.2 and 5.3, p. 97). First, the listeners were asked to state their age (*'leeftijd'*), sex (*'vrouw' / 'man'*), highest education (*'Wat is uw hoogst genoten opleiding'*), and whether they resided in one of the cities of the 'randstad' (*'Woont u in een van de randsteden'*). Next, they were asked about hearing loss (*'Is gehoorverlies bij u bekend'*), whether Dutch was their mother tongue (*'Is Nederlands uw moedertaal'*), and whether they had ever been phonetically trained (*'Heeft u een opleiding gehad in de fonetiek'*).

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

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

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

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

Luisterexperiment

Leeftijd: vrouw man

Is gehoorverlies bij u bekend?: ja nee

Is Nederlands uw moedertaal?: ja nee

Woont u in een van de randsteden?: ja nee

Wat is uw hoogst genoten opleiding?: wo hbo vwo havo mbo mavo vmbo lbo

Heeft u een opleiding gehad in de fonetiek?: ja nee

Tijdens dit experimentje gaat u luisteren naar een aantal woorden uit de spontane spraak van verschillende personen.

Bij ieder voorbeeld kunt u het geluid horen door op de tekst-knop te klikken. U moet daarna uw antwoord geven in een van de antwoordvelden.

U kunt de voorbeelden meerdere keren beluisteren. Door middel van de 'volgende' knop, gaat u naar de volgende site waar u weer hetzelfde doet.

Verder is het belangrijk dat u in een zo stil mogelijke omgeving zit en gebruik maakt van een koptelefoon. Zet het geluidsnivo op een level dat u prettig vindt.

Uw data worden uiteraard geanonimiseerd. Mocht u vragen hebben, dan kunt u mailen naar i.jacobi@uva.nl

Alvast bedankt!

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

my_name: 60 nog volgende vragen

Instructie

Bij dit experiment gaat u twee sprekers vergelijken.
 U moet hierbij letten op de uitspraak van de klinkers die in GROTE LETTERS zijn aangegeven.
 Na het luisteren geeft u aan of de twee sprekers overeenkomen (gelijk) in hun uitspraak van de klinkers, of niet (verschillend).

hOOgste bOvenop gekOOkt |

nOdig geslOten diplOma |

Verschillend Gelijk

volgende |

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

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

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

5.3 Results

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

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

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

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

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

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

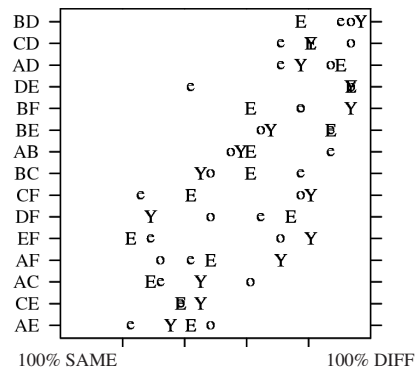


Figure 5.4: 30 listeners’ mean response to vowel phoneme realizations of the 15 speaker pairings. ‘e’ represents /e:/, ‘o’ /o:/, ‘E’ /ɛi/, and ‘Y’ /æy/.

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

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

/ɛi/	DIFF	SAME	/æy/	DIFF	SAME	/e:/	DIFF	SAME	/o:/	DIFF	SAME
DE	** 94	06	BD	** 97	03	BD	** 90	10	BD	** 94	06
AD	** 90	10	DE	** 94	06	BE	** 87	13	DE	** 94	06
BE	** 87	13	BF	** 94	06	AB	** 87	13	CD	** 94	06
CD	** 81	19	EF	** 81	19	BC	** 77	23	AD	** 87	13
BD	** 77	23	CD	** 81	19	BF	** 77	23	BF	** 77	23
DF	* 74	26	CF	** 81	19	CD	* 71	29	CF	** 77	23
AB	61	39	AD	** 77	23	AD	* 71	29	EF	* 71	29
BC	61	39	AF	* 71	29	DF	65	35	BE	65	35
BF	61	39	BE	68	32	DE	42	58	AC	61	39
AF	48	52	AB	58	42	AF	42	58	AB	55	45
CF	42	58	AC	45	55	CE	39	61	AE	48	52
AE	42	58	BC	45	55	AC	32	68	BC	48	52
CE	39	61	CE	45	55	EF	29	* 71	DF	48	52
AC	29	* 71	AE	35	65	CF	26	* 74	CE	39	61
EF	23	** 77	DF	29	* 71	AE	23	** 77	AF	32	68

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

5.3.1 Overall Response Behavior and Acoustic Distances

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

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

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

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

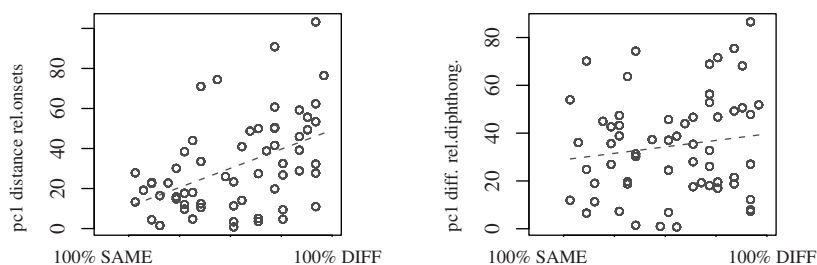


Figure 5.5: The distance of the relative onsets (left) and the relative degrees of diphthongization (right) in the pc1 dimension of all 60 stimuli versus the mean same-different ratio of the 30 listeners.

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

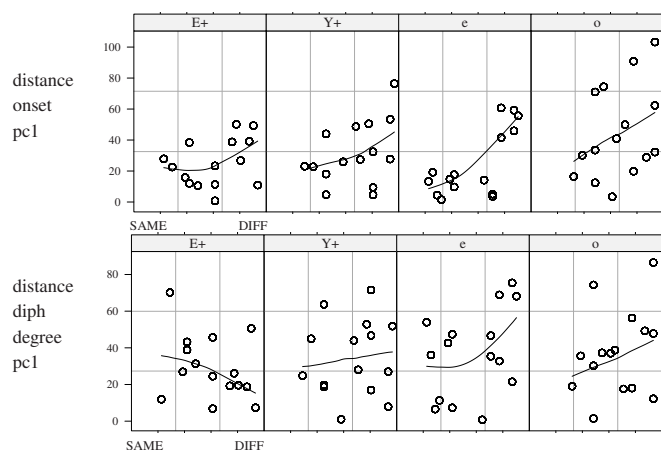


Figure 5.6: Listener mean response (*x*-axis) to the phoneme realizations of the 15 speaker-pairings versus acoustic distances (*y*-axis). 'E+' for /*ɛi*/, 'Y+' for /*æy*/, 'e' for /*e*/, 'o' for /*o*/.

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

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

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

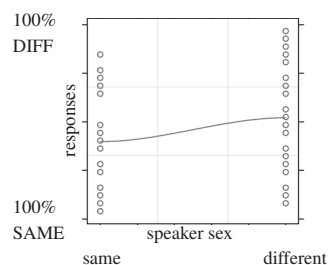


Table 5.5: Logistic regression coefficients with standard errors of the best fitting models for the response behavior to acoustic distances for /e:/.

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

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

5.3.2 Age Dependent Response Behavior and Acoustic Distances

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

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

some variation in response behavior.

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

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

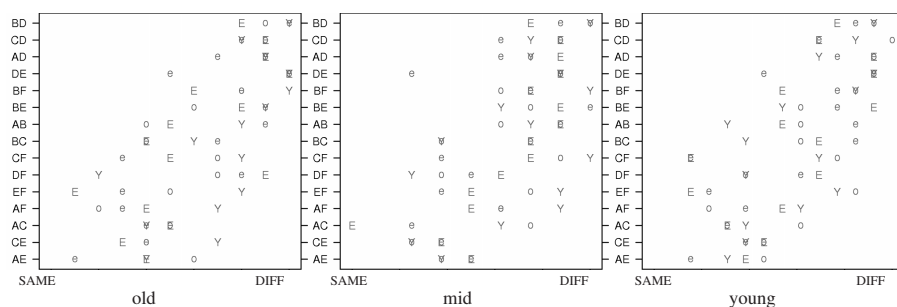


Figure 5.8: Mean response (*x*-axis) per listener age group (*old*, *mid*, *young*) to each vowel phoneme of the 15 speaker pairs. The speaker pairs on the *y*-axis are ordered according to the listeners’ overall mean response, with *BD* being judged as most different and *AE* as most similar.

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

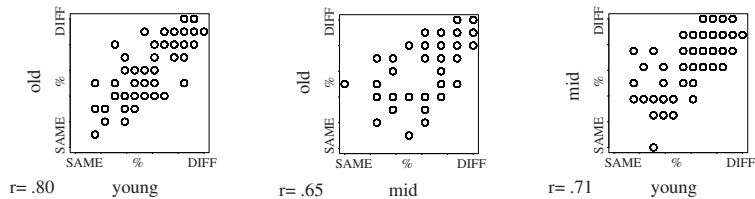


Figure 5.9: Mean responses and correlation of the three listener age groups (old, mid, young) to all 60 stimulus pairings.

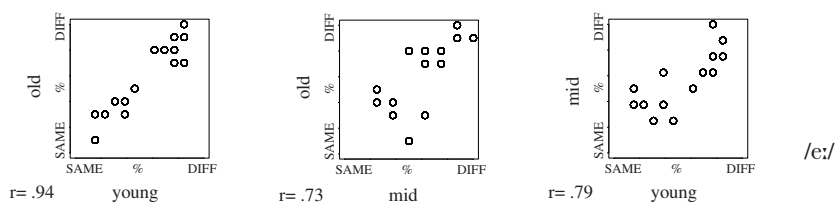


Figure 5.10: Mean responses and correlation of the three listener age groups to the 15 /e:/ stimulus pairings.

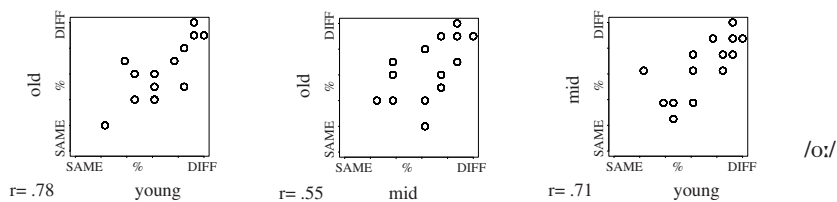


Figure 5.11: Mean responses and correlation of the three listener age groups to the 15 /o:/ stimulus pairings.

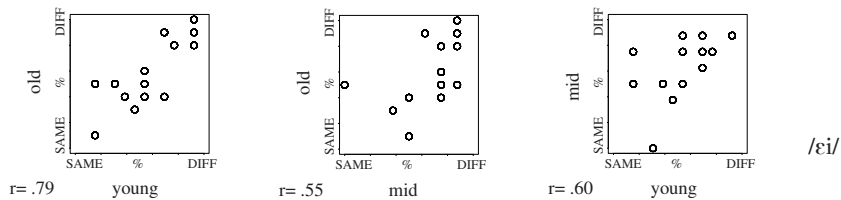


Figure 5.12: Mean responses and correlation of the three listener age groups to the 15 /ɛi/ stimulus pairings.

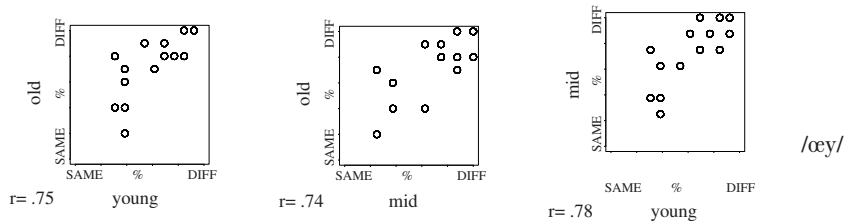


Figure 5.13: Mean responses and correlation of the three listener age groups to the 15 /œy/ stimulus pairings.

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

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

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

5.4 Summary

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

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

distances in vowel realizations, the predictability of the response behavior had to be investigated separately for each phoneme. /e:/ was the only phoneme for which listener responses could be predicted by the stimulus distance in the pɔ1 onsets. As for the other vowel phonemes, /o:/, /ɛi/, and /æy/, even though the acoustic distances in pɔ1 onsets and degrees of diphthongization between the speakers added significantly to the predictability of the response behavior in the regression models, they could explain only little of the variance in the outcome.

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

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

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

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

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