An acoustic description of the vowels of Northern and Southern Standard Dutch

Patti Adank[a)]
School of Psychology, University of Liverpool, Bedford Street South, L69 7ZA, Liverpool, United Kingdom

Roeland van Hout
Department of Linguistics, University of Nijmegen, PO box 9103, 6500 HD, Nijmegen, The Netherlands

Roel Smits
Max Planck Institute for Psycholinguistics, PO box 310, 6500 AH Nijmegen, The Netherlands

(Received 9 December 2003; revised 1 May 2004; accepted 15 June 2004)

A database is presented of measurements of the fundamental frequency, the frequencies of the first three formants, and the duration of the 15 vowels of Standard Dutch as spoken in the Netherlands (Northern Standard Dutch) and in Belgium (Southern Standard Dutch). The speech material consisted of read monosyllabic utterances in a neutral consonantal context (i.e., /sVs/). Recordings were made for 20 female talkers and 20 male talkers, who were stratified for the factors age, gender, and region. Of the 40 talkers, 20 spoke Northern Standard Dutch and 20 spoke Southern Standard Dutch. The results indicated that the nine monophthongal Dutch vowels /a ɑ ɛ i ɪ ɔ u y ʏ/ can be separated fairly well given their steady-state characteristics, while the long mid vowels /e o ø/ and three diphthongal vowels /ɛɪ ɔu œy/ also require information about their dynamic characteristics. The analysis of the formant values indicated that Northern Standard Dutch and Southern Standard Dutch differ little in the formant frequencies at steady-state for the nine monophthongal vowels. Larger differences between these two language varieties were found for the dynamic specifications of the three long mid vowels, and, to a lesser extent, of the three diphthongal vowels. [DOI: 10.1121/1.1779271]

PACS numbers: 43.70.Fq, 43.70.Kv, 43.72.Ar [AL] Pages: 1729–1738

I. INTRODUCTION

This paper presents a description of the acoustic characteristics of the 15 vowels of Standard Dutch as spoken in the Netherlands and in Belgium (Flanders). Previous descriptions of the vowels of Standard Dutch are given in Pols *et al.* (1973) and Van Nierop *et al.* (1973).[1] These studies describe the Dutch vowels in terms of the average frequencies and standard deviations of the first three formants. Pols *et al.* recorded 1 token of the 12 monophthongal vowels in an /hVt/ context, produced by 50 male talkers, and Van Nierop *et al.* recorded one token of each of the 12 monophthongal vowels in an /hVt/ context, produced by 25 female talkers.

The acoustic measurements by Pols *et al.* (1973) and Van Nierop *et al.* (1973) have been recognized as representing the standard formant values for the vowels of Dutch and were used in studies on vowel perception and talker normalization (e.g., Disner, 1980; Syrdal and Gopal, 1986). However, in our view the description of the acoustic characteristics of the vowels of Dutch as presented by Pols *et al.* and Van Nierop *et al.* is limited in four respects:

(a) Pols *et al.* and Van Nierop *et al.* do not provide information about dynamic properties such as vowel duration and spectral change (formant measurements were presented for a single time slice); information that has been found to play a central role in vowel perception (Di Benedetto, 1989a, b; Hillenbrand and Gayvert, 1993; Nearey, 1989, Strange *et al.* 1983; Strange, 1989), and that reflects differences in language varieties (Hagiwara, 1997).

(b) Not all vowels of Dutch were included by Pols *et al.* and Van Nierop *et al.*, as no description was given of the acoustic characteristics of the three diphthongs /ɛɪ ɔu œy/.

(c) No information is given in the two studies about the regional background of the talkers; the formant measurements described in Pols *et al.* and Van Nierop *et al.* may therefore display uncontrolled regional variation.[2]

(d) No recordings were made of talkers from Flanders, the Dutch-speaking part of Belgium, thus excluding the southern variety of Standard Dutch from the description.

The purpose of the present study was twofold. First, we aim to give a description of the Dutch vowel sounds. In this description we want to improve on the four limitations mentioned above. Second, we aimed to give an overview of the similarities and dissimilarities of the vowel systems of Northern Standard Dutch and Southern Standard Dutch.

The present study was set up as follows. A total of 1200 vowel tokens were recorded, spoken by 40 talkers, 20 male talkers and 20 female talkers. These talkers were screened for their age, gender, regional background, and socioeconomic status. Each talker produced two tokens of all 15 vowel sounds of Dutch, /ɑ a ɛ ɪ i ɔ u ʏ y e o ø ɛɪ ɔu œy/. Each token was produced in a carrier sentence, in a /sVs/

a)Electronic mail: adank@liv.ac.uk

context. For each token, measurements were made of the duration, the fundamental frequency, and the first three formants at nine time slices in the vowel duration.

II. MATERIALS

A. Database design

Our materials were taken from a large database, comprising recordings of 160 talkers of standard Dutch who were stratified for the following sociological variables: speech community (country), regional background, gender, and age. All talkers were teachers of Dutch at secondary education institutes at the time the interview was recorded. They were required to be teachers of the Dutch language for three reasons. First, Dutch teachers are professional language users; they are expected to speak standard Dutch on a daily basis. Second, they are instructors of the standard language and can thus be regarded as having a normative role. Third, Dutch teachers' speech was expected to show more regional variation than broadcasters', whose speech is used in other pronunciation studies of the standard language (cf. Bell, 1983; Deterding, 1997).

There is a long-standing discussion whether the speech communities of the Netherlands and Flanders have one or two Dutch standard languages. It is evident that there is a Netherlandic and a Flemish standard variety, with clear differences on all linguistic levels, certainly on the phonetic level (cf. Van de Velde *et al.*, 1997). The historical background of the development of the Dutch standard language in relation to the speech communities of the Netherlands and Flanders is too complex to summarize here. Van de Velde *et al.*, (1997) present a short overview, including the relevant literature. Research data available produce convincing evidence that many differences exist between the two speech communities on the level of pronunciation. The data set was therefore split into a Dutch and a Belgian component, referred to as Northern Standard Dutch (NSD) and Southern Standard Dutch (SSD), respectively.

In each community, four regions were selected: a central region, an intermediate region, and two peripheral regions. The central region is the economically and culturally dominant region in each of the speech communities. For the Netherlands, the central region is the west, consisting of the provinces of Northern-Holland, Southern-Holland, and Utrecht, also known as the ''Randstad,'' referred to as ''N-R'' (Netherlands-Randstad). The cities Amsterdam, Rotterdam, Utrecht, and The Hague are part of the Randstad. In Flanders, the central region is ''Brabant,'' or ''F-B'' (Flanders-Brabant). Brabant comprises the Belgian provinces ''Antwerpen'' and ''Flemish-Brabant,'' with the cities of Antwerpen and Leuven, respectively.

In this paper, we limit our description of the acoustic characteristics of the vowel tokens produced by the 40 Dutch teachers from the two ''central'' regions N-R and F-B, respectively. Therefore, whenever we refer to Northern Standard Dutch, the N-R region is meant, and whenever we refer to Southern Standard Dutch, the F-B region is meant. See Adank (2003) for a description of the remaining six regions.

Several towns were selected per region, following three criteria. First, the selected towns in each region had a comparable socioeconomic profile. Second, the towns within each region belonged to the same dialect group. Third, the Dutch spoken in that town was regarded as characteristic of that region. No major cities were selected, because it was expected that the Dutch spoken in major cities is influenced by dialects (or languages) other than the dialects spoken in the surrounding region, due to migration. For N-R, two towns were selected: Alphen aan de Rijn and Gouda; for F-B the two towns selected were Lier and Heist-op-den-berg.

The teachers who participated in the interview taught at schools for secondary education in the selected towns. They had to meet the following requirements. First, at the time of the interview, they all lived in one of the selected towns, or near that town in the dialectal region characteristic for that region. Second, they were born in the region or moved there before their eighth birthday. Third, they had lived in the region for at least eight years prior to their 18th birthday. This last requirement was formulated on the basis of studies by Payne (1980) and by Scovel (1988). Payne stated that children younger than 8 years old have no difficulty acquiring the phonological system of the place they moved to. Scovel reported that learners of a second language generally do not acquire near-native pronunciation of this language after puberty. This last requirement was therefore used to make sure that the talkers had lived in the town/region from an age at which they had no difficulties in learning the language variety spoken in that region/town. Finally, the talkers were divided into two age groups, a younger group and an older group. The talkers in the younger group were between 22 and 44 years old at the time of the interview and talkers in the older group were between 45 and 60 years old. NSD and SSD thus both consisted of 20 talkers: five young men, five older men, five younger women, and five older women.

B. Recordings

Dutch vowels are traditionally divided into phonologically short vowels, /ɑ ɛ ɪ ɔ ʏ/, phonologically long vowels /a e i ø o u y/, diphthongs, /ɛɪ ɔu œy/, and schwa, /ə/ (Booij, 1995). Recordings were made for all vowels except for schwa, because it does not occur in stressed syllables in Dutch.[3] All target vowels were produced in a carrier sentence. The sentences had the following generic structure for the short vowels (''V'' indicates the target vowel):

In sVs en in sVsse zit de V
/ɪn sVs ən ɪn sVsə zɪt də V/
[*In sVs and in sVsse is the V*]

The sentences had the following structure for the long vowels and the diphthongs:

In sVs en in sVze zit de V
/ɪn sVs ən ɪn sVzə zɪt də V/
[*In sVs and in sVze is the V*]

Of the three different consonantal contexts (CVC, CVCV, or V), the CVC contexts were selected for further processing. The CVC structure /sVs/ can be regarded as a neutral context for Dutch vowels. It was decided not to use the ''traditional'' /hVt/ structure (/hVd/ for English), because it could not be predicted how people from Flanders would

pronounce the word-initial /h/. Almost all people interviewed in Flanders are originally dialect talkers; they usually speak Southern Standard Dutch as well as their (native) dialect. Some Flanders dialects do not have /h/ in their phoneme inventory. When asked to produce a word-initial /h/, Flemish dialect talkers tend to replace /h/ by either a glottal stop, or by /χ/. The /sVs/ context was adopted to make sure that no additional sources of variation were added to the data.

The vowel tokens were recorded as a task in a so-called ''sociolinguistic interview'' in which vowels and consonants were elicited in a wide variety of tasks. The carrier sentences were presented to the talker on a computer screen, with a three-second interval between sentences. When the talker made a mistake, the interviewer interrupted the computer program and went back at least two sentences and asked the talker to read these sentences again, to make sure that all sentences were recorded correctly. The carrier sentences task was performed twice during the interview; each vowel token was therefore available twice in each syllabic structure. A total of 1200 vowel tokens were recorded: two tokens of each of the 15 vowel categories of Dutch, produced by 40 talkers.

The recordings were made on DAT with a TASCAM DA-P1 portable DAT-recorder, with an AKG C420 Headset condenser microphone. The recordings were digitized through a Lucid Technology PCI24 digital audio card, and stored at 48 kHz on a PowerMac 7500/100. The neutral context sentences were down-sampled to 16 kHz, using an non-causal FFT anti-aliasing filter set to 8 kHz.

Recording conditions were different for each of the talkers. Some were interviewed in an empty classroom and others were interviewed at their own home. Due to these differences in recording conditions, in rare cases, background noises were audible. Whenever this was the case, the speech segment was excluded from further analysis.

C. Acoustic measurements

1. Duration

The start and end times for the duration of each token were labeled manually in the digitized speech wave, using the program Praat, version 4.02 (Boersma, 2001). We define the duration as the interval between the onset and offset of the glottal vibrations of the vocalic portion of the /sVs/ syllable. When labeling, it was ensured that the surrounding speech sounds were not audible in the remaining signal. Segment labels were placed at zero crossings. The duration of each vowel segment was defined as the interval between the segment labels at the start and end of the vocalic portion.

2. Fundamental frequency

For each vowel token, $F0$ was extracted automatically with Praat using an autocorrelation-based procedure that was evaluated as the best option available in Praat (Boersma, 1993). The upper and lower limits for the autocorrelation peaks were set at 50 and 300 Hz for male talkers and at 100 and 500 Hz for female talkers, respectively. The size of the Hanning window was chosen such that it included at least three glottal pulses and was set to 60 ms for the male talkers and to 33.3 ms for the female talkers. $F0$ was extracted at the temporal mid-point of the vowel token's duration.

The resulting measurements were checked for outliers (cases 1.5–3 times the interquartile range, or IQR), and extreme values (cases >3 IQR) for each talker. Every outlier and extreme value was manually verified. Eleven cases were replaced by the mean $F0$ for the talker, usually because voice characteristics (e.g., hoarseness) of the talker prevented reliable $F0$-estimation.

3. Formant contours

The formant contours for $F1$, $F2$, and $F3$ were estimated using a program for automatic formant tracking (Nearey *et al.*, 2002).[4] Nearey *et al.*'s program consists of two parts: a formant tracking algorithm and a user interface that allows the user to verify, and, where necessary, adjust the formant tracks generated by the tracking algorithm. Its preprocessing consists of applying a 25-ms cosine4 window with a time-step of 2 ms. Subsequently, three formant candidates are estimated by means of root extraction, using a version of Markel and Gray's (1976) ''FORMNT'' algorithm, followed by a five-point running median smoothing. The number of LPC coefficients is fixed at nine. For each vowel token, these settings were identical.

Some settings for Nearey *et al.*'s tracking algorithm are to be chosen by the user. The tracking program cycles through several cutoff frequencies for each individual vowel token. The frequency range consists of two parts, a lower range and an upper range. The lower range is fixed between 0 and 3000 Hz, while the upper range is to be set by the user between 3000 Hz and the highest cutoff frequency (hc) to be evaluated. Within this upper range (3000-hc) the user must also provide the total number of cutoff frequencies (nc) to be evaluated. Finally, the user has to decide whether the distance (d) between the cutoff frequencies is spaced logarithmically (Log) or linearly (Hz) across the upper range.

We evaluated several combinations of settings for hc, nc and d and found that the best results were obtained with hc set to 4200 Hz, nc set to 5, and d set to Log (Adank, 2003). Formant tracks for all vowel tokens were measured with these settings.

The resulting formant tracks for each vowel token were verified by hand in the user interface of the program. Here, the tracks were plotted on the smoothed spectrogram. The course of each track could be altered, which was done in 20%–25% of the cases.

Once the experimenter was satisfied with the position of the formant tracks, the frequencies of the first three formant tracks were stored at nine points ($t1-t9$) of the vowel's duration, with the first point, $t1$, at the start of the vocalic portion of the token, and the ninth point, $t9$, at the end of the vocalic portion, and the remaining points spaced at equal-sized intervals, relative to the absolute duration of the token. Whenever we refer to the vowel's temporal midpoint, the fifth point of the total of nine points, or $t5$, is meant. We only present results for points $t3$ (25%) through $t7$ (75%), because we suspected that, the measurements at that point were influenced less by the consonantal context than at $t1$ and $t9$, or even at $t2$ and $t8$.

TABLE I. Average durations (ms), fundamental frequencies (Hz), and formant frequencies (Hz) for the first three formants for the vowels tokens produced by the 20 NSD talkers and the 20 SSD talkers. F, female talkers; M, male talkers, Dur, duration. *N* per cell is 20. *F*0, *F*1, *F*2, and *F*3 were sampled at 50% (*t*5) of the steady-state portion of each vowel token.

			/ɑ/	/a/	/ɛ/	/e/	/ø/	/ɪ/	/i/	/ɔ/	/u/	/o/	/ʏ/	/y/	/ɔu/	/ɛɪ/	/œy/
NSD	Dur	F	94	214	101	177	184	89	92	96	98	183	89	96	205	199	200
		M	96	203	95	181	184	82	94	90	98	184	88	93	195	192	189
	*F*0	F	226	194	220	207	201	221	248	218	249	201	246	245	201	200	201
		M	149	134	154	131	142	154	157	152	164	139	154	162	127	133	136
	*F*1	F	758	912	535	442	445	399	294	419	286	445	417	305	715	659	693
		M	578	670	475	400	375	361	278	402	259	412	366	259	580	543	574
	*F*2	F	1280	1572	1990	2343	1713	2276	2524	918	938	964	1830	1918	1280	2097	1679
		M	1172	1425	1739	1995	1563	1919	2162	821	805	929	1595	1734	1117	1804	1515
	*F*3	F	2895	2852	2871	2908	2550	2883	2911	3013	2736	2417	2711	2635	2586	2816	2618
		M	2435	2485	2492	2583	2241	2536	2665	2851	2253	2306	2345	2205	2284	2498	2280
SSD	Dur	F	107	240	101	192	200	88	147	97	128	210	89	153	245	238	241
		M	90	204	86	169	175	76	96	83	99	182	77	109	212	199	212
	*F*0	F	225	203	224	219	217	256	234	233	237	215	249	236	196	202	201
		M	126	116	128	119	121	135	148	136	149	125	138	144	113	124	114
	*F*1	F	725	868	581	436	439	455	317	475	321	418	457	337	696	670	696
		M	555	717	475	384	374	364	278	398	266	369	353	265	549	545	583
	*F*2	F	1262	1640	1932	2420	1804	2115	2647	987	1019	968	1785	2077	1282	2159	1762
		M	1066	1429	1616	1993	1539	1745	2179	850	978	862	1492	1825	1127	1779	1484
	*F*3	F	3041	3031	2978	3021	2666	2948	3312	3133	2871	2992	2884	2634	2808	3040	2855
		M	2655	2651	2572	2616	2377	2566	2787	2665	2422	2540	2514	2348	2444	2533	2415

III. RESULTS

Table I gives the average measurement results for the true monophthongal vowels, for the five acoustic variables (duration, *F*0, *F*1, *F*2, and *F*3), taken at the temporal midpoint *t*5 (50%), for the NSD and the SSD talkers. The average values at *t*3 (25%) and *t*7 (75%) for /ɔu œy ɛɪ/ and the long mid vowels /e ø o/ are displayed in Table II.

1. Duration

Figure 1 shows the average durations for all 15 vowels for the male talkers and Fig. 2 for the female talkers. A first observation is that the vowels in Figs. 1 and 2 can roughly be divided into two groups based on their duration: long ones, i.e., the three diphthongal vowels /ɛɪ ɔu œy/, the three long mid vowels /e o ø/ plus /a/, and short ones, i.e., all full monophthongal vowels /ɑ ɛ i ɪ ɔ u y ʏ/ except /a/. A series of *t*-tests was carried out to establish which vowel pairs differed significantly in duration. A total of 105 pairwise comparisons were carried out for the 15 vowels. These analyses were done separately for the NSD men, the NSD women, the SSD men, and the SSD women. Because of the large number of analyses, a Bonferroni correction was carried out, setting the significance level to $\alpha = 0.001$. The results showed that /a e ɛɪ ɔu o ø œy/ all have significantly longer durations than /ɑ ɛ ɔ u y ʏ/.

Figures 1 and 2 further show longer durations for female SSD talkers, for a larger subset of the vowels. To test the significance of possible effects of regional background and gender, we carried out a repeated measures analysis of variance (ANOVA) on the duration measurements. In the ANOVA, the within-subject factor VOWEL consisted of the duration measurements for the 15 vowels, while the regional

TABLE II. Average formant frequencies (Hz) for the first three formants for the vowels tokens produced by the 20 NSD talkers and the 20 SSD talkers. F, female talkers; M, male talkers. *N* per cell is 20. *F*1, *F*2, and *F*3 were sampled at 25%: "*t*3" and 75%: "*t*7" of the steady-state portion of each vowel token.

			/ɔu/		/ɛɪ/		/œy/		/e/		/ø/		/o/	
NSD			*t*3	*t*7	*t*3	*t*7	*t*3	*t*7	*t*3	*t*7	*t*3	*t*7	*t*3	*t*7
	*F*1	F	763	565	692	527	740	530	452	402	468	418	485	407
		M	633	466	557	450	592	457	415	361	407	329	444	365
	*F*2	F	1419	1128	2019	2225	1675	1675	2241	2425	1726	1737	1078	937
		M	1254	994	1733	1920	1485	1555	1887	2075	1544	1592	1015	848
	*F*3	F	2656	2616	2817	2823	2679	2594	2890	2940	2574	2552	2489	2417
		M	2320	2340	2502	2586	2357	2261	2545	2597	2254	2229	2258	2371
SSD			*t*3	*t*7	*t*3	*t*7	*t*3	*t*7	*t*3	*t*7	*t*3	*t*7	*t*3	*t*7
	*F*1	F	741	564	666	587	679	603	438	430	437	424	435	408
		M	612	428	575	478	588	488	386	376	377	355	385	347
	*F*2	F	1398	1158	2111	2270	1742	1856	2349	2425	1796	1828	1077	927
		M	1225	985	1704	1899	1449	1584	1942	2020	1534	1562	965	844
	*F*3	F	2806	2888	3042	3081	2868	2825	3007	2997	2697	2685	2922	3040
		M	2441	2544	2539	2553	2439	2404	2616	2611	2422	2387	2472	2576

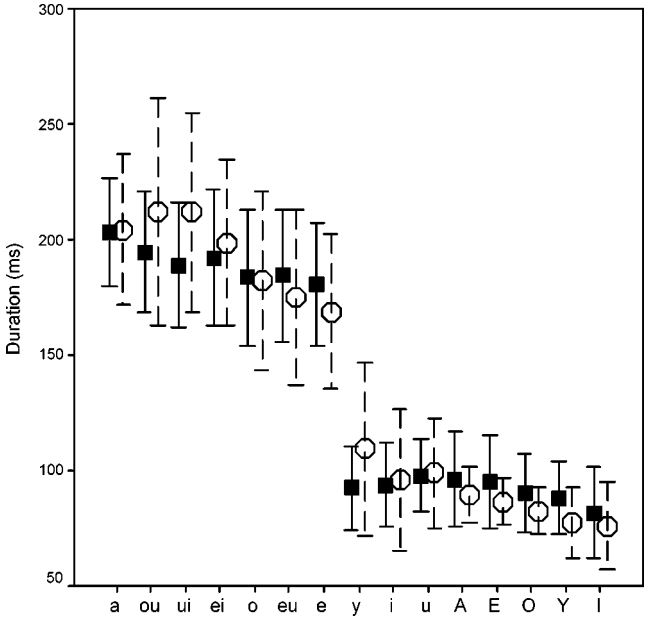

FIG. 1. Error plot of average duration (ms) per vowel, for the men for the two regions: NSD (filled squares) and SSD (open circles) (“ou”=/ɔu/, “ui”=/œy/, “eu”=/ø/, “A”=/ɑ/, “O”=/ɔ/, “Y”=/ʏ/, “I”=/ɪ/). Bars represent one standard deviation.

background of the talker (region) and the talker’s gender (gender) served as between-subject factors. The analysis showed that the longer durations for the female talkers in Figs. 1 and 2 are significant: gender had a significant effect on duration ($F[1,36]=5.77$, $p<0.05$). Region, the interaction of gender×region, and the three-way interaction of vowel×region×gender were just not significant. Restricting the ANOVAs to pairs out of the four groups clearly indicated that one group was different from the other three: SSD women, who on average produce longer durations. *Post-hoc*

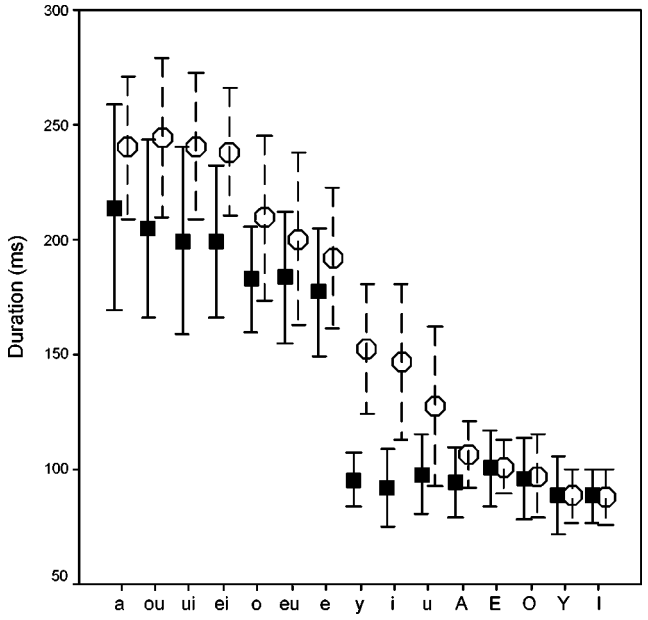

FIG. 2. Error plot of average duration (ms) per vowel, for the women for the two regions: NSD (filled squares) and SSD (open circles) (“ei”=/ɛi/, “ou”=/ɔu/, “ui”=/œy/, “eu”=/ø/, “A”=/ɑ/, “O”=/ɔ/, “Y”=/ʏ/, “I”=/ɪ/). Bars represent one standard deviation.

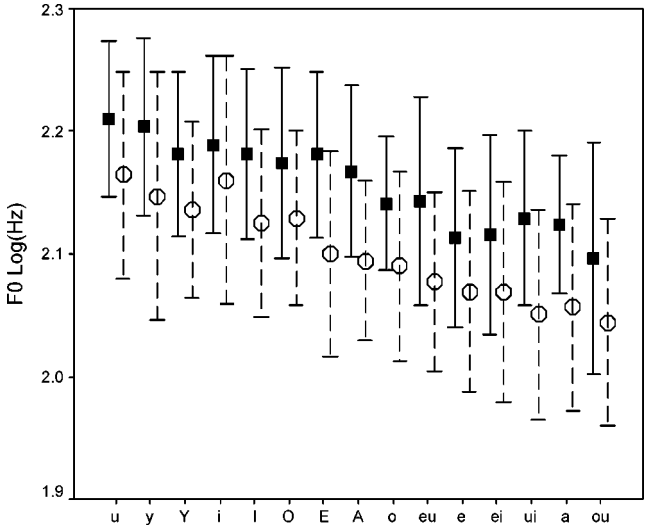

FIG. 3. Error plot of average $F0$ in Log(Hz) per vowel, for the men for the two regions: NSD (filled squares) and SSD (open circles) (“ei”=/ɛi/, “ou”=/ɔu/, “ui”=/œy/, “eu”=/ø/, “A”=/ɑ/, “O”=/ɔ/, “Y”=/ʏ/, “I”=/ɪ/). Bars represent one standard deviation.

analyses on the four groups for all 15 vowels reveal a clear pattern (Tukey, $\alpha=0.05$): SSD women have significantly longer durations compared to the other three groups for the diphthong “ɛi” and for the three high vowels /y ʏ u/.

The above results for the duration measurements can be summarized as follows. First, the 15 vowels fall into two groups based on their duration: longer ones /a ɛɪ ɔu œy e o ø/ and shorter ones /ɑ ɛ i ɪ ɔ u y ʏ/. Second, female SSD talkers produced longer durations than male talkers for several vowels, indicating a more general effect for the whole set of long vowels and the high vowels. Longer vowel durations for female talkers were reported earlier in Hillenbrand *et al.* (1995). We, like Hillenbrand *et al.*, cannot account for these gender-specific differences and we cannot predict whether a similar difference between male and female talkers in SSD would also be found in spontaneous speech.

Finally, it is not feasible to compare our results with those reported in Pols *et al.* (1973) and Van Nierop *et al.* (1973), because they did not include duration measurements in their description of the Dutch vowels.

2. Fundamental frequency

The average $F0$ values for the male talkers are displayed in Fig. 3, whereas Fig. 4 gives the values for the female talkers. First, a regional difference can be observed in Fig. 3: average $F0$ values for all vowels are lower for SSD, with differences ranging between 8 Hz (for /i/) and 25 Hz (for /ɛ/). No such pattern can be observed for the female talkers in Fig. 4. A repeated measures ANOVA was carried out for the male and female talkers separately to evaluate whether there was an effect of region on $F0$. In this ANOVA, vowel was composed of the 15 vowels and the regional background of the talker (region) served as a between-subject factor. The results show no significant effects for region. Further analysis showed that $F0$ and duration correlate: vowels with a low average $F0$ in Figs. 3 and 4 generally show longer durations in ms in Figs. 1 and 2. Pearson’s r was computed for $F0$ and

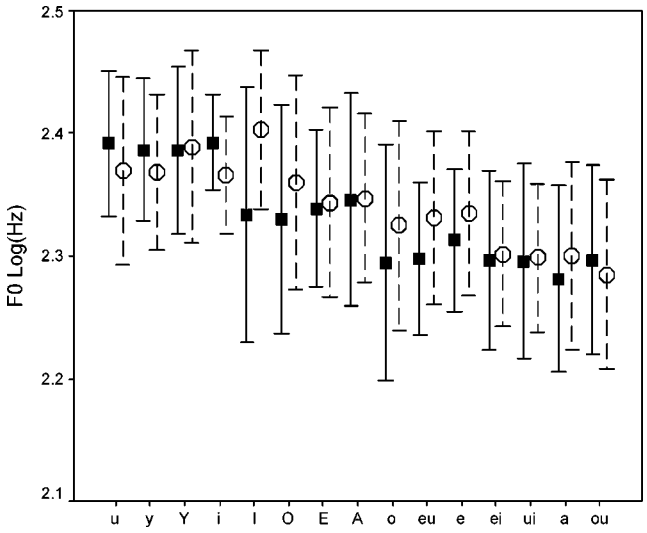

FIG. 4. Error plot of average $F0$ in Log(Hz) per vowel, for the women for the two regions: NSD (filled squares) and SSD (open circles) (''au''=/ɔu/, ''ui''=/œy/, ''eu''=/ø/, ''A''=/ɑ/, ''O''=/ɔ/, ''Y''=/ʏ/, ''I''=/ɪ/). Bars represent one standard deviation.

duration for the NSD women (−0.468), NSD men (−0.480), SSD women (−0.284), and SSD men (−0.331), across all vowel tokens. All correlations were significant at $p<0.05$. This phenomenon is also known as vowel-intrinsic $F0$, and it was reported to occur in many languages [cf. Whalen and Levitt (1995) for an overview]. Finally, it was again not possible to compare our results with those in Pols *et al.* (1973) and Van Nierop *et al.* (1973), because they did not include fundamental frequency measurements in their description.

3. Formant frequencies

a) Steady-state. Figures 5 (male talkers) and 6 (female talkers) show acoustic vowel diagrams for the nine monophthongal vowels at 50% ($t5$); the long mid vowels /e ø o/ and the three diphthongal vowels /ɛɪ ɔu œy/ were excluded, because we suspected that these six vowels show considerable diphthongization.

It is not easy to provide a concise description of the variation patterns in Figs. 5 and 6. Still, some general tendencies can be observed. Overall, Fig. 5 shows that the shape and size of the vowel systems for both language varieties for the male talkers are roughly similar, although /u/ for NSD shows a higher average $F2$. Large differences can also be seen for /ɪ/, /a/, /ɛ/, and /i/, for $F1$ as well as for $F2$. For the female talkers in Fig. 6, the overall shape and size of the vowel systems for NSD and SSD appear to differ little; the locations of the point vowels /i a u/ are roughly equal for both varieties. Larger differences can be observed for /ɪ/ and /y/; average $F2$ is higher and average $F1$ is lower for /ɪ/ for NSD, and average $F1$ and $F2$ are lower for NSD for /y/. Finally, Fig. 6 shows that /ɪ/ and /ɛ/ differ mainly in their average $F1$ and less in their average $F2$.

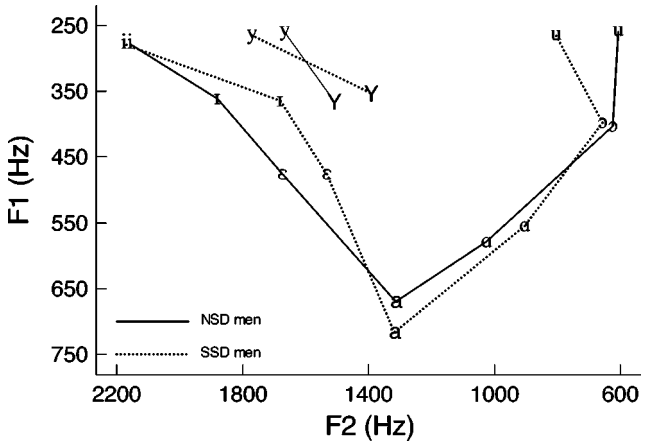

FIG. 5. Acoustic vowel diagram showing average formant frequencies for the nine monophthongal vowels /a ɑ ɛ i ɪ ɔ u y ʏ/ for NSD and SSD for the men. The averages were taken at 50% of the vowel duration.

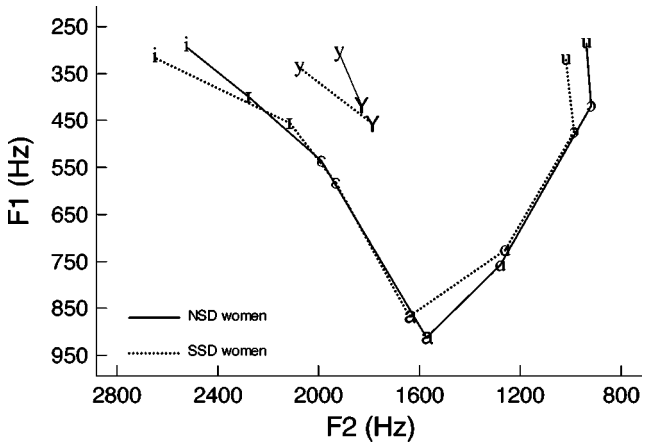

FIG. 6. Acoustic vowel diagram showing average formant frequencies for the nine monophthongal vowels /a ɑ ɛ i ɪ ɔ u y ʏ/ for NSD and SSD for the women. The averages were taken at 50% of the steady-state portion of the vowel duration.

b) Spectral change. Figure 7 gives the average frequencies for $F1$ and $F2$ at 25% and 75% in the vowel duration ($t3$ and $t7$, respectively). Figure 7 presents spectral change for the three full diphthongs, for the male and female talkers, respectively. Figure 8 shows the spectral change for the three long mid vowels, for the male and female talkers.

To evaluate whether there are systematic differences in the spectral change patterns of the two regions and the two genders depicted in Figs. 7 and 8, five ANOVAS (repeated measures) were carried out, three for the diphthongal vowels /ɛɪ ɔu œy/ and two for the long mid vowels /e o ø/. The within-subject factor consisted of a measure for the spectral change in each vowel token, which was defined as the absolute difference of the formant frequency between 25% and 75% of the vowel duration in Hz. Thus $\Delta F1$ is the absolute difference in Hz between the values of $F1$ at $t3$ and at $t7$, and $\Delta F2$ is the absolute difference in Hz between the values of $F2$ at $t3$ and at $t7$. The first ANOVA for the diphthongal vowels tested possible effects of the between-subject factors region and gender and the within-subject factor vowel, on the change in the first formant frequency $\Delta F1$. The results showed an effect for region ($F[1,36]=7.15$, $p<0.05$) and an interaction effect for gender×region ($F[1,36]=4.42$, $p<0.05$). The significance of the interaction between gender and region indicates that the effect of region on $\Delta F1$ is

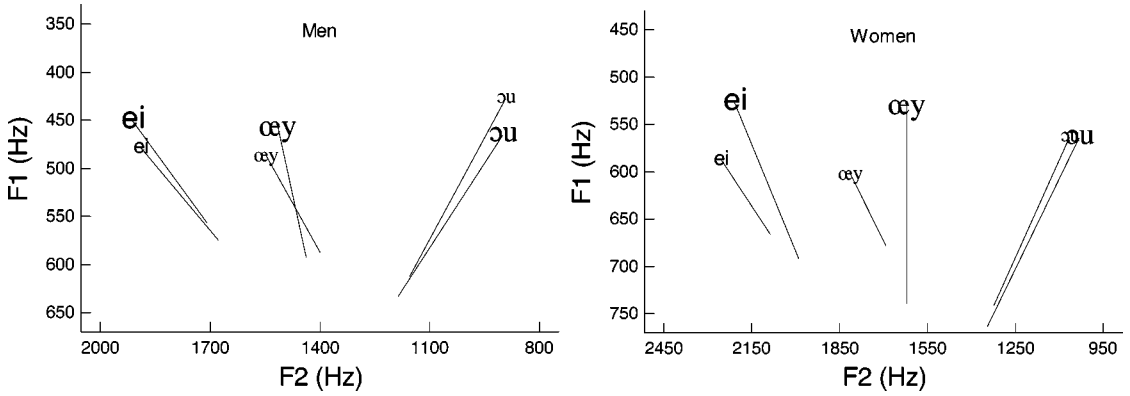

FIG. 7. Spectral change patterns for the three diphthongal vowels /ɛi œy ɔu/ for NSD and SSD for the men (left panel) and the women (right panel). The phonetic symbol at the end of each line is plotted at the average formant frequency at 75% of the vowel duration and each line originates from the average formant frequencies at 25% of the vowel duration. The larger symbols refer to NSD and the smaller symbols refer to SSD. ei refers to /ɛi/.

gender-specific. To investigate the interaction further, the last ANOVA was repeated for the male and female talkers separately, with region as the sole between-subject factor. The results showed an effect of region for the female talkers ($F[1,18]=11.32$, $p<0.05$) for $\Delta F1$, but not for the male talkers. In the next analysis, vowel represented measurements of $\Delta F2$ for the vowels /ɛɪ ɔu œy/, and region and gender as between-subject factors. The results showed no significant effects for $\Delta F2$. The final two analyses were set up as the first two, only this time the analyses were carried out on the vowels /e o ø/ for $\Delta F1$ and $\Delta F2$, respectively. The results for the ANOVA with $\Delta F1$ for /e o ø/ showed an effect for region ($F[1,36]=33.53$). An effect for region was also found for the ANOVA with $\Delta F1$ for /e o ø/ ($F[1,36]=18.51$, $p<0.05$).

In summary, the following spectral change patterns were found. First, for the diphthongal vowels /ɛɪ ɔu œy/, greater diphthongization of $F1$ was found for the female NSD talkers as opposed to the female SSD talkers, whereas no such region effect was found for the male talkers. Second, the long mid vowels /e o ø/ showed more diphthongization of $F1$ and $F2$ for NSD than for SSD, for both genders.

c) Comparison with Pols *et al.* (1973) *and* Van Nierop *et al.* (1973). Figure 9 gives the acoustic vowel diagram for the nine monophthongal vowels at 50% ($t5$) for our male talkers plus the Pols *et al.*'s male talkers. Figure 10 gives the acoustic vowel diagram for the nine monophthongal vowels at 50% ($t5$) for our female talkers plus Van Nierop *et al.*'s female talkers. A first general observation from Fig. 9 is that all average formant values for Pols *et al.*'s data are higher for $F1$. The vowels /a ɑ ɔ/ show lower average values for $F2$ for Pols *et al.*'s data. Figure 10 shows a different variation pattern in the vowel systems for the women than the one observed for the men in Fig. 9. Higher average values for $F1$ for /a ɛ ɪ ɔ ʏ/ and lower values for $F2$ for the vowels /a ɑ ɛ ɔ u y ʏ/ can be observed in Fig. 10. Overall, the vowel diagrams for our data appear smaller than Pols *et al.*'s and Van Nierop *et al.*'s, possibly indicating a more ''casual'' speaking style in our data. It is unclear what caused the other differences between the two data sets. One possible cause may be that the pronunciation of the Dutch vowels has changed in the three decades since Pols *et al.* and Van Nierop *et al.* made their recordings. In our opinion, however, it seems more likely that the observed differences are due to either differences in the consonantal context of the vowels (/hVt/ vs. /sVs/), or to uncontrolled regional variation in Pols *et al.*'s and Van Nierop *et al.*'s data, or to differences in the techniques used to estimate the formant frequencies. Given these variation sources, we decided not to further analyze differences between our data and data described in Pols *et al.* and Van Nierop *et al.*

IV. DISCRIMINANT ANALYSIS

We carried out a series of quadratic discriminant analyses (QDAs) to establish how well the vowels could be separated based on various combinations of acoustic characteristics. We evaluated steady-state characteristics, i.e., $F0$ and

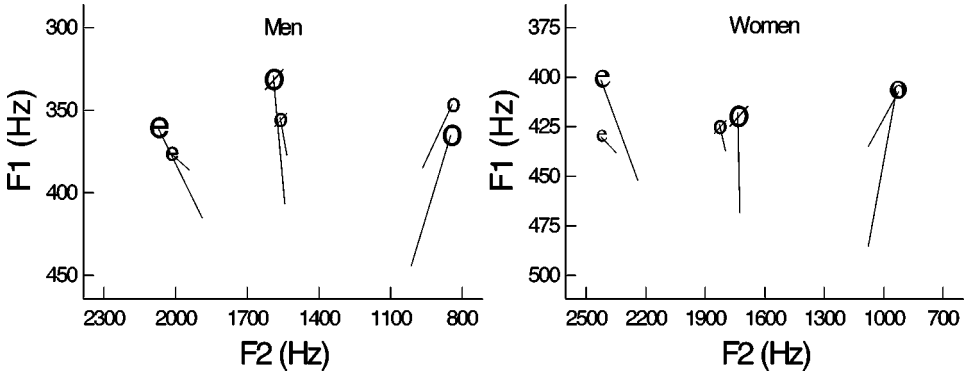

FIG. 8. Spectral change patterns for the three long mid vowels, /e ø o/, for NSD and SSD for the men (left panel) and the women (right panel). The phonetic symbol at the end of each line is plotted at the average formant frequency at 75% of the vowel duration and each line originates from the average formant frequencies at 25% of the vowel duration. The larger symbols refer to NSD and the smaller symbols refer to SSD.

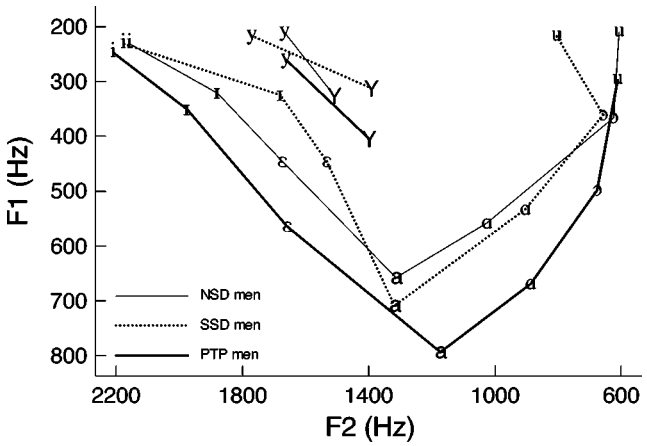

FIG. 9. Acoustic vowel diagram showing average formant frequencies for the nine monophthongal vowels /a ɑ ɛ i ɪ ɔ u y ʏ/ for NSD, SSD, and from Pols, Tromp, and Plomp (1973), "PTP," for the men. All averages were taken at 50% of the vowel duration. The averages for NSD and SSD were based on 20 talkers and the averages from Pols *et al.* were based on 50 talkers.

$F1-F3$ sampled at 50% of the vowel duration ($t5$), as well as dynamic specifications, i.e., the duration and $F1-F3$ sampled at 25% and 75% of the duration ($t3$ and $t7$). Table III shows that including only $F1$ and $F2$ at $t5$ yields 59.0% for NSD and 54.2% for SSD. Including $F0$ in the set led to a small improvement for NSD and a considerable improvement for SSD. Entering the parameter set consisting of $F1$, $F2$, and $F3$ led to improvements for both language varieties as compared to entering only $F1$ and $F2$. Entering $F0$, $F1$, $F2$, and $F3$ led to another improvement in the scores. To find out whether the results would improve when the three mid vowels and the three full diphthongal vowels are excluded, the QDA for the parameters $F1$ and $F2$ was repeated using only the nine monophthongal vowels. The results show a remarkable improvement over those carried out using all 15 vowels, for both language varieties. However, the improvement is considerably higher for NSD. Adding $F3$ led to a further improvement, especially for SSD. The next QDA was

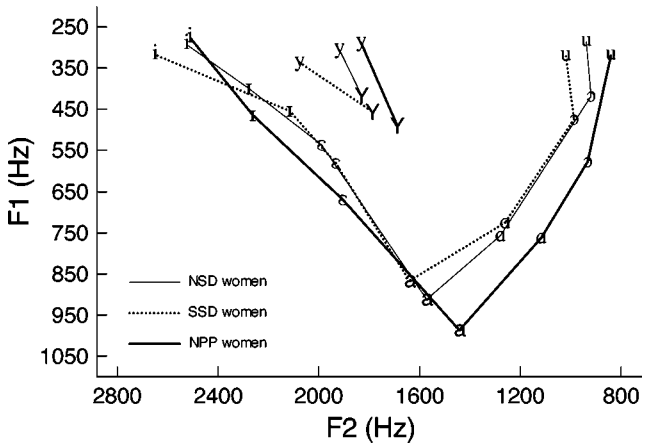

FIG. 10. Acoustic vowel diagram showing average formant frequencies for the nine monophthongal vowels /a ɑ ɛ i ɪ ɔ u y ʏ/ for NSD, SSD, and from Van Nierop, Pols, and Plomp (1973), "NPP," for the women. All averages were taken at 50% of the vowel duration. The averages for NSD and SSD were based on 20 talkers and the averages from Van Nierop *et al.* were based on 25 talkers.

TABLE III. Percent correctly classified vowel tokens for the quadratic discriminant analyses for the NSD vowel tokens ($N=600$) and the SSD vowel tokens ($N=600$). The data were classified into 1 of the 15 vowel categories, or into 1 of 9 monophthongal vowel categories /a ɑ ɛ i ɪ ɔ u y ʏ/ ("9 vowels"). The parameters were sampled at one point in the vowel duration: "$t5$" (50% of the vowel duration), or at two points "$t3+t7$" (25% and 75%). For NSD, all percentages higher than 62.9% are significantly higher than the baseline (59.0%), for SSD all percentages higher than 58.2% are significantly higher than the baseline (54.2%).

Parameters	Categories	Sample	NSD	SSD
$F1$, $F2$	15 vowels	$t5$	59.0	54.2
$F0$, $F1$, $F2$	15 vowels	$t5$	65.5	67.5
$F1$, $F2$, $F3$	15 vowels	$t5$	68.0	66.3
$F0$, $F1$, $F2$, $F3$	15 vowels	$t5$	71.5	71.5
$F1$, $F2$, $F3$, dur	15 vowels	$t3+t7$	90.8	90.0
$F1$, $F2$	9 vowels	$t5$	83.6	76.4
$F1$, $F2$, $F3$	9 vowels	$t5$	86.1	83.3
$F1$, $F2$, $F3$, dur	15 vowels	$t5$	90.0	85.3
$F0$, $F1$, $F2$, $F3$, dur	15 vowels	$t5$	89.3	88.8

carried out using $F1$, $F2$, and $F3$ and duration as predictor variables, using all 15 vowels at $t5$, to evaluate the role of duration. For this parameter set, a considerable improvement was found as compared to including only $F1$, $F2$, and $F3$ as parameters. Furthermore, when $F0$ is added as well as $F1$, $F2$, $F3$, and duration, some improvement is again found, but this time only for SSD (85.3% to 88.8%); for NSD the percentage is lower (90.0% to 89.3%). It is unclear why the percentage for NSD is lower, but since it only is a small difference (0.7%) not much weight should be attributed to it. The improvement for SSD is attributable to including $F0$. It appears that $F0$ plays a greater role in distinguishing vowels in SSD than in NSD, whenever $F0$ is added to a parameter set, the improvement is largest for SSD. Duration, on the other hand, seems to play a greater role in distinguishing vowels for NSD; improvements for NSD are largest for NSD whenever duration is added to a parameter set. Finally, when the last analysis is repeated with all parameters at $t3$ and $t7$, the highest percents correctly classified are obtained for both language varieties. Note that the relative improvements are highest for SSD, which is remarkable given the finding that SSD shows less diphthongization of $F1$ and $F2$ for the long mid vowels and the diphthongal vowels.

The results in Table III can be interpreted as follows: first, it appears that the nine monophthongal vowels could be separated reasonably well using only steady-state $F1$ and $F2$, although a higher percent correctly classified vowel tokens was obtained for NSD. Second, the three long mid vowels and the three full diphthongal vowels could be separated fairly well when dynamic properties such as duration and dynamic spectral information were included in the analysis, although again the percent correctly classified vowel tokens was highest for NSD. It could thus be the case that the vowel system for NSD requires more information about dynamic properties (especially for the three long mid vowels and the three diphthongs) in order to be separated acoustically.

V. DISCUSSION

The purpose of this paper was twofold: to give a description of the acoustic characteristics of all 15 Dutch vowel

sounds and to provide an overview of similarities and dissimilarities of the vowel systems of Northern Standard Dutch and Southern Standard Dutch. The results for the first aim were as follows.

First, the 15 vowels of Dutch could be divided into two groups depending on their duration. The group of shorter vowels consists of all full monophthongal vowels /y i u ɑ ɛ ɔ ʏ ɪ/ except for /a/, while the group of long vowels consists of the three diphthongal vowels /ɔu œy ɛɪ/, the three long mid vowels /o ø e/, and /a/. This phonetic pattern in the relative duration is only partially compatible with the phonological characteristics of Dutch vowels as described in Booij (1995), where the set of long vowels includes, besides /o ø e a/, also /i y u/. Our results also deviate slightly from the description provided by Koopmans-van Beinum (1980) and Rietveld *et al.* (in press). Koopmans-van Beinum and Rietveld *et al.* suggest that Dutch vowels can be divided into three groups: short, /ɑ ɛ ɪ ɔ ʏ/, half-long, /i y u/, and long, /a e ø o/ plus /ɛɪ ɔu œy/. If we pool the short and half-long vowels, the system by Koopmans-van Beinum and Rietveld *et al.* is compatible with our findings. This applies to three of the four groups of talkers. The results for the female SSD talkers conform the pattern found by Koopmans-Van Beinum and Rietveld *et al.* Nevertheless, it remains uncertain whether the pattern we reported for the relative average durations would also occur for data in other consonantal contexts and in spontaneous speech. Further research is required to investigate which patterns would occur in other speech samples.

Second, the nine monophthongal vowels of Dutch, /ɑ a ɛ ɪ i ɔ u ʏ y/, could be separated fairly well based on their steady-state characteristics for their first two formants frequencies alone, while the three long mid vowels and three diphthongal vowels required information about their dynamic characteristics as well. The long mid vowels cannot be described adequately acoustically by their steady-state characteristics alone. We suggest that the three long mid vowels for Dutch should therefore not be treated as monophthongal vowels, but instead as semi-diphthongal vowels, when describing Dutch vowels acoustically. This is especially the case for NSD. The diphthongization of /e ø o/ is most likely the result of a vowel shift. It is not clear whether the long mid vowels showed (early) signs of the process of diphthongization in Pols *et al.*'s and Van Nierop *et al.*'s descriptions three decades earlier. Pols *et al.* excluded the full diphthongs /ɛɪ ɔu œy/, but included the three long mid vowels. However, in their Fig. 3, considerable overlap can be observed in the vowel plot between /e/ and /ɪ/, and between /ø/ and /ʏ/ (denoted in Pols *et al.* by ''œ''), and /ɔ/ and /ø/. Furthermore, Pols *et al.* remark, ''...the vowels in these three pairs have very similar formant frequencies and levels. The main difference between them is duration.'' Nevertheless, it is not possible to establish whether /e ø o/ can be distinguished from /ɪ ʏ ɔ/, respectively, through their dynamic characteristics as well as by their duration, because Pols *et al.* and Van Nierop *et al.* do not present the formant frequencies at different time slices.

Regarding the second aim of this study, the difference between the two language varieties, the results were as follows. A first difference between the NSD and SSD was found for the duration of the three full diphthongs /ɛɪ ɔu œy/; the durations for the SSD talkers were longer for all three diphthongs, especially for the female talkers. The comparisons indicated further that NSD and SSD differ little in the steady-state characteristics of the nine monophthongal vowels; the main difference between both varieties was found for the three long mid vowels and the three full diphthongs. For the three diphthongs, more diphthongization (i.e., larger excursions) of the first formant was found for the female NSD talkers than for the female SSD talkers. No such effect was found for the male talkers. Overall, vowels produced by all NSD talkers show more diphthongization than vowels produced by SSD talkers. The differences between NSD and SSD were largest for the three long mid vowels. The results for the duration and dynamic spectral characteristic may be related: the longer durations for the diphthongal vowels for SSD may be used to compensate for the smaller amount of diphthongization for the SSD diphthongal vowels. Further research on other speech samples is necessary to evaluate the hypothesis that NSD and SSD use different quantity distinctions.

A comparison of our data with the data described in Pols *et al.* and Van Nierop *et al.* showed differences in the average frequencies of the nine monophthongal vowels. Most vowels show higher values for $F1$ for Pols *et al.*'s and Van Nierop *et al.*'s data. In addition, Pols *et al.* and Van Nierop *et al.* show lower values for $F2$ for some vowels (especially for /a ɑ ɔ/). We suggested three possible causes for the observed differences: a change in the pronunciation of the vowels, differences between the two data sets, such as the surrounding consonants (/sVs/ vs. /hVt/), or uncontrolled regional variation in Pols *et al.*'s and Van Nierop *et al.*'s data.

The present study is limited in that it describes the acoustic characteristics of the vowels of Dutch in read speech and in a fixed consonantal context only. It would be interesting to compare the acoustic characteristics of the vowels of the present study with other vowel tokens produced by the same talkers in different consonantal context and speaking styles. During the sociolinguistic interview in which our vowel data was obtained, recordings were also made of (guided) spontaneous conversations. We plan to extend our study in the near future and analyze the acoustic characteristics of vowel in these spontaneous speech samples as well. Finally, it was not feasible to provide a complete overview of all the variation patterns in all eight regional varieties in our data set in the present study. However, the individual measurements for all 160 speakers can be obtained by contacting the first or second author.

ACKNOWLEDGMENTS

We are grateful to two anonymous reviewers for their useful comments. We wish to thank James Hillenbrand for his help with Figs. 5–10. We are greatly indebted to Terry Nearey for writing the program and user interface that allowed us to perform our formant measurements. This research was supported by the Netherlands Organization for

Research (NWO) through the Flemish Netherlands Committee (VNC) under Project No. 205-41-069 (PI Roeland van Hout).

[1]Pols, Trom and Plomp (1973) used the vowel set of the 50 male talkers described in Klein, Plomp, and Pols (1970).

[2]Klein, Plomp, and Pols' (1970) talkers are described as ''The 50 speakers were young male adults with a pronunciation representative of correct Dutch.''

[3]No recordings were made for the so-called ''marginal'' vowels of Dutch, i.e., the vowels /i: y: ɛ: œ:/ and the nasalized versions of the vowels /ɛ: ɑ: ɔ:/, which all occur only in loan words in Dutch (Gussenhoven, 1999).

[4]It was not feasible to include measurements of *F*4, because these could not be obtained in the program used for the formant estimation.

Adank, P. (**2003**). ''Vowel Normalization: A perceptual-acoustic study of Dutch vowels,'' Ph.D. thesis, University of Nijmegen.

Bell, A. (**1983**). ''Broadcast news as language standard,'' Int. J. Soc. Lang. **40**, 29–42.

Boersma, P. (**1993**). ''Accurate short-term analysis of fundamental frequency and the harmonics-to-noise ratio of a sampled sound,'' in *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, Vol. 17, pp. 97–110.

Boersma, P. (**2001**). ''Praat, a system for doing phonetics by computer,'' Glot Int. **5(9/10)**, 341–345.

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

Deterding, D. (**1997**). ''The formants of monophthong vowels in Standard Southern British English Pronunciation,'' J. Int. Phon. Assoc. **27**, 47–55.

Di Benedetto, M.-G. (**1989a**). ''Vowel representation: Some observations on temporal and spectral properties of the first formant frequency,'' J. Acoust. Soc. Am. **86**, 55–66.

Di Benedetto, M.-G. (**1989b**). ''Frequency and time variations of the first formant: Properties relevant to the perceptions of vowel height,'' J. Acoust. Soc. Am. **86**, 67–77.

Disner, S. (**1980**). ''Evaluation of vowel normalization procedures,'' J. Acoust. Soc. Am. **67**, 253–261.

Gussenhoven, C. (**1999**). ''Illustrations of the IPA: Dutch,'' in *Handbook of the International Phonetic Association* (Cambridge U.P., Cambridge), pp. 74–77.

Hagiwara, R. (**1997**). ''Dialect variation and formant frequency: The American English vowels revisited,'' J. Acoust. Soc. Am. **102**, 655–658.

Hillenbrand, J., and Gayvert, R. T. (**1993**). ''Identification of steady-state vowels synthesized from the Peterson and Barney measurements,'' J. Acoust. Soc. Am. **94**, 668–674.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). ''Acoustic analysis of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.

Klein, W., Plomp, R., and Pols, L. C. W. (**1970**). ''Vowel spectra, vowel spaces and vowel identification,'' J. Acoust. Soc. Am. **48**, 999–1009.

Koopmans-van Beinum, F. (**1980**). ''Vowel contrast reduction,'' Ph.D. thesis, University of Amsterdam.

Markel, J. D., and Gray A. H. (**1976**). *Linear Prediction of Speech* (Springer-Verlag, Berlin).

Nearey, T. M. (**1989**). ''Static, dynamic, and relational properties in speech perception,'' J. Acoust. Soc. Am. **85**, 2088–2113.

Nearey, T. M., Assmann, P., and Hillenbrand, J. (**2002**). ''Evaluation of a strategy for automatic formant tracking,'' J. Acoust. Soc. Am. **112**, 2323.

Payne, A. (**1980**). ''Factors controlling the acquisition of the Philadelphia dialect by out-of-state children,'' in *Locating Language in Time and Space*, edited by W. Labov (Academic, New York).

Pols, L. C. W., Pols, L. C. W., and Plomp, H. R. C. (**1973**). ''Frequency analysis of Dutch vowels from 50 male speakers,'' J. Acoust. Soc. Am. **53**, 1093–1101.

Rietveld, T., Kerkhoff, J., and Gussenhoven, C. (**in press**). ''Word prosodic structure and vowel duration in Dutch,'' to appear in J. Phonetics.

Scovel, T. (**1988**). *A Time to Speak: A Psycholinguistic Inquiry into the Critical Period for Human Speech* (Newbury House, Rowley).

Strange, W. (**1989**). ''Dynamic specification of coarticulated vowels spoken in sentence context,'' J. Acoust. Soc. Am. **85**, 2135–2153.

Strange, W., Jenkins, J. J., and Johnson, T. L. (**1983**). ''Dynamic specification of coarticulated vowels,'' J. Acoust. Soc. Am. **74**, 695–705.

Syrdal, A. K., and Gopal, H. S. (**1986**). ''A perceptual model of vowel recognition based on the auditory representation of American English vowels,'' J. Acoust. Soc. Am. **79**, 1086–1100.

Van de Velde, H., van Hout, R., and Gerritsen, M. (**1997**). ''Watching Dutch Change,'' J. Sociolinguistics **1**, 361–391.

Van Nierop, D. J. P. J., Pols, L. W. C., and Plomp, R. (**1973**). ''Frequency analysis of Dutch vowels from 25 female speakers,'' Acustica **29**, 110–118.

Whalen, D. H., and Levitt, A. G. (**1995**). ''The universality of intrinsic F0 of vowels,'' J. Phonetics **23**, 349–366.