ACOUSTIC PROPERTIES OF DUTCH STEADY-STATE VOWELS: CONTEXTUAL EFFECTS AND A COMPARISON WITH PREVIOUS STUDIES

Jan-Willem van Leussen[a], *Daniel Williams*[b] *& Paola Escudero*[c]

[a]University of Amsterdam, the Netherlands; [b]University of Sheffield, UK;
[c]University of Western Sydney, Australia
j.w.vanleussen@uva.nl; gep09dpw@sheffield.ac.uk; paola.escudero@uws.edu.au

ABSTRACT

Recent vowel corpora show that there are often clear acoustic differences between vowels produced in different phonetic contexts. We expand on a recent corpus of Northern Standard Dutch (NSD) vowels by including a variety of consonantal contexts. Our results show that there are very clear contextual effects on the spectral and temporal properties of NSD vowels. The most striking effect is the apparent 'fronting' of vowels in alveolar contexts, which has not previously been reported for Dutch. Classification with a supervised learning algorithm reveals some substantial differences between our acoustic measurements and those reported in earlier studies on NSD vowels.

Keywords: Dutch, vowels, production, coarticulation, fronting

1. INTRODUCTION

A number of vowel corpora have appeared in recent years that include vowels that are not limited to one specific phonetic context, but include a variety of phonetic contexts, e.g. [3, 5, 7, 11]. Such corpora reveal that there are often clear acoustic differences between vowels produced in different contexts. Moreover, the significance of this acoustic variation may extend to native, e.g. [7], and cross-language vowel perception, e.g. [10].

However, a recent study on Northern Standard Dutch (NSD), the standard variety of Dutch spoken in the Netherlands, recorded vowels in only the sVs and sVsə [1] contexts, and two older studies on NSD vowels used only the hVt context [8, 9].

The present study presents acoustic data from a new corpus of NSD vowels that have been produced in six consonantal contexts in monosyllabic and disyllabic words and in three sentence positions. Our findings will be compared to the three previous NSD vowel corpora through the use of a phonological learning algorithm.

2. METHOD

2.1. Participants

22 native NSD speakers (11 female, 11 male) participated. They grew up and lived in the central region of the Netherlands (the provinces of North Holland, South Holland and Utrecht) for most of their lives. As young speakers of Dutch have a good command of English, speakers were excluded if they rated their fluency in any other foreign language at 4 or higher on a scale from 0 to 7.

Participants were 18 to 28 years of age and were students or recent graduates. None reported any speech or hearing problems.

2.2. Task and recording

Recordings were made in a soundproof chamber at the University of Amsterdam, using a Sennheiser microphone and Edirol UA-25 sound card with a sampling rate of 44.1 kHz.

Sentences were read from prompt sheets in a randomized order. They had the format "*<u>CVC</u>. In <u>CVC</u> en <u>CVC</u>ə zit de <u>V</u>*" ("<u>CVC</u>. In <u>CVC</u> and <u>CVC</u>ə is <u>V</u>"), where V was one of 15 Dutch vowels /i, y, ɪ, ʏ, ø, e, ɛ, a, ɑ, ɔ, o, u, ɛi, œy, ɔu/ and CVC was one of six consonantal contexts /fVf, sVs, pVp, tVt, kVk, tVk/ [1]. This format was adapted from recent vowel production studies [3, 5]. Each possible CVC permutation was included twice, for a total of 180 sentences (15x6x2).

We asked participants to read at a speed and style close to their normal speaking rate, and to re-read a sentence if they made a mistake. Participants were given a practice session before recording to ensure they understood the task and read the intended vowels correctly.

2.3. Formant, pitch and duration analysis

We used Praat [2] to manually mark the beginning and end of the first three vowels of each recorded sentence. Vowels that were the result of a reading error were discarded.

F0, F1 and F2 values were then extracted for each vowel at 50% of its duration using the Burg algorithm in Praat. The upper bound of the search space for formants was determined through the "optimal formant ceilings" method of [5]: F1-F3 of all vowel *tokens* were measured with a formant ceiling ranging from 4000 to 6500 Hz in 10 Hz steps. Then, for each vowel of each speaker, a formant ceiling was chosen such that it yielded the lowest variation for the first two formants within the set of 36 tokens annotated for this vowel. Table 1 summarizes the results of the formant and duration analyses.

Table 1: Acoustic values of our corpus, logarithmically averaged over all M/F tokens.

		F1 (Hz)	F2 (Hz)	F3 (Hz)	Duration (ms)
/a/	F	963.28	1622.4	2628.09	171.99
	M	728.32	1338.12	2326.63	173.38
/ɑ/	F	781.96	1335.23	2647.63	98.55
	M	617.33	1178.74	2230.55	98.37
/ɛ/	F	656.29	1968.35	2724.48	100.22
	M	528	1620.02	2369.03	97.42
/i/	F	354.58	2644.04	3229.29	90.86
	M	316.85	2127.72	2782.6	90.46
/ɪ/	F	460.55	2275.89	2922.68	89.33
	M	405.18	1852.35	2566.99	87.55
/ɔ/	F	477.36	1016.6	2739.28	91.04
	M	455.51	904.26	2283.44	91.58
/u/	F	388.77	1046.84	2693.97	90.66
	M	339.26	953.26	2186.96	92.09
/y/	F	372.71	2031.29	2731.46	95.91
	M	325.15	1697.23	2264.94	95.55
/ʏ/	F	462.79	1813.64	2673.17	91.88
	M	420.12	1498.84	2225.88	89.88

3. RESULTS

In the following statistical analyses, we only report the results for the nine monophthongal (steady-state) Dutch vowels /i, y, ɪ, ʏ, ɛ, a, ɑ, ɔ, u/ and exclude the three potential diphthongs /ø, e, o/ and the three diphthongs /ɛi, œy, ɔu/. Our analyses are largely based on those of [3]. We tested for the effects of gender, vowel category and consonantal context on log-transformed duration values and F0, F1 and F2 values measured at 50% duration.

3.1. Effects of vowel category and gender

We performed separate repeated-measures analyses of variance (ANOVA) on the F0, F1, F2 and duration values (medianized over 36 tokens) for each vowel of every speaker, with vowel category as within-subjects factor (nine levels) and gender as between-subjects factor. Table 2 summarizes the results.

Table 2: Effects of vowel category (V) and gender.

V	F0	$F(3.45\varepsilon, 69.08\varepsilon) = 72.96\varepsilon$	$p < 0.001$
V	F1	$F(2.26\varepsilon, 45.09\varepsilon) = 364.79\varepsilon$	$p < 0.001$
V	F2	$F(3.37\varepsilon, 67.31\varepsilon) = 497.41\varepsilon$	$p < 0.001$
V	dur.	$F(3.91\varepsilon,78.27\varepsilon) = 355.59\varepsilon$	$p < 0.001$
gender	F0	$F(1,20)= 124.74$	$p < 0.001$
gender	F1	$F(1,20) = 66.29$	$p < 0.001$
gender	F2	$F(1,20) = 110.43$	$p < 0.001$

There is a significant effect for vowel category on pitch, formant frequencies and duration. Unsurprisingly, we also find a main effect of gender on F0, F1 and F2. There is no effect of gender on duration, indicating that male and female speakers do not differ significantly in their realization of duration in monophthongs.

3.2. Exploratory analysis of consonant context

For each vowel of every speaker, we take the median F0, F1, F2 and duration over all 6 tokens produced in each consonant context (/fVf, sVs, pVp, tVt, kVk, tVk/).

We find a significant main effect of consonantal context, as well as a significant interaction with vowel category, on all measures (Table 3).

Table 3: Effects of consonantal context (C) and interactions with vowel category (V).

C	F0	$F(3.88\varepsilon, 77.67\varepsilon)\ \varepsilon = 5.33$	$p = 0.001$
C	F1	$F(3.90\varepsilon, 77.92\varepsilon)\ \varepsilon = 26.85$	$p < 0.001$
C	F2	$F(3.30\varepsilon, 66.09\varepsilon)\ \varepsilon = 155.13$	$p < 0.001$
C	dur.	$F(3.97\varepsilon, 79.45\varepsilon)\ \varepsilon = 23.15$	$p < 0.001$
C*V	F0	$F(7.88\varepsilon, 157.61\varepsilon)\ \varepsilon = 2.69$	$p = 0.009$
C*V	F1	$F(10.23\varepsilon, 204.60\varepsilon)\ \varepsilon = 4.66$	$p < 0.001$
C*V	F2	$F(9.77\varepsilon, 196.48\varepsilon)\ \varepsilon = 48.42$	$p < 0.001$
C*V	dur	$F(11.63\varepsilon, 232.56\varepsilon)\ \varepsilon = 6.30$	$p < 0.001$

3.3. Effects of consonantal context per vowel

One-way ANOVAs on the vowel categories separately reveal no significant effect of context on F0 for any vowel. For F1, we find a significant effect of context for the vowels /a/ and /u/. For F2, we find significant effects of consonantal context for all vowels except /i/ and /ɛ/ (the latter was near-significant).

For duration, significant effects are found for /ɑ, a, i, ɔ, u/. Table 4 summarizes these results.

No interaction effects were found between gender and consonantal context for any of the measures on any vowel.

Figure 1: average male steady-state vowels in different contexts. Shapes encode all six contexts;

ellipses show one standard deviation from the mean for alveolar (solid) and non-alveolar (dotted) contexts.

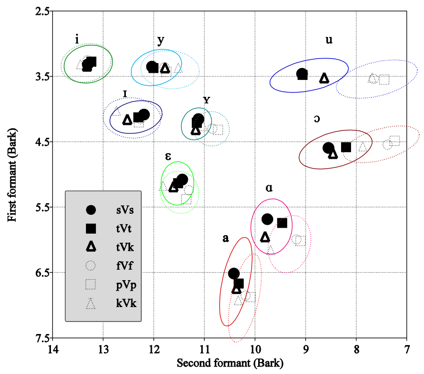


Table 4: Effects of cons. context per vowel.

/a/	F1	*F*(5, 120) = 5.99	$p < 0.001$
/u/	F1	*F*(5, 120) = 2.37	$p = 0.043$
/a/	F2	*F*(5, 120) = 2.71	$p = 0.023$
/ɑ/	F2	*F*(5, 120) = 20.22	$p < 0.001$
/ɪ/	F2	*F*(5, 120) = 3.79	$p = 0.003$
/ɔ/	F2	*F*(5, 120) = 23.95	$p < 0.001$
/u/	F2	*F*(5, 120) = 46.77	$p < 0.001$
/y/	F2	*F*(5, 120) = 4.28	$p = 0.001$
/ʏ/	F2	*F*(5, 120) = 6.40	$p < 0.001$
/a/	dur	(*F*(5, 120) = 3.18	$p = 0.010$
/ɑ/	dur	*F*(5, 120) = 3.14	$p = 0.011$
/i/	dur	*F*(5, 120) = 4.73	$p = 0.001$
/ɔ/	dur	*F*(5, 120) = 5.32	$p < 0.001$
/u/	dur	(*F*(5, 120) = 2.81	$p = 0.020$

4. CLASSIFICATION SIMULATIONS

4.1. Learning algorithm and data sets

To investigate the classification of our newly recorded vowels and previous data, we trained virtual learners on one of three data sets of male NSD monophthongs: the data from [9], from [1] and from the corpus of this study. Our data are further split for the six different consonantal contexts.

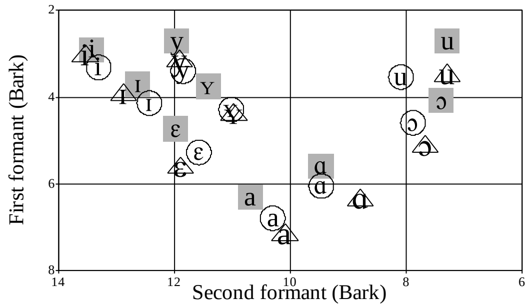
Our virtual learners are represented by an Optimality Theoretic *cue constraint grammar* (see e.g. [4]) which takes pairs of Bark-scale <F1,F2> values as input, and maps it to one of the nine vowel phonemes /i, y, ɪ, ʏ, ɛ, a, ɑ, ɔ, u/. The winning output is determined by a set of ranked constraints which forbid mapping a certain F1 or F2 value to a given vowel. At each learning step, the ranking of these constraints is stochastically convolved with some *evaluation noise*; the output of the grammar for an <F1,F2> dyad is then checked against that of the training data. If the wrong candidate wins, constraints are re-ranked to reduce the probability of an error in the future.

Possible input formant values are discretized in steps of 0.25 Bark: F1 may range from 2 to 9 Bark, F2 from 5 to 15 Bark. Thus, the grammar has (9x(29+41) =) 630 constraints militating over (9x29x41 =) 1198 input values. All constraints are set to an equal ranking of 100.0 before training.

Measurements from the three (sub)sets were converted to pairs of discretized phonetic [F1, F2] Bark values, coupled to one of nine phonological categories. Each vowel measurement is assigned to the four closest values on the [F1, F2] grid, weighted by its inverse Euclidean distance to each of these four points. This 'smoothed' binning prevents problems with sparse input data for data sets with relatively few tokens.

We train the OT grammar on each of the data sets using the standard Praat settings also employed in [4]: 100,000 inputs per training set, with the *plasticity* parameter decaying from 1.0 to 0.001 in 4 steps, and *evaluation noise* set to 2.0 during learning. Next, we test the trained grammars by feeding it the F1 and F2 values of the other sets to see how well the grammar predicts the vowel intended by the speaker.

Figure 2: Average positions of male vowels on the F1/F2 plane for data from Pols [9] (triangles), Adank, et al. [1] (rectangles) and the current study (circles).



4.2. Results

Table 5 gives the average results for 10 runs of the training and testing procedures over all sets unsurprisingly, having a grammar classify the data set it was trained on gives the lowest testing error. However, no data set or subset classifies 100% correct on itself (bold values in Table 5), indicating that the data are not entirely separable on the basis of the first two formants. The within-corpus results indicate that the clearest division between the different consonantal contexts is between alveolar and non-alveolar place of articulation: mutual classification error is higher between these two contexts.

Table 5: Confusion matrix for all training (rows) and testing combinations of the various data sets, averaged over 10 runs. Underlined values are highest classification rate between different data sets.

↓Input	P'73	A'04	all	fVf	kVk	pVp	sVs	tVk	tVt
P'73	**89.2**	57.7	73.2	80.5	78.0	80.5	61.4	69.9	67.7
A'04	51.1	**95.3**	45.3	45.7	49.7	47.3	42.2	39.7	47.3
all	74.7	54.5	**89.8**	88.8	87.6	92.5	90.8	89.8	89.0
fVf	71.2	50.2	84.3	**93.4**	83.5	91.0	74.6	80.1	83.3
kVk	70.3	51.0	81.7	80.5	**95.1**	81.1	72.8	82.7	77.5
pVp	69.5	53.1	82.6	87.9	81.4	**94.6**	71.9	77.4	81.5
sVs	53.0	42.0	79.7	71.0	74.4	69.8	**92.8**	85.4	85.4
tVk	63.7	38.9	83.0	79.6	81.7	78.5	82.9	**92.6**	82.2
tVt	61.4	51.5	83.0	80.3	71.4	82.3	86.9	82.2	**93.4**

Looking at between-corpus classification, we find a number of surprising results. First, mutual classification performance between data set in [1] and all other sets is quite low: less than half of the tokens are classified correctly across the other sets. This poor performance even extends to our data produced in an sVs context, which was the same context as that in Adank et al. Finally, while our subjects did not produce vowels in the hVt context that [9] used, we see that vowels produced in a labial context pattern most closely with these data.

5. DISCUSSION

We collected a corpus of Dutch vowels produced in various consonantal, syllabic and prosodic contexts. Statistical analyses of the spectral and durational properties of Dutch monophthongal vowels show that there are a number of context-dependent effects previously shown for other languages: a significant effect of consonantal context on the first two formants, most pronounced in the 'fronting' of higher back vowels in an alveolar context.

Our results also confirm the results of production studies conducted for other languages (e.g. [3, 11]), which showed large realization differences within vowel categories, especially regarding consonantal context. Unless controlled for, these coarticulatory effects severely limit the extent to which vowel production studies may be compared, either across or within languages.

We used a supervised learning algorithm to test the similarity of our male data with two other Dutch vowel production studies. The results reveal a large mismatch in the values of the first two formants when our data is compared to a recent study, and a smaller mismatch with an older study. These results underscore the difficulty of comparing vowels produced in different consonantal contexts. Even when context is controlled for, however, our formant measurements of Dutch monophthongs differ markedly from those reported in a recent study of the same dialect of Dutch. A larger-scale production study, complemented by a perception study, would be needed to test whether age or unforeseen social/regional pronunciation variation is behind these differences.

Finally, the present study only investigates static spectral properties of monophthongs at the vowel midpoint. To gain a more complete understanding of the various contextual effects on vowel production, dynamic properties of the formants of all 15 Dutch monophthongs and diphthongs must be studied.

6. REFERENCES

[1] Adank, P., van Hout, R., Smits, R. 2004. An acoustic description of the vowels of Northern and Southern Standard Dutch. *JASA* 116, 1729-1738.
[2] Boersma, P., Weenink, D. 2010. PRAAT: Doing phonetics by computer. *http://www.praat.org/*
[3] Chládková, K., Escudero, P., Boersma, P. In press. Context-specific acoustic differences between Peruvian and Iberian Spanish vowels. *JASA*.
[4] Escudero, P., Boersma, P. 2004. Bridging the gap between L2 speech perception research and phonological theory. *Stud. Second Lang. Acquis.* 26, 551-585.
[5] Escudero, P., Boersma P., Rauber, A., Bion, R. 2009. A cross-dialect acoustic description of vowels: Brazilian versus European Portuguese. *JASA* 126, 1379-1393.
[6] Harrington, J., Kleber, F., Reubold, U. 2008. Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *JASA* 123, 2825-2835.
[7] Hillenbrand, J.M., Clark, M.J. 2001. Effects of consonant environment on vowel formant patterns. *JASA* 109, 763-748.
[8] van Nierop, D.J.P.J., Pols, L.C.W., Plomp, R. 1973. Frequency analysis of Dutch vowels from 30 female speakers. *Acustica* 29, 110-118.
[9] Pols, L.C.W., Tromp, H.R.C., Plomp, R. 1973. Frequency analysis of Dutch vowels from 50 male speakers. *JASA* 53, 1093-1101.
[10] Strange, W., Levy, E.S., Law II, F.F. 2009. Cross-language categorization of French and German vowels by naïve American listeners. *JASA* 126, 1461-1476.
[11] Strange, W., Weber, A., Levy, E.S., Shafiro, V., Hisagi, M., Nishif, K. 2007. Acoustic variability within and across German, French, and American English vowels: Phonetic context effects. *JASA* 122, 1111-1129.

[1] The sentences were presented orthographically using unambiguous spellings for Dutch speakers, e.g. "P*o*p. In p*o*p en p*o*ppe zit de *o*" for the vowel /ɔ/.