THE DIPHTHONGAL VOWEL SPACE PARADOX

Sander van der Harst[a,b], *Hans Van de Velde*[a] *& Roeland van Hout*[b]

[a]UiL-OTS; [b]CLS Nijmegen, the Netherlands
S.vanderHarst@uu.nl; H.VandeVelde@uu.nl; r.vanhout@let.ru.nl

ABSTRACT

This paper investigates style differences in the distribution of Dutch diphthongal vowels over the acoustic vowel space. It aims to find out whether a paradox that arises in monophthongs, i.e. that variation between vowels decreases in informal speech and variation within vowels increases without a loss of identifiability, is also found for diphthongal vowels. Our data reveal only a partial paradox, for which we propose an explanation.

Keywords: vowel space size, style, Dutch, diphthongs, sociophonetics

1. INTRODUCTION

A widely reported finding in sociolinguistic studies on stylistic vowel variation is that variation is larger in informal than in formal speech (e.g. [9, 10]). In phonetic studies, it has been shown that the size of the acoustic vowel space is smaller in faster, spontaneous or more informal speech styles [5, 11] and that listeners can still successfully recognize these vowels [3].

When these results are combined, a paradox arises, which has been coined as the ‘Vowel Space Paradox’ [7]: whereas the variation *between* vowels decreases in informal speech, the variation *within* vowels increases, which does not result in a lower identifiability of these vowels.

The abovementioned phonetic studies mainly investigate monophthongs (an exception is [6]). It is therefore unclear whether the paradox can be extrapolated to diphthongal vowels. The first aim of the present paper is to find this out. The vowels selected for this study are the Dutch diphthongs /ɛi, œy, ɔu/ and the long mid vowels /e, ø, o/, which have a diphthongal realization in the Netherlands. The vowels have been elicited from speakers of Standard Dutch in three tasks that were part of a sociolinguistic interview. The three tasks differ with respect to the amount of attention paid to the vowel that the researcher intends to elicit. In logatome reading the amount of attention is high (i.e. the speech reflects more formal speech, cf. [9]), in word list reading it is slightly lower, and in spontaneous speech it is low (i.e. the speech reflects more informal speech). The largest differences in pronunciation are expected between the two reading styles and spontaneous speech. In this paper we will focus on regional variation.

As will be shown in Section 3, the paradox only partially applies to diphthongal vowels. The second aim of this study, therefore, is to explain the paradox for these vowels. Thus, we want to resolve how diphthongal vowels remain identifiable in spontaneous speech with a decreasing vowel space and increasing sociogeographic variation. The third and final aim of this study is to explain why the paradox is only partial.

2. MATERIALS

2.1. Speakers

The 20 speakers in the current study are Dutch language teachers. Ten of them come from the Randstad region (N-R), which is the economic and cultural center of the Netherlands, and the other ten from Netherlands-South (N-S), which is a geographically peripheral region. In each region, we selected 5 male and 5 female speakers, between 22 and 40 years old.

2.2. Logatome reading (LOG)

The vowels that are selected from the logatome reading task were analyzed acoustically in [1]. Her data are used here. In the current task, the speakers had to read out carrier sentences in which the target vowels (indicated by “_” below) occurred three times:

<In s_s en in s_ze zit de _ >
/ ɪn s_s ɛn ɪn s_zə zɪt də _ /
In s_s and in s_ze is the _

Only the token in the first logatome (i.e. s_s) was analyzed. Every speaker performed the reading task twice, resulting in a total of 240 vowel tokens: 2 tokens of 6 vowels produced by 20 speakers.

2.3. Word list reading (WL)

Two sets of six monosyllabic words are selected from a word list that the speakers had to read out. In the first set, the target vowels are followed by /s/ (*mees, neus, boos, ijs, huis, kous*) and in the second set (*beet, neut, boot, spijt, fluit, fout*) they are followed by /t/. These codas are selected (i) to have the same coda as in the logatomes and (ii) to study the influence of the coda on the realization of the vowel. As for the logatomes, a total of 240 vowel tokens were obtained.

2.4. Spontaneous speech (SP)

In the final part of the interview the interviewer had a spontaneous conversation (15 minutes) with the participant. We tried to select 5 tokens per vowel per speaker. In order to keep the vowel tokens in spontaneous speech comparable to the tokens in the word list and logatomes, we tried to select tokens in stressed syllables followed by the same coda (/s/). However, the differences in frequency distribution of vowels and codas in spontaneous speech, urged us to select also vowel tokens followed by other codas (Table 1). A comparison of the vowel tokens in different codas indicated that the differences in codas did not affect our results.

Table 1: The distribution of the selected tokens over the different codas. The total number of tokens is 565.

	/t,d/	/s/	/k/	/p,b/	/f/	/ɣ/	no coda	Total
/e/	95	5	-	-	-	-	-	100
/ø/	-	9	36	-	-	6	20	71
/o/	39	9	1	27	15	2	7	100
/ɛi/	96	4	-	-	-	-	-	100
/œy/	64	32	-	-	-	-	1	97
/ɔu/	30	1	-	-	-	-	66	97

2.5. Recordings

The speakers were interviewed at home or at work (in 1999). The speech was recorded on digital audiotape with a portable TASCAM DA-P1 recorder and an AKG C420 headset microphone. The recordings were digitalized on computer and down-sampled to 16 kHz (16 bits).

2.6. Acoustic measurements

The words containing the selected vowel tokens were segmented at the phoneme level, using the protocol of [12]. The duration of each token was defined as the interval between segment labels.

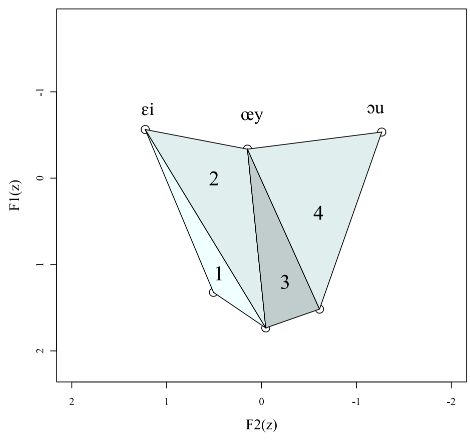
F1 and F2 were measured, using Praat [4], at seven equidistant time points, between 12.5% and 87.5% of the duration of the vowel. All word list formant data were hand checked and approximately 30% of the formant values of the spontaneous tokens were hand checked using the method proposed by [7].

F1 and F2 were normalized using Lobanov's procedure (e.g. [1]), on the basis of the formant values at the midpoint of all tokens of both the monophthongs and the diphthongal vowels of a speaker, from all three styles. By using formant values from all styles, the differences between styles were preserved after normalization.

2.7. Computing vowel space sizes

For each style, the size of the area covered by the vowels was computed. The vowel space sizes were calculated separately for the long mid vowels and the diphthongs, since the respective vowel spaces overlap for some speakers, but not for others. The area covered by the vowels is divided into four smaller areas (cf. [8], see for example Figure 1). The corner points of the first triangle, for instance, are the onset and offset of /ɛi/ and the onset of /œy/. The size of each of the 4 triangles was computed using Heron's method. This method first computes *s* by taking half of the sum of the three sides *a*, *b* and *c* (see (1)), and then computes the size of the triangle area, as in (2). Finally, the size of the whole vowel space is computed by summing the sizes of the four smaller triangle areas.

Figure 1: The triangles used for the calculation of the size of the vowel space area of the diphthongs.



(1) $$s = \frac{(a+b+c)}{2}$$

(2) $$\text{Triangle Area} = \sqrt{s(s-a)(s-b)(s-c)}$$

3. RESULTS

In this section, the role of style in the vowel space size (3.1), positioning of diphthongal vowels (3.2), amount of sociogeographic variation (3.3) and the identifiability of diphthongal vowels (3.4) is investigated. Except for Section 3.3, regional differences will not be discussed. The logatomes will be referred to as "LOG", the word list as "WL" and spontaneous speech as "SP".

3.1. Vowel space sizes

Two repeated-measures ANOVAs were run (i.e. one for the long mid vowels and one for the diphthongs), with vowel space size as the dependent variable, style as the within-subjects factor and region and gender as the between-subjects factors (p<.05).

For both the long mid vowels ($F_{1.5,24.6}$=9.812, p=.002, partial η^2=.380, Huynh-Feldt corrected) and the diphthongs ($F_{2,32}$=63.634, p=.000, partial η^2=.799), a larger vowel space is found in WL than in the other two styles (Table 2). Surprisingly, LOG yielded a smaller vowel space size than SP for the long mid vowels. For the diphthongs, LOG and SP do not differ significantly.

Table 2: Average vowel space size (in squared z-units) of the long mid vowels and diphthongs, by style.

	Long mid	Diphthong
Logatomes	0.91	1.73
Word list	1.45	3.62
Spontaneous	1.11	2.31

3.2. Position of diphthongal vowels

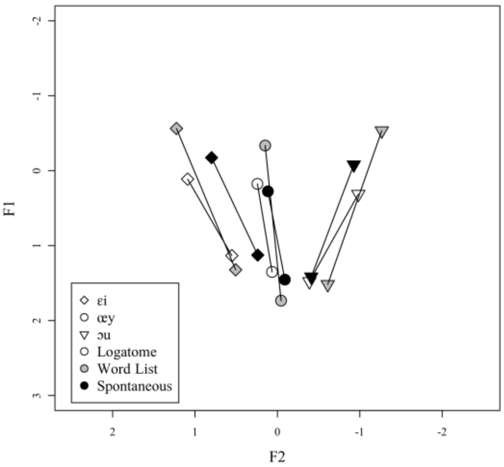
In order to find out which vowels were responsible for the changes in vowel space size, the positions of the diphthongal vowels were compared across styles. Six series of four repeated-measures ANOVAs were carried out: for each vowel, F1 and F2 at both onset (at 25% of the vowel duration) and offset (at 75%) were analyzed. For each analysis the speaker mean F1 and F2 values per style were computed first, since the number of tokens differed between speakers and styles. The same independent variables were used as in Section 3.1. Only partial η^2 values (given as 'η^2' below) are reported for the significant style effects.

At the onset, only shifts in F2 showed up. For /ɔu/ (η^2=.606), the onset F2 in SP and WL was lower than the LOG onset F2. /e/ (η^2=.579) and /ɛi/ (η^2=.601) showed a lower onset F2 in SP than in LOG and WL.

At the offset, F1 was lower in WL than in LOG for /e/ (η^2=.343) and it was lower in WL than in LOG and SP for /ɛi/ (η^2=.662), /œy/ (η^2=.564) and /ɔu/ (η^2=.526). F2 was lower in SP than in (one of) the reading styles for the front vowels /e/ (LOG only; η^2=.454), /ø/ (η^2=.508), /ɛi/ (η^2=.724), /œy/ (WL only, η^2=.371). For /ɔu/ (η^2=.359), F2 in SP was higher than in WL.

Figure 2, which shows the onset and offset positions of the N-R diphthongs, serves as an illustration of the observed differences.

Figure 2: The onset and offset normalized F1 and F2 of the N-R diphthongs, given for each style.



3.3. Sociogeographic variation

The differences in the amount of sociogeographic variation across styles was investigated by first computing the amount of acoustic (i.e. F1 and F2) variation *at the level of the speaker* in each style and running a series of ANOVAs (using the same factors as in 3.2). An increase of acoustic variation for the onset or offset of a vowel was interpreted as an increase of sociogeographic variation when (i) a significant difference between N-R and N-S for the same time point and for the same formant of that vowel was found in our data (cf. Bell's [2] Style Axiom), or (ii) when for that vowel and that formant sociogeographic variation has been observed in the literature.

Table 3 shows that all diphthongal vowels, except /ø/, showed an increase of acoustic variation in SP. The increase of F1 variation at the onset was interpreted as an increase of sociogeographic variation, since onset F1 showed regional differences for these vowels in our data. Since the literature reports on stronger gliding on the height dimension in certain speaker groups

[13], the offset F1 differences are interpreted in the same way.

Table 3: Differences in the amount of acoustic (marked as X>Y) or sociogeographic (idem, in grey cells only) variation between styles.

	F1		F2	
	onset	offset	onset	offset
/e/	-	SP>LOG	-	-
/ø/	-	-	-	-
/o/	SP>LOG	-	SP>WL	SP>WL,LOG
/ɛi/	SP>WL,LOG	-	-	SP>LOG
/œy/	SP>LOG	-	-	-
/ɔu/	SP>WL,LOG	SP>WL	-	SP>WL,LOG

3.4. Identifiability

For each speaker, a linear discriminant analysis with F1 and F2 at the onset and offset of the diphthongal vowels as predictors, was run to find out whether the reported changes across styles affected the identifiability of the vowels. The analyses showed that in LOG and WL, the vowels are (nearly) perfectly recognizable (success rates of 99.6% and 100%, respectively). In SP the success rate has dropped (92.9%), which is mainly caused by confusions between (i) long mid vowels and their diphthong counterpart (25 out of 40 confusions, of which 13 for the back vowels) and (ii) front unrounded diphthongal vowels and their rounded counterparts (11 confusions).

4. DISCUSSION AND CONCLUSION

In order to test whether the paradox is found for the diphthongal vowels in our data, the results presented above are combined. First, it was found that a smaller vowel space was only found when SP was compared to WL. This result is different from the monophthongs, which yield a larger vowel space for both WL and LOG. The small diphthongal vowel space of LOG, in which the level of monitoring is highest, in our data suggests that strong glides in F1 are still stigmatized [13]. Second, for a subset of vowels an increase of sociogeographic variation is found.

These two changes combined resulted in a relatively small loss of identifiability of diphthongal vowels in SP. We may therefore conclude that the paradox at least partially (i.e. when WL and SP are compared) holds for the diphthongal vowels.

In order to explain the paradox that a shrunken vowel space and an increase in vowel-specific sociogeographic variation did not result in a strong loss of identifiability, we take a closer look at changes at the level of the vowel. It appears that the changes surface mainly on different dimensions. That is, whereas the vowels mainly shift positions in F2, the increase of sociogeographic variation is found in F1. In this way, the overlap between vowels is kept minimal in SP. Only for the offset of /ɔu/, a shift in F1 is combined with an increase of variation, which results in the highest amount of confusions (i.e. between /ɔu/ and /o/). These confusions are relatively unproblematic, since diphthongs and long mid vowels can still be distinguished on the basis of their duration. Note as well that the front unrounded and rounded vowels can still be distinguished by also including F3.

In sum, the paradox can be explained by the fact that the changes in vowel space and the amount of variation show up on two different dimensions. The cases in which the changes take place on the same dimension can be resolved by including other factors than F1 and F2.

5. REFERENCES

[1] Adank, P. 2003. *Vowel Normalization*. *A Perceptual-Acoustic Study of Dutch Vowels*. Ph.D. dissertation. Katholieke Universiteit Nijmegen.

[2] Bell, A. 1984. Language style as audience design. *Language in Society* 13(2), 145-204.

[3] van Bergem, D. 1995. *Acoustic and Lexical Vowel Reduction.* Ph.D. dissertation, University of Amsterdam.

[4] Boersma, P., Weenink, 2007. D. Praat: doing phonetics by computer (Version 4.5.11). *http://www.praat.org/* Retrieved by January, 2007.

[5] Bradlow, A. 2002. Confluent talker- and listener-oriented forces in clear speech production. In Gussenhoven, C., Warner, N. (eds.), *Laboratory Phonology* 7, 241-274.

[6] Gay, T. 1968. Effects of speaking rate on diphthong formant movements. *J. Acoust. Soc. Am.* 44, 1570-1573.

[7] van der Harst, S. 2011. *The Vowel Space Paradox. A Sociophonetic Study on Dutch*. Ph.D. dissertation, Radboud University Nijmegen.

[8] Jacewicz E., Fox, R.A., Salmons, J. 2007. Vowel space areas across dialects and gender. *Proc. 16th ICPhS* Saarbrücken, 1465-1468.

[9] Labov, W. 1972. *Sociolinguistic Patterns*. Oxford: Blackwell.

[10] Labov. W. 2001. *Principles of Linguistic Change: Social Factors*. Malden: Blackwell.

[11] Lindblom, B. 1963. Spectrographic study of vowel reduction. *J. Acoust. Soc. Am.* 35(11), 1773-1781.

[12] van Son R.J.J.H., Bolotova, O., Lennes, M., Pols, L.C.W. 2004. Frequency effects on vowel reduction in three typologically different languages. *Proc. Interspeech 2004*, 1277-1280.

[13] van de Velde, H. 1996. *Variatie en verandering in het gesproken Standaard-Nederlands*. Ph.D. dissertation, Katholieke Universiteit Nijmegen.