DUTCH DIPHTHONG AND LONG VOWEL REALIZATIONS AS CHANGING SOCIO-ECONOMIC MARKERS

Irene Jacobi, Louis C.W. Pols, Jan Stroop

Amsterdam Center for Language and Communication, University of Amsterdam, Netherlands
i.jacobi@uva.nl

ABSTRACT

To judge the influence of speaker background on the quality of five long vowels and diphthongs /o^{ʊ}/, /e^{ɪ}/, /ɑu/, /ɛɪ/, and /ʌy/ in Standard Dutch, the spectra of these vowel realizations in spontaneous speech were measured for 70 subjects, and analyzed with regard to the subjects' regions of education and residence, their level of education and occupation, as well as their sex and age. Our preliminary analyses indicated that high educated speakers lowered (higher F1, articulated more open) the long vowels and diphthongs and diphthongized them stronger than low educated speakers. There seemed to be effects of age but more data were needed to specify these.

The new data show that besides the level of education or occupation, the factor 'age group' has a major effect on the variations in speech production. The vowel attributes 'onset' and 'degree of diphthongization' were affected variably by speaker background data. Speakers of the older generation hardly differed in their realization behavior, contrary to the younger and middle aged speakers. The higher educated displayed systematic age patterns, the lower educated did not. A slight effect of region of residence was found for the females from the north peripheral residence region, who had higher onsets for /o^{ʊ}/. An effect of sex was found within the youngest age group, where higher educated males differed from higher educated females by their stronger diphthongization of /e^{ɪ}/ and /o^{ʊ}/. The vowel variations that were related to age reflected several pronunciation changes in progress.

Keywords: Dutch diphthongs, acoustic quality, speaker background, variation, change

1. INTRODUCTION

The Dutch long vowels /eː/, /øː/, and /oː/ are traditionally transcribed as steady-state vowels, though being realized slightly closing /e^{ɪ}/, /ø^{ʏ}/, /o^{ʊ}/. In a phonetic description from 1983 [7], these long vowels were alluded to as retaining a narrow glide within conservative Standard Dutch (ABN), but being realized increasingly more extended by younger mainstream speakers of Standard Dutch. While steady-state realizations were said to be restrained to areas outside the central conurbation of the Netherlands, the popular speech in the central conurbation (the 'Randstad' speech) was mentioned to be socially marked by wide diphthongs.

At the end of the 1990s, attention was drawn to the ear-catching lowering of the diphthong /ɛɪ/, called Polder Dutch (also avant-garde Dutch). For a homogenous high educated group, the more open variety of /ɛɪ/ with higher F1 onset values was found to predominate the females rather than the males [4] (also cf. [1]). Our recent study revealed that not only /ɛɪ/ shows class-dependent differences for males and females, but that /ɑu/, /ʌy/, /e^{ɪ}/, and /o^{ʊ}/ also differ in terms of onset and diphthongization, depending on the speaker's level of education [5]. Speakers who lowered the genuine diphthongs /ɑu/, /ɛɪ/, /ʌy/ also lowered /o^{ʊ}/, and /e^{ɪ}/ and diphthongized them to a larger extent. These speakers differed from other speakers in having a higher level of education. No significant effects of age were found. In that study, speaker regions had been neglected under the assumption that the corpora included only Standard Dutch speakers and no dialectal speakers.

The present study is also based on Standard Dutch speakers. Yet, in an informal test some of the speakers from our data pool were perceived as speakers with some dialect coloring, and so we also wanted to consider the influence of regional factors. Here, the research on the long vowels and diphthongs of spoken Standard Dutch is continued by taking into account more meta data, including information on the speakers' regions of residence, and many more speakers to form properly filled age groups.

2. DATA

2.1. Speech material

The speech material was taken from the Corpus Gesproken Nederlands (CGN), a Standard Dutch speech corpus with more than 5.6 million words, collected and recorded in the Netherlands around the year 2000 [8]. From various subcorpora available within the CGN we selected spontaneous speech of recorded gatherings, interviews, discussions or private conversations of 70 speakers altogether. All stressed realizations of /ɑu/, /ʌy/, /ɛɪ/, /ø^{Y}/, /o^{u}/, and /e^{I}/, were measured, as well as all stressed realizations of the anchor vowels /a/, /i/, /u/ for a speaker normalization of the data. The criterion of stressed realizations of the vowels was based on lexical stress and a minimally required duration of the assigned vowel segment. Vowel labels and boundaries were based on the available CGN segmentations and annotations [3], using a broad transcription based on SAMPA. With one symbol representing one Dutch phoneme, each symbol thus comprises all variants of the assigned phoneme. We checked all data manually and very rarely we had to exlude suspect segmentations or transcriptions. Furthermore, the speech data were checked for speaker overlap or signal distortions. The vowels were taken from various phonetic contexts, and only vowels from potentially strongly coarticulating environments were excluded, e.g. those followed by final velarized /l/ or retroflexal /r/.

2.2. Meta data

All 70 speakers were attributed Standard Dutch as the language variety they used during the recorded samples, as well as at home, at work, and as the language variety they were raised in. Of the 70 speakers, 35 speakers were females and 35 males. Within the males, 17 speakers had a high and 18 a low education. Within the females 18 were highly educated and 17 lowly. From the meta data of the CGN we further retrieved each speaker's year of birth, the region of residence, and the region of education.

The regions of residence and education taken from the CGN are split into four: the central region, a transitional region, a south peripheral region, and a north east peripheral region. A speaker's region of education was assigned by the region where the speaker lived during secondary education (at the age of 4 to 16). The level of education (high or low) matched the level of occupation (high or low) in almost all cases and thus only one of the two factors was chosen, the level of education. The speakers were born between 1926 and 1981, and split into three age groups: speakers older than 55, speakers up to 35 years of age, and those inbetween.

3. METHOD

3.1. Spectral analysis

The vowel segments of the 70 speakers were digitally analyzed by means of a principal component analysis (PCA) on the barkfiltered spectra, and intensities were equalized at 80 dB. The measurements were done with the Praat program [2]. The three corner vowels /a/, /i/, /u/ were analyzed at the mid of the steady state phase. The diphthongs and long vowels were analyzed at one tenth and nine tenth of their total duration, thus preserving rather unidirectional spectral transitions, and leaving out the very first and last frames and the strongest coarticulatory effects.

The spectrum of each sound segment was filtered up to ca. 4200 Hz by using 18 bark filters. This overall bandwidth covers the important information concerning the vowel quality, including the area of the first, second and third formant (also e.g. F3 of a high female /i/). Higher formants include mainly speaker-specific information. Each filter covered an area of one bark and subsequent filters overlapped at -3 dB. Due to possible strong variance caused by the speakers' varying fundamental frequency, the first two filters were represented by one mean intensity. As the stressed anchor vowels are hardly influenced by sound changes and by the speakers' individual speech style, we took all speakers' /a/, /i/, and /u/ means to calculate the PCA dimensions that were used to analyze all vowels furtheron. No hand corrections were applied.

3.2. Normalization

Male and female speakers were comparable in the calculated pc-dimensions. The CGN includes a diversity of recording qualities which resulted in shifted locations of our speakers' /a/-/i/-/u/ triangles in the pc-plane. To make the triangles comparable, each speaker's /a/-/i/-/u/-triangle focal point was set to 0. The first dimension then explained 65% of the variance, the second added another 25%, and by the third dimension no more than 3% was added. Spectra recalculated from the pc-values display the pc-relations with formant and antiformant areas and allow for an interpretation in articulatory terms. For the data set of more than 12,400 measured long vowels and diphthongs, the resulting values of the first dimension correlated with F1 in bark (r=.70), the second with F2 in bark (r=.72).

Although no significant sex differences were found in the first dimension, there were some sex differences in the second dimension concerning high (in terms of tongue position) vowels. These differences could as well have been caused by the recording quality, as the males included more recordings with bad quality than the females. To split between the effects of sex characteristics and recording quality, the influence of noise was tested by degrading the quality of recorded speech samples. The higher the minimum dB in the filters was set, the more the high vowels shifted into the direction of low front vowels, ultimately resulting in a mere point in the plot for experimentally increased minimum filter values higher than 60 dB. The area of the back vowel /u/, characterized by a spectrum with little energy in the lower part of the spectrum and the absence of energy in higher parts, was the first to change its position by added noise. Whereas the first eigenvector of the PCA on the barkfiltered anchor vowels explained most of the variance in the lower parts of the spectrum, the second eigenvector explained most of the variation in the higher part of the spectrum. In conclusion, the differences in the second dimension were mostly due to recording quality.

To normalize for the remaining differences in individual vowel space sizes, the 'relative onset' of a speaker's long vowel or diphthong was defined by its distance to his/her anchor vowels in each dimension (see below equation (1)). For front vowels, anchors were /a/ and /i/, for back vowels the choice also fell on /a/ and /i/, to be uninfluenced by possibly deviant /u/-values due to poor signal to noise ratios. The 'degree of diphthongization' was defined by the distances the vowel's movement covered per dimension in relation to the speaker's /a/-/i/ distance (see below equation (2)). The normalized values were calculated as follows:

$$(1) \quad relative\ onset = \frac{pc_a - pc_{onset}}{pc_a - pc_i}$$

$$(2) \quad rel.\ diphthongization = \frac{pc_{onset} - pc_{offset}}{pc_a - pc_i}$$

The normalization procedures thus ruled out most of the differences between the speakers' vowel spaces that were caused by physical characteristics or diverse recording qualities.

4. RESULTS

Speaker-independently, a linear relation was found for the onset and diphthongization of /oʊ/ and /eɪ/: the lower their onset positions, the stronger their diphthongization. The other vowels showed less speaker-independent linearity.

Statistical analyses on the speakers' relative mean vowel positions and diphthongizations in pc1/pc2 revealed numerous effects of level of education, age, region, and sex. Besides multivariate analyses of variance, Anova's and post-hoc's were used on the data to further establish differences between subgroups. Effects found in the Manova's were mainly caused by the first dimension. In the following we concentrate on the relative values based on pc1, which explained most variance in the /a/-/i/-/u/ space. The normalized pc1-value correlates strongly with the first formant in bark and so it is the most important indicator for lowering. Since some factors affected the same vowel values, the speakers were split into subgroups. The further analyses on the first dimension then displayed different effects for 'diphthongization' and 'onset': the level of education significantly influenced the onsets of /oʊ/, /eɪ/, and /ɑu/, and affected the degree of diphthongization of /oʊ/, /eɪ/, /ɛɪ/ and /ʌy/. The only effect of region was found for females from the north peripheral region, who had higher onset values for /oʊ/.

4.1. Differences within age groups

4.1.1. Elder speaker group

Within the oldest age group (aged > 55), there were no significant pronunciation differences for the single vowels and diphthongs, neither between higher or lower educated speakers, nor between females and males. Yet, overall, the higher educated speakers displayed slightly lower onsets and a stronger degree of diphthongization for all vowels. The higher educated speakers from the central region had the lowest onsets and strongest diphthongizations of /oʊ/ and /eɪ/.

4.1.2. Middle aged speaker group

The middle aged group (speakers between 35 and 55 years of age) showed highly significant differences between lower and higher educated speakers for the degree of diphthongization of all vowels; the higher educated diphthongized them much stronger than the lower educated (for /oʊ/ compare Fig.1). Of all vowels only the diphthongization of /ɑu/ within the males did not reach significance.

Regarding the onset values, high significance was found for the onsets of the long vowels /oʊ/ and /eɪ/: the higher educated articulated them more open than the lower educated, and, concerning /oʊ/, females more than males. Other significant onset value differences were only found for the females and their onsets of /ɑu/ and /ʌy/; higher educated females lowered more than lower educated females.

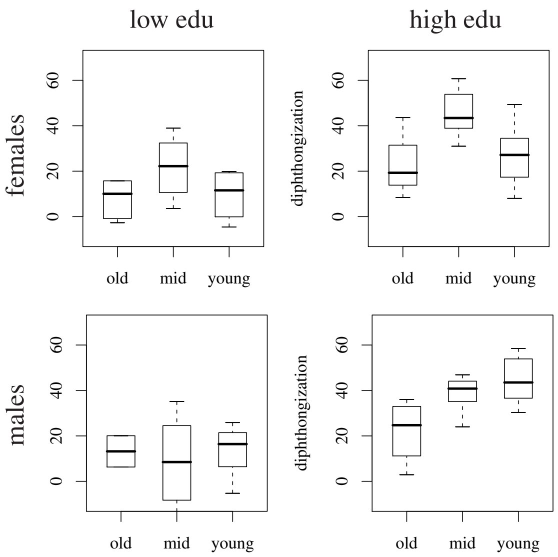
4.1.3. Young speaker group

Within the youngest group of speakers (aged up to 35) there were significant differences between the educational groups, and differing patterns between the sexes. The higher educated males diphthongized all vowels, except for /ɑu/ to a stronger extent than the lower educated males; also, their onset of /oʊ/ was significantly lower. For the females, there were no significant differences between higher and lower educated speakers. Within high educated speakers, the differences between the females' and males' diphthongization of /oʊ/ and /eɪ/ reached almost significance.

4.2. Differences between age groups

For the higher educated speakers, differences between the old and mid generation in the degree of diphthongization were significant for all vowels but /ʌy/ (for /oʊ/ compare Fig.1, plots on the right side). For the lower educated speakers, differences between the old and mid genereration were only found for the females; their diphthongization of /ɑu/ decreased significantly. Concerning the vowel onsets, differences from the old to the mid generation were significant for the males and their /ɛɪ/ and /ɑu/ onsets, which were lowered.

Figure 1: The degree of diphthongization of /oː/ (/oʊ/) illustrated by boxplots for females (top) and males (bottom): To the left, plots of the lower educated speakers, to the right the higher educated, each split into age groups 'old', 'mid', 'young'.



From the mid to the young speaker group, significant differences were found for the high educated females' diphthongization of /oʊ/ and /eɪ/: whereas for the males the degree of diphthongization increased from generation to generation, the females' degree decreased again. The vowels /oʊ/ and /eɪ/ in contrast were diphthongized to a even stronger extent by the young males than by the males of the mid generation (for /oʊ/ see lower right panel in Fig.1). The high educated females' onsets of /oʊ/ and /ɑu/ decreased significantly from the mid to the young generation, whereas the males' onsets were stable.

5. DISCUSSION AND CONCLUSION

The analyses displayed several differences and changes over time for the two educational (occupational) groups and sexes. The higher educated were affected stronger by the tested factors than the lower educated, who were hardly affected at all. The influence of age on the realization behavior showed up as an interaction effect and did not appear overall regularly for all vowels and speaker groups. The effect of age is thus not interpreted as a degrading one, but as one reflecting changes in progress. These changes show, that, with the generations, pronunciation differences between higher and lower socio-economic classes increased. Contrary to the dispersed possibilities of lower education, the high density of higher education in the big cities of the central area ('Randstad') could have strengthened the effect of a uniform expression pattern for the higher social classes. Interestingly, within the oldest generation, it were the higher educated speakers of the central region who showed the lowest onsets and strongest diphthongizations, that later on became the general characteristic of higher educated speakers. Though both sexes showed similar age patterns, sex differences started to appear within the younger generations. The greater sex differentiation in the higher socio-economic hierarchy that has been stated frequently (e.g. [6]), was also seen in our data. Altogether, age, education/occupation, and sex significantly affected the realizations of the long vowels and diphthongs, and explain for their variation in Standard Dutch.

6. REFERENCES

[1] Adank, P., van Hout, R., van de Velde, H. 2007. An acoustic description of the vowels of northern and southern standard Dutch II: Regional varieties. *J. Acoust. Soc. Am.*, Vol. 121, No. 2, 1130-1141.
[2] Boersma, P., Weenink, D. 2006. *Praat* (Version 4.5.02) [Computer program]. http://www.praat.org/.
[3] Demuynck, K., Laureys. T., Gillis, S. 2002. Automatic Generation of Phonetic Transcriptions for Large Speech Corpora. *Proc. ICSLP,* Denver, Vol. 1, 333–336.
[4] van Heuven, V.J., van Bezooijen, R., Edelman, L. 2002. The pronunciation of /ɛɪ/ by male and female speakers of avant- garde Dutch. *LIN,* 19, 61-72.
[5] Jacobi, I., Pols, L.C.W., Stroop, J. 2006. Measuring and Comparing Vowel Qualities in a Dutch Spontaneous Speech Corpus, *Proc. ICSLP,* Pittsburgh PA, 701–704.
[6] Labov, W. 1990. The intersection of sex and social class in the course of linguistic change. *Language Variation and Change*, Vol. 2, 205-254.
[7] Mees, I., Collins, B. 1983. A phonetic description of the Vowel System of Standard Dutch (ABN), *JIPA,* Vol. 13, 64–75.
[8] Oostdijk, N., Goedertier, W., van Eynde, F., Boves, L., Martens, J.P., Moortgat, M., Baayen, H. 2002. Experiences form the Spoken Dutch Corpus project, *Proc. LREC,* Las Palmas, 340–347.