Production and perception of the Dutch /ɑ/-/a/ continuum: variation within and between speakers

Marieke Ebeling-van den Heuvel

University of Amsterdam

Thesis for the Research Master's in Linguistics

Student number:	6198783
Supervisor:	Prof. Dr. P.P.G. (Paul) Boersma
Second Reader:	Dr. S.R. (Silke) Hamann

23 June 2014

~ To Bas

Contents

Preface

I would like to use this preface to thank everyone who has contributed to this thesis:

- The 22 people who have lent me their voices and choices. Thank you very much for participating in this study! I am a data-gatherer and without you there would not have been data.

- My supervisor Paul Boersma, for forcing me to try to find things out by myself, for letting me make my own choices throughout the study, and for helping me to get funding for the experiments.

- Dirk Jan Vet, for providing me with a useful Praatscript and helping me in making sure the sound recordings went well.

- The people at the Phonetics Lab meetings (26 March and 7 May 2014) and the Current Issues sessions at the University of Amsterdam, where I presented earlier versions of my work, for asking questions and giving me new ideas. Especially the people at the Lab meeting of 7 May have given me renewed enthusiasm.

- Bas, for many reasons. For pronouncing everything, including my name, with [ɑ] (which gave me the idea for this thesis), for listening to me when I was happy about this thesis, for trying to make me laugh and helping me when I was frustrated about it, and for perceiving the world in box plots. You are the person who has helped me the most throughout my work on this thesis. Therefore, I dedicate this thesis to you (not that that is of any use to you, but I thought I'd seize the opportunity).

Abstract

In Dutch, the vowels /a/ and /ɑ/ behave as separate phonemes, but in particular contexts this contrast does not exist as obviously as in minimal pairs. The focus in this study lies on the question of whether the alternation of the vowels in such contexts lies in between-speaker and/or within-speaker variation. A production experiment and a classification experiment were carried out, containing control words, where no alternation of [a] and [ɑ] was expected (words from minimal pairs such as <ban> (/bɑn/) and <baan> (/ban/)), and target words, where alternation was expected (such as <ˈpa.pri.ˌka> and <ka(s).ˈstan.je>). Measurements carried out on the production data were the vowels' F1, F2 and duration relative to the word's duration. It was found that while the vowel distributions of the control words are bimodal, the distributions of the vowels in the target words are monomodal. Concerning between-speaker variation, compared to their own produced control vowels, some speakers tend to use more sounds that are more similar to [a] for the target words, another group of speakers tends to use more sounds more similar to [ɑ] for those words and yet others pronounce most sounds in between [a] and [ɑ]. When production data is compared to classification data, it seems that there exist two largely overlapping vowel categories on the surface. Concerning within-speaker variation, it was found that there is more variation within part of the target words than within control words and other target words. It is argued that on the level of the language as a system there is no clear underlying vowel for the target words, but that individual speakers have an underlying representation of either /a/ or /ɑ/ for most of these words. Which vowel is the underlying one differs per speaker per word. The fact that the target vowels are often in unstressed position or under secondary stress, instead of primary stress as is the case for the control words, together with the lack of necessity of expressing a semantic contrast causes the underlying categorical distinction to be less clearly visible on the surface for the target words than for the control words, leading to the monomodal distributions. Instead of seeing the variation in terms of reduction from [a] to [ɑ], it is argued that the alternation is the result of archiphonemic neutralization on the level of the language as a system.

1 Introduction

The goal of this study is to investigate the variation in the usage of the Dutch vowels /a/ and /ɑ/. These two sounds normally behave as separate phonemes, but this contrast seems to be neutralized in particular contexts. The most important question addressed in this study is whether the observed alternation of the vowels lies in between-speaker and/or within-speaker variation. The other question that is addressed concerns the way the vowels may be represented in the brain on a phonological level.

Examples of the observed alternation are given in (1-7)[1]. As illustrated by examples of minimal pairs in (1-2), long/tense /a/ and short/lax /ɑ/[2] are used contrastively in Dutch.

(1)	a.	/man/	b.	/mɑn/
		'moon'		'man'
(2)	a.	/van/	b.	/vɑn/
		'banner'		'of/from'

The words in (1a) and (2a) are always pronounced with [a], while the words in (1b) and (2b) are always pronounced with [ɑ]. The same holds for a group of words that are not part of minimal pairs. The examples in (3) always surface with [a] and never with [ɑ], while the examples in (4) always surface with [ɑ], and never with [a].

(3)	a.	[kʊat.ˈar.dəg]	b.	*[kʊɑt.ˈar.dəg]
		'evil'		*'evil'
	c.	[ˈwa.tər]	d.	*[ˈwɑ.tər]
		'water'		*'water'
(4)	a.	*[ˈkan.də.ˌlar]	b.	[ˈkɑn.də.ˌlar]
		*'candlestick'		'candlestick'
	c.	*[ˈa.pəl]		[ˈɑ.pəl]
		*'apple'		'apple'

On the other hand, there exists a group of words in which [a] and [ɑ] may be used in free variation, such as the words in examples (5-7) below. In some cases, such as <pa.ˈtat> ('fries'), the vowel may also be realized as schwa (see (7c)).

(5)	a.	[ˈpa.pri.ˌka]	b.	[ˈpɑ.pri.ˌka]
		'paprika'		'paprika'

1 Examples (1-7) are based on informal observations made from the pronunciation by some native speakers and their judgements.

2 The question of whether the Dutch /a/-/ɑ/ contrast should be described in terms of lax vs. tense or long vs. short has been widely discussed in the literature (see e.g. Botma & van Oostendorp 2012). This question is beyond the scope of this paper and the contrast will be described as an opposition between tense /a/ and lax /ɑ/ because of the simple reason that the opposition is not only one of quantity, but also of quality (see e.g. Botma & van Oostendorp 2012, and measurements by Koopmans-van Beinum (1980) and Van Leussen et al. (2011)).

(6) a. [ˌa.po.ˈtek] b. [ˌɑ.po.ˈtek]
'pharmacy' 'pharmacy'

(7) a. [pa.ˈtɑt] b. [pɑ.ˈtɑt] c. [pə.ˈtɑt]
'fries' 'fries' 'fries'

To summarize these data, there are (at least) three groups of words that are relevant for describing the context of the alternation of /a/ and /ɑ/ in Dutch. Firstly, there are words that are part of minimal pairs that always contain either [a] or [ɑ] on the surface. Secondly, there is a group of words that are not part of minimal pairs and still always contain either [a] or [ɑ] on the surface. Finally, there is a group of words in which the two low vowels can be used in free variation. Similar alternations occur for other vowel pairs, be it less commonly, namely /o/ and /ɔ/, /e/ and /ɪ/, and /i/ and /ɪ/ (Kager 1989).

Many interesting questions arise from these data. The topics and questions that are handled in the current study are:

1) Is the variation present between speakers or within speakers (or both)?

The hypothesis is that there exists variation both within and between speakers. Concerning between-speaker variation, for almost every word, each speaker is expected to use either mainly an [a]-similar sound, or an [ɑ]-similar sound. As for within-speaker variation, for some words it is expected that some speakers use very variable sounds, both [a]-similar and [ɑ]-similar sounds, or sounds 'in between' [a] and [ɑ]. The part of this hypothesis about between-speaker variation is falsified if every speaker uses the same vowel for each specific word, i.e., if everyone uses e.g. [a] for <ˈpa.pri.ˌka>, or [ɑ] for <ˌa.bri.ˈkoos>. The part of this hypothesis about within-speaker variation is falsified if speakers consistently use one type of sound for a word, i.e., if they always use an [a]-similar sound, or always use an [ɑ]-similar sound or always a sound in between. Their target vowels are seen to be consistent if the speakers do not show more variation for the target words than for the control words.

2) How is the sound in the words that allow variation represented in the brain?

The hypothesis is that for each speaker, some words have an underlying phoneme /a/, while other words have an underlying phoneme /ɑ/. For some speakers, some words have an underlying archiphoneme /A/ or an underlying feature [+low]. The first part of this hypothesis is falsified if speakers classify the A-vowel in the words inconsistently. Inconsistent classification could indicate the absence of an underlying phoneme contrast. Consistent classification could indicate that a vowel in a word is not represented only by a feature [+low], but by a set of features or a phoneme.

In the next section the literature on the alternation is discussed. Then section 3 discusses the methodology of this study. Two experiments are described: a production and a classification experiment. Next, in section 4 the results of these experiments are described and analyzed. Sequentially, section 5 gives conclusions by answering the research questions. Finally, section 6 discusses the findings of the current study in relation to earlier literature on the topic and provides ideas for future research.

2 Literature

The existing literature on the alternation of [a] and [ɑ] mainly focuses on answering the question of what is the exact context in which the variation is possible. Authors link the phenomenon to theories of stress and prosody of Dutch (e.g., Kager 1989, Geerts 2008). The current study focuses on a less well studied aspect of the phenomenon, by investigating variation in pronunciation and perception between different speakers and within speakers (question 1). Furthermore, while previous studies assume it to be straightforward which vowel is the underlying one in different types of words, this study will consider other possibilities (question 2).

According to Kager (1989), the alternations are caused by vowel reduction from long to short vowels, and in some cases to schwa. As can be clearly seen from the tense-lax vowel pairs mentioned by him, namely /o/ and /ɔ/, /e/ and /ɪ/, and /i/ and /ɪ/, an explanation in terms of durational reduction only does not suffice, since, if it were to be true that the lax vowels are reduced varieties of the tense vowels, the quality of the vowel is affected in this reduction, as well. Kager refers to Martin (1968), who points out that there is a reduction in quality. Martin, focusing on the pair of vowels as pronounced in <bek> and <beek> (/bɛk/ and /bek/, respectively), maintains that [ɛ] is an intermediate stage in the transition from [e][3] to [ə]. Referring to Scharpé (1912), he argues that evidence for this observation is that [e]-[ɛ]-[ə] form a line in the vowel space. Kager implies this is also the reason for the other vowel alternations. However, the problem with this argument is that the vowel pairs actually do not

3 Transcriptions [ɛ] and [e] are mine. Martin refers to the sounds in Dutch <bek> and <beek>, respectively. Kager discusses a vowel pair that he indicates with /e/ and /ɪ/. It is not entirely clear whether Kager and Martin are referring to the same vowel pair with these different transcriptions. There are some words that occur in both of their writings, such as <liberaal> and <acetyleen>. Martin discusses the possibility of 'reduction' of the <beek> (/bek/) vowel to the vowel as in <bid> (/bɪt/) shortly in a separate paragraph. Kager and Martin seem to discuss the same process for similar vowels, but to classify the vowels differently from each other.

lie on a straight line with schwa in the vowel space. See for example F1 and F2 measurements by Koopmans-van Beinum (1980) or more recent measurements by van Leussen et al. (2011), who also compare their measurements to earlier data from Pols et al. (1973) and Adank et al. (2004). Such measurements indicate that the formation of a line from, for example, /a/ to /ɑ/ to the center of the vowel space would require a considerably sharp curve. The questionability of this argumentation leads to the question of whether this phenomenon of alternation really is a result of reduction from long to short (or tense to lax) vowels (and in some cases sequentially to schwa) or whether there is another reason for the observed variation. Furthermore, an explanation in terms of reduction suggests that the tense vowel is the original underlying vowel in these words, while the question is not explicitly answered why this should be the case. These issues need to be investigated more thoroughly. The problem of which vowel is the underlying one is taken into consideration in the current study by addressing question 2.

Aside from the question of how the pattern should be explained, it is, as Kager mentions, not agreed upon in what contexts exactly these alternations occur. A problem that is discussed in the literature, is in what contexts the alternation between /a/ and /ɑ/ exists and in what contexts reduction to schwa may occur. Booij (1999) and Geerts (2008) make a distinction between syllables in which the vowel can be reduced to [ə] on the one hand, and syllables in which the vowel can be laxed (or "shortened") on the other hand, arguing that these are two processes that occur in different syllabic contexts, where laxing feeds reduction to schwa (Geerts 2008). In an attempt to shed some more light on these issues and to find out how this works in present Dutch, A (meaning: [a] and/or [ɑ] similar vowels) is investigated in the current study in several different segmental and suprasegmental contexts (see section 3.2.2).

A final important issue concerns the gathering of data. In the literature on the topic, often no distinctions are made between different speakers. Martin (1960) has based his observations mainly on pronunciation dictionaries in addition to the production of two informants and it does not seem to be the case that other researchers (Kager 1989, Booij 1999, Geerts 2008) have based their data on a clearly described set of informants. In contrast, the current study will provide such experimental data, in order to investigate the extent of variation in pronunciation within and between speakers (question 1), and to hypothesize about the underlying form of the sounds (question 2). Observations that can be made from the resulting data may shed new light on theories previously discussed in the literature. For example, on how stressing and destressing interact (Kager 1989), what a model of speech production should look like, e.g. in what order stress assignment and reduction are applied (Geerts 2008), how words are to be divided into feet (Geerts 2008), and how loan words behave in relation to the phonological system of Dutch (Martin 1968). Finally, since the focus of the current study lies on finding out whether the observed variation is of the within-speaker and/or between-speaker type, the results may be of use for the field of forensic linguistics, because the study might reveal speaker-characteristic alternation patterns.

3 Methodology

Two experiments were carried out, namely a production experiment (section 3.2) followed by a classification experiment (section 3.3), both with the same participants (section 3.1). Before gathering the actual data, a pilot experiment was carried out with two participants in order to find out whether the goal of the production experiment was not transparent and whether the duration of the experiment was feasible for the participants. In the actual experiment, there was a short break between the production part and the classification part of approximately five minutes in which participants filled out a form about their geographical information and languages they speak. In addition, there was a questionnaire (section 3.4) about the classification experiment, carried out by the participants directly after the classification experiment. The distributions of target vowels were compared to the distributions of control vowels, which were compared between speakers, and standard deviations were investigated for finding out about within-speaker variation (section 3.5).

3.1 Participants

The participants were 20 native speakers of Dutch (10 M, 10 F). Each participant received a Dutch Boekenbon (book token) of €10,- for their participation. Since there are no clear indications in the literature as to whether and how factors such as age and region could affect the investigated alternation, it was decided to test people of various ages, who had been living in various regions of the Netherlands, without explicitly forming the sample in such a way that a comparison could be made between, e.g., two age groups or two region groups. The male participants' mean age is 37.7 years, range [20, 74] and the female participants' mean age is 35.7 years, range [20, 68]. An overview of the participants' ages and geographical information is given in Table 1.

Participant	**Gender**	**Age**	**Current residence**	**Earlier places of residence in Netherlands**
1	F	21	Amsterdam	Nijmegen
2	F	22	Amsterdam	Wieringerwerf
3	F	68	Amsterdam	Amsterdam, Utrecht, Driebergen
4	F	23	Amsterdam	Hoorn, Andijk
5	F	20	Amsterdam	Langedijk, Diemen
6	M	20	Amsterdam	Biezenmortel
7	M	23	Amsterdam	Amsterdam, Maastricht
8	F	61	Amsterdam	Bungenbrug, Wieringerwaard, Heemskerk
9	F	23	Amsterdam	Arcen
10	M	20	Amsterdam	Losser
11	M	21	Amsterdam	Leeuwarden
12	M	29	Amersfoort	Tynaarlo
13	M	46	Badhoevedorp	Amsterdam, Utrecht
14	M	60	Amsterdam	-
15	F	33	Hoofddorp	Amsterdam
16	F	48	Amsterdam	Amsterdam, Meerssen, Maastricht, Groningen
17	F	38	Haarlem	Laren, Venhuizen, Amsterdam
18	M	74	Amsterdam	-
19	M	59	Amsterdam	Arnhem, Maastricht, Mheer, Steenbergen
20	M	25	Amsterdam	Zwanenburg

Table 1: Participant overview: gender, age, places of residence.

3.2 Production experiment

3.2.1 Goal & task

The goal of the production experiment is to obtain acoustic measurements of the A-vowels in various words in order to measure the extent of variation for the different words, both within and between speakers. In order to achieve this goal, participants read a set of 800 isolated target words, control words and distractors one by one from a computer screen. Each word appeared for two seconds. Participants were told to pronounce the words. They could take a short break after every 100 words, until all 800 words had appeared. This task cost each participant a total of approximately 30 minutes. Before starting the experiment, the participants carried out an example experiment with 10 words that did not occur in the real experiment, to get used to the setting. The recordings were made in a soundproof room with a Marantz PMD660 Professional Solid State Recorder and a Samsom PM6 microphone.

3.2.2 Words

There were 160 different words, each occurring five times, leading to a total of 800 words. The set of 160 different words is shown in Table 2 and consisted of:

1) 60 target words in which alternation between /a/ and /ɑ/ was expected to be possible. These words varied in frequency (information acquired from the CELEX – Max Planck Institute for Psycholinguistics 2001), segmental context (sonorant vs. obstruent – Kager 1989), and spelling (double vs. single following consonant, because Dutch vowels are usually lax before double written consonants and tense before single written consonants). The set of target words consisted of:
    - 15 words of the pattern σˈσ
    - 15 words of the pattern ˌσσˈσ
    - 15 words of the pattern σˈσσ
    - 15 words of the pattern ˈσσˌσ
2) 20 control words (10 minimal pairs) in which alternation between /a/ and /ɑ/ was not expected to be possible, so that no doubt would exist as to which vowel was meant by the speakers (assuming the existence of a phoneme /a/ and a phoneme /ɑ/). There were five monosyllabic minimal pairs and five bisyllabic minimal pairs. These words were included to compare the pronunciation of A-vowels in the target words to.
3) 80 distractors that were randomly picked from a large list of words obtained from the CELEX (Max Planck Institute for Psycholinguistics 2001). The CELEX was searched for words with a minimum of one and a maximum of three syllables and a frequency of more than 1 in 42 million (to prevent the list from containing a large amount of very unusual words that participants might not know)[4]. This query resulted in a list of 87493 words out of which 160 words were randomly picked with the Linux *shuf*

4 Database: *Dutch Wordforms*, columns: *frequency:Inl, orthography:Word, and phonology: SylCnt*, query: *(SylCnt <= 3) && (SylCnt > 0) && (Inl > 1)*.

Targets				Controls	
σˈσ	ˌσσˈσ	σˈσσ	ˈσσˌσ	σ	ˈσσ
mai.ˈllot	ˌla.te.ˈraal	pla.ˈce.bo	ˈal.ma.ˌnak	ban	ˈa.dder
pa.ˈsseer	ˌra.di.ˈcaal	ka(s).ˈstan.je	ˈboe.ka.ˌrest	baan	ˈa.der
ka.ˈneel	ˌja.loe.ˈzie	ca.ˈsse.tte	ˈal.fa.ˌbet	as	ˈba.kken
na.ˈïef	ˌa.bri.ˈkoos	la.ˈwi.ne	ˈboe.da.ˌpest	aas	ˈba.ken
ka.ˈnaal	ˌa.me.ˈthist	ka.ˈlen.der	ˈbun.ga.ˌlow	al	ˈkwa.llen
pa(s).ˈstoor	ˌa.ller.ˈgie	ga.ˈra.ge	ˈka.zach.ˌstan	aal	ˈkwa.len
ra.ˈpport	ˌa.nar.ˈchie	pa.ˈssa.ge	ˈpa.pri.ˌka	gas	ˈga.ppen
ma.ˈnier	ˌa.de.ˈquaat	ka.ˈrak.ter	ˈma.jes.ˌteit	gaas	ˈga.pen
a.ˈnaal	ˌman.da.ˈrijn	na.ˈtuur.lijk	ˈpa.gi.ˌna	kwal	ˈba.nnen
a.ˈbuis	ˌmo.za.ˈïek	a.ˈkoes.tisch	ˈa.ni.ˌmo	kwaal	ˈba.nen
a.ˈzijn	ˌper.ma.ˈnent	a.ˈman.del	ˈa.fri.ˌka		
a.ˈpplaus	ˌsi.ga.ˈret	a.ˈppen.dix	ˈpa.na.ˌma		
a.ˈdres	ˌca.na.ˈdees	a.ˈffi.che	ˈma.ra.ˌthon		
a.ˈpril	ˌpa.ra.ˈllel	a.ˈpos.tel	ˈa.na.ˌnas		
a.ˈlleen	ˌa.ppa.ˈraat	a.ˈdop.tie	ˈfa.ra.ˌo		
Distractors					
aandoet	geest	klaarstaat	peilende	strandjutters	vlogen
aanschoot	gemetseld	knipoog	perplexe	streekplannen	voetpad
aardig	grimaste	kraaien	pikte	theetafel	volksgroepen
absentie	herfstkleur	kromming	plafond	uithoren	volslanke
afgelast	hoeden	kwellen	receptor	uitweken	voltooit
bedarend	hoezeer	leraar	ringwegen	varkentjes	wegdenken
bedrukken	ingedeeld	maatje	roosteren	vastbijtend	wegreden
berust	ingevoerd	mango	rustplaats	verbruik	wenteling
biecht	inwijden	muurtjes	schriftbeeld	verdenken	werkgroep
bokser	jaarcyclus	neerkwamen	slap	verschrikte	wijsmaken
bruidsfoto	kantelend	nok	smaakvol	versterkt	
contrasten	kelderde	onderging	spa	verstrekt	
doeltreffend	kiesdeler	oorarts	straatlantaarn	verwerk	
duiden	kinderzorg	openstaand	stralingsbron	verziende	

Table 2: Words used in the production experiment: 60 target words, 20 control words, and 80 distractors. Stress pattern and syllable division are shown for the target words and the control words. Consonants that are double in orthography may be ambisyllabic in pronunciation.

command[5]. Words that were too unusual (based on subjective assessment) were replaced by other words randomly obtained from the large list.

The words were presented in randomized order. The three alphabetically sorted lists of target words, control words, and distractor words were all randomized separately[6] and then concatenated into one file (in the mentioned order). This file was then randomized again. This was done five times so that all words would occur five times in the experiment, equally distributed over the whole experiment. This procedure was done for each participant separately, so that the order of the words had a different random distribution for each participant.

3.2.3 Segmentation & measurements

The target words and control words were segmented and annotated using Praat (Boersma & Weenink 2013). Boundaries were set at word boundaries and at relevant vowel boundaries. The spectrogram was used as the main source of information during annotation. If that information did not suffice, the waveform and the sound were taken into consideration. If the waveform showed clearer transitions than the spectrogram, the waveform was taken as main segmentation criterion. Since Dutch /a/ and /ɑ/ differ from each other both quantitatively and qualitatively, measurements were made of the first formant, the second formant, and duration. For each target vowel F1 and F2 were measured automatically using the Praat function *To Formant (burg)...* (with as *Maximum formant* 5000 Hz for men and 5500 Hz for women, as recommended in the Praat manual), followed by *Get mean...*. The duration of the vowel was measured with *Get duration...*. In order to make duration a variable on which words can be compared, the measurement taken was the duration of the vowel relative to the duration of the token of the word that vowel occurred in, normalized for number of syllables in the word:

$$(8) \qquad RelDur_{[vowel]} = \frac{Dur_{[vowel]}}{Dur_{[token]}} \cdot No.Syllables_{[token]}$$

3.3 Classification experiment

3.3.1 Goal & task

The goal of this experiment is to find out about intuitions that speakers have about the vowels /a/ and /ɑ/ in various positions. These intuitions can be compared with their actual

5 Linux command *shuf -n 160 87493words.txt > distractors.txt*
6 Linux command *shuf -n 160 distractors_sorted.txt > distractors_random.txt*

pronunciation as recorded in the production experiment. The results of this experiment may provide some indications as to how the vowels are represented underlyingly by the participants.

Participants were told that the experiment concerned /a/-like sounds. On a computer screen, a single written word containing a bold vowel appeared. Below this word there were two buttons, one containing the word <**a**der> (/ˈa̲.dər/, ‘vein’) (with bold <a>), the other containing the word <**a**dder> (/ˈɑ̲.dər/, ‘viper’) (with bold <a>). These two words form a minimal pair, only differing in the pronunciation of the bold vowel. The task was described to the participant as follows: “imagine whether you, yourself, pronounce the bold vowel of the upper word in a way that is more similar to the way you pronounce the vowel in <**a**dder> or more similar to the way you pronounce the vowel in <**a**der>. Click the word with the most similar vowel. There are no right or wrong answers.” The next word appeared after one of the two response options was clicked. The participants could take a break every 100 words until all 435 words had appeared. This task took each participant a total of approximately 15 minutes. Before starting the experiment, they did an example experiment of 10 words with as response options the minimal pair <w**e**ker> (/ˈwe̲.kər/, ‘softer’) and <w**e**kker> (/ˈwɛ̲.kər/, ‘alarm clock’).

3.3.2 Words

The words used in this experiment were the same 20 control words and 60 target words as used in the production experiment. Since <**a**der> and <**a**dder> formed the response options, they were not used as stimuli (leading to a total of 78 different words). Each word occurred five times, except for the seven words that contained two target vowels, which occurred ten times, leading to a total of 435 stimuli. Since the participants were told that the experiment concerned /a/-like sounds, no distractors were included. The words were presented to each participant in randomized order, using the <PermuteBalancedNoDoublets> strategy of the Praat ExperimentMFC program.

3.3.3 Measurements

For each word, the number of <**a**der> responses and the number of <**a**dder> responses were counted. For separate analysis of the classification experiment, one of three labels was given for each participant*word combination. The label “/a/ preference” was given when a participant had chosen the <**a**der> response four or five (out of five) times for a particular word, the label "/ɑ/ preference" was given when a participant had chosen the <**a**dder> response four or five times for a particular word. Otherwise, the label "variable response" was assigned. The results are shown in section 4.1.

3.4 Questionnaire

In order to get more insight into the meaning of the choices the participants made in the classification experiment, there was a questionnaire consisting of six questions immediately following the classification experiment. The questions concerned the origin of the choices the participants had made and what they thought about the completeness of the response options. A few of these questions are shortly discussed later on in this paper where their outcomes turned out to be relevant (in section 6). The questions were the following:

1) On what basis did you make your choices?
2) If you took into account pronunciation, did you take into account your own pronunciation or the pronunciation of others?

   (Answer options: my own pronunciation; others' pronunciations; mainly my own pronunciation, but also that of others; mainly others' pronunciations, but also my own; I did not pay attention to pronunciation.)
3) Did you also take into account the way the words were written?

   (Answer options: yes, I paid attention to …; no)
4) What did you think of the response options?
5) Was one of the two options always the right one according to you?

   (Answer options: yes; no, because …)
6) Did you miss an option for some of the words?

   (Answer options: yes, namely …; no)

3.5 Comparisons & statistical analysis

3.5.1 Production & classification: defining an /ɑ/-/a/ scale

The main goal of this study is to compare the pronunciation of [a]/[ɑ]-like vowels, as pronounced within various words, between speakers and within speakers (over several tokens of words). The group of words of interest consists of words where variable pronunciation is expected. In order to be able to draw conclusions about whether the vowels in this group of target words are more similar to [a] or more similar to [ɑ], the pronunciation of the vowels /a/

and /ɑ/ in minimal pairs (control words) is used as a reference. To compare the pronunciation of A in a particular target word between speakers, or within speakers between tokens of a word, several types of acoustic measurements should be taken into account. As was mentioned in section 3.2.3, three types of measurements were taken: F1, F2 and relative vowel duration. In this way, the extent of "/a/-similarity" or "/ɑ/-similarity" was taken to be a combination of these three factors. As it is rather complex to interpret comparisons between so many three-dimensional objects, it was decided to convert the combinations of the three measured values to a scale from 0 to 1, so that the extent of /a/-similarity or /ɑ/-similarity of a vowel token could be expressed by a single value. This conversion was carried out for each speaker separately, comparing a speaker's target words to their own control words. This form of speaker normalization makes it possible to compare a speaker's vowel realizations to another speaker's vowel realizations, by comparing at what points on the continuous /ɑ/-/a/ scale the vowels in the words are situated for the different speakers.

In order to obtain a single A-value for each token, the first step was to normalize the values of the three factors F1, F2 and RelDur. This was done for each participant separately, in the following way. For each control word, F1 was averaged over the five[7] tokens of that word. The lowest average F1 of all controls words for that speaker was set to 0 and the highest average F1 was set to 1. Based on their average F1's, all tokens of all words (control as well as targets) could be placed on this scale by:

(9) $$F1Norm_{[token]} = \frac{F1_{[token]} - minF1_{[allControlWords]}}{maxF1_{[allControlWords]} - minF1_{[allControlWords]}}$$

In this way, most tokens were located on the scale somewhere between 0 and 1, while some were allowed to be somewhat lower than 0 or higher than 1. The same procedure was followed for F2 and relative duration, so that for each token there were three normalized values: F1Norm, F2Norm, and RelDurNorm.

In order to decide on the weights that should be given to the three factors in forming the single A-values for each token, a logistic regression analysis was carried out on the combination of the normalized production values and the classification data for the control words, with dependent variables /a/ and /ɑ/ and factors F1Norm, F2Norm and RelDurNorm. This logistic regression analysis was carried out separately for the different speakers, so that

7 For some speakers there were less than five tokens for some of the words. 63 out of a total of 8265 (vowel in 87 contexts * 5 tokens * 19 participants = 8265 – one participant was excluded from analysis) tokens were excluded from analysis. Reasons for exclusion were usage of a word different from the word that was meant or usage of a non-existent word, background noise, sighing or yawning during pronunciation of the word, missing of the word by the participant, and failure to find clear segmentation boundaries. Words that were mispronounced at first but were then corrected by the participant were included.

the weighing of the three factors in the final calculation of the A-values differed between speakers. In this way, a speaker's own production was interpreted as having been of relevance for their classification choices (in accordance with the classification experiment's intention of speakers making their classification choices based on their own pronunciation). The A-values were then calculated using the coefficients (β_1, β_2 and β_3) found in the logistic regression for ln(P(/a/)/P(/ɑ/)). The calculations of the final normalized A-values were done as in (10), for each participant separately. In this calculation, the coefficients from the logistic regression are divided by the total of the three coefficients so that they are used as weights for the three factors.

(10)

$$ANorm_{[token]} = \left(\frac{\beta_1}{\beta_1+\beta_2+\beta_3}\right)\cdot F1Norm_{[token]} + \left(\frac{\beta_2}{\beta_1+\beta_2+\beta_3}\right)\cdot F2Norm_{[token]} + \left(\frac{\beta_3}{\beta_1+\beta_2+\beta_3}\right)\cdot RelDurNorm_{[token]}$$

By calculating the A-values in this way, most tokens were located on the scale somewhere between 0 and 1, while some were allowed to be somewhat lower than 0 or higher than 1. Because of this normalization procedure it is possible to make between-speaker comparisons. Since the logistic regression was based on the control words only, the /ɑ/-/a/ scale is also control word-based and the targets are thus placed on a control word scale, so that the A-ness of target vowels is decided relatively to the vowels [a] and [ɑ] as in minimal pairs.

3.5.2 Between-speaker variation: vowel distributions & ANOVA

One of the goals of this study is to find out whether there is variation in the pronunciation of A-vowels in specific words between speakers. A method of investigating this is by looking at a speaker's distribution of the vowels of the target words, as compared to their distribution of the control words, and then comparing the speakers' distributions to each other. This was done in the current study and the results are provided in section 4.2.1.

In order to make a more explicit between-speaker comparison, an ANOVA was carried out on the target words and the control words separately (see also section 4.2.1 for the results), so that the findings for the target words could be compared to the findings for the control words. The factors were Speaker and Word the dependent variable was the A-value as calculated in section 3.5.1. Effects tested for are the following. Firstly, an effect of Speaker on A-value would mean that speakers have different distributions of sounds (the distribution of a speaker's target words being based on the distribution of that speaker's control words). This would indicate a type of between-speaker variation. For the test on the control words it would mean that speakers have different distributions of sounds. For the test on the target words it would mean that speakers use different sound distributions for the target words as compared to their own range of sounds used for the control words. Secondly, an effect of Word on A-value would mean that different words have different sounds, which does not indicate a type of between speaker variation. Finally, an effect of the interaction between Speaker and Word on the A-values would mean that the relative position of words to each other differs between speakers. The ANOVA was carried out in R, with Speaker and Word as fixed variables.[8]

3.5.3 Within-speaker variation: comparison of standard deviations

Another of the goals of this study is to find out whether there is variation in the pronunciation of the A-vowels in specific words within speakers over tokens (i.e., within words). To this end, standard deviations could be investigated. Standard deviations were calculated for each word for each participant separately, on the basis of the A-values that were calculated for each token as described in section 3.5.1. The question is whether the standard deviations of the target words are higher than the standard deviations of the control words. If so, it could be said that there is more variation within target words than within control words. It may be the case that some speakers show more variation anyway than others (in their control words), which is why for each participant the target variation should be compared to their own control variation. Normalization of the targets based on the controls (as was done in calculating the A-values – see section 3.5.1) also makes it possible to compare within-speaker variation between speakers. The results are shown in section 4.2.2.

4 Analysis

In the following sections, the results of the experiments described in section 3 will be discussed and analysed. The classification experiment is taken into consideration in section 4.1, before the results of the production experiment are discussed in section 4.2, because the calculations of the A-values for the produced vowels involve classification data. In this way, the production data can also be compared to the findings for the classification experiment.

8 With the command *anova(lm(table$A ~ table$Speaker * table$Word))*

Section 4.3 discusses the question of underlying vowels in relation to the findings about the vowels on the surface.

4.1 Classification experiment

When investigating the variation within and between speakers in the classification data, two important aspects of the data can be explored. Firstly, it could be investigated for how many words the different participants have an /a/ preference, /ɑ/ preference, and variable response. The participants can be compared to each other concerning how many words they classify as /a/, how many as /ɑ/, and how many variably. This shows something about between-speaker variation (some speakers willl classify more words as /a/ than other speakers), as well as within-speaker variation (some speakers will classify some words variably). Secondly, looking at differences between words, it could be investigated what the number of participants is that has an /a/ preference, or an /ɑ/ preference for particular words. This also relates to the extent of between speaker variation, since there will be a group of words for which speakers agree on the classification, but also a group of words for which they do not.

Table 3 shows, for each participant, the percentages of target words for which there is a preference for /a/, a preference for /ɑ/, and a variable response, females in the top part of the table and males in the bottom part of the table. The table for the control words is not shown, because all participants, except for P8, have an /a/ preference for 50% of the words and an /ɑ/ preference for the other 50% of the words (P8: 22% /a/ preference, 50% /ɑ/preference, 28% variable response) and they did in the expected way, i.e., they had an /a/ preference for words expected to be classified as /a/ and an /ɑ/ preference for words expected to be classified as /ɑ/. This means that, in accordance with expectations, all of the control words were classified consistently. This is not the case for the target words. On average, the participants classified 12% of the target words in a variable manner, range [1%, 25%] (excluding P8). Since no control words were classified variably, it may be concluded that there is at least some within-speaker variation present in the way participants classify the target words.

Concerning between-speaker variation, the participants can be divided into groups based on the proportion of words they classify as /a/ or /ɑ/:

1) Some speakers classify the largest part of the words as containing the vowel /a/. This group includes P1, P2, P3, P4, P5, P9, P12, P16, P19, and perhaps P14. Percentages of words classified as /a/ range from approximately 50% to approximately 75% in this

Participant	P1	P2	P3	P4	P5	P9	P15	P16	P17	
Sex;Age	*F;21*	*F;22*	*F;68*	*F;23*	*F;20*	*F;23*	*F;33*	*F;48*	*F;38*	
% words /a/ pref.	70	66	73	51	72	55	46	61	43	
% words /ɑ/ pref.	22	25	22	36	19	30	40	31	37	
% words var. resp.	7	9	4	13	9	15	13	7	19	
Participant	**P6**	**P7**	**P10**	**P11**	**P12**	**P13**	**P14**	**P18**	**P19**	**P20**
Sex;Age	*M;20*	*M;23*	*M;20*	*M;21*	*M;29*	*M;46*	*M;60*	*M;74*	*M;59*	*M;25*
% words /a/ pref.	6	28	24	18	54	10	48	13	73	22
% words /ɑ/ pref.	93	61	66	79	21	81	28	78	10	54
% words var. resp.	1	10	10	3	25	9	24	9	16	24

Table 3: The percentages of target words for which participants have a preference for /a/, /ɑ/, and a variable response. P8 excluded.

group.

2) Some speakers classify the largest part of the words as containing the vowel /ɑ/. This group consists of P6, P7, P10, P11, P13, P18, and P20. The proportion of words classified as /ɑ/ by the speakers of this group range from approximately 55% to almost 95%.

3) Some speakers do not seem to have an overall preference for either of the two vowels in classification. P15, P17, and perhaps P14 form this group.

It is notable that the speakers in group 2 are only male speakers and that the speakers in group 1 are mainly female speakers, except for P12 and P19 (and P14). Thus, next to within-speaker variation, there appears to exist variation between speakers in classification of the target vowels.

Taking a closer look at the different words that occurred in the experiment, Table 4 shows, for each word, the number of participants that had a preference for /a/, a preference for /ɑ/, and a variable response. As can be derived from this table, there is a large group of words for which part of the speakers has an /a/ preference while the other part has an /ɑ/ preference, showing that there is variation between speakers concerning the classification of specific target words. As can also be seen in Table 4, more than 2/3 of the words (45 words) are only classified variably by zero, one or two participants. A group of 11 words is variably classified

word	#P /a/ pref.	#P /ɑ/ pref.	#P var.	#tok. /a/	#tok. /ɑ/	word	#P /a/ pref.	#P /ɑ/ pref.	#P var.	#tok. /a/	#tok. /ɑ/
ˈa̲.ni.ˌmo	19	0	0	94	1	ˌra̲.di.ˈcaal	8	3	8	55	40
ˈa̲.fri.ˌka	18	0	1	91	4	pla̲.ˈce.bo	8	7	4	53	42
ˈpa̲.gi.ˌna	18	0	1	90	5	a̲.ˈkoes.tisch	8	7	4	48	47
ˈma̲.jes.ˌteit	17	0	2	89	6	ma̲i.ˈllot	8	10	1	43	52
ˈfa̲.ra.ˌo	17	2	0	85	10	ˈa.na̲.ˌnas	8	10	1	43	52
ˈfa.ra̲.ˌo	16	1	2	83	12	a̲.ˈdop.tie	7	7	5	49	46
ˌmo.za̲.ˈïek	15	4	0	70	25	ˈal.ma̲.ˌnak	7	7	5	48	47
na̲.ˈïef	14	3	2	72	23	a̲.ˈpril	7	7	5	44	51
ˈboe.da̲.ˌpest	14	5	0	72	23	a̲.ˈpos.tel	7	9	3	45	50
ˌsi.ga̲.ˈret	13	2	4	73	22	ka̲.ˈnaal	7	9	3	44	51
ˌca.na̲.ˈdees	13	4	2	70	25	a̲.ˈdres	6	8	5	47	48
ˌper.ma̲.ˈnent	13	4	2	67	28	ˌpa̲.ra.ˈllel	6	8	5	40	55
la̲.ˈwi.ne	13	5	1	66	29	ˌca̲.na.ˈdees	6	13	0	31	64
ˈpa̲.pri.ˌka	13	5	1	65	30	ˌla̲.te.ˈraal	5	8	6	41	54
a̲.ˈzijn	13	5	1	62	33	ka̲.ˈrak.ter	5	10	4	35	60
ˈboe.ka̲.ˌrest	13	6	0	66	29	a̲.ˈman.del	5	11	3	37	58
ˈma.ra̲.ˌthon	12	5	2	66	29	ˈma̲.ra.ˌthon	4	8	7	38	57
ˌja̲.loe.ˈzie	12	5	2	59	36	ˌa̲.de.ˈquaat	4	11	4	32	63
ˈpa.na̲.ˌma	11	6	2	63	32	ˈpa̲.na.ˌma	4	13	2	26	69
ˈbun.ga̲.ˌlow	11	6	2	62	33	a̲.ˈpplaus	2	14	3	17	78
ˌman.da̲.ˈrijn	11	6	2	59	36	a̲.ˈppen.dix	2	15	2	17	78
ma̲.ˈnier	11	6	2	56	39	ˈa̲.na.ˌnas	2	15	2	14	81
ga̲.ˈra.ge	11	6	2	56	39	a̲.ˈlleen	2	17	0	8	87
ˌpa.ra̲.ˈllel	11	7	1	58	37	ca̲.ˈsse.tte	1	15	3	14	81
a̲.ˈbuis	11	7	1	57	38	ra̲.ˈpport	1	17	1	12	83
na̲.ˈtuur.lijk	11	7	1	56	39	ˌa̲.nar.ˈchie	1	17	1	8	87
ˈal.fa̲.ˌbet	10	4	5	62	33	ka̲(s).ˈstan.je	1	17	1	8	87
ka̲.ˈlen.der	10	5	4	59	36	ˌa̲.ppa.ˈraat	1	18	0	7	88
a̲.ˈnaal	10	7	2	53	42	a̲.ˈffi.che	0	14	5	11	84
ˌa̲.bri.ˈkoos	9	5	5	55	40	pa̲.ˈssa.ge	0	18	1	6	89
ka̲.ˈneel	9	8	2	52	43	ˌa̲.ller.ˈgie	0	18	1	4	91
ˌa̲.me.ˈthist	9	8	2	52	43	pa̲.ˈsseer	0	18	1	3	92
ˌa.ppa̲.ˈraat	9	8	2	52	43	pa̲(s).ˈstoor	0	19	0	1	94
ˈka̲.zach.ˌstan	9	9	1	46	49						

Table 4: The number of participants with a preference for /a/, a preference for /ɑ/, and a variable response, for each word separately (P8 excluded), and the total number of tokens classified as /a/ and /ɑ/ over all participants (except P8), for each word separately.

by five or more (>25%) of the participants. When looking at the total number of times the words were classified as /a/ and as /ɑ/, it becomes clear that some words are almost always classified as /a/. Many of these words are words where the target vowel carries primary stress, for example <ˈa.ni.ˌmo> (94 out of 95 times classified as /a/). On the other hand, there is also a group of words that are almost always classified as /ɑ/, such as <pa(s).ˈstoor> (94 out of 95 times classified as /ɑ/). These are often words in which the target vowel is in prestress position and the vowel is followed by a double orthographic consonant. Finally, there is a large group of words that is categorized as /a/ approximately half of the time and as /ɑ/ the other half of the time, such as <ˈka.zach.ˌstan> (46 times classified as /a/, 49 times as /ɑ/).

In short, in perception, some words seem to be more susceptible to within-speaker variation than others. Moreover, some words tend to be categorized as containing /a/, while others tend to be categorized as containing /ɑ/, and yet others are classified variably over participants. Thus, in contrast with the classification of the control words, there is both between-speaker variation and within-speaker variation in the classification of the target words.

4.2 Production experiment

4.2.1 Between-speaker variation

Figure 1 shows the distributions of A-values (as calculated in section 3.5.1) for four participants, for the control words and target words separately, all tokens shown separately. Not all speakers' graphs are shown here because of space issues. The graphs for all speakers (except P8, who was excluded) can be found in Appendix A. As will become clear in the following paragraphs, the graphs of the speakers represented in the figure are representative of all participants. P8 was excluded from the production analysis because of her variable responses in the classification experiment to the control stimuli (because the classification data were used in the calculation of the A-values).

A first observation that should be made from the distribution graphs is that the control word distributions are bimodal, while the target word distributions are monomodal. This suggests the existence of two clear categories for A-vowels in the control words, but the existence of only one category of A-vowels in the target words, in production. The second observation takes into account the classification data. The black bars in the graphs represent words for which the participant has a preference for /ɑ/ in classification, while the gray bars represent words for which the participant has a preference for /a/ in classification, and the

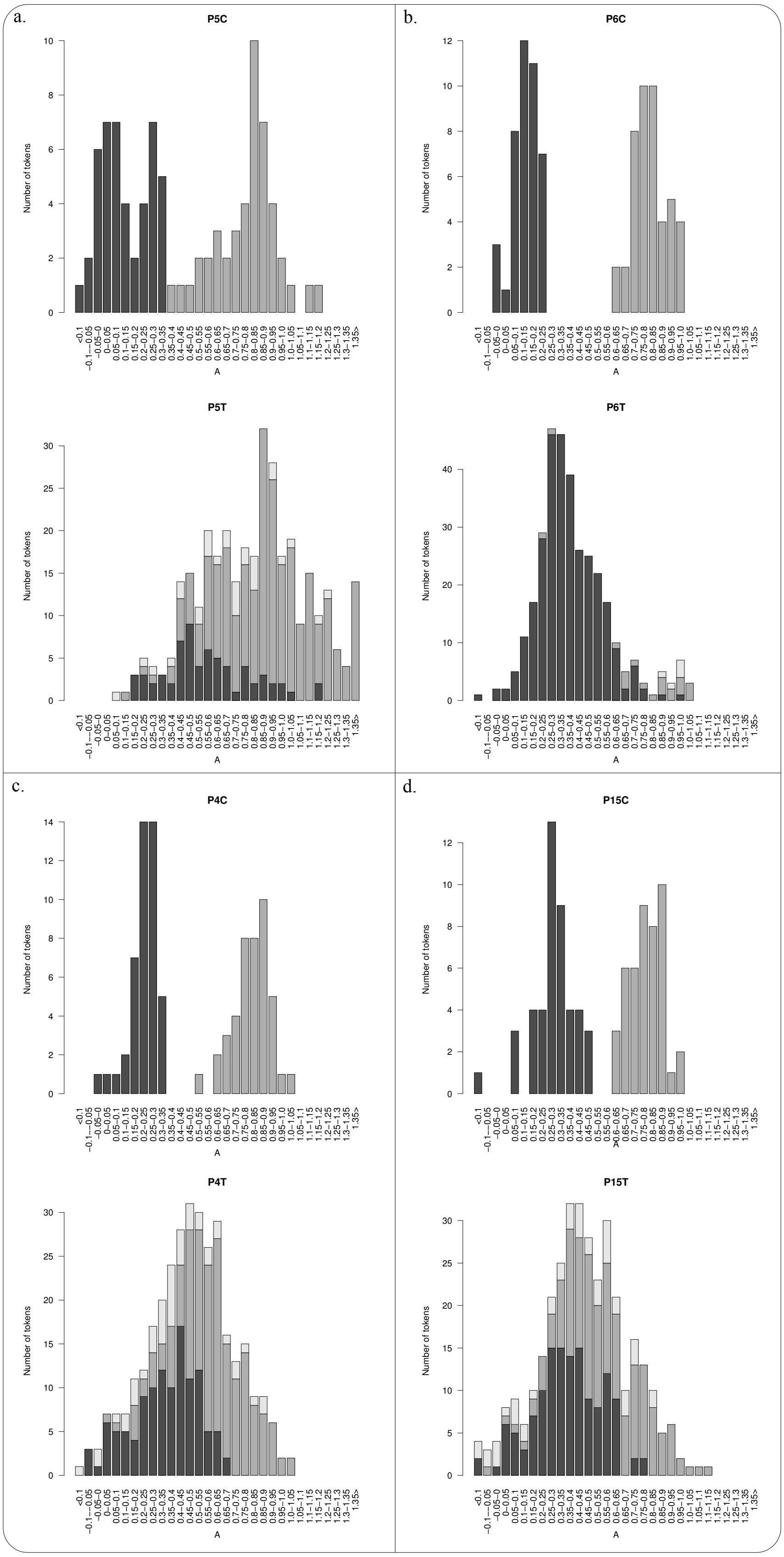


Figure 1: A-value distributions for four speakers, control words and target words separately. a) P5, b) P6, c) P4, d) P15. Black bars: tokens classified as /ɑ/, gray bars: tokens classified as /a/, white bars: tokens classified variably.

white bars represent words that the participant classified variably. When taking into account the way the participants categorize the vowels of the different words, it seems to be the case that the participants might have two categories of vowels for the target words, as well, be it categories that overlap to a large extent. This is shown by the fact that the A-value distributions of the words represented in the black bars (the bars for the words that received an /ɑ/ preference), are situated slightly more to the left end of the scale as a group (the [ɑ]-end) than the A-value distributions of the words represented in the gray bars (the bars for the words that received an /a/ preference). When looking at the white bars (the bars for the words that received variable classification), there appears to be no uniform way in which participants pronounce the words that are variably classified. Some participants pronounce such words all over the scale (such as P4 – Figure 1.c. and P5 – Figure 1.a.), whereas others pronounce them mostly in the range where there is overlap between /a/ classification and /ɑ/ classification (such as P15 – Figure 1.d.; although P15 has overlap on almost the whole scale) and yet others pronounce these vowels to the right or left of the overlapping part (such as P6 – Figure 1.b.). Summarizing the overall vowel distributions, while control vowel distributions are clearly bimodal, indicating the existence of two vowel categories, target vowel distributions are monomodal. However, classification data suggest the existence of two vowel categories in the target words that overlap to a large extent in production.

When comparing the production data with the classification data, it becomes clear that not all participants behave in the same manner. It appears to be the case that for part of the participants the classification data are consistent with their production data for the vowels of the target words, although to a lesser extent than for the control words. The group of participants whose classification data correlates with their production data consists of three types of participants. Firstly, participants P1, P2, P5, P12 and P19 have an overall preference for /a/ in classification and also have their distributions of target vowels lying slightly on the right side of the scale in production (the [a]-side) (see Figure 1.a. for P5) when compared to their placement of control words on the scale. Secondly, P6, P7 and P18 have an overall preference for /ɑ/ in classification and correspondingly have their distributions of target vowels tending slightly towards the left side of the scale in production (the [ɑ]-side) (see Figure 1.b. For P6) when compared to their own control word categories. Finally, P17 and maybe P14 do not have a clear overall preference in classification for either /a/ or /ɑ/ and, correspondingly, their production A-value distributions lie in the middle of the scale as compared to their own control words. So, for half of the participants, classification seems to correlate with production, some speakers preferring sounds more similar to [a], others preferring sounds more similar to [ɑ], and others not having an overall preference.

On the other hand, the other participants do not show such correspondence between production and classification. A group of participants, including P3, P4, P9, P10, P13, P20, and perhaps P14 have target vowel distributions in the middle of the scale when compared to their control distributions, even though they have an overall preference for either /a/ or /ɑ/ in classification (see P4 in Figure 1.c. as an example). For part of these participants (P4, P9, P20 and P14), their overal preference in classification is not as clear as for the participants for whom classification and production correlate. In other words, the difference between the number of words for which the participants in the final group have an /a/ preference and the words for which they have an /ɑ/ preference is not as large as for the participants in the first group, which could be an explanation for the absence of correspondence between classification and production on either of the sides of the scale for these participants. However, the other participants (P3, P10 and P13) have a clearer overall preference for either of the two vowels in classification, but do not show this in their overall target vowel distributions. Nevertheless, since for these participants there still seem to be two overlapping distributions of sounds when comparing classification and production, this observation does not take away the possibility of the existence of two vowel categories in the target words.

Finally, the remaining three participants behave differently from these two groups. P11 has an overall preference for /ɑ/ in classification, but his target vowel distribution is situated slightly towards the right when compared to his control word vowels. P16 shows the opposite pattern. P15 does not seem to have a preference in classification for either of the two vowels, but her target vowel distribution is situated slightly to the left when compared to her control vowels. As can be seen in Figure 1.d. for P15, the classification and production data do not conflict very extremely. Again, there seem to be two overlapping distributions of sounds, indicating the possibility of the existence of two vowel categories after all.

Summarizing, there seems to exist between-speaker variation concerning the overall distribution of sounds used for the target words. Based on the distributions of target A-values on the normalized scale, in combination with information about classification, the participants can be divided into three groups:

1) Speakers whose production and classification data correlate with each other, consisting of:

a) Speakers who produce more sounds that are more similar to /ɑ/ and also classify the majority of the words as containing /ɑ/.

b) Speakers who produce more sounds that are more similar to /a/ and also classify the majority of the words as containing /a/.

c) Speakers whose produced sounds are distributed around the middle of the scale and do not have a clear preference in classification.

2) Speakers whose produced sounds are distributed around the middle of the scale and who have a less clear preference for either /a/ or /ɑ/ in production.

3) Speakers with slightly opposite patterns in production and classification.

In an attempt to draw more solid conclusions about whether between-speaker variation truly exists here, an ANOVA was carried out on the data (as described in section 3.5.2). It was expected that the ANOVA for the control words would not show a significant effect of the Speaker*Word interaction, because there are two clear phoneme categories in the control words, that are used by all speakers. However, both the ANOVA for the control words and the ANOVA for the target words showed significant results for Speaker (controls: $F(18,1498)=21.34$, $p<0.001$; targets: $F(18,5051)=255.87$, $p<0.001$), Word (controls: $F(19,1498)=1434.79$, $p<0.001$; targets: $F(66,5051)=106.86$, $p<0.001$) and the Speaker*Word interaction (controls: $F(342,1498)=7.19$, $p<0.001$; targets: $F(1187,5051)=5.40$, $p<0.001$). The found interaction effect for the control words shows that, on an acoustic/phonetic level, the difference between words is different for different speakers. This could be explained by an effect of coarticulation, which is that one speaker coarticulates to a larger extent than another speaker. The found Word effect for the control words can be explained by the different types of stress the vowels were under or by the different segmental contexts of the different words. The found Speaker effect means that speakers have different sound distributions.

On the basis of these findings, it is impossible to draw conclusions about between-speaker variation in the target words, since between-speaker variation was also found for the control words (on a phonetic level). A method to find out more about the Speaker*Word effect, would be to filter out the influences of the factors stress and coarticulation. The influence of stress could be filtered out by testing for a Speaker*Word interaction for separate groups of words, e.g., a group of words where the vowel of interest is under primary stress, a group with secondary stress, a group where the vowel is in post-stress position, and a group where the vowel is in pre-stress position. The influence of coarticulation could be filtered out by calculating the A-values for different sets of words separately, based on, for example, the consonant following the A-vowel. Unfortunately, the consonants following the vowel of interest were not controlled for in this study to such an extent that large enough groups can be made for calculating the y-values separately, because of the relatively small amount of control words. Therefore, this ANOVA turned out not to be a sufficient method for finding out whether there is between-speaker variation for the target words in this study. Future research could focus on finding a method to make the desired comparison between target words and control words.

4.2.2 Within speaker variation

Graphs of the standard deviations per word are shown in Figure 2 for two speakers (P5 and P4). The graphs for all speakers (except P8) can be found in Appendix B. All speakers have larger standard deviations for part of the target words than for the control words (usually about 0.05-0.1 A higher). Some speakers, such as P5 (Figure 2.a.) show this for more targets words than others, such as P4 (Figure 2.b.). How can the observation that there is more variation within part of the target words than within control words be explained? The expectation was that target words would indeed show more variation than control words, because it was hypothesized that speakers do not have clearly defined phoneme categories for all target words. In the previous section, it was suggested that speakers actually do have two phoneme categories for the target words, even though they overlap to a large extent. It could be the case that for some target words the speaker does not have a clear underlying vowel defined, and that such words show more variation than others. If this is the case, the words that show most variation might be the same words as the words that are classified inconsistently by the speaker. However, a closer look at the data does not suggest such a correlation. When sorting the words per participant from highest to lowest standard deviation there does not appear to be a general tendency for variably classified words to be among the words with the highest standard deviations. But, of course, the fact that variation is possible does not mean it is necessarily present. In other words, inconsistently classified words should not necessarily show more variation in production than consistently classified words (they might vary more when more tokens are produced, for example in different speech styles), and words produced variably should not necessarily be classified inconsistently (they could vary within a category boundary).

Another reason for more variation in some target words than in control words and other target words could originate in the segmental and suprasegmental make-up of the target

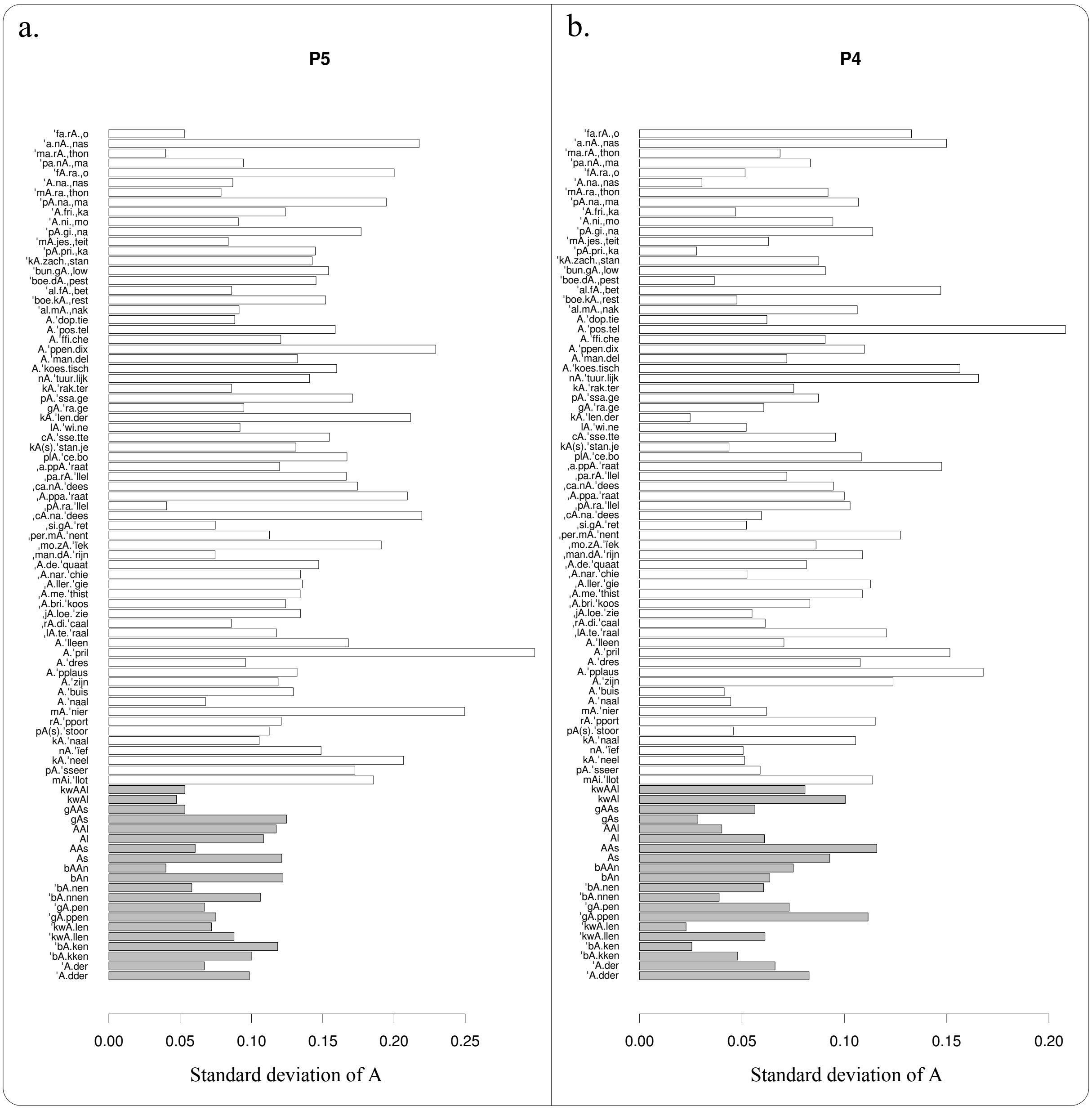


Figure 2: Standard deviations of the A-values of target words (white) and control words (gray) for two speakers. a) P5, b) P4.

words. It could be the case that some stress patterns and/or consonantal contexts allow for more variation than others. For example, stressed vowels (such as in <ˈa.ni.ˌmo>) might allow less variation than unstressed vowels (such as in <ˈboe.ka.ˌrest>), where schwa is often possible. However, when sorting the standard deviation data from low to high (per speaker), it does not seem to be the case that the target vowels with primary stress or the target vowels with secondary stress have smaller standard deviations than the target vowels that are unstressed. For now, the question remains to what extent (supra-)segmental differences between the words affect the extent of variation in production of those words. This question will be discussed more thoroughly in section 4.3.

Another explanation for the larger standard deviations in part of the target words as opposed to the control words, could be related to the fact that the targets do not contrast with other words they minimally differ from, while such a lexical contrast does exist for the control words. The argument is then that the controls do not allow for as much variation as the targets because for the control words a contrast between two words needs to be maintained, whereas this is not necessary for the target words. A potential method to check whether this could be a reason for differences in standard deviations, is to test for variation in another group of words, namely words that belong to the same group as the words in (3-4), such as <ˈɑ.ppəl> and <kʊat.ˈar.dəg>. Such words do not allow variation of the type that occurs for the target words, but they do not contrast minimally with other words either. If it were to be the case that those words group with the target words rather than the control words concerning within-speaker variation, this is evidence for the idea that the control words are more restrictive in their variation than the target words because of contrast issues. If those words group with the control words rather than the target words concerning within-speaker variation, this may indicate a special status of the target words concerning underlying vowels. Nevertheless, it should be kept in mind that, if variation within a word for a speaker is minimal, this does not mean variation is not possible. The reverse holds, as well: if variation is possible, it does not mean that it occurs. The latter could be a reason for difference in extent of variation between target words. The main point is that, next to the influence of stress, segmental context, and the presence or absence of an underlying vowel contrast, the necessity of expressing lexical contrast may be a factor in the presence of a vowel contrast on the surface. This leads to a discussion of question 2, which is provided in the next section.

4.3 Underlying and surface vowels

As a starting point for the question of which vowels are the underlying ones, the classification data needs to be taken into consideration. The largest part of the stimuli is classified consistently by the participants (88% of the words, averaged over participants). Some participants classify more words variably than others. Consistent classification can mean either of two things:

1) the participants truly represent the A-vowel in each target word as either /a/ or /ɑ/.

2) the participants remembered for most of the target words how they classified them the first time and chose the same response option the next times, leading to consistent classification. This option would mean that it is not clear whether speakers truly have the vowels in the target words represented with either of the two low vowel phonemes.

The production data show that within the targets words some realizations are similar to [a], other realizations are similar to [ɑ], and many realizations lie in between the distributions of control word [a] and [ɑ]. These data can be interpreted in several ways. One way is to conclude that, since the target word distributions are monomodal, there exists only one vowel category for the target words, namely a low vowel. The observed consistent classification is then explained by the willingness of the participant to respond in a consistent fashion and does not represent two truly existing phoneme categories. However, as was shown in section 4.2.1, for half of the participants, the overall distributions of sounds tend to be similar to [a] for the participants who often choose /a/ in the classification task and the overall distributions of sounds tend to be similar to [ɑ] for those participants who often choose /ɑ/ in the classification task, and there were not many participants for whom classification and production preferences were in conflict. Moreover, when comparing the placement on the scale of the words classified as /a/ to the placement on the scale of the words classified as /ɑ/, all participants seem to have two (largely) overlapping distributions of sounds. This implies the existence of some sort of vowel contrast not only for the control words, but also for the target words. In this case, the observed consistent classification would be a reflection of this underlyingly represented contrast on a phonological level.

While the distributions of the target vowels tend to lie on either the right part or the left part of the scale for part of the participants, at the same time many target vowel realizations are situated on the scale in between the two categories formed by the control words. This suggests the usage of sounds that lie somewhere in between typical [a] and [ɑ] (as in the control words) for this specific group of words, leading to the monomodal distributions. At first sight, this observation seems to be in conflict with the observation from the classification experiment (section 4.1) that most target words are classified consistently, which suggests that speakers have clear underlying sounds for most of the words. A potential explanation for this paradox is that underlyingly, speakers indeed have two vowel categories for these target words, but that this opposition is not shown on the surface because of the suprasegmental contexts the vowels occur in, i.e., mostly in unstressed positions. The two phoneme categories are not clearly visible on the surface (i.e., in acoustic measurements) because the unstressed positions the target vowels occur in do not allow expression of this contrast to as wide an extent as stressed positions.

The idea that unstressed positions do not allow for vowel contrast expression as much as stressed positions is in line with findings in the dissertation of Koopmans-van Beinum (1980). Her study concerns the reduction of vowel contrasts in Dutch in various conditions, varying type of speech (isolated words, read text, retold stories, and free conversation) and type of stress (lexical+sentence stress vs. no sentence stress). She introduces a measure of "acoustic system contrast", which proves to be lower for unstressed vowels in read text than for stressed vowels in read text, because the vowels are more central (within the vowel system of a speaker) in the former condition than in the latter. Moreover, unstressed vowels in retold stories and free conversation show less acoustic system contrast than stressed vowels in any of the conditions. It is not implausible that such a difference in degree of contrast also holds within isolated words for unstressed as opposed to stressed vowels, which could explain the large amount of overlap in the target words even if there are two underlying vowel categories.

Related to the phenomenon of vowel contrast reduction, Beckman (1998) discusses several languages, such as Catalan and Brazilian Portuguese, in which the vowel inventory is reduced in unstressed syllables as opposed to the vowel inventory in stressed syllables. According to Beckman, this happens by loss of certain features or segments. The vowel inventory in unstressed positions forms a subset of the vowel inventory in stressed positions, containing vowels that are less marked articulatorily or acoustically, leading either to a set of only central vowels or a set of only peripheral vowels. In the case of the Dutch /a/-/ɑ/ alternation, it seems that the vowel tends to be more central in the target words (where the vowel is often unstressed). The same might happen for the other vowel pairs (/o/ and /ɔ/, /e/ and /ɪ/, and /i/ and /ɪ/). Such a more central vowel might be said to be less marked articulatorily because it is less extreme. A potential way of analyzing the data found in the current study about Dutch in terms of feature loss is that on the language level features are lost in unstressed positions as compared to stressed positions, leading to a single low vowel category in the former, while on the individual speaker level there are still two categories for most of the words. However, since there are also target words where the vowel of interest bears secondary or primary stress, the explanation of loss of features in unstressed position cannot account for all the data. Therefore, semantic contrast must also play a role. It may be argued that the contrast in the vowel system is only maintained in cases where it is used as a means for semantic contrast. In other words, features get lost diachronically if they are not necessary for expressing meaning.

A question that arises now, is how exactly the contrast between the vowels disappears. As was shown by the results of the production experiment in the current study (section 4.2), the contrast between the vowels /a/ and /ɑ/ falls apart on the surface, on a combination of the dimensions of F1, F2, and duration, in such a way that it often leads to a vowel in between /a/ and /ɑ/ as they are in minimal pairs. At the same time, the contrast disappears in more radical way, namely on the level of the vowel system as a whole, i.e., between /a/, /ɑ/ and other vowels in the system, such as /o/ and /e/, when reduced to schwa. In contrast to Heeroma (1960), Martin (1968) and Kager (1989), who maintain that there is reduction from a long or tense vowel to a short or lax vowel that is in turn reduced to schwa (for example from /a/ to /ɑ/ to /ə/; as discussed in section 2), I argue that the two processes (taking place synchronously) are: 1) neutralization of /a/ and /ɑ/ on the language level, and 2) reduction from /a/ to /ə/ and from /ɑ/ to /ə/. How this works becomes more clear when taking into consideration Crosswhite (2001).

Crosswhite (2001) describes contrast neutralization in unstressed positions in terms of two processes, namely contrast enhancement and prominence reduction. In contrast enhancement, features are deleted (leading to schwa) or both deleted and inserted (leading to non-schwa) in unstressed positions, in order to reach a system that is maximally dispersed, so that constrasts are clear in unstressed positions. Crosswhite discusses the special case of Slovene neutralization of tense and lax vowels, where the neutralized vowels seem to lie in between the tense and lax vowels as they are when not neutralized, so that it is a case of "archiphonemic neutralization". The Dutch production data may well be described as undergoing a similar process (on the level of the language as a system), since in production many of the sounds lie in between control word /a/ and /ɑ/. In Slovene, the distinction between tense and lax vowels also disappears in short stressed syllables, showing that it is not only unstressed syllables that can undergo this type of neutralization. This accounts for the possibility of alternation between [a]-similar and [ɑ]-similar vowels under primary or secondary stress in Dutch, as well (as in <ˈpa̲.pri.ˌka> and <ˌa̲.bri.ˈkoos>). The reason for the possibility of neutralization in such words where the vowel is stressed, as opposed to the control words (where the vowel is stressed, as well), could be that in the target words a vowel contrast is not required for the expression of meaning.

Crosswhite's explanation of prominence reduction is that vowel qualities that are salient are avoided in unstressed positions, because the prominence of a position should not be in conflict with the prominence of a segment. In Dutch, such a phenomenon may be argued to occur for reduction of vowels to schwa. This also happens for other vowels than /a/ and /ɑ/, such as /e/ and /o/ (e.g. Martin 1968, Booij 1999). In this manner, the contrast between those vowels gets neutralized or reduced in unstressed positions. Thus, the reduction of contrast on the surface between /a/ and /ɑ/ and also between these low vowels and other Dutch vowels can be explained by a combination of two different processes that take place synchronously, namely contrast enhancement (by neutralization) and prominence reduction (reduction to schwa).

Finally, a problem needs to be discussed that arises if it is assumed that there are two underlying vowel categories. The problem is how the occurrence of inconsistently classified words can be explained. It cannot be the case that this inconsistency is solely caused by mistakenly clicking on the wrong button in the experiment, since otherwise it would be expected that the same pattern exists for control words, which is not the case. In that case, it would seem plausible to assume that for some words speakers do not have a clearly defined underlying vowel. As was shown in section 4.1, there are 11 target words that are variably classified by five or more participants, but there are also many target words that are only classified variably by one participant. Apparently, for some words it seems to be harder to decide on the vowel it contains than for other words. Next to this, speakers differ from each other in which exact words they classify variably, so that the absence or presence of a vowel category for a specific word may be different for different speakers.

In short, speakers can differ in the underlying vowel they have represented for specific target words. This means that on the level of the language as a whole, such a word does not have one unambiguous underlying vowel, while that is the case for the control words. It could be said that on the level of the language the underlying vowel in such a word is only specified as 'low', while, for most of the words, individual speakers represent the A-vowel more specifically, namely as either of the two phonemes /a/ and /ɑ/. This analysis can also account for the fact that the words that are classified variably differ per speaker. The underlying contrast is not expressed clearly on the surface because of contrast neutralization for contrast enhancement, and through prominence reduction, taking place not only in unstressed positions, but also in some words where the vowel is in a stressed position, where no lexical contrast needs to exist.

5 Conclusions

This paper focused on the variation occurring in perception and production in a specific group of words containing vowels on the Dutch /ɑ/-/a/ continuum. The first question that was posed is whether this variation occurs between speakers and whether it occurs within speakers. When taking into account both production and classification data, it seems that speakers have two vowel categories for (most of) the target words, because they classify many of the words consistently, because there seem to exist two overlapping distributions of sounds in production, and because for part of the speakers their production A-values seem to agree with their overall preference in classification for either /a/ or /ɑ/. However, the categorical distinction was argued to be less clearly visible on the surface for the target words than for the control words because of the type of stress the target vowels undergo (the vowels often occur in unstressed positions) and because the contrast is not necessary for distinguishing between the meaning of such words.

Concerning between-speaker variation, compared to their own produced control vowels, some speakers tend to use sounds similar to [a] for the target words, another group of speakers tends to use sounds similar to [ɑ] for those words and yet others pronounce most sounds in between [a] and [ɑ]. It was observed in section 4.1 that all the speakers who classify the majority of the target words as /ɑ/ are males and almost all of the speakers who classify the majority of the target words as /a/ are females, which is remarkable. As for the question of whether one speaker uses a different vowel for a specific word than another speaker, this seems to be true when looking at classification data. There is a considerably large proportion of words for which the opinion is variable as to which of the two low vowels is used in those words. It may be concluded that concerning their vowel distributions and their ideas about underlying vowels, there is variation between speakers. Thus, between-speaker variation clearly exists in classification and also in production, be it less clearly.

Concerning within-speaker variation, classification data show that some speakers are not sure about a small part of the target vowels' categories in their own speech. In production, speakers have larger standard deviations for part of the target words than for the control words, suggesting that there exists within-speaker variation for such words to some extent. Potential explanations for this observation are that 1) the sound category for such words is not defined, 2) that the (supra-)segmental make-up of those target words allows for more variation than for the control words and other target words and 3) that the target words allow for more variation because expression of contrast is not relevant for the meaning of such words. These possibilities do not rule each other out but may supplement each other. It needs to be kept in mind here that the possibility of variation does not necessarily mean the occurrence of variation, so that in reality there might exist more variation than the variation that was measured in the experiment.

The second question that was posed in this paper is how the sounds of the investigated words are represented in the brain, i.e. whether they are represented as two separate phonemes or as a single low vowel. It was argued that for most of the words speakers have either an underlying /a/ or an underlying /ɑ/, but for some of the words they do not have a clear underlying phoneme. For many words, there is no clear preference for either of the categories on the language level. In production, the underlying contrast is less visible on the surface for the target words than for the control words, because of the (supra-)segmental context the vowels occur in, in which the contrast is neutralized, and because a contrast is not necessary in such positions.

6 Discussion and further research

This section discusses earlier research on the alternation of [a] and [ɑ] and relates that research to the findings of the current study. Furthermore, suggestions for improvement of this study and for further research are provided. A first point of interest, is that it seems that Kager (1989) and Heeroma (1960) assume it to be straightforward which words have an underlying /a/ and which words have an underlying /ɑ/. Kager (1989:306) provides a list of words of which he assumes that the vowels are underlyingly short. These are all words in which the vowel is followed by a double orthographic consonant (e.g., <ca.ˈsse.tte>). Heeroma (1960:191) seems to take this position, as well, by stating that in the name "Matthijs" the vowel has a stronger variety [a] and a reduced variety [ə], while the normal variety is [ɑ]. Although the classification results of the current study (Table 4) indeed show that many of the target words in which the vowel is followed by a double consonant are often categorized as /ɑ/, not all speakers do this. For example, the vowel in the word <a.ˈppen.dix> was consistently classified as /a/ by two of the participants and this word was categorized as /a/ 17 out of 95 times in total. At the same time, words in which vowels are followed by a single consonant are not always classified as /a/, but also as /ɑ/. An example is the word < ˌa.nar.ˈchie>, the vowel of which is classified consistently as /ɑ/ by 17 of the participants, and 87 out of 95 times in total. Thus, the definition of an underlying vowel based on orthography only does not appear to be a valid one in this case. Of course, the participants were asked in the classification task how they *pronounce* the vowels themselves. There might be better methods to find out about their intuitions about the underlying sounds in these words, since the question of pronunciation might lead to answers in terms of sounds on the surface. For example, the participants could have been asked how the vowels *should* be pronounced, in order to find out more about their ideas of a norm, which may be closer to a real underlying form.

Concerning actual production of the words, the question is how different segmental contexts affect the surface forms of the vowels. In order to find out in what way speakers differ from each other in the pronunciation of the vowels in specific words, a comparison could be made between speakers concerning the order of the words on the /ɑ/-/a/ scale. It would be interesting to find out whether the location on the scale of a specific word in relation to the location on the scale of the other words is different for different speakers and whether that location correlates with the way speakers classify the vowel in that word. To this end, further research could try to make a comparison between the speakers of the orders of the words on the scale. This could provide more clarity on the manner in which different (supra-)segmental contexts affect the vowel.

Another interesting point is that in the literature on the Dutch vowel system, discussions have been going on about how the vowels can be divided into groups of long, short, and sometimes half-long vowels (e.g., Nooteboom 1972, Koopmans-van Beinum 1980, Rietveld et al. 2004), where /a/ and other non-high tense monophthongs seem to behave differently from high tense monophthongs (/i/, /y/, /u/). One of the findings of Rietveld et al. (2004) is that the duration of Dutch vowels in syllables with primary stress are longer than vowels in syllables with secondary stress, which are in turn longer than unstressed vowels. Gussenhoven (2014) maintains that, although in earlier literature /a/ in penultimate syllables was analyzed as a long vowel [a:], it is actually a short vowel [a], which has to do with the type of stress the vowel undergoes. According to Gussenhoven, this vowel is only long when stressed or when occurring in a closed syllable. A point that is shown by the current study, is that it is not only the quantity of the vowel that is different in different types of syllables, but that the quality may also vary in different types of positions, and also *within* positions. As was shown in section 4.2.1, the vowels in the target words may be more similar to [a] in some cases and more similar to [ɑ] in other cases, often lying in between [a] and [ɑ] on the A-scale in production. In the questionnaire carried out in the current study (section 3.4), there were three participants who said that in some cases they thought the sound was one in between /a/

and /ɑ/. The sounds that are used also appeared to be dependent on the speaker, and for many words there is no clear indication as to what the underlying vowel would be on the language level, since the underlying vowel seems to differ between speakers. In positions where the vowel is unstressed, or where a semantic contrast need not be expressed, there appears to exist neutralization of tense /a/ and lax /ɑ/. This is not only a matter of quantity, but also one of quality. Of course, the data gathered in the current study could be used for investigating the issues of vowel duration (on a phonetic level) and length (on a phonological level) in Dutch more thoroughly.

Concerning the relation between the two processes of alternation between [a] and [ɑ] on the one hand and reduction to schwa on the other hand, I disagree with Kager (1989), Heeroma (1960) and Martin (1968) that there is by definition reduction from tense to lax (or "long" to "short") vowels. These authors argue that [a] reduces to [ɑ], which in some cases can then reduce to schwa. I do not agree with the idea that [ɑ]-like sounds in words such as the target words used in the current study are reduced forms of underlying /a/'s. The analysis provided in the current study is that some of such words have an underlying /a/ and others have an underlying /ɑ/ and that this contrast is not clearly visible on the surface because sounds in between [a] and [ɑ] are used, which is explained by the type of stress the vowels undergo and by the lack of necessity for semantic contrast expression. So, the underlying /a/'s and underlying /ɑ/'s surface in various ways as both [a]-similar and [ɑ]-similar sounds. Contrast between the vowels of Dutch is in this way enhanced in particular positions. At the same time, these underlying /a/'s and /ɑ/'s can in some contexts surface as schwa, with prominence reduction as a reason. So it is not the case that [ɑ] is a reduced variety of [a], but that the surface vowels are less extreme than either [a] or [ɑ]. The variation is not caused by a series of reduction steps from [a] to [ɑ] to [ə], but rather by a co-occurrence of two processes, namely one in which /a/ and /ɑ/ are neutralized in specific contexts and one in which /a/ and /ɑ/ are reduced to schwa. To be clear, this is not to deny the possibility of the existence of a series of reduction steps, but rather to oppose the view that [ɑ] is a reduced variety of [a]. The possibility remains that schwa is not a direct reduced form of [a] or [ɑ], but that there indeed is a series of reduction steps, namely first from [a] or [ɑ] to a vowel in between [a] and [ɑ] and then from such a more central vowel to schwa.

It may well be the case that phenomena similar to the ones described in the current study also occur for the other vowel pairs mentioned in the literature (/o/-/ɔ/, /e/-/ɪ/, and /i/-/ɪ/), although the relation between the members of these pairs is not exactly the same as the relation between /a/ and /ɑ/, in terms of how they differ from each other on each dimension. Future research could focus on one or more of these other vowel pairs, in order to get a better idea of how the Dutch vowel system as a whole works and may shed more light on whether speakers really have two clear categories for words in which such vowels alternate.

Concerning the design of the experiments, several aspects could be improved and extended. Firstly, future research may use more target words and control better for stress and coarticulation, so that a simpler model can be constructed of the production data, which can then be used for statistical tests that can compare the target words to the control words. In order to make the picture clearer about the precise role of stress, words where no vowel alternation seems to exist, but where no lexical contrast exists either, could also be investigated. Secondly, further research could try to avoid the potential unwanted effects of the orthographic stimuli used in the classification experiment, because the occurrence of double consonants in orthography may have affected the participants' choices in the classification task. In the questionnaire about the classification experiment (section 3.4), there were five participants who indicated that they had used double consonants in orthography in making their response choices. A task in which orthography is absent can avoid this potential effect, but it has to be kept in mind that the use of sounds instead of written stimuli may interfere with the speakers ideas of their own pronunciation and representations, as well. Finally, a method of investigating underlying sounds would be to ask participants to pronounce the words in a formal way. It would be interesting to see whether there would be a clearer expression of a vowel contrast in a more formal speech style, since that may indicate the existence of two underlying vowel categories.

The current study has taken a phenomenon discussed in the literature and approached it from a different point of view. Instead of focusing on describing the contexts in which /ɑ/-/a/ vowels can alternate, a start was made at finding the origin of this variation in terms of individual speakers. This was done by gathering production and perception data from 20 native speakers, which had not been done before for this topic.

References

Adank, P., R. van Hout & R. Smits (2004). "An Acoustic Description of the Vowels of Northern and Southern Standard Dutch." In: *Journal of the Acoustical Society of America, **116***, 1729-1738.

Beckman, Jill (1998). *Positional Faithfulness.* University of Massachusetts: doctoral dissertation.

Boersma, Paul & David Weenink (2013). Praat: doing phonetics by computer [Computer program]. Version 5.3.56, retrieved 15 September 2013 from http://www.praat.org.

Booij, Geert (1999). *The Phonology of Dutch.* Oxford: Oxford University Press.

Botma, Bert & Roland Noske (Eds.) (2012). *Phonological Explorations. Empirical, Theoretical and Diachronic Issues.* Berlin/Boston: De Gruyter.

Botma, Bert & Marc van Oostendorp. "A propos of the Dutch vowel system 21 years on, 22 years on." From: Botma & Noske (Eds.) (2012), 135-154.

Crosswhite (2001). *Vowel Reduction in Optimality Theory.* New York/London: Routledge.

Geerts, Antonius Cornelis Johannes (2008). *More about Less. Fast Speech Phonology: the Case of French and Dutch.* Radboud Universiteit Nijmegen: doctoral dissertation.

Gussenhoven, Carlos (2014). "Possible and impossible exceptions in Dutch word stress." From: Van der Hulst (ed.) (2014), 276-296.

Heeroma, K. (1960). "De *ie* als plus-foneem van de reductievocaal." In: *Tijdschrift voor Nederlandse Taal- en Letterkunde, 77*, 187-202.

Kager, René W.J. (1989). *A Metrical Theory of Stress and Destressing in English and Dutch.* Universiteit Utrecht: doctoral dissertation.

Koopmans-van Beinum, F. J. (1980). *Vowel Contrast Reduction. An Acoustic and Perceptual Study of Dutch Vowels in Various Speech Conditions.* University of Amsterdam: Doctoral Dissertation.

Martin, W. (1968). "De verdoffing van gedekte en ongedekte *e* in niet-hoofdtonige positie bij Romaanse leenwoorden in het Nederlands." In: *De Nieuwe Taalgids, **61**,* 162-181.

Max Planck Institute for Psycholinguistics (2001). WebCelex. Available at http://celex.mpi.nl/. Visited 20 January 2014.

Nooteboom, S.G. (1972). *Production and perception of vowel duration: A study of durational properties of vowels in Dutch.* University of Utrecht: doctoral dissertation.

Pols, L.C.W., H.R.C. Tromp & R. Plomp (1973). "Frequency Analysis of Dutch Vowels from 50 Male Speakers." In: *Journal of the Acoustical Society of America, **53***, 1093-1101.

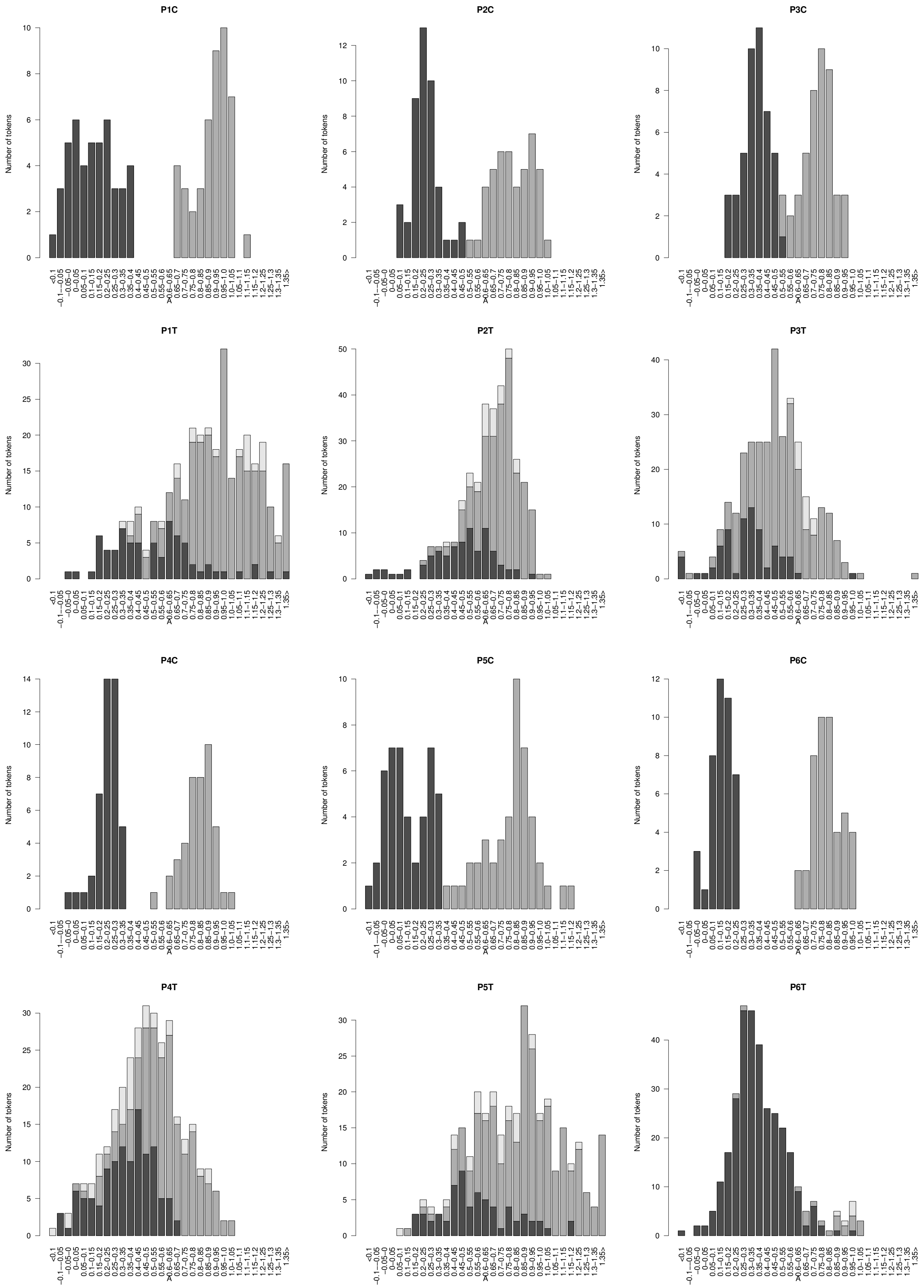
Rietveld, T., J. Kerkhoff & C. Gussenhoven (2004). "Word prosodic structure and vowel duration in Dutch." In: *Journal of Phonetics, **32***, 349-371.

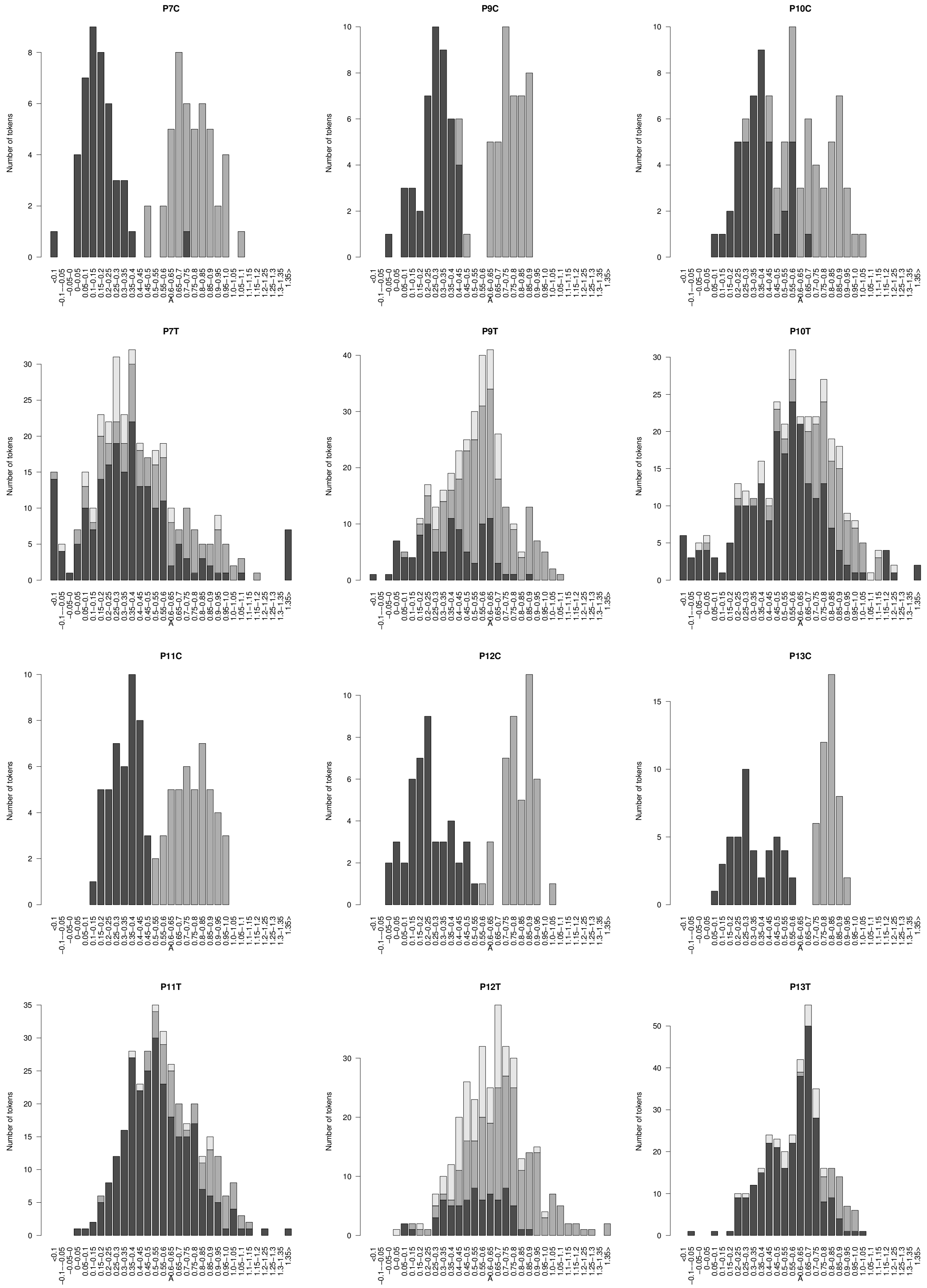
Scharpé, L. (1912). *Nederlandsche Uitspraakleer.*

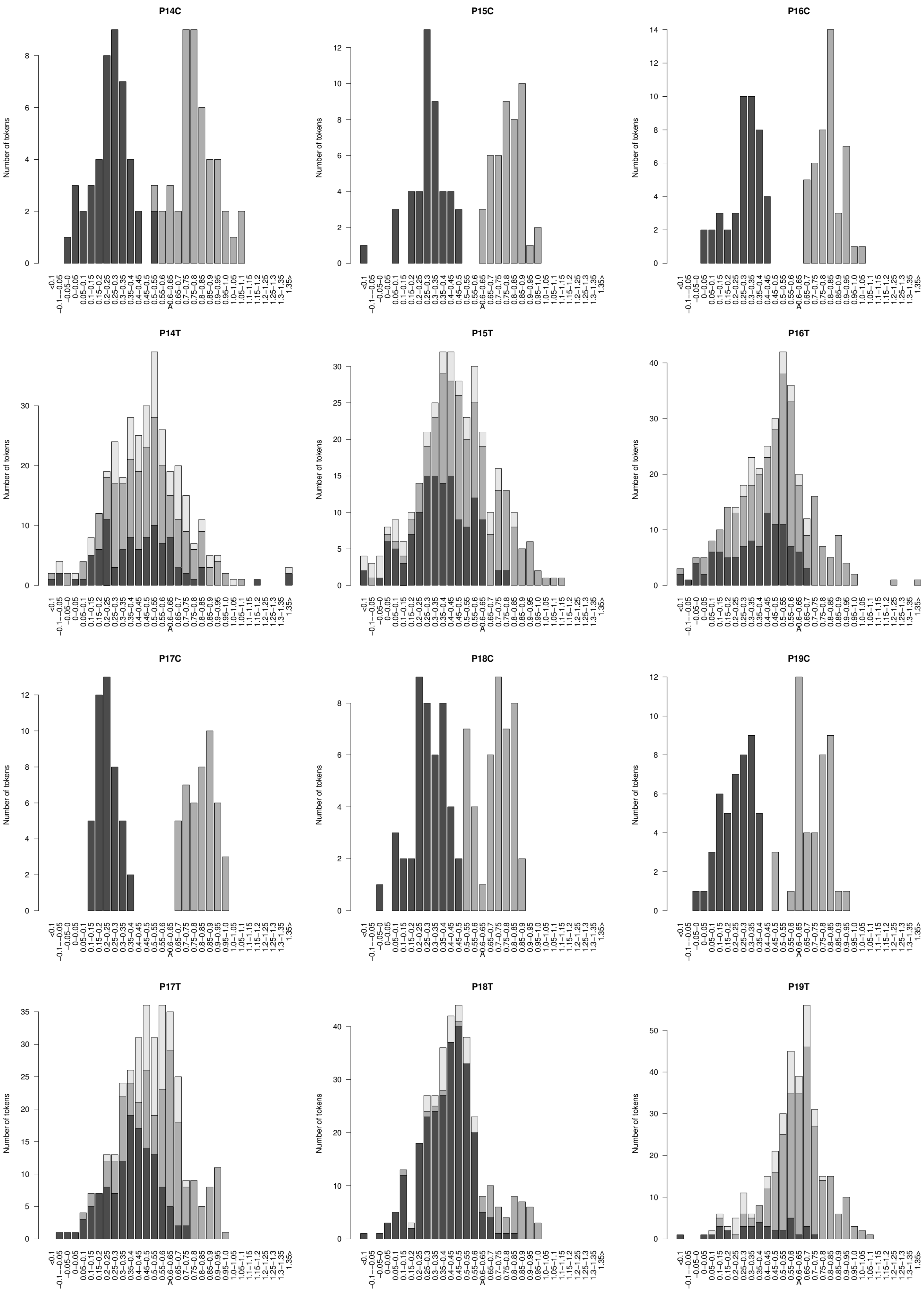
Van der Hulst (ed.) (2014). *Word Stress: Theoretical and Typological Issues.* Cambridge: Cambridge University Press.

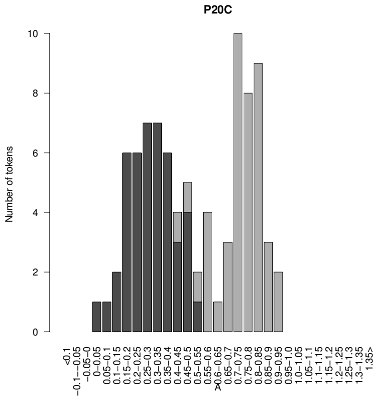
Van Leussen, Jan-Willem, Daniel Williams & Paola Escudero (2011). "Acoustic Properties of Dutch Steady-State Vowels: Contextual Effects and a Comparison with Previous Studies." In: *Proceedings of the International Congress of Phonetic Sciences XVII*, 1194-1197

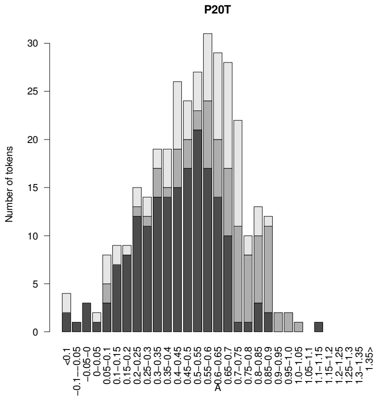
Appendix A: A-value distribution graphs for all participants (excluding P8)



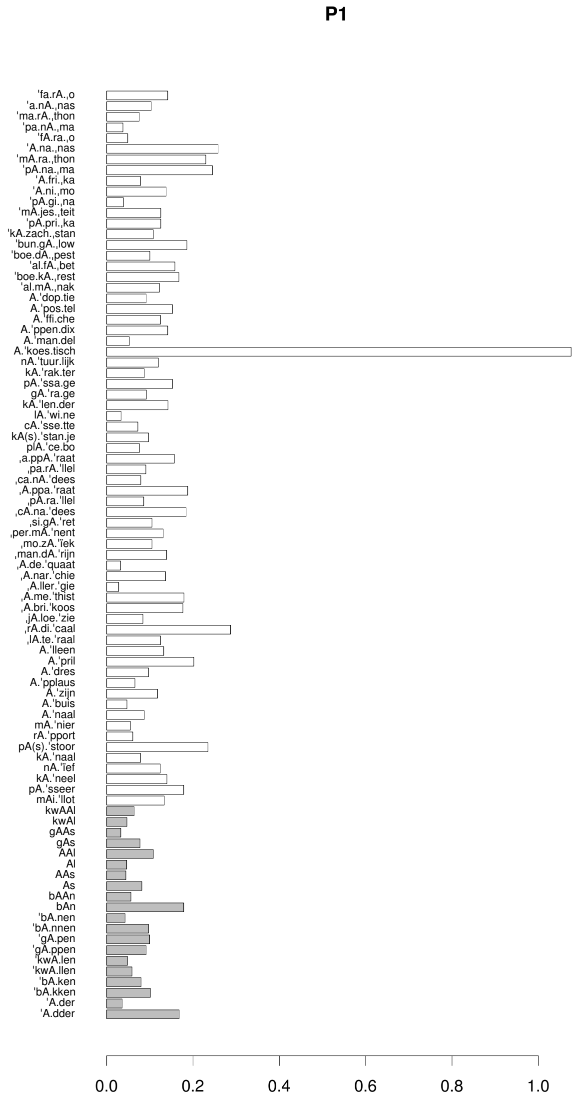

P7C
P9C
P10C
P7T
P9T
P10T
P11C
P12C
P13C
P11T
P12T
P13T
Number of tokens

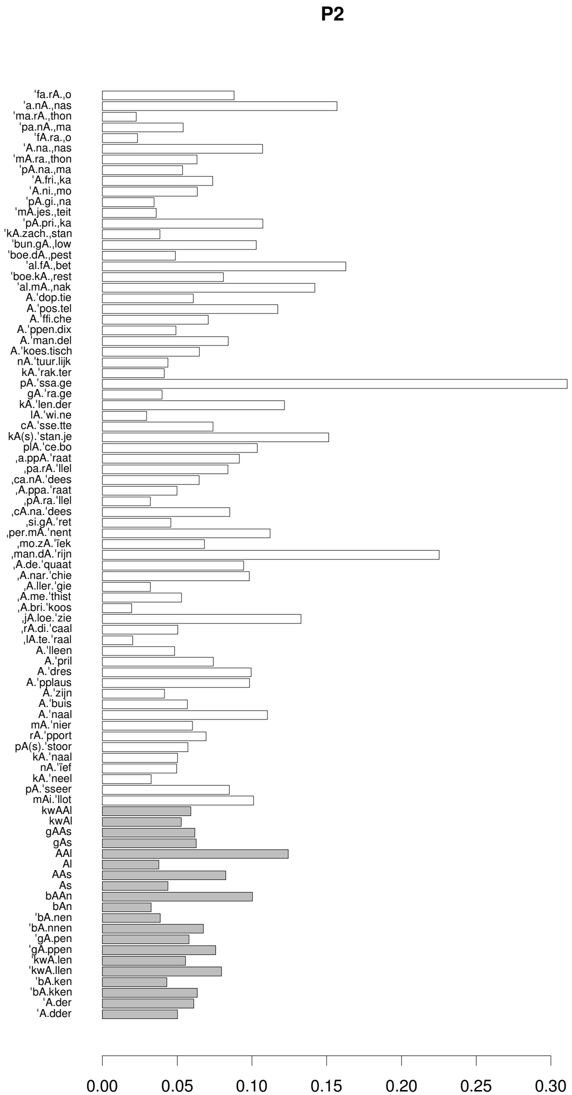

P14C
P15C
P16C
P14T
P15T
P16T
P17C
P18C
P19C
P17T
P18T
P19T
Number of tokens

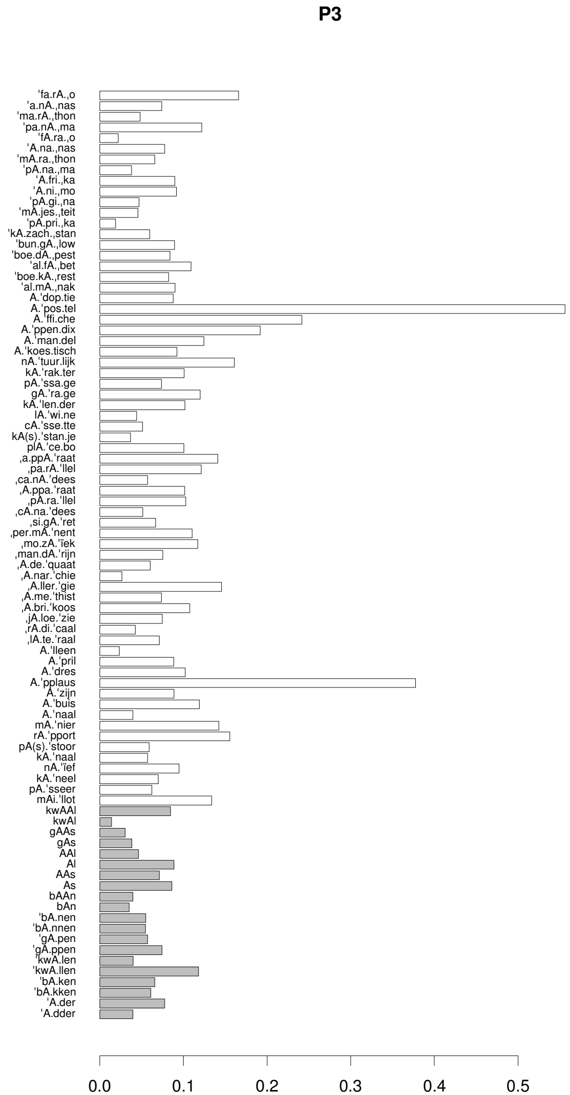

P20C
Number of tokens

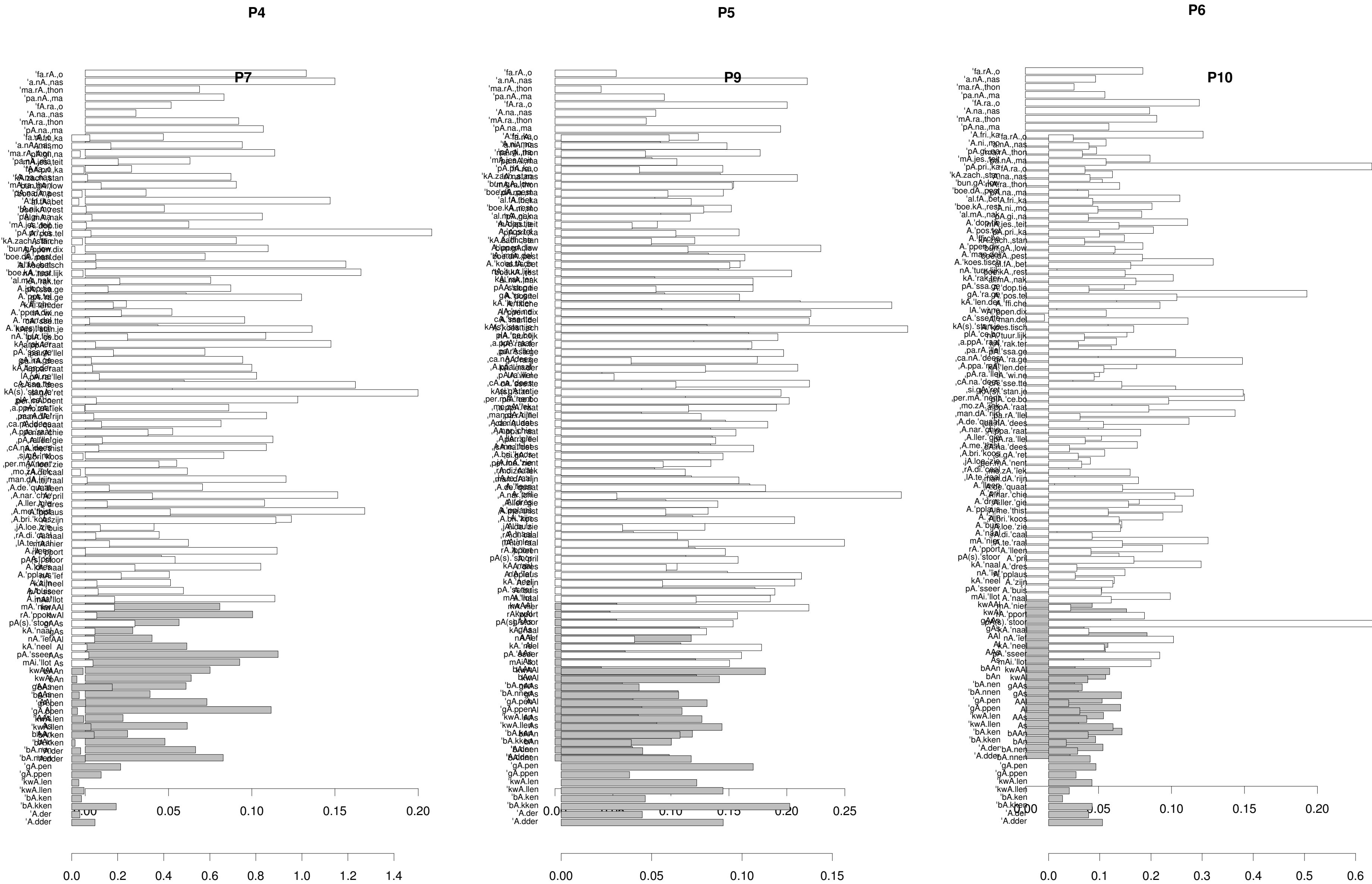

P20T
Number of tokens

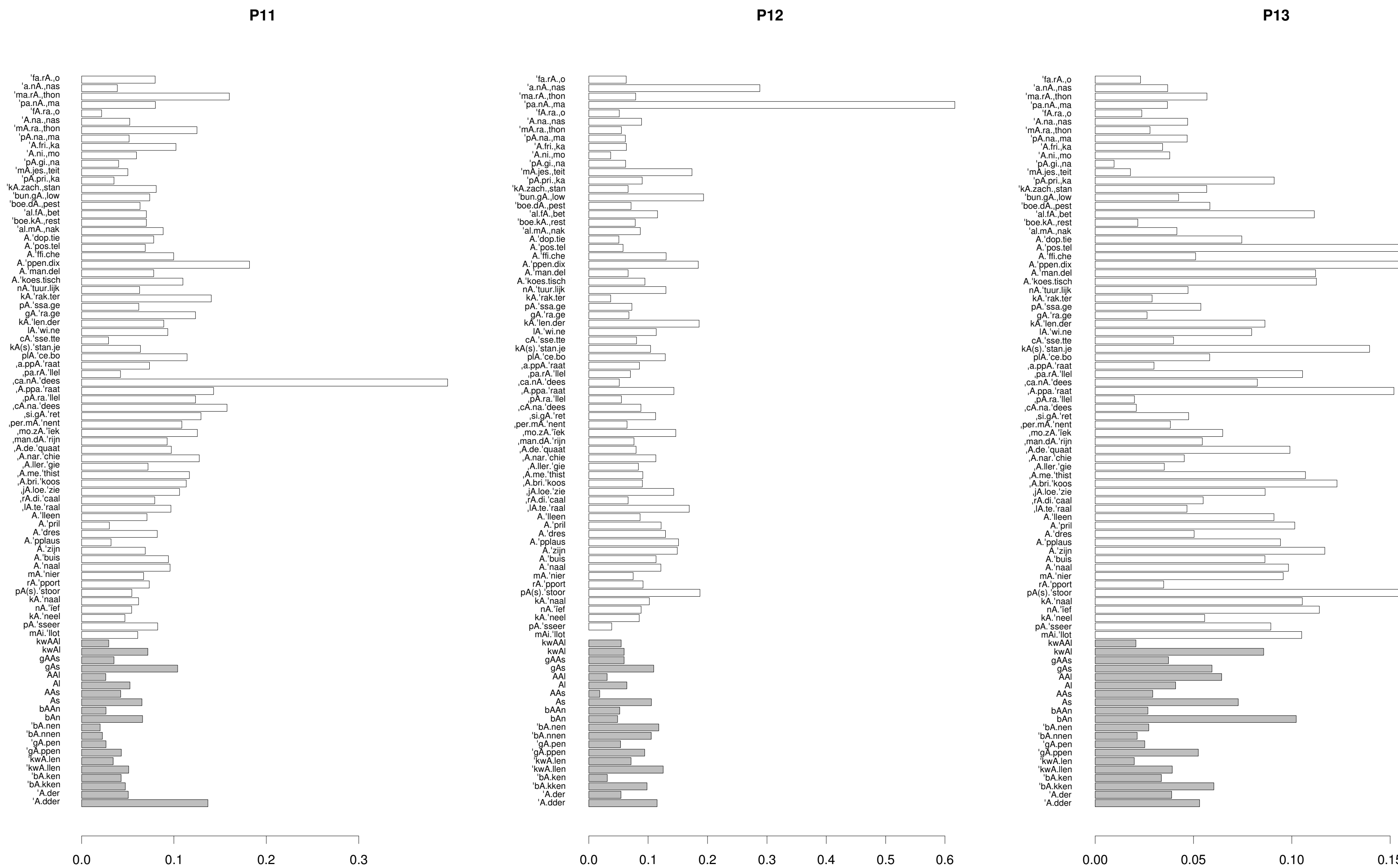
Appendix B: Standard deviation graphs for all participants (excluding P8)



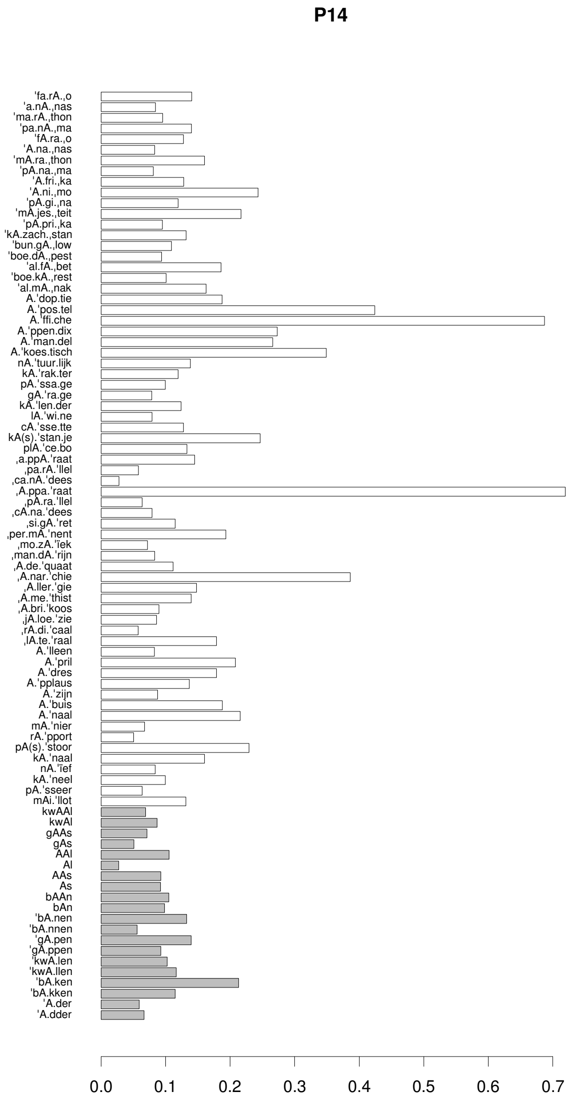


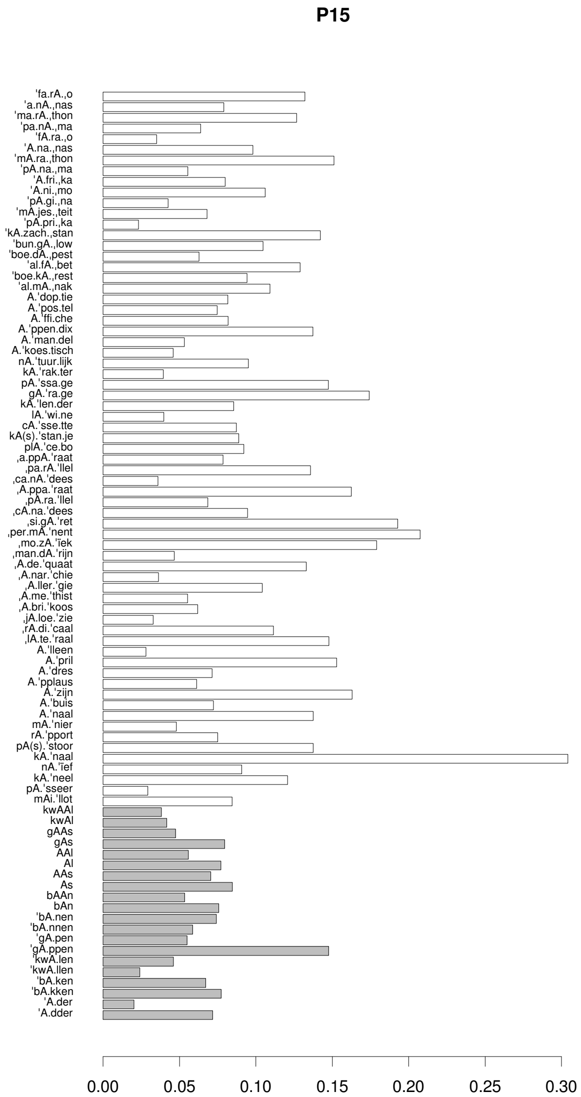


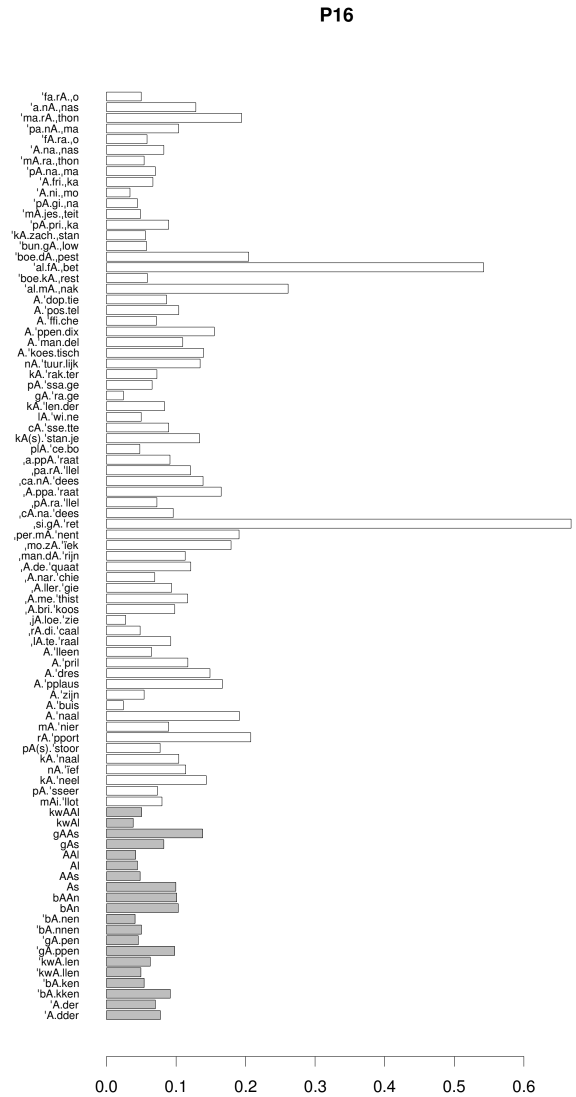

P4
P5
P6
P7
P9
P10

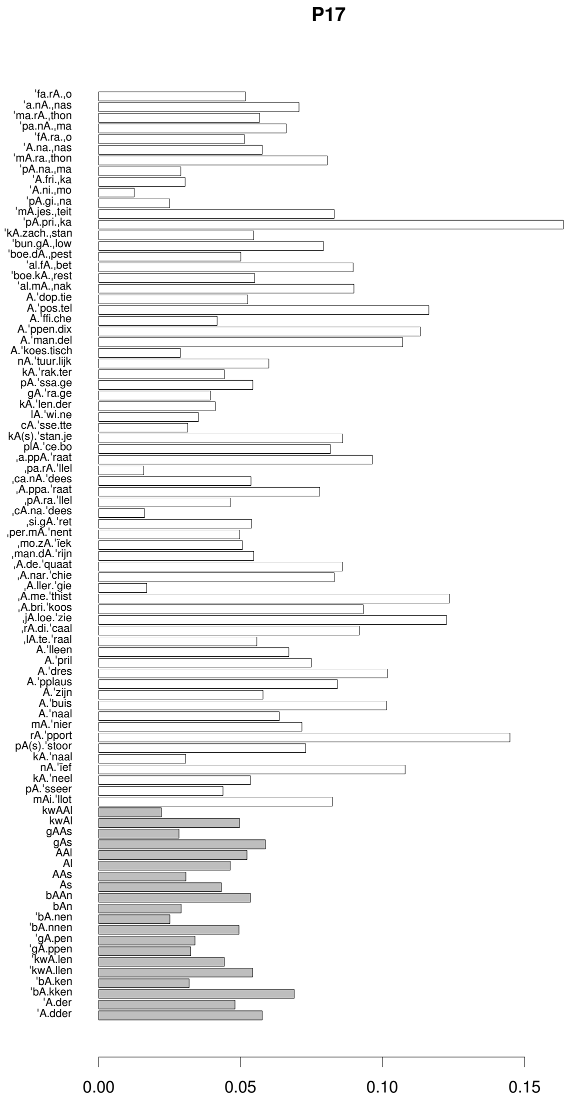


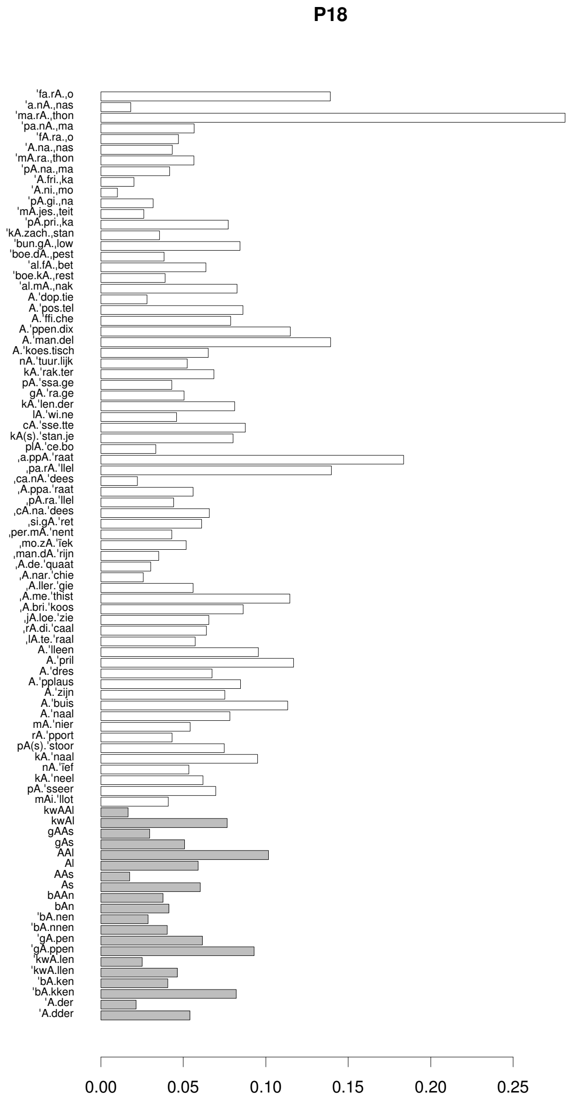
P11
P12
P13
'fa.rA.,o
'a.nA.,nas
'ma.rA.,thon
'pa.nA.,ma
'fA.ra.,o
'A.na.,nas
'mA.ra.,thon
'pA.na.,ma
'A.fri.,ka
'A.ni.,mo
'pA.gi.,na
'mA.jes.,teit
'pA.pri.,ka
'kA.zach.,stan
'bun.gA.,low
'boe.dA.,pest
'al.fA.,bet
'boe.kA.,rest
'al.mA.,nak
A.'dop.tie
A.'pos.tel
A.'ffi.che
A.'ppen.dix
A.'man.del
A.'koes.tisch
nA.'tuur.lijk
kA.'rak.ter
pA.'ssa.ge
gA.'ra.ge
kA.'len.der
lA.'wi.ne
cA.'sse.tte
kA(s).'stan.je
plA.'ce.bo
,a.ppA.'raat
,pa.rA.'llel
,ca.nA.'dees
,A.ppa.'raat
,pA.ra.'llel
,cA.na.'dees
,si.gA.'ret
,per.mA.'nent
,mo.zA.'ïek
,man.dA.'rijn
,A.de.'quaat
,A.nar.'chie
,A.ller.'gie
,A.me.'thist
,A.bri.'koos
,jA.loe.'zie
,rA.di.'caal
,lA.te.'raal
A.'lleen
A.'pril
A.'dres
A.'pplaus
A.'zijn
A.'buis
A.'naal
mA.'nier
rA.'pport
pA(s).'stoor
kA.'naal
nA.'ïef
kA.'neel
pA.'sseer
mAi.'llot
kwAAl
kwAl
gAAs
gAs
AAl
Al
AAs
As
bAAn
bAn
'bA.nen
'bA.nnen
'gA.pen
'gA.ppen
'kwA.len
'kwA.llen
'bA.ken
'bA.kken
'A.der
'A.dder
0.0 0.1 0.2 0.3
0.0 0.1 0.2 0.3 0.4 0.5 0.6
0.00 0.05 0.10 0.15

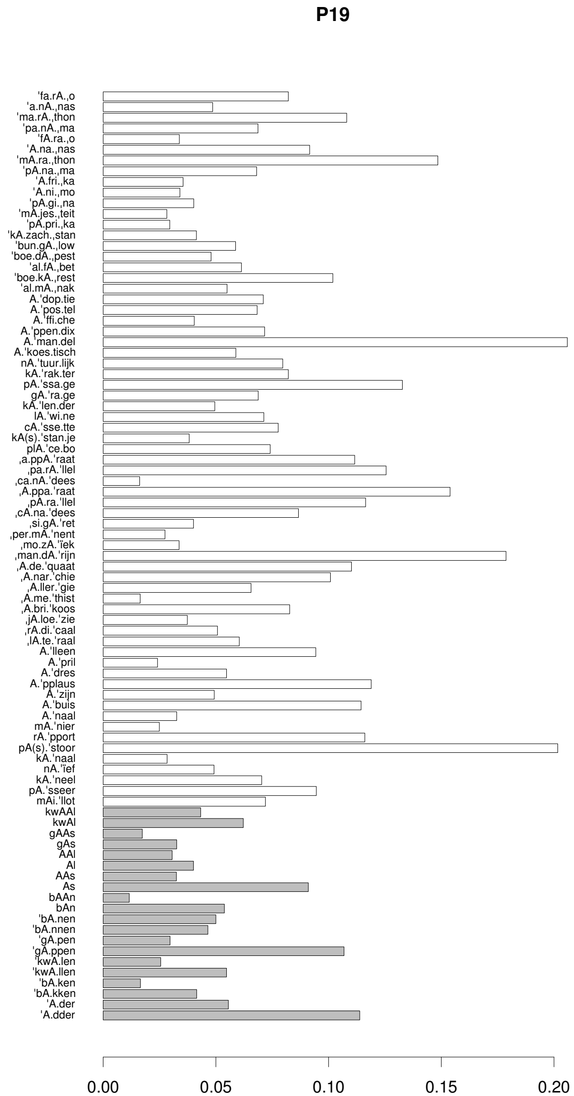

P14
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7

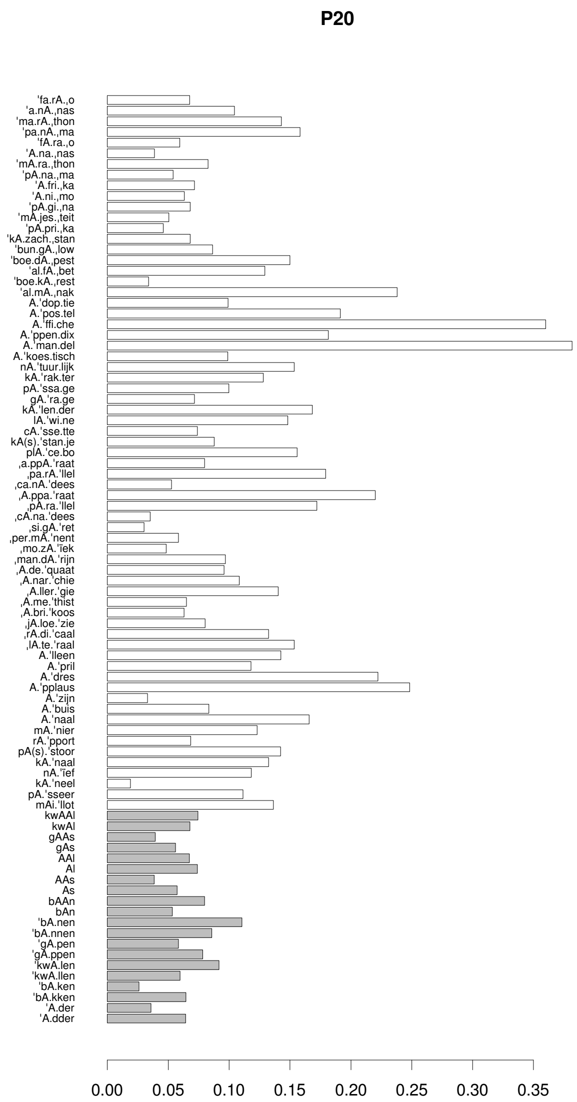

P15
0.00
0.05
0.10
0.15
0.20
0.25
0.30


P16
0.0
0.1
0.2
0.3
0.4
0.5
0.6


P17
'fa.rA.,o
'a.nA.,nas
'ma.rA.,thon
'pa.nA.,ma
'fA.ra.,o
'A.na.,nas
'mA.ra.,thon
'pA.na.,ma
'A.fri.,ka
'A.ni.,mo
'pA.gi.,na
'mA.jes.,teit
'pA.pri.,ka
'kA.zach.,stan
'bun.gA.,low
'boe.dA.,pest
'al.fA.,bet
'boe.kA.,rest
'al.mA.,nak
A.'dop.tie
A.'pos.tel
A.'ffi.che
A.'ppen.dix
A.'man.del
A.'koes.tisch
nA.'tuur.lijk
kA.'rak.ter
pA.'ssa.ge
gA.'ra.ge
kA.'len.der
lA.'wi.ne
cA.'sse.tte
kA(s).'stan.je
plA.'ce.bo
,a.ppA.'raat
,pa.rA.'llel
,ca.nA.'dees
,A.ppa.'raat
,pA.ra.'llel
,cA.na.'dees
,si.gA.'ret
,per.mA.'nent
,mo.zA.'iek
,man.dA.'rijn
,A.de.'quaat
,A.nar.'chie
,A.ller.'gie
,A.me.'thist
,A.bri.'koos
,jA.loe.'zie
,rA.di.'caal
,lA.te.'raal
A.'lleen
A.'pril
A.'dres
A.'pplaus
A.'zijn
A.'buis
A.'naal
mA.'nier
rA.'pport
pA(s).'stoor
kA.'naal
nA.'ief
kA.'neel
pA.'sseer
mAi.'llot
kwAAl
kwAl
gAAs
gAs
AAl
Al
AAs
As
bAAn
bAn
'bA.nen
'bA.nnen
'gA.pen
'gA.ppen
'kwA.len
'kwA.llen
'bA.ken
'bA.kken
'A.der
'A.dder
0.00
0.05
0.10
0.15


P18
'fa.rA.,o
'a.nA.,nas
'ma.rA.,thon
'pa.nA.,ma
'fA.ra.,o
'A.na.,nas
'mA.ra.,thon
'pA.na.,ma
'A.fri.,ka
'A.ni.,mo
'pA.gi.,na
'mA.jes.,teit
'pA.pri.,ka
'kA.zach.,stan
'bun.gA.,low
'boe.dA.,pest
'al.fA.,bet
'boe.kA.,rest
'al.mA.,nak
A.'dop.tie
A.'pos.tel
A.'ffi.che
A.'ppen.dix
A.'man.del
A.'koes.tisch
nA.'tuur.lijk
kA.'rak.ter
pA.'ssa.ge
gA.'ra.ge
kA.'len.der
lA.'wi.ne
cA.'sse.tte
kA(s).'stan.je
plA.'ce.bo
,a.ppA.'raat
,pa.rA.'llel
,ca.nA.'dees
,A.ppa.'raat
,pA.ra.'llel
,cA.na.'dees
,si.gA.'ret
,per.mA.'nent
,mo.zA.'iek
,man.dA.'rijn
,A.de.'quaat
,A.nar.'chie
,A.ller.'gie
,A.me.'thist
,A.bri.'koos
,jA.loe.'zie
,rA.di.'caal
,lA.te.'raal
A.'lleen
A.'pril
A.'dres
A.'pplaus
A.'zijn
A.'buis
A.'naal
mA.'nier
rA.'pport
pA(s).'stoor
kA.'naal
nA.'ief
kA.'neel
pA.'sseer
mAi.'llot
kwAAl
kwAl
gAAs
gAs
AAl
Al
AAs
As
bAAn
bAn
'bA.nen
'bA.nnen
'gA.pen
'gA.ppen
'kwA.len
'kwA.llen
'bA.ken
'bA.kken
'A.der
'A.dder
0.00
0.05
0.10
0.15
0.20
0.25


P19
'fa.rA.,o
'a.nA.,nas
'ma.rA.,thon
'pa.nA.,ma
'fA.ra.,o
'A.na.,nas
'mA.ra.,thon
'pA.na.,ma
'A.fri.,ka
'A.ni.,mo
'pA.gi.,na
'mA.jes.,teit
'pA.pri.,ka
'kA.zach.,stan
'bun.gA.,low
'boe.dA.,pest
'al.fA.,bet
'boe.kA.,rest
'al.mA.,nak
A.'dop.tie
A.'pos.tel
A.'ffi.che
A.'ppen.dix
A.'man.del
A.'koes.tisch
nA.'tuur.lijk
kA.'rak.ter
pA.'ssa.ge
gA.'ra.ge
kA.'len.der
lA.'wi.ne
cA.'sse.tte
kA(s).'stan.je
plA.'ce.bo
,a.ppA.'raat
,pa.rA.'llel
,ca.nA.'dees
,A.ppa.'raat
,pA.ra.'llel
,cA.na.'dees
,si.gA.'ret
,per.mA.'nent
,mo.zA.'iek
,man.dA.'rijn
,A.de.'quaat
,A.nar.'chie
,A.ller.'gie
,A.me.'thist
,A.bri.'koos
,jA.loe.'zie
,rA.di.'caal
,lA.te.'raal
A.'lleen
A.'pril
A.'dres
A.'pplaus
A.'zijn
A.'buis
A.'naal
mA.'nier
rA.'pport
pA(s).'stoor
kA.'naal
nA.'ief
kA.'neel
pA.'sseer
mAi.'llot
kwAAl
kwAl
gAAs
gAs
AAl
Al
AAs
As
bAAn
bAn
'bA.nen
'bA.nnen
'gA.pen
'gA.ppen
'kwA.len
'kwA.llen
'bA.ken
'bA.kken
'A.der
'A.dder
0.00
0.05
0.10
0.15
0.20


P20
'fa.rA.,o
'a.nA.,nas
'ma.rA.,thon
'pa.nA.,ma
'fA.ra.,o
'A.na.,nas
'mA.ra.,thon
'pA.na.,ma
'A.fri.,ka
'A.ni.,mo
'pA.gi.,na
'mA.jes.,teit
'pA.pri.,ka
'kA.zach.,stan
'bun.gA.,low
'boe.dA.,pest
'al.fA.,bet
'boe.kA.,rest
'al.mA.,nak
A.'dop.tie
A.'pos.tel
A.'ffi.che
A.'ppen.dix
A.'man.del
A.'koes.tisch
nA.'tuur.lijk
kA.'rak.ter
pA.'ssa.ge
gA.'ra.ge
kA.'len.der
lA.'wi.ne
cA.'sse.tte
kA(s).'stan.je
plA.'ce.bo
,a.ppA.'raat
,pa.rA.'llel
,ca.nA.'dees
,A.ppa.'raat
,pA.ra.'llel
,cA.na.'dees
,si.gA.'ret
,per.mA.'nent
,mo.zA.'ïek
,man.dA.'rijn
,A.de.'quaat
,A.nar.'chie
,A.ller.'gie
,A.me.'thist
,A.bri.'koos
,jA.loe.'zie
,rA.di.'caal
,lA.te.'raal
A.'lleen
A.'pril
A.'dres
A.'pplaus
A.'zijn
A.'buis
A.'naal
mA.'nier
rA.'pport
pA(s).'stoor
kA.'naal
nA.'ïef
kA.'neel
pA.'sseer
mAi.'llot
kwAAl
kwAl
gAAs
gAs
AAl
Al
AAs
As
bAAn
bAn
'bA.nen
'bA.nnen
'gA.pen
'gA.ppen
'kwA.len
'kwA.llen
'bA.ken
'bA.kken
'A.der
'A.dder
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35