



Word frequency, vowel length and vowel quality in speech production: An EMA study of the importance of experience

Fabian Tomaschek, Martijn Wieling, Denis Arnold and Harald Baayen

Quantitative Linguistics, University of Tübingen

fabian.tomaschek@uni-tuebingen.de, wieling@gmail.com,
denis.arnold@uni-tuebingen.de, harald.baayen@uni-tuebingen.de

Abstract

A frequently replicated finding is that higher frequency words tend to be shorter and contain more strongly reduced vowels. However, little is known about potential differences in the articulatory gestures for high vs. low frequency words. The present study made use of electromagnetic articulography to investigate the production of two German vowels, [i] and [a], embedded in high and low frequency words. We found that word frequency differently affected the production of [i] and [a] at the temporal as well as the gestural level. Higher frequency of use predicted greater acoustic durations for long vowels; reduced durations for short vowels; articulatory trajectories with greater tongue height for [i] and more pronounced downward articulatory trajectories for [a]. These results show that the phonological contrast between short and long vowels is learned better with experience, and challenge both the Smooth Signal Redundancy Hypothesis and current theories of German phonology.

Index Terms: articulography, vowels, word frequency, length, quality, learning.

1. Introduction

It is well known that how often a word is pronounced affects its phonetic form. High frequency (HF) words tend to have a lower number of segments [1] and shorter acoustic durations [2] as well as a higher probability of deleting a segment [3, 4] than low frequency (LF) words.

In English, HF words have been shown to contain more centralized and shorter vowels than LF words [5]. In addition, the vowel space in HF words is more contracted than in LF words [6, 7]. These effects are explained by the Smooth Signal Redundancy hypothesis [8, 9]: More frequent words have a higher probability of occurrence and hence carry less information. Under the assumption that it is important to keep the amount of information per time unit constant in a noisy channel, frequent words would have shorter acoustic durations.

Previous studies concentrated on acoustic measurements such as the duration of a segment/word or indirect measures of vowel height such as formant values in the center of the vowel. In order to assess directly the dynamics of vowel production (see e.g. [10] and [11] for the effect of formant changes as a function of time in English) the present study focused on the articulatory movements of the tongue associated with the production of the vowel in HF and LF words. The central question motivating the present study is: Do tongue movements reveal less articulatory detail with increasing frequency of occurrence? We addressed this question through an investigation of the articulatory gestures of the German vowels [i] and [a].

[a] short HF	Wasser, lassen, Staffel
[a] long HF	Vater, Name, Laden
[i] short HF	Bitte, sitzen, Zimmer
[i] long HF	wieder, diese, niemand
[a] short LF	Masse, Tarnung, Stamm
[a] long LF	Pfad, Status, Tafel
[i] short LF	Bischof, Stille, Schilde
[i] long LF	Meeting, Dealer, Liebste

Table 1. *Word material*

2. Stimuli and methods

2.1. Stimuli

A total of 24 real German words (Table 1) with stress on the first syllable were used as stimulus material and selected according to the following parameters:

- The vowel in the stressed syllable was either high [i] or low [a] and was crossed with phonological length (short, long). The two variants of [i] are distinguished both by acoustic duration and by vowel height. The two variants of [a] are distinguished by duration only [12, 13].
- The consonantal context of the vowel in each of the words was CVC with the places of articulation being coronal-V-coronal, coronal-V-labial or labial-V-coronal (e.g. *Laden* 'shop', *Name* 'name', *Vater* 'father').
- Furthermore, words were differentiated by frequency. For each vowel, three HF and three LF words were selected based on the spoken frequency extracted from the CELEX database [14]. Frequency was log transformed (range: 1.8 to 7.4). The median log-frequency from CELEX served as the boundary between HF and LF words.

2.2 Recording method

All recordings were conducted in a sound proof booth at the Department of Linguistics of the University of Tübingen. A total of 11 native German subjects (mean age: 22.9, sd = 3.1, number of females = 6) were instructed to read aloud the stimuli presented in pseudo random order. Each word appeared three times in the list.

Articulatory movements of the tongue were recorded with the NDI wave articulograph at a sampling frequency of 100 Hz. Simultaneously, the audio signal was recorded (Sampling rate: 22.05 kHz, 16bit) and synchronized with the articulatory recordings. To correct for head movements and to define a local coordinate system, a reference sensor was attached to the subjects' forehead. In addition, three sensors were placed on

10.21437/Interspeech.2013-347

the tongue: slightly behind the tongue tip (TT), at the tongue middle (TM) and the tongue body (distance between each sensor: ~2cm). Before the tongue sensors were attached, a bite plate recording was made to determine the rotation from the local reference to a standardized coordinate system (defined by the bite plate). During this recording, the subject was instructed to bite on a plastic card to which three sensors in a triangular configuration were attached.

2.3. Preprocessing

The recorded positions of the tongue sensors were centered at the midpoint of the bite plate and rotated in such a way that the front-back direction of the tongue was aligned to the x-axis with more positive values towards to the front of the mouth, and more positive z-values towards the top of the oral cavity. Movement trajectories were low-pass filtered at 15 Hz with a Butterworth-filter in line with [15].

To determine segment boundaries, the audio signal was automatically aligned with phonetic transcriptions by means of a Hidden-Markov-Model-based forced aligner for German [16]. Alignments were manually verified and corrected where necessary. The onset and offset time points of each vowel in every word were used to identify the movement trajectories of the three tongue sensors.

3. Analysis and results

3.1. Vowel duration

In order to investigate the effect of word frequency on vowel duration, a mixed-effect regression analysis was conducted with vowel duration as the response variable using the lme4 package for R [17], and as predictors word frequency (**HF**, **LF**), length (short, **long**) and vowel (**[i]**, **[a]**), using treatment coding (reference levels in bold). Subjects and words were included as a random-effect factors. The results of the analysis are presented in Table 2 and Figure 1. Length yielded a significant main effect. Short vowels were 83 ms shorter than long vowels. Additionally, vowel category affected vowel duration with [i] vowels being 45 ms shorter than [a] vowels. These effects were in line with previous findings and therefore expected [12, 18]. Interestingly, length interacted with word frequency: in HF words, long vowels became longer whereas short vowels became shorter as compared to LF words. As a consequence, there was a large durational distinction between long and short vowels in HF words. No other interactions have been found.

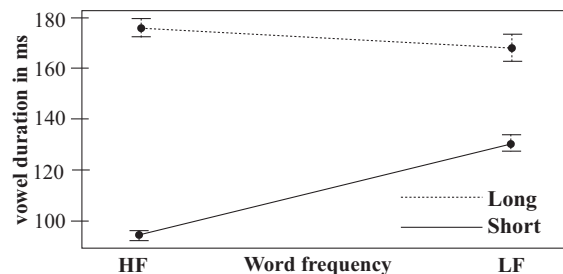


Figure 1: Effects of word frequency on vowel duration. HF=high frequency, LF=low frequency words

	Estimate	Std.Error	t-value
Intercept	198	13ms	14.4
Length	-82ms	15ms	-5.3
VowCat	-44ms	11ms	-3.7
WordFreq	-21ms	16ms	-1.3
WF*Length	55ms	23ms	2.4

Table 2: Mixed-effect model predicting vowel duration

Factor level	edf [i]	p-value [i]	edf [a]	p-value [a]
TB long HF	10.7	<0.0001	12.9	<0.0001
TM long HF	6.7	<0.0001	6.0	<0.0001
TT long HF	12.7	<0.0001	11.9	<0.0001
TB short HF	11.6	<0.0001	10.7	<0.0001
TM short HF	1.0	0.012557	4.6	<0.0001
TT short HF	11.6	<0.0001	12.5	<0.0001
TB long LF	13.1	<0.0001	13.2	<0.0001
TM long LF	1.0	0.0002	5.0	<0.0001
TT long LF	11.3	<0.0001	11.2	<0.0001
TB short LF	13.7	<0.0001	11.6	<0.0001
TM short LF	6.6	<0.0001	4.1	<0.0001
TT short LF	9.4	<0.0001	11.3	<0.0001
Random Effect	37.0	<0.0001	61.0	<0.0001

Table 3: GAM results fitting the tongue movements of [i] and [a] across the length/frequency conditions.

3.2. Vertical movement trajectories

Vertical movement trajectories were analyzed by means of generalized additive models (GAMs) [19]. So far, GAMs have not been used to analyze articulation data. They are, however, perfectly suited to analyze this type of data because they allow the modeling of nonlinear relationships between the predictor and the dependent variable. The non-linearity is modeled by a cubic spline, i.e. a set of smooth, connected cubic polynomials. Increasing the number of knots – the points where the polynomials are connected – allows the spline to be more ‘wiggly’. The exact wigglyness is measured by the estimated degrees of freedom (edf) invested in the spline. Since the duration of each vowel will differ from utterance to utterance per person and depending on vowel height [18] and vowel length [12], vowel duration was normalized between 0 and 1. Separate splines were fitted by using of the mgcv package for R [19] to the vertical sensor position in normalized time for each of the combinations of sensors (TB, TM, TT), vowels ([i], [a]), phonological length (short, long) and word frequency (HF, LF). A maximum of 20 knots was allowed and proved to be sufficient.

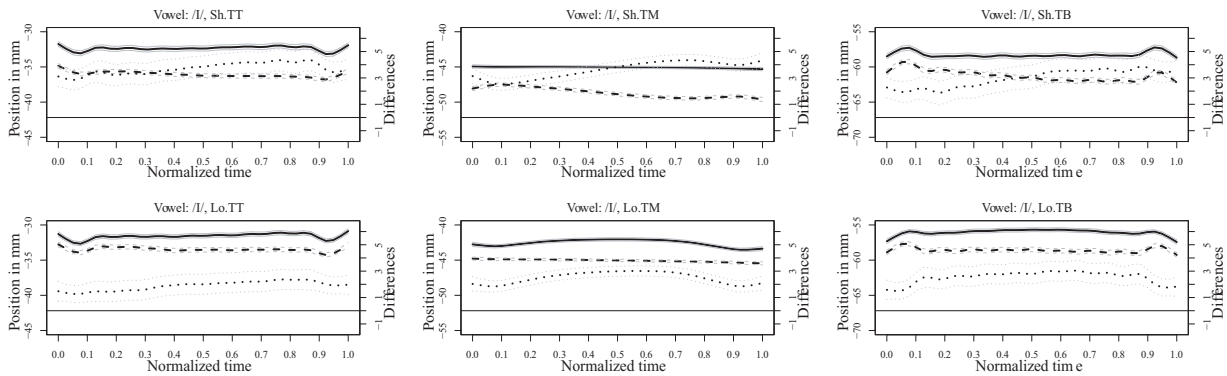


Figure 2: Vertical movement trajectories for [i]. Lo = long, Sh = short vowel. TT: tongue tip, TM: tongue middle, TB: tongue body. Solid line: high frequency, dashed line: low frequency, dotted line = differences between high and low frequency. Grey lines: 95% CI. The scale for the dotted line is shown on the right vertical axis.

Since the exact tongue movements might differ across subjects due to different morphologies of the oral cavity, by-subject random smooths for normalized time were included. These random smooths have the same function as the combination of random intercepts and random slopes in a standard linear regression analysis. The model is summarized in Table 3, where edf measures the effective degrees of freedom, providing an index of the wigglyness of each spline. More wiggly curves require more edfs. The GAM for [i] explained 88.9% of the variance in vowel height and revealed mostly non-linear tongue trajectories (two linear relationships: for TM short HF and TM long LF). The GAM for [a] explained 88.3% of the variance in vowel height and supported only non-linear trajectories.

To clarify the effect of word frequency, Figures 2 and 3 show tongue height trajectories for the different constellations of sensors and vowel length. In each figure, the solid line represents the trajectory associated with HF words, while the dashed line shows the trajectory belonging to LF words. The dotted line visualizes the difference between the two curves. Each regression curve is presented together with its 95% confidence interval (CI). The difference between HF and LF trajectories is significant wherever its 95% CI does not contain zero. A detailed description of this approach can be found in [20-22].

3.2.1. Effects of vowel length

As expected, vowel length predicted the height of the [i] trajectory. The tongue position was lower while producing

short vowels than long vowels across all sensors (Figure 2). The effect was strongest for the TM sensor. In line with [23], the differences in tongue height for [a] were much reduced as compared to [i]. Nevertheless, short [a] is consistently higher than long [a]. In addition, the two length categories differ by the gesture toward the following consonant: in the short vowel it started earlier than in the long vowel. In other words: the vowel target was reached earlier. For a thorough analysis of vowel length in articulatory, see [24, 25].

3.2.2. Effects of word frequency on trajectories

Word frequency affected the [i] trajectories in both length categories (short, long) and across all sensors. The 95% CI around the dotted line never crosses zero (Figure 2). In other words, [i] was produced higher, i.e. more tense, in HF words than in LF words. The vertical difference between the trajectories for high and low frequency words were larger for short [i] than for long [i] (Figure 2, upper vs. lower panel). For the [a] trajectories, the main effect of word frequency was much reduced. Differences can still be observed at most sensors at the onset and offset of the vowel, with a larger displacement of the tongue in HF words compared to LF words. In short vowels, differences are present only in the offset. Furthermore, the gesture towards the following consonant was initiated earlier in HF words than in LF words. The tongue reached nearly the same minimum at the center of the vowel for both HF and LF words. As a consequence, high frequency words have a more pronounced articulatory trajectory with greater tongue displacement.

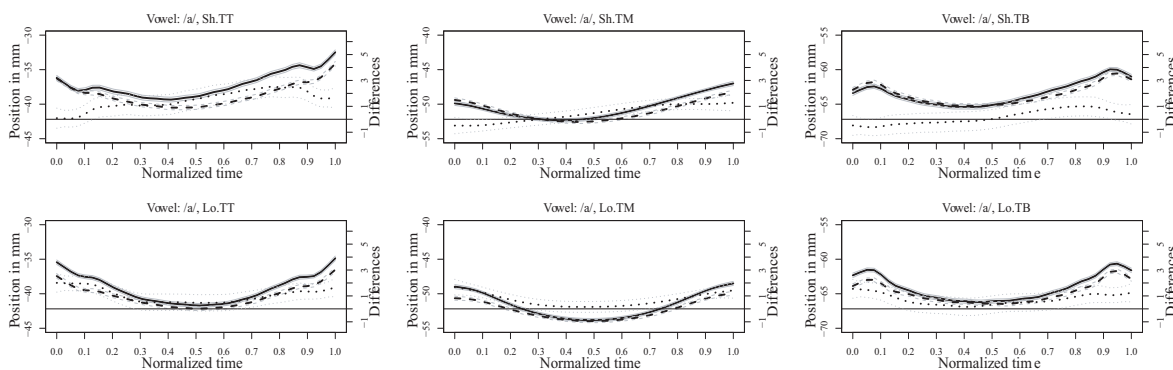


Figure 3: Vertical movement trajectories for [a]. Lo = long, Sh = short vowel. TT: tongue tip, TM: tongue middle, TB: tongue body. # Solid line: high frequency, dashed line: low frequency, dotted line = differences between high and low frequency. Grey lines: 95% CI. The scale for the dotted line is shown on the right vertical axis.

4. Discussion

It has been widely observed that more frequent words tend to have shorter acoustic durations and have fewer segments [1, 2, 26]. According to the Smooth Signal Redundancy Hypothesis, shorter realizations of higher frequency words are necessary to smooth the amount of information carried by the speech signal [8, 9]. The present study investigated differences in the realizations of vowels in high and low frequency German words using articulography. The pattern of results obtained challenges both the Smooth Signal Redundancy Hypothesis as well as current phonological theories of the German vowel system. The pattern of results points to the importance of experience for learning phonological contrasts, on a word by word basis.

With respect to acoustic duration, we found that the distinction between long and short vowels is enhanced in high frequency words and reduced or even neutralized in low frequency words (Figure 1). The effect of greater frequency is that long vowels become longer and short vowels become shorter. In other words: the contrast between long and short vowels is enhanced by experience.

With respect to the articulator trajectories, we observed higher positioned trajectories for [i] in high frequency words in contrast to low frequency words, and more pronounced U-shaped trajectories for [a]. This pattern also supports the idea that the contrast between short and long vowels becomes enhanced with experience. It is well known that children underarticulate as compared to adults due to the lack of practice (see [27] and citations therein). It takes considerable practice to master the fine muscle control of the articulatory apparatus. Our data suggests that the effect of practice is visible within an age group as a function of experience with individual words.

The Smooth Signal Redundancy Hypothesis predicts that higher frequency words should have a shorter acoustic durations. However, we observed more precise differentiation between long and short vowels in high frequency words, which gave rise to longer instead of shorter acoustic durations for [i] and [a]. Orthogonal to the principle of least effort [1] and orthogonal to the Smooth Signal Redundancy principle, the present data indicates that experience is a double edged sword. On the one hand, experience allows for reduction. But on the other hand, experience also trains for more precise execution. Our articulatory record shows that for highly practiced words more extreme articulatory trajectories are executed in terms of height ([i]) and displacement ([a]).

Phonological theories of German have sought to take one feature as primary (e.g. duration) and infer the other feature by rules (e.g. quality) [13, 28]. The way that experience modulates acoustic duration and tongue height suggests that the phonological feature of vowel length and vowel quality are intrinsically interconnected. Furthermore, from a phonological perspective one would expect that that the reduction of the durational distinction for low frequency [i] would be compensated for by a greater distinction in vowel height. However, our data do not provide strong support for such a compensatory process: the only sensor for which tongue height is lower in low frequency words with short vowels is at the tongue tip, which has little effect on the location of the first formant.

Our finding that the phonological contrast between long and short vowels is realized differentially across the lexicon depending on frequency of use fits well with theories of

language change according to which sound changes spread through a language's lexicon on a word by word basis [29]. The susceptibility of specifically lower frequency words to undergo phonological change [30] may in part be due to the lesser degree to which native contrasts have been acquired for these words.

5. References

- [1] Zipf, G.K., *The Psycho-Biology of Language. An Introduction to Dynamic Philology*. 1935, Cambridge, Mass: The M.I.T. Press.
- [2] Gahl, S., "Thyme" and "Time" are not homophones. *Word durations in spontaneous speech*. *Language*, 2008. **84**(3): p. 474-496.
- [3] Jurafsky, D., et al., *Reduction of English function words in Switchboard*. Proceedings of ICSLP-98, 1998: p. 3111-3114.
- [4] Pluymaekers, M., M. Ernestus, and R.H. Baayen, *Lexical frequency and acoustic reduction in spoken Dutch*. *The Journal of the Acoustical Society of America*, 2005. **118**(4): p. 2561-2569.
- [5] Aylett, M. and A. Turk, *Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei* *Journal of the Acoustical Society of America*, 2006. **119**(5): p. 3048-3058.
- [6] Munson, B. and N.P. Solomon, *The effect of phonological neighborhood density on vowel articulation*. *Journal of speech and hearing research*, 2004. **47**: p. 1048-1058.
- [7] Munson, B., *Lexical access, lexical representation, and vowel production*. *Laboratory Phonology* 2007. **9**: p. 201-228.
- [8] Jaeger, F., *Redundancy and reduction: Speakers manage syntactic information density*. *Cognitive Psychology*, 2010. **61**(1): p. 23-62.
- [9] Aylett, M. and A. Turk, *The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech*. *Language and Speech*, 2004. **47**(1): p. 31-56.
- [10] Nearey, T.M. and P.F. Assmann, *Modeling the role of inherent spectral change in vowel identification*. *The Journal of the Acoustical Society of America* 1986. **80**(5): p. 1297-1308
- [11] Nearey, T.M., *Static, dynamic, and relational properties in vowel perception*. *The Journal of the Acoustical Society of America* 1989 **85**(5): p. 2088-2113
- [12] Tomaschek, F., H. Truckenbrodt, and I. Hertrich, *German Vowel Quantity: Categorical Perception or Perceptual Magnet Effect?*, in *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)* 2011: Hong Kong. p. 2002-2005.
- [13] Wiese, R., *The phonology of German*. *The phonology of the world's languages 2000*, Oxford: University Press.
- [14] Baayen, H., R. Piepenbrock, and H.v. Rijn, *The CELEX Lexical Database, Release 2 (CD-ROM)*, in *Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA*. 1995.
- [15] Yunusova, Y., et al., *Positional targets for lingual consonants defined using electromagnetic articulography*. *Journal of the Acoustical Society of America*, 2012. **132**(2): p. 1027-1038.
- [16] Rapp, S. *Automatic phonemic transcription and linguistic annotation from known text with Hidden Markov Models / An Aligner for German*. in *Proceedings of ELSNET goes east and IMACS Workshop "Integration of Language and Speech in Academia and Industry"*. 1995. Moscow.
- [17] Baayen, R.H., D.J. Davidson, and D.M. Bates, *Mixed-effects modeling with crossed random effects for subjects and items*. *Journal of Memory and Language*, 2008. **59**(4): p. 390-412.
- [18] Gussenhoven, C. *A vowel height split explained: Compensatory listening and speaker control*. in *Laboratory Phonology 9*. 2007.
- [19] Wood, S., *Generalized additive models: an introduction with R* 2006: CRC Press.
- [20] Kryuchkova, T., et al., *Danger and usefulness are detected early in auditory lexical processing: Evidence from electroencephalography*. *Brain and Language*, 2012. **122**: p. 81-91.

- [21] van Rij, J., *Pronoun processing: Computational, behavioral, and psychophysiological studies in children and adults*, 2012, University of Groningen.
- [22] Kösling, K., et al., *Prominence in triconstituent compounds: Pitch contours and linguistic theory*. *Language and Speech*, 2013. **0**(0): p. 1-26.
- [23] Hoffmann, A., *A comparison of native and non-native vowels featuring German as a target language: a statistical analysis using corpora*, 2011, <http://nbn-resolving.de/urn:nbn:de:bsz:21-opus-54494>: Tübingen, Germany.
- [24] Hertrich, I. and H. Ackermann, *Articulatory control of phonological vowel length contrasts: Kinematic analysis of labial gestures*. *The Journal of the Acoustical Society of America*, 1997. **102**(1): p. 523-536.
- [25] Hoole, P. and C. Mooshammer, *Articulatory analysis of the German vowel system*. *Silbenschnitt und Tonakzente*2002.
- [26] Bell, A., et al., *Predictability effects on durations of content and function words in conversational English*. *Journal of Memory and Language*, 2009. **60**(1): p. 92-111.
- [27] White, K., et al., *Adults show less sensitivity to phonetic detail in unfamiliar words, too*. *Journal of Memory and Language*, 2013. **68**(4): p. 362--378.
- [28] Reis, M., *Lauttheorie und Lautgeschichte : Untersuchungen am Beispiel d. Dehnungs- u. Kürzungsvorgänge im Deutschen ("Theory and History of phonemes")*, in *Internationale Bibliothek für allgemeine Linguistik ; 141974*, Fink: München, Germany. p. 336 S.
- [29] W, W., *Competing changes as a cause of residue*. *Language*, 1969. **45**: p. 9-25.
- [30] Wieling, M., J. Nerbonne, and R.H. Baayen, *Quantitative Social Dialectology: Explaining Linguistic Variation Geographically and Socially*. *PLoS ONE*, 2011. **6**(9): p. e23613.