What do Finnish and Central Bavarian have in common? Towards an acoustically based quantity typology

Markus Jochim[1], *Felicitas Kleber*[1]

[1]Institute of Phonetics and Speech Processing, Ludwig-Maximilians-Universität München, Germany

{markusjochim|kleber}@phonetik.uni-muenchen.de

Abstract

The aim of this study was to investigate vowel and consonant quantity in Finnish, a typical quantity language, and to set up a reference corpus for a large-scale project studying the diachronic development of quantity contrasts in German varieties. Although German is not considered a quantity language, both tense and lax vowels and voiced and voiceless stops are differentiated by vowel and closure duration, respectively. The role of these cues, however, has undergone different diachronic changes in various German varieties. To understand the conditions for such prosodic changes, the present study investigates the stability of quantity relations in an undisputed quantity language. To this end, recordings of words differing in vowel and stop length were obtained from seven older and six younger L1 Finnish speakers, both in a normal and a loud voice. We then measured vowel and stop duration and calculated the vowel to vowel-plus-consonant ratio (a measure known to differentiate German VC sequences) as well as the geminate-to-singleton ratio. Results show stability across age groups but variability across speech styles. Moreover, VC ratios were similar for Finnish and Bavarian German speakers. We discuss our findings against the background of a typology of vowel and consonant quantity.

Index Terms: vowel and consonant quantity, stability, typology, Finnish, speech production

1. Introduction

The aim of this study was to corroborate acoustically a typology of quantity usage in different languages as part of a large-scale project studying the diachronic development of quantity contrasts in German varieties. With German varieties using both durational and non-durational cues to mark the vowel length and the so-called voicing contrast, we chose Finnish – undisputedly a quantity language – as a solid basis for comparison. Beyond the Finnish data already available [1, 2, 3, 4], we need to further establish the phonetic detail pertaining to its quantity contrasts. For our typological aim, we need to be able to compare the strength of the durational cues in German phonological systems not only to the other cues inside the same systems, but also to the strength of durational cues in quantity-heavy phonological systems like Finnish. Additionally, since the use of durational cues in German varieties has changed diachronically, both on a historic timescale (from Old High German to Modern High German) [5] and within recent generations [6, 7], we also investigated the stability of the relevant durational cues in Finnish across generations and under the influence of system-internal variation (here different speech styles).

Bannert [8] has proposed a typology that classifies languages based on where they allow quantity contrasts: in vowels only (e. g. Czech); in consonants only (e. g. Italian); in both vowels and consonants, independently of each other (e. g. Finnish [2, 8]); in both vowels and consonants, but interdependently (e. g. Central Bavarian [8]); or not at all.

Many languages, including those under investigation in this paper, have two quantities: long vs. short. In Finnish, this leads to four possible types of vowel-consonant (VC) sequences: VC, Vː C, VCː, and Vː Cː (here and hereafter ː indicates phonologically long vowels and consonants, respectively). Central Bavarian employs complementary length with long vowels always preceding lenis (i. e. short) consonants and short vowels only fortis (i. e. long) consonants. That is, in this variety only two types are possible: Vː C and VCː. However, there is evidence that it has started to allow a vowel length contrast before fortis stops [6, 7], presumably influenced by Standard German.

While Finnish and Central Bavarian are part of the typology proposed in [8], Standard German is not. Standard German poses a challenge for the quantity typology, because it uses durational cues for both vowel and consonant contrasts (independently, like Finnish), but it also uses non-durational cues to support them. In particular, vowel contrasts are cued by duration and by quality [9] (note though that there is one vowel pair, /a aː/ that is distinguished solely by duration). For stops, German has a two-way contrast that is variously termed fortis/lenis or voicing contrast (see [10] for a discussion). Its main cue is aspiration, but in the absence of aspiration (e. g. before nasals as in [beːtn̩] 'to pray'), the most important cue becomes relative duration of the stop's closure phase and the preceding vowel [11]. In the remainder of the paper we refer to this cue as proportional vowel duration (PVD). Since the described vowel and consonant contrasts are (1) to some extent quantity contrasts and (2) freely combinable, it therefore seems plausible to take the same four types of VC sequences as described above for Finnish as a basis for Standard German VC sequences.

PVD has been shown to separate Vː Cː and Vː C sequences (e. g. [boːtn̩] 'messengers' vs. [boːdn̩] 'floor', [12, 13]) as well as Vː Cː and VCː sequences (e. g. [biːtn̩] 'to offer' vs. [bɪtn̩] 'to request'], [6]) in different varieties of Standard German. The present study asks whether PVD is a good measure (1) to demonstrate the phonemic four-way length contrast in Finnish VC sequences and (2) for an acoustically based quantity typology as suggested by [8]. Table 1 gives a first impression of how similar PVD values (recalculated from previous studies) are across languages.

Three major questions arise from table 1. (a) German and Finnish appear to implement the four types differently in terms of PVD. Is this due to the different usage of non-durational cues in these two languages, and should they therefore be treated differently (i. e. due to the different phonetic implementation) or the same (because of similar kinds of phonemic categories) typology-wise? (b) If Bavarian is developing a third category Vː Cː (which may be governed by dialect leveling with Standard German but is certainly not governed by an assimilation to Finnish), will it adopt Standard German's phonetic imple-

Table 1: *Proportional vowel duration (PVD) in VC sequences; n/a refers to missing data for a particular sequence in the respective analysis. Vowel duration is either given as a proportion of the vowel+closure sequence (German, Central Bavarian upper row) or of the vowel+stop sequence (Finnish, Central Bavarian lower row).*

Language [Source]	V:C	V:C:	VC	VC:
German [11]	0.76	0.58	n/a	n/a
German [14]	0.75	0.68	0.61	0.50
Finnish [1]	0.68	0.47	0.45	0.35
Central Bavarian [6, 13]	0.76	0.56	n/a	0.31
	0.69	0.44	n/a	0.21

mentation (as the values in the upper row suggest), or will it be more similar to Finnish (as the values in the lower row give reason to expect, and perhaps for the same typological reasons that set Finnish and German apart)? There are two differently calculated sets of PVD values for Bavarian: One includes aspiration in the calculation and the other does not. The difference between the two sets clearly demonstrates that an acoustically based quantity typology needs to consider both lower-level units such as phones (i. e. the closure phase) and higher-level units such as phonemes (i. e. the entire stop). And (c), since VC and V:C: are very close in Finnish but further apart in German, can this measure be used to separate all four categories in a language like Finnish? Such a separation depends largely on the dispersion of a given data set, but from [1] we only know the mean. As a first step towards an acoustically based quantity typology, we will therefore focus on question (c) in the present study.

Thus, our first research question is whether the four Finnish quantity categories can be separated by means of PVD and how this measure performs in relation to absolute duration and geminate to singleton ratio that have been investigated in previous studies [1, 2, 3, 4]. In order to compare (in future studies) the outcome of the present study to ongoing diachronic developments in Germanic languages, our second research question is how stable the observed patterns remain under the influence of system-internal variation. We chose to test the difference between younger and older speakers. We do not, however, expect any substantial age differences in Finnish, since we are not aware of any instability reports regarding the language. Moreover, as a within-speaker type of variation, we chose to test the difference between normal and loud speech. Differences in loudness are known to correlate with speech rate (with louder speech being slower than normal speech [15]; we did not vary rate directly to allow future comparison with children's data [16]).

2. Method

2.1. Material

We analyzed 13 words (table 2) of a 45-word corpus. All but two of the corpus words, and all of the analyzed words, were structured $C_1V_1C_2V_2$. Within the 13 target words, V_1, C_2, and V_2 were either short or long, and both C_1 and C_2 were stops.

Special emphasis has been put on the words *taka* and *taakka*, which form the only minimal pair in our corpus that contrasts VC and V:C:; and on the words *kota/koota/tutti/tuutti*, which contrast all four types of VC while preserving the identity of C_2 and the quantity of V_2. Moreover, all V_1 in these four

words are high/mid-high back vowels and show strong overlap in their formant frequencies F1 and F2.

Table 2: *Target words analyzed in the current study.*

Finnish	English
kota	'capsule'
koota	'put together, collect'
tutti	'pacifier'
tuutti	'cone'
taka	'back, rear or hind (prefix)'
taakka	'burden'
takka	'chimney'
kaato	'bull's eye'
katto	'roof'
kiitää	'to race'
kiittää	'to thank'
tapaa	'to meet sb.'
tappaa	'to kill'

2.2. Participants and Recording Procedure

13 native speakers of Finnish took part in the experiment (9 female, 4 male). They were assigned to one of two age groups: younger (born 1995-1997) and older (six born 1950-1962, one born 1971). The recordings were made in 2016. The apparent-time design was used to test the (in-)stability of the cues involved. Moreover, the young group allows for a real-time comparison with data of then-young speakers described in Lehtonen [1]. Ten participants lived in the region of Uusimaa (located in South Finland and including Helsinki) at the time of recording. The other three had also lived there, but had moved to Munich, Germany, within two years before the recordings were made. Participants were paid.

The speakers were recorded at their own homes, using a laptop computer, mobile recording equipment (Beyer-Dynamic headset microphone, M-Audio audio interface) and SpeechRecorder [17] (version 3.4.2). The digital audio signals were sampled at 44.1 kHz, with a 16-bit resolution.

Each of the 45 target words was embedded into the carrier sentence *Sano X yhden kerran* 'say X once'. Six repetitions of each sentence were presented one at a time and in randomized order on the laptop screen. The recording sessions were divided into six blocks, each consisting of all 45 words. The participants were asked to read the sentences in a normal voice in blocks 1, 3, and 5, and in a loud voice in blocks 2, 4, and 6.

2.3. Analysis

The recordings were automatically segmented using Web-MAUS [18]. Segment boundaries were then corrected manually where necessary. Because WebMAUS has not yet incorporated Finnish training data, we used its language-agnostic mode.

All manual corrections and the analysis were conducted using the EMU Speech Database Management System [19] (version 0.2.1) and R [20] (version 3.3.2).

The dependent variables we investigated were the absolute duration of V_1 and C_2, the respective proportional vowel duration (PVD), defined as $\frac{V}{V+C}$ (like [1] we included aspiration in the consonant duration to allow for direct comparison of all Finnish data available and based on the assumption that aspi-
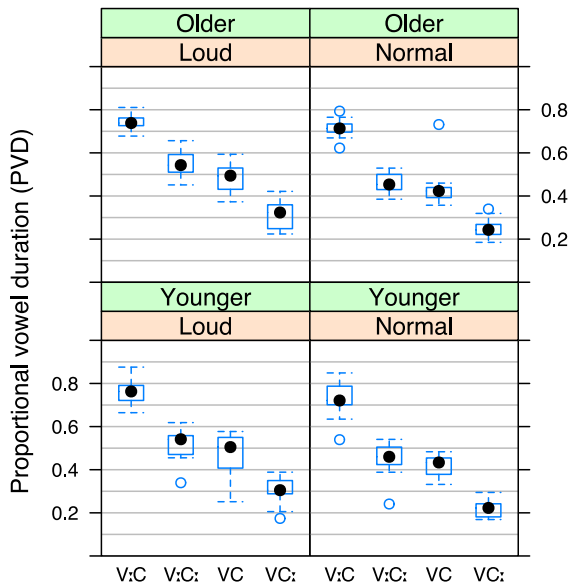
Figure 1: *PVD for kota/koota/tutti/tuutti tokens. N per boxplot is 18 (younger), 21 (older); overall N = 312.*

Figure 2: *PVD for taka/taakka tokens. N per boxplot is 18 (younger), 21 (older); overall N = 156.*

ration only plays a marginal role in Finnish), and the ratio of long vs. short segments (V_1 ratio: $\frac{V:}{V}$, C_2 ratio: $\frac{C:}{C}$).

Our independent variables were age (younger/older), speech style (normal/loud), and category (V:C/V:C:/VC/VC:). For some tests, category was reduced to two factors V_1 and C_2 quantity (long/short). All factors except age were varied within-subjects.

3. Results

3.1. Proportional Vowel Duration (PVD)

Commensurate with fig. 1, a repeated measures ANOVA with PVD as the dependent variable revealed significant main effects for category ($F[3, 33] = 435.5, p < 0.001$) and speech style ($F[1, 11] = 30.6, p < 0.001$), as well as a significant interaction effect for category × speech style ($F[3, 33] = 5.2, p < 0.01$). To prevent any potential effects of vowel height or position on PVD, the analysis was run on o/u-word tokens only.

In order to specifically test VC against V:C:, and again to ensure best comparability, we ran another analysis on the tokens of *taka* and *taakka*. Commensurate with fig. 2, the ANOVA revealed main effects for category ($F[1, 11] = 25.9, p < 0.001$) and speech style ($F[1, 11] = 42.9, p < 0.001$), but no statistically significant interactions.

These findings suggest a difference between the categories VC and V:C: that is subtle, yet robust and statistically significant. The difference in mean (fig. 1 and 2 show the median) between *taka* and *taakka* is between 5 and 6 % for loud speech (younger and older) and for the younger speakers' normal speech, and about 9 % for the older speakers' normal speech.[1] While the younger speakers show substantial overlap

---

[1] We also calculated the PVD measure with two types of logarithmic transformations, defining it as $\frac{ln(V)}{ln(V)+ln(C)}$ or $\frac{ln(V)}{ln(V+C)}$, respectively.
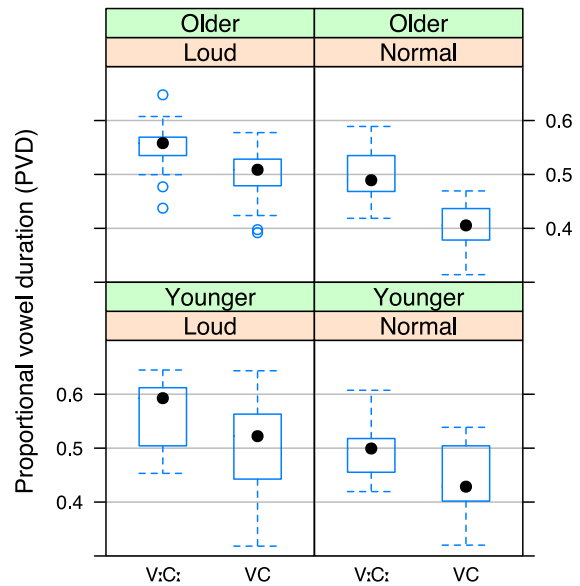
between the two categories, the older speakers show very little.

In general, PVD appears to increase in loud speech in all four types of VC sequences, but it does so to the same extent in all four categories.

The two endpoint categories in normal speech show PVD means of 22 % (younger, VC:), 25 % (older, VC:), 73 % (younger, V:C), and 71 % (older, V:C), respectively. For V:C, this is similar to Lehtonen's data, but for VC:, it differs substantially (see table 1).

3.2. Absolute V_1 and C_2 duration

The absolute durations of V_1 and C_2 are shown in a scatter plot in fig. 3. We observe four clearly separated clusters, one for each type of VC sequence. The overlap between them is remarkably small. However, it increases in loud speech. If the durations were completely independent of each other, we would expect the four clusters to form a rectangle along the two dimensions. This, however, is not the case. While V:C and VC: sequences show greater dispersion along the dimensions of vowel and consonant duration, respectively, V:C: and VC sequences vary along both dimensions although they differ greatly in the degree of dispersion. These category-dependent distributions suggest that (1) variation is greater in long than in short phonemes (see [21] for similar results in German), (2) the two vowel categories overlap to a greater extent when preceding long as opposed to short consonants, and (3) the overlap between the two consonant categories is not affected by V_1 quantity. This observation is in line with previous accounts of a language-independent tendency of vowels being influenced by adjacent consonants, but not the other way round [14, 22]. In section 3.3 we will evaluate this observation numerically.

With neither of them did the degree of separation between the two categories diminish.
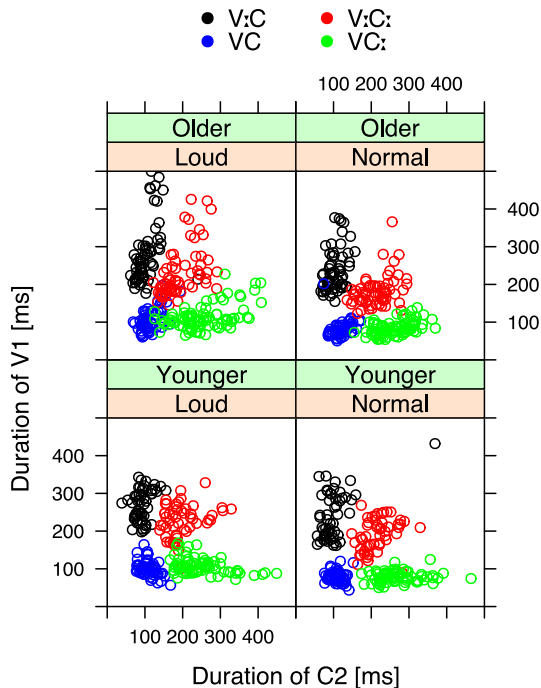
Figure 3: *Absolute duration of V_1 and C_2. Included are tokens of all 13 target words (table 2), separated by speech style and age group. The colors encode the kind of VC sequence the respective token appears in. Overall N = 1,010.*

3.3. Geminate to Singleton ratio (GSR)

We calculated the geminate to singleton ratio for both V_1 and C_2, as a function of C_2 or V_1 quantity, respectively. The GSR in V_1 is higher before short C_2 than before long C_2 (mean: 2.87 before short, 2.22 before long C_2); this effect turned out to be statistically significant ($F[1,11] = 18.8, p < 0.01$). Not only the mean values, but also the dispersion of values in fig. 3 point in this direction. On the other hand, and again commensurate with fig. 3, GSR in C_2 did not differ significantly between short V_1 and long V_1 tokens (2.33 after short, 2.26 after long V_1).

4. Discussion

This study comprised two main aims: Firstly, to test whether PVD is a useful acoustic measure for an acoustically based quantity typology, and secondly, to establish the stability of duration cues used for signaling phonemic vowel and consonant quantity in Finnish. The three main findings were as follows:

(1) PVD is able to separate all four types of VC sequences in Finnish. This suggests that it may be a useful acoustic correlate for the typology. In comparison with absolute duration, the main advantage of PVD is that it constitutes a uni-dimensional measure for all four types. Absolute durations, while providing perhaps a better separation of the four categories (cf. fig. 3 vs. fig. 1 and 2), need two dimensions to achieve the same.

(2) The category separation provided by absolute durations is slightly reduced in loud speech. This appears not to be the case for PVD. This suggests that PVD is slightly more robust as a cue in terms of normalization across speech styles/rates (see [23] for similar results in Italian), which would seem plau-

sible because the measure itself may integrate normalization for rate. It would be very interesting to specifically test the perceptual relevance of PVD in Finnish, especially in light of Kohler's finding [11] that PVD is a strong perceptual cue in German.

(3) As expected, we did not find any substantial differences between the two age groups. This suggests that the acoustic basis of Finnish quantity contrasts, namely duration, has not changed within recent generations.

One problem regarding PVD, however, remains: Why would the vowel proportion in VC be smaller than in V:C:? This appears to be so in both German and Finnish and it suggests that the difference between long and short is stronger for vowels than for consonants. This might be explained in terms of the non-durational cues employed in the respective contrasts. Finnish is often regarded not to use cues such as aspiration or vowel quality to distinguish its quantity contrasts – neither in vowels nor in consonants, which would make it likely for vowel and consonant lengthening to be the same. [4], however, did investigate and find some additional cues for the stop length contrast. This could explain the bias towards more vowel lengthening and thus a higher vowel proportion in V:C:. In German, non-durational cues are known to play an important role in both vowels and consonants. However, [14] only investigated the /a a:/ contrast, where vowel quality plays a minor role, making duration especially important for vowels. This could explain why in those data, PVD is particularly high for V:C:.

Finally, how does Central Bavarian – the German variety that motivated the current study – fit in the pattern? Depending on the exact definition of PVD, the Bavarian PVD values in table 1 are either closer to the Finnish or the German PVD values: When PVD marks the vocalic proportion of a vowel+closure sequence, the Bavarian temporal patterns of the three VC categories resemble more closely those of Standard German but such a measure leaves aside an important part of the stop (namely the aspiration phase) that may very well be a relevant factor in the auditory processing of the vowel and consonant length contrast (note that [6, 13] did, unlike [11], include words with oral releases). In fact, when PVD marks the vocalic proportion of a vowel+stop sequence then the temporal patterns of the three categories measured for Bavarian are closer to those found for Finnish. In particular, the VC: category – where Central Bavarian and Finnish according to Lehtonen [1] diverge the most – in our Finnish data was much closer to the Bavarian values (our data yielded a mean of 22–25 % for Finnish VC:).

We are currently conducting further analyses of durational and non-durational cues in Central Bavarian and other German varieties to better understand the timing relations in VC sequences and their typological characteristics. Considering the entire stop in the PVD value might be the more appropriate measure for an acoustically based typology because it appears to better allow for generalization – both within (e. g. when comparing orally vs. nasally released stops) and across languages (e. g. when comparing languages that use aspiration with those that do not). After all, Standard German temporal patterns may also have something in common with Finnish temporal pattern when accounting for the entire stop in the PVD measure.

5. Acknowledgements and Data

An R script and a data frame containing the 1,010 observations this report is based on is permanently available at [24].

6. References

[1] J. Lehtonen, *Aspects of Quantity in Standard Finnish*, ser. Studia Philologica Jyväskyläensia. Jyväskylä: K. J. Gummerus, 1970, no. VI.

[2] K. Suomi, E. Meister, R. Ylitalo, and L. Meister, "Durational patterns in Northern Estonian and Northern Finnish," *Journal of Phonetics*, vol. 41, no. 1, pp. 1–16, Jan. 2013.

[3] O. Engstrand and D. Krull, "Durational Correlates of Quantity in Swedish, Finnish and Estonian: Cross-Language Evidence for a Theory of Adaptive Dispersion," *Phonetica*, vol. 51, no. 1-3, pp. 80–91, Jul. 1994.

[4] C. S. Doty, K. Idemaru, and S. G. Guion, "Singleton and geminate stops in Finnish - acoustic correlates." in *INTERSPEECH*. ISCA, 2007, pp. 2737–2740.

[5] G. Seiler, "On the development of the Bavarian quantity system," *Interdisciplinary Journal for Germanic Linguistics and Semiotic Analysis*, vol. 10, no. 1, pp. 103–129, 2005.

[6] F. Kleber, "Complementary length in vowel-consonant sequences: Acoustic and perceptual evidence for a sound change in progress in Bavarian German." *Journal of the International Phonetic Association*, in press.

[7] S. Moosmüller and J. Brandstätter, "Phonotactic information in the temporal organization of Standard Austrian German and Viennese dialect," *Language Sciencies*, vol. 46, pp. 84–95, 2014.

[8] R. Bannert, *Mittelbairische Phonologie auf akustischer und perzeptorischer Grundlage*, ser. Travaux de l'institut de linguistique de Lund, B. Malmberg and K. Hadding, Eds. CWK Gleerup, 1976, vol. 10.

[9] R. Wiese, *The Phonology of German*. Oxford University Press, 2000.

[10] A. Braun, *Zum Merkmal "Fortis/Lenis". Phonologische Betrachtungen und instrumental - phonetische Untersuchungen an einem mittelhessischen Dialekt*. Stuttgart: Steiner, 1988.

[11] K. J. Kohler, "Dimensions in the Perception of Fortis and Lenis Plosives," *Phonetica*, vol. 36, no. 4-5, pp. 332–343, 1979.

[12] ——, "The production of plosives," *Arbeitsberichte des Instituts für Phonetik der Universität Kiel*, vol. 8, pp. 30–110, 1977.

[13] F. Kleber, "Partielle Neutralisierung des Stimmhaftigkeitskontrasts in zwei varietäten des Deutschen," in *Perzeptive Linguistik: Phonetik, Semantik, Varietäten*, T. Krefeld and E. Pustka, Eds. Stuttgart: Steiner, 2014, pp. 19–32.

[14] N. Braunschweiler, "Integrated Cues of Voicing and Vowel Length in German: A Production Study," *Language and Speech*, vol. 40, no. 4, pp. 353–376, Oct. 1997.

[15] C. Dromey and L. O. Ramig, "Intentional Changes in Sound Pressure Level and RateTheir Impact on Measures of Respiration, Phonation, and Articulation," *Journal of Speech, Language, and Hearing Research*, vol. 41, no. 5, pp. 1003–1018, Oct. 1998.

[16] F. Kleber, "On the role of temporal variability in the acquisition of the german vowel length contrast," in *INTERSPEECH*. ISCA, 2017.

[17] C. Draxler and K. Jänsch, "SpeechRecorder - a Universal Platform Independent Multi-Channel Audio Recording Software," in *Proc. of the IV. International Conference on Language Resources and Evaluation*, Lisbon, Portugal, 2004, pp. 559–562.

[18] T. Kisler, U. Reichel, F. Schiel, C. Draxler, B. Jackl, and N. Pörner, "BAS Speech Science Web Services - an Update of Current Developments," in *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*. Portorož, Slovenia: European Language Resources Association (ELRA), 2016.

[19] R. Winkelmann, J. Harrington, and K. Jänsch, "EMU-SDMS: Advanced speech database management and analysis in R," *Computer Speech & Language*, in press.

[20] R Core Team, "R: A Language and Environment for Statistical Computing," Vienna, Austria, 2016. [Online]. Available: https://www.R-project.org/

[21] C. Mooshammer and C. Geng, "Acoustic and articulatory manifestations of vowel reduction in german," *Journal of the International Phonetic Association*, vol. 38, no. 02, pp. 117–136, 2008.

[22] M. Chen, "Vowel Length Variation as a Function of the Voicing of the Consonant Environment," *Phonetica*, vol. 22, no. 3, pp. 129–159, 1970.

[23] E. R. Pickett, S. E. Blumstein, and M. W. Burton, "Effects of speaking rate on the singleton/geminate consonant contrast in italian," *Phonetica*, vol. 56, no. 3-4, pp. 135–157, 1999.

[24] M. Jochim and F. Kleber, *What do Finnish and Central Bavarian have in common? Towards an acoustically based quantity typology*. Open Data LMU, 2017, doi: 10.5282/ubm/data.106.