TEMPORAL ORGANIZATION OF VOWEL PLUS STOP SEQUENCES IN PRODUCTION AND PERCEPTION: EVIDENCE FROM THE THREE MAJOR VARIETIES OF GERMAN

Nicola Klingler[1], Felicitas Kleber[2], Markus Jochim[2], Michael Pucher[1], Stephan Schmid[3] Urban Zihlmann[3]

[1]Acoustics Research Institute, Austrian Academy of Sciences; [2]Institute of Phonetics and Speech Processing, Ludwig-Maximilians-University Munich, [3]Phonetics Laboratory, University of Zurich
[nklingler|mpucher]@kfs.oeaw.ac.at, [kleber|markusjochim]@phonetik-uni-muenchen.de,
[stephan.schmid|urban.zihlmann]@uzh.ch

ABSTRACT

This study investigated cross-linguistically the temporal organization of short vs. long vowels (V vs. V:) with following lenis vs. fortis (C vs. C:) stops in disyllabic trochees in the three major varieties of Standard German (Austrian, German, and Swiss). A total of 51 speakers of the three varieties participated in a production and perception experiment. Acoustic analyses revealed that Austrian speakers take up an intermediate position regarding a stop duration contrast which was clearly present in Swiss but absent in German speakers. In perception, however, Austrians and Germans but not Swiss participants judged stimuli from a vowel/(vowel+closure) duration (VC-ratio) continuum more often as containing V:C, i.e., intermediate VC-ratios were reinterpreted as long vowels by Austrians and Germans, but as fortis stops by Swiss participants. Findings suggest a sometimes less stable temporal organization in Austrians. Moreover, results imply a greater diversity in phonological structure between German varieties than previously assumed.

Keywords: Phonemic duration contrasts, Major varieties

1. INTRODUCTION

The length opposition in stressed vowels and the postvocalic voicing contrast in stops (henceforth fortis/lenis, cf. [10]) are both well-studied for so-called standard German [5, 8, 9], which is, however, regionally bound to the north of Germany [5]. It is less clear how the standard variety spoken in the south of Germany or the standard varieties in other German speaking countries like Austria and Switzerland implement the vowel length contrast before fortis and lenis stops. This contribution focuses on the three major standard varieties of German spoken in Austria, Germany, and Switzerland. All three but in particular the Austrian and Swiss variety have been apparently shaped by the different phonological systems of the various regionally restricted dialects (Bavarian → Austrian, Alemannic → Swiss) spoken by the majority of the then bilingual population but which hitherto has been omitted from phonetic descriptions of standard German all too often. These dialects all trace back to Old High German but developed differently over the course of time. The development from Old High German to Modern High German (varieties) led, among others, to a change from a syllable-timed language with a length contrast in vowels (long vs. short) and consonants (geminates vs. singletons) to a stress-timed language with a phonemic length contrast in vowels and a strength contrast in stops [14, 11, 12]. Alemannic and Bavarian, however, have preserved (at least to some extent) the consonant length contrast [16].

One of the aims is therefore to investigate whether the Austrian and Swiss standard varieties, too, show patterns of a consonant length distinction. With respect to standard German as spoken in the north of Germany, phonemic vowel length is mainly cued via duration, but also by differences in vowel quality (e.g. long /iː, eː/ vs. short /ɪ, ɛ/ [21]) and the fortis/lenis contrast primarily by the presence/absence of aspiration (i.e., /pʰ/ vs. /b̥/, etc.) and only secondarily by the length of the closure phase. The Alemannic dialects spoken in Switzerland (but not the standard variety), on the other hand, have only a very limited set of aspirated stops while the fortis/lenis distinction is one of closure duration (cf. [13]). This has led some scholars [11, 12] to interpret these acoustic duration differences in terms of a phonological contrast between singletons and geminates. The vowel length opposition in Swiss dialects is even more systematic regarding the duration cue than in northern German with non-remarkable height differences as a function of underlying length

[20]. The phonological systems of German standard German and Swiss dialects (and the Swiss standard variety for that matter) allow for a free combination of long and short vowels with fortis and lenis stops, respectively (e.g., /biːtʰɐ/ 'bidder' /bɪtʰɐ/ 'bitter', /viːdɐ/ 'again', /vɪdɐ/ 'ram'). Such a four-way combination is prohibited in the dialects of Bavarian which are spoken in the southeast of Germany and Austria and which have shaped the Austrian standard variety considerably (cf. e.g. [15, 3, 17]). Bavarian and the Austrian standard variety are said to feature a duration-based fortis (≙ long, C:)/lenis (≙ short, C) contrast (aspiration only plays a secondary role) from which – at least in Bavarian – vowel length is predictable: Lenis stops are always preceded by long vowels and fortis stops by short vowels. The Austrian standard variety, however, appears to pattern with the German standard variety in that it allows combinations of long vowels and fortis stops [15].

Another aim is to test this claim for the Austrian standard variety in this cross-linguistic study that specifically and for the first time investigates the temporal organization of the combined vowel length and fortis/lenis-contrast in the production and perception of the standard varieties of German spoken in the south of Germany, in Austria, and in Switzerland. The production results served as a basis for the perception experiment; the perception data in turn complements the phonemic analyses of the two contrasts. The overall goals are to further our understanding of (1) the implementation of the two contrasts in German and (2) the development of major varieties in geographically close countries (as opposed to the major varieties of English).

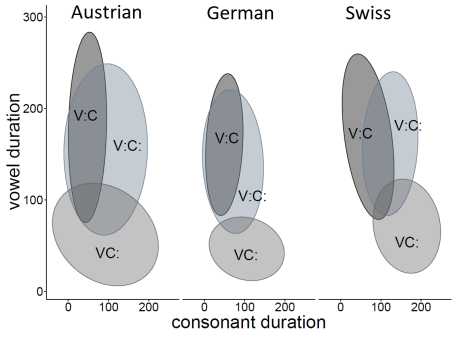
2. PRODUCTION EXPERIMENT

2.1. Method

16 speakers of the Austrian standard variety from Vienna, 16 bilingual Swiss participants from Zurich and 19 speakers of the German standard variety from Munich (n=51, 25 male) were recorded in sound-attenuated booths using the SpeechRecorder Software [4] with a sampling rate of 44.1 kHz. The participants were asked to read silently sentences presented in standard German orthography on a computer screen and to repeat them loudly after the written text had disappeared from the screen. With this procedure we expect to diminish the influence of the orthographic representation and to elicit a natural pronunciation of the standard variety. The target words in the production experiment analyzed here were five repetitions each of the words *Hagen* (/haːgən/, VːC, a proper name), *Haken* (/haːkən/, VːCː, 'hook'), and *hacken* (/hakən/, VCː, 'to hack') taken from a larger corpus. The data were automatically segmented with WebMAUS [19] and analysed in R [18] using emuR [22]. As this study focuses on durational cues, the acoustic measurements include solely the vowel and consonant duration (in ms). We then calculated the V/(V+C) ratio where V corresponds to the duration of the vowel and C to that of the consonant (henceforth VC-ratio). A similar measure (where C = closure phase) has been shown to capture the voicing [9] and the vowel length contrast in German near standard varieties [7]. Since WebMAUS only marks the on- and offset of the entire consonant the VC-ratio as defined here yields smaller values in words with aspirated stops than previous studies. The statistical analyses were carried out by fitting Linear Mixed Effect Models (LMM, [1]) with either VC-ratio or vowel or consonant duration as dependent factor. Region (three levels), VC sequence (VːC, VːCː, VCː) were entered as fixed factors and speaker as random factor.

2.2. Results

Figure 1: Distribution of vowel and consonant duration (in ms) separately for the VC sequences and regions. The ellipses comprise 95 % of the underlying data points.

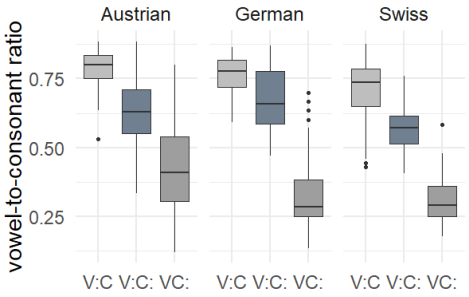


Commensurate with Fig. 1 all three groups produced a clear duration-based contrast between phonologically long and short vowels. Vowel duration was significantly affected by region ($F[2,51] = 4.27, p = .019$) and VC sequence ($F[2,50] = 385.10, p < .0001$) but differently so as the interaction effect between the two factors ($F[4,50] = 2.73, p = .039$) and posthoc tests reveal.

While the largely overlapping VːC and VːCː distributions of German speakers along the x-axis in Fig. 1 suggest the absence of a consonant duration contrast, the tendency towards such a contrast was much more pronounced in Austrian speak-

ers but not as clear as for Swiss speakers. That is, Austrian German takes up an intermediate position between the Swiss duration-based and the German non-durational fortis/lenis contrast. The LMM with consonant duration as the dependent variable revealed a significant effect for VC sequence ($F[2,51] = 138.37$, $p < .0001$), region ($F[2,51] = 33.26$, $p < .0001$) as well as a significant interaction between region and VC sequence ($F[4,51] = 9.49$, $p < .0001$).

Figure 2: VC-ratios in the productions of *Hagen* (V:C), *Haken* (V:C:), *hacken* (VC:) separately for the three varieties.



As shown in Fig. 2 all speakers produce the three categories by means of different VC-ratios despite a tendency towards a less pronounced distinction between *Hagen* and *Haken* in German as opposed to Austrian and Swiss productions. In general, however, high VC-ratios signalled V:C-sequences, low VC-ratios VC:-sequence and V:C: was characterized by intermediate VC-ratios ranging between 0.5 and 0.6. The statistical analysis performed on the VC-ratio as the dependent factor revealed a main effect of region ($F[2,51] = 12.10$, $p < .0001$) and VC sequences ($F[2,50] = 443.69$, $p < .0001$) and a significant interaction between region and VC sequence ($F[4,50] = 4.83$, $p = .002$). Post-hoc tests showed that Austrian and German speakers did not differ significantly from each other in V:C and V:C:-sequences but both groups differed significantly from the production of Swiss speakers (V:C:, $t(51) = 4.1$, $p = .005$, V:C, $t(46) = 3.6$, $p = .018$) except for non-significant distinctions between German and Swiss speakers in VC:-sequences and Austrian and Swiss speakers in V:C:-sequences.

3. PERCEPTION EXPERIMENT

Based on the findings in production showing similarities in VC-ratio, on the one hand, but differences regarding consonant duration, on the other, we tested in a second experiment how the same speakers turned listeners use these acoustic cues in the perception of the combined contrasts. Following the experimental procedure in [7] we were interested in whether or not Swiss participants perceive less stimuli from a *Hagen–hacken* continuum (encompassing *Haken*) as *hacken* than listeners from the other two groups given their (tendency towards) lower VC-ratios in all sequences in production.

3.1. Method

The stimuli used in the perception experiment were derived from one natural production each of the lexeme *Hagen* uttered by two phonetically-trained approximately same-aged male speakers from Switzerland and Germany, respectively. The lexeme was embedded in a carrier phrase (*Maria hat Hagen gesagt*, lit.: 'Maria has Hagen said') and uttered by the two speakers ten times. The recordings were made in sound-attenuated booths. Pretests had revealed that the German speaker was an acceptable model talker of the standard variety for Austrian but not for Swiss listeners (mainly due to vowel quality differences in the carrier phrase). One production per speaker was chosen as the first stimulus of the continuum. The other stimuli were then derived by (1) successively lengthening the closure duration and shortening the vowel duration using Praat's [2] duration manipulation function and (2) resynthesizing each duration-specific stimulus using the overlap and add function in Praat. The participants were the same as in the production experiment. Six repetitions of each stimulus were presented to the listeners in randomized order in two settings: the listeners first judged in the 2AFC task whether a stimulus sounded more like *Hagen* or *hacken*. In the second setting – a control task – the listeners were again presented with the same stimuli but rated them now within a three alternative forced choice (3AFC) test, i.e. the in-between category *Haken* was now among the response options.

Due to differences in the speaker's fundamental frequency (Swiss: 136 Hz; German: 102 Hz) the resynthesis procedure had led to slight differences in step size between the two continua. Prior to the statistical analysis, the stimuli from the Swiss and the Austro-German continuum, respectively, were therefore matched according to the respective VC-ratio values. Sigmoid functions were fitted to the responses to the 13 stimuli from the matched continuum (cf. Tab. 1) using binary logistic regression separately per listener using equation (1)

$$p_a = \frac{e^{(m \cdot Stim + k)}}{1 + e^{(m \cdot Stim + k)}} \quad (1)$$

where p_a is the proportion of *hacken*-responses,

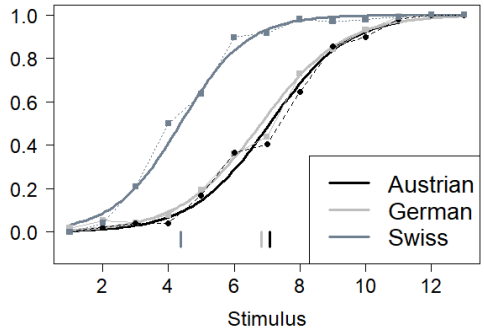
Stimulus the number of the stimulus along the continuum, and m and k the listener-specific slope and intercept, respectively. The category boundaries, calculated for each listener by $-k/m$, were then subjected to statistical analysis as described below.

Table 1: VC-ratios for each of the 13 stimuli from the matched *Hagen-hacken*-continuum.

Stimulus	1	2	3	4	5	6	7	8	9	10	11	12	13
VC-ratio	.76	.73	.69	.65	.62	.59	.57	.54	.51	.47	.45	.41	.38

3.2. Results

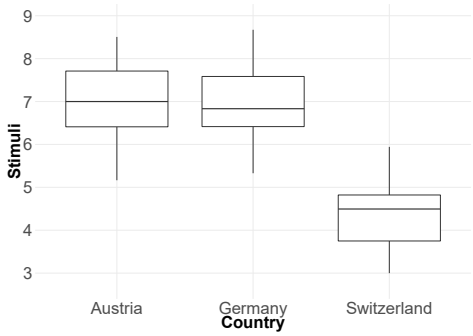
Figure 3: Proportion of *hacken*-responses in the 2AFC task aggregated across listeners per region. Superimposed are the sigmoids and mean group category boundaries.



The average response curves per listener group in the 2AFC task (Fig. 3) indicate that the three groups separated the VC-ratio continuum into two distinct categories but that the Swiss category boundary between a long and a short vowel was, unexpectedly, left-shifted compared to the Austrian and German category boundaries. Stimuli that were identified as *Haken* in the control 3AFC task, e.g. stimulus 6, were unambiguously assigned to *hacken* by Swiss listeners but lay in between categories for Austrian and German participants. Stimuli with a VC-ratio below 0.65 (i.e. from stimulus 5 up) were already perceived as *hacken* while the ratio had to be below 0.57 for Austrians and Germans to clearly indicate a short vowel.

Commensurate with Fig. 4 a repeated measures ANOVA with listener-specific category boundary as dependent variable, region as fixed factor (three levels) and listener as random factor revealed a significant effect for country ($F[2,48] = 20.1$, $p < .001$). Post-hoc Tukey tests showed no significant difference between Austrians and Germans but a significant difference between listeners from the two countries to Swiss participants (A: $t(27.3) = 7.6$, $p < .001$; G: $t(26.4) = 5.3$, $p < .001$).

Figure 4: Distribution of the listener-specific category boundaries separately for the three countries.



4. DISCUSSION AND CONCLUSION

The first finding was that in production Austrians took up an intermediate position regarding a duration-based fortis/lenis contrast. Secondly, all three varieties produced the three VC sequences by means of different VC-ratios but Swiss speakers tended towards overall lower proportional vowel durations. These findings indicate that VC-ratio masks some of the regional differences that emerged in the consonant duration (cf. [6]). Furthermore, the second finding stands in contrast to the perception results where Swiss as opposed to Austrian and German speakers perceived stimuli with a higher VC-ratio as *hacken*. That is, despite the lower VC-ratios in production Swiss listeners interpreted stimuli with in-between VC-ratios as indicating a long consonant while German and Austrian listeners heard the same VC-ratios predominantly as containing a long vowel. This discrepancy might be linked to the phonology of Alemannic dialects which in contrast to German and Bavarian allows for combinations of independent phonemic quantity contrasts in vowels and consonants. Despite the proportionally shifted V to VC duration listeners appear to focus either on the vowel or the consonant contrast depending on the regional background. This suggests that the standard varieties spoken in the three countries diverge to a greater extent on the phonological level than previously assumed and highlights the development of different major varieties in geographically close countries.

5. REFERENCES

[1] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1), 1–48.

[2] Boersma, P., Weenik, D. 2017. Praat: Doing phonetics by computer [Computer program]. Version 6.0.24. http://www.fon.hum.uva.nl/praat/.

[3] Brandstätter, J., Kaseß, C., Moosmüller, S. 2016. Quality and quantity in high vowels in Standard Austrian German. In: Leemann, A., Kolly, M.-J., Schmid, S., Dellwo, V., (eds), *Trends in Phonetics and Phonology*. Peter Lang 79–92.

[4] Draxler, C. 2011. Speech recorder quick start and user manual.

[5] Jessen, M. 1998. *Phonetics and phonology of tense and lax obstruents in German*. Amsterdam and Philadelphia: John Benjamins Pub. Co.

[6] Jochim, M., Kleber, F. 2017. What do Finnish and Central Bavarian Have in Common? Towards an Acoustically Based Quantity Typology. *Proc. of Interspeech 2017* Stockholm. ISCA 3018–3022.

[7] Kleber, F. 2017. 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* 1–22.

[8] Kohler, K. J. 1977. The production of plosives. *Arbeitsberichte des Instituts für Phonetik der Universität Kiel* 8, 30–110.

[9] Kohler, K. J. 1979. Dimensions in the perception of fortis and lenis plosives. *Phonetica* 36(4-5), 332–343.

[10] Kohler, K. J. 1984. Phonetic explanation in phonology: the feature fortis/lenis. *Phonetica* 41(3), 150–174.

[11] Kraehenmann, A., Lahiri, A. 1999. Phonological quantity contrast in Swiss German stops: history and acoustics. *Ms, University of Konstanz*.

[12] Kraehenmann, A., Lahiri, A. 2008. Duration differences in the articulation and acoustics of Swiss German word-initial geminate and singleton stops. *The Journal of the Acoustical Society of America* 123(6), 4446–4455.

[13] Ladd, D. R., Schmid, S. 2018. Obstruent voicing effects on f0, but without voicing: Phonetic correlates of Swiss German lenis, fortis, and aspirated stops. *Journal of Phonetics* 71, 229–248.

[14] Lahiri, A., Riad, T., Jacobs, H. 1999. Diachronic prosody. In: van der Hulst, H., (ed), *Word Prosodic Systems in the Languages of Europe*. Mouton de Gruyter 335–422.

[15] Moosmüller, S., Brandstätter, J. 2014. Phonotactic information in the temporal organization of Standard Austrian German and the Viennese dialect. *Language sciences (Oxford, England)* 46, 84–95.

[16] Moosmüller, S., Ringen, C. 2004. Voice and Aspiration in Austrian German Plosives. *Folia Linguistica* 38(1-2).

[17] Moosmüller, S., Schmid, C., Brandstätter, J. 2015. Standard Austrian German. *Journal of the International Phonetic Association* 45(03), 339–348.

[18] R Core Team, 2017. R: A language and environment for statistical computing.

[19] Schiel, F. 1999. Automatic phonetic transcription of non-prompted speech. *Proc. of the ICPhS 1999* San Francisco. 607–610.

[20] Schmid, S. 2004. Zur Vokalquantität in der Mundart der Stadt Zürich. *Linguistik online* 20(3), 93–116.

[21] Wiese, R. 1996. *The phonology of German*. Oxford: Clarendon Press.

[22] Winkelmann, R., Harrington, J., Jänsch, K. 2017. Emu-sdms: Advanced speech database management and analysis in R. *Computer Speech & Language* 45, 392–410.