Full title

Analyzing dialectal variation in articulation rate using crowdsourced speech data

Short title

Geospatial variation in articulation rate in Swiss German

Adrian Leemann

al764@cam.ac.uk

Phonetics Laboratory

Department of Theoretical and Applied Linguistics

University of Cambridge

Sidgwick Avenue

Cambridge, CB3 9DA

United Kingdom

+44 77 153 999 45

ABSTRACT

Most recent studies on the geographical distribution of acoustic features analyze comparatively few speakers and localities, both of which may be unrepresentative of the diversity found in larger or more spatially fragmented populations. In the present study we introduce a new paradigm that enables the crowdsourcing of acoustic features through smartphone devices. We used *Dialäkt Äpp*, a free iOS app that allows users to record themselves, to crowdsource audio data. Nearly 3,000 speakers from 452 localities in German-speaking Switzerland provided recordings; we measured articulation rates for these speakers using a metric based on duration intervals between consecutive vowel onsets. Results revealed distinct regional differences in articulation rate between major dialect regions and individual localities. The specification of 452 localities enabled analyses at an unprecedented spatial resolution. Results further revealed a robust effect of gender, with women articulating significantly more slowly than men. Both the geographical patterns and the effect of gender found in this study corroborate similar findings on Swiss German previously reported in a very limited set of localities, thus verifying the validity of the crowdsourcing framework. Given the application of this new framework, a large bulk of the discussion is devoted to discussing methodological caveats.

ACKNOWLEDGMENTS

This research was supported by the Swiss NSF (grant nr. P300P1_151210, http://p3.snf.ch/project-151210). We thank our colleague Beat Siebenhaar from the University of Leipzig, who provided insight and expertise that greatly assisted the research. We thank Tam Blaxter from the University of Cambridge for assistance with GIS, Yang Li for support on the statistics used, and Thomas Kettig for comments that greatly improved the manuscript. We would also like to show our gratitude to Marie-José Kolly, Daniel Wanitsch, Iwar Werlen, Fiona Hasler, and Carina Eisenblatt for co-developing *Dialäkt Äpp*, the tool that laid the basis for the research presented. Any errors are our own and should not tarnish the reputations of the persons acknowledged here.

1. INTRODUCTION

Speech rate can be measured in terms of speaking rate – the number of units, typically syllables or words, and silent intervals produced in a given time – or articulation rate, the number of units produced in a given time after silent intervals have been removed (Robb et al., 2004). Pauses and silent intervals have been reported to exhibit distinct between-speaker variation (cf. Kolly et al., 2015), which is why articulation rate has been said to provide a more sensitive measure of speech rate (Robb et al., 2004). Studies have demonstrated inherent speech rate differences between languages as well as between varieties of the same language. Extensive research has been conducted on within-language dialectal speech rate variation in English: Robb et al. (2004) compared speaking and articulation rates of 40 New Zealand English (Christchurch) speakers to 40 American English (Connecticut) speakers and reports that NZ English speakers overall articulated 30 syllables per minute faster. They report a higher degree of vowel reduction and elision for the NZ English speakers. For American English dialects, Jacewicz et al. (2009) and Jacewicz, Fox and Wei (2010) examined 192 speakers from two dialect regions, the North (Wisconsin) and the South (North Carolina). Northerners showed significantly higher articulation rates than Southerners, and men spoke slightly faster than women. In a more large-scale fashion – again within American English – Byrd (1994) studied the speaking rates of 630 speakers from eight different dialect regions. She found a significant effect of region, with Southerners speaking the slowest and Northeasterners and Army Brats speaking the fastest. Byrd (1994) also reports an effect of gender, with men speaking more than 6% faster than women, as well as speaker-specific speech rates.

Differences in speech rate have been reported within other languages, too. Verhoeven et al. (2004), for example, analyzed the spontaneous speech of 160 Dutch speakers from four regions in the Netherlands as well as four regions in northern Belgium (Flanders), providing evidence of significant differences in articulation rate between the two countries: 5.05 syll/sec for the Netherlands, and 4.23 syll/sec for Belgium. Men were also shown to speak more quickly than women and younger speakers spoke somewhat more quickly than older ones. Schwab and Avanzi (2015) investigated regional differences in speech rate in French varieties, examining seven regions in France, Belgium, and Switzerland (eight speakers per variety); they report that Swiss speakers, on average, have longer syllable durations (indicated in ms/syll).

Despite abundant anecdotal evidence that the German-speaking regions of Europe differ in how quickly or slowly people speak, comparatively little scientific research has been devoted to the spatial distribution of speech rate in varieties of German. Much of the research on speech rate in German is found in applied fields, such as speech technology (Siebenhaar et al., 2001; Trouvain, 2003) or forensic speaker identification (Jessen, 2007; Künzel, 1997). From a more forensic angle, Jessen (2007), for example, provided an initial description of 100 male German speakers' average articulation rates. Mean articulation rate was reported to be 5.21 syll/sec in the read condition and 5.41 and 5.19 syll/sec in the spontaneous conditions. A large bulk of the literature on speech rate in German addresses methodological factors surrounding the examination of speech rate (Kohler et al. 1981, 1982; Trouvain et al. 2001) or addresses speech perception-related aspects of speech rate (Pfitzinger, 1998, 1999). Ulbrich (2004) presents one of the few analyses on regional variation in articulation rate in German, examining the speech rate of a total of 28 news reporters from Germany, Austria, and Switzerland based on read material. She reports significantly higher syllable rates for the German reporters as opposed to the Austrian reporters, with the Swiss nested between the two. The duration of pauses, too, exhibited significant differences between the three varieties, with the German speakers exhibiting the fewest pauses per sentence and the Swiss the most. Hahn and Siebenhaar (2016) provide a preliminary study of regional variation in speech rate in German dialects. By examining speech rate in two conditions – normal and fast reading speeds – with speakers from 67 localities (one speaker per locality) across German-speaking Europe, they found a North/South divide in articulation rate. The study is currently being expanded to include 1494 recordings from 192 localities across German-speaking Europe. With a particular focus on Swiss German varieties, Leemann and Siebenhaar (2007, 2010) and Leemann (2012) have found that the dialects of Bern (Northwest), Winterthur (Northeast), Chur (Southeast), and Brig (Southwest) differ in articulation rate in spontaneous speech. Results from ten speakers per dialect revealed that speakers from Winterthur and Brig articulated nearly one syllable per second faster than speakers from Bern (5.8 syll/sec vs. 5.0 syll/sec). Chur German was nestled between the two extremes. Leemann and Siebenhaar (2010) speculate that differences in phrase-final lengthening between the dialects contribute to these findings: Bern German exhibited distinctly longer mean vowel durations in all positions, but especially phrase-finally.

As this review indicates, the factors that affect speech rate are numerous, ranging from regional affiliation to age, gender, speaker, and speaking style (with this list and literature review being far from exhaustive; see Section 4). Some of the studies presented above used comparatively small numbers of homogeneous speakers that may in reality be unrepresentative of the diversity of speakers found in a larger population of a specific language. In the present study we put forth a new paradigm for how prosodic features, including speech rate, can be crowdsourced through smartphone devices. We present how this approach can yield area-covering geographical distributions of linguistic parameters, such as speech rate. We examine regional differences in the articulation rates of nearly 3,000 speakers of Swiss German based on data that was collected through the recording capability of smartphones. In a first proof-of-concept study, Leemann, Kolly and Dellwo (2014) presented a pilot of this new method: data from 100+ Bern city and 200+ Zurich city German speakers, i.e., speakers in two localities, were collected through *Dialäkt Äpp* (henceforth *DÄ*, Leemann & Kolly, 2013). Six recorded words per speaker were analyzed for articulation rate, which was captured using a metric measuring intervals between consecutive vowel onsets in disyllabic words. The shorter the duration, the faster articulation rate; the larger the duration, the slower the articulation rate. Leemann, Kolly and Dellwo (2014) robustly showed that the 300+ speakers of the two dialects differed in speech rate: Bern German speakers articulated significantly more slowly than Zurich German speakers, corroborating previous findings that exhibited the same trend (Leemann, 2007, 2012; Leemann & Siebenhaar, 2010). They further reported differences between men and women, the latter articulating more slowly. In the present contribution we expand this paradigm by applying the interval measure on the full corpus presented in Leemann, Kolly and Dellwo (2014), i.e. about 3,000 speakers from 452 localities. Before the predictions of this study are introduced, we first provide relevant background information on the sociolinguistics of Swiss German and on a number of technicalities concerning crowdsourcing speech parameters with smartphone applications.

1.1. The linguistic situation in Switzerland

Switzerland has four official languages: German, French, Italian, and Romansh (Federal Constitution of 1999, Article 4). However, the term ‘German’ does not adequately represent the dialects of German-speaking Switzerland: the variety of German spoken is referred to as ‘Swiss German’, of which there are approximately 5.1 million speakers (Federal Department of

Statistics, 2016a). 63.3% of the Swiss population indicated German (incl. the dialects) as their primary language (ibid.). Swiss German is comprised of a number of dialects that are not clearly delineated (Lötscher, 1983). Commonly, lay people name the dialects according to the canton or half-canton – administrative subdivisions – in which they are spoken. Figure 1 shows the 26 cantons of Switzerland along with major localities. Twenty-one of the cantons (incl. half cantons) have German as an official language. Abbreviations for the cantons are explained in the leftmost column of Table 1.

@@@INSERT FIGURE 1 HERE@@@

French is spoken in the West, Italian in the South, and Rumansh in the Southeast. Typically, mutual intelligibility is assured due to extensive dialect contact and exposure to the national mass media. A majority (an estimated 60%) of Swiss National Television programs are broadcast in dialect (Siebenhaar & Wyler, 1997). The language use situation is one of diglossia: both Standard German and Swiss German dialects are used (Ferguson, 1959). The Standard variety and a local vernacular coexist and each have their specific domains of application: dialect is the common means of communication and meets high approval in society, to the point that it is viewed as more prestigious than Standard German (Sieber & Sitta, 1986). The use of the standard is largely restricted to writing and reading, as Swiss German does not have a formal writing system. Contemporary technology, such as email and text messages, has fostered the use of written Swiss German, characterized by idiosyncratic orthography (Christen, 2004). The oral use of Standard German is restricted to school, mass media, and – most commonly – public speeches (Lötscher, 1997). Speakers of Swiss German are well aware of regional variation and many dialects are stereotyped: Zurich German, for example, is perceived as fast. Bern German, which has been dubbed Switzerland's most popular regional variety (Schwarzenbach, 1969), is perceived as very slow (Berthele, 2006; Ris, 1992; Werlen, 1978).

1.2. Crowdsourcing speech parameters with smartphone applications

'Crowdsourcing' refers to 'the practice of obtaining needed services, ideas, or content by soliciting contributions from a large group of people and especially from the online community rather than from traditional employees or suppliers' (Merriam-Webster, 2016). The use of crowdsourcing applications for studying linguistic phenomena has received relatively little attention until recently. This is extraordinary, given that iPhone microphones, for example, feature wide frequency responses of 50Hz-20kHz that enable high-quality audio recordings (faberacoustical, 2009). Even a first generation iPhone from 2007 has been shown to prove very useful for speech analysis and enables reliable acoustic measurements – particularly for F1 and F2 (De Decker & Nycz, 2011). A number of smartphone applications are in use or in development for crowdsourcing linguistic data; for instance, Hughes et al. (2010) and de Vries et al. (2014) developed Android applications as a means to collect speech for the training of acoustic models, and *Ma! Iwaidja* (de Moiser 2016) is an application in development for the purpose of documenting endangered languages. Such apps are primarily used for acoustic modeling, dictionary building, text collection, translation, and dialect mapping. With *DÄ* we present a novel method for crowdsourcing acoustic data to conduct large-scale spatial analyses.

The objective of the present study is to examine regional differences in articulation rate in Swiss German. Given our development of a new paradigm for collecting and analyzing speech data for these purposes, a large bulk of the discussion will be dedicated to addressing methodological caveats. Going into the study, we expected the geographical patterns emergent from our data to corroborate previously reported evidence on speech rate variation in Swiss German, though in this case such patterns would be be shown in significantly higher spatial resolution (452 localities) and with a much larger and a more representative dataset (nearly 3,000 speakers).

2. METHODS

2.1. iOS application: 'Dialäkt Äpp'

The data for the present paper was crowdsourced through *DÄ* (Leemann & Kolly, 2013). *DÄ* has two core functionalities: on the one hand, users can localize their Swiss German dialect by indicating their dialectal pronunciation of 16 words – that is, listening to pre-recorded items and then tapping on the screen to select one. The app then tells users which dialects they most likely speak. Secondly, users can anonymously record these 16 words in their dialect, re-listen to these recordings, and listen to the recordings of other users by navigating an interactive map. Data used in the current study stem from this second functionality. In this recording function, the user interface prompts speakers first to indicate, i.e. self-declare, their dialect, out of 550 possible localities mirroring those used in the *Sprachatlas der Deutschen Schweiz* (1962–2003) (henceforth referred to as *Atlas*, 1962–2003) as well as their age and gender (Figure 2, left panel), before proceeding to the recording instructions (Figure 2, right panel).

@@@INSERT FIGURE 2 HERE@@@

The right panel in Figure 2 reads: 'Please record your voice in as quiet an environment as possible. Keep an approximate distance of about 15 cm between your device and your lips. Please articulate the text loudly and clearly in your own dialectal pronunciation'. Users then record the tokens shown on the screen (see Figure 3, left panel). Recordings are anonymously uploaded to servers where each audio file is given a unique ID. It is explained that in recording their voices, the users consent to providing their acoustic data and information about their dialectal origin, gender, and age (Figure 2, left panel). None of these pieces of information individually or in combination allow for the identification of a user. Users also have the opportunity to opt out of this procedure at any time (i.e. by clicking on 'back' in the top left corner in Figure 2, left panel). This procedure for collecting and analyzing anonymous user data conforms to the regulations of the Zurich cantonal ethics committee (http://www.kek.zh.ch/internet/gesundheitsdirektion/kek/de/home.html) and the accompanying federal laws on experimentation on humans in Switzerland (http://www.admin.ch/opc/de/classified-compilation/20061313/index.html). Once their recording has been uploaded, users can navigate to an interactive map of Switzerland (Figure 3, central panel) where they can listen to their own recordings and those of other users (Figure 3, right panel).

@@@INSERT FIGURE 3 HERE@@@

In Switzerland, *DÄ* became the number one downloaded free app for iPhone after its release on March 22, 2013 (App Annie, 2013). It received major media attention and so far has >91,000 downloads. More than 3,000 users from all over German-speaking Switzerland have uploaded voice recordings.

2.2. Material

For this study we selected six out of a total of 15 recorded *DÄ* words (see Section 2.1) for analysis of speech rate. Each token consisted of two syllables, given that we were measuring the temporal distance between adjacent vowel onsets (see Section 2.7). Half of our selected words featured phonologically long vowels or diphthongs (as defined by their Middle High German roots) while the other half featured historically short vowels. This distribution of long and short vowels holds for nearly all of the 452 localities examined in the present study; only in the word *fragen* does the *Atlas* (1962–2003, variable 140.10) document underlyingly short vowels in four localities in the canton of Solothurn and four localities in the canton of Bern. The selected words with underlying long vowels were *Abend* ‘evening’, *Augen* ‘eyes’, and *fragen* ‘to ask’; those with short vowels were *Donnerstag* ‘Thursday’, *heben* ‘to lift’, and *trinken* ‘to drink’. Typical articulations of these words in two dialects, Bern and Zurich German, are as follows:

Long vowels/diphthongs:

Abend: Bern German: [ˈɑːb̥ə]; play audio file #1 here; Zurich German: [ˈɒːb̥ig̥]; play audio file #2 here.

Augen: Bern German: [ˈɔug̥ə]; play audio file #3 here; Zurich German: [ˈæug̥ə]; play audio file #4 here.

fragen: Bern German: [ˈfrɑːg̥ə]; play audio file #5 here; Zurich German: [ˈfrœːg̥ə]; play audio file #6 here.

Short vowels:

Donnerstag: Bern German: [ˈd̥ɔnʃt̥i]; play audio file #7 here; Zurich German: [ˈd̥unʃt̥ig]; play audio file #8 here.

heben: Bern German: [ˈlʏp͡fə]; play audio file #9 here; Zurich German: [ˈlup͡fə]; play audio file #10 here.

trinken: Bern German: [ˈtrɪŋk͡xə]; play audio file #11 here; Zurich German: [ˈtriŋk͡xə]; play audio file #12 here.

The vast majority of recordings were usable, i.e. demonstrated little background noise interference. Instances of unfavorable audio quality or otherwise unusable material were discarded from the analysis (with approximately 5–10% discarded tokens, cf. Leemann et al., 2015). The word *heben* (‘to lift’) has much fewer recordings than the other five words (cf. Table 1) because the prompt was frequently misinterpreted as meaning ‘to hold’ and thus articulated as [ˈhɛb̥ə], whereas we intended to elicit [ˈlup͡fə], i.e. ‘to lift up’. *lupfen*, however, is not a word in Standard German, which is why would we would not have been able to use *lupfen* as a prompt.

2.3. Localities

Users who submitted the information described in Section 2.1 served as subjects. Figure 4 shows the total of 452 localities represented in the current study. The map only shows German-speaking Switzerland. Each locality is illustrated with a yellow dot, with *Bing areal* used as a base map on QGIS (QGIS, 2016).

@@@INSERT FIGURE 4 HERE@@@

The white lines in Figure 4 denote canton borders. The net of localities is quite evenly spread across German-speaking Switzerland. Alpine areas, e.g. the Southeast and Southwest, are not as densely captured as the Midland areas (in particular the cantons of Bern, Aargau, and Zurich), reflecting their lower population densities.

2.4. Speakers

The number of recordings analyzed varied for each of the six words, since not all speakers read all of the presented words. Table 1 shows the number of recordings for each word by canton as well as the percentage for each word by canton; this sums to 100% for each canton from left to right.

@@@INSERT TABLE 1 HERE@@@

The number of recordings ranged from 1,820 for *heben* ‘to lift’ to 3,149 for *Donnerstag* ‘Thursday’. Relative proportions of words varied between the cantons (X^2=265, $p<.0001$***): the word with the highest number of tokens in Aargau, for example, was *trinken* (N=576, 23.25%), while in Bern this was *fragen* (N=545, 18.54%). Altogether, the database consists of 17,260 audio recordings. Figure 5 shows the number of recordings, pooled for all words, for each canton – each dot represents one locality; Figure 6 shows the number of recordings for each of the 452 localities. The categorizations in Figures 5 and 6 are based on five and ten natural breaks (Jenks). Polygons in Figure 6 are Voronoi polygons (10 buffer). Layers that demarcate the German-speaking area of Switzerland were retrieved from the Federal Department of Statistics (Federal Department of Statistics, 2016b). The canton of Zurich by far shows the most recordings (N=3,845), followed by the cantons of Bern (N=2,939) and Aargau (N=2,477). This is unsurprising, as the major hubs of Zurich, Bern, Aarau, and Baden have some of the highest populations in German-speaking Switzerland. Central Swiss cantons like Uri (N=89) or Obwalden (N=128) have comparatively few respondents. Overall, there is an evident bias toward urban, Midland regions with more respondents than Alpine or central Swiss localities.

@@@INSERT FIGURE 5 HERE@@@

@@@INSERT FIGURE 6 HERE@@@

2.5. Gender

Overall, 48.55% of the recordings were from women and 51.45% from men. Table 2 shows the distribution of recordings by gender and canton.

@@@INSERT TABLE 2 HERE@@@

The ratio of women to men differed between the cantons (X^2=218, p<.0001***). Uri, for example, supplied 21.35% recordings from women and 78.65% from men. Aargau, on the other hand, demonstrate the opposite trend with 58.26% of recordings from women and 41.74% from men. Figure 7 shows the relative percentage of recordings for each canton as % female. The five categories are based on equidistant breaks. The majority of cantons feature a percentage of 43.5–50.9% female speakers. Aargau, the Northeast, and some central Swiss localities exhibit disproportionate numbers of female respondents (dark red). Uri and Nidwalden in Central Switzerland demonstrate a lower proportion of female speakers (yellow or bright orange).

@@@INSERT FIGURE 7 HERE@@@

2.6. Age

On average, speakers were 31.4 years old (SD=15.97); the median age was 28, with a range from 5 to 89 (upper quartile: 43; lower quartile: 18). Datasets with age indications below 5 or above 89 were not included; it is likely that these speakers were not felicitous when answering the questionnaire. Figure 8 shows the distribution of ages.

@@@INSERT FIGURE 8 HERE@@@

The relatively low age mean and median were expected, given that apps particularly target a younger audience.

2.7. Procedure

Traditionally, speech rate has been assessed by measuring a given linguistic unit per second (such as words, syllables, segments, consonantal intervals, and vocalic intervals; cf. Roach, 1998). Since the *DÄ* corpus contains words that exhibit cross-dialectal differences in syllable structure (e.g. *Abend*: Bern German V.CV [ˈɑːb̥ə] vs. Zurich German V.CVC [ˈɒːb̥ig̊], for example, or *Donnerstag*: Bern German CVC.CCV [ˈd̥ɔnʃt̥i], vs. Zurich German CVC.CCVC [ˈd̥unʃt̥ig]), we refrained from applying conventional speech rate measures such as number of syllables per second. Instead, we measured the temporal duration between the two vowel onsets in these disyllabic words. P-Center theory motivates this choice of measure: Allen (1972) examined the perceived location of stress in spontaneous speech by having subjects assess where in the signal they perceived rhythmic beats. He found that stress beats were closely associated with the onsets of vowels in stressed syllables, thus representing perceptually prominent syllable centers. What Allen (1972) called ‘stress beats’, Morton, Marcus and Frankish (1976) deemed ‘P-centers’ – psychological moments of occurrence of a word. Both Allen (1972) and Morton, Marcus and Frankish (1976) found that the perceived beat of a syllable was located near the onset of voicing of the vowel in the stressed syllable. However, perception of the beat is pushed earlier as a function of the length of the preceding onset consonant cluster (the longer the cluster, the earlier the percept of the beat), and later as a function of the length of the syllable coda (Cummins & Port, 1998). The measure applied here – the temporal distance between the vowel onsets in disyllabic words – we call *durVonVon* (cf. Leemann, Kolly & Dellwo, 2014), *duration of vowel-onset-to-vowel-onset*. Ideally, all of the six words used would have featured the same consonants in the first syllable onset and the same consonants in the second syllable coda. For practical reasons, this was not realistic. Figure 9 shows a schematic of the measurement technique applied in the present study.

@@@INSERT FIGURE 9 HERE@@@

Figure 9 shows the oscillogram of a Bern German speaker articulating *fragen* as [ˈfrɑːɡə] (see 1st tier – play audio file #13 here). The 2nd tier shows the boundaries placed at the vowel onsets, and ‘sil’ indicates silence. The shorter *durVonVon*, the faster the temporal progression between the two syllabic beats; the longer the measure, the slower the rate. Altogether there were 17,260 measurement points (34,520 boundaries). *durVonVon* was measured in Praat (Boersma & Weenink, 2016). All labeling was carried out manually by trained phoneticians.

Data were analyzed using R (R Core Team, 2016) and the R packages *lme4* (Bates & Maechler, 2009) and *languageR* (Baayen, 2008, 2009). Unless indicated otherwise, we analyzed data using linear mixed effect models. Normality was checked by visual inspection of quantile plots. Effects were tested by model comparison between a full model, in which the factor in question is entered as an effect, and a reduced model without this effect. We obtained p-values by comparing the results from the two models using standard ANOVAs. Only p-values that are considered significant at the α=0.05 level are reported. We used QGIS (QGIS, 2016) for the spatial visualization of speech rate and count data.

3. RESULTS

Section 3.1 presents findings on the individual words, 3.2 on effects of gender, 3.3 on age, 3.4 on the cantons, and 3.5 on individual localities.

3.1. Word

We ran a linear model that tested for the effect of WORD (six levels). The models were computed as model_full<-lm(durvonvon ~ gender + canton + age + word, data=dataX); model_reduced<-lm(durvonvon ~ gender + canton + age, data=dataX). The differences between the two models, as established by a simple ANOVA comparing the two models, was significant (p<.0001***). The lowest estimated *durVonVon* (as retrieved from the model output of the full model) was found in *fragen* (estimate=0.3305, SE=0.0018), followed by a higher score for *Abend* (this was the default level, which is why we report the raw mean and standard deviation – raw mean=0.3427, SD=0.0787), *Augen* (estimated mean=0.370, SE=0.0019), *heben* (estimated mean=0.374, SE=0.0021), *trinken* (estimate=0.408, SE=0.002), and *Donnerstag* (estimate=0.45, SE=0.0018). Table 3 shows the coefficients of the linear model, using WORD as a fixed factor.

@@@INSERT TABLE 3 HERE@@@

3.2. Gender

We then tested for an effect of GENDER (two levels), using the model specified here: model_full<-lmer(durvonvon ~ gender + canton + age + (1|word), data=dataX); model_reduced<-lmer(durvonvon ~ canton + age + (1|word), data=dataX); i.e. with WORD as a random effect. An ANOVA revealed significant differences between the two models. The estimated mean for men was 0.374 (SE=0.001105), the raw mean of women 0.389 (SD=0.0863). Women overall spoke significantly more slowly than men (see Table 3).

3.3 Age

We tested for an effect of AGE with the same full model used in 3.2. The comparison to the reduced model (which excludes AGE as a fixed effect) revealed a significant effect (p=.00315**). The scatterplot shown in Figure 10 shows *durVonVon* as a function of AGE; Table 3 shows the coefficient obtained for the factor AGE.

@@@INSERT FIGURE 10 HERE@@@

The regression line in Figure 10 suggest a marginal linear relationship between the two variables: the older the speaker, the lower *durVonVon*, i.e. the slower the speech rate.

3.4 Canton

To test for an effect of CANTON (21 levels), we ran the same model used in 3.2. The comparison to the reduced model (excluding CANTON as a fixed effect) was highly significant (p<.0001***). We interpret this as evidence of distinct variation in articulation rate between the cantons. Table 4 shows the estimated *durVonVon* scores and standard errors for each canton in descending order, as retrieved from the full model output. We used the values of the canton of Lucerne as the default level, as it exhibits a nearly balanced 50%/50% distribution of males and females in the dataset (see Table 2). Therefore, the values for Lucerne in Table 4 are the raw mean *durVonVon* and standard deviation (marked with '*' in Table 4).

@@@INSERT TABLE 4 HERE@@@

The canton of Bern exhibits the lowest speech rates (as indicated by the high estimated *durVonVon* scores), followed by Obwalden, Uri, Glarus, and Schwyz. At the other end of the spectrum is the canton of Zurich with the lowest scores (i.e. fastest speech rates), followed by Basel-Landschaft, Valais, Schaffhausen, and Basel Stadt – as well as much of Eastern Switzerland. Figure 11 shows the overall distribution of *durVonVon* by German-speaking cantons (averaged over all recordings for each canton), using seven natural breaks (Jenks). The means displayed are the estimated means shown in Table 4 (adjusted for gender, age, and word effects). The bluer the canton, the faster the articulation rate; the greener the canton the slower the articulation rate.

@@@INSERT FIGURE 11 HERE@@@

3.5 Locality

In this section we turn to more localized articulation rate differences in Swiss German dialects. Because AGE, GENDER, and WORD are all confounds in our analyses of the geographical distribution of articulation rate, we again present estimated *durVonVon* values for the localities. These values are adjusted for all the effects in the model, retrieved from model_full<- lmer(durVonVon ~ gender + city + age + (1|word), data=dataX). Figure 12 shows the estimated *durVonVon* scores for each of the 452 localities, thus depicting more fine-grained differentiation and spatial resolution for the individual localities than Figure 11. We used Voronoi polygons (10 buffer), showing eight natural breaks (Jenks). We applied a nearest neighbor normalization that averages each estimated *durVonVon* with the ten geographically nearest neighbors (Blaxter, 2016). Cantons are shown to differ as to the homogeneity of rates in the individual localities: in the canton of Zurich, for example, it seems that the localities examined uniformly exhibit similarly fast articulation rates, as indicated by virtually all blue-colored polygons. In the western canton of Aargau, however, we find a hybrid, with the East articulating much faster than the West. The canton of Bern seems to be relatively homogeneous as well, though not as homogenous as the canton of Zurich; the former, however, spans a much vaster area (5,959 km^2, vs. 1,729 km^2 in the canton of Zurich). To show these local patterns in more detail, Figures 13 and 14 display the cantons of Zurich, Aargau, and Bern.

@@@INSERT FIGURE 12 HERE@@@

@@@INSERT FIGURE 13 HERE@@@

@@@INSERT FIGURE 14 HERE@@@

Figure 13 reveals relatively fast articulation rates for Zurich across all the major sub-dialect regions in Zurich (cf. Weber & Dieth, 1987): this includes city dialect speakers, including Zurich city, Rümlang, Utikon, Stallikon, Zumikon, and Bassersdorf, as well as the Unterländer dialects (Windlach, Glattfelden, Niederweningen) and Oberländer dialects (Pfäffikon, Bäretswil, Fischenthal). The dialect regions around Winterthur also exhibit relatively fast articulation rates. There are a few exceptions to this pattern: Neftenbach, Uster, and Maur exhibit somewhat slower articulation rates. When looking at the adjacent canton, Aargau, it is noticeable that southwestern localities (e.g. Oftringen, Zofingen, Safenwil, Brittnau) tend to have slower articulation rates than, for example, places in the center or or east of the canton such as Birr, Niederrohrdorf, and Würenlos. The geographical patterning of speech rate is much more heterogeneous than in Zurich. Figure 14 reveals a less heterogenous pattern for the canton of Bern as well, though speech rate seems to be somewhat more coherently distributed across the localities than in Aargau: for localities such as Blumenstein, Ried, Thun, and Seftigen – towards the Bernese Oberland – we find very slow articulation rates. Localities such as Köniz, Rosshäusern, or Grindelwald reveal somewhat faster articulation rates. The northwest part of the canton in general – including Bern – seems to exhibit somewhat slower rates than in the northeast or the southwest. Overall, however, slower articulation rates are clearly dominant for this canton. To more closely investigate individual cities, Figure 15 and Table 5 show the estimated *durVonVon* as well as standard errors for major cities in German-speaking Switzerland. For this data, no nearest neighbor smoothing was applied. Only cities for which there were at least 100 recordings are shown; Brig-Glis is included to represent Southwestern Switzerland despite having only 68 recordings in total. We used the values of Aarau as the default level, which is why we report raw means and standard deviations in Table 5, not estimated means and the standard errors (marked with '*'). The estimated means are shown in descending order, i.e. cities with the lowest articulation rates appear on top and those with highest rates at the bottom.

@@@INSERT FIGURE 15 AND TABLE 5 HERE@@@

The slowest articulation rates in cities with more than 100 recordings are found in Langenthal (Bern), Burgdorf (Bern), Thun (Bern), and Solothurn (Solothurn), all of which exhibited estimated means larger than in Bern (Bern) – i.e. with slower rates than the city of Bern. The fastest articulation rates in major cities show up in Wil (St. Gallen), Meilen (Zurich), Zurich (Zurich), and Uster (Zurich). All of these localities are in the canton of Zurich or St. Gallen. Chur (Graubünden), in Southeastern Switzerland, is found to articulate neither especially slow or especially fast. Brig-Glis (Valais) seems to have rather fast speech, but again does not lie at either extreme.

4. DISCUSSION

Based on a controlled set of words spoken by a large number of speakers, the current study found distinct differences in articulation rate for Swiss German dialects. The findings also revealed an effect of WORD, GENDER, and AGE. Here we discuss the findings in the same sequence as presented in Section 3. Because we used a new methodological paradigm of collecting speech as a basis for analyzing the geographical distribution of articulation rate, a large bulk of this discussion is devoted to critical reflection on this new methodology.

4.1. Word, gender, and age

In Section 3.1, we presented the effect of WORD; the highest estimated means of *durVonVon* were found for *Donnerstag*, followed by *trinken*, *heben*, *Augen*, *Abend*, and *fragen*. This effect was expected, as the number of consonants differed between the measured interval durations. Typical realizations of the six words in question are shown below for Bern German, in descending order of *durVonVon*. We also show the syllable structure of these prototypical realizations; affricates and diphthongs are counted as two segments each. On the very right we present the number of segments over which *durVonVon* – the interval duration of vowel onsets in disyllabic words – spans.

Donnerstag [ˈd̥ɔnʃt̥i] – CVC.CCV – 4 segments
trinken [ˈtrɪŋk͡xə] – CCVC.CCV – 4 segments
heben [ˈlʏp͡fə] – CV.CCV – 3 segments
Augen [ˈɔug̊ə] – VV.CV – 3 segments
Abend [ˈɑːb̥ə] – VV.CV – 3 segments
fragen [ˈfrɑːg̊ə] – CCVV.CV – 3 segments

VC.CC sequences resulted in the highest scores of *durVonVon*, followed by V.CC and VV.C. This adds to the growing body of literature that shows the sensitivity of speech rate not only to different styles (such as in read vs. spontaneous speech, with read speech frequently showing slower rates given fewer reductions; Crystal & House, 1982; Jacewicsz et al., 2009) but also to speech material. Similarly, Quené (2008) has shown that articulation rate is strongly affected by the length of the phrase in Dutch; speakers shorten syllables when they anticipate more syllables in a phrase, a process known as anticipatory shortening. Because WORD was a confounding factor when examining regional difference in articulation rate, we used estimated means when plotting regional patterns in Sections 3.4 and 3.5.

Our results further revealed a robust effect of GENDER, with women overall exhibiting lower articulation rates than men. A number of studies have shown similar trends in British English (Whiteside, 1996) and American English dialects (Byrd, 1994; Jacewicz et al., 2009). Simpson (1998) has also demonstrated that in German, the durations of women's vowels are systematically longer than men's. This finding may be connected to the fact that women tend to be more preoccupied with status, making greater use of social capital mediated through the manipulation of socially marked features of speech (Trudgill, 1972). Women may be more likely to attempt to avoid speaking quickly because they do not view speaking quickly as having overt prestige. Our results in Section 3 further revealed an effect of AGE, with a relatively weak trend of increased age suggesting a lower *durVonVon* score, i.e. speaking faster. This weak effect may be an artifact of the large data set we are using (cf. Kilgarriff, 2005); it is not consistent with previous findings, which have shown that elderly speakers tend to articulate more slowly than younger speakers (Quené, 2008; Yuan, Cieri & Liberman, 2006). Given that both GENDER and

AGE revealed significant effects in the models calculated, we adjusted the means in the geographical distribution of articulation rate accordingly in Figures 11–15.

4.2. Regional differences

Section 3 revealed distinct regional differences in articulation rate at both the canton and individual locality scale. Given that *durVonVon* is a rather abstract measure of articulation rate, let us extrapolate these results to a more realistic scenario for illustrative purposes: say, having speakers from Bern, Zurich, Valais, Obwalden, and Basel-Landschaft German each read Aesop's fable 'The North Wind and the Sun'. The fable consists of 129 syllables, i.e. 128 vowel-onset-to-vowel-onset intervals (if we take Zurich German as an example, cf. Fleischer & Schmid, 2006). Based on the findings of the current study (as shown for each canton in Table 4), disregarding contextual factors such as phrase-final lengthening, differences in style, between-speaker differences in reading fluencies, pauses and so on, the Bern German speaker should take 53.15 seconds to read the text, the Obwalden speaker 52.9 seconds, the Valais speaker 46.18 seconds, Basel-Landschaft 46.01 seconds, and Zurich only 45.69 seconds. The findings of the current study reveal that the temporal information contained in a few isolated words alone is sufficient to tell apart the dialects on a cantonal level.

These findings, however, need some form of quality control given the noisy nature of the data. To scrutinize the validity of these results, we can compare them to those of previous studies that have examined speech rate across a few Swiss German localities. As mentioned earlier, based on the analyses of the spontaneous speech of ten speakers per locality, Leemann & Siebenhaar (2007, 2010) and Leemann (2012) reported that speakers from Winterthur (Zurich) articulated nearly one syllable more per second than speakers from Bern city (5.8 syll./sec. vs.

5.0 syll./sec). They further found that speakers of Brig (Valais) articulated as quickly as Zurich German speakers (5.8 syll./sec), with speakers from Chur (Graubünden) between these two extremes (5.2 syll./sec). Our current data largely mirrors this trend for these four cantons. For Graubünden, the present study does not reveal as slow articulation rates as reported in the previous studies; however, overall, speech rates captured by the two methods thus seems to be robustly related, underscoring the validity of the crowdsourcing method applied.

The strength of the current study is its very high, indeed unprecedented, spatial resolution for the distribution of articulation rate across 452 localities. In sum, our results demonstrate three previously unreported trends in the distribution of speech rates in Swiss German:

(a) An East/West divide: Eastern Swiss German dialects have quite rapid articulation rates, from the North (Schaffhausen) down to Zurich, Thurgau, and St. Gallen; speakers in St. Gallen demonstrate slower rates the further south in the canton. The southeastern canton Graubünden overall exhibits somewhat slower articulation rates long with much of the West; Basel and its urban areas as well as the western part of Valais exhibit the fastest articulation rates in Western Switzerland.

(b) Aargau – lying in the crossover zone between the East and the West – is clearly split into localities that speak more quickly in the East (bordering the canton of Zurich) and localities in the West, particularly in the Southwest, that speak much more slowly (bordering the canton of Bern).

(c) Not only Bern German exhibits particularly slow speech rates, but also speech in a large number of central Swiss localities in Obwalden, Uri, Glarus, and Schwyz.

The East/West divide found in speech rate in the current dataset aligns with other linguistic isoglosses separating the East from the West: phonetically, this divide is reflected in the degree of opening in vowels – with more closed vowels in the East, e.g. [oː] (e.g. ‘bread’ Zurich German [b̥roːt], Bern German [b̥rɔːt]), [ɛ] (‘to die’ Zurich German [ʃtɛrb̥ə], Bern German [ʃtærb̥ə]), and [e] (‘bed’ Zurich German [b̥et], Bern German [b̥ɛt]). The divide is also apparent in morphological features: plural forms of verbs, for example, pattern homogeneously in the East, while Western Swiss German dialects use a more differentiated plural system (Siebenhaar & Wyler, 1997). This East/West contrast is frequently referred to as the Brünig-Napf-Reuss line, reflecting not only linguistic but also cultural contrasts such as the use of different decks of playing cards (French versions in the West, German in the East) (Weiss, 1947). The line cuts through the canton of Aargau, almost exactly as found in our data. These differences between the two dialect regions show the influence of the past 200–500 years of political history, especially regarding the political and linguistic influence of Bern and Zurich on their hinterlands (Lötscher, 1983). Siebenhaar (2000) has robustly shown that Aarau, the capital city of Aargau, uses phonetic features that are typical for Eastern and Western Swiss German dialects.

More generally, it has been claimed that speech rate may have urban/rural correlates (cf. Hewlett & Rendall, 1998; Jacewicz et al., 2009). Hewelett and Rendall (1998) did not find a robust effect of speech rate between their urban (Edinburgh) vs. rural (Orkney Islands) speakers, failing to find evidence for the usual stereotype of faster urban speakers and slower rural speakers. In the present data, we also do not find an urban vs. rural effect: even rural areas like so many in Valais in the Southwest and many Eastern Swiss German localities articulate comparatively fast. The urban area around Bern, in turn, has quite slow speakers. Figure 15 demonstrates how urban centers can range from very slow (e.g. Langenthal, Burgdorf, and Thun) to very quick articulation (as in Zurich, Meilen, and Wil).

As for potential linguistic explanations for the differences in speech rate, Leemann & Siebenhaar (2010) speculated that the Bern/Zurich contrast may lie in differences in how speakers behave in terms of phrase-final lengthening; they report Bern German speakers to produce distinctly longer mean duration of vowels particularly in phrase-final position. However, overall mean vowel durations were longer in Bern German as well. Potentially, in reading isolated words out loud, speakers may have been treating the utterances as if they were isolated sentences. In the future, more linguistically-driven analyses might better establish why the dialects pattern geographically the way they do. Kohler (2001) notes, for example that the realization particularly of unstressed syllables seems to heavily contribute to regional variation.

4.3. Methodological caveats

Given the novelty of our paradigm for crowdsourcing regional articulation rate distributions, a number of methodological issues merit further discussion. Both the use of *durVonVon* as a metric as well as the use of isolated words as speech material must be discussed further. In addition, the user's self-declared dialect serves as the basis for the analysis of speech rate, and this is not without problems. Finally, some general methodological concerns in using mobile applications to crowdsource speech data deserve further mention.

durVonVon is a metric that – to our knowledge – has not been used before except for the proof-of-concept study conducted by us (Leemann, Kolly & Dellwo 2014). In this sense, comparisons to previous studies such as Leemann and Siebenhaar (2007, 2010) and Leemann (2012) need to be taken with a grain of salt. The latter used the conventional measure of syllables per second to quantify articulation rate despite the confound of syllable structures differing between the Swiss German dialects examined (Keller, 2008). The *durVonVon* metric provides a new approximation of speech rate in the examination of between-dialect differences, and is especially useful if the dialects exhibit the same syllable structures for the words in question. For the six words we examined, this is the case in the vast majority of dialects. There are a few exceptions, however: in the word *trinken*, Bernese Oberland, Southern Fribourgeois localities, the Valais, and a number of localities in the southeast of the canton of Graubünden would not use [ˈtrɪŋk͡xə] (with dialect-specific alternations of vowel quality) but rather [ˈtrɛːxə] (again, with dialect-specific alternations of the vowel). The elision of nasals before homorganic fricatives and the lengthening or diphthongization of the preceding vowel is referred to as Staub's Law (Staub, 1874). This phonological process triggers a change in syllable structure, which affects the Cs and Vs captured by *durVonVon*: *trinken* without Staub's law applied would be [ˈtrɪŋk͡xə] – CCVC.CCV – with 4 segments, while *trinken* with Staub's Law applied would be [ˈtrɛːxə] – CCVV.CV – with 3 segments; the underlined Cs and Vs would be captured by *durVonVon* (if affricates and diphthongs are again counted as two segments each). In a pilot study, Werlen (2012) examined the effect of Staub's Law on the temporal structure of words in eight Swiss German dialects. He reported that the relative duration of the vowel + nasal (Staub's Law not applied) as a percentage of the word duration, i.e. normalized for speech rate, was not significantly different in duration from the lengthened monophthong/diphthong (Staub's Law applied) as a percentage of the word duration. Not only is the use of *durVonVon* new, but also using isolated words as a basis for measurement of articulation rate has – to our knowledge –

never before been done. Using isolated words is unusual in the sense that normally, analyses of speech rate involve spontaneous speech and/or read speech (see literature review). Reading isolated words is, in a sense, situated between reading and spontaneously speaking: the word appears on the screen and triggers an immediate articulation which does not involve a significant amount of planning or cognitive performance as is required in reading. Traditional dialectology, though, has long relied on the observation of single, isolated words in lists.

A second issue is the user's self-declaration of dialect. As mentioned in Section 2.1 (Figure 2, left panel), users manually indicated their age, gender, and dialect before recording the audio. This self-declaration of dialect forces us to assume that users have an understanding of their own linguistic origins (Leemann et al., 2016). It is conceivable, however, that users tried to imitate a more 'model' target dialect when doing the audio recordings, perhaps because they felt it would sound more prestigious. This, in turn, would cause speakers to be more homogeneous than is really the case. Labov (1996) has shown that lay speakers have relatively poor intuitions about a number of aspects of their own non-standard dialect use, however. The results presented here need to be interpreted against the background of this limitation.

There are some more general limitations to using crowdsourced audio data that must be further addressed. For one, speakers may have submitted audio recordings multiple times (Birnbaum, 2004): app-based research allows for multiple submissions, and we as researchers do not know whether a speaker used the app to participate repeatedly or used the app on multiple smartphones. Reips (2002) found, though, that the rate of repeated participation does not seem to pose a relevant threat to the trustworthiness of app-based research, given that it is <3% in the studies investigated. Secondly, a sampling bias underlies the results of the current study: only iOS users were able to participate. At the time of the launch of *DÄ* (2013), running on iOS only was deemed most useful as the vast majority of smartphone users were iOS users. Targeting these users only, of course, means neglecting a potentially different social substratum who favor devices running Android or other platforms. Thirdly, the audio recordings collected may feature a response bias in that the order in which the words were presented was not randomized; the word *trinken* appeared last in the 15 words presented to users. Since users are essentially reading off a list of words (with each word shown on the screen in isolation), the dataset could possibly feature particularly distinct vocalic lengthening of the /i/ of *trinken*. Finally, the trustworthiness of participants in web or app-based crowdsourcing is often cited as a problem (McGraw, 2013). How are researchers to know that they are being given meaningful responses from the crowd? There are, however, clear benefits that – in our opinion – balance out the limitations. For instance, subject recruitment involves very low costs; conducting this experiment with paid researchers going to 452 localities to elicit several speakers per locality would have been very expensive.

5. CONCLUSIONS

This study set out to determine articulation rate differences between dialects using a large, crowdsourced audio database of roughly 3,000 speakers. The most obvious finding to emerge is that Swiss German dialects exhibit distinct regional patterns in speech rate. Previous studies examining only a handful of the dialects scrutinized here validate our results. The results of this research thus support the idea that smartphone apps enabling audio recordings can provide a valid alternative for collecting audio data. The principal strength of this approach is its high spatial resolution; no other study has shown this degree of precision, using hundreds of localities, to examine a prosodic parameter like articulation rate. This parameter represents only one of countless possible speech prosody or phonetic variables for which a crowdsourcing approach can provide new and useful insights. Indeed, the present region-wide *DÄ* corpus enables further analyses of the spatial distribution of formant frequencies, intonation patterns, and speech rhythm properties, to name just a few examples. Such phenomena can be explored multi-dimensionally, enabling us to test for effects of speaker, age, and gender in addition to geographical distribution. Crowdsourcing applications for British English have just been released for iOS and Android (Leemann, Kolly & Britain, 2016) inspired by the *DÄ* framework. Future analyses with this corpus are planned, modeled on those done for *DÄ*.

6. REFERENCES

Allen, G. D. 1972. The location of rhythmic stress beats in English: An experimental study I. *Language and Speech* 15. 72-100.

App Annie. 2013. http://www.appannie.com/ (19 April 2016).

Baayen, R. H. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press

Baayen, R. H. 2009. *languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics using R"*. R package version 0.955.

Bates, D. M. & M. Maechler. 2009. *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-32.

Berthele, R. 2006. Wie sieht das Berndeutsche so ungefähr aus? Über den Nutzen von Visualisierungen für die kognitive Laienlinguistik. In H. Klausmann (ed.), *Raumstrukturen im Alemannischen. Beiträge der 15. Arbeitstagung zur alemannischen Dialektologie, Schloss Hofen (Vorarlberg) vom 19.-21.9.2005*, 163-176. Graz-Feldkirch: Neugebauer.

Birnbaum, M. H. 2004. Human research and data collection via the Internet. *Annual Review of Psychology*. 55. 803-832.

Blaxter, T. 2016. *Geospatial temporal visualisation*. Manuscript draft. Cambridge, UK: University of Cambridge Ph.D. thesis.

Boersma, P. & D. Weenink. 2016. *Praat: doing phonetics by computer*. http://www.praat.org/ (19 April 2016).

Byrd, D. 1992. Preliminary results on speaker-dependent variation in the TIMIT database. *Journal of the Acoustical Society of America* 92(1). 593-596.

Christen, H. 2004. Dialekt-Schreiben oder sorry ech hassä Text schribä. In E. Glaser, P. Ott & R. Schwarzenbach (eds.), *Alemannisch im Sprachvergleich: Beiträge zur 14. Arbeitstagung für alemannische Dialektologie in Männedorf (Zürich) vom 16.–18.9.2002*, 71-87. Stuttgart: Franz Steiner.

Crystal, T. H. & A. S. House. 1982. Segmental durations in connected speech signals: Preliminary results. *Journal of the Acoustical Society of America* 72. 705-716.

Cummins, F. & R. Port. 1998. Rhythmic constraints on stress timing in English. *Journal of Phonetics* 26(2). 145-171.

De Decker, P. & J. Nycz. 2011. For the Record: Which Digital Media Can be Used for Sociophonetic Analysis? *University of Pennsylvania Working Papers in Linguistics* 17(2). 51-59.

de Moiser, M. 2016. Ma! Iwaidja. https://itunes.apple.com/au/app/ma-iwaidja/id557824618?mt=8 (19 April 2016).

faberacoustical. 2009. http://blog.faberacoustical.com/2009/ios/iphone/iphone-microphone-frequency-response-comparison/ (19 April 2016).

Federal Department of Statistics. 2016a. http://www.bfs.admin.ch/bfs/portal/de/index/themen/01/05/blank/key/sprachen.html (19 April 2016).

Federal Department of Statistics. 2016b. http://www.bfs.admin.ch/bfs/portal/de/index/dienstleistungen/geostat/datenbeschreibung/generalisierte_gemeindegrenzen.html (19 April 2016).

Ferguson, C. 1959. Diglossia. *Word* 15. 325-340.

Fleischer, J. & S. Schmid. 2006. Zurich German. *Journal of the International Phonetic Association* 36(2). 243-253.

Hahn, M.& B. Siebenhaar 2016. Sprechtempo und Reduktion im Deutschen (SpuRD). In O. Jokisch (ed.) *Elektronische Sprachsignalverarbeitung 2016*, 198-205. Dresden: TUDpress.

Hewlett, N. & M. Rendall. 1998. Rural versus urban accent as an influence on the rate of speech. *Journal of the International Phonetic Association* 28. 63-71.

Hughes, T., K. Nakajima, L. Ha, A, Vasu, P. Moreno & M. LeBeau. 2010. Building transcribed speech corpora quickly and cheaply for many languages. *Proceedings of Interspeech 2010*. 1914-1917.

Jacewicz, E., R. A. Fox, C. O'Neill & J. Salmons. 2009. Articulation rate across dialect, age, and gender. *Language Variation and Change* 21(2). 233-256.

Jacewicz, E., R. A. Fox & L. Wei. 2010. Between-speaker and within-speaker variation in speech tempo of American English. *The Journal of the Acoustical Society of America* 128(2). 839-850.

Jessen, M. 2007. Forensic reference data on articulation rate in German. *Science & Justice*, 47(2). 50-67.

Keller, K. 2008. „Hützt's z'Zuzwil?" Zu den Silbenstrukturen des Schweizerdeutschen, empirisch analysiert an zwei Dialekten. *Lizentiatsarbeit am Institut für Sprachwissenschaft der Universität Bern 2. Mai 2008.*

Kilgarriff, A. 2005. Language is never, ever, ever, random. *Corpus Linguistics and Linguistic Theory* 1(2). 263-276.

Kohler, K. J., K. Schafer, W. Thon & G. Timmermann. 1981. Sprechgeschwindigkeit in Produktion und Perzeption. *Arbeitsberichte Kiel* 16. 137-205.

Kohler, K. J. 1982. Rhythmus im Deutschen in Experimentelle Untersuchungen von Zeitstrukturen im Deutschen. Le rythme en allemand. *Arbeitsberichte-Institut für Phonetik* 19. 89-105.

Kohler, K. J. 2001. The investigation of connected speech processes. Theory, method, hypotheses and empirical data. *Arbeitsberichte des Instituts für Phonetik der Universität Kiel* 35. 1-32.

Kolly, M.-J., A. Leemann, V. Dellwo & P. Boula de Mareuil 2015. Speaker-idiosynrasy in pausing behavior: evidence from a cross-linguistic study. *Proceedings of ICPhS* (Glasgow).

Künzel, H. J. 1997. Some general phonetic and forensic aspects of speaking tempo. *International Journal of Speech Language and the Law* 4(1). 48-83.

Labov W. 1996. When Intuitions Fail. *Proceedings of the 32nd Regional Meeting of the Chicago Linguistic Society* 32. 77-106.

Leemann, A. 2012. *Swiss German Intonation Patterns*. Amsterdam: Benjamins.

Leemann, A. & M.-J. Kolly. 2013. *Dialäkt Äpp*. https://itunes.apple.com/ch/app/dialakt-app/id606559705?mt=8 (19 April 2016).

Leemann, A., M.-J. Kolly, & D. Britain, D. 2016. *English Dialects*. https://itunes.apple.com/gb/app/english-dialects/id882340404?mt=8, https://play.google.com/store/apps/details?id=ch.uk_regional&hl=en_GB (19 April 2016).

Leemann, A., M.-J. Kolly & V. Dellwo. 2014. Crowdsourcing regional variation in speaking rate through the iOS app 'Dialäkt Äpp'. *Speech Prosody* 7: 217-221.

Leemann, A., M.-J. Kolly, J. P. Goldman, V. Dellwo, I. Hove, I. Almajai & D. Wanitsch. 2015. Voice Äpp: a mobile app for crowdsourcing Swiss German dialect data. *Proceedings of of Interspeech 2015*. 2804-2808.

Leemann, A., M. J. Kolly, R. Purves, D. Britain & E. Glaser 2016. Crowdsourcing Language Change with Smartphone Applications. PloS ONE 11(1). e0143060.

Leemann, A. & B. Siebenhaar. 2007. Intonational and temporal features of Swiss German. *Proceedings of the ICPhS, Saarbrücken*. 957-960.

Leemann, A. & B. Siebenhaar. 2010. Statistical Modeling of F0 and Timing of Swiss German Dialects. *Proceedings of Speech Prosody*.

Lötscher, A. 1983. *Schweizerdeutsch: Geschichte, Dialekt, Gebrauch*. Frauenfeld: Huber.

McGraw, I. 2013. Collecting speech from crowds. In M. Eskenazi, G. A. Levow, H. Meng, G. Parent & D. Suendermann (eds), *Crowdsourcing for speech processing: Applications to data collection, transcription and assessment*, 38-71. Hoboken, NJ: John Wiley & Sons.

Merriam-Webster. 2016. *Crowdsourcing*. http://www.merriam-webster.com/dictionary/crowdsourcing (19 April 2016).

Morton, J., S. Marcus & C. Frankish. 1976. Perceptual centers (P-centers). *Psychological Review* 83. 405-408.

Pfitzinger, H. R. 1998. Local Speech Rate as a Combination of Syllable and Phone Rate. *Proceedings of the ICSLP* 3. 1087-1090.

Pfitzinger, H. R. 1999. Local speech rate perception in German speech. *Proc. of the XIVth Int. Congress of Phonetic Sciences* 2. 893-896.

QGIS. 2016. QGIS *Geographic Information System. Open Source Geospatial Foundation Project.* http://qgis.osgeo.org (19 April 2016).

Quené, H. 2008. Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *Journal of the Acoustical Society of America* 123. 1104-1113.

R Core Team. 2016. *R: A Language and Environment for Statistical Computing*, version 3.0.0. R Foundation for Statistical Computing. http://www.R-project.org (19 April 2016).

Ris, R. 1992. Innerethik der deutschen Schweiz. In P. Hugger (ed.) *Handbuch der schweizerischen Volkskultur*, vol. 2, 749-766. Zürich: Offizin.

Reips, U. D. 2002. Standards for Internet-based experimenting. *Experimental Psychology* 49. 243-256.

Roach, P. 1998. *Myth 18: Some languages are spoken more quickly than others*. In L. Bauer & P. Trudgill (eds.), *Language Myths*, 150-158. London: Penguin.

Robb, M. P., M. A. Maclagan & Y. Chen. 2004. Speaking rates of American and New Zealand varieties of English. *Clinical Linguistics & Phonetics* 18(1). 1-15.

Schwab, S. & M. Avanzi. 2015. Regional variation and articulation rate in French. *Journal of Phonetics* 48. 96-105.

Schwarzenbach, R. 1969. *Die Stellung der Mundart in der deutschsprachigen Schweiz. Studien zum Sprachgebrauch der Gegenwart*. Frauenfeld: Huber.

Sieber, P. & H. Sitta. 1986. *Mundart und Standardsprache als Problem der Schule*. Aarau: Sauerländer.

Siebenhaar, B. 2000. *Sprachvariation, Sprachwandel und Einstellung. Der Dialekt der Stadt Aarau in der Labilitätszone zwischen Zürcher und Berner Mundartraum.* Stuttgart: Franz Steiner Verlag.

Siebenhaar, B. & A. Wyler. 1997. Dialekt und Hochsprache in der deutschsprachigen Schweiz. Zürich: Pro Helvetia.

Siebenhaar, B., B. Zellner Keller & E. Keller. 2001. *Phonetic and timing considerations in a Swiss High German TTS system. Improvements in Speech Synthesis*, 165-175. Chichester: John Wiley.

Simpson, A. P. 1998. Phonetische Datenbanken des Deutschen in der empirischen Sprachforschung und der phonologischen Theoriebildung. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel* 33.

Sprachatlas der deutschen Schweiz. 1962–2003. Bern (I-VI), Basel (VII-VIII): Francke.

Staub, F. 1874. Die Vokalisierung des N bei den schweizerischen Alemannen. Halle 1874 (unter dem Titel: Ein schweizerisch-alemannisches Lautgesetz auch in Deutsche Mundarten 7, 1877, S. 18–36, 191–207, 333–389).

Trouvain, J. 2003. *Tempo variation in speech production. Implications for speech synthesis*. Saarbrücken, Germany: Universität des Saarlandes Ph.D. thesis.

Trouvain, J., J. Koreman, A. Erriquez & B. Braun. 2001. Articulation Rate Measures and Their Relation to Phone Classification in Spontaneous and Read German Speech. *Proceedings of the ISCA Workshop on Adaptation Methods for Speech Recognition 2001*. 155-158.

Trudgill, P. 1972. Sex, Covert Prestige and Linguistic Change in the Urban British English of Norwich. *Language in Society* 1(2). 179-195.

Ulbrich, C. 2005. *Phonetische Untersuchungen zur Prosodie der Standardvarietäten des Deutschen in der Bundesrepublik Deutschland, in der Schweiz und in Österreich.* Frankfurt: Peter Lang.

Verhoeven, J., G. De Pauw & H. Kloots. 2004. Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands. *Language and Speech* 47(3). 297-308.

de Vries, N., M. H. Davel, J. Badenhorst, W. D. Basson, F. de Wet, E. Barnard & A. de Waal. 2014. A smartphone-based ASR data collection tool for under-resourced languages. *Speech Communication* 56. 119-131.

Weber, A. & E. Dieth. 1987. *Zürichdeutsche Grammatik: ein Wegweiser zur guten Mundart*, vol. 1. H. Rohr.

Weiss, R. 1947. Die Brünig-Napf-Reuss-Linie als Kulturgrenze zwischen Ost- und Westschweiz auf volksmundlichen Karten. *Geographica Helvetica* 2(3). 153-175.

Werlen, I. 1978. Zur Einschätzung von schweizerdeutschen Dialekten. In I. Werlen (ed.), *Probleme der schweizerdeutschen Dialektologie 2. Kolloquium der Schweizerischen Geisteswissenschaftlichen Gesellschaft*, 195-257. Freiburg: University of Freiburg.

Werlen, I. 2012. *Zu Staub zurückkehren. Oder: warum Hanf nicht Haif ist - Gedanken zum Staubschen Gesetz*. Presentation, University of Bern.

Whiteside, S. P. 1996. Temporal-based acoustic-phonetic patterns in read speech: some evidence for speaker gender differences. *Journal of the International Phonetic Association* 26. 23-40.

Yuan, J., C. Cieri & M. Liberman. 2006. *Towards an integrated understanding of speaking rate in conversation.* Paper presented at the International Conference on Spoken Language Processing (Interspeech 2006), Pittsburgh, PA.

TABLES

Count, Row %	*Abend*	*Augen*	*Donnerstag*	*fragen*	*heben*	*trinken*	Total by canton
Aargau (AG)	250,10.09	537,21.68	557,22.49	289,11.67	268,10.82	576,23.25	2477
Appenzell Innerrhoden (AI)	13,16.46	14,17.72	14,17.72	11,13.92	12,15.19	15,18.99	79
Appenzell Ausserrhoden (AR)	20,16.67	22,18.33	22,18.33	21,17.50	12,10.00	23,19.17	120
Bern (BE)	488,16.60	490,16.67	516,17.56	545,18.54	368,12.52	532,18.10	2939
Basel-Landschaft (BL)	101,17.32	107,18.35	107,18.35	110,18.87	39,6.69	119,20.41	583
Basel-Stadt (BS)	99,16.53	111,18.53	109,18.20	116,19.37	46,7.68	118,19.70	599
Freiburg (FR)	41,16.87	41,16.87	39,16.05	44,18.11	35,14.40	43,17.70	243
Glarus (GL)	36,18.09	35,17.59	38,19.10	38,19.10	16,8.04	36,18.09	199
Graubünden (GR)	90,16.16	95,17.06	98,17.59	110,19.75	51,9.16	113,20.29	557
Luzern (LU)	194,17.91	197,18.19	187,17.27	204,18.84	102,9.42	199,18.37	1083
Nidwalden (NW)	18,16.98	17,16.04	19,17.92	20,18.87	14,13.21	18,16.98	106
Obwalden (OW)	22,17.19	24,18.75	22,17.19	21,16.41	18,14.06	21,16.41	128
St. Gallen (SG)	219,16.99	224,17.38	226,17.53	245,19.01	134,10.40	241,18.70	1289
Schaffhausen (SH)	40,17.86	40,17.86	40,17.86	40,17.86	24,10.71	40,17.86	224
Solothurn (SO)	150,17.63	150,17.63	154,18.10	155,18.21	81,9.52	161,18.92	851
Schwyz (SZ)	85,18.12	78,16.63	87,18.55	81,17.27	46,9.81	92,19.62	469
Thurgau (TG)	116,16.60	123,17.60	125,17.88	126,18.03	69,9.87	140,20.03	699
Uri (UR)	16,17.98	16,17.98	16,17.98	16,17.98	10,11.24	15,16.85	89
Wallis (VS)	56,16.18	57,16.47	56,16.18	76,21.97	42,12.14	59,17.05	346
Zug (ZG)	62,18.51	60,17.91	56,16.72	62,18.51	34,10.15	61,18.21	335
Zürich (ZH)	687,17.87	670,17.43	661,17.19	701,18.23	399,10.38	727,18.91	3845
Total by word	2803	3108	3149	3031	1820	3349	17260

Table 1: Number of recordings for words and cantons.

Count, Row %	*Female*	*Male*	Total by canton
Aargau (AG)	1443,58.26	1034,41.74	2477
Appenzell Innerrhoden (AI)	45,56.96	34,43.04	79
Appenzell Ausserrhoden (AR)	58,48.33	62,51.67	120
Bern (BE)	1277,43.45	1662,56.55	2939
Basel-Landschaft (BL)	292,50.09	291,49.91	583
Basel-Stadt (BS)	296,49.42	303,50.58	599
Freiburg (FR)	96,39.51	147,60.49	243
Glarus (GL)	88,44.22	111,55.78	199
Graubünden (GR)	270,48.47	287,51.53	557
Luzern (LU)	542,50.05	541,49.95	1083
Nidwalden (NW)	38,35.85	68,64.15	106
Obwalden (OW)	58,45.31	70,54.69	128
St. Gallen (SG)	617,47.87	672,52.13	1289
Schaffhausen (SH)	118,52.68	106,47.32	224
Solothurn (SO)	381,44.77	470,55.23	851
Schwyz (SZ)	242,51.60	227,48.40	469
Thurgau (TG)	366,52.36	333,47.64	699
Uri (UR)	19,21.35	70,78.65	89
Wallis (VS)	131,37.86	215,62.14	346
Zug (ZG)	195,58.21	140,41.79	335
Zürich (ZH)	1807,47.00	2038,53.00	3845
Total by gender	8379,48.55	8881,51.45	17260

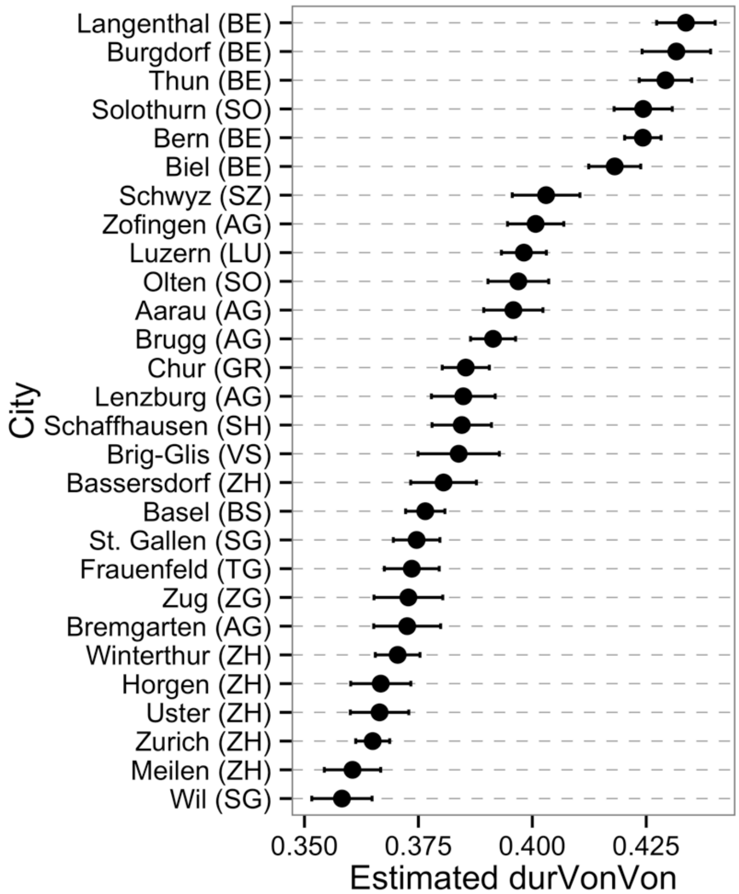
Table 2: Number of recordings by gender.

Coefficient	*Estimate*	*Std. Error*	*t value*	*Pr(>|t|)*
(Intercept)	0.35580	0.00223	159.47000	< 2e-16***
gender: male	-0.01545	0.00111	-13.98400	< 2e-16***
age	-0.00010	0.00003	-2.95100	0.003169**
word: Augen	0.02820	0.00187	15.10100	< 2e-16***
word: Donnerstag	0.10730	0.00186	57.61400	< 2e-16***
word: fragen	-0.01213	0.00187	-6.47100	0.0000000001***
word: heben	0.03162	0.00216	14.66100	< 2e-16***
word: trinken	0.06584	0.00184	35.87400	< 2e-16***

Table 3: Coefficients of linear model, using WORD as a fixed factor.

Canton	*Estimated durVonVon*	*Standard error*
Bern (BE)	0.415284	0.002544
Obwalden (OW)	0.413354	0.006685
Uri (UR)	0.405964	0.007892
Glarus (GL)	0.403934	0.005517
Schwyz (SZ)	0.397304	0.003954
Freiburg (FR)	0.396149	0.005079
Solothurn (SO)	0.392696	0.003277
Nidwalden (NW)	0.390219	0.007281
Luzern (LU)*	0.386934*	0.0913*
Appenzell Innerrhoden (AI)	0.386451	0.008349
Aargau (AG)	0.381489	0.002612
Zug (ZG)	0.376664	0.004472
Appenzell Ausserrhoden (AR)	0.373684	0.006881
Graubünden (GR)	0.373384	0.003729
St. Gallen (SG)	0.367254	0.002948
Thurgau (TG)	0.365614	0.00347
Basel-Stadt (BS)	0.364694	0.003642
Schaffhausen (SH)	0.364364	0.005249
Wallis (VS)	0.360824	0.004419
Basel-Landschaft (BL)	0.359494	0.003674
Zürich (ZH)	0.357004	0.002462

Table 4: Estimated means and standard errors for each canton.



City	*Estimated durVonVon*	*Standard error*
Langenthal (BE)	0.433709	0.006431
Burgdorf (BE)	0.431609	0.007522
Thun (BE)	0.429219	0.005764
Solothurn (SO)	0.424329	0.006385
Bern (BE)	0.424259	0.003999
Biel (BE)	0.418089	0.005707
Schwyz (SZ)	0.403025	0.007431
Zofingen (AG)	0.400726	0.006179
Luzern (LU)	0.39815	0.004932
Olten (SO)	0.396931	0.006668
Aarau (AG)*	0.395839*	0.0065*
Brugg (AG)	0.391386	0.004946
Chur (GR)	0.385399	0.005173
Lenzburg (AG)	0.384859	0.006998
Schaffhausen (SH)	0.384519	0.006528
Brig-Glis (VS)	0.383839	0.008925
Bassersdorf (ZH)	0.380519	0.007213
Basel (BS)	0.376509	0.004309
St. Gallen (SG)	0.374609	0.005113
Frauenfeld (TG)	0.373549	0.006023
Zug (ZG)	0.372799	0.00755
Bremgarten (AG)	0.372539	0.007349
Winterthur (ZH)	0.370449	0.004903
Horgen (ZH)	0.366739	0.006599
Uster (ZH)	0.366479	0.006414
Zurich (ZH)	0.364969	0.003727
Meilen (ZH)	0.360539	0.006186
Wil (SG)	0.358219	0.006615

Figure 15 and Table 5: Estimated means and standard errors for cities with at least 100 recordings.

FIGURES

Figure 1: The 26 Swiss cantons (map credit http://www.4allpc.ch/karte_kantone.jpg).

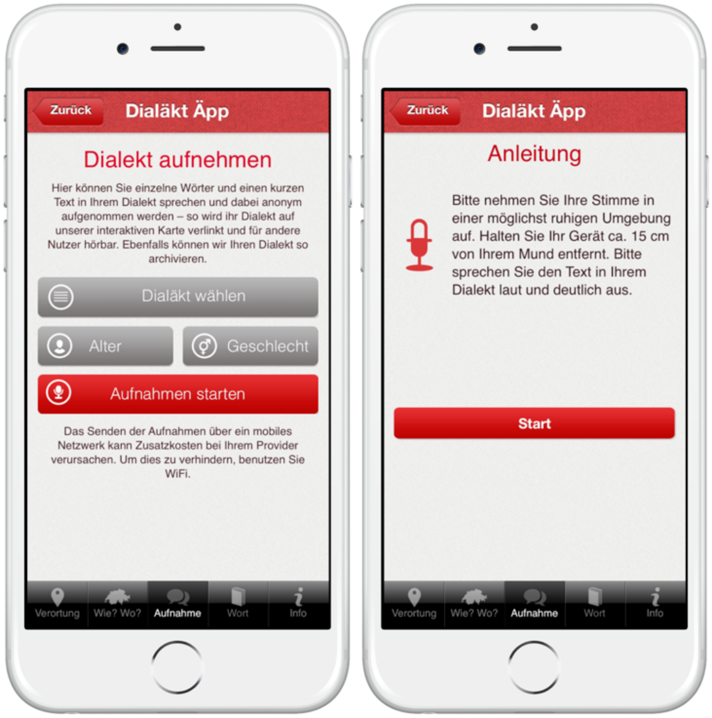


Figure 2: User interface for dialect, age, and gender selection (left panel) and recording instructions (right panel).

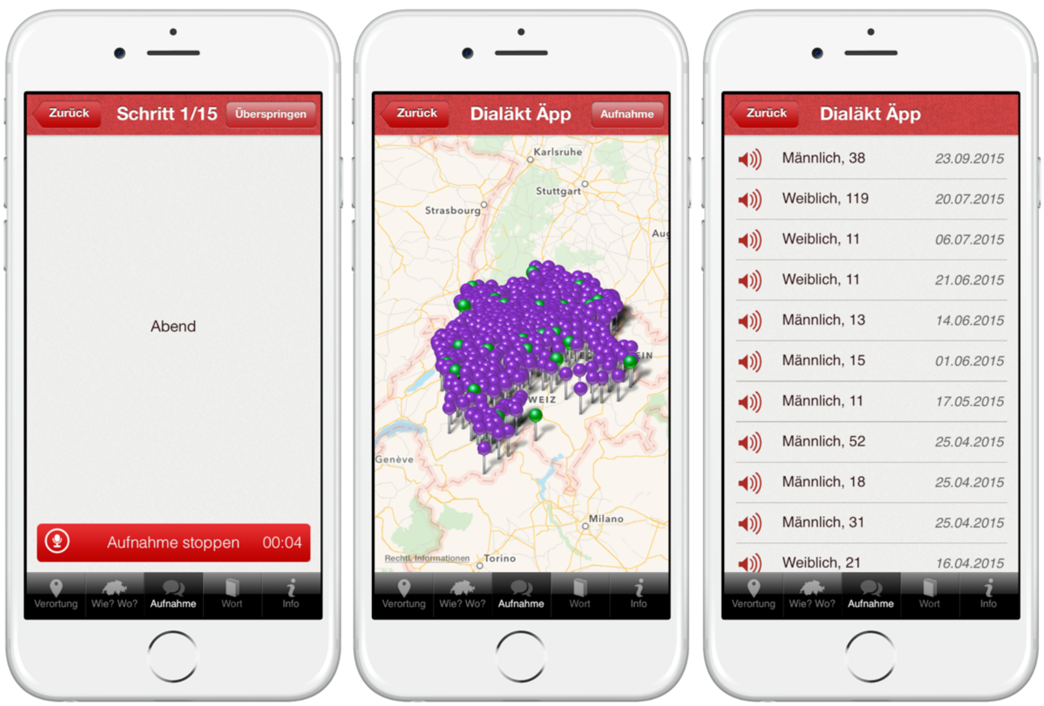
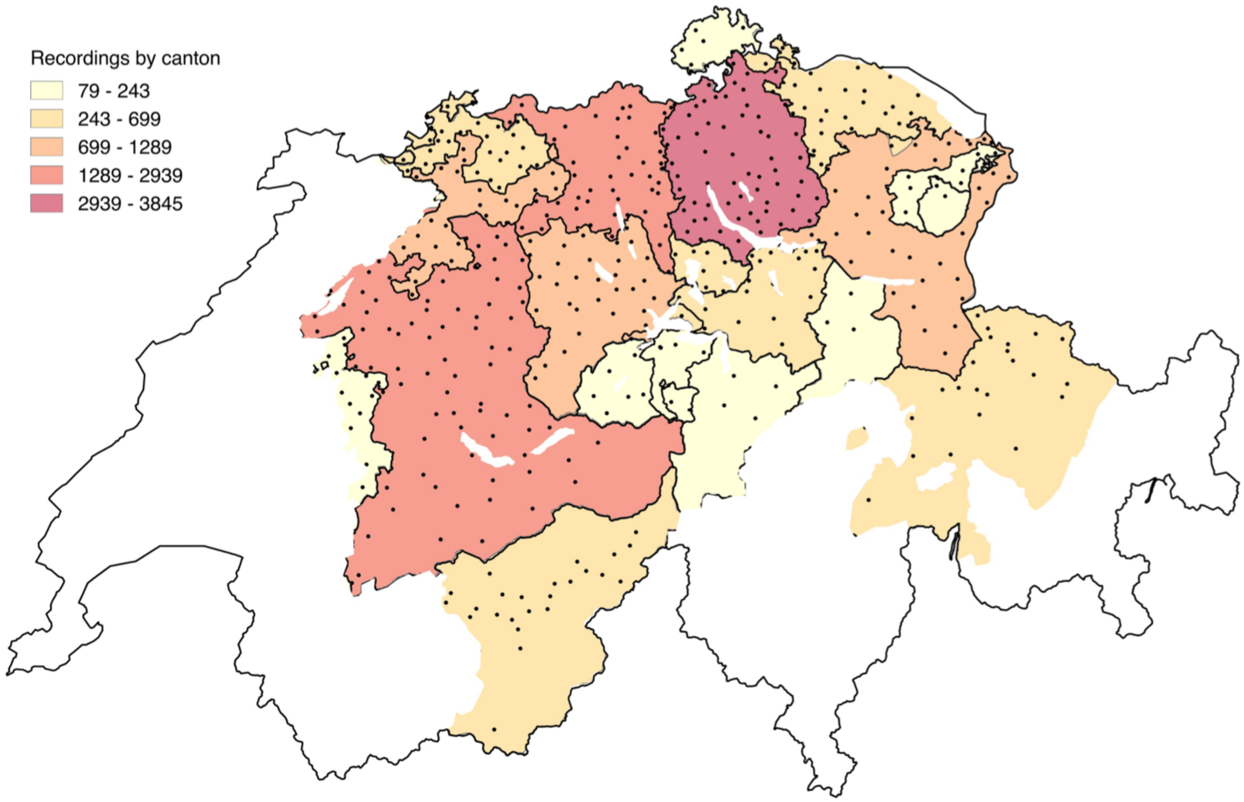


Figure 3: User interface for word recording (left panel), localities shown as pins (central panel), audio playback interface of one's own and other users' recordings when clicking on a given locality (right panel).

Figure 4: 452 elicited localities in German-speaking Switzerland.



926

927 **Figure 5**: Number of recordings by canton.

928

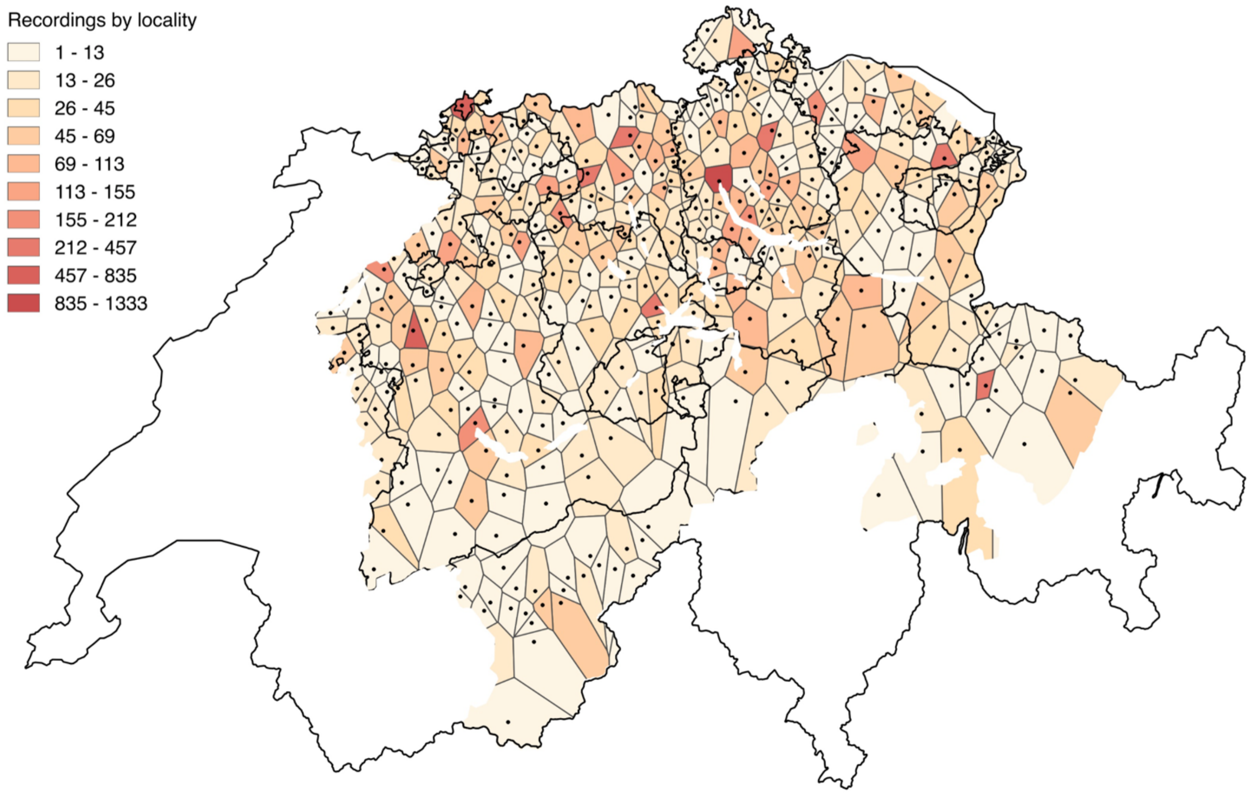


Figure 6: Number of recordings by locality.

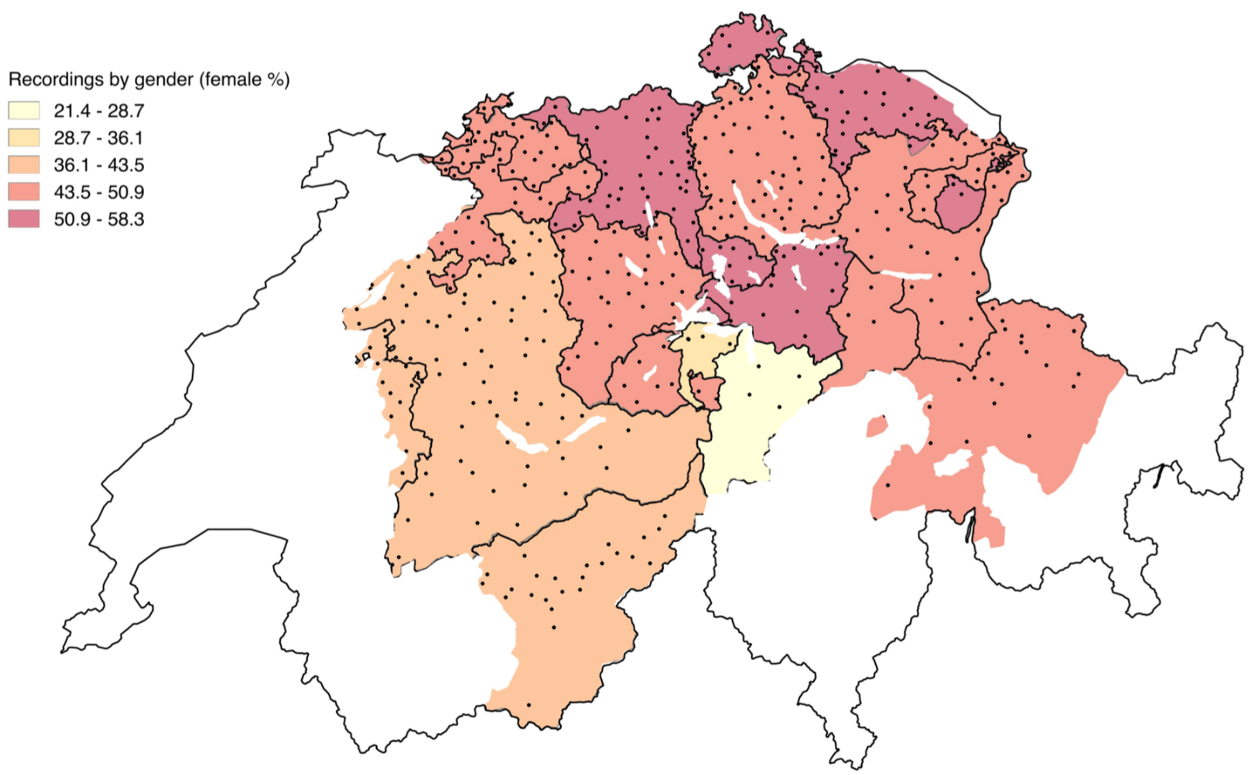


Figure 7: Number of recordings by canton and gender (as % female).

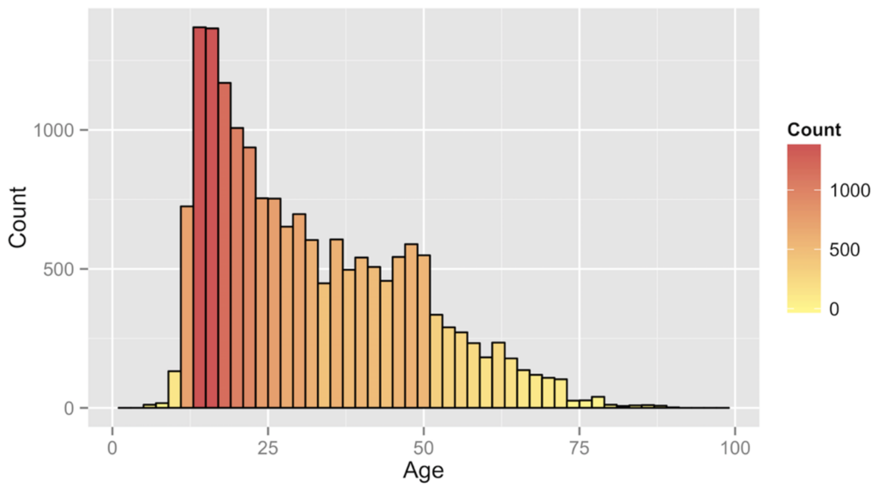


Figure 8: Distribution of speaker ages.

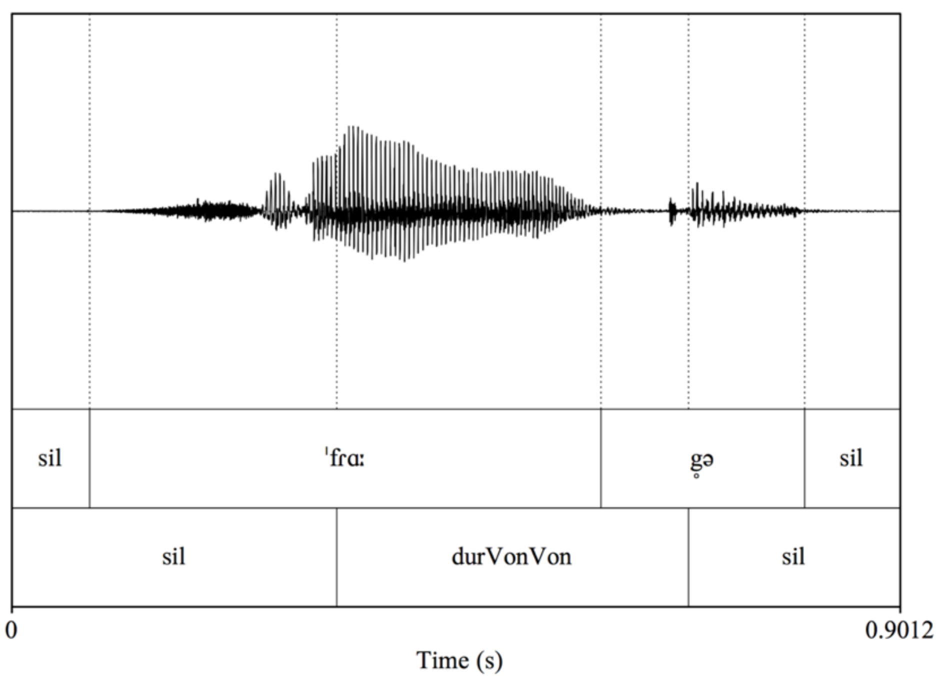


Figure 9: Schematic of vowel-onset-to-vowel-onset measurement (2nd tier).

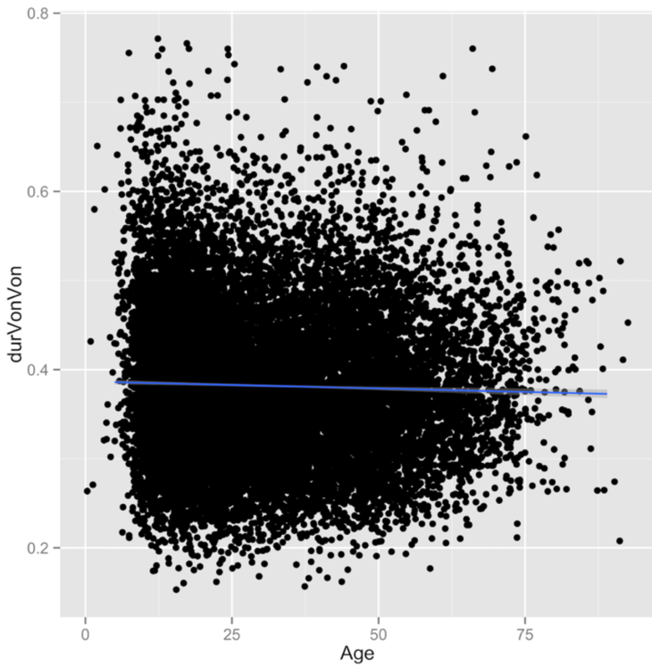


Figure 10: Scatterplot of *durVonVon* as a function of AGE.

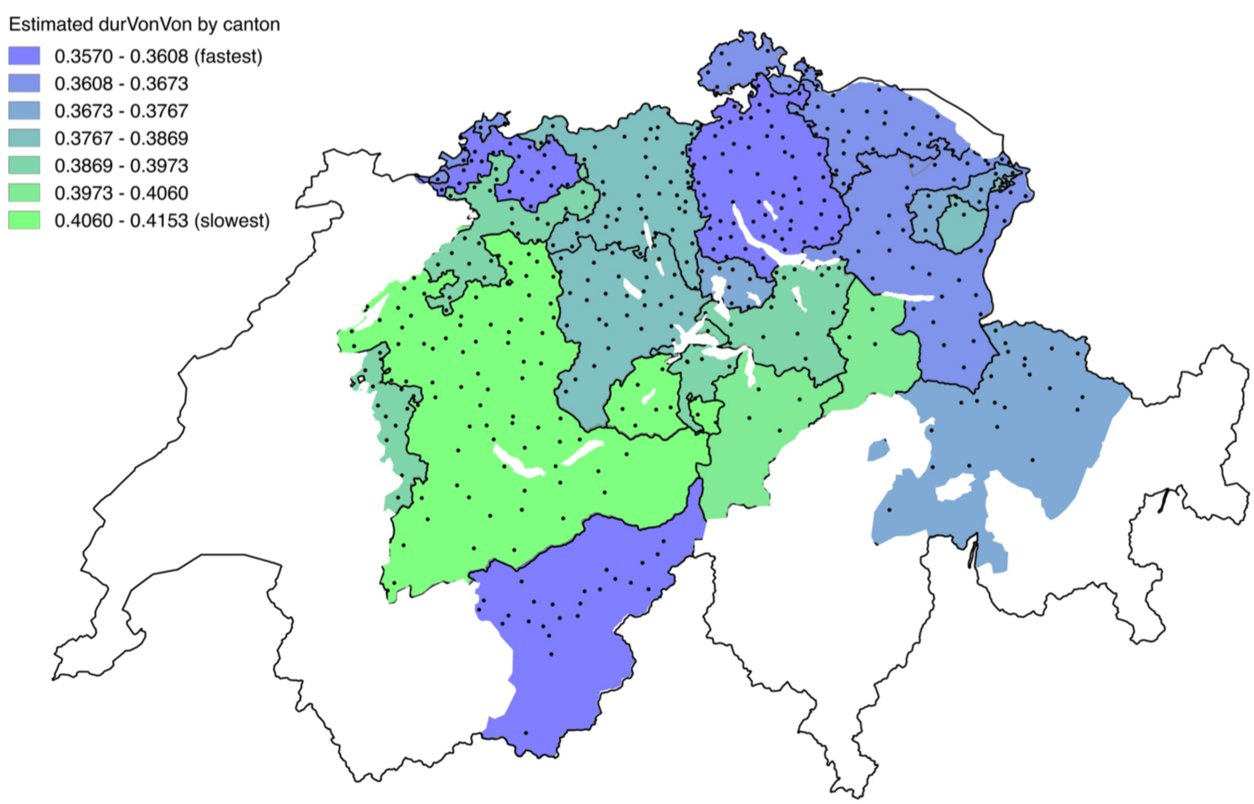


Figure 11: *durVonVon* by canton. The darker blue the region, the lower *durVonVon* (i.e. the faster); the lighter green the region, the higher *durVonVon* (i.e. the slower).

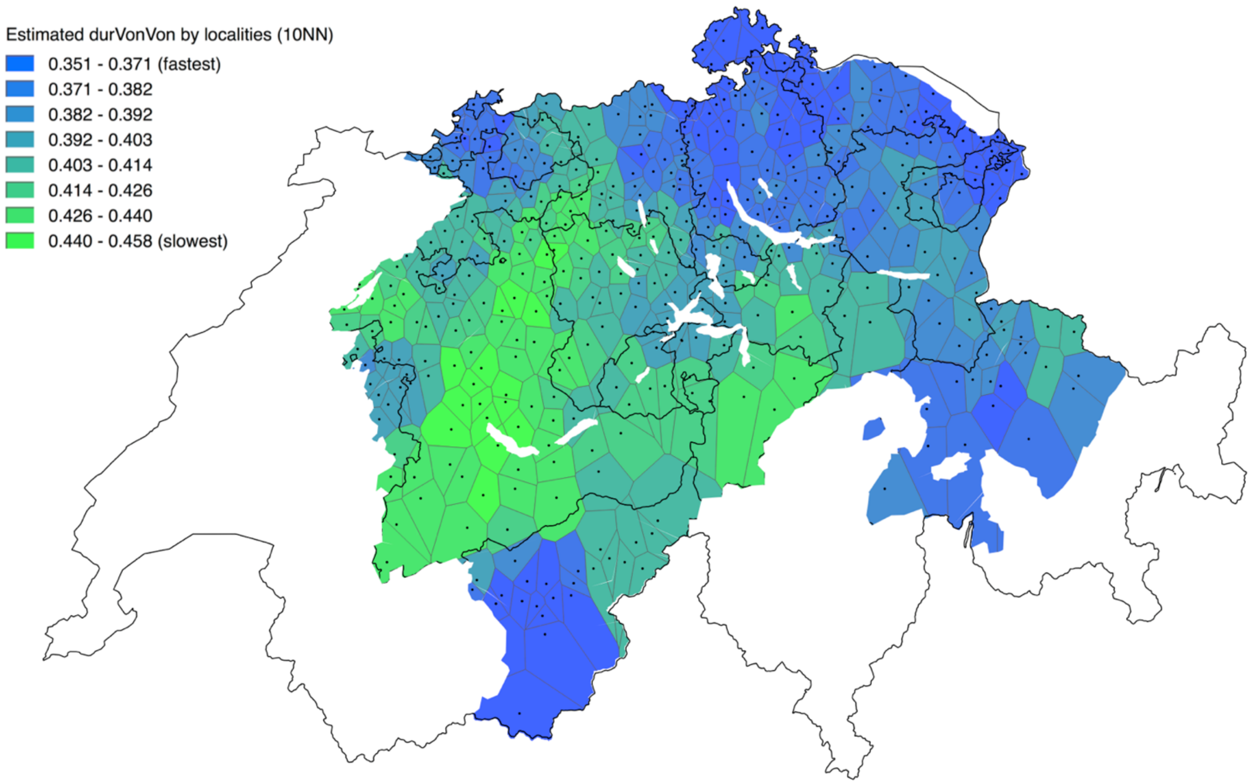


Figure 12: *durVonVon* by locality. The darker blue the locality, the lower *durVonVon* (i.e. the faster); the lighter green the locality, the higher *durVonVon* (i.e. the slower).

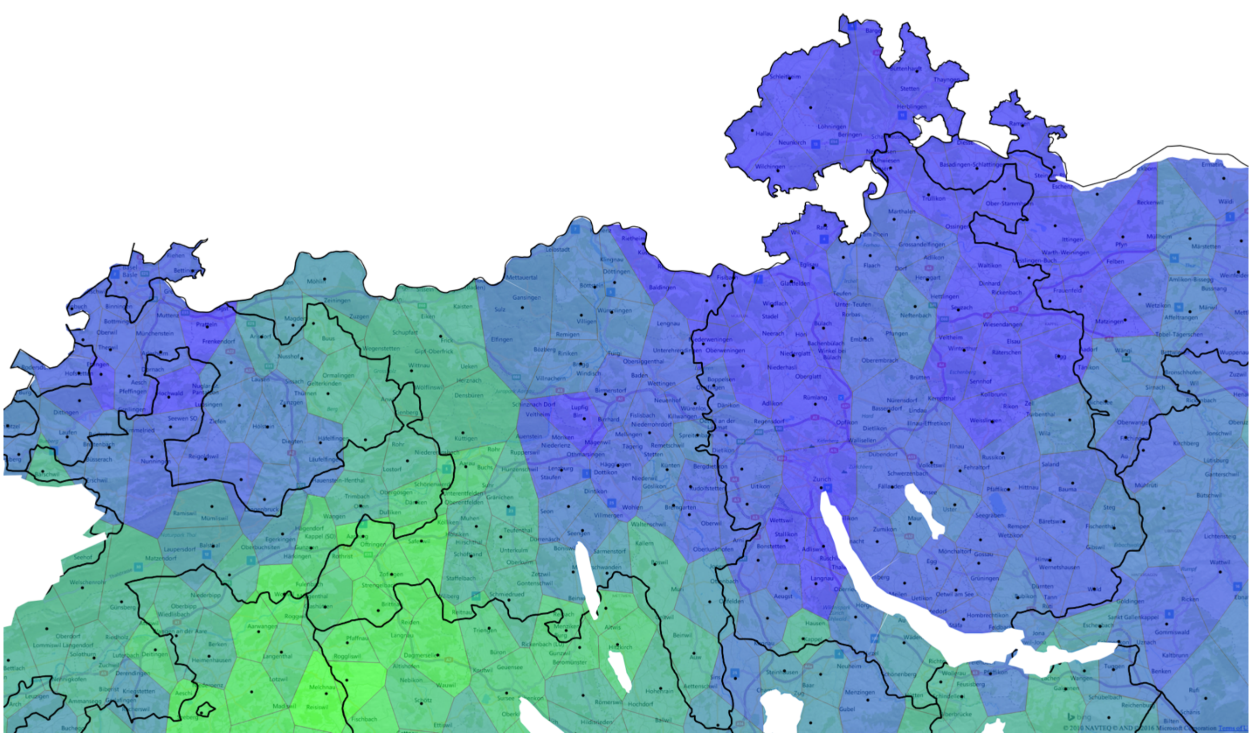


Figure 13: Close-up map of between-locality variation in *durVonVon* for the cantons of Zurich and Aargau.

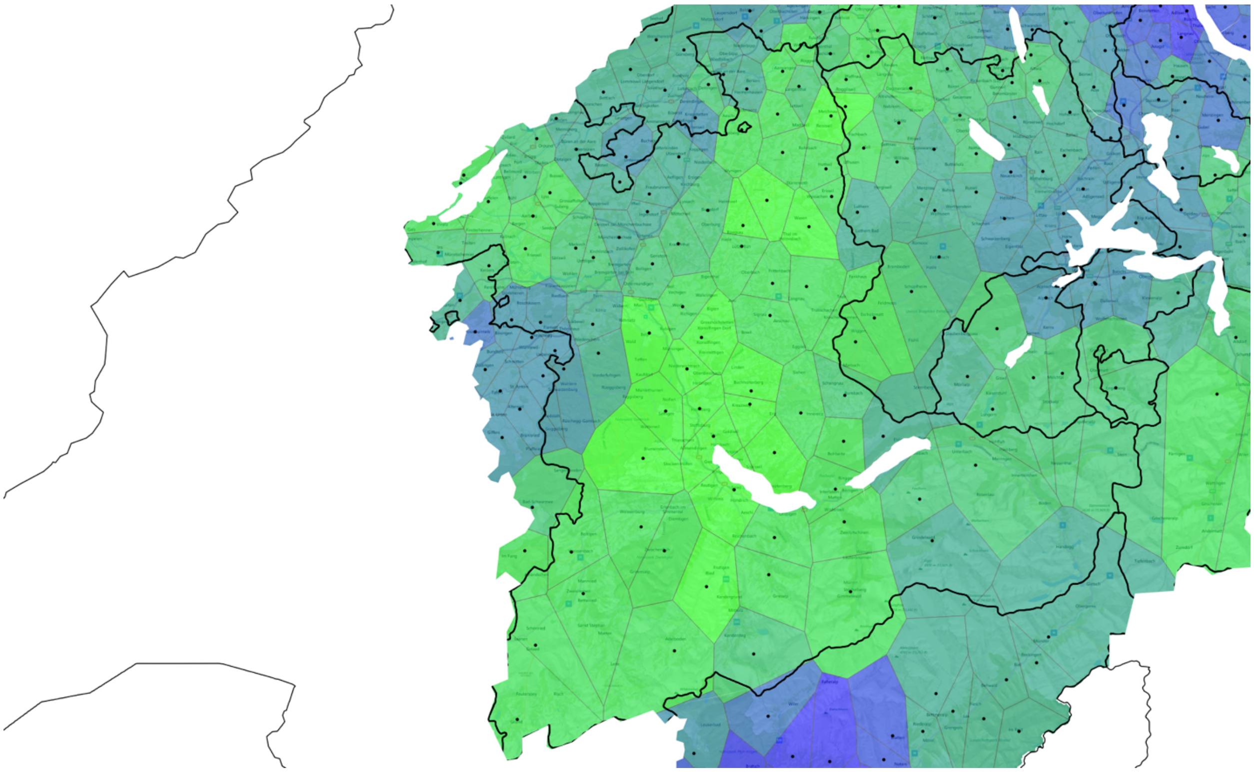


Figure 14: Close-up map of between-locality variation in *durVonVon* for the canton of Bern.

AUDIO FILES

AudioFile_1.wav

AudioFile_2.wav

AudioFile_3.wav

AudioFile_4.wav

AudioFile_5.wav

AudioFile_6.wav

AudioFile_7.wav

AudioFile_8.wav

AudioFile_9.wav

AudioFile_10.wav

AudioFile_11.wav

AudioFile_12.wav

AudioFile_13.wav

DATA FILE

Full_Set.txt