FROM PASTA TO P^{H}IZZA: VARIATION AND CHANGE IN VOT IN SWISS GERMAN

Carina Steiner[1,2], Péter Jeszenszky[3], Adrian Leemann[1,2]

[1]University of Bern, Switzerland; [2]University of Zurich, Switzerland, [3]Independent researcher
carina.steiner@unibe.ch, pjeszenszky@gmail.com, adrian.leemann@unibe.ch

ABSTRACT

In contrast to Standard German (StG) or English, word-initial bilabial and alveolar stops are traditionally not aspirated in Swiss German (SwG), and typically exhibit short voice onset times (VOT; e.g. StG [ˈpʰitsa] vs. SwG [ˈpitsa]). Throughout the last century, however, more and more aspirated forms have been reported for SwG, but empirical evidence on the factors driving this change is thin. This study is the first to investigate SwG VOT in a large and balanced sample of 1000 speakers from 125 localities. The results suggest an apparent-time change, depending on factors related to the specific word as well as to the region. Furthermore, our models revealed that more frequent use and more favorable attitudes towards StG trigger longer VOT. This effect is particularly interesting against the backdrop of the special diglossic situation in German-speaking Switzerland and suggests that linguistic practices and attitudes can be important motivators of sound change.

Keywords: VOT, aspiration, voiceless stops, sound change, dialect-standard convergence

1. INTRODUCTION

The time interval between the release of a stop and the onset of periodicity in the following vowel (voice onset time; VOT) is a distinguishing typological feature. Across different languages, VOT has been found to increase as a function of front to back in place of articulation ([1–3]; but cf. also [4]). Further, VOT may vary in terms of prosodic features related to the position in a phrase [5, 6] or lexical frequency [6–8].

Two recurring sociolinguistic factors which have been found to cause variation in VOT are age and linguistic repertoire, encompassing the range of a person's linguistic competences and resources. Concerning age, previous studies have yielded mixed results [e.g. 3, 9]. Regarding linguistic repertoire, interesting insights come from [10], who reported convergence trends in VOT in the spontaneous code-switching of Spanish-English bilinguals (cf. also [11]).

In Standard German, VOT has been described as the most important cue to contrast between word-initial voiceless and voiced stops [12][1]. In German dialects spoken in Switzerland, however, fortis and lenis are traditionally not contrasted by VOT, but by closure duration and potential F0 effects in the following vowel, as shown in [13] (cf. [4, 14] for similar phenomena in other German varieties and [3, 15] for other dialect–standard contexts). The following example illustrates the /d, t/ contrast in both varieties in *danken* 'to thank' vs. *tanken* 'to refuel':

- Standard German (StG): [ˈ**d̥**aŋkn̩] vs. [ˈ**tʰ**aŋkn̩]
- Swiss German (SwG): [ˈ**d̥**ɑŋ͡kxə] vs. [ˈ**t**ɑŋ͡kxə]

Thus, except for some rare minimal pairs (cf. e.g. [13]), word-initial /p, t/ are generally unaspirated in SwG.[2] In recent years, however, it has been argued that the dialects might show convergence trends to StG in that – depending on a set of intra- and extralinguistic factors – longer VOTs in word-initial bilabial and alveolar fortis stops may occur.

The main *intralinguistic* factors driving this change are assumed to be related to lexical criteria. For example, [8] report aspirated stops in loan words from StG (e.g. [tʰeː] 'tea') as well as from English (e.g. [ˈpʰɑrti] 'party', cf. also [16]), and in proper nouns (e.g. [pʰɑʊl] 'Paul'). Based on an analysis of grammars, linguistic atlases, and personal field notes, [2] also documents more aspirated forms in proper nouns, referring not only to persons, but also to places (e.g. *Thun* [ˈtʰuːn]). In relation to the last example, [2] argues that the graphic representation <Th> might also trigger aspiration. Furthermore, according to [13], the letter names of <p> and <t> are nowadays usually aspirated. Due to the strong lexical determination of aspiration in SwG, [13] even discuss a potential slow lexical diffusion of a sound change which might lead to a phonological contrast as in StG or English in the distant future.

Regarding *sociolinguistic* variation in VOT in SwG, regional stratification has been investigated to some degree, while other factors have remained empirically uncharted. Previous reports point towards a North–South divide with speakers from urban, Northern regions exhibiting longer VOTs than those from the Alpine South [2, 17]. Besides region, [2] further speculates that aspiration may convey cosmopolitanism, urbanity, and education, which leads to the conclusion that style and social meaning

may be even more important than regional origin in explaining variation in VOT.

In sum, while VOT has been broadly investigated in various languages, empirical evidence for variation and change due to sociolinguistic factors in SwG is thin. The goal of this study is to bridge this gap by investigating VOT in word-initial /p, t/ in a controlled sample of 1000 speakers from 125 localities, taking into account large-scale metadata. The special diglossic situation and the atypically high prestige attributed to the dialect (cf. e.g. [18]) further make SwG an interesting context in which to explore effects of linguistic practices and attitudes on VOT.

2. METHODS

The current study draws on data from the SDATS corpus [19]. Its materials, speakers, and procedures are detailed in sections 2.1–2.3.

2.1. Materials

Table 1 presents the four stimuli selected for the current study, two with a word-initial /p/ and two with a /t/, in both isolated and phrase-embedded contexts.

Item	Place o.A.	Context
Parlament	bilabial	isolated
Patrick	bilabial	in phrase
Thermometer	alveolar	isolated
Tee	alveolar	in phrase

Table 1: Properties of the four VOT stimuli.

Besides place of articulation and co-articulation context, the items were selected due to lexical properties discussed in the literature: first, they vary in lexical frequency. Second, *Tee* was chosen as a typical example of a StG loanword, as opposed to *Parlament* as a French borrowing; *Patrick* represents a proper noun, and with *Thermometer*, graphemic influences beyond place names can be tested.

2.2. Speakers

The SDATS database comprises a total of 1000 speakers from 125 localities in German-speaking Switzerland, balanced across age (four younger and four older speakers) and gender (two female and two male speakers per cohort) in each locality. All participants are autochthone speakers of the local dialect with a comparatively low long-term mobility and are competent and daily users of (Swiss) StG, while their English skills vary.

For the current analysis, 11 speakers were excluded due to missing data, reducing the sample to 989 speakers.

2.3. Procedures

Data collection took place in 2020 and 2021. The four VOT items were collected as part of 2–3h oral dialect interviews. *Parlament* and *Thermometer* were elicited via picture-naming tasks, while *Patrick* and *Tee* were embedded in sentence translation tasks.

VOT was coded manually in Praat [20]. The period of aperiodic friction between stop release and the initiation of periodicity was identified based on acoustic and visual inspection of the oscillogram and spectrogram. To normalize VOT for articulation rate, the time between stop release and the end of the following vowel (RVT) was coded and normVOT was calculated as the relative proportion of RVT (normVOT = VOT/RVT*100).

Statistical analyses were conducted in R [21]. After descriptive and visual data inspection, mixed effects modeling was performed using lme4 [22] with normVOT as outcome and the predictors listed in Table 2 as fixed effects. Random intercepts were allowed for speakers. We started with a model comprising all main effects plus potential two-way interactions, of which only those reaching $p<.01$ were retained in the final model.

Fixed Effects	Specifications
Item	4 levels, cf. Table 1
Age Cohort	2 levels: older (60+), younger (20–35)
Region	8 levels, clustered based on [23]
DSP	numeric -10 to 10: Dialect Standard Profile (cf. below)
EN Competence	Numeric from 1–7, based on self-assessments
Education	4 levels: (voc.) baccalaureate, secondary voc. education, tertiary voc. education, university degree
Gender	2 levels: female, male

Table 2: Structure of fixed effects. EN=English, voc.=vocational.

A note on the DSP in Table 2: the Dialect Standard Profile (DSP) was designed following the principles of the Bilingual Language Profile (BLP [24]), an established language dominance index for bilingual speakers. Analogous to the BLP, the DSP was based on questionnaire data about productive (e.g. varieties spoken with family, friends, etc.) and receptive use (e.g. TV, radio), as well as attitudes towards both varieties (for more details, cf. analysis report: https://osf.io/rq5gk/).

3. RESULTS

Our final model revealed significant effects for item, age, region, DSP, and education, as well as item*age and item*region interactions (the full model output can be consulted in the analysis report, cf. link in 2.3). All effects are presented in more detail in sections 3.1–3.4. Figures 1–3 do not represent raw data, but predicted values based on the final model with all fixed and random effects taken into account (formula: normVOT ~ Item + Age Cohort + Region + DSP + EN Competence + Education + Gender + Age Cohort:item + item:Region + (1 | Speaker)).

3.1. Item and age

In general, normVOT was higher in the younger cohort (1.84(±0.61), t=3.03, p<0.01) and higher for alveolar compared to bilabial stops (most pronounced in the case of *Thermometer* compared to *Parlament*, 23.26(±.93), t=25.11, p<0.001).

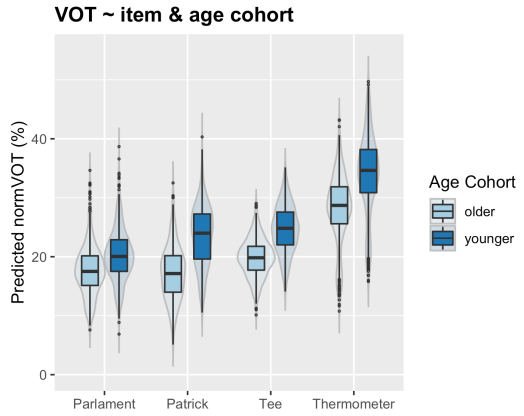

Figure 1: Predicted VOT according to item and age.

However, age and item interacted, as shown in Figure 1: while only slight age-related differences were identified in *Parlament*, the gap widened in *Tee* and *Thermometer* and was the biggest in the case of *Patrick*, where the highest between-subject variation within the younger cohort was also observed.

3.2. Item and Region

Figure 2 illustrates the interaction between item and region. With regard to *Parlament*, the North-Western region showed the highest predicted normVOT, and for the remaining territory, a slight increase from west to east was revealed. While *Patrick* also exhibited the highest values in the North-West, normVOT was predicted to be higher in central regions compared to the Alpine South. Almost no regional differences were identified for *Tee*. A sharp North–South contrast, paired with the highest normVOTs overall, was revealed for *Thermometer*.

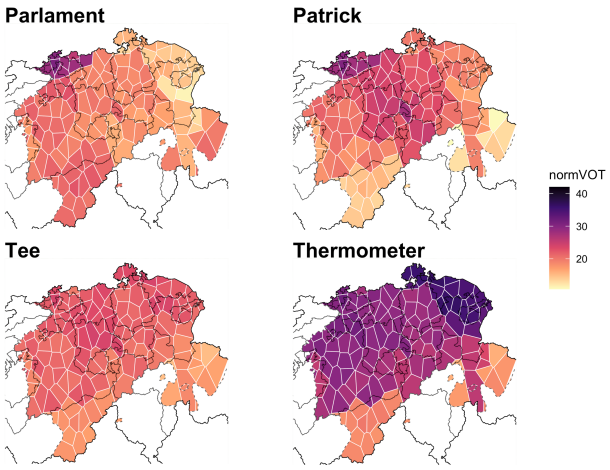

Figure 2: Predicted VOT according to region and item. Darker colors suggest higher values.

3.3. Language attitudes and use

The effect for DSP (cf. Figure 3) can be read as follows: The stronger the dialect orientation in terms of attitudes and use, the lower the normVOT (-0.23(±0.10), t=-2.41, p=0.016). Higher normVOT, in turn, was associated with more positive attitudes towards and more frequent use of StG.

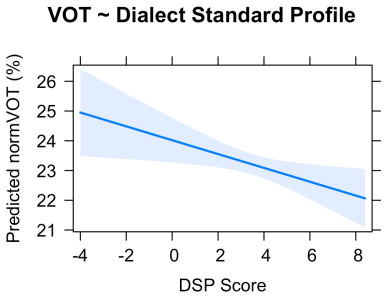

Figure 3: Predicted VOT according to DSP.

Further, our model revealed a positive trend slightly above the p=.05-level for higher English competence increasing normVOT (0.26(±0.14), t=1.86, p=0.063).

3.4. Education and gender

While no gender effects were detected, speakers with a (vocational) baccalaureate were shown to exhibit slightly higher normVOTs compared to those with other educational backgrounds; the effect was most pronounced in comparison with speakers with secondary vocational education (-1.14(±0.50), t=.2.27, p=0.024).

4. DISCUSSION

4.1. Intralinguistic effects

Our study corroborates previous findings in that VOT is longer in alveolar than in bilabial stops [1–3]. Furthermore, the notably longer VOTs in *Thermometer* support the argument presented in [2], showing that the graphemic representation may trigger aspiration, regardless of whether it is a place name or not. Our study is also in line with previous work regarding longer VOTs in lexical borrowings from StG, and – comparing the two bilabial stops – longer VOTs for *Patrick* as a proper noun [8, 13] (cf. also interaction with age discussed in section 4.2 below). Lastly, frequency effects discussed in [6–8] might be at play, since the longest VOTs were identified for the least frequent word according to [25] (*Thermometer*; 3/7) and consecutively longer VOTs for the more frequent *Tee* (4/7) and *Parlament* (5/7) (*Patrick* was not integrated since no frequency statistics existed for this proper noun). This result suggests that, while more commonly used words are pronounced in a rather traditional manner, less typical words might trigger more standard-like, aspirated forms.

While these are the main factors discussed in previous literature, they only account partly for variation in VOT and further specificities related to the particular words might be at play (e.g. syllabic structure or position, cf. also 4.3. limitations).

4.2. Sociolinguistic effects

While age effects generally showed the expected standard-convergent apparent-time trend and did not interact with other social variables (as e.g. in [3, 9]), the age*item interaction requires further discussion. The fact that the biggest age difference as well as the highest variability among the younger speakers were related to *Patrick* suggests that – in line with [2, 8] – the ongoing change might be most obvious in proper nouns. The reason for the least age-related differences being found in *Parlament* might be twofold: The high frequency discussed in section 4.1 above might not only trigger the shortest VOT, but also the slowest change. Second, the short-lag VOT might also be explained by the fact that this lexeme was borrowed from French rather than StG or English.

Regarding regional variation, the North–South divide documented in [2, 17] was replicated for *Thermometer* and – to some degree – for *Patrick*, whereas the other items exhibited different regional distributions. Similar VOTs across all regions in *Tee* suggests that, while the first instances of aspirated forms were reported at the beginning of the 20th century [2], this sound change has advanced to all regions in German-speaking Switzerland. At the same time, the apparent-time effect indicates that it has not yet completed. A surprising regional effect was related to lower VOTs in *Parlament* in the East, because standard convergence trends are usually most obvious in these regions.

The effect related to the Dialect Standard Profile showed that cross-linguistic influences in VOT may not only be at play among different *languages* in a repertoire (as in [10, 11]), but also among different *varieties*. For example, a person who frequently watches StG TV shows and who feels comfortable speaking the standard language may aspirate word-initial stops more often. More generally speaking, this effect suggests convergence to StG for speakers with more favorable attitudes towards and more frequent productive and receptive use of the standard language. The importance of this index as a predictor in our model, as well as the identified positive trend for English competence, suggests that it would be worthwhile for future studies to investigate influences of language competence, practices, and attitudes in dialect–standard situations more deeply.

Lastly, while – as expected – no differences between male and female speakers arose, the effect for education is surprising to a certain extent, given that SwG is usually not expected to vary in terms of educational background. However, this effect reflects an age trend, i.e., speakers with a (vocational) baccalaureate in our study are also the youngest ones.

4.3. Limitations and conclusion

The main limitation of the current study is that more items would be needed to allow for a more in-depth investigation of the phonetic environment. Furthermore, the four items were coded by four different people, and although the coding procedure was objective and problematic cases were cross-checked by a second coder, influences on between-item variation cannot be ruled out completely.

Notwithstanding these limitations, this is the first large-scale multilocality study on variation and change in VOT in SwG, yielding generalizable results due to its big sample size, controlled speaker selection, and supervised data collection. We showed that increasing VOT in word-initial stops can be seen as a standard-converging sound change in progress, which, on the one hand, may heavily depend on word-specific characteristics, and, on the other hand, is influenced by the speakers' age, regional origin, and their linguistic practices and attitudes. However, if this phonetic change may lead to a potential standard-like phonological contrast, as discussed in [13], remains open for future research.

5. ACKNOWLEDGEMENTS

This research was funded by the Swiss National Science foundation (Grant no. PCEFP1_181090).

6. REFERENCES

[1] Cho, T., Ladefoged, P. 1999. Variations and universals in VOT: Evidence from 18 languages. *J. Phon.*, vol. 27, 207–229.

[2] Schifferle, H. 2010. Zunehmende Behauchung. Aspirierte Plosive im modernen Schweizerdeutsch. In: Christen, H., German, S., Haas, W., Montefiori, N., Ruef, H. (eds), *Alemannische Dialektologie: Wege in die Zukunft. Beiträge zur 16. Tagung für alemannische Dialektologie in Freiburg/Fribourg vom 07.–10.09.2008*, 43–55.

[3] Stuart-Smith, J., Sonderegger, M., Rathcke, T., Macdonald, R. 2015. The private life of stops: VOT in a real-time corpus of spontaneous Glaswegian. *Lab. Phonol.* 6(3–4), 505–549. doi: 10.1515/lp-2015-0015.

[4] Hödl, P. 2019. Perception of voice onset time by Austrian and German listeners, in *Proc. 19th ICPhS* Melbourne, 3686–3690.

[5] Cole, J., Kim, H., Choi, H., Hasegawa-Johnson, M. 2007. Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech. *J. Phon.*, 35(2), 180–209. doi: 10.1016/j.wocn.2006.03.004.

[6] Yao, Y. 2009. Understanding VOT Variation in Spontaneous Speech. *UC Berkeley PhonLab Annu. Rep.* 5(5), 29–43.

[7] Yu, A. C., Abrego-Collier, C., Sonderegger, M. 2013. Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and "autistic" traits. *PloS One* 8(9), e74746.

[8] Fleischer, J., Schmid, S. 2006. Zurich German. *J. Int. Phon. Assoc.* 36(2), 243–253. doi: 10.1017/S0025100306002441.

[9] Torre, P., Barlow, J. A. 2009. Age-related changes in acoustic characteristics of adult speech. *J. Commun. Disord.* 42(5), 324–333. doi: 10.1016/j.jcomdis.2009.03.001.

[10] Balukas, C., Koops, C. 2015. Spanish-English bilingual voice onset time in spontaneous code-switching. *Int. J. Biling.* 19(4), 423–443. doi: 10.1177/1367006913516035.

[11] Sancier, M. L., Fowler, C. A. 1997. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *J. Phon.* 25(4), 421–436.

[12] Beckman, J., Jessen, M., Ringen, C. 2013. Empirical evidence for laryngeal features: Aspirating vs. true voice languages. *J. Linguist.* 49(2), 259–284. doi: 10.1017/S0022226712000424.

[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. *J. Phon.* 71, 229–248. doi: 10.1016/j.wocn.2018.09.003.

[14] König, W. 2001. Ein Lautwechsel T>K. Oder wie die Tartoffel zur Kartoffel wurde. *Ger. Pragensia* XVIII, 95–101.

[15] Scobbie, J. M. 2006. Flexibility in the face of incompatible English VOT systems. *Lab. Phonol.* 8, 367–392.

[16] Dalcher, P. 1986. Anglicisms in Swiss German: The evaluation by computer of a survey conducted in 1964/5. In: Viereck, W., Bald, W.-D. (eds), *English in Contact with other Languages*. Budapest: Akadémiai Kiadó, 179–206.

[17] SDS = Hotzenköcherle, R., Baumgartner, H. (eds). 1962–2003. *Sprachatlas der deutschen Schweiz*. Bern / Basel: Francke.

[18] Berthele, R. 2004. Vor lauter Linguisten die Sprache nicht mehr sehen – Diglossie und Ideologie in der deutschsprachigen Schweiz, In: Christen, H. (ed), *Dialekt, Regiolekt und Standardsprache im sozialen und zeitlichen Raum. Beiträge zum 1. Kongress der Internationalen Gesellschaft für Dialektologie des Deutschen*. Wien: Praesens, 111–136.

[19] Leemann, A., Jeszenszky, P., Steiner, C., Messerli, J., Studerus, M. 2020. *SDATS Corpus – Swiss German dialects across time and space*. [Online]. Available: osf.io/s9z4q.

[20] Boersma, P.,Weenink, D. 2022. *Praat: doing phonetics by computer*. Version 6.3.02. [Online]. Available: http://www.praat.org/.

[21] R Core Team. 2022. *R: A language and environment for statistical computing*. Version 4.2.2. R Foundation for Statistical Computing, Vienna, Austria, 2022. [Online]. Available: https://www.R-project.org/.

[22] Bates, D., Maechler, M., Bolker, B., Walker, S. 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67(1), 1–48. doi: 10.18637/jss.v067.i01.

[23] Scherrer, Y., Stoeckle, P. 2016. A quantitative approach to Swiss German – Dialectometric analyses and comparisons of linguistic levels. *Dialectol. Geolinguistica* 24(1), 92–125.

[24] Birdsong, D., Gertken, L. M., Amengual, M. 2022. *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism*. Center for Open Educational Resources and Language Learning (COERLL), University of Texas Austin. [Online]. Available: https://sites.la.utexas.edu/bilingual/.

[25] DWDS = Berlin-Brandenburgischen Akademie der Wissenschaften (eds). 2022. *DWDS – Digitales Wörterbuch der deutschen Sprache. Das Wortauskunftssystem zur deutschen Sprache in Geschichte und Gegenwart*. [Online]. Available: https://www.dwds.de/.

[1] Findings from [4, 13], however, suggest that the reality is more complex than a mere dichotomy based on VOT.

[2] Velar stops are not considered in this case since they follow different dialectal distributions, e.g. word-initial /k/ is realized as [χ] in most SwG dialects.