ANALYZING VARIABILITY IN CLOSURE VOICING AND CO-INTRINSIC F0 IN CENTRAL STANDARD SWEDISH

James Kirby[1, a)], Maryann Tan[2, b)]

[1]Institute for Phonetics and Speech Processing (IPS), LMU Munich
[2]Centre for Research in Bilingualism, Stockholm University
a) jkirby@phonetik.uni-muenchen.de, b) maryann.tan@biling.su.se

ABSTRACT

The voicing contrast of Central Standard Swedish has been argued to be typologically unusual, on the grounds that it contrasts phonetically prevoiced with voiceless aspirated plosives. However, reports vary regarding how often voiced plosives are actually realized with closure voicing. In this study, we measure how often phonologically voiced plosives /b d/ are devoiced, along with F0 trajectories and aspiration durations following release of closure. Results reveal considerable variation in voicing rates: some speakers consistently prevoice /b d/, while others consistently devoice them. Phonologically voiced plosives are frequently realised with short-lag VOT, regardless of whether or not they are also realized with closure voicing. Finally, F0 trajectories following nasals and phonologically voiced plosives are indistinguishable, regardless of whether plosives are prevoiced or devoiced, but differ significantly from those following phonologically voiceless plosives. F0 may thus be a more reliable phonetic indicator of laryngeal status in Swedish than the presence of closure voicing.

Keywords: Swedish, microprosody, voicing, VOT

1. INTRODUCTION

Central Standard (Stockholm) Swedish (CSS) has been argued to have a typologically unusual, phonetically overspecified two-way laryngeal contrast between prevoiced lenis /b d g/ and postaspirated fortis /p t k/ plosives [1, 2]. However, reports vary regarding the extent to which /b d g/ are actually prevoiced, at least in utterance-initial position. While some researchers [1, 2] have reported primarily strong prevoicing for both males and females (as do [3] for Umeå Swedish), others [4, 5] maintain that they are typical voiceless unaspirated, especially in the speech of female talkers, presumably due to a smaller supralaryngeal vocal tract disfavoring the aerodynamic conditions required for voicing [6].

This reported variability in the realization of prevoicing makes CSS an interesting language to consider in terms of the microprosodic or *co-intrinsic* pitch perturbations (hereafter CF0) conditioned by onsets. First observed in so-called "aspirating" languages such as German [7] and English [8], CF0 effects have also been documented in "true voicing" languages such as French [9, 10], Italian [10], and Spanish [11]. The basic observation is that a CF0 dichotomy exists between phonologically voiced and voiceless (or lenis and fortis) obstruents, F0 being higher following the voiceless/fortis member compared with the voiced/lenis member, regardless of other aspects of their phonetic realization [12, 13].

If CF0 effects are primarily the result of a gesture or gestures designed to support or enhance voicing [12, 14], CF0 might be expected to differ following devoiced and prevoiced tokens of phonologically voiced stops, particularly if the effect is contingent on the realization of glottal pulsing during the closure phase. If the dichotomy is primarily driven by gestures implemented to suppress voicing during the closure phase of voiceless plosives [13, 15, 16], on the other hand, F0 trajectories following phonologically voiced plosives would not be expected to differ regardless of whether or not plosives are realized with closure voicing.

Because CF0 effects have not previously been reported for CSS, and because reports differ regarding the frequency with which CSS voiced plosives are realized with closure voicing, in this study we revisit the acoustic realization of this laryngeal contrast in a sample of CSS speakers. We focus on the questions of how often phonologically voiced plosives in CSS are realized with/out glottal pulsing during the closure phase (Q1), as well as how prevoiced and devoiced plosives differ in terms of their post-release voicing lag (Q2) and F0 trajectories (Q3).

2. METHODS AND MATERIALS

2.1. Participants

Forty native speakers of Standard Central (Stockholm) Swedish (24 female, ages 20-44, median 28;

16 male, ages 18-43, median 34) were recruited for this study. They were paid 100SEK in the form of cash or gift vouchers for their effort. No participants reported a history of speech or hearing disorders. Data from one female speaker was removed as she was later revealed not to speak the Stockholm variety, leaving data from 39 speakers for analysis.

2.2. Speech materials

Forty-eight words with /b p d t m n/ onsets were selected for recording (see Appendix) along with 22 real and non-words of the form /hVd/ or /hVr/ for a separate vowel analysis not reported here. We aimed to include as many monosyllabic voicing pairs as possible, but did not control for word frequency or other lexical characteristics. Nasal onsets were included to provide a baseline for assessing differences in the F0 trajectories [13]. There were 19 items with long vowels /ɑː oː iː/ and 29 with short vowels /a u i ʉ/. All disyllables bore the grave accent (accent 2).

2.3. Procedure

Speakers were recorded in a sound-attenuated room in the Multilingualism Lab of the Centre for Research in Bilingualism at Stockholm University, using an audio-Technica AT 3505 microphone at a sampling rate of 44.1 kHz. Words were distributed across four blocks such that word triplets did not appear in the same block. Each word within each block was repeated five times, with the order of words within each block pseudo-randomized so that repetitions of a word did not appear consecutively. Block presentation across participants followed a latin-square order. Items were displayed on a monitor and the researcher controlled the presentation.

2.4. Annotation

Recordings were manually annotated and stored as an EMU speech database [17]. Items were annotated to indicate oral (c)losure and (o)pen intervals and points indicating the onset (v) and possible cessation (cv) of voicing during the closure, along with a point indicating the onset or resumption (r) of voicing following the release burst (see Figs. 1-2).

2.5. Analysis

In what follows, we use different terms to distinguish the *phonological* voicing specification of an onset (*voiced* = /b d/, *voiceless* = /p t/) from its *phonetic* realization (*prevoiced* = [b d], *devoiced* = [b̥ d̥], *aspirated* = [$\mathrm{p^h}$ $\mathrm{t^h}$]). While in theory /p t/ could be realized as [b d], no tokens of /p t/ showed evidence of voicing during the closure in our data. Duration of voicing preceding the release burst was measured for each token containing a phonetically prevoiced plosive [b d] as either the time of the interval *c*, or the time of the difference between the points *cv* and *v*, whichever was smaller. In addition, we calculated the After Closure Time (ACT) [18] as the time from closure release (start time of *o* interval) to the onset of regular glottal pulsing of the following vowel (time of point *r*; see Figs. 1-2). For voiceless aspirated plosives, this corresponds to VOT, but while VOT for a prevoiced plosive may be negative, its ACT may be positive (as in Fig. 1). F0 was measured at 5 msec intervals throughout the open phase using the `ksvF0` estimator in the *wrassp* package [19].

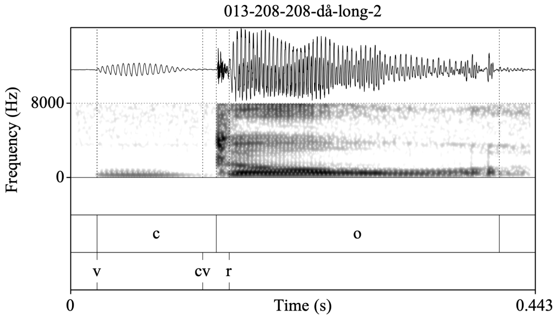

Figure 1: Example of prevoiced *då* 'then', speaker F13, repetition 2, illustrating both (negative) voice lead and (positive) voice lag. See Sec. 2.4 for explanation of labels.

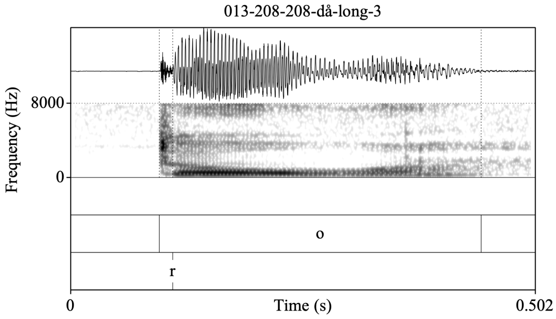

Figure 2: Example of devoiced *då* 'then', speaker F13, repetition 3, illustrating post-release voice lag but no voice lead.

3. RESULTS

3.1. Proportion of prevoicing

Out of 3282 tokens of phonologically voiced plosives (/b d/), slightly over half (1842) had at least

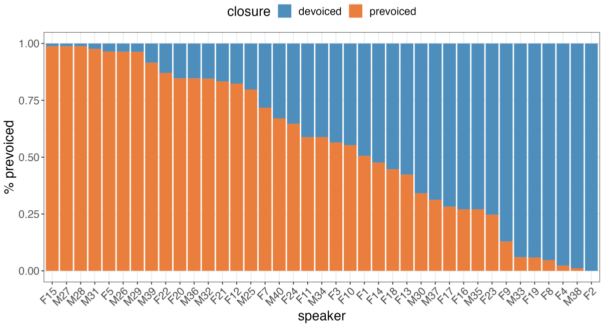

Figure 3: Proportion of /b d/ tokens realized with at least some prevoicing by speaker.

some closure voicing; we refer to these tokens as *prevoiced*. As shown in Fig. 3, however, the distribution of prevoicing across speakers is not uniform: some speakers prevoiced nearly every instance of every /b d/ token, while others consistently devoiced them. Visual examination did not reveal any obvious effects of lexical item, although impressionistically, tokens produced with closure voicing by primarily "devoicing" speakers were often coronals with high vowels (e.g. *dill, dimm, dipp*).

To explore the factors influencing devoicing rates, we fit a generalized linear mixed model (GLMM) with a logistic link function using the *lme4* package [20] to predict the probability of devoicing from log word *frequency*, vowel *height* (high or low), *place* of articulation of the onset (bilabial or coronal), and speaker *age* and *sex*. Interactions were explored but dropped after model criticism. Word frequencies (occurrence per million) were taken from the Swedish PAROLE corpus [21]. Random intercepts were included for items and speakers, with (uncorrelated) by-speaker random slopes for *place*, *height*, and *frequency*. Onset *place*, speaker *sex*, and vowel *height* were Helmert-coded. Only *sex* and *age* were significant predictors ($p < 0.05$) and both significantly improved model fit (*sex*: $\chi^2 = 5.55, Pr(> \chi^2) = 0.018$; *age*: $\chi^2 = 4.88, Pr(> \chi^2) = 0.027$). Conditional and marginal R^2 values computed with the `MuMIn` package [22] indicate that much of the variance is explained by the random effects ($R^2_m = 0.13, R^2_c = 0.66$), primarily due to the random term for *speaker* (without which $R^2_m = 0.14, R^2_c = 0.15$).

3.2. After Closure Time

Fig. 4 shows the distribution of prevoiced, devoiced, and voiceless plosives by place of articulation. To assess magnitude differences, we fit a generalized linear mixed-effects regression model predicting ACT from onset *place* (bilabial, coronal), *phonetic realization* (devoiced, prevoiced, aspirated), and their interaction, along with syllable *nucleus*, speaker *sex*, and the interaction of *place*, *sex*, and *phonetic realization*. Post-hoc pairwise comparisons of the estimated marginal means [23] showed a significant difference of 55 msec in ACTs between devoiced and aspirated plosives for both males ($SE = 2.1, z = 25.2$) and females ($SE = 2.1, z = 26.9$, both $p < 0.001$), but ACTs of devoiced and prevoiced plosives differed only for females and then only by 3 msec ($SE = 0.9, z = 3.6, p = 0.001$). Differences (averaged over sex) between levels of *place* were small but significant for devoiced (9 msec, $SE = 2.9, z = -2.9, p = 0.004$) and prevoiced (7 msec, $SE = 3, z = -2.5, p = 0.01$), but not voiceless (aspirated) plosives.

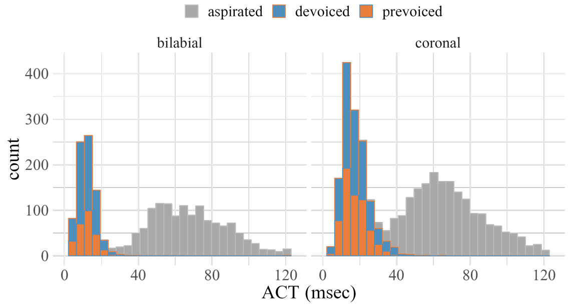

Figure 4: Distributions of After Closure Time (in msec) by phonetic realization of obstruent.

3.3. F0 trajectories

A generalized additive mixed model (GAMM: [24]) was fit to the speaker-centered F0 values, with *phonetic realization* (prevoiced, devoiced, aspirated, or nasal) of the onset as a predictor variable and smooth terms for F0 trajectory over the open phase for each of the four onset types. Item- and speaker-level differences were captured using factor smooths, analogous to random slopes and intercepts in a linear mixed model. A scaled-t family was selected on account of the heavy-tailed distribution of residuals.

Deviance explained by the GAMM model was 49.6%. Fig. 5 shows the predicted trajectories over the open phase (measured from closure release rather than vowel onset). The difference smooths (Fig. 6) show that while F0 following devoiced [b̥ d̥] and aspirated [pʰ tʰ] are significantly different for most of their trajectories, the same is not true of devoiced [b̥ d̥] and prevoiced [b d]: while there is a small difference immediately following release, the trajectories rapidly converge and remain similar throughout the remainder of the vowel. A similar pattern is observed for the difference smooth between devoiced [b̥ d̥] and the baseline nasals [m n].

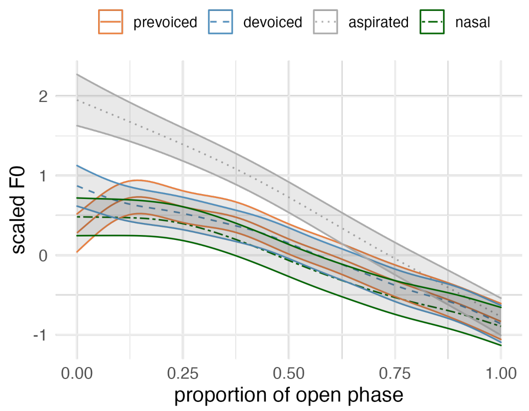

Figure 5: GAMM predicted F0 trajectories by phonetic realization of onset.

4. DISCUSSION

Our results indicate that in CSS citation forms, realization of /b d/ as [b̥ d̥] is common (Q1), particularly (but by no means exclusively) in the speech of female talkers, a trend previously reported elsewhere [4, 5], cf. [1, 2, 3]. However, devoicing of /b d/ did not substantially affect either the realization of ACTs (Q2) nor of the following F0 trajectories (Q3). Barring a small perturbation immediately following closure release (also documented for English by [25]), F0 trajectories following devoiced [b̥ d̥] are comparable to those following prevoiced [b d] and nasal [m n], consistent with previous reports of the (lack of) effect of devoicing on co-intrinsic F0 (e.g. [26].) These findings are what we would expect if the CF0 dichotomy is primarily driven by gestures implemented to suppress voicing during the closure phase of voiceless plosives [13, 15, 16].

Over 50 years ago (at the 7th ICPhS), Jan Lindqvist presented fiberoptic data showing that while Swedish lenis plosives were often acoustically devoiced utterance-initially, the glottis was nonetheless invariably in a voicing position [4]. The acoustic devoicing we observe here is thus likely the result of a failure to overcome the Aerodynamic Voicing Constraint [6, 27] rather than a categorical difference in articulatory posture. Taken together with Lindqvist's findings, the present results suggest that the Swedish plosive contrast may indeed be phonetically overspecified [1, 2], but they also demonstrate that presence or absence of glottal pulsing during the closure is not a reliable acoustic-phonetic correlate of phonological voicing in this language. Rather, the co-intrinsic F0 dichotomy may be a more reliable indicator, since it persists regardless of whether or not closure voicing is achieved. This may also be true of other so-called "true voicing" languages in which spontaneous devoicing is common such as Tokyo Japanese [26], Dutch [28, 29], and European Portuguese [30].

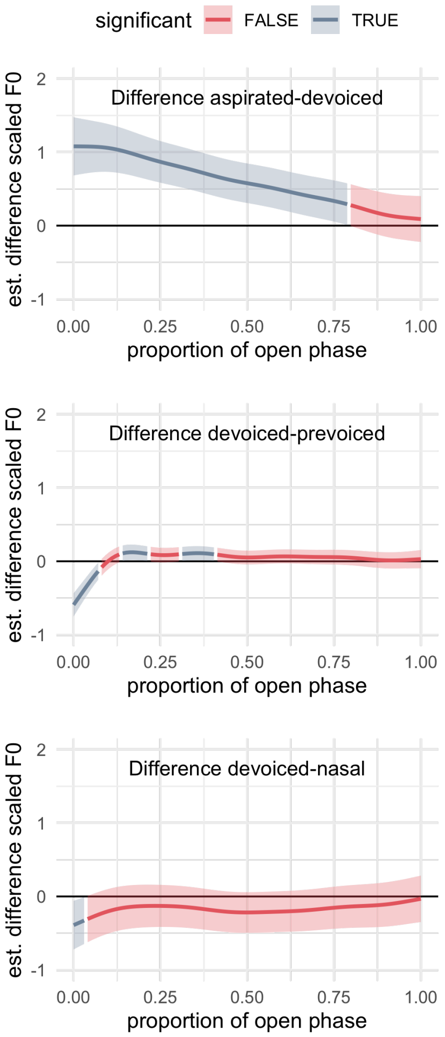

Figure 6: Difference smooths for F0 between different onset types.

5. ACKNOWLEDGMENTS

This work was funded in part by ERC Starting Grant 758605 to James Kirby and a Helge Ax:son Johnson Foundation grant to Maryann Tan. Thanks to Nicole Benker for assistance with data annotation.

6. APPENDIX: WORDLIST

bål, ball, bar, bil, biff, borr, buss, då, damm, dal, dill, dina, ding, dipp, ditt, dopp, dugga, mål, mall, mat, mil, mitt, morr, must, nå, namn, nav, nick, Nina, norr, nunna, Pål, pall, par, pil, piff, porr, puss, tå, tand, tal, till, ting, tipp, titt, topp, tugga, Tina.

7. REFERENCES

[1] Beckman, J., Helgason, P., McMurray, B., Ringen, C. 2011. Rate effects on Swedish VOT: Evidence for phonological overspecification. *Journal of Phonetics* 39(1), 39–49.

[2] Helgason, P., Ringen, C. 2008. Voicing and aspiration in Swedish stops. *Journal of Phonetics* 36(4), 607–628.

[3] Karlsson, F., Zetterholm, E., Sullivan, K. P. H. 2004. Development of a gender difference in Voice Onset Time. *Proceedings of the 10th Australian International Conference on Speech Science Technology* 316–321.

[4] Lindqvist, J. 1972. Laryngeal articulation in Swedish. *Proceedings of the seventh International Congress of Phonetic Sciences* 361–365.

[5] Keating, P. A., Linker, W., Huffman, M. 1983. Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics* 11, 277–290.

[6] Ohala, J. J. 1983. The origin of sound patterns in vocal tract constraints. In: MacNeilage, P. (ed), *The Production of Speech*. New York: Springer Verlag, 189–216.

[7] Meyer, E. A. 1897. Zur Tonbewegung des Vokals im gesprochenen und gesungenen Einzelwort. *Phonetische Studien (Beiblatt zu der Zeitschrift die neuren Sprachen)* 10, 1–21.

[8] House, A. S., Fairbanks, G. 1953. The influence of consonant environment upon the secondary acoustic characteristics of vowels. *The Journal of the Acoustical Society of America* 25(1), 105–113.

[9] Di Cristo, A., Hirst, D. J. 1986. Modelling French micromelody: Analysis and synthesis. *Phonetica* 43(1-3), 11–30.

[10] Kirby, J., Ladd, D. R. 2016. Effects of obstruent voicing on vowel F0: evidence from “true voicing” languages. *Journal of the Acoustic Society of America* 140(4), 2400–2411.

[11] Dmitrieva, O., Llanos, F., Shultz, A. A., Francis, A. L. 2015. Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics* 49, 77–95.

[12] Kingston, J., Diehl, R. L. 1994. Phonetic knowledge. *Language* 70(3), 419–454.

[13] Hanson, H. M. 2009. Effects of obstruent consonants on fundamental frequency at vowel onset in English. *The Journal of the Acoustical Society of America* 125(1), 425–441.

[14] Honda, K., Hirai, H., Masaki, S., Shimada, Y. 1999. Role of vertical larynx movement and cervical lordosis in F0 control. *Language and Speech* 42(4), 401–411.

[15] Halle, M., Stevens, K. N. 1971. A note on laryngeal features. *MIT Research Laboratory of Electronics Quarterly Research Report* 101, 198–213.

[16] Löfqvist, A., Baer, T., McGarr, N. S., Story, R. S. 1989. The cricothyroid muscle in voicing control. *The Journal of the Acoustical Society of America* 85, 1314–1321.

[17] Winkelmann, R., Harrington, J., Jänsch, K. Sep 2017. EMU-SDMS: Advanced speech database management and analysis in R. *Computer Speech Language* 45, 392–410.

[18] Mikuteit, S., Reetz, H. 2007. Caught in the ACT: The timing of aspiration and voicing in East Bengali. *Language and Speech* 50(2), 247–277.

[19] Bombien, L., Winkelmann, R., Scheffers, M. 2022. wrassp: an R wrapper to the ASSP Library. R package version 1.0.2.

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

[21] Borin, L. 2014. The Swedish Parole corpus. http://hdl.handle.net/11372/LRT-421.

[22] Bartoń, K. 2022. MuMIn: Multi-model inference. R package version 1.47.1.

[23] Lenth, R. V. 2022. emmeans: Estimated marginal means, aka least-squares means. R package version 1.8.2.

[24] Wood, S. N. 2017. *Generalized Additive Models: An Introduction with R*. New York: Chapman and Hall/CRC 2 edition.

[25] Xu, Y., Xu, A. 2021. Consonantal F0 perturbation in American English involves multiple mechanisms. *The Journal of the Acoustical Society of America* 149(4), 2877–2895.

[26] Gao, J., Arai, T. 2019. Plosive (de-)voicing and f0 perturbations in Tokyo Japanese: Positional variation, cue enhancement, and contrast recovery. *Journal of Phonetics* 77, 100932.

[27] Ohala, J. J. 2011. Accommodation to the aerodynamic voicing constraint and its phonological relevance. *Proceedings of The 16th International Congress of Phonetic Sciences* Hong Kong 64–67.

[28] van Alphen, P. M., Smits, R. 2004. Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: the role of prevoicing. *Journal of Phonetics* 32(4), 455–491.

[29] Pinget, A.-F., Kager, R., Van de Velde, H. 2020. Linking variation in perception and production in sound change: evidence from Dutch obstruent devoicing. *Language and Speech* 63(3), 660–685.

[30] Pape, D., Jesus, L. M. 2015. Stop and fricative devoicing in European Portuguese, Italian and German. *Language and Speech* 58(2), 224–246.