10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

Effects of prosodic prominence on obstruent-intrinsic F0 and VOT in German

James Kirby[1], *Felicitas Kleber*[2], *Jessica Siddins*[2], *Jonathan Harrington*[2]

[1]Linguistics and English Language, University of Edinburgh, Scotland (U.K.)
[2]Institute of Phonetics and Speech Processing, Ludwig-Maximilians-Univeristät, Munich, Germany

j.kirby@ed.ac.uk, kleber@phonetik.uni-muenchen.de, jessica@phonetik.uni-muenchen.de, jmh@phonetik.uni-muenchen.de

Abstract

We consider how lexical stress and phrasal accent influence the acoustic realization of cues to phonological voicing in German plosives. 22 native speakers of Standard German were recorded producing a total of 3168 utterances in both strong (stressed/focused) and weak (unstressed/unfocused) prosodic contexts, while holding prosodic domain constant. Both Voice Onset Time (VOT) and obstruent-intrinsic F0 (CF0) were analyzed. We found that differences in the magnitude of CF0 between voiced and voiceless plosives were greatest in the strong prosodic context, but were not always obliterated in the weak prosodic context. However, individual differences were also observed, with speakers broadly patterning into four groups with respect to the interaction of micro- and macroprosody. VOT differences were also more pronounced in strong prosodic contexts. We consider the implications of our findings for sound changes involving the reanalysis of obstruent-intrinsic F0.

Index Terms: German, intrinsic F0, microprosody, stress, accent, tone

1. Introduction

When perceiving an F0 contour, listeners track both the gross movements of the global intonation contour (macroprosody) while also calculating the effects of segmental perturbations (microprosody) [1]. Microprosodic effects are of two primary types: CF0, whereby vowels following voiceless consonants are usually higher than those following voiced consonants or sonorants [2, 3]; and VF0, whereby high/close vowels tend to have higher F0 than low/open vowels [4, 5].

While frequently smoothed over in intonation research [6], both CF0 and VF0 are relevant for the perception of segmental contrasts [7, 8, 9]. There is also considerable evidence that microprosodic effects—primarily, but not exclusively, CF0—correlate with tonogenesis and subsequent tonal splits [3, 10]. In the simplest cases, a laryngeal contrast that is signaled primarily through presence vs. absence of voicing (*ba* $\sim$ *pa*) may come instead to be signaled by a difference in F0 height (*pà* $\sim$ *pá*). This process is well-documented in Southeast Asian languages such as Kammu [11] but has also been found in Indo-European languages such as Punjabi [12] and Afrikaans [13].

Despite the ubiquity of this shift, it remains unclear exactly how it takes place. Presumably, some listener/learner needs to arrive at a novel parse of the signal, either due to error [14], mode of perception [15], or ambiguity of the coarticulatory source [16, 17]. This leads us to ask if there are contexts in which the salience of microprosodic cues might increase or even come to dominate the perceptual parse.

1.1. Contextual sources of microprosodic variation

One variable that we might expect to influence the acoustic realization of microprosody is *prosodic strength*, either at the level of the word (stress) or phrase (accent). If microprosody is enhanced in strong prosodic contexts, such as a stressed syllable under focus, this may serve to enhance its perceptual salience. Alternatively, listeners may come to give microprosodic effects greater perceptual weight in weak prosodic contexts, such as an unstressed syllable in an unfocused position, due to the greater difficulty of associating coarticulatory source and effect [17]. But before we can ask whether, or how, parsing of microprosody is affected by prosodic strength, we need a better understanding of how micro- and macroprosody interact acoustically.

Previous work suggests two competing hypotheses for how prosodic strength might influence the acoustic realization of intrinsic F0. In some (primarily non-tonal) languages, CF0 effects have been found to be most prominent in high-pitch environments [18, 19, 20], suggesting that CF0 will be enhanced in strong prosodic contexts. The physiological basis for this would be gestural overlap between the vocal-fold stiffening gesture for the (voiceless) obstruent and the high F0 target for the following vowel [21]. In other (primarily tonal) languages, the magnitude and temporal extent of CF0 effects have been found to be most prominent in low-pitch environments, such as when co-occurring with low or rising lexical tones [22, 23] (but cf. [20]). If CF0 is fundamentally an aerodynamic effect [23, 24], this would make sense: when intended pitch is low, the vocal folds will be lax, and aerodynamic factors will be able to exert a more noticeable influence.

1.2. The present study

Here, we consider how prosodic strength influences the realization of two acoustic cues, CF0 and Voice Onset Time (VOT), in Standard German, a language which contrasts long-lag and short-lag voiceless plosives. We selected German for two reasons: the existence of documented CF0 effects [25, 26] and its use of F0, among other cues, to mark prosodic prominence [27]. Furthermore, although the direction of the global F0 contour has been shown to affect microprosody in German [25], we are not aware of any studies that have compared the realization of CF0 in accented and unaccented contexts.

Based on previous work on English [18], and given previous work showing that F0 is higher in lexically stressed and phrasally accented syllables in German [28, 29], we expect CF0 effects in German to be most visible in stressed syllables under prosodic focus, and less visible in unstressed syllables that do not bear phrasal accent.

Because CF0 often takes over from Voice Onset Time (VOT) as the primary acoustic cue to a laryngeal contrast, we also consider how VOT distributions differ in these two con-

10.21437/SpeechProsody.2020-43

texts. Longer VOTs for voiceless (aspirated) stops in stressed syllables have been observed in several languages, including German [28]. This effect is often attributed to a general tendency for hyperarticulation and lengthening of segments in stressed or accented syllables [30] and/or boundary-related strengthening of the glottal abduction gesture, although (at least in English) the effect may also interact with lexical stress (see [31] for a recent review). For German, there is some evidence that VOT is affected by prosodic boundary strength [32], but perhaps not so much by phrasal accent [33]. As for CF0, we predict enhancement of VOT for voiceless plosives in stressed and accented syllables when prosodic domain is controlled for.

Note that because stress and accent are not fully crossed, this design will not allow us to unambiguously attribute any effect to the presence/absence of either stress *or* accent. However, as our primary interest is in seeing how CF0 and VOT behave under different prosodic strengths, we opted for a 1 x 2 design with the two possible prosodic extremes to provide optimal environments for the effect, if any, to emerge.

2. Methods and materials

2.1. Participants

22 native speakers of German (14 female, 8 male) aged between 19 and 36 participated in the experiment. All speakers were from Munich and surroundings and spoke a southern variety of standard German.

2.2. Speech materials

We selected 24 Standard German lexical items containing phonologically voiced and voiceless target syllables at three places of articulation (Table 1). All items were embedded in both accented and unaccented carrier phrases: "Anna/Timo wollte **X** sagen" ("Anno/Timo wanted to say **X**") or "**Kerstin/Martin** wollte X sagen" ("**Kerstin/Martin** wanted to say X"), where X represents the target and the bold-faced word carries the phrasal accent. We hereafter refer to stressed items in the accented position as the *strong* context, and to unstressed (or secondarily stressed) items in the unaccented position as the *weak* context. Each phrase was produced 3 times with each name, for a total of 144 utterances per speaker.

Table 1: *Lexical items used in production experiment. Target segments are shown in italics.* Strong *targets head lexically stressed syllables;* weak *targets head lexically unstressed/secondarily stressed syllables.*

Place	Strong		Weak	
p/b	*p*arke(n)	ˈpaʁkə	*P*arkett	paʁˈkɛt
	*B*arke	ˈbaʁkə	*B*aguette	baˈgɛt
	*p*acken	ˈpakn̩	aus*p*acken	ˈaʊspakn̩
	*b*acken	ˈbakn̩	aus*b*acken	ˈaʊsbakn̩
t/d	Quar*t*ett	kvaʁˈtɛt	Ka*t*er	ˈkaːtɐ
	Ka*d*ett	kaˈdɛt	Ka*d*er	ˈkaːdɐ
	Pro*t*est	pʁoˈtɛst	Bo*t*e	ˈboːtə
	Po*d*est	poˈdɛst	Bo*d*e	ˈboːdə
g/k	*K*ern	ˈkeʁn	im*k*ern	ˈimkɐn
	*g*ern	ˈgeʁn	zö*g*ern	ˈt͡søːgɐn
	*K*amm	ˈkam	*K*amele	kaˈmeːlə
	*G*arn	ˈgaʁn	*G*arnele	gaʁˈneːlə

2.3. Recording procedures and data pre-processing

Recordings were made in the sound-attenuated booth at the LMU Institute of Phonetics and Speech Processing in Munich using the SpeechRecorder software [34] and a Sennheiser USB headset at a sampling rate of 44.1 kHz. Prior to the recording, participants were familiarized with each item, in order to ensure that sentences would be produced with the desired accent profile. During the recording session, sentences were presented in randomized order on a computer screen, and speakers read each sentence aloud once. Participants were asked to repeat a sentence when it was obviously mispronounced and/or when the researcher noted a deviation from the intended prominence pattern. 3168 utterances were recorded in total (22 speakers x 144 tokens) which were then automatically segmented on two different tiers into words and segments using WebMAUS [35] and stored as an EMU speech database [36].

2.4. Acoustic analysis

The closure phase of the target obstruent and the following open phase that included the stop's VOT (always positive) and the postconsonantal vowel, as well as the onset of periodic voicing, were annotated on two further tiers using the EMU-webApp [36]. In some cases in the accented condition, participants would insert a pause before the target item; if the annotator judged a pause to have introduced a prosodic boundary, the item was not analysed (6%, 189 out of 3168 utterances). This was to ensure that the context for all targets syllables was comparable (utterance-medial, rather than some utterance-medial and others post-pausal). F0 estimates were made at 5 ms intervals using the `ksvf0` pitch estimator of the `wrassp` package [37]. Speaker-scaled F0 (z-scores) were then calculated across all voiced frames of the syllable nucleus, and each target nucleus was length-normalized to 21 timepoints.

3. Results

All coefficient estimates and standard errors were estimated using linear mixed-effects models with predictors VOICE (voiced, voiceless), PLACE (labial, alveolar, velar), CONTEXT (strong, weak) and SEX (male, female) and all 2-, 3- and 4-way interactions. Reported models included subject- and item-specific intercepts and by-subject random slopes for VOICE and CONTEXT. We report pairwise comparisons of the model-based estimated marginal means [38].

3.1. Global effect of context manipulation

As a check on whether the manipulation had the intended effect of altering prosodic prominence, we first compared the difference in the duration of the nucleus and F0 at 50% of the nucleus for each utterance by prosodic context. The model-estimated difference in marginal means of duration between *strong* and *weak* contexts was 44 ms (SE 13 ms) while the difference in F0 was 44 Hz (SE 6 Hz).This suggests that our manipulation had the desired effect on the local F0 context: items in the *strong* context were in fact produced as stressed/accentuated, with longer voicing lags and higher F0.

3.2. Effects on CF0

Fig. 1 shows the speaker-centered F0 excursions by context and voicing, averaged over speakers, items, and repetitions. To aid interpretation, we used raw F0 differences between contexts over the first 25% of the nucleus for modeling, but results us-

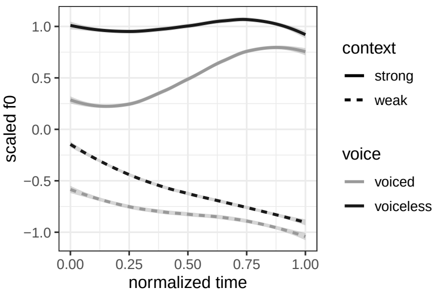

Figure 1: *Loess-smoothed F0 excursions by voicing and context, averaged over items, speakers, and repetitions.*

ing speaker-scaled F0 are comparable. In the *strong* context, F0 was on average 20 Hz higher following voiceless plosives than following voiced plosives $(\mathrm{SE} = 4.52, \mathrm{df} = 16, t = 4.52, p < 0.001)$. In the *weak* context, the average difference was just 9 Hz $(\mathrm{SE} = 4.53, \mathrm{df} = 16.2, t = 2.04, p = 0.08)$. The difference was smallest for alveolars (2 Hz in weak context, 17 Hz in strong) but comparable for labials and velars (12-13 Hz in weak, 21-22 Hz in strong).

We also found individual differences in the magnitude of the CF0 difference across contexts as well as the overall effect of context. Fig. 2 shows F0 excursions for four individual participants, chosen to illustrate the four types of qualitative patterns in the F0 data. *Type I* (illustrated by speaker VP10) shows the most common case: a larger CF0 effect in the strong than in the weak context (13 speakers). *Type II* (illustrated by speaker VP09) shows a CF0 effect of roughly comparable magnitude in both prosodic contexts, as well as an overall difference between contexts (5 speakers). *Type III* (speaker VP21) shows a comparable difference in CF0 within contexts, but little difference between contexts (3 speakers). Finally, *Type IV* (speaker VP22) provides the sole example in our data where there is little CF0 difference in either context, but a large between-context difference (1 speaker).

3.3. Effects on Voice Onset Time

Fig. 3 illustrates the effects of our stress/accent manipulation on VOT. Within PLACE and CONTEXT, the expected differences between voiced and voiceless plosives are observed. On average, VOT for voiceless plosives was 54 ms longer than VOT for voiced plosives when stressed/accented $(\mathrm{SE} = 2.43, \mathrm{df} = 28.1, t = 22.14, p < 0.0001)$ and 29 ms longer when unstressed/unaccented $(\mathrm{SE} = 2.42, \mathrm{df} = 27.8, t = 12.13, p < 0.0001)$. Averaging over PLACE and SEX, VOT was longer when accented for voiceless plosives $(\beta = 24.67, \mathrm{SE} = 2, \mathrm{df} = 18.4, t = 12.25)$ but not appreciably for voiced plosives $(\beta = -0.03, \mathrm{SE} = 2, \mathrm{df} = 18.4, t = 0.01, p = 0.99)$. Note that phonologically voiced plosives were always realized with voicing lag between closure release and vowel onset.

3.4. VOT-CF0 covariance

Fig. 4 shows the relationship of VOT to mean speaker-scaled CF0 over the first 25% of the nucleus. For ease of visualization, we averaged over repetitions of each item by speaker. In the *strong* context, there was a weak negative correlation between VOT and CF0 for both voiced (Pearson's $r = -0.22$) and voiceless $(r = -0.24)$ plosives. No meaningful correlations were observed in the *weak* contexts for either voiced

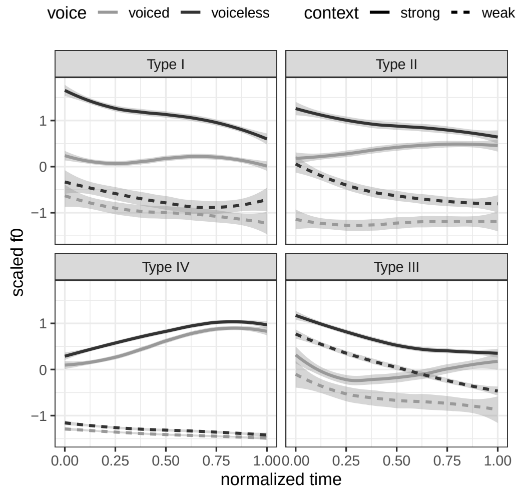

Figure 2: *Exemplars of F0 patterns in our data sample. Clockwise from top left: (I) comparable difference across voicing and context; (II) overall effect of context but larger CF0 effect in strong context; (III) overall difference in context, but little CF0 effect in either context; (IV) comparable CF0 effect across contexts, but little overall difference in context.*

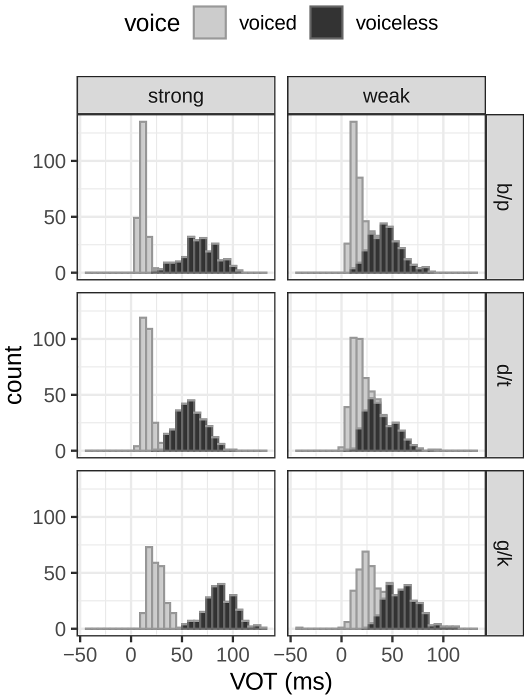

Figure 3: *VOT distributions by place, voice, and context.*

$(r = 0.03)$ or voiceless $(r = 0.08)$ plosives.

To assess the extent to which the relative cue weights differed between contexts, we computed Cohen's d (difference in means divided by average standard deviation) for VOT and mean speaker-scaled CF0 over the first 25% of the nucleus by speaker and context. On average, VOT became over three times less informative in distinguishing voicing in the *weak* context

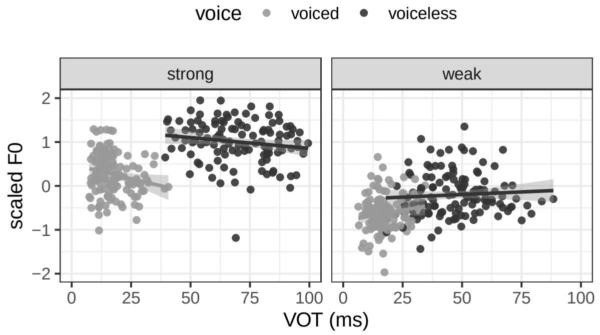

Figure 4: *VOT-CF0 covariance by voicing and context. Each data point represents the VOT and speaker-scaled CF0 for a single item and speaker, averaged over repetitions of that item. Lines indicate linear trends modeled by voicing and context.*

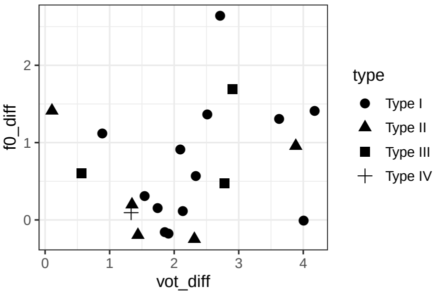

Figure 5: *Differences in Cohen's* d *for VOT and CF0 (*strong *minus* weak *contexts) by qualitative CF0 pattern (see Fig. 2).*

(mean reduction $= 2.19$, SD $= 1.1$) than CF0 (mean reduction $= 0.62$, SD $= 0.66$). That is, while both VOT and CF0 become less informative in distinguishing [$\pm$voice] in the *weak* context, the reduction is much greater for VOT than for CF0.

Next, we computed the difference in Cohen's d between contexts, to try and identify speakers who showed a large difference in VOT, but little or no difference in CF0 between contexts. The results are shown in Fig. 5. Points near the ordinate zero line are speakers for whom F0 weights in the *weak* context were comparable to those in the *strong* context. While we observe a sizable cluster of speakers around the ordinate zero, they are not exclusively members of one of our qualitatively assessed CF0 patterns. This indicates that, although a speaker might produce CF0 of similar magnitude in both contexts, this does not directly correlate with the magnitude of their VOT reduction across contexts.

4. Discussion

Our first goal with this work was to see how prosodic strength influenced the realization of two acoustic cues to voicing in German, VOT and CF0. We hypothesized that both cues would be enhanced in strong prosodic contexts. Our data are broadly consistent with this hypothesis. For most speakers, CF0 differences between voiced and voiceless plosives were more pronounced in the *strong* contexts (Type I in Fig. 2), although a number of speakers also showed differences of a similar magnitude in the *weak* context (Type II). VOT differences were also more pronounced in strong contexts, consistent with previous work showing lengthened VOT for voiceless plosives in prosodically prominent contexts in English [39, 40].

This study was also motivated by our desire to be able to inform predictions about how perceptual parsing of CF0 (and VOT) might interact with prosodic context. Based on our current results, we predict that German listeners will be more likely to compensate for microprosody when parsing F0 in prosodically strong contexts, because the perturbations are greater in these contexts, where CF0 is also (weakly) correlated with VOT (Sec. 3.4). In weak prosodic contexts, on the other hand, perceptual compensation should be less apparent.

While the effectiveness of both VOT and CF0 for distinguishing between [$\pm$voice] were reduced in the weak context, the reduction was far greater for VOT than for CF0. Thus, these results point to the potential for cue reweighting between the two prosodic contexts such that the importance of the primary VOT cue in relation to the secondary CF0 cue for the separation between [$\pm$voice] was diminished in the prosodically weak context. This type of cue reweighting is just the scenario that has the potential to lead to sound change, as demonstrated by the recent work of Beddor and colleagues [13, 41].Although we do not think that German is undergoing such a change, the fact that there is a greater reduction in the informativeness of VOT compared to CF0 also suggests that tonogenetic sound changes, where pitch cues take on a more prominent role, may be more likely to be initiated in weak prosodic contexts.

Our initial hypothesis—that CF0 would be most visible in phrasally accented stressed syllables—predicted the Type I CF0 pattern (Fig. 2), which was indeed the most frequent. Accounting for the Type II and Type III patterns is somewhat trickier. One possibility is that in the weak context, some speakers were actively enhancing the CF0 difference to compensate for the effects of the prosodic context, independent of the degree to which they employed F0 to indicate prominence more generally. Conversely, the Type IV speaker may simply not have employed whatever combination of laryngeal maneuvers gives rise to CF0 effects in the first place, or may have timed them differently relative to the closure release, leaving behind a small aerodynamic effect attenuated by the global F0 program [25]. Finally, we would not be surprised if differences in overall effect of context were due to individual differences in the timing/anchoring of the intonational pitch accent. These possibilities, as well as the extent to which individual differences in production might predict different degrees of compensation in perception, are topics for future research.

5. Conclusions

When produced in a strong prosodic context, German voiceless aspirated plosives had longer VOTs and the following vowels higher F0. Conversely, VOTs were shorter, and F0 differences less pronounced, when the prosodic context was weak. However, for some individuals, F0 differences were comparable in both prosodic contexts. Our results suggest that weak prosodic contexts may be an environment in which, all else being equal, listeners may be more likely to reinterpret intrinsic F0 perturbations as primary cues to a segmental contrast.

6. Acknowledgements

Thanks to Brisa Speier-Brito for assisting with annotation of the recordings. This research was funded by ERC grant 758605 to J. Kirby, DFG grant KL2697/1-1 to F. Kleber, and an LMU Center for Advanced Studies grant to J. Harrington.

7. References

[1] K. E. A. Silverman, "F0 segmental cues depend on intonation: the case of the rise after voiced stops," *Phonetica*, vol. 43, pp. 76–91, 1986.

[2] A. S. House and G. Fairbanks, "The influence of consonant environment on secondary acoustical characteristics of vowels," *Journal of the Acoustical Society of America*, vol. 25, pp. 105–113, 1953.

[3] J.-M. Hombert, J. J. Ohala, and W. G. Ewan, "Phonetic explanations for the development of tones," *Language*, vol. 55, no. 1, pp. 37–58, 1979.

[4] I. Lehiste and G. E. Peterson, "Some basic considerations in the analysis of intonation," *Journal of the Acoustical Society of America*, vol. 33, pp. 419–425, 1961.

[5] D. H. Whalen and A. G. Levitt, "The universality of intrinsic F0 of vowels," *Journal of Phonetics*, vol. 23, pp. 349–366, 1995.

[6] U. D. Reichel and R. Winkelmann, "Removing micromelody from fundamental frequency contours," in *Speech Prosody 2010*, 2010.

[7] D. H. Whalen, A. S. Abramson, L. Lisker, and M. Mody, "Gradient effects of fundamental frequency on stop consonant voicing judgments," *Phonetica*, vol. 47, p. 36–49, 1990.

[8] D. H. Whalen, L. Lisker, A. S. Abramson, and M. Mody, "F0 gives voicing information even with unambiguous voice onset times," *The Journal of the Acoustical Society of America*, vol. 93, no. 4, p. 2152–2159, 1993.

[9] C. A. Fowler and J. M. Brown, "Intrinsic f0 differences in spoken and sung vowels and their perception by listeners," *Perception & Psychophysics*, vol. 59, no. 5, pp. 729–738, 1997.

[10] J. A. Matisoff, "Tonogenesis in Southeast Asia," in *Consonant types and tone*, L. Hyman, Ed. Los Angeles: University of Southern California, 1973, pp. 71–95.

[11] J.-O. Svantesson and D. House, "Tone production, tone perception and Kammu tonogenesis," *Phonology*, vol. 23, pp. 309–333, 2006.

[12] J. Kanwal and A. Ritchart, "An experimental investigation of tonogenesis in Punjabi," in *Proceedings of the 18th International Congress of Phonetic Sciences*, 2015.

[13] A. W. Coetzee, P. S. Beddor, K. Shedden, W. Styler, and D. Wissing, "Plosive voicing in afrikaans: Differential cue weighting and tonogenesis," *Journal of Phonetics*, vol. 66, p. 185–216, Jan 2018.

[14] J. J. Ohala, "The listener as a source of sound change," in *CLS 17-2: Papers from the parasession on language and behavior*, C. Masek, R. Hendrick, and M. Miller, Eds. Chicago: Chicago Linguistic Society, 1981, pp. 178–203.

[15] B. Lindblom, S. Guion, S. Hura, S.-J. Moon, and R. Willerman, "Is sound change adaptive?" *Rivista di Linguistica*, vol. 7, no. 1, pp. 5–37, 1995.

[16] P. S. Beddor, "A coarticulatory path to sound change," *Language*, vol. 85, no. 4, pp. 785–821, 2009.

[17] J. Harrington, F. Kleber, U. Reubold, and J. Siddins, "The relationship between prosodic weakening and sound change: evidence from the German tense/lax vowel contrast," *Laboratory Phonology*, vol. 6, no. 1, p. 87–117, 2015.

[18] H. M. Hanson, "Effects of obstruent consonants on fundamental frequency at vowel onset in English," *The Journal of the Acoustical Society of America*, vol. 125, no. 1, pp. 425–441, 2009.

[19] J. Kirby and D. R. Ladd, "Effects of obstruent voicing on vowel f0: evidence from 'true voicing' languages," *The Journal of the Acoustical Society of America*, vol. 140, no. 4, pp. 2400–2411, 2016.

[20] J. Kirby, "Onset pitch perturbations and the cross-linguistic implementation of voicing: Evidence from tonal and non-tonal languages," *Journal of Phonetics*, vol. 71, pp. 326–354, 2018.

[21] M. Halle and K. N. Stevens, "A note on laryngeal features," *MIT Quarterly Progress Report*, vol. 101, p. 198–212, 1971.

[22] J. T. Gandour, "Consonant types and tone in Siamese," *Journal of Phonetics*, vol. 2, pp. 337–350, 1974.

[23] C. X. Xu and Y. Xu, "Effects of consonant aspiration on Mandarin tones," *Journal of the International Phonetic Association*, vol. 33, no. 2, pp. 165–181, 2003.

[24] K. J. Kohler, "F0 in the perception of lenis and fortis plosives," *The Journal of the Acoustical Society of America*, vol. 78, no. 1, p. 21–32, 1985.

[25] ——, "F0 in the production of fortis and lenis plosives," *Phonetica*, vol. 39, p. 199–218, 1982.

[26] M. Jessen, *The phonetics and phonology of tense and lax obstruents in German*. Amsterdam: John Benjamins, 1999.

[27] M. Grice, S. Baumann, and R. Benzmüller, "German intonation in autosegmental-metrical phonology," in *Prosodic typology: The phonology of intonation and phrasing*, S.-A. Jun, Ed. Oxford: Oxford University Press, 2005, pp. 55–83.

[28] M. Jessen, K. Marasek, K. Schneider, and K. Clahßen, "Acoustic correlates of word stress and the tense/lax opposition in the vowel system of German," in *Proceedings of the 13th International Congress of Phonetic Sciences*, vol. 4, 1995, pp. 428–431.

[29] G. Dogil, "The phonetic manifestation of word stress," in *Word prosodic systems in the languages of Europe*, H. van der Hulst and B. Williams, Eds. Berlin: Mouton de Gruyter, 1999, pp. 273–334.

[30] K. de Jong, "The supraglottal articulation of prominence in English," *Journal of the Acoustical Society of America*, vol. 97, pp. 491–504, 1995.

[31] T. Cho, "Prosodic boundary strengthening in the phonetics–prosody interface," *Language and Linguistics Compass*, vol. 10, no. 3, pp. 120–141, 2016.

[32] C. Kuzla and M. Ernestus, "Prosodic conditioning of phonetic detail in German plosives," *Journal of Phonetics*, vol. 39, no. 2, pp. 143–155, 2011.

[33] P. Hoole and L. Bombien, "A cross-language study of layngeal-oral coordination across varying prodosic and syllable-structure conditions," *Journal of Speech, Language, and Hearing Research*, vol. 60, p. 525–539, Mar 2017.

[34] C. Draxler and K. Jänsch, "SpeechRecorder - a universal platform independent multi-channel audio recording software," in *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)*, Lisbon, 2004, p. 559–562.

[35] T. Kisler, U. Reichel, and F. Schiel, "Multilingual processing of speech via web services," *Computer Speech & Language*, vol. 45, pp. 326–347, 2017.

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

[37] L. Bombien, R. Winkelmann, and M. Scheffers, *wrassp: an R wrapper to the ASSP Library*, 2018, R package version 0.1.8. [Online]. Available: https://github.com/IPS-LMU/wrassp

[38] R. Lenth, *emmeans: Estimated Marginal Means, aka Least-Squares Means*, 2018, R package version 1.1.2. [Online]. Available: https://CRAN.R-project.org/package=emmeans

[39] L. Lisker and A. S. Abramson, "Some effects of context on Voice Onset Time in English stops," *Language and Speech*, vol. 10, no. 1, pp. 1–28, 1967.

[40] J. Cole, H. Kim, H. Choi, and M. Hasegawa-Johnson, "Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech," *Journal of Phonetics*, vol. 35, pp. 180–209, 2007.

[41] P. S. Beddor, A. W. Coetzee, W. Styler, K. B. McGowan, and J. E. Boland, "The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change," *Language*, vol. 94, no. 4, pp. 931–968, 2018.