Interspeech 2018
2-6 September 2018, Hyderabad, India

# Speaker-specific structure in German voiceless stop voice onset times

*Marc A. Hullebus*[1], *Stephen J. Tobin*[12], *Adamantios I. Gafos*[12]

[1]Department für Linguistik, Universität Potsdam, Germany
[2]Haskins Laboratories, New Haven, USA

hullebus@uni-potsdam.de, tobin@uni-potsdam.de, gafos@uni-potsdam.de

## Abstract

Voice onset time (VOT), a primary cue for voicing in many languages including English and German, is known to vary greatly between speakers, but also displays robust within-speaker consistencies, at least in English. The current analysis extends these findings to German. VOT measures were investigated from voiceless alveolar and velar stops in CV syllables cued by a visual prompt in a cue-distractor task. Comparably to English, a considerable portion of German VOT variability can be attributed to the syllable's vowel length and the stop's place of articulation. Individual differences in VOT still remain irrespective of speech rate. However, significant correlations across places of articulation and between speaker-specific mean VOTs and standard deviations indicate that talkers employ a relatively unified VOT profile across places of articulation. This could allow listeners to more efficiently adapt to speaker-specific realisations.

**Index Terms**: speech production, speech variability, voice onset time

## 1. Introduction

Our perceptual systems manage to categorise speech sounds with relative ease, despite the seemingly unstructured variability in natural speech production. Many characteristics of individual speech are paralinguistic and serve to facilitate talker identification. However, even acoustic parameters that are indicative of contrastive speech categories vary to a remarkable degree both across and within speakers. Spectral characteristics such as formant frequencies have been shown to vary irrespective of phonemic context and overall pitch [1]. Durational parameters, specifying differences in speech properties relating to time, are also not immune to variability. One of the most extensively studied durational parameters conveying linguistic contrasts is voice onset time (VOT) [2]. In many languages, VOT serves as a primary cue for the voicing contrast in stops (e.g. voiced versus voiceless). VOT is defined as the time in milliseconds between the burst release of the stop and the onset of periodicity indicating vocal fold vibration for the vowel. The relative timing of these two events is crucial for distinguishing stop categories, with longer VOT values mostly associated with voiceless stops and shorter or negative VOTs with voiced stops. However, VOT values can differ widely between speakers [3], with mean VOT ranges for an English velar stop for instance ranging from 52 to 80 ms across studies [4], [5]. Voiceless stops in particular tend to display more variability [1], which could possibly be attributed to the lack of a contrastive category boundary in the high VOT range as opposed to the presence of a lower boundary where the threshold for the voiced category lies. Another reason might be that the diminishing synchronicity of oral and laryngeal gestures associated with longer aspiration is associated with greater variability [6].

Despite this variability, VOT is also known to be influenced systematically by a number of well-known parameters. These can reduce some of the apparent randomness of the variability and expose some of its underlying structure. A well-known effect is that of place of articulation (PoA), with VOT generally increasing with backer PoAs [5]. Another variable influencing durational properties (as well as many other speech characteristics) is speech rate [7]. VOT is also greatly influenced by speech rate, which for sentences is often measured as the number of syllables per second. For single syllables this can be measured as the duration of the vowel. As more rapid speech leads to a compressed time frame, voice onset times will necessarily be shortened [7]. Another possible factor is hyperarticulation - a phenomenon hard to exclude in experimental settings. Hyperarticulation can result in longer speech planning times which in turn can translate into longer aspiration and consequently longer voice onset times [8]. Further potential influences are prosodic context [9] and sex differences [10].

Nevertheless, even when taking factors such as speaking rate into account, there is still much individual speaker variability that remains [11], [12]. For instance, the PoA effect does not appear as clear-cut with respect to alveolar and velar stops [13]. Although both have consistently longer VOT than bilabial stops, alveolars sometimes rank higher than velars in VOT length. As much as speakers might differ with regards to their VOT productions, there are indications that they do behave relatively consistent with respect to their VOT productions when looking at a single speaker's stop categories. Previous studies such as those of Zlatin [14], Koenig [15] demonstrated significant correlations among VOT distributions of stop categories at different PoAs, suggesting that stops across different places of articulation are produced in similar ways by single speakers.

Although speakers might produce relatively consistent VOTs across stop categories, these VOTs can be influenced by situational factors as well, such as the speech of their interlocutor. Several studies have found that speakers adjust their speech production in response to the phonetic realisation of speech that they hear. These adjustments, called "phonetic accommodation", have been observed in different phonetic parameters, including VOT. Aside from the effects of long-term exposure to different VOT ranges [16], [17], speakers tend to produce longer VOTs shortly after hearing words with longer VOTs and even generalize the lengthening of VOTs to different PoAs from those words that they have just heard [17]. The effect of such a change has often been found to be on the order of 5 ms [6], [16], [17]. Newer evidence suggests that interactions between perception and production can occur


10.21437/Interspeech.2018-2288

on the millisecond scale and can directly influence speech planning within ~150 ms of presentation [18]. The cue-distractor paradigm has been used to probe the link between perception and production [19], [20]. Participants are asked to produce a response following a corresponding visual cue, e.g. a CV syllable, but are 'distracted' during the speech planning phase immediately following the cue, by a syllable presented over headphones, that has one or multiple features in common with the target syllable, e.g. voicing or PoA. Even subphonemic features such as VOT could induce these perceptuo-motor effects, as some speakers show response time effects when distractor VOTs were manipulated [21]. A clearer understanding of which parameters most influence VOT values provides insights into what information needs to be extracted out of VOT variability.

In previous work on German VOT sample sizes were small ($6 \leq N \leq 10$) [22]–[24], making analyses of individual differences not meaningful. Using a larger number of speakers, the current analysis aims to verify whether the patterns along which speaker-specific VOT variability have been observed in English hold true for German syllables in isolation. It is hypothesised that both stop category means will be significantly correlated with each other. Furthermore, mean VOTs will also be significantly correlated with standard deviations. Further hypotheses are that the main predictor of VOT will be vowel length, that PoA will play a relatively minor role affecting VOT, and that the main source of random talker variation will be attributable to mean VOT differences between speakers. These differences should remain even when taking into account the effect of vowel length in the model.

# 2. Methods and results

## 2.1. Methods

### *2.1.1. Participants*

40 undergraduate students at Universität Potsdam, Germany, participated in the study. They were all native speakers of German with no history of auditory or speech disorders. Students received credit in return for participation. The data of 25 participants that are currently available for analysis are presented. 23 of the participants were female.

### *2.1.2. Stimuli*

The analysed VOT values were obtained from a cue-distractor dataset. Recordings were made in a sound-treated booth at a sampling rate of 44.1 KHz. The task involved participants responding to a visual cue by producing a CV-syllable consisting of either an alveolar or a velar voiceless stop and an open back unrounded vowel. This meant responding [tʰɑ] after two asterisks (**) or [kʰɑ] after two hashes (##). Additionally, participants heard a 'distractor' syllable played over headphones, 150 ms after the visual cue appeared on screen. The distractor syllable either matched the cued response syllable (match condition) or was the alternative (non-match condition). The distractors varied in VOT length along 11 steps ranging between 45-95 ms in a "low distractor" set and between 70-95 ms in a "high distractor" set. Before the data in the cue-distractor conditions were registered, participants produced the two response syllables in a baseline condition, allowing for the registration of a participant-specific baseline VOT profile. On the basis of this task, participants were grouped into either a "low VOT baseline" or a "high VOT baseline" with a cut-off of 65 ms. Each participant produced 100 syllables in the baseline task, which were randomly divided between the two stops (resulting in 50 baseline tokens per stop category). They produced an additional 720 stops in the experimental task, totalling 820 responses per participant. A total of 18663 out of 20500 responses were subjected to analysis after excluding incorrect responses, leaving on average 747 responses per participant.

In what follows, we make use of the complete set of VOT productions from each measured participant (that is, both baseline and response VOTs in the cue-distractor conditions of the experiment) rather than just the baseline task in order to maximise the size of the dataset and enable potentially more reliable results to emerge. In this analysis, we aim to uncover general patterns in VOT variability. The effect of the distractors does not lie within the scope of this analysis. An independent analysis will be devoted to assessing the existence of effects of the distractors.

### *2.1.3. Measurements*

Voice onset times were obtained both automatically with acoustic landmark detection software [24] and semi-automatically in custom labelling software in the context of validating our own automatic landmark detection software. VOTs were calculated as the time between the onset of the burst and the onset of periodicity in the waveform. Vowel length was measured using the same onset of periodicity and the time point where the periodicity ended and the waveform dropped to the level of background noise. Statistical analyses were performed in R [25], including the lme4 package developed for linear mixed models [26].

## 2.2. Results

An overview of the VOT productions as shown in Table 1 reveals VOT profiles that span over an extensive range of the VOT continuum. The largest range between VOT means is over 90 ms. The mean VOT of the velar stops is approximately 3 ms higher than the mean VOT of the alveolar stops, whereas standard deviations are similar.

### *2.2.1. Correlations between stop categories*

Previous work has found that VOTs tend to covary between stop categories [13]. Correlations between the two stop categories were computed with Pearson's product-moment correlation coefficient. All but 3 participants show a positive correlation between the two stop categories, ranging between $r = 0.046$ and 0.4499. The negative correlations ranged between $r = -0.0528$ and -0.002.

When investigating the means per participants for either of the stop categories specifically, a correlation analysis found a close to perfect significant correlation between the means for alveolar and velar PoAs ($r = 0.955$, $p < 0.001$). This is illustrated by Figure 1, containing a scatterplot with per participant VOT means for both stops. The overlaps between stop categories are also evident in Figure 2, which shows density plots for each PoA per participant.

Table 1: *Summary of descriptive statistics.*

| Stop | Mean range | SD range | Mean | SD |
|---|---|---|---|---|
| [tʰ] | 42-116 ms | 8.3-21.8 | 67 | 12.4 |
| [kʰ] | 39-131 ms | 7.71-21.76 | 70 | 12.9 |

### 2.2.2. *Correlations between means and standard deviations*

Another correlation that has been explored in previous literature concerns the link between VOT means and standard deviations within stop categories. In the current dataset, the VOT means and standard deviations for both velar and alveolar places of articulation are highly significantly correlated. For alveolar [tʰɑ], the Pearson's correlation is $r$ = 0.799 ($p<0.001$), whereas for velar [kʰɑ] it is lower ($r$ = 0.694, $p<0.001$). Both of these correlations are higher than those reported by [3] on English. Figure 3 illustrates the correlation of mean and SD for [tʰɑ].

These findings are in agreement with the well-known relation between mean and variability, i.e. higher means generally result in more variation [27]. This is apparent when viewing the distributions of individual participants in the density plots in Figure 2. The participant plots are ordered by increasing VOT means, and viewing from top left to bottom right the distributions grow increasingly heavy tails as VOT increases.

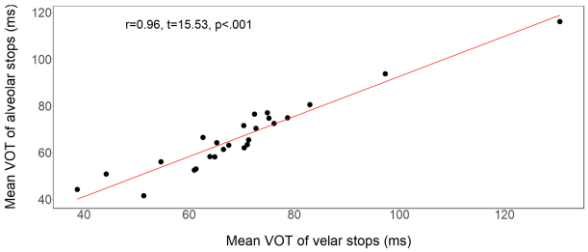


Figure 1: *Correlation of participants' mean VOTs across PoA.*

### 2.2.3. *Statistical assessment of effects on VOT*

In order to assess the relative influence of specific factors on VOT production, a linear mixed-effects analysis was performed. This is a type of regression analysis that has the advantage that both fixed and random effects can be included. Fixed effects correspond to the experimentally manipulated independent variables that can influence VOT overall, whereas the random effects are not experimentally controlled sources of variability that are, for instance, introduced by inter-speaker differences. In this case, linear mixed models allow us to interpret the factors that are known to influence VOT such as speech rate and place of articulation from the fixed effects, while the amount of variability that is introduced due to individual speaker differences can be gauged by the proportion of variance attributed to the various random effects. To account for the random effects, restricted likelihood estimation (REML) was used [29]. An effect was judged as being significant if t-values exceeded 2, but an additional adjusted Wald test [24] was performed to report tentative $\chi^2$- and p-values for the fixed effects.

The chosen model included fixed effects of speaking rate (whose proxy is taken to be vowel length), place of articulation (or response syllable; the alveolar stop is the reference category) and an interaction between the two. The random effects included intercepts for participants and speech rate and random slopes for place of articulation. Other models varied with regards to their random effects structure and were compared on the basis of Akaike's Information Criterion (AIC) [30]. A second model excluded random slopes for PoA, a third excluded the participant intercepts and PoA slopes leaving only random intercepts for vowel length, and a fourth excluded intercepts for vowel length but kept participant-specific intercepts and PoA slopes. The first (fullest) model performed significantly better on the likelihood ratio test and had lower AIC scores than all of the alternative models except the last, which excluded vowel length. Model 1 did not, however, perform significantly worse than model 4 and had a very similar AIC score (AIC=149206 vs 149208). Hence, the inclusive model was kept in order to investigate the effect of vowel length on the random effect structure. A random-effects principal component analysis (using the rePCA function from the RePsychling package [31]) revealed none of the dimensions of the random effects model to lack any variance, indicating the model structure is not too complex for the dataset under analysis. Finally, since gender is a possible influencer of VOT [10], an analysis excluding the minority of male participants was done, which showed the same pattern of results as the chosen model.

The model returned significant main effects for place of articulation and vowel length. As a baseline intercept, the model estimated 65 ms. Furthermore, the fixed effect of PoA suggests that, in our dataset, alveolar stops have significantly shorter VOTs than velars and that a change in place of articulation results in a 3 ms decrease in VOT. The place of articulation effect was also significant (t= -2.190, $\chi^2$=14.7, df=1, $p<0.001$), though not as far above the threshold value of 2 as the effect of vowel length. This vowel length effect (which we attribute to speech rate) was highly significant (t=7.86, $\chi^2$=106.7, $p<0.001$). The model estimated a 2 ms increase in VOT per 100 ms increase in vowel length, which is consistent with previous findings in English [32]. The interaction between response syllable and vowel length did not reach significance.

The random effect structure of the model attributed the biggest proportion of variance to the per-speaker intercept (SD=16.5), followed by the PoA slopes (SD=5.2) and finally the intercept for vowel length (SD=2.5). Hence, the largest differences in means are at the level of individual speakers, where intercept differences are several times larger than those for the other random effects (more than 3 times that of the PoA and almost 8 times that of vowel length). The results of the model are displayed in Table 2.

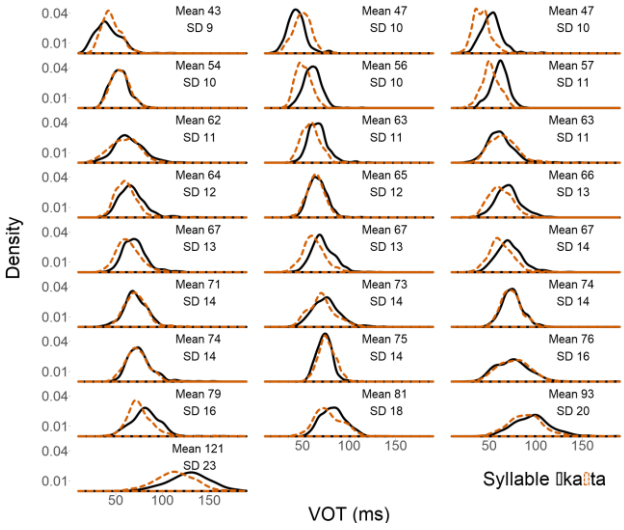


Figure 2: *Density plots per participant ordered by VOT mean, from shortest, upper left panel, to longest, bottom right panel.*

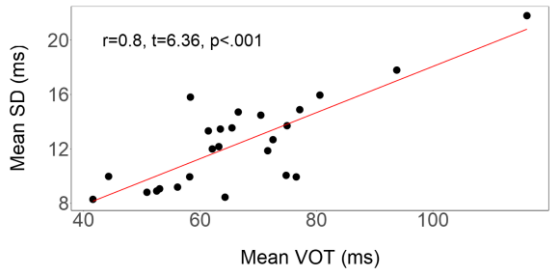


Figure 3: *Correlation of [tʰa] VOT mean and SD.*

Table 2: *Results of the linear mixed model.*

| **Random effects** | **Name** | **Variance** | **St. Dev** | **Corr** |
|---|---|---|---|---|
| **Vowel length** | Intercept | 6.3 | 2.5 | |
| **Subject** | Intercept | 269.1 | 16.4 | -0.49 |
| | PoA | 27.2 | 5.2 | |
| **Residual** | | 171.2 | 13 | |

| **Fixed effects** | **Estimate** | **Standard Error** | **t-value** |
|---|---|---|---|
| **Intercept** | 65 | 3.3 | 19.2 |
| **PoA(ta)** | -3 | 1.4 | -2.2 |
| **vowel length** | 0.02 | 0.003 | 7.8 |
| **PoA*v length** | -0.004 | 0.004 | -0.97 |

## 3. Discussion

There is remarkable variability between speakers of the same language, as can be seen in the current dataset from voiceless German stops in the large differences between participant means and in the shapes of their VOT distributions. Despite these individual differences, however, variability seems to be tempered by speaker-specific relationships. Specifically, positive correlations between speakers' places of articulation indicate that idiosyncratic VOT patterns are applied across voiceless stops, with the exception of only 3 participants. A very strong correlation between the means across places of articulation confirms this. Additionally, high correlations between the means and standard deviations in the two places of articulation provide further evidence that higher overall VOT means more variability. A possible explanation for this could be that the greater asynchronicity between oral and laryngeal gestures underlying longer VOTs is also associated with greater variability.

Our results are also in line with previous findings for English VOTs. As is expected for a durational property in speech, the largest main influencer of VOT was the length of the vowel following the stop. Previous work found vowel length to be the single most important predictor of VOT, with PoA not being significant, e.g. [11], [7]. The current analysis provides further evidence supporting the claim that an effect of speech rate on VOT is already present at the syllable level and even in mere CV-syllables and also extends this relation to German stops. In contrast to the findings of Allen et al. [11], the effect of PoA was significant in this investigation, which is consistent with velar stops generally being found to have longer VOTs. Likewise the relatively small effect size (3 ms) is consistent with the similarities found between stop category VOTs. Most of the random variability in the data is revealed by the random effects to be linked to speaker-specific differences regarding VOT intercepts. Corroborating previous findings, our results indicate that in German too, there is considerable speaker-specific variance after factoring out the influence of speech rate (vowel length) and place of articulation.

The co-occurrence of high between-speaker VOT variability on the one hand and highly correlated stop category distributions on the other hand could serve to facilitate both speaker identification and adaptation. Idiosyncratic VOT distributions could support listeners' recognition of individual speakers as well as speaker recognition algorithms. Furthermore, a predictable relationship among stops belonging to the same voicing category would allow listeners (or automatic VOT detection algorithms [25]) to make inferences about an individual speaker's stop at different PoAs by extracting a more general VOT profile from the speaker-specific realisation of a stop at just one PoA without necessarily having prior experience with that stop [32], [33]. With respect to listener and talker adaptation, such generalisations based on lower-dimensional features could facilitate greater efficiency in tuning perception or even production at fast time-scales, where evidence for speaker adaptation has been presented or sought before [21], [33]–[36].

## 4. Conclusions

The current paper has confirmed findings of a consistent structure in talkers' VOT variability. Observations of strong linear relations between the VOT values of voiceless velar and alveolar stops in English are extended to German, indicating similar VOT distributions for both stop categories. Validating prior findings about higher VOT means being associated with more variability in VOT productions, the correlation analyses also demonstrated high correlations between talker VOT means and their standard deviations. Vowel length is found to have the greatest impact on VOT length, but even taking significant factors like speech rate and place of articulation into account, the greatest portion of random variability is attributable to individual differences [13].

Future research should extend the range of stop categories considered, as well as include negative VOT ranges, which have not yet received much attention, but could potentially exhibit similar patterns to those found in aspirated stops as there is no other category boundary at the extremum of the negative VOT range (like in the positive long-lag range). Perception studies investigating speaker recognition and adaptation on the basis of speaker-specific VOT distributions would also extend our understanding of structured variation in VOT.

## 5. Acknowledgements

The authors gratefully acknowledge support by ERC Advanced Grant 249440 and by the Deutsche Forschungsgemeinschaft (DFG) Collaborative Research Centre SFB1287.

# 6. References


[1] D. H. Klatt and L. C. Klatt, "Analysis, synthesis, and perception of voice quality variations among female and male talkers", *the Journal of the Acoustical Society of America*, vol. 87, no. 2, pp. 820–857, 1990.

[2] L. Lisker and A. S. Abramson, "A cross-language study of voicing in initial stops: Acoustical measurements", *Word*, vol. 20, no. 3, pp. 384–422, 1964.

[3] P. J. Bailey and M. P. Haggard, "Perception and production: Some correlations on voicing of an initial stop", *Language and speech*, vol. 16, no. 3, pp. 189–195, 1973.

[4] D. Byrd, "54,000 American stops", *UCLA Working Papers in Phonetics*, vol. 83, pp. 97–116, 1993.

[5] T. Cho and P. Ladefoged, "Variation and universals in VOT: evidence from 18 languages", *Journal of Phonetics*, vol. 27, no. 2, pp. 207–229, 1999.

[6] S. Tobin, "Phonetic accommodation in Spanish-English and Korean-English bilinguals", *Proc. Mtgs. Acoust.*, vol. 19, no. 1, p. 060087, 2013.

[7] J. L. Miller and L. E. Volaitis, "Effect of speaking rate on the perceptual structure of a phonetic category", *Perception & Psychophysics*, vol. 46, no. 6, pp. 505–512, 1989.

[8] C. Kirov, "A Bayesian approach to speech production". Ph.D. dissertation, John Hopkins University, Baltimore, MD, USA, 2014.

[9] C. Fougeron and P. A. Keating, "Articulatory strengthening at edges of prosodic domains", *The Journal of the Acoustical Society of America*, vol. 101, no. 6, pp. 3728–3740, 1997.

[10] S. P. Whiteside and C. J. Irving, "Speakers' sex differences in voice onset time: some preliminary findings", *Perceptual and motor skills*, vol. 85, no. 2, pp. 459–463, 1997.

[11] J. S. Allen, J. L. Miller, and D. DeSteno, "Individual talker differences in voice-onset-time", *The Journal of the Acoustical Society of America*, vol. 113, no. 1, pp. 544–552, 2003.

[12] E. Chodroff and C. Wilson, "Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English", *Journal of Phonetics*, vol. 61, pp. 30–47, Mar. 2017.

[13] G. J. Docherty, The timing of voicing in British English obstruents. Walter de Gruyter, 1992.

[14] M. A. Zlatin, "Voicing contrast: perceptual and productive voice onset time characteristics of adults", *The Journal of the Acoustical Society of America*, vol. 56, no. 3, pp. 981–994, Sep. 1974.

[15] L. L. Koenig, "Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds", *Journal of Speech, Language, and Hearing Research*, vol. 43, no. 5, pp. 1211–1228, 2000.

[16] M. L. Sancier and C. A. Fowler, "Gestural drift in a bilingual speaker of Brazilian Portuguese and English", *Journal of phonetics*, vol. 25, no. 4, pp. 421–436, 1997.

[17] J. E. Flege and W. Eefting, "Cross-language switching in stop consonant perception and production by Dutch speakers of English", *Speech Communication*, vol. 6, no. 3, pp. 185–202, Sep. 1987.

[18] S. Tobin, M. A. Hullebus & A. Gafos. "Phonetic Convergence in VOT in a Cue-Distractor Paradigm", Presented at HISPhonCog 2018, Seoul, South Korea. 2018.

[19] B. Galantucci, C. A. Fowler, and L. Goldstein, "Perceptuomotor compatibility effects in speech", *Attention, Perception, & Psychophysics*, vol. 71, no. 5, pp. 1138–1149, 2009.

[20] K. D. Roon and A. I. Gafos, "Perceptuo-motor effects of response-distractor compatibility in speech: beyond phonemic identity", *Psychonomic bulletin & review*, vol. 22, no. 1, pp. 242–250, 2015.

[21] K. D. Roon, E. Klein, and A. I. Gafos, "Perceptuo-motor interactions across and within phonemic categories", in *Proceedings of the 18th international congress of phonetic sciences*, 10-14 August, 2015, Glasgow, U.K. Glasgow: the University of Glasgow, 2015.

[22] H. G. Piroth and P. M. Janker, "Speaker-dependent differences in voicing and devoicing of German obstruents", *Journal of Phonetics*, vol. 32, no. 1, pp. 81–109, 2004.

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

[24] W. K. Haag, "An articulatory experiment on voice onset time in German stop consonants", *Phonetica*, vol. 36, no. 3, pp. 169–181, 1979.

[25] [1]S. R. Kuberski, S. J. Tobin, and A. I. Gafos, "A landmark-based approach to automatic voice onset time estimation in stop-vowel sequences", in *Signal and Information Processing (GlobalSIP)*, 2016, *IEEE Global Conference on Signal and Information Processing*, 2016, pp. 60–64.

[26] R. C. Team, "R: A language and environment for statistical computing", Austria: R Foundation for Statistical Computing, Vienna, 2013.

[27] R. R. Sokal and C. A. Braumann, "Significance tests for coefficients of variation and variability profiles", *Systematic Biology*, vol. 29, no. 1, pp. 50–66, 1980.

[28] D. Bates, M. Mächler, B. Bolker, and S. Walker, Fitting Linear Mixed-Effects Models using lme4, 1–51. *Computation*, 2015.

[29] M. G. Kenward and J. H. Roger, "Small sample inference for fixed effects from restricted maximum likelihood", *Biometrics*, pp. 983–997, 1997.

[30] H. Akaike, "A new look at the statistical model identification", *IEEE transactions on automatic control*, vol. 19, no. 6, pp. 716–723, 1974.

[31] H. Baayen, D. Bates, R. Kliegl, and S. Vasishth, "RePsychLing: data sets from psychology and linguistics experiments", R package version 0.0, vol. 4, 2015.

[32] E. Chodroff, J. Godfrey, S. Khudanpur, and C. Wilson, "Structured variability in acoustic realization: A corpus study of voice onset time in American English stops", in *Proceedings of the 18th international congress of phonetic sciences*, 10-14 August, 2015, Glasgow, UK: the University of Glasgow, 2015.

[33] K. Nielsen, "Specificity and abstractness of VOT imitation", *Journal of Phonetics*, vol. 39, no. 2, pp. 132–142, 2011.

[34] M. Sato, K. Grabski, M. Garnier, L. Granjon, J.-L. Schwartz, and N. Nguyen, "Converging toward a common speech code: imitative and perceptuo-motor recalibration processes in speech production", *Frontiers in psychology*, vol. 4, p. 422, 2013.

[35] C. A. Fowler, J. M. Brown, L. Sabadini, and J. Weihing, "Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks", *Journal of Memory and Language*, vol. 49, no. 3, pp. 396–413, 2003.

[36] D. Kerzel and H. Bekkering, "Motor activation from visible speech: Evidence from stimulus response compatibility.", *Journal of Experimental Psychology: Human Perception and Performance*, vol. 26, no. 2, pp. 634–647, 2000.