

The effect of predictability on German stop voicing is phonologically selective

Omnia Ibrahim, Ivan Yuen, Bistra Andreeva, Bernd Möbius

Department of Language Science and Technology, Saarland University, Germany

omnia@lst.uni-saarland.de

Abstract

Cross-linguistic evidence suggests that syllables in predictable contexts have shorter duration than in unpredictable contexts. However, it is not clear if predictability uniformly affects phonetic cues of a phonological feature in a segment. The current study explored the effect of syllable-based predictability on the durational correlates of the phonological stop voicing contrast in German, viz. voice onset time (VOT) and closure duration (CD), using data in Ibrahim et al. [1]. The target stop consonants /b, p, d, k/ occurred in stressed CV syllables in polysyllabic words embedded in a sentence, with either voiced or voiceless preceding contexts. The syllable occurred in either a low or a high predictable condition, which was based on a syllable-level trigram language model. We measured VOT and CD of the target consonants (voiced vs. voiceless). Our results showed an interaction effect of predictability and the voicing status of the target consonants on VOT, but a uniform effect on closure duration. This interaction effect on a primary cue like VOT indicates a selective effect of predictability on VOT, but not on CD. This suggests that the effect of predictability is sensitive to the phonological relevance of a language-specific phonetic cue.

Index Terms: predictability, voicing, stress, word-medial position, German

1. Introduction

Contextual predictability has pervasive effects on the acoustic realisation of speech [2, 3, 4]. Its effect is quite common and general as evident in a survey of 600 languages [5]. Several studies have reported predictability effects at the level of the word [6], syllable [7, 1], or phoneme [7, 8]. Generally speaking, duration is shortened in more predictable contexts, and conversely lengthened in less predictable contexts. It has been suggested that long duration results from explicit encoding to improve intelligibility of hard-to-understand units [9, 10]. There are several measures to quantify predictability or the amount of information conveyed in a message [11]. One of them is surprisal. Surprisal captures the intuition that linguistic expressions that are highly predictable in a given context convey less information than those that are unexpected. Surprisal is defined as the contextual predictability of a unit and can be used as a measure of the amount of information that is conveyed by that unit in terms of bits, using Equation (1) where S stands for surprisal and P for probability:

$$S(unit_i) = -log_2 P(unit_i | Context)$$
(1)

A recent study by [1] examined the combined effects of syllable-based surprisal and noise in German CV syllables and found that syllable duration was shorter in predictable than unpredictable contexts, in line with existing literature [7]. Such surprisal effect at the syllable level could percolate directly downstream to the sub-syllabic level. However, [1] did not analyse the acoustic duration of sub-syllabic units, partly because the target syllables were comprised of initial stop consonants with different places of articulation and different vowels. Despite these constraints, half of the initial stops were voiced and the other half voiceless. This allowed us to use the voicing contrasts in these word-initial stops and explore the question as to how a syllable-based surprisal effect might change sub-syllabic phonological feature, e.g. voicing. It is possible that phonological voicing contrasts are simply lengthened because of the cascaded effect of surprisal-related syllable lengthening. The current study aims to address this issue.

Phonological voicing of stop consonants is distinguished as [+voice] vs. [-voice] in German (e.g. [12, 13]). Jessen [14] identified 8 acoustic correlates of [voice], viz. voice onset time (VOT), closure duration (CD), closure voicing, fundamental frequency onset, first formant onset, preceding vowel duration, following vowel duration, and the difference between the amplitude values of the first and second harmonics. Among these correlates, German relies primarily on VOT to signal the phonological voicing contrasts for initial stop consonants, while closure duration has been reported as another though not systematic cue [15]. German stops are characterized by long VOT for voiceless initial stops /ptk/ vs. short-lag VOT for their voiced counterparts/bdg/. In terms of laryngeal realism [16], German is an 'aspirating' language which specifies the feature [spread glottis] [17, 18]. Hence, voicing in German stop consonants is claimed to be along the fortis-lenis dimension (e.g. [19, 20]), with long-lag VOT for the former and short-lag VOT for the latter. Across languages, a range of factors has been identified to moderate VOT as a correlate of stop voicing, including preceding phoneme [21, 22, 23, 24], place of articulation [25, 21, 26], word position [25], vowel context [27, 15], and speaking rate [28], with some degree of variability between speakers [29].

Using data from a previously collected experiment [1], the present study explores how the effect of syllable-based predictability, defined as surprisal, will influence the acoustic realisation of the durational correlates for voicing contrasts of syllable-initial stops, viz. voice onset time (VOT) and closure duration (CD), in word-internal position. We hypothesize that speakers will lengthen VOT and CD for both voiced and voiceless initial stops in less predictable (high surprisal) syllable contexts, because syllable duration was lengthened in these contexts (see [1]).

2. Methods

The current study analysed the acoustic realisation of voicing contrasts in syllable initial stop consonants, using data collected in Ibrahim et al. for a different purpose [1].

2.1. Participants

Twenty-six native German female speakers were recruited (mean age = 27; age range 19–60 years). None of the speakers reported any hearing and speaking impairments.

 Table 1: Number of tokens in each condition

	Target [+voice]	Target [-voice]
Preceded by [+voice]	487	487
Preceded by [-voice]	238	85

2.2. Stimuli

The syllable stimuli were taken from [1] which investigated the effect of syllable-level surprisal and noise. In that study, 60 stimuli in a sentence context were selected from the DeWaC corpus [30, 31] with estimated high vs. low surprisal bins (HS vs. LS). Surprisal, the information-theoretic factor, was calculated as the negative probability of the syllable to occur in a specific context (two preceding syllables), and estimated by means of a trigram syllable-level language model trained on part of the DeWaC corpus using Equation 1. Each target CV syllable was lexically stressed (i.e., in a prosodically strong position) and occurred in a word-internal position of a polysyllabic word, where C beginning with a /p, k, b, d/ stop was combined with 5 vowels (/aː, eː, iː, oː, uː/). The polysyllabic target word was embedded in a sentence (Figure 1) in 3 noise conditions (baseline = no noise, 0 dB and -10 dB SNR with white noise). The combination of 5 vowels in 2 surprisal contexts in 3 noise conditions with 2 repetitions resulted in a total of 60 stimuli. The current study grouped those previously collected data into voiceless vs. voiced initial target stop consonant in low or high surprisal syllables, with a voiced or voiceless segment preceding the target stop consonant. Since the effect of noise did not interact with that of surprisal in Ibrahim et al [1], we collapsed the data across 3 noise conditions to increase statistical power. This resulted in two homogeneous voice sequences and two heterogeneous voice sequences in low or high surprisal syllables (Table 1).

2.3. Experimental Procedure

As described in [1], the data were recorded in a soundproof booth. Speakers wore a DPA 4067-F Omni headset microphone to record the speech signal and AKG K271 MKII over-ear headphones to hear the white noise signal during the recording session. The stimuli were visually presented on screen as a slide one at a time. A research assistant remotely controlled the advancement of the stimuli outside the recording booth. A practice session was provided for speakers before testing. During practice, speakers were instructed to read a different set of German sentences and the research assistant calibrated the equipment. The test phase consisted of three blocks. Speakers were informed about the presence of background noise in the first and last blocks. They stood upright and read the stimuli using their habitual reading pace. The order of the noise conditions (0 dB vs. 10 dB) was counterbalanced, retaining the middle block for the no-noise condition (baseline). Productions were recorded and stored as a mono wav file with a sampling rate of 48 kHz and 24 bits per sample.

2.4. Data annotation

The target words/syllables/segments were manually annotated by two trained phoneticians using Praat [32], while the nontarget words/syllables/segments in each sentence were automatically annotated using WebMAUS [33]. Apart from the annotations reported in Ibrahim et al [1], additional manual annotations of the following acoustic events were included for the



Figure 1: Example of manual annotation for the target consonant /k/in a context /ake:/

target stop consonants: onset of stop closure, stop burst release, onset of the following vowel as defined by periodicity and clear F2 (Figure 1). Positive VOT was defined as the interval from the burst release to the beginning of the onset of voicing in the following vowel, while negative VOT was defined as the interval from the burst release to the start of voicing during stop closure [25]. Only fully realised stops were included, as defined by the presence of closure and burst release visually detectable from the waveform and spectrogram.

2.5. Data analysis

Two durational measures were extracted from the target consonants (the focus of this study) using in-house Python and Praat scripts: voice onset time (VOT) and closure duration (CD). Linear mixed-effects modelling was used to evaluate the hypothesised effect(s) of surprisal and its possible interaction with the voicing status of the target stop consonants and their preceding segmental context using R lmer package [34]. The fixed factors were 3 two-level factors; (1) Voicing status of target stops (voiced or voiceless), (2) Surprisal group (LS or HS) and (3) Preceding context (voiced or voiceless). All factors were coded as simple contrasts. Backwards model selection procedure was applied to arrive at a final model as reported below. According to this procedure, a maximal random structure was first formulated to identify the model that best fit our data [35]. We included random intercepts and random slopes for all fixed effects. The random effects were speaker, syllable, and consonant. In case of convergence errors we reduced the maximal random structure step-wise. First, we removed random slopes, and then, if necessary, random intercepts. Significance of fixed effects was evaluated by performing maximum likelihood t-tests using Satterthwaite approximations to degrees of freedom. After model comparisons, the final model for VOT and CD was a model with interaction terms among fixed factors:

Feature ~ Surprisal * Preceding context voicing * Target consonant voicing + (1 | Speaker) + (1 | Syllable) + (1 | Consonant)

Table 2: Fixed effect results of the linear mixed-effects model for voice onset time (VOT) and closure duration (CD)

Comparison	Coeff.	SE	t-value	p-value			
VOT (interaction model)							
Surprisal Group	-3.98	1.42	-2.80	.005**			
Voicing of Preceding Context	-3.6	1.47	-2.46	.014*			
Voicing of Target consonant	-51.3	8.31	-6.17	.023*			
Surprisal*V. Preceding	-1.64	2.85	-0.57	.566			
Surprisal*V. Target	5.63	2.85	1.98	.048*			
V. Preceding*V. Target	-20.1	2.95	-6.81	<.0001***			
Surprisal*V. Preceding*V. Target	-4.9	5.7	-0.86	0.38			
CD (inte	eraction m	odel)					
Surprisal Group	-5.56	1.57	-3.53	<.0001***			
Voicing of Preceding Context	-3.43	1.64	-2.09	.037*			
Voicing of Target consonant	0.77	11.7	0.06	.953			
Surprisal*V. Preceding	-2.35	3.16	-0.74	.456			
Surprisal*V. Target	1.26	3.15	0.40	.687			
V. Preceding*V. Target	-22.1	3.28	-6.73	<.0001***			
Surprisal*V. Preceding*V. Target	-7.47	6.3	1.2	0.23			

3. Results

3.1. Effect of syllable-based surprisal on voice onset time

Figure 2 shows VOT values (in milliseconds) of our target consonants: voiced /b, d/ and voiceless /p, k/in low/high surprisal syllables with either a preceding voiced or voiceless context. As expected, VOT is higher for voiceless than voiced stops. The statistical model revealed a significant main effect of the voicing status of the target stops, confirming the expected pattern. Significant main effects of Surprisal, and Preceding context (voiced / voiceless) were also observed. Moreover, there were two significant interactions: (1) Surprisal by Voicing status of target stops, and (2) Voicing status of target stops by Preceding context (Table 2). The interaction between Surprisal and Voicing status of target stops is mostly attributable to higher VOT values for voiceless target stops in high surprisal syllables, but not for voiced target stops. The interaction between Voicing status of target stops and Preceding context arises, because VOT values increase for voiced target stops when they are preceded by a voiceless segmental context (i.e., in a heterogeneous sequence). But such pattern was not observed for voiceless target stops.

3.2. Effect of syllable-based surprisal on closure duration

Figure 3 shows closure duration values (in milliseconds) of the target consonants: voiced /b, d/and voiceless /p, k/ in low/high surprisal syllables. CD is longer in high than in low surprisal syllables, irrespective of the voicing status of the target stops or the preceding contexts. The statistical model revealed significant main effects of Surprisal and Preceding context, with a significant interaction of Voicing status of target stops and Preceding context (Table 2). This interaction arises because CD increases for voiced target stops when they are preceded by a voiceless segmental context, and CD increases for voiceless target stops when they are preceded by a voiceless target stops when they are preceded by a voiceless target stops when they are preceded by a voiceless target stops when they are preceded by a voiceless target stops when they are preceded by a voiceless target stops when they are preceded by a voiceless target stops when they are preceded by a voiceless target stops when they are preceded by a voiceless target stops when they are preceded by a voiceless target stops when they are preceded by a voiceless target stops when they are preceded by a voiceless target stops when they are preceded by a voiceless target stops when they are preceded by a voiceless target stops when they are preceded by a voiceless target stops when they are preceded by a voiceless target stops when they are preceded by a voiceless target stops when they are preceded by a voiceless target stops when they are preceded by a voiceless target stops are preceded by a voiceless target stops are preceded by a voiceless target stops when they are preceded by a voiceless target stops are preceded by a vo

Taking the results of VOT and CD together, it seems that the effect of syllable-based surprisal on VOT is conditional on

Voice onset time (milliseconds) Target voiced Target voiceless 100 50 (ms) -50 ₽ 100 50 NIC NOIC -50 Low surprisal High surprisal Low surprisal High surprisal

Figure 2: VOT of target stop consonants (voiced/voiceless) with preceding voiced/voiceless contexts. The blue points are the mean values.

the voicing status of the target initial stops, but its effect on CD is not. This suggests that the lengthening effect of syllablebased surprisal does not uniformly extend to durational cues associated with voicing contrasts in syllable-initial stops.

4. Discussion and Conclusions

The current paper has extended previous analysis in Ibrahim et al [1] by exploring the effect of syllable-based surprisal on sub-syllabic feature: phonological voicing contrasts in German syllable-initial stop consonants. We hypothesized that the surprisal effect will percolate downstream to the acoustic reali-



Figure 3: Closure duration of target stop consonants (voiced/voiceless) with preceding voiced/voiceless contexts. The blue points are the mean values.

sation of durational cues for phonological voicing contrasts. Based on 1297 syllable-initial German stops, our results yielded surprisal effects on both voice onset time (VOT) and closure duration (CD). However, the surprisal effect is uniform on CD, but interacts with Voicing status of target stops on VOT. CD is overall longer in high than low surprisal syllables, irrespective of the voicing status of the target stops or the voicing status of the preceding segmental context. For VOT we found interactions of the voicing status of the target stops with surprisal as well as with the preceding context. Voiceless target stops increase VOT in high surprisal syllables, but voiced target stops do not. This suggests that the syllable-based surprisal effect is sensitive to the phonological feature of a segment. At the same time, voiced target stops increase VOT as well as CD when they are preceded by a voiceless segmental context. This suggests that the voicing contrast between two adjacent segments is exaggerated in the heterogeneous voice sequence, even when there is an intervening syllable boundary.

Despite the constraint of mixed places of articulation in our target stops, we attempted to group our data into 2 sets, one with bilabial stops, and the other with non-bilabial stops and then examined the mean and standard deviation of VOT and CD for these 2 sets with voiced or voiceless target initial stops, preceded by a voiced or voiceless segmental context in high or low surprisal syllables. No data was available for the homogeneous voiceless sequence in low surprisal syllables. On the basis of the available descriptive statistics, different places of articulation do not seem to deviate from the general durational patterns of VOT and CD reported above. If the syllable-based surprisal effect on phonological voicing contrasts is a cascaded effect from the surprisal-induced syllable lengthening, one would have expected the same lengthening pattern in CD and VOT for voiced and voiceless target stop consonants, irrespective of whether they are preceded by a voiced or voiceless segmental context. But our results did not support such a simple interpretation. Instead, our results suggest that syllable-based surprisal might directly affect sub-syllabic phonological voicing contrasts, not mediated through lengthening the target syllable, at least for VOT. At the same time, the pattern from CD seems to suggest otherwise, with a cascaded effect from syllable-based surprisal. In other words, the surprisal effect on VOT might not be mediated through surprisal-induced lengthening, but its effect on CD might be.

Taking the results of VOT and CD together, it seems that the way syllable-based surprisal affects phonological voicing contrasts might depend on the acoustic-phonetic cues. A possible explanation might be the phonological relevance of the respective cue. Although both VOT and CD are durational cues to voicing contrasts, they differ in their primacy and perceptual distinctness. VOT has been shown to be a primary cue to phonological voicing contrasts in German, and CD to be a secondary cue. If the syllable-based surprisal effect is to emphasize a linguistic unit's informativity, such function or need might be better met by using a primary phonetic cue rather than a secondary cue. When a cue is primary and phonologically relevant, the syllable-based surprisal effect is sensitive to the phonological status of the target initial stops; however, when a cue is secondary, the surprisal effect is mediated through surprisalinduced lengthening.

Admittedly, the current study is limited by using data that were collected for a different research question. Potential confounds such as accentuation of the target syllable, word length, sentence length and places of articulation for initial stops should be considered in further investigations of the effect of syllablebased vs. phone-based surprisal on voicing contrasts in German. Despite these potential confounds, it is quite unexpected to observe a selective syllable-based surprisal effect on VOT and CD for sub-syllabic features, suggesting that the flow of the surprisal effect between linguistic levels might be selective.

To conclude, syllable-based surprisal both directly and indirectly affects the phonological feature [voice], as it affects phonetic cues differentially, depending on how relevant and distinct each cue is to the phonological contrast.

5. Acknowledgements

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project–ID 232722074 – SFB 1102. We thank Kirstin Kolmorgen and Anna Konobelkina for annotation and technical support.

6. References

- O. Ibrahim, I. Yuen, M. van Os, B. Andreeva, and B. Möbius, "The effect of lombard speech modifications in different information density contexts," in *Elektronische Sprachsignalverarbeitung* 2021, Tagungsband der 32. Konferenz (Berlin). Dresden: TUDpress, 2021, pp. 185–191.
- [2] M. Aylett and A. Turk, "The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech," *Language and Speech*, vol. 47, no. 1, pp. 31–56, 2004, https://doi.org/10.1177/00238309040470010201.
- [3] A. F. Frank and T. Jaeger, "Speaking rationally: Uniform information density as an optimal strategy for language production," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, 2008.
- [4] M. W. Crocker, V. Demberg, and E. Teich, "Information Density and Linguistic Encoding (IDeaL)," *KI - Künstliche Intelligenz*, vol. 30, no. 1, pp. 77–81, 2016, https://doi.org/10.1007/s13218-015-0391-y.

- [5] T. Pimentel, C. Meister, E. Salesky, S. Teufel, D. Blasi, and R. Cotterell, "A surprisal-duration trade-off across and within the world's languages," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021, pp. 949–962.
- [6] E. Buz and T. F. Jaeger, "The (in)dependence of articulation and lexical planning during isolated word production," *Language*, *Cognition and Neuroscience*, vol. 31, no. 3, pp. 404–424, 2016, https://doi.org/10.1080/23273798.2015.1105984.
- [7] M. Aylett and A. Turk, "Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei," *The Journal of the Acoustical Society of America*, vol. 119, no. 5, pp. 3048–3058, 2006, https://doi.org/10.1121/1.2188331.
- [8] J. Bybee, "Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change," *Language Variation and Change*, vol. 14, no. 3, p. 261–290, 2002.
- [9] T. F. Jaeger, "Redundancy and reduction: speakers manage syntactic information density," *Cognitive Psychology*, vol. 61, no. 1, pp. 23–62, 2010, https://doi.org/10.1016/j.cogpsych.2010.02.002.
- [10] S. Gahl, Y. Yao, and K. Johnson, "Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech." *Journal of Memory and Language*, vol. 66, no. 4, pp. 789–806, 2012.
- [11] J. Hale, "Information-theoretical complexity metrics," Language and Linguistics Compass, pp. 1–16, 2016, https://doi.org/10.1111/Inc4.12196.
- [12] R. Wiese, *The Phonology of German*. Oxford: Oxford University Press, 2000.
- [13] J. Rubach, "Final devoicing and cyclic syllabification in german," *Linguistic Inquiry*, vol. 21, pp. 79–94, 1990.
- [14] M. Jessen, Phonetic implementation of the distinctive auditory features [voice] and [tense] in stop consonants. De Gruyter Mouton, 2001, pp. 237–294, https://doi.org/10.1515/9783110886672.237.
- [15] D. Pape, C. Mooshammer, P. Hoole, and S. Fuchs, "Devoicing of word-initial stops: A consequence of the following vowel?" in *Towards a better understanding of speech production processes*, J. Harrington and M. Tabain, Eds. New York: Psychology Press, 2006, pp. 211–226.
- [16] P. Honeybone, Diachronic evidence in segmental phonology: the case of obstruent laryngeal specifications. De Gruyter Mouton, 2012, pp. 317–352. [Online]. Available: https://doi.org/10.1515/9783110890402.317
- [17] M. Jessen, B. Wiesbaden, and C. Ringen, "Laryngeal features in german," *Phonology*, pp. 1–30, 2002.
- [18] J. Beckman, M. Jessen, and C. Ringen, "Empirical evidence for laryngeal features: Aspirating vs. true voice languages," *Journal* of *Linguistics*, vol. 49, no. 2, p. 259–284, 2013.
- [19] K. J. Kohler, *Einführung in die Phonetik des Deutschen*. Erich Schmidt Verlag, Berlin, 1995.
- [20] M. Jessen, Phonetics and Phonology of Tense and Lax Obstruents in German. John Benjamins, Amsterdam, 1998.
- [21] G. J. Docherty, The Timing of Voicing in British English Obstruents. De Gruyter Mouton, 1992, https://doi.org/10.1515/9783110872637.
- [22] Y. Yao, "Understanding vot variation in spontaneous speech," UC Berkeley PhonLab Annual Report, vol. 5, pp. 29–43, 2009.
- [23] C. Shih, B. Möbius, and B. Narasimhan, "Contextual effects on consonant voicing profiles: A cross-linguistic study," in *Proceedings of the 14th International Congress of Phonetic Sciences (San Francisco, CA)*, vol. 2, 1999, pp. 989–992.
- [24] B. Möbius, "Corpus-based investigations on the phonetics of consonant voicing," *Folia Linguistica*, vol. 38, no. 1–2, pp. 5–26, 2004, https://doi.org/10.1515/flin.2004.38.1-2.5.

- [25] 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, https://doi.org/10.1080/00437956.1964.11659830.
- [26] 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, https://dx.doi.org/10.1006/jpho.1999.0094.
- [27] N. Abdelli-Beruh, "The stop voicing contrast in French sentences: Contextual sensitivity of vowel duration, closure duration, voice onset time, stop release and closure voicing," *Phonetica*, vol. 61, no. 4, pp. 201–219, 2004.
- [28] R. M. Theodore, J. L. Miller, and D. DeSteno, "Individual talker differences in voice-onset-time: Contextual influences," *The Journal of the Acoustical Society of America*, vol. 125, no. 6, pp. 3974– 3982, 2009, https://doi.org/10.1121/1.3106131.
- [29] J. Tanner, M. Sonderegger, and J. Stuart-Smith, "Structured speaker variability in japanese stops: Relationships within versus across cues to stop voicing," *The Journal of the Acousti*cal Society of America, vol. 148, no. 2, pp. 793–804, 2020, https://doi.org/10.1121/10.0001734.
- [30] E. Brandt, F. Zimmerer, B. Andreeva, and B. Möbius, "Melcepstral distortion of German vowels in different information density contexts," in *Proceedings of Interspeech (Stockholm, Sweden)*, 2017, pp. 2993–2997.
- [31] M. Baroni and A. Kilgarriff, "Large linguistically-processed web corpora for multiple languages," in *Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics (EACL-2006)*, 2006, pp. 87–90.
- [32] P. Boersma and D. Weenink, "Praat: doing phonetics by computer (version 6.1. 08) [computer program], retrieved May 1," 2019.
- [33] F. Schiel, "Automatic phonetic transcription of non-prompted speech," in Proc. of the 14th International Congress of Phonetic Sciences (San Francisco, CA), 1999, pp. 607–610.
- [34] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.
- [35] D. Barr, "Random effects structure for testing interactions in linear mixed-effects models," *Frontiers in Psychology*, vol. 4, p. 328, 2013.