



Glottal Stops in Upper Sorbian: a Data-Driven Approach

Ivan Kraljevski¹, Maria Paola Bissiri^{2*}, Frank Duckhorn¹, Constanze Tschoepe¹, Matthias Wolff³

¹Fraunhofer IKTS, Fraunhofer Institute for Ceramic Technologies and Systems, Dresden, Germany

²Dipartimento di Scienze umane e dell'innovazione per il territorio (DiSUIT),
Università degli Studi dell'Insubria, Como, Italy

³Chair of Communications Engineering, Brandenburg University of Technology (BTU)
Cottbus-Senftenberg, Cottbus, Germany

{ivan.kraljevski, frank.duckhorn, constanze.tschoepe}@ikts.fraunhofer.de,
mariapaola.bissiri@uninsubria.it, matthias.wolff@b-tu.de

Abstract

We present a data-driven approach for the quantitative analysis of glottal stops before word-initial vowels in Upper Sorbian, a West Slavic minority language spoken in Germany. Glottal stops are word-boundary markers and their detection can improve the performance of automatic speech recognition and speech synthesis systems.

We employed cross-language transfer using an acoustic model in German to develop a forced-alignment method for the phonetic segmentation of a read-speech corpus in Upper Sorbian. The missing phonemic units were created by combining the existing phoneme models. In the forced-alignment procedure, the glottal stops were considered optional in front of word-initial vowels.

To investigate the influence of speaker type (males, females, and children) and vowel on the occurrence of glottal stops, binomial regression analysis with a generalized linear mixed model was performed. Results show that children glottalize word-initial vowels more frequently than adults, and that glottal stop occurrences are influenced by vowel quality.

Index Terms: glottal stops, acoustic modeling, under-resourced languages

1. Introduction

Recent advances in speech technology introduced improved tools for phonetic research. However, many of them are strictly language-dependent and require the acquisition of carefully prepared and transcribed acoustic data. That can be problematic for endangered and under-resourced languages, for which few or no resources exist.

The acquisition and transcription of speech data are time-consuming and costly processes and significant limiting factors in developing speech-enabled applications. The manual transcription performed by different annotators can result in inconsistent and error-prone text and speech alignments.

Automatic text-speech alignment (forced-alignment) is successfully applied to the phonetic segmentation of widely spoken and well-studied languages. However, this method is seldom used for under-resourced or endangered languages [1], due to the inability to create a reliable forced-alignment tool suited for the language.

In recent times, novel approaches appeared for the rapid creation and development of speech-enabled systems for under-resourced languages [2]. Depending on the availability

of language resources, there are two main research directions: zero-resource approaches [3],[4], where it is assumed that there are no language resources at all, and universal cross-language speech recognition systems [5],[6]. The typical approach exploits acoustic similarities between the sounds in different languages to perform a cross-language transfer, unsupervised training, and bootstrapping.

In this study, we present a data-driven approach for the quantitative analysis of glottal stops before word-initial vowels in Upper Sorbian. We employed a trained acoustic model in German to develop a forced-alignment method for the phonetic segmentation of a read-speech corpus in Upper Sorbian.

Although the target and the source language (German) belong to different language families, the phoneme inventories share a significant portion of their units. We created the missing units by combining existing phoneme models. The forced-alignment used the lexicon and the orthographic transcriptions, with optional glottal stops in front of word-initial vowels. A generalized linear mixed model for binomial regression was used to investigate the effects of speaker type (males, females, and children) and vowel on the occurrence of glottal stops.

2. Sorbian Languages

Upper and Lower Sorbian belong to the West Slavic branch of the Indo-European language family with Polish, Czech, and Slovak; they are spoken in Germany, in the region of Upper Lusatia and Lower Lusatia, respectively. Although closely related, they are only partially mutually intelligible. Upper Sorbian is considered closer to Czech and Slovak, whereas Lower Sorbian is considered closer to Polish [7].

Upper and Lower Sorbian are both endangered and under-resourced languages [8]. There are no reliable figures about the number of speakers of the Sorbian languages. Different sources provide a wide range of estimations which vary between 5 to 10 thousand for Lower Sorbian speakers and 15 and 30 thousand speakers for Upper Sorbian [9], which makes them a minority (1%) in a predominantly German-speaking area [10]. Most native speakers do not use the language in daily communication; therefore, the language is threatened despite being under the protection of the European Charter for Regional or Minority Languages. A substantial effort has been made to preserve the Sorbian languages by various projects conducted by the Foundation for the Sorbian People¹ and by the Sorbian Institute². Other research groups contributed to collecting and

*For academic purposes, Maria Paola Bissiri is responsible for sections 3, 6.2, and 7.

¹www.stiftung.sorben.com

²www.serbski-institut.de

documenting speech and language in Sorbian, mainly focusing on the digitalization of the resources.

3. Glottal Stops

Glottal stops are articulated by closing and abruptly opening the vocal folds. They are not always realized as full stops and can be substituted by other forms of glottalization, such as irregular and low-frequency glottal fold vibrations known as creaky voice.

Glottal stops are present in most world languages, albeit their function varies across different language groups. They can be phonemic, non-phonemic, and in certain language groups they can be both [11]. Glottal stops can also carry relevant linguistic information by marking morpheme, word, phrase, or utterance boundaries, depending on the language.

In German, Czech and Polish glottal stops are not considered phonemes since they cannot differentiate word meaning. In German and Czech glottal stops or creaky voice before word-initial vowels are very frequent word-boundary markers (s. [12] for German and [13, 14] for Czech), while in Polish they are less frequent and more likely to occur at phrase boundaries [15].

Similarly as in Czech and German, word-initial vowels in Upper Sorbian are normally preceded by a glottal stop [7]. In Upper Sorbian lexical stress always falls on the first syllable [10]. Vowel-initial words are rare in Upper Sorbian, mostly appearing in recent borrowings. However, the initial vowel /a/ occurs in the function words *a* “and”, *ale* “but”, *abo* “or” and *ani* “nor” [7]. In Lower Sorbian, the occurrence of glottal stops can be interpreted as the increasing influence of the German language [16].

Most studies on Sorbian report qualitative analyses of glottal stops; however, when it comes to a quantitative analysis, employing tools for automatic detection is necessary. Such tools are commonly based on automatic speech recognition systems used in the forced-alignment mode to provide phonetic segmentation, like MAUS [17] and the UASR (Unified Automatic Speech Recognition and Synthesis) toolkit [18] which was used in this study.

The prerequisite is that the employed acoustic model, apart from phoneme models, include also glottal stop models. A reliable alignment of glottalized segments would provide a valuable pre-segmentation for basic phonetic research; moreover, the acoustic modeling of glottalizations, which are good indicators of word boundaries, can improve the overall performance of an Automatic Speech Recognition (ASR) system [19].

Modeling glottal stops also improves naturalness in speech synthesis [20] and could be employed for the automatic creation of speech stimuli in perceptual experiments for basic research on glottalizations [21].

4. Speech Data

The HSB speech corpus was collected by Fraunhofer IKTS and the Brandenburg University of Technology (BTU) at the premises of the Foundation for the Sorbian People in Bautzen. The recording sessions used three different phonetically balanced prompt sets (HSB-1, 2, and 3).

Around 2/3 of the prompts are taken from the Common Voice [22] dataset (version: hsb.2h.2020-06-22) and the rest from domain-specific sentences which include examples of command and control actions, e.g., turning on/off an intelligent light bulb, setting the brightness, and the color.

4.1. Recording Prompts

The prompts should be phonetically rich and balanced, resembling the phoneme units’ statistics of a larger textual corpus to maximize the benefits of the controlled recording sessions.

For that purpose, we additionally used the textual corpus “Monolingual Upper Sorbian Data” from the “Shared Task: Unsupervised MT and Very Low Resource Supervised MT” [23] with a vocabulary of 251,358 words.

The frequencies of the phones, diphones, and triphones were calculated and the prompts were selected from the domain-general and the domain-specific sentences, according to the scoring algorithm presented in [24] applied on diphones.

4.2. Lexicon

The Upper Sorbian language has simple grapheme-to-phoneme mapping rules with several exceptions [25]. The lexicon was created using the vocabulary of the prompts’ textual content, and each word had only one pronunciation variant.

A native speaker checked the lexicon, thus the pronunciation rules were further improved and non-suitable words removed. This lexicon was used to filter out inappropriate sentences, which could lead to inconsistencies in the experiments.

Pronunciation variants were added for the words with initial vowels by adding an optional glottal stop in front of the phoneme sequence. This allowed the forced-alignment labeling to detect a mark glottal stops before word-initial vowels.

The lexicon has 3547 entries, 3457 of them with canonical pronunciation, and 90 with an optional word-initial glottal stop. Most words, as mentioned in section 3, are recent borrowings, such as *insolwencu*, *italsku*, *exponaty*, and *organizacija*.

4.3. Speech Recordings

We recorded 11 hours, 30 minutes, and 40 seconds of speech with almost equal distribution across the three datasets (HSB-1, 2 and 3). Recordings from 30 speakers of different gender and age were collected: ten females, ten males, and ten children. One additional male speaker was recorded at the BTU studio as part of the testing phase. The children participants are minors attending either higher elementary or lower classes in high school, which implies good reading skills.

Each speaker was instructed to read the prompts precisely as displayed, to reduce the effort for post-processing and to avoid manual transcriptions of the recordings.

A native speaker of Upper Sorbian was present for the whole duration of the recording session to ensure a proper pronunciation of the speech prompts. Whenever a speaker would mispronounce a word, the prompt recording was repeated. The order of the prompts was randomized, and the speakers were free to read as many as they liked.

The number of prompts recorded by each speaker varied in the range between 100 and 250, with an average of 191 prompts per speaker. A coverage of 100% of the available prompts (1210) was achieved. Each prompt was read from 1 to 10 times (5.2 times on average).

5. Experimental Setup

5.1. Acoustic Modeling

The acoustic model used for forced-alignment was created by combining phone models from the existing German acoustic model. The baseline German model was trained using the

Table 1: Grapheme (*G*), Upper Sorbian X-Sampa (*HSB*) and German Sampa (*UASR*)

| G | HSB | UASR | G | HSB | UASR |
|-----|------|---------|------|------|--------|
| ⟨A⟩ | a | a | ⟨N⟩ | n | n |
| ⟨B⟩ | b | b | ⟨Ń⟩ | jn | j,n |
| ⟨C⟩ | ts | t,s | ⟨O⟩ | o | O |
| ⟨Č⟩ | tS | t,S | ⟨Ó⟩ | uU | U |
| ⟨Ć⟩ | t.s\ | t,S | ⟨P⟩ | p | p |
| ⟨D⟩ | d | d | ⟨R⟩ | r | r |
| ⟨E⟩ | E | E or e: | ⟨Ř⟩ | S | S |
| ⟨Ě⟩ | iI | I | ⟨S⟩ | s | s |
| ⟨F⟩ | f | f | ⟨Š⟩ | S | S |
| ⟨G⟩ | g | g | ⟨T⟩ | t | t |
| ⟨H⟩ | h | h | ⟨U⟩ | u | u: |
| ⟨I⟩ | i | i: | ⟨W⟩ | w | U,v |
| ⟨J⟩ | j | j | ⟨Y⟩ | l | Y |
| ⟨K⟩ | k | k | ⟨Z⟩ | z | z |
| ⟨Ł⟩ | w | U,v | ⟨Ž⟩ | Z | Z |
| ⟨L⟩ | l | l | ⟨CH⟩ | x | x or C |
| ⟨M⟩ | m | m | ⟨DŽ⟩ | d.Z\ | d,Z,S |

UASR (Unified Automatic Speech Recognition and Synthesis) toolkit [18] on part of the German Verbmobil Database [26]. The acoustic model structure consists of 40 monophonic HMMs for phonemic units, with separate models for pauses, non-speech and glottal stops; it was successfully employed in similar studies [27, 21]. The employed grapheme-phoneme mapping is derived from the Upper Sorbian alphabet and phoneme inventory reported in [28, 10, 29, 7]. The target phoneme inventory was created by selecting and combining the nearest German phoneme models.

Table 1 presents the adopted grapheme and phoneme inventories with the corresponding German equivalents used in this study. A separate model “Q”, representing the glottal stops, and appropriate models for silence “.” and non-speech “#” are also included in the resulting acoustic model (in total, 36 models).

The resulting acoustic model was not adapted to the recorded corpora’s speech: adaptation might improve the overall forced-alignment performance for most phonemes, but it would distort the glottal stops model, introducing lower detection performance.

5.2. Forced-Alignment

The forced-alignment procedure uses the aforementioned acoustic model, the created lexicon, and the orthographic transcriptions.

The feature extraction takes the raw speech recordings in Wave format with 16 kHz, 16-bit PCM signals. The primary feature analysis was performed by employing a short-time Fourier transformation with a Blackman window of 512, and a MEL filter-bank with a triangular transfer function. The frame features have a dimension of 30 elements.

The secondary features analysis was done over the primary features by standardizing the feature elements to zero mean and unit variance, computing the delta features, and performing Principal Component Analysis (PCA), effectively reducing the frame features’ dimension to 24 elements.

5.3. Detection of Glottal Stops

The whole recorded speech corpus was aligned, and for each sentence, the original word-initial vowels were identified by referring to the lexicon. The presence of glottal stops was detected by observing the recognized label.

For each word-initial vowel, we store in a data frame the information about the word, the vowel, the presence or absence of the glottal stops, along with speaker type (female, male or child), speaker identifier, and prompt identifier.

An expert phonetician inspected the detection’s reliability by checking the segmentation quality and correctness for all the sentences containing at least one word-initial vowel.

6. Results and Discussion

6.1. Exploratory Statistics

The corpus consists of 6313 speech recordings by 31 speakers. The distribution of the recordings per speaker type was; 24.57 % from children, 30.94 % from female, and 44.50 % from male speakers.

Table 2: Observations per speaker type (%)

| children | females | males |
|----------|---------|-------|
| 22.33 | 28.40 | 49.27 |

After applying the FA procedure, we identified the 1557 sentences of the corpus that contain at least one word where an initial glottal stop could appear. In total, 1864 cases compose the observations to be analyzed. Table 2 shows the distribution of the observations per speaker type, which roughly resembles the recordings’ distribution.

Table 3 presents the distribution of the word-initial vowels in the observations. The most frequent vowel is /a/ due to the

Table 3: Observations per word-initial vowel (%)

| | a | E | i | O |
|-----------|-------|------|-------|------|
| canonical | 79.50 | 2.15 | 10.36 | 7.94 |

very high frequency of the function word *a* “and”, appearing in 61.6% of the observations (Table 4).

Table 4: Frequencies of the top 5 words (%)

| word | frequency |
|-------------------|-----------|
| <i>a</i> | 61.57 |
| <i>ale</i> | 4.72 |
| <i>abo</i> | 2.9 |
| <i>oranżowu</i> | 1.83 |
| <i>optikarjej</i> | 0.97 |

Glottal stops occurred significantly more often in children (85.82%) than in female (72.97%) and male (61.55%) speakers ($\chi^2 = 83.311$, $df = 2$, $p < 0.001$), as shown in Figure 1.

Similarly, the occurrence of glottal stops was also significantly dependent on the word-initial vowels ($\chi^2 = 104.44$, $df = 3$, $p < 0.001$) as shown in Figure 2.

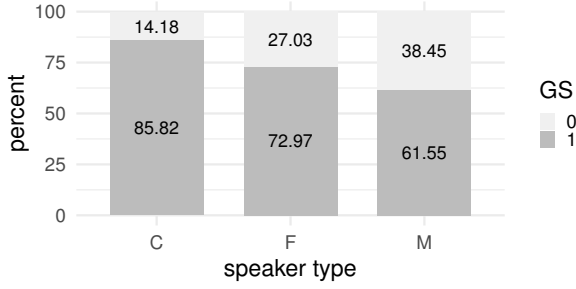


Figure 1: Glottal stop frequencies across speaker type

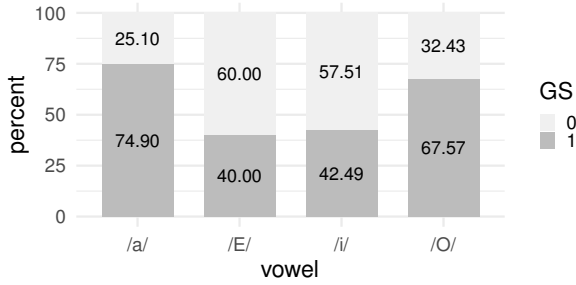


Figure 2: Glottal stop frequencies across word-initial vowel

6.2. Generalized Linear Mixed Model

A binomial regression analysis by means of a generalized linear mixed model was carried out, using the R package **lme4** [30]. The binomial dependent variable (*gs*) was defined as presence (1) or absence (0) of the glottal stop before the word-initial vowel. Speaker type (*type*), word-initial vowel (*vowel*), and their interaction were given as fixed effects. Speaker and word were the random effects, as given in the model formula:

$$gs \sim type + vowel + type : vowel + (1|speaker) + (1|word)$$

Stratified 5-fold cross-validation was employed to assess the glottal stops detection performance; the model achieved a balanced accuracy rate [31] of 71.90% with 95% CI (69.67, 74.13), and an F1 score of 85.62%.

Table 5 presents the analysis of deviance table of Type III with Wald chi-square tests with speaker type, vowel, and the type-vowel interaction as factors. All three are significant factors for the glottal stop occurrences.

Figure 3 presents the odds ratios for each fixed effect, the exponential of the model coefficients (the fixed effects estimates). The intercept in the model contains speaker type “C”

Table 5: Analysis of Deviance Table (Type III Wald chi-square tests)

| | Chi-square | Df | Pr(> Chi-square) |
|-------------|------------|----|------------------|
| (Intercept) | 9.5172 | 1 | 0.002036 ** |
| type | 10.6288 | 2 | 0.004920 ** |
| vowel | 10.6626 | 3 | 0.013697 * |
| type:vowel | 14.9171 | 6 | 0.020911 * |

*Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

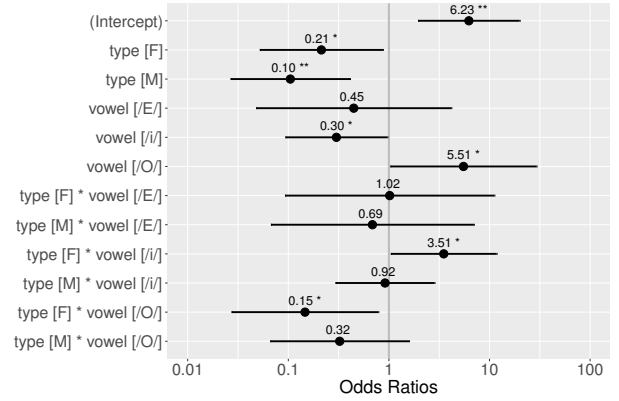


Figure 3: Odds ratios of fixed effects.

(children) and the vowel /a/. The effects *type*[M] ($p < 0.01$) and *type*[F] ($p < 0.05$) are significant, confirming that male and female speakers use fewer glottal stops than children. No difference was found between female and male speakers. The literature reports conflicting results whether males glottalize more frequently than females or vice-versa [32, 33]. Several variables could influence glottalization rate: anatomical, sociolinguistic and structural factors [32].

Glottalizations of word-initial vowels by children are underinvestigated. In our study, speech rate could play a role in children’s higher glottal stop rates: the recorded children read at a lower speech rate compared to adults, and more frequent glottalizations have been associated to slower speech [34, 33]. It could also be hypothesized that higher F0, as in children’s speech, influences glottal stop rate, since many languages have glottalization associated to high tones (s. references in [35]).

The vowel /O/ is more likely ($p < 0.05$) while the vowel /i/ is less likely ($p < 0.05$) to be preceded by a glottal stop than /a/ and /E/. These results are in line with [34, 33], with open vowels being more glottalized than close vowels, and back vowels being more glottalized than front and central vowels. However, the significant interactions *type*[F] * *vowel*[i/] ($p < 0.05$) and *type*[F] * *vowel*[O/] ($p < 0.05$) in the opposite direction show that female speaker’s glottal stop rates are less influenced by vowel quality.

7. Conclusions

This paper presents a data-driven study of glottal stops before word-initial vowels in Upper Sorbian. To phonetically transcribe a read-speech corpus we employed a forced-alignment tool by means of cross-language acoustic modeling.

The resulting phoneme segmentation and the detected presence of glottal stops were used for further statistical analysis. Binomial regression with a generalized linear mixed model showed more frequent glottal stops by children than by adults. Glottal stops were also more frequent before open than close vowels, and before back than central and front vowels, although glottal stop rates by female speakers were less influenced by vowel quality. Further research will investigate not only glottal stops but also other glottalization types.

8. Acknowledgments

This study was supported by the Foundation for the Sorbian People in Bautzen, Germany.

9. References

- [1] C. DiCanio, H. Nam, D. H. Whalen, H. T. Bunnell, J. D. Amith, and R. C. García, "Using automatic alignment to analyze endangered language data: Testing the viability of untrained alignment," *The Journal of the Acoustical Society of America*, vol. 134, no. 3, pp. 2235–2246, 2013.
- [2] O. Scharenborg, F. Ciannella, S. Palaskar, A. Black, F. Metze, L. Ondel, and M. Hasegawa-Johnson, "Building an ASR system for a low-resource language through the adaptation of a high-resource language ASR system: Preliminary results," *Proceedings of ICNLSP, Casablanca, Morocco*, 2017.
- [3] X. Li, S. Dalmia, D. Mortensen, J. Li, A. Black, and F. Metze, "Towards zero-shot learning for automatic phonemic transcription," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 05, pp. 8261–8268, Apr. 2020.
- [4] M. Wiesner, O. Adams, D. Yarowsky, J. Trmal, and S. Khudanpur, "Zero-shot pronunciation lexicons for cross-language acoustic model transfer," in *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2019, pp. 1048–1054.
- [5] T. Schultz and A. Waibel, "Language-independent and language-adaptive acoustic modeling for speech recognition," *Speech Communication*, vol. 35, no. 1-2, pp. 31–51, 2001.
- [6] L. Besacier, E. Barnard, A. Karpov, and T. Schultz, "Automatic speech recognition for under-resourced languages: A survey," *Speech Communication*, January 2014, vol. 56, pp. 85–100, 2014.
- [7] G. Stone, "Sorbian (Upper and Lower)," in *In Comrie, Bernard & Corbett, Greville G. (eds.), The Slavonic languages*. London & New York: Routledge, 1993, pp. 593–685.
- [8] C. Moseley, *The UNESCO atlas of the world's languages in danger: Context and process*. World Oral Literature Project, 2012.
- [9] R. Jodlbauer, "Die aktuelle Situation der niedersorbischen Sprache : Ergebnisse einer soziolinguistischen Untersuchung der Jahre 1993–1995," Bautzen, 2001.
- [10] P. Howson, "Upper Sorbian," *Journal of the International Phonetic Association*, vol. 47, no. 3, pp. 359–367, 2017.
- [11] I. Maddieson and K. Precoda, "Updating UPSID," *The Journal of the Acoustical Society of America*, vol. 86, no. S1, pp. S19–S19, 1989.
- [12] K. J. Kohler, "Glottal stops and glottalization in German," *Phonetica*, vol. 51, no. 1-3, pp. 38–51, 1994.
- [13] M. P. Bissiri and J. Volín, "Prosodic structure as a predictor of glottal stops before word-initial vowels in Czech English," in *Proceedings of the 20th Czech-German Workshop on Speech Processing*, 2010, pp. 23–28.
- [14] M. P. Bissiri, M. L. García Lecumberri, M. Cooke, and J. Volín, "The role of word-initial glottal stops in recognizing English words," in *Twelfth Annual Conference of the International Speech Communication Association*, 2011.
- [15] M. Żygis and B. Pompino-Marschall, *Glottale Markierungen vokallinitialer Wörter: Deutsch und Polnisch im Kontrast*. De Gruyter, 2012, pp. 75–98.
- [16] R. Marti, B. Andreeva, and W. Barry, "GENIE: The corpus for spoken Lower Sorbian (GESprochenes NIEdersorbisch)," *The Phonetician*, vol. 101/102, pp. 47–59, 01 2010.
- [17] T. Kisler, U. Reichel, and F. Schiel, "Multilingual processing of speech via web services," *Computer Speech & Language*, vol. 45, pp. 326–347, 2017.
- [18] R. Hoffmann, M. Eichner, and M. Wolff, "Analysis of verbal and nonverbal acoustic signals with the Dresden UASR system," in *International Workshop on Verbal and Nonverbal Communication Behaviours. COST Action 2102*, ser. Lecture Notes in Computer Science, A. Esposito, M. Faundez-Zanuy, E. Keller, and M. Marinaro, Eds., vol. 4775. Vietri sul Mare, Italy: Springer-Verlag, mar 2007, pp. 200–218.
- [19] T.-J. Yoon, X. Zhuang, J. Cole, and M. Hasegawa-Johnson, "Voice quality dependent speech recognition," in *International Symposium on Linguistic Patterns in Spontaneous Speech*. Citeseer, 2006.
- [20] T. Drugman, J. Kane, and C. Gobl, "Modeling the creaky excitation for parametric speech synthesis," in *Thirteenth Annual Conference of the International Speech Communication Association*, 2012.
- [21] I. Kraljevski, M. P. Bissiri, G. Strecha, and R. Hoffmann, "Analysis and synthesis of glottalization phenomena in German-accented English," in *Speech and Computer*, A. Ronzhin, R. Potapova, and V. Delic, Eds. Cham: Springer International Publishing, 2014, pp. 97–104.
- [22] R. Ardila, M. Branson, K. Davis, M. Henretty, M. Kohler, J. Meyer, R. Morais, L. Saunders, F. M. Tyers, and G. Weber, "Common Voice: A massively-multilingual speech corpus," in *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, 2020, pp. 4211–4215.
- [23] L. Barrault, O. Bojar, F. Bougares, R. Chatterjee, M. R. Costa-jussà, C. Federmann, M. Fishel, A. Fraser, Y. Graham, P. Guzman, B. Haddow, M. Huck, A. J. Yepes, P. Koehn, A. Martins, M. Morishita, C. Monz, M. Nagata, T. Nakazawa, and M. Negri, Eds., *Proceedings of the Fifth Conference on Machine Translation*. Online: Association for Computational Linguistics, November 2020.
- [24] J. Berry, L. Fadiga *et al.*, "Data-driven design of a sentence list for an articulatory speech corpus," in *INTERSPEECH*, 2013, pp. 1287–1291.
- [25] I. Kraljevski, M. Rjelka, F. Duckhorn, C. Tschoepe, and M. Wolff, "Cross-lingual acoustic modeling in Upper Sorbian - preliminary study," in *Studentexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2021*, S. Hillmann, Ed. TUDpress, Dresden, 2021.
- [26] T. Bub and J. Schwinn, "Verbmobil: The evolution of a complex large speech-to-speech translation system," in *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96*, vol. 4. IEEE, 1996, pp. 2371–2374.
- [27] M. P. Bissiri, I. Kraljevski, and R. Hoffmann, "Improved phoneme segmentation of German-accented English by means of lexicon and acoustic model adaptation," in *Studentexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2013*, P. Wagner, Ed. TUDpress, Dresden, 2013, pp. 231–238.
- [28] J. Šoćina and E. Wornar, *"Obersorbisch im Selbststudium/Hornjoserbšćina za samostudij"*, ser. Ein Sprachkurs für Unerschrockene (inkl. CD). Aufl., Bautzen: Domowina-Verlag, 2012, vol. 3.
- [29] H. Schuster-Šewc, *Grammar of Upper Sorbian language: phonology and morphology*, ser. LINCOM Studies in Slavic Linguistics. München: Lincom Europa, 1996, vol. 3.
- [30] 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.
- [31] K. Brodersen *et al.*, "The balanced accuracy and its posterior distribution," in *20th International Conference on Pattern Recognition*. IEEE, 2010, pp. 3121–3124.
- [32] L. Redi and S. Shattuck-Hufnagel, "Variation in the realization of glottalization in normal speakers," *Journal of Phonetics*, vol. 29, no. 4, pp. 407–429, 2001.
- [33] A. Markó, A. Deme, M. Bartók, T. E. Grácsi, and T. G. Csapó, "Speech rate and vowel quality effects on vowel-related word-initial irregular phonation in Hungarian," *Challenges in Analysis and Processing of Spontaneous Speech*, pp. 49–74, 2018.
- [34] Z. Malisz, M. Żygis, and B. Pompino-Marschall, "Rhythmic structure effects on glottalisation: A study of different speech styles in Polish and German," *Laboratory Phonology*, vol. 4, no. 1, pp. 119–158, 2013.
- [35] M. P. Bissiri and M. Zellers, "Perception of glottalization in varying pitch contexts across languages," in *INTERSPEECH*, 2013, pp. 253–257.