FILOZOFICKÁ FAKULTA
Univerzita Karlova

MASTER THESIS

Bc. Richard Mazúr

Acoustic measures of glottalization in Czech

Institute of Phonetics

Supervisor of the master thesis: Ing. Tomáš Bořil, Ph.D.

Study programme: Philology

Study branch: Phonetics

Prague 2023

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In Prague, Date: January 9th, 2023 signature of the author

First and foremost I would like to thank my supervisor Ing. Tomáš Bořil, Ph.D. for all the knowledge, patience, and guidance he has shown me. I would also like to thank my mum Zdeňka, my dad Richard, my sisters Karin and Anette, my girlfriend Tereza, my hakun-takun friends Kryštof, Adam, Honza, my friends Marco Serafini and Jakub Steinbach. Besides, I would like to thank all the amazing teachers and classmates for an incredible journey.

Title: Acoustic measures of glottalization in Czech

Author: Bc. Richard Mazúr

Department: Institute of Phonetics

Supervisor: Ing. Tomáš Bořil, Ph.D., Institute of Phonetics

Abstract: Word-initial vowel glottalization is used in Czech as a boundary signal. While its contextual occurrence is predictable, it is not always realized. In addition, it is not always realized as a canonical glottal stop. Acoustic parameters which would quantify and identify the various types of glottalizations are selected based on literature research and exploratory data analysis. These acoustic parameters are then used in a machine learning categorization model. Results show that with a small number of parameters, satisfactory results can be obtained, and thus these parameters are deemed suitable in characterizing these glottalizations.

Keywords: acoustics parameters categorization glottalization glottal stop

Název: Akustické parametry hlasivkových rázů v češtině

Autor: Bc. Richard Mazúr

Ústav: Fonetický ústav

Supervisor: Ing. Tomáš Bořil, Ph.D.

Abstrakt: Hlasivkový ráz na začátku slova před vokálem je v češtině užívaný jako hraniční signál. Zatímco jeho kontextuální výskyt je predikovatelný, není vždy realizován. Navíc není vždy realizován jako kanonická hlasivková explozíva. Akustické parametry, které by kvanitifkovaly a identifikovaly rozličné typy rázů, byly vybrány na základě průzkumu literatury a exploratorní datové analýzy. Tyto akustické parametry jsou následně použity v kategorizačním modelu strojového učení. Výsledky ukazují, že i s menším počtem parametrů je možné získat uspokojivé výsledky, a tím pádem jsou tyto parametry považovány za vhodné pro charakterizaci rázů.

Klíčová slova: akustika parametry kategorizace ráz hlasivková explozíva

Contents

Introduction

> *"The inability of Czech speakers to be conscious about glottal stop stems from the type of linguistic education in which phonemic function dominates the consciousness of language users, setting aside nearly all other functions of segments and other sound structures, whether these functions are phonological or non-phonological."*[1]
>
> Romportl (1984)

Brief history of Czech glottal stop

The oldest mention of the glottal stop, commonly referred to as *ráz* in contemporary Czech linguistics, which I could trace was as far as the early 20th century by Antonín Frinta (1909). As the initial quote alludes, usually it has been at the center of attention by linguists while being on the periphery of the attention of the general public. The level of interest given by the general public about the glottal stop should not be mistaken as reflective of its frequency or importance in spoken communication, but simply to the fact that it has no written representation in Czech script (Romportl, 1984). In fact, despite being invisible in written texts to common users of the Czech language, it is an inherent part of the language system. Similarly (but not as consistently) to the velarization of /n/ in Czech - it exists and is being used regardless of whether the speakers are conscious of the fact.

According to Chlumský (1933, p.172), the glottal stop was introduced to the Czech language (especially in Prague) through German. This claim is stated *en passant* merely as a footnote. While the idea of this being the case is not counter-intuitive and lends itself to an investigation, no other source that would confirm this notion was found by the author.

[1] Author´s translation, the original text: *"Neschopnost českého mluvčího uvědomovat si ráz pramení z toho, že v našem typu jazykového vzdělání dominuje ve vědomí uživatelů jazyka fonematická funkce hlásek a stranou jsou ponechány takřka všechny ostatní funkce hlásek i jiných zvukových útvarů, ať už jde o funkce fonologické nebo nefonologické"* Romportl (1984)

During a glottal stop's attested history (longer than a century) in the Czech language, it has been targeted by purists for removal (Havránek and Weingart, 1932), later it has been incorporated into the Czech orthoepic norm (Štěpánová, 2019), and recently it has been studied more closely from an acoustic perspective (Skarnitzl, 2004).

Goals

This thesis sets out to identify acoustic parameters relevant to the phenomena of word-initial vowel glottalizations in Czech language. In order to do so, terminology, its usage and development is researched. This step serves for facilitation for proper research-paper investigation on the topic of how this phenomena is studied. Another intermediary step is cleaning and general pre-processing of phonetically labeled speech material of Czech. Theoretical knowledge from research-paper investigation is then applied on the real data via selection of promising candidate parameters. Finally, exploratory analysis and categorization machine learning methods are employed in order to achieve the goal of identifying important acoustic parameters of said glottalizations.

1. Theory

1.1 Ráz, glottal stop, glottalization & others

In early 2000, Palková et al. (2004) published a paper called *"Stabilization of some terms for the phonetic description of Czech in relation to new results of research"*[1]. It recommends, amongst others, making a difference between the term of *ráz* and *hlasivková explozíva*. Reserving the former for an umbrella term, while the latter for it's specific realization. This recommendation stems from the fact that the glottal stop (i.e. the IPA /ʔ/) is referred to as *ráz*, *hlasivkový ráz*, *hlasiková explozíva*, *laryngální explozíva*, *glotální jev*, *glotalizace*, and possibly others in Czech linguistics as is described below. Specifically *ráz*, *glotalizace* and *glotální jev* have a unique position in the linguistic terminology which had to adapt due to new scientific findings.

These terms are prone to be used interchangeably but not always felicitously. Depending on the author, the era, and the topic it may refer to either:

- a canonical segment of [ʔ]
- a canonical segment of [ʔ] in a specific context
- an umbrella term for any type of glottalization
- an umbrella term for any type of glottalization in a specific context.

Another level of ambiguity is added when these terms are translated into English. The following section is an overview of how these terms were used throughout time and how their meaning might have changed using contemporary research findings as a comparative reference.

1.1.1 Usage in Czech literature

Frinta (1909) is the originator of the term *ráz*. According to Chlumský (1928, p.144), the term was a result of the translation of the French term *coup de glotte*.

[1] Author´s translation, the original text: *"Stabilizace některých termínů pro fonetický popis češtiny v závislosti na nových výsledcích výzkumu"* (Palková et al., 2004)

However, it is interesting to see that plosive, at the time, was called *ražená*[2], and that the term *ráz* is coined directly after classifying it as one (Frinta, 1909, 41). Chlumský (1933, p.144) mentions that *ráz* is an ambivalent term and that he recommends the usage of *hlasivková explozíva*. However, the nature of this ambivalence is not addressed, and the reader is instead redirected to his work *Kvantita* (Chlumský, 1928, p.144), where the issue is reinstated but not further developed. In both of these cases, *ráz* and *hlasivková explozíva* are used in the sense of a segment occurring in a specific context, but there is no ambiguity about the nature of the segment being canonical [ʔ]. For both authors, the terminology issue lies purely in whether one term is better than the other one. At this point in time, it seems that the need for an umbrella term is not present.

Romportl (1984) uses exclusively the term *ráz*, but he is aware that the canonical realization [ʔ] is not the only possible one. He brings to attention a study by Hlaváč and Pech (1981) who reported that a more open realization is more frequent than the canonical one. It is, therefore, possible that the awareness of multiple realizations amongst Czech scholars was spread in the 90s. In the English version of this text, *ráz* is translated as a glottal stop.

In the more recent works, Palková (1994, p.55) cautiously uses "so-called *ráz*", classifies it as *hlasivková explozíva*, and relates it with the boundary signals which can be manifested in diverse ways. It is not clear whether this relation was meant to bring the reader to the conclusion that also the "so-called *ráz*" can be manifested in diverse ways, because this relation is made between different chapters. Nevertheless, a decade later, Palková et al. (2004) advocate establishing the problematic term *ráz* as the *"umbrella term for various realizations of boundary signal...perceived as initiation...of a vowel"*[3]. *"Various realizations"* refer to the fact that we need to be able to refer not only to canonical realization, *"boundary signal"* refers to the function of these realizations, and *"initiation... of a vowel"* refers to the context in which these realizations fulfill the given function. Another term used here is *glotalizační jevy*, which is presumably synonymous with *glotalizace* (glottalizations). This term is an umbrella term for the glottal stop, creaky voice, breathy voice, and whisper. The author's understanding of this is

[2] full classification of this sound is *"ražená, neznělá souhláska hlasivková"* Frinta (1909)

[3] Author´s translation

that a boundary signal before a vowel has many realizations (not necessarily in form of glottalization) which can be (for the sake of efficiency) called *ráz* (1.1).

Research by Skarnitzl (2004) precedes chronologically the work by Palková et al. (2004), however, it was published later. It is written in English and does not operate with Czech terms. Nonmodal phonation and glottalization(s) are terms used to describe the means by which a boundary is signaled. This is in line with the previously mentioned research by Palková et al. (2004). The fact that it is in the context preceding a vowel is explicitly stated in the title and reinstated multiple times - *"in the context of the Czech conjunction 'a'"*. This is not a very compact way, however, after being declared once, "glottalizations" can be used similarly to how *ráz* was defined by Palková et al. (2004). Otherwise, confusion is possible because the term "glottalizations" does not automatically relate to any specific context or boundary signals.

Volín (2012) uses *ráz* as an umbrella term for multiple possible realizations including *laryngální okluzíva* (and thus, synonymous to *hlasivková explozíva* or glottal stop). He also points out that, in accordance with international terminology, *glotalizace* (glottalization) is used to describe both glottal stop and creaky voice. He then proceeds to investigate the glottalizations manifested in the word-initial vowel context. Volín (2012) directly addresses the concept of boundary signal. He extends the definition of *ráz* with a description of what constitues the boundary signal: *"physiologically and acoustically different events that fulfill the same function"*[4].

It would seem that the terminological state of *ráz*, *hlasivková explozíva*, *glotalizace* and others is not transparent. In order to avoid this, it seems to be necessary to be specific about the segmental context, and to respect the recommended terminology.

- glotalization(s): glotalizace, *glotální jevy*
- umbrella term for various realizations of boundary signal: *ráz*
- glottal stop: *hlasivková explozíva*, *laryngální okluzíva*, *hlasivkový ráz*, *ráz*

[4] Author´s translation, the original text: *"fyziologicky i akusticky různé události plnící stejnou funkci"* Volín (2012)

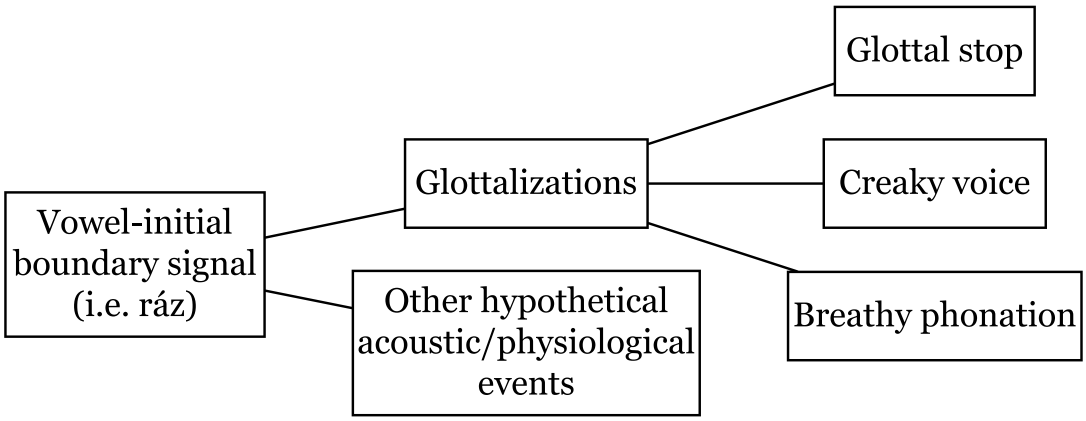

Figure 1.1: Representation of *ráz*

1.1.2 Usage in international literature

In the mid-70s, Lehiste (1965) conveyed research on boundary signals of Finnish, Czech, and Serbo-Croat. She reported that the canonical glottal stop does not appear in the word-initial vowel context as often as laryngalization does. Laryngalization is used as an umbrella term for any creaky voice - like realizations which are verified for example by Ashby and Przedlacka (2014, p.283). Since context and considered realizations are presented in addition with the connection to juncture, this study complies with contemporary terminology.

The study by Ashby and Przedlacka (2014) overviews the terminology regarding glottalization and glottal stops in English. Ashby specifies that in his research, the terminology is used in a specific phonological context. Glottalization, in the sense employed in this paper, serves as a proxy to two different processes found in English: glottal reinforcement and glottal replacement (Ashby and Przedlacka, 2014, p.284). Those are defined as *"glottalization...as an accompaniment to, or as a complete replacement for, certain voiceless oral stops"* (Ashby and Przedlacka, 2014, p.283). Ashby then mentions that the articulation of the concerned glottalization is usually associated with the canonical glottal stop, however, states that this is not necessarily reflective of reality, and that other realizations may be found or even more frequent. This is strikingly similar to how *ráz* is usually described or explained with glottal reinforcement being the closest term found so far by the author.

Bissiri et al. (2014) uses the term glottalization as an umbrella term for glottal stops and creaky-voice. She continues to note that: *"glottalizations can be good indicators of word boundaries"* (Bissiri et al., 2014). This is in agreement with the Czech conception established by Palková et al. (2004) and Volín (2012), that it cannot be automatically linked with the role boundary signal.

1.1.3 Recommended usage

There are even more terms to be specified such as the difference between laryngalization, creaky-voice, and creak, or the special position of hard onset (*tvrdý hlasový začátek*, *předráz*). Only hard onset will be mentioned since it is linked with methodology further in this thesis 2.1.5. It is another term that can be used instead of glottal stop [ʔ]. However, in this thesis it is considered as contextually specific to the beginning of an utterance or by being preceded by a pause within which articulators stop their activity (Volín, 2012). Again, such a definition is not flawless and is not used uniformly in Czech (Palková, 1994, p.55) nor in English (Ashby and Przedlacka, 2014, p.285). For the remainder of this thesis, glottalization will be used in the sense of nonmodal phonation found in the word-initial vowel context in the Czech language, unless specified differently. The context in which glottalization usually occurs (or should occur) is specified in 1.2.2. The next section sets out to clarify the term of boundary signal, which was already used often.

1.2 Boundary signal, juncture & diaereme

Glottalization in the word-initial vowel position serves a boundary signal (*hraniční signál*) or so-called juncture (*předěl*). Palková (1994, p.129) notes there is a difference between the two terms depending on conceptual preferences, with boundary signal being associated with the work of Trubeckoy and juncture with Bloomfield, however, she decides to use these terms interchangeably. Bičan (2006) summarizes these differences in approaches by stating that the Bloomfieldien juncture was deemed as a phoneme - a view non-compatible with the functional phonology of Trubeckoy.

In this thesis the descriptions as in Palková (1994), which seem to be inclined

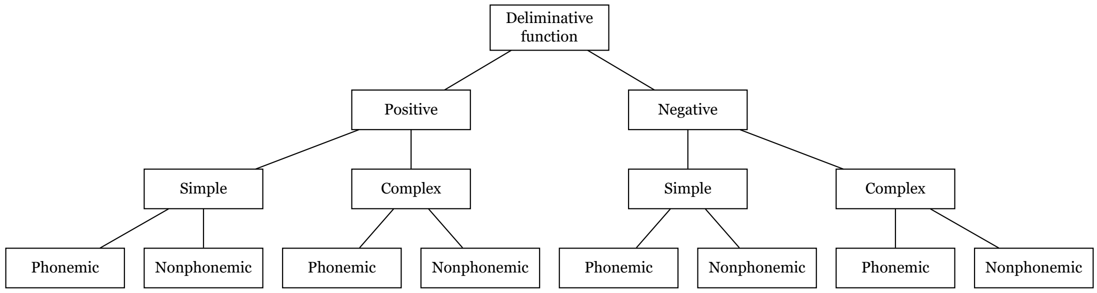

Figure 1.2: Trubeckoys categorization of boundary signals adapted after Bičan (2006, p.47)

towards the definition of boundary signal by Trubeckoy, are followed. Using this definition, it would classify Czech glottalizations as positive simple non-phonemic[5] boundary signals. This is done knowing that there are indeed terminological and conceptual difficulties to be taken into account. As reported by Šedivá (2022, p.15), this has been investigated by Bičan (2006, 2008). He proposes new terminology of *diaereme* that would reflect boundary signals phonologically in a more complete manner, while not being burdened by concepts associated with the term juncture. *Diaereme* is defined by Bičan as para-phonotactic feature[6]: *"whose base is generally a phonological form of a word - it accompanies the word so that it determines its boundaries"* (2006, pp.22-23).

This term might prove useful in future research, however, as this thesis is not phonologically oriented, the term boundary signal will be used for simplicity.

1.2.1 Functions

Deliminative

The presence of a boundary signal serves primarily to fulfill the deliminative (demarcation) function (Palková, 1994, p.128). In the case of this thesis, this boundary signal is glottalization and as such it facilitates speech processing through chunking (Skarnitzl et al., 2016, p.143) (e.g. [tak?ahoj], [ne?ustále]). So far, the deliminative function of glottalizations in relation to prosodic units in Czech was only studied before the conjunction "a" by Skarnitzl (2004) to the best knowledge

[5]Even though a handful of minimal pairs does exist, glottalizations are not considered phonemes: /proudil/x/pro?udil/; /potokem/x/pot?okem/

[6]This term is employed by Bičan as an alternative to prosody and suprasegmentals

of the author. It does imply that the glottal stop occurs more often on higher prosodic boundaries than other types of glottalizations, while creaky-voice type of glottalization occurs more often on the lower levels. In this section, it should be mentioned that glottalizations are not the only means which can be employed to signal boundaries in Czech. Amongst others, when at least two potentials of stress are realized, then one of the accentuated syllables will be relatively more prominent than the other. In a language with fixed stress, such as Czech, this prominence effectively serves the deliminative function (Skarnitzl et al., 2016, p.143).

Sociocultural

The presence or absence of glottalization (and its effect on the surrounding segments) can also be indicative of dialect. It is described by Romportl (1984) or reflected in the orthoepic rules by Palková (1994, pp.325-327). Recently, this phenomenon was more closely studied by Bortlík (2014) who, in this pilot study, investigated whether the difference between Bohemian and Moravian dialects in glottalization usage could be generalized. Due to the limited number of participants and experimental setting, no conclusion was made.

Affective

At the end of his research, Volín (2012) mentions a few examples of when glottalizations, outside of its usual context, can serve the affective function: [neʔ] or [cojeʔ].

1.2.2 Theoretical distribution

It was already mentioned that the attitudes towards glottalizations in the prevocalic context were diverse throughout the 20th century . Whether this was due to the alleged origin from Germany or some other reason is beyond the scope of this work. However, it can be stated that the amount of material produced regarding orthophony and orthoepics in this era was quite high. An extensive overview of the orthoepic material was created by Štěpánová (2019). According to this overview an important work in the field of orthoepy was done through the publication of (Hála, 1967) where the rules for Czech orthoepy can be found. It is however mentioned that these rules are better organized in Palková (1994) while staying true to the source material (Štěpánová, 2019, p.198). She also states that most of the academic texts dealing with this topic quote the work of Palkova or of Hurkova (Štěpánová, 2019, p.96). Taking this into account, Palková (1994) is used as a source for an outline of the orthoepic distribution of glottalizations. It is noteworthy that it is hard to judge whether the source material was a descriptive or prescriptive endeavor, and how the situation changed after nearly 30 years.

In theory, most of the time *ráz* is facultative, but due to its deliminative function is usually recommended in the proper context. This "proper context" is always conditioned by whether or not a word starts with a vowel.

Preceding context	**Usage**	**Example**
Non-syllabic preposition	Obligatory	/kʔoknu/
Unaccented mono-syllabic word	Recommended	/bilʔospaliː/
V-final preposition	Recommended	/poʔulitsi/
V-final prefix	Recommended	/neʔustaːle/

Table 1.1: Orthoepic rules regarding usage of glottalization adapted after Palková (1994, pp.325-327)

An overview of how voicing is influenced, depending on whether or not glottalization is used, is included for the sake of completion.

The question of voicing of C-final words when glottalization is absent brings us back to the socio-cultural function. The variant realization of *"dub opadal"* may be /dubopadal/ which is one of the realizations associated with Moravia (Palková, 1994, p.326).

1.2.3 Realizations

In general, glottalizations are classified on a continuous scale from the most open to the most closed with breathy, modal, and creaky phonation in-between serving as points of reference along the scale (Gordon and Ladefoged, 2001). The canonical glottal stop corresponds to one of the extremes where the glottis is in a fully

Voicing when glottalization is present		
Preceding context	**Final voicing**	**Example**
pair C-final word	Unvoiced	/dupʔopadal/
Voicing when glottalization is absent		
Preceding context	**Final voicing**	**Example**
C-final prefix	Voiced	/bezotkladně/
C-final monosyllabic preposition	Voiced	/podoknem/
C-final word	Unvoiced	/dupopadal/

Table 1.2: Orthoepic rules regarding assimilation of voicing depending on glottalization usage adapted after Palková (1994, p.325-327)

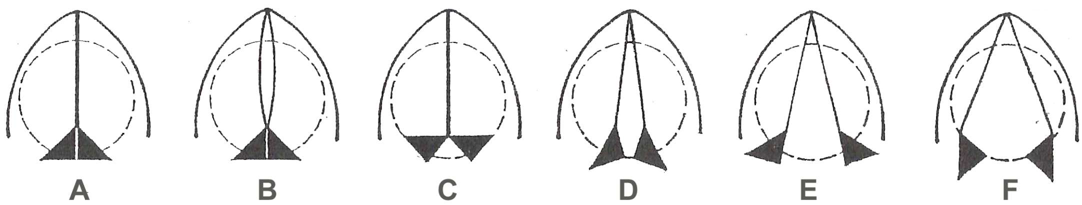

Figure 1.3: Continuum of phonation types. From closed glottis to open glottis - (WikimeadiaCommons, 2005)

closed position.

The fact that canonical glottal stop might not be the most frequent realization of the boundary signal created by glottalizations in the Czech language was, as far as the author can tell, first reported by Lehiste (1965). The more frequent variant seems to be the creak - which is the more open variant of the canonical glottal stop. Confirmation and extension of this finding can be found in a study by Skarnitzl (2004) which introduced a detailed classification. It was inspired by findings of Redi and Shattuck-Hufnagel (2001) and Pompino-Marschall and Żygis (2010), however, it differs in the fact that each realization is assigned multiple labels which results in a simple hierarchy and thus more detailed categorization. 126 glottalized segments were inspected in this study. They were divided into glottal (51) and creak (75) types and then further subdivided. The final category schema is adapted after Skarnitzl (2004).

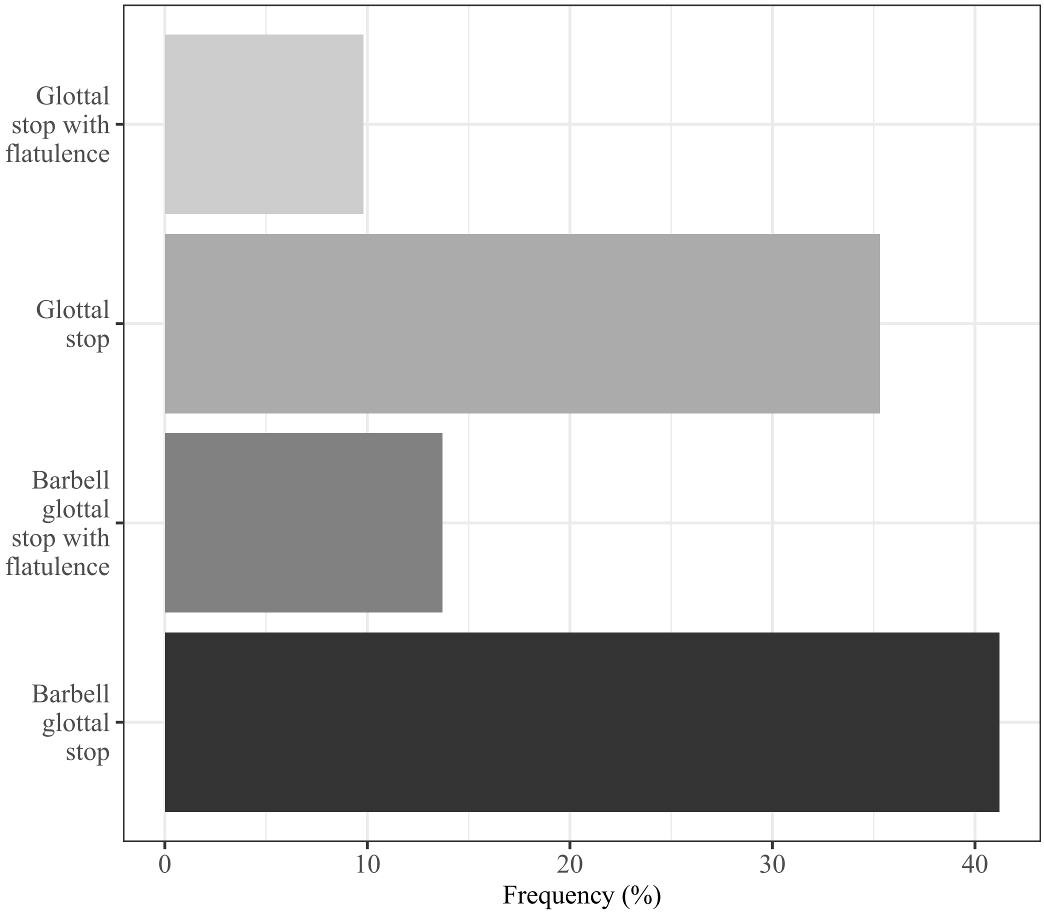

Figure 1.4: Distribution and categorization of glottal type

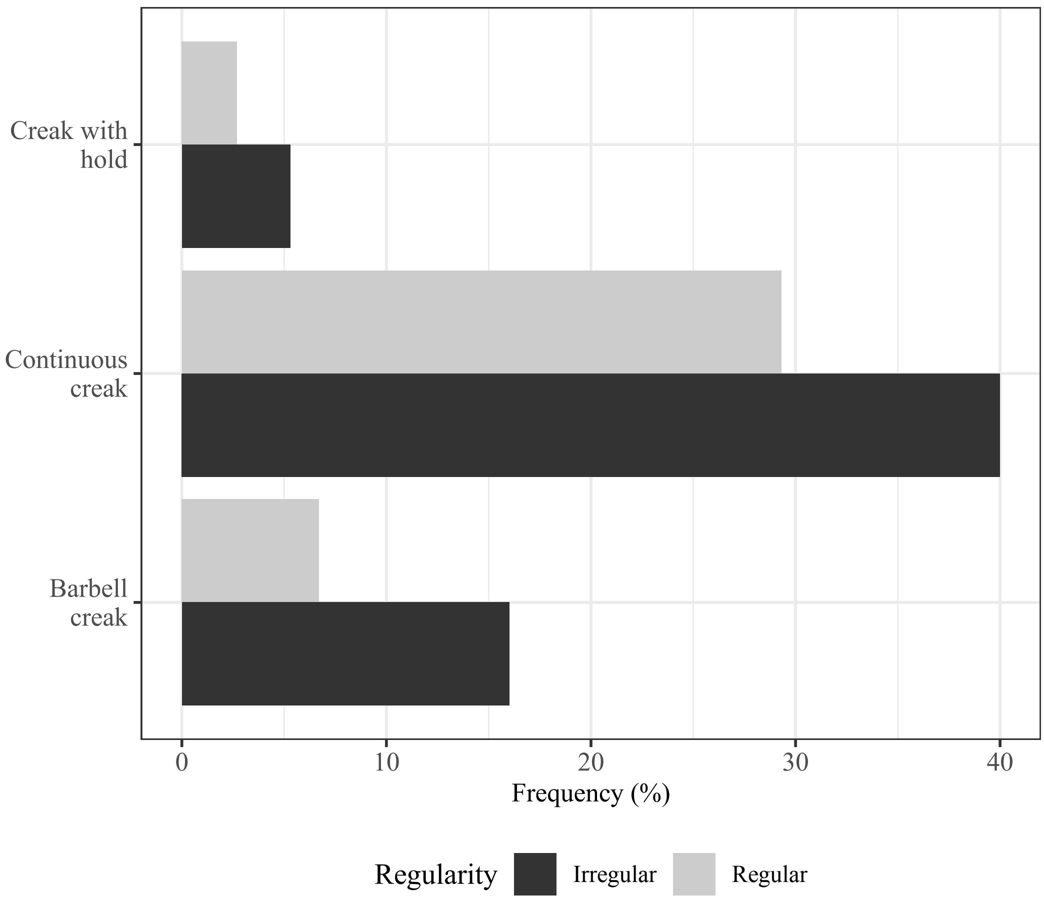

Figure 1.5: Distribution and categorization of creak type

1.3 Research on glottal activity

This thesis presents an overview of studies that were concerned with glottalizations in word-initial vowel positions. In addition two publications which study this phenomena within a different context are present Redi and Shattuck-Hufnagel (2001) and (Ashby and Przedlacka, 2014). The former is noted due to its influence on research by Skarnitzl (2004) and the latter for the relevance for this study despite its different scope. Many of these studies were focused on another language and so the results might not be one-fit-them all, nevertheless they might serve as an important source of inspiration.

In principle, the aim of these studies can be divided into those concerned with distribution (where and how often do glottalizations occur), categorization (what different types of glottalizations can we distinguish), and parametrization (what acoustic parameters can best describe glottalization), however, such division is only orientational, because they are often intertwined.

1.3.1 Distribution

In a study that was already mentioned, Volín (2012) sets to investigate the difference in distribution between speech style, sex, and rate of glottalization in Czech. Working with two groups of 10 participants each, where sex was represented equally (each group had 5 male and 5 female participants), one group of the professional broadcast announcers and one of the university students. Results were based on read text retrieved from the broadcasted news, and on semi-spontaneous material retrieved from a pair activity. They showed that both speech style and sex have an impact on the rate of usage of glottalizations. Female participants and read texts achieved both higher rates than males and semi-spontaneous speech. The highest results were, therefore, achieved by females who read (broadcasters), and the lowest by male semi-spontaneous speech.

A study of English glottalization by Dilley et al. (1996) used a pre-existing prosodically-labeled corpus of five professional broadcasters (2 males and 3 females). In combination with subsequent glottalization labeling investigation of the effect of prosodical boundaries, pitch stress, and segments on the distribution

Sex	Read (%)	Unprepared dialogue (%)
Male	88,30	41,10
Female	96,90	64,90

Table 1.3: Findings of the study adapted after Volín (2012)

of glottalization. The conclusion drawn from this result was that the rate of glottalization rises when vowels occur at the start of a new intonational phrase, when the target word is marked with a pitch accent, and that it is influenced by the preceding segment at phrase internal boundaries (as opposed to phrase initial).

Besides some of the already mentioned tools (such as phrasal position and word accent), Pompino-Marschall and Żygis (2010) investigated the rate of glottalization in German also using word type (content and functional), speech rate, and following vowel. Pre-existing spontaneous speech material of three famous German citizens was used. There were 9 recordings in total - 3 per subject. After annotating the appropriate occurrences of glottalizations and having the number of syllables at disposal, the analysis of speech rate in relation to glottalizations was possible. The speech rates were divided into four groups - slow, slow medium, fast medium, and fast. The general trend for each subject was that with increased speech rate, canonical glottal stop decreased while creaky voice and the absence of any glottalization increased. This rate was still true, but less strongly when the context of the glottalization was separated into content and functional words. In addition, low vowels were identified as being glottalized by the glottal stop more than mid and high vowels. These three effects were tested also by Skákal (2015) for Czech, however, the results of this study were not statistically significant.

1.3.2 Categorization

The previous studies were concerned with the distribution of glottalizations in general - mostly differentiating only one or two types - canonical glottal stop and creaky voice. In the work by Redi and Shattuck-Hufnagel (2001) on glottalization in English, in order to help find glottalized regions (which were not reserved to word-initial-vowels), four acoustic categories were proposed: aperiodicity, creak, diplophonia and glottal squeak. Any given token could have exhibited none or all of the characteristics assigned to the categories, however, it seems that in the end each token was given only one final label. These characteristics included variation/(ir)regularities in f0 and period-to-period duration/shape/amplitude. Working with two different corpora - one with 6 professional radio announcers (3 female, 3 male), which were already labeled prosodically, and 4 nonprofessional readers (2 female, 2 male); second with 4 nonprofessional readers (3 female, 1 male). Amongst other results, it provided an important insight into the individual preference of different glottalization types of speakers. Directly inspired by this study, Skarnitzl (2004) also set out to categorize the glottalizations in Czech. Although the findings for Czech are already described in more detail in another section of this thesis, a few points should be elaborated on. Mainly, categories of squeak and diplohonia were not attested in the study by Skarnitzl, and contrarily to the study by Redi and Shattuck-Hufnagel, the categories were given a hierarchy. As was mentioned, the methodology about how many attributes could be assigned to one label was not clear, but it seemed there were 4 glottalization types in the end - one for each proposed acoustic category. Skarnitzl (2004) assigns multiple attributes to each category - such as periodic continuous creak or barbell glottal stop with flatulence.

1.3.3 Acoustic properties

The various categories emerging from research set themselves apart not only visually but also acoustically. Various tools and measures have to be employed to quantify their acoustic difference.

While the glottalization context of the study by Ashby and Przedlacka (2014) is not word-initial-vowels, but rather voiceless oral stops, the need for measuring the acoustic properties of glottalizations remains. In this study of English, special processes are being investigated - replacement and accompaniment of glottalizations at some voice-less stops. A pre-existing segmentally labeled corpus, that was created for a sociolinguistic study, was re-used. It contained words and phrases elicited from one-to-one interviews with 22 teenage speakers (12 female, 10 male). The most promising feature differentiating glottalized and non-glottalized segments was the use of the autocorrelation function (f0 and energy were reported as not as effective) which was hypothesized to represent the regularity of the shape of the soundwave.

By far the most extensive research when it comes to the number of used parameters was done by Garellek (2012), ranging from prosody related to lexical or segmental. However, these are already covered in the previous section of the thesis - the main focus of this section are the acoustic parameters. The study was made on the case of glottal-stop and other glottalizations occurrence before word-initial-vowels in English. The used material was the same as by Dilley et al. (1996) and Redi and Shattuck-Hufnagel (2001), and as such already prosodically labeled. This corpus was subsetted to 4 speakers (2 females, 2 males). In general, Garellek followed the methodology of Dilley et al. and Redi and Shattuck-Hufnagel, however, any type of glottalization was kept - with the exception of allophone of /t/; glottal squeak as a category was also discarded similarly as done by Skarnitzl (2004). The acoustic parameters used are presented in 1.4. Garellek (2012, p.11) states, that these parameters are not measurable on a voiceless glottal stop and thus, uses them to characterize vowels that follow it.

Parameter	**Description**
f0	fundamental frequency
Duration	length of vowel
H1*-H2*	difference between amplitudes of first two harmonics
H1*-H4*	difference between amplitudes of second and fourth harmonics
H1*-A1*	difference in amplitudes of first harmonic and harmonic nearest F1
H1*-A2*	difference in amplitudes of first harmonic and harmonic nearest F2
H1*-A3*	difference in amplitudes of first harmonic and harmonic nearest F3
CPP	noise measure
Energy	measure of loudness

Table 1.4: Acoustic parameters used by Garellek (2012)

2. Practical part

2.1 Method

2.1.1 Chosen parameters

There are four main types of measures we can specify depending on which so-called domain we operate in.

When the sound is being recorded by a microphone, the pressure of air molecules being pushed by sound waves is measured in time. Through digitalization (sampling and quantization), it is turned into a time domain which we see as amplitude changes in time on an oscillogram.

Using a Fourier transformation we can change our point of view, and see a static composition of the frequencies distribution and their relative intensity via the spectrum. We can also view the change through time of spectra by dividing the time domain into short windows, composing their spectra, and concatenating them using a pseudo-3D representation of them on a spectrogram - a time-frequency domain.

Last but not least, we can also do a special operation known as the inverse Fourier transform of the logarithm of the power spectrum in order to gain a representation of the periodicity of the spectrum - known as the cepstral domain.

Each of these domains has its own benefits and introduces different ways on how to measure parameters that in general, relate to the basic acoustic domains: frequency, duration, intensity, and spectrum.

Duration

While not being compared to another segment, Ashby and Przedlacka (2014) shows that the duration of glottalizations in the specific context of his study was clustered together around 125ms. Skarnitzl (2004, p.65) also measured duration - his results for average duration of considered categories were between 62,90 ms and 96,40 ms. Should glottalizations exhibit different duration distribution than other segments, potentially as a parameter, it might contribute to its identification.

Fundamental frequency

As it is stated by Ashby and Przedlacka (2014), even the mean f0 of glottalizations is allows to differentiate men from women (even though it was not conclusive for females in his study, it was attributed to the failure of f0 estimation). Since our analysis does not utilize any categorical parameters, using an acoustic parameter characterizing sex might be helpful even if it seems inconclusive.

Jitter

"Fundamental frequency perturbation measure" is a more explanatory term for this parameter - sound waves differ in how regular their cycles are, and jitter quantifies how a sound fluctuates in terms of the distance between its cycles (Titze et al., 1987). When it comes to the creak type of glottalization, as defined in 1.2.3 irregularity, it is one of the main acoustic cues. Ashby and Przedlacka (2014) used auto-correlation as an acoustic indicator of glottalization which, as he observes, showed minimum when jitter peaked. Since jitter is easily interpretative and is also recommended by Skarnitzl (2004, p.67), it is included as a parameter for our dataset. Jitter can be, however, unreliable to some extent in practice. It is reported by Boersma (2009) or by Titze and Liang (1993) that in order to acquire its measure, one must first somehow extract the onset time of the glottal pulse - for this purpose, there are various algorithms, but they have their limitations. These algorithms tend to fail when the jitter surpasses 1,00 %, i.e. when it enters the classification of very aperiodic (Titze and Liang, 1993, p.1132)(Boersma, 2009, p.308).

Skarnitzl (2004) counts the classification with the fact that some creaks are more irregular than others. For the extraction, it might fail for any glottal type since voicing is not expected there (and so absent f0). Higher jitter is also to be expected between regular creak type and potential context of glottalization. The failed extraction of f0 and consequently of jitter might, therefore, be indicative of either the glottal type or the irregular creak type.

Glottal noise measures

Harmonics-to-noise ratio (HNR) is one of the parameters used by Garellek (2012) - it quantifies the degree to which periodic and noise components contribute to a sound wave signal (Yumoto et al., 1982). This is usually calculated for multiple frequency bands: 0-500 Hz, 0-1500 Hz, 0-2500 Hz and 0-3500 Hz. HNR's resulting value is in dB.

Values within the 0-1500 Hz band were reported by Garellek (2012) as much lower between vowels preceding a non-preceding glottalization. As for the 0-2500 Hz band, he found that values for these segments are higher. HNR can be calculated in different domains and by various algorithms. Detailed explanation is provided in section 2.1.5. It is used in our dataset as a parameter that might differentiate the presence from absence of glottalization since it seems it has the ability to do so indirectly through vowels.

Spectral tilt

In general, there are two types of spectral tilt - long-term and short-term. For specific segments, short-term spectrum spectral tilt (in literature, also harmonic amplitude measures) is used (Skarnitzl, 2014). As the name implies, it is calculated from the spectral domain. Peaks of different harmonics and formants are compared, resulting in a measure representing the tilt between them. These tilts or slopes can be steep or mild. Different slopes of various combinations of the harmonics and formants are correlated with different voice qualities. For example, the H1*-H2* (first and second harmonic) is considered correlate of the open quotient (OQ) (Hanson, 1995, p.44). This open quotient relates to the openness of glottis (see 1.3) - since glottalizations are known to be inclined towards the closed extreme of this continuum, it could be used to differentiate between vowels and glottalizations.

These parameters were also employed by Garellek (2012), and showed that vowels following a glottalization have lower H1*-H2* than those which are not preceded by it. Since H1*-H2* is correlated with OQ, lower H1*-H2* measures represent more closed position of the glottis which is to be expected from glottalizations. These parameters are used as they could be characteristics of glottalizations. The calculation of these parameters are explained more in detail in section 2.1.5.

Cepstral peak prominence

As described in Hillenbrand et al. (1994, pp.771-773), cepstral domain can provide important information similar to the other domains, with the added benefit of not relying on extraction of f0. When sound wave is converted into this domain, different quefrencies are on x axis while the cepstral magnitude is on the y axis. The maximum magnitude - the peak of some quefrency is indicative of the most periodic part of the spectrum. This is usually fundamental frequency. In order to compare different sounds or segments, comparing absolute values of these magnitudes is not used. Instead, some average tendency is inferred from the cepstrum, which is then compared with the maximum magnitude - the cepstral peak coefficient (CPP). Using this coefficient, different sounds may be compared.

This parameter was also used by Garellek (2012), but was studied in relation to phrasing. It is, however, another way how to look at periodicity which is important aspect of glottalization. CPP for glottalized context might be lower (which represents lesser periodicity of the frequency), especially for the glottalized segment itself. Thus, it is included as another parameter for the dataset.

Mel-frequency cepstral coefficients

While not being addressed by any of the over-viewed works in section 1.3, Mel-frequency cepstral coefficients (MFCCs) provide another way on how to look at the sound wave. The same principle of non-reliance as in 2.1.1 applies. The method of how to arrive at the cepstral domain is, however, a little bit different. Godino-Llorente et al. (2006, p.1945) demonstrates that frequencies are transformed using mel-frequency banks, and instead of using inverse Fourier transformation, discrete cosine transformation is used. The resulting coefficients (usually 12-13) then describe the spectral shape of the segment. This is similar to what short-term tilt parameters also approximate. MFCCs are also used as parameters in the dataset since they may contain complementary information about the shape of glottalizations.

2.1.2 Hardware and software configuration

The entirety of the work was done on hardware with these specifications:

- Processor: 11th Gen Intel(R) Core(TM) i5-1135G7 @ 2.40GHz 1.38 GHz
- RAM memory: 8,00 GB
- 64bit OS, processor for x64 platform
- Edition: Windows 10 Pro

In order to make the necessary calculations, the following tools were used - their settings were kept to default unless stated otherwise:

- VoiceSauce v1.37, an implementation of Matlab for phonetic research
- Praat version 6.2.10, specialised tool for phonetic research
- Python version 3.8.10, programming language
- R version 4.2.2, programming language

The libraries used within respective programming languages:

- Parselmouth version 0.4.3 based on Praat version 6.1.38, library for controlling Praat from within Python (Jadoul et al., 2018)
- rPraat version 1.3.2.1 (Bořil and Skarnitzl, 2016), library for controlling Praat from within R including specialised function not found in the source Praat tool
- Tidyverse version 1.3.2 (Wickham et al., 2019), library for working with and visualising tabular data
- caret version 6.0-93, library for machine learning (Kuhn, 2019)
- corrplot version 0.92, library for visualisation of a correlation matrices (Taiyun et al., 2021)

The core of the work was done using Praat and R (namely rPraat library), but some calculations were done, for reasons of convenience, by using VoiceSauce and Python (namely using Parselmouth library).

2.1.3 Overview of speech material

In 2009, the efforts of the Institute of Phonetics to study speech characteristics using phonetic corpora were re-affirmed by the creation of new material (Skarnitzl, 2009). A subset of the same material was made available for this thesis, and is used to create a dataset of glottalization, relevant parameters and their respective values.

The original material contains 24 pre-prepared dialogues and out of those, 16 have been annotated (and made available for this thesis). These dialogues are marked always as "R{number of dialogue}_{number of turn}_{subject ID}". Each dialogue had 3-5 turns. Not every turn had an expected context for glottalization to occur - in fact, 9 dialogues had at least one turn in which it was not expected. The dialogues were first read and then acted out by paired participants from the ranks of students.

dialogue ID	**R01**	**R07**	**R08**	**R09**	**R11**	**R12**	**R13**	**R14**	**R16**
total turns	4	4	4	4	5	4	3	5	5
turns without glottalization	1	2	1	1	1	1	2	2	1

Table 2.1: Number of turns in which there was no reason to expect glottalization occurrence

There were 40 different participants in the recording of these 16 files. 32 of these participants were of the female sex, remaining 8 of the male sex. Each participant had an average of 55 recordings, minimum 51, maximum 57.

These recordings were then annotated. This annotation was done first using Prague Labeller (Pollak et al., 2008), and then adjusted by 5 annotators manually. Since these annotators were students, errors are to be expected. The material was annotated using X-SAMPA.

The total amount of sound files was initially 2534 while there were only 2172 TextGrid files. Upon verification, not every TextGrid file had a sound file counterpart: 362 sound files did not have a corresponding TextGrid. Mostly, these missing annotations were from dialogues R04, R08, R10, and R16.

Since there were many different factors to verify and fix manually, such as word alignment (even after employing automatic boundary alignment), these missing

dialogue ID	R01	R04	R06	R08	R09	R10	R15	R16
missing files	1	80	40	78	1	1	19	142

Table 2.2: Number of missing TextGrid files per dialogue

TextGrids were not (with a few exceptions) created by the author. The justification for this is that due to the amount of analyzed data, these missing TextGrids should not play a significant role in the final distribution of values. Nevertheless, from the TextGrids, which were available, more than 25,00 % were checked manually. These were files that appeared often within the replacement of labels or word misalignment - namely R1, R2, R10, R14 and R15.

During the processing, several files were found as somehow erroneous. One sound file (R14_3_SOBA) was cut before the turn was over - this was an error within the TextGrid itself. Another sound file (R10_4_KADA) appeared to suffer from the same issue, however, could not be fixed using the same method. The last sound file (R10_1_BURA) did not have its text grid available which was, therefore, added. The final number of paired files was 2171. Every step described below is taken only on these paired files.

The descriptive statistics of the sound file duration are shown in the table below.

Measure	Value
Min.	0,78
Median	3,21
Mean	3,23
Max.	7,93

Table 2.3: Descriptive statistics of audio files duration in seconds

2.1.4 Cleaning speech material

The next step is to correct and normalize labels in word and phone TextGrid tiers as well as to verify their time alignment.

Segment labels

In the phone tier, segment labels usually contain typographical errors such as leading/trailing space, leading/trailing tab, or undesired capitalization. This can be checked using a free-access script (Bořil, 2019) which makes an overview of all used labels over all files. This script was modified by the author so that any found errors may be re-written with the correct value.

There were around 80 wrong labels. Considering the total number of segment labels was around 109 000, this seems like a reasonable error rate. In addition, for reasons of compatibility with the next steps, all empty labels were replaced with a #SP# label (Bořil and Skarnitzl, 2016).

Word labels

Part of the subsequent step 2.1.5 consists of searching for surrounding segment labels of contexts in which glottalization should occur but did not. This is done because we wish to compare the measures between segments surrounding glottalization, and those in which glottalization was not realized. In order to make the said step resistant to typographical errors, each word tier label was normalized in terms of:

- case: lower case
- interpunction: removed.

Time alignment

Alignment is another type of cleaning that had to be utilized. There are two alignments that had to be taken into consideration: whether phone tier and word tier boundaries correspond to each other, and whether segments are properly aligned.

The mutual alignment of phone and word tier can be done in part using a function from Bořil and Skarnitzl (2016) called boundary magnet, which aligns the closest boundaries together. This does not necessarily result in correct boundaries being aligned since the error might be too large - boundary segments may be aligned with the previous or following words instead of the actual word they

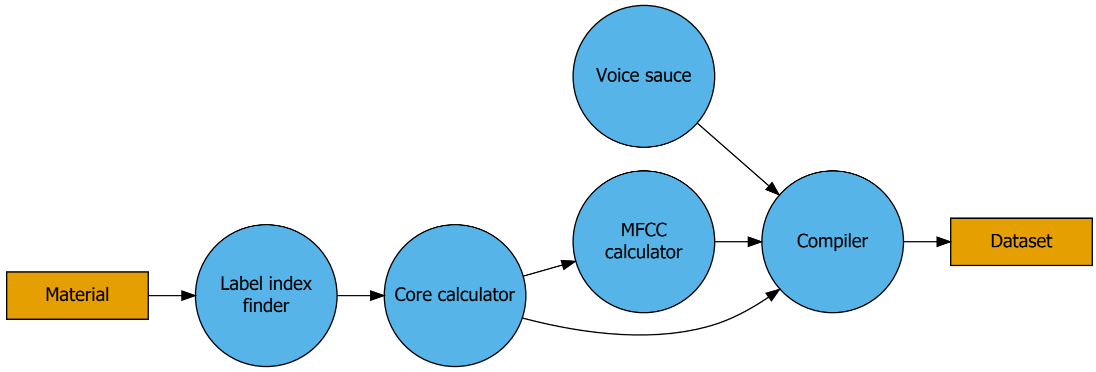

Figure 2.1: A schema of the steps taken after the material was checked for various types of errors and to some extent corrected

belong to. This was often the case within the R10 dialogue. Knowing that boundaries may not be properly aligned even after taking this step, manual check was done as well for those dialogues and turns, which seemed to be afflicted the most.

While phone and segment tier might be properly aligned between themselves, segments on their own might still misaligned. They were checked manually using Machač and Skarnitzl (2009) as a general guide. This was done primarily for those dialogues and turns which seemed to suffer from boundary misalignment in the previous step.

2.1.5 Turning speech material into data

Many precautions were taken regarding the quality of material. In order to have a table with desired measures of parameters - data, many steps have to be taken. They are represented in figure 2.1.

Label index finder

Before starting any computations, we first have to know what the index of the segment within its respective TextGrid is. We are interested in the segment of glottalization and its surrounding segments as well as the segments surrounding a potential place of glottalization occurrence which was not realized. Potentially, phrase-initial occurrences of glottalization (or occurrences after pause) could be removed as mentioned in 1.1.3, this was not done as it is assumed, due to the size of the dataset, this will not impact the results.

Each glottalization has the potential to occur in a context that is mainly predictable (see 1.2.2). The surrounding segments of a realized glottalization can be easily found by looking at the preceding and following segments of a glottalization label: [ʔ].

When glottalization is not realized, we still need to access the segments which would normally precede and follow it - but we cannot use the glottalization label to find these segments since it is not present. Instead, we can search on the word tier level for previous and following word labels of these contexts. Because the dialogues are pre-made, we know upfront what text is uttered in each turn of each dialogue. Thus, we also know between which words glottalization should occur - e.g. *"stačila otočit"*. Then, we can simply take the last segment of the first word, and the first label of the last word to have the desired indices for contexts in which glottalization should have occurred but did not.

These indices are saved alongside a unique identifier allowing us to use each index only for the relevant participant, dialogue, and turn. In other words, we are able to link indices with their respective TextGrids.

Core calculator

This module utlizes pure Praat script for its caluclations. The calculations include following parameters:

- Duration
- Jitter
- Average pitch
- CPP.

These are measured only for relevant indices and only for segments longer than 0.01 s. This duration was selected because many parameters are not measurable if the segment is shorter. All of these measures were done using Praat default settings.

MFCCs calculator

This parameter was calculated utilizing the Parselmouth Python library. It uses Praat for the calculations including its settings. The default setting for calculating MFCCs was used except for the window length, which was overridden by using a length of 0,01 s instead of 0,02 s. This is due to the fact that the segments start at 0,01 s duration. The calculation is done directly on the segment taken from the audio. Each coefficient is calculated for each time step. Therefore, longer segments would result in having more calculations within each coefficient (for example resulting in having 100 values per coefficient). Since there should be one MFCC value per coefficient, these length-dependent values had to be averaged. Each segment, therefore, has exactly one measure per MFCC.

VoiceSauce

VoiceSauce is a less used tool within the Institute of Phonetics in Charles University. Some explanation about the difference with Praat are in order. More detail information can be found on the website of VoiceSauce mnual:
http://www.phonetics.ucla.edu/voicesauce/documentation/contents.html.

HNR in Praat is recommended by the cross-correlation method while in VoiceSauce it is done in using a modification of the algorithm develop by Krom (1993), which is cepstral domain algorithm. The number of glottal pulses for this algorithm is 5 against 12, as indicated by Fernandes et al. (2018).

Harmonic-amplitude measures of the short-term spectral tilt are normalized or corrected, as is the standard. However, instead of using an asterisk (which is the standard) Skarnitzl (2014), a suffix "c" is used in VoiceSauce. This is due to the underlying Matlab restraints on the usage of "*".

As opposed to default settings, Snack was chosen as the f0 estimator since it was needed for the underlying calculations of other parameters. The f0, which is used as a parameter in this study, is calculated from Praat. Labelled TextGrids were not used in the VoiceSauce estimation process.

2.1.6 Cleaning speech data

Intermediary columns such as indices, or those generated by VoiceSauce (which are required for the desired parameters), are deleted.

Since measures such as f0 or jitter were often impossible to calculate, the resulting values such as ”–undefined–” by Praat or ”NA” by VoiceSauce were replaced by 0. An alternative would have been to replace it with another relatively extremely high or extremely low number (such as 9999 or -9999). Irretrievable value should have its representation in the dataset, since the inability to calculate a measure for a segment bears information.

Further, the columns are declared into appropriate types: factors integers, and floats.

2.2 Results

The final data contained **53 parameters** (columns) and 9409 rows. Out of these rows, 931 were for the context before and after the potential occurrence of glottalizations, 2519 for the preceding context of glottalization, 2508 for glottalizations, and 2520 for following it. The discrepancy between the glottalized/context of glottalized segments is the condition of minimal duration. Some of the segments were, therefore, too short and were not used in subsequent steps of the processing.

Position	Value
Before potential	931,00
After potential	931,00
Before glottalization	2 519,00
Glottalization	2 508,00
After glottalization	2 520,00

Table 2.4: Frequency count of segments within each type category

Many studies mentioned in this thesis reported that in general, women glottalize more often than men. As a form of checking, whether our data makes sense, we will check this. All the unrealized segments plus the number of glottalizations, together create the total amount of potential occurrences for the subjects to glottalize. If we take the number of glottalizations and divide it by this total per sex, a percentage-based result is obtained. Having the results in relative measures instead of absolute should prevent misrepresentation due to the imbalanced number of females against males in the dataset.

Sex	**Frequency (%)**
Male	71,70
Female	73,30

Table 2.5: Relative frequency of glottalizations from all potential occurence per sex

While greater differences were expected, females glottalize in general more often. The ratio itself is between the expected ranges of previous study by Volín (2012), as summarized in 1.3. The final results are summarized using averages and standard deviations for each type of segment in 2.6. Since raw data in a table are the best tool for data anylysis, exploratory data analysis using visualizations is used.

Parameter	Mean	SD	Parameter	Mean	SD
Duration	52,63	60,22	X1	1 118,60	301,13
CPP	65,03	13,49	X1	360,00	127,72
Mean f0	207,40	59,34	X3	17,26	78,82
Jitter	0,02	0,03	X4	28,69	52,08
HNR05	18,16	11,88	X5	4,69	44,88
HNR15	23,41	12,11	X6	−6,98	39,72
HNR25	26,69	11,43	X7	−41,26	40,34
HNR35	28,80	10,98	X8	9,75	34,41
H1A1c	14,60	8,22	X9	−30,36	28,25
H1A2c	13,54	9,85	X10	−2,21	27,47
H1A3c	7,19	10,57	X11	−7,64	23,61
H1H2c	4,09	5,79	X12	4,55	22,24
H2H4c	0,88	6,92	X13	−2,06	21,27

Table 2.6: Descriptive statistics of observed parameters

2.2.1 Exploratory analysis of data

As we want to consider reducing the amount of parameters before we start, we need to follow an informed process. Therefore, correlation graph is used. If we find two different parameters that are highly correlated with each other, we could decide to discard one of them - in theory, we should not loose too much of data variability explained.

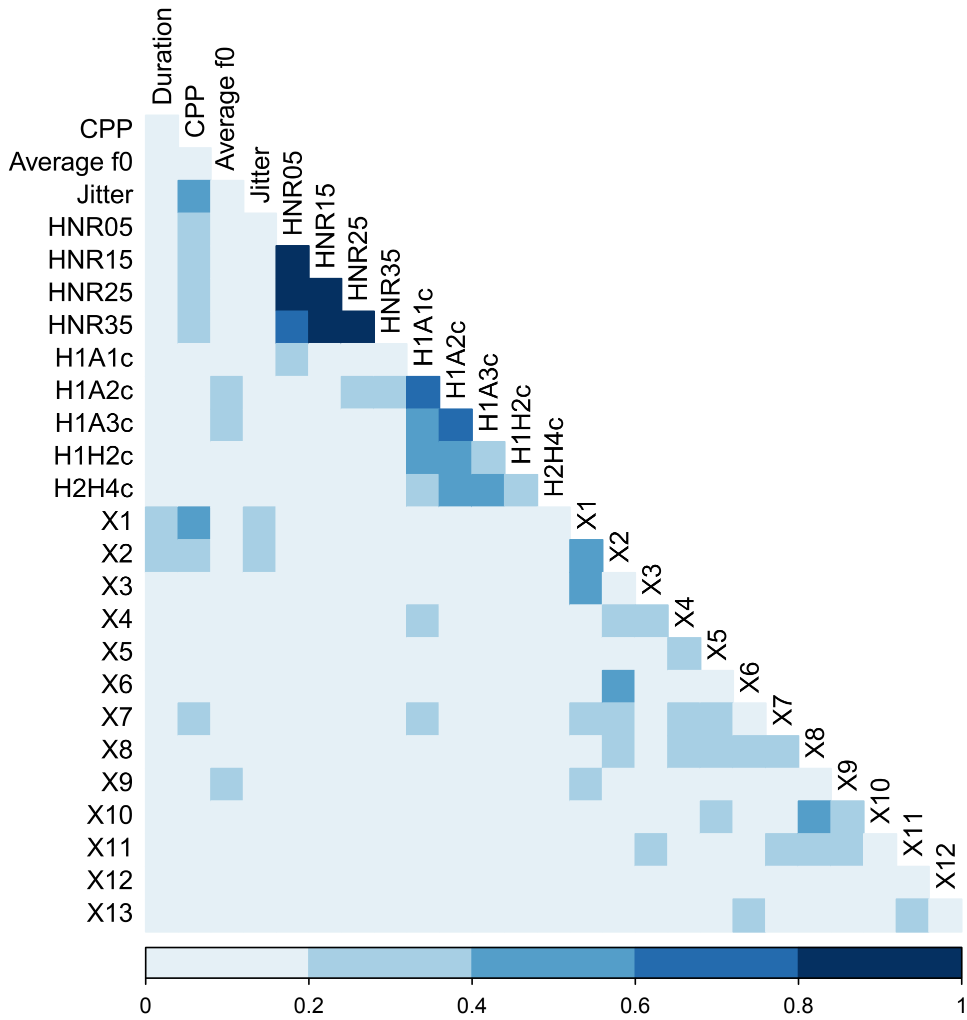

Figure 2.2: Pairwise-correlation of selected parameters

There are 3 distinct areas which pop up - each for the big group of parameters - the HNR, the short-term spectral tilt parameters and the MFFCs. The ones that are highly correlated are the prime candidates for being removed. These parameters are HR15 with HR25, and H1A1C with H1A2c.

Having 27 parameters, one of those being a category of each segment as found in 2.4, we can view each parameter as per its position: before or after, or based on its context: unrealized or realized. Glottalization itself is in a grey zone, and the data is split into two separate subsets - one only containing the data about glottalized segments, and second with the other segments. Two more columns are created for this second subset - one for the position and one for the context. This way we can plot the various combinations of segments while having the glottalization in the background (and for each plot the underlying glottalization values are the same across the combinations).

Using this setup, we start with the candidate for parameter reduction found by the correlation graph. While the result was not very different, perceptually HNR15 seemed to have a more distinct distribution. The bimodality found in the Realized Before quadrant might be indicative of some type of segment not having a HNR15 at all - since these non-existent values were replaced with 0. While not being very distinct, it does provide us with a better alternative than HNR25.

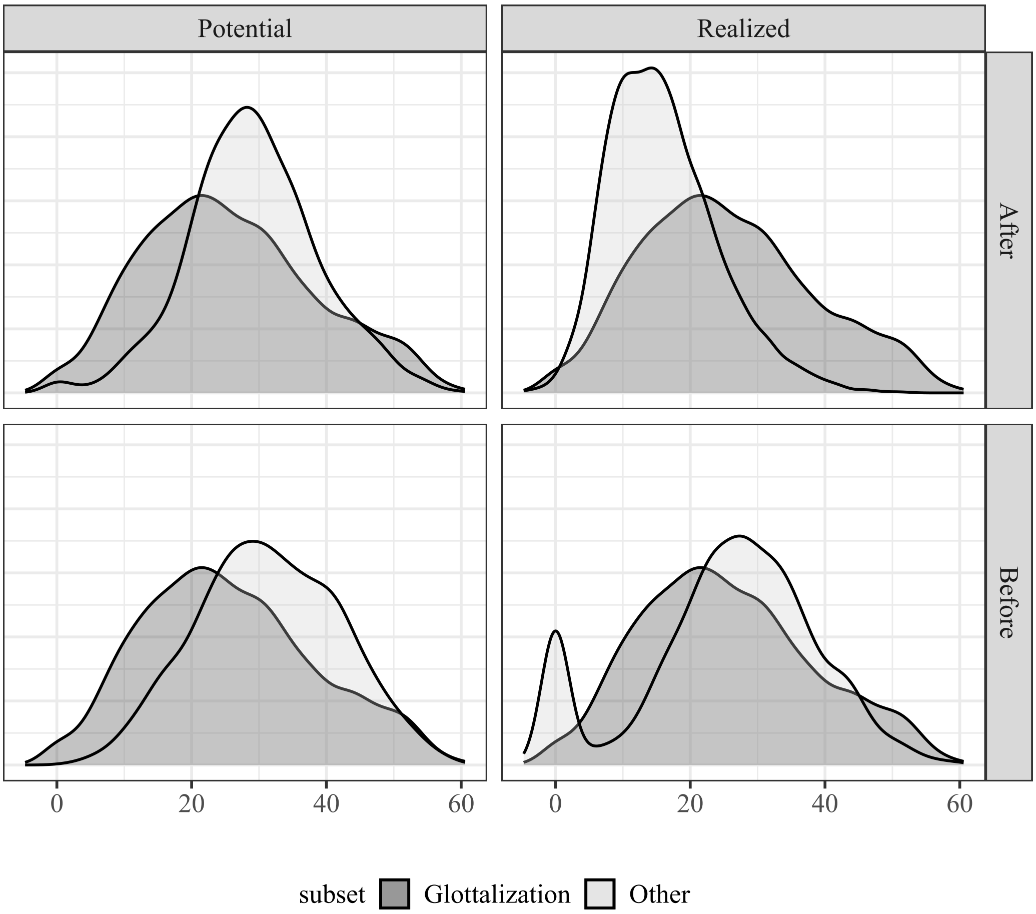

Figure 2.3: HNR15 per position and context

The second pair of parameters which were highly correlated were H1A1c and H1A2c. The differences between these two were also not substantial, however, again in the Realized Before quadrant, bimodality was more prominent. Once more, Being indicative either very low or non-existent values which could be useful in the distinction of glottalization against other segments.

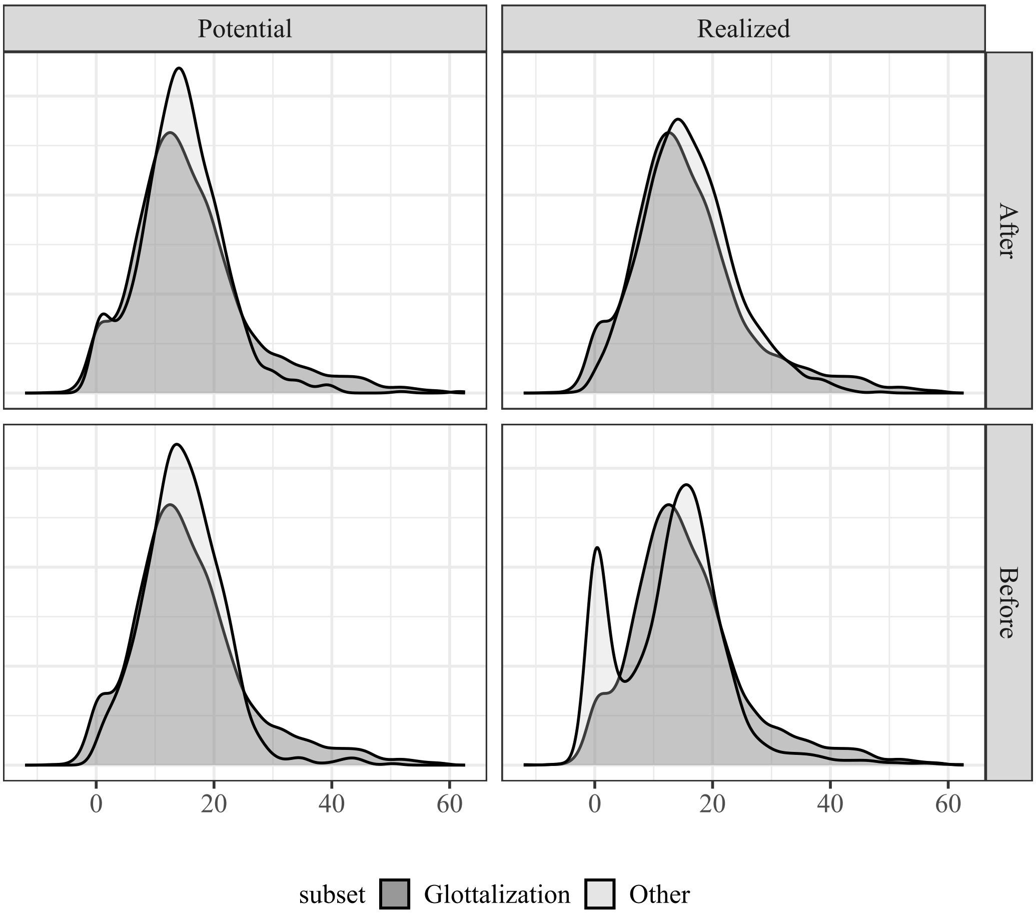

Figure 2.4: H1A1c per position and context

For the remainder of the parameters, only those that showed at least some level of ability to set apart glottalization and other segments, were kept (with the previous two serving as a benchmark).

Mean f0 showed an interesting property - being trimodal. This time the values larger than 300Hz seemed to be more indicative of glottalization. They are kept again for the sake of having a parameter that is known to differentiate between sexes.

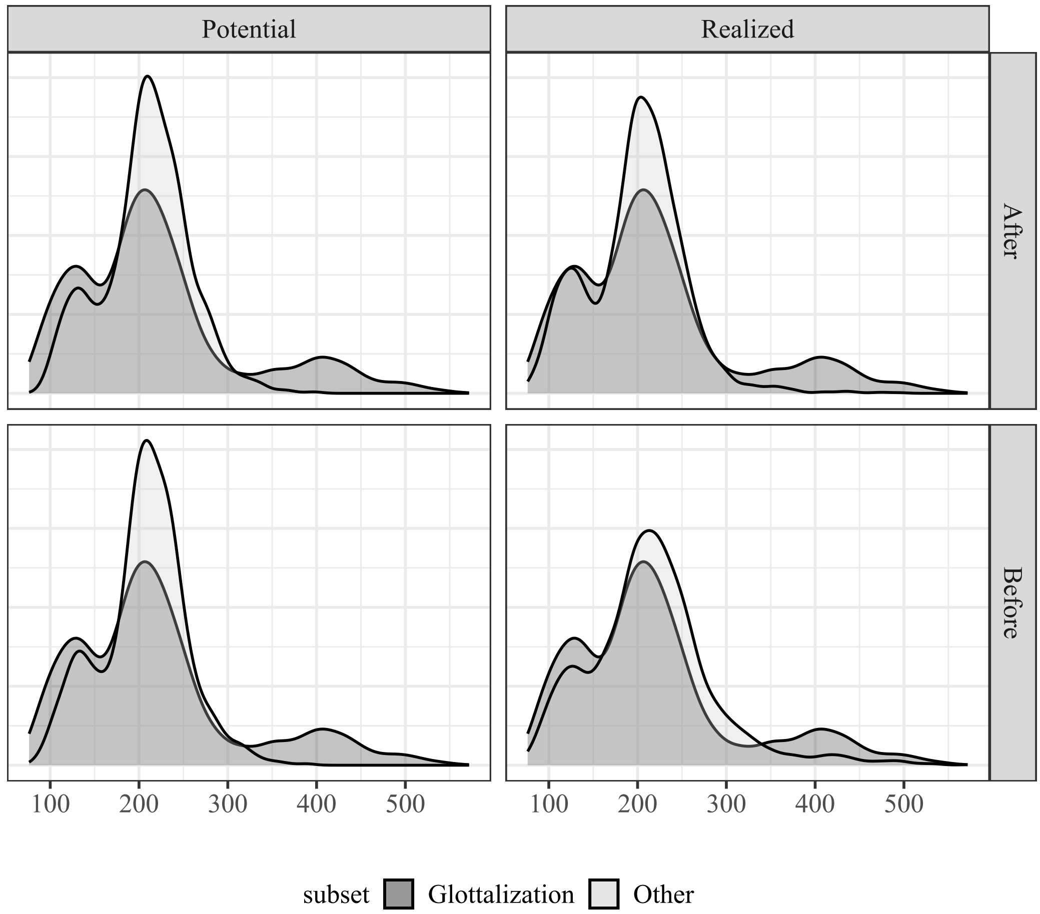

Figure 2.5: Mean f0 in Hz per position and context

While not being very convincing, the Potential Before quadrant showed some promise of being useful. For better resolution, the plot was limited on the x axis, showing only values from 0.0 to 0,20 %. This removed 113 measures.

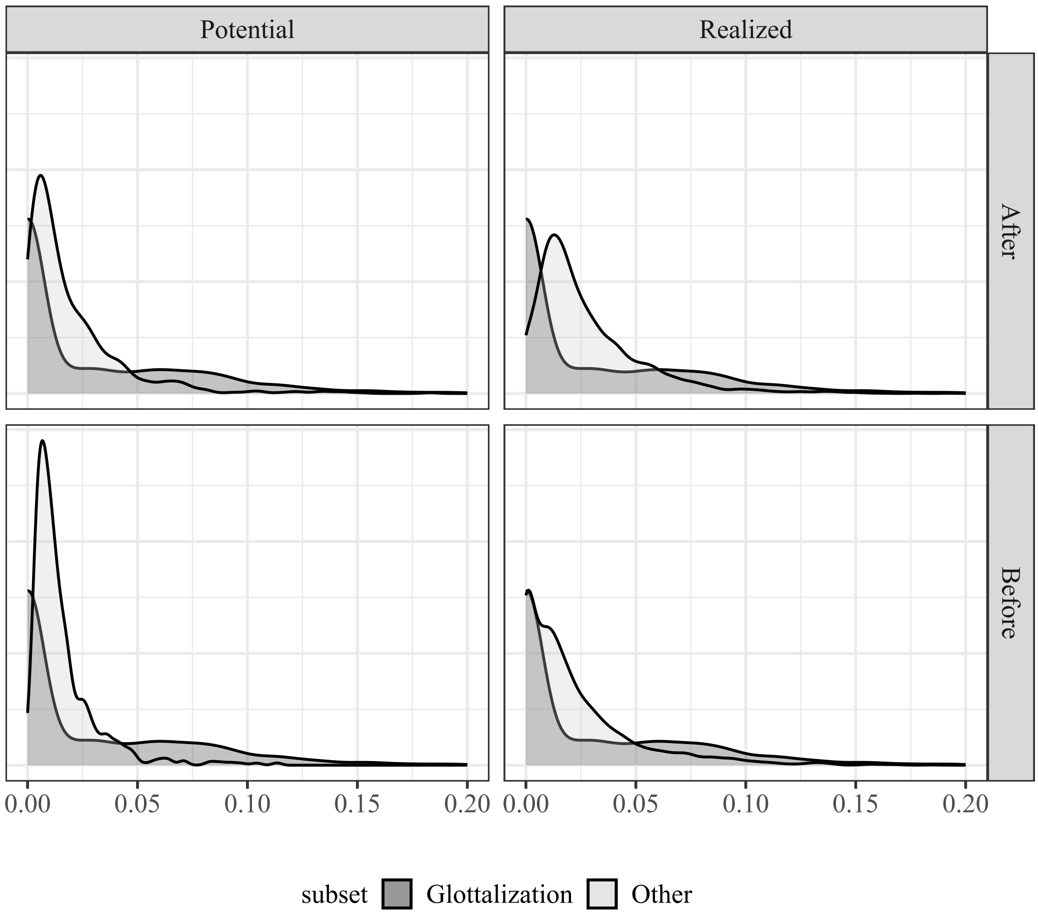

Figure 2.6: Jitter in % per position and context

The first of the MFFCs seemed more clustered towards higher values in general for other segments.

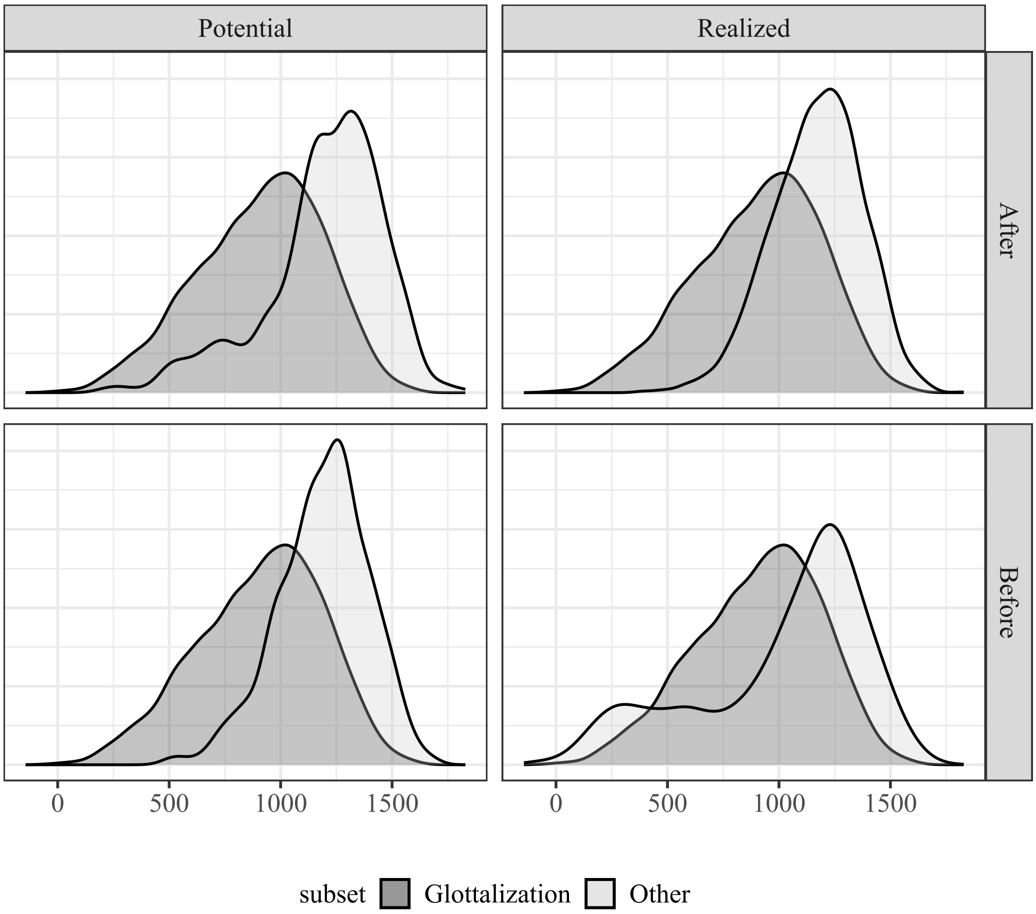

Figure 2.7: X1 per position and context

While overlapping considerably for the After Potential and Before Realized quadrant, for the others, glottalizations seemed to have a tendency towards smaller values.

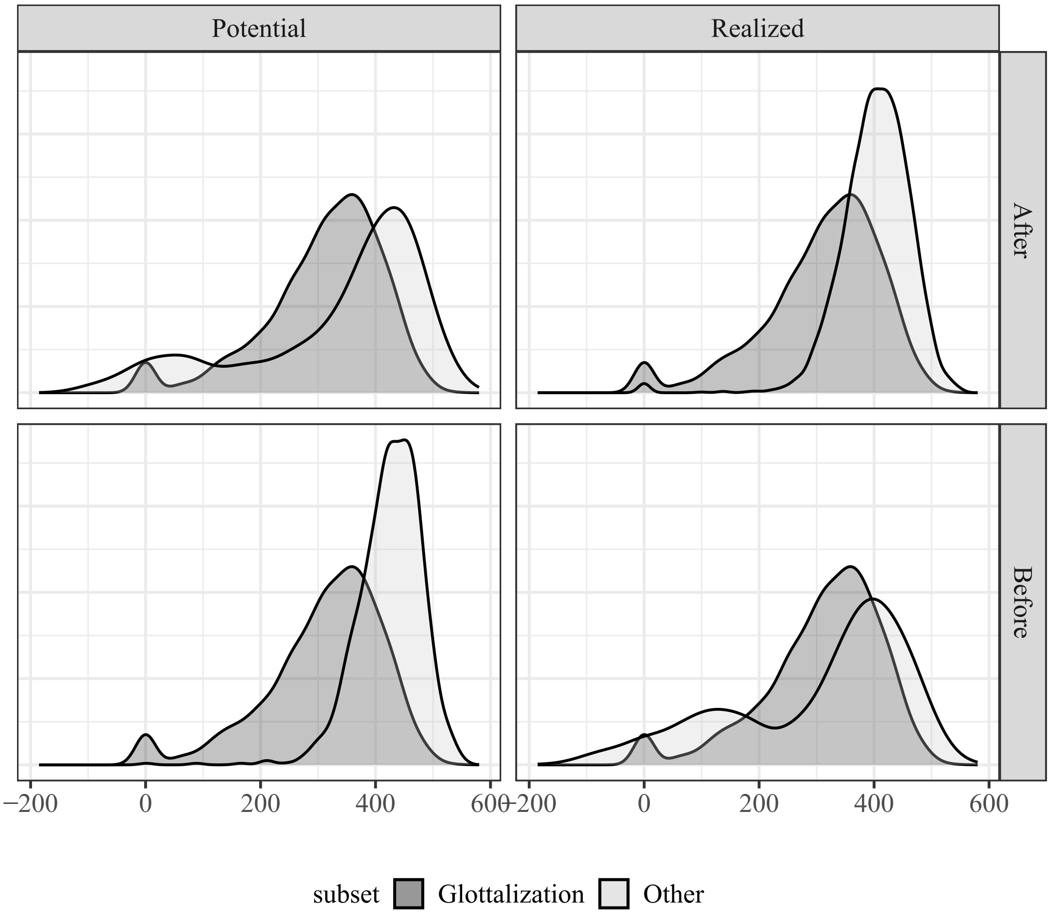

Figure 2.8: X2 per position and context

CPP seemes as the most potent parameter so far. Glottalization is very centered, while other segments are much more uneven to the left.

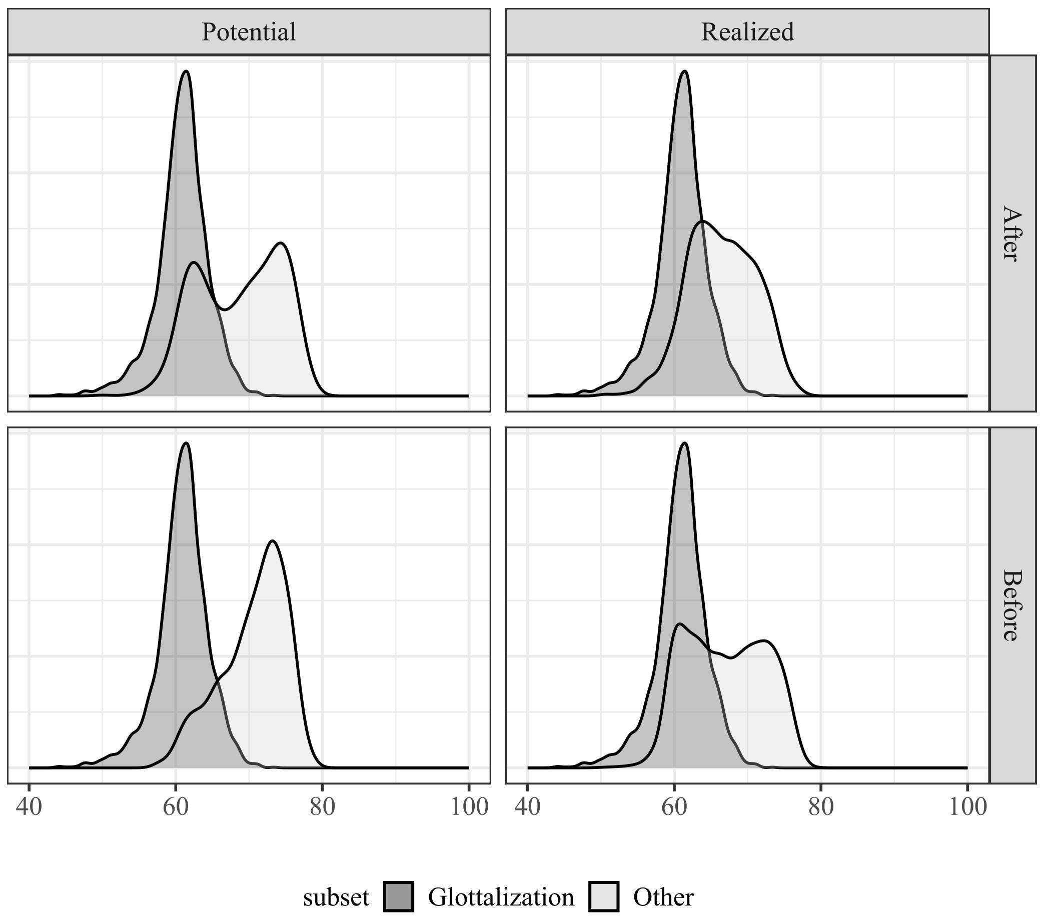

Figure 2.9: CPP per position and context

The remainder of the parameters was overlapping almost identically. The final count of parameters is 7: HNR15, H1A1c, mean f0, Jitter, X1, X2 and CPP.

2.2.2 Categorization

In order to quantify at least some degree of our findings, machine learning algorithm of random forest are trained and tested.

The data is split into training set of 0,80 and testing set of 0,20 of the total. The split is automatically done as random stratification of the target parameter. Cross-validation of 4-folds was used, alongside with Accuracy as the main metric. Random seed 3456 was used.

The number of parameters is 27 in total, including the target parameter.

The target parameter is type. There are 5 different types: before unrealized, after unrealized, before glottalization, glottalization and after glottalization. The same as in the graphs of the previous section.

Firstly, we use all of the parameters to train the model - even those we discarded using correlation. The accuracy was the mean over the 3 decision trees: 68,05 %.

Then, the same process is done only for those parameters which we have selected as interesting - this is 8 including our target parameter. The final mean accuracy is 61,98 %. Around 6,07% decrease of accuracy by removing 19 parameters.

For this reduced dataset, we will iterate for each parameter, and remove it from it. Train the model again and in the end see which missing parameter reduced the accuracy the most.

The final results are presented in 2.7. While the full dataset was more succesfull, a difference of 6,07 % as opposed almost the same be removing HNR15 parameter is striking. These parameters function together and so removing to much of them might cause the interconnections to stop existing, and thus not be able to explain variability in the data. The benchmark for this model could be to guess that the given segment is always the most common segment. The most common segment occurs 2 520,00 times. The total amount of segments is 9 409,00. The benchmark of such guessing would have been 26,78% accuracy.

Dataset	**Accuracy (%)**	**Difference (%)**
Full	68,05	
Reduced	61,98	6,07
Reduced-HNR15	55,63	6,17
Reduced-Jitter	58,67	3,12
Reduced-X2	59,92	1,89
Reduced-f0	60,58	1,23
Reduced-CPP	61,00	0,81
Reduced-H1A1c	61,07	0,74
Reduced-X1	61,15	0,66

Table 2.7: Comparison of various machine learning models with varying input data against the full model

Conclusion & Discussion

When it comes to terminology, it is important to be very specific about what exactly we are talking about - not only in Czech due to plethora of synonyms and theoretical approaches, but also in English, where the need to name specific processes such as glottal reinforcement and glottal replacement have to be addressed, too. Only after getting familiar with the terminological situation, the proper research could have been done.

In terms of typology, it seems that the most common approach is to study the distribution of glottalizations in various contexts and in different languages by labeled data (to study, for example, where glottalizations occur more frequently, or even in which styles they occur more often). Categorization and research on the acoustic side was found less frequently. This is due to the special context in which the phenomena is studied. Study of glottalization is an important topic in the study of pathology of voice, however, these studies usually search for signs of glottalizations within the entire speech wave, or for sustained vowels. Despite not being the most frequent topic, categorization of Czech glottalizations was done, which served as an important conceptual asset. The selection of acoustic parameters was primarily inspired by international studies that were concerned with English.

Additional research was done in order to understand the underlying acoustic topics. This was important since the utilized parameters were selected from works concerning other languages, where the glottalizations could manifest themselves differently. It was in connection with the understanding of acoustic parameters, and in combination with the previous research on categorization of Czech glottalization.

Working with a large amount of material required a statistical approach towards data quality - which had to be ensured, but could not be meticulously controlled. Normalization on segment and word level of TextGrids was done as well as on the boundary level of these two levels. Informed selection of problematic TextGrids was used in order to prevent extra manual labour. This was described as a linear process during the thesis in 2.1, however, in reality, it was a cycle of reiterations.

When material was transformed into data, exploratory analysis and familiarization with the data was in order - this was done through visual inspection. Based on these, the parameters were reduced into what is believed as a reasonable subset. Further subdivision per other categories - per segment type (C, V), per sex, per vowel quality, and other - was considered. Ultimately, the only categorical variable used was the target parameter of whether the segment is a glottalization or not.

The final model showed that it has the potential to be functional even with a handful of parameters. These parameters, without any categorical parameter, were able to successfully identify 60,00 % of the segments.

From the terminological point of view, it would be best to respect the recommendations for Czech language. However, in English texts, it might be beneficial to create a term which would be more specifically linked with the function of glottalization as a boundary signal. The terms such as glottal reinforcement or glottal replacement alludes possible terms such as glottal insertion or intrusive glottalization. It could be interesting to conceptualize glottalization even during labelling as marking a boundary instead of a segment - that way, other realizations could be marked. Having worked with hundreds of examples of glottalization, replacement for schwa or a reduction was another way how to signal a boundary. Nevertheless, enriching this dataset with more specific sub-types of glottalizations could be considered in order to gain further insight in how glottalizations appear, and even to further specify the acoustic parameters per each sub-type. Additional work can be done also for example in terms of prosodic boundaries or whether glottalization occurs more often before some specific segments (see 1.3.1).

Bibliography

Ashby, M. and Przedlacka, J. (2014). Measuring incompleteness: Acoustic correlates of glottal articulations. *Journal of the International Phonetic Association*, 44:pp. 283–296. [online]. [cit. 2023-01-01]. https://doi.org/10.1017/S002510031400019X.

Bissiri, M. P., Kraljevski, I., and Hoffmann, R. (2014). HMM-based classification of glottalization phenomena in german-accented english. pages 185–192. In: 25. Konferenz Elektronische Sprachsignalverarbeitung (ESSV) 2014. [online]. [cit. 2023-01-02]. https://www.researchgate.net/publication/267728788_HMM-based_classification_of_glottalization_phenomena_in_German-accented_English.

Bičan, A. (2006). *Accent and Diaereme and Their Position in Functional Phonology*. [online]. [cit. 2023-01-02]. Master thesis, Masaryk University, Faculty of Arts, Brno. Supervisor : Ondřej Šefčík. https://is.muni.cz/th/tcm2w/.

Bičan, A. (2008). Accent and Diaereme in Czech. *La linguistique*, 44(2):p. 45. [online]. [cit. 2022-12-02]. https://doi.org/10.3917/ling.442.0045.

Boersma, P. (2009). Should Jitter Be Measured by Peak Picking or by Waveform Matching?. *Folia Phoniatrica et Logopaedica*, 61(5):pp. 305–308. [online]. [cit. 2022-12-20]. https://doi.org/10.1159/000245159.

Bortlík, J. (2014). Pre-Vocalic Glottalization vs Resyllabification in Regional Varieties of Czech (A Pilot Study). In *Language Use and Linguistic Structure*, pages 293–302. Palacký University Press, Olomouc, 1 edition. https://dx.doi.org/10.5507/ff.14.24440606.21.

Bořil, T. (2019). Skripty a nástroje | Fonetický ústav. *Čištění TextGridů (rPraat)*. [online]. [cit. 2023-01-08]. https://fonetika.ff.cuni.cz/vyzkum/skripty-a-nastroje/.

Bořil, T. and Skarnitzl, R. (2016). Tools rPraat and mPraat. In Sojka, P., Horák, A., Kopeček, I., and Pala, K., editors, *Text, Speech, and Dialogue: 19th International Conference, TSD 2016, Brno, Czech Republic, September 12-16, 2016,* *Proceedings*, pages 367–374. Springer International Publishing, Cham. [online]. [cit. 2022-01-08]. https://dx.doi.org/10.1007/978-3-319-45510-5_42.

Chlumský, J. (1928). *Česká kvantita, melodie a přízvuk = Quantité, la mélodie et l'accent d'intensité en tchéque : avec un résumé en français.* Rozpravy České akademie věd a umění. Třída 3 ; č. 65. Česká akademie věd a umění, Praha. OCLC 19973552.

Chlumský, J. (1933). Poznámky k Weingartově Zvukové kultuře českého jazyka. *In: Naše řeč.* volume 17, pages 161–178. [online]. [cit. 2023-01-06]. https://kramerius.lib.cas.cz/uuid/uuid:366000b7-4c3e-11e1-1418-001143e3f55c.

Dilley, L., Shattuck-Hufnagel, S., and Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *In: Journal of Phonetics.* volume 24, pages 423–444. [online]. [cit. 2023-01-01]. https://doi.org/10.1006/jpho.1996.0023.

Fernandes, J., Teixeira, F., Guedes, V., Junior, A., and Teixeira, J. P. (2018). Harmonic to Noise Ratio Measurement - Selection of Window and Length. *In: Procedia Computer Science.* volume 138 of *CENTERIS 2018 - International Conference on ENTERprise Information Systems / ProjMAN 2018 - International Conference on Project MANagement / HCist 2018 - International Conference on Health and Social Care Information Systems and Technologies, CENTERIS/ProjMAN/HCist 2018*, pages 280–285. [online]. [cit. 2023-01-02]. https://doi.org/10.1016/j.procs.2018.10.040.

Frinta, A. (1909). *Novočeská výslovnost: pokus o soustavnou fonetiku jazyka českého.* Praha. Česká akademie císaře Františka Josefa pro vědy, slovesnost a umění.

Garellek, M. (2012). Glottal stops before word-initial vowels in American English: distribution and acoustic characteristics. *In: UCLA Working Papers in Phonetics*, 110:pp. 1–23. [online]. [cit. 2023-01-05]. https://escholarship.org/content/qt7m55b8bb/qt7m55b8bb_noSplash_ad63f821e520155ec5195eaaa435e10f.pdf.

Godino-Llorente, J., Gomez-Vilda, P., and Blanco-Velasco, M. (2006). Dimensionality Reduction of a Pathological Voice Quality Assessment System Based on Gaussian Mixture Models and Short-Term Cepstral Parameters. *IEEE Transactions on Biomedical Engineering*, 53(10):pp.1943–1953. Conference Name: IEEE Transactions on Biomedical Engineering. https://dx.doi.org/10.1109/TBME.2006.871883.

Gordon, M. and Ladefoged, P. (2001). Phonation types: a cross-linguistic overview. *In: Journal of Phonetics*, 29(4):pp.383–406. https://doi.org/10.1006/jpho.2001.0147.

Hanson, H. (1995). *Glottal characteristics of female speakers*. PhD thesis, Harvard University. The Division of Applied Sciences. Advisor: Kenneth N. Stevens.

Havránek, B. and Weingart, M. (1932). *Spisovná čeština a jazyková kultura*. Vol. 14. Melantrich, Praha.

Hillenbrand, J., Cleveland, R. A., and Erickson, R. L. (1994). Acoustic Correlates of Breathy Vocal Quality. *In: Journal of Speech, Language, and Hearing Research*, 37(4):pp.769–778. https://doi.org/10.1044/jshr.3704.769.

Hlaváč, S. and Pech, K. (1981). Slovo a slovesnost – lidský hlas jako pojem a fyzikální veličina (k některým otázkám produkce hlasu v souvislé řeči). *In: Slovo a slovesnost*, 42(2):134–142. [online]. [cit. 2023-01-05]. http://sas.ujc.cas.cz/archiv.php?art=2746.

Hála, B. (1967). Výslovnost spisovné češtiny.

Jadoul, Y., Thompson, B., and de Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *In: Journal of Phonetics*, 71:pp.1–15. https://doi.org/10.1016/j.wocn.2018.07.001.

Krom, G. d. (1993). A Cepstrum-Based Technique for Determining a Harmonics-to-Noise Ratio in Speech Signals. *In: Journal of Speech, Language, and Hearing Research*, 36(2):pp.254–266. Publisher: American Speech-Language-Hearing Association. https://doi.org/10.1044/jshr.3602.254.

Kuhn, M. (2019). [online]. [cit. 2023-01-09]. http://topepo.github.io/caret/index.html.

Lehiste, I. (1965). Juncture. In *Phonetic Sciences*, pages 172 – 200, New York: University of Münster. Karger Publishers, Zwirner Eberhard and Bethge Wolfgang, ed. [online]. [cit. 2023-01-04]. http://tscheer.free.fr/interface/Lehiste%201965%20-%20Juncture.pdf.

Machač, P. and Skarnitzl, R. (2009). *Fonetická segmentace hlásek*. Epocha, Praha. ISBN 978-80-7425-031-6.

Palková, Z. (1994). *Fonetika a fonologie češtiny : s obecným úvodem do problematiky oboru*. Karolinum, Praha, 1. edition. ISBN 80-7066-843-1.

Palková, Z., Veroňková, J., Volín, J., and Skarnitzl, R. (2004). Stabilizace některých termínů pro fonetický popis češtiny v závislosti na nových výsledcích výzkumu. *In: Sborník z Konference česko-slovenské pobočky ISPhS*. pages 65–74, Praha: Fonetický ústav FF UK. T. Duběda. [online]. [cit. 2023-01-03]. https://sites.ff.cuni.cz/fonetika/wp-content/uploads/sites/104/2015/10/Pal04-terminy.pdf.

Pollak, P., Volín, J., and Skarnitzl, R. (2008). Phone segmentation tool with integrated pronunciation lexicon and Czech phonetically labelled reference database. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA). [online]. [cit. 2023-01-06]. http://www.lrec-conf.org/proceedings/lrec2008/pdf/421_paper.pdf.

Pompino-Marschall, B. and Żygis, M. (2010). Glottal marking of vowel-initial words in German. *In: ZAS Papers in Linguistics*, 52:pp.1–17. [online]. [cit. 2023-01-02]. https://doi.org/10.21248/zaspil.52.2010.380.

Redi, L. and Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *In: Journal of Phonetics*, 29(4):pp.407–429. https://doi.org/10.1006/jpho.2001.0145.

Romportl, S. (1984). Slovo a slovesnost – Úloha rázu a jeho ekvivalentů při signalizování předělu. *In: Slovo a slovesnost*, 45(2):pp.104–120. [online]. [cit. 2023-01-07]. http://sas.ujc.cas.cz/archiv.php?art=2953.

Skarnitzl, R. (2004). Acoustic categories of nonmodal phonation in the context of the Czech conjunction "a". *In: Phonetica Pragensia*, 2004(10):pp.57–68.

Skarnitzl, R. (2009). Prague Phonetic Corpus: status report. *In: Acta Universitatis Carolinae - Philologica*, XII(1):65–67.

Skarnitzl, R. (2014). *Fonetická identifikace mluvčího*. Univerzita Karlova v Praze, Filozofická fakulta, 1. edition. ISBN 978-80-7308-548-3.

Skarnitzl, R., Šturm, P., and Volín, J. (2016). *Zvuková báze řečové komunikace: fonetický a fonologický popis řeči*. Univerzita Karlova v Praze, Nakladatelství Karolinum, Praha, 1. edition. ISBN 978-80-246-3272-8.

Skákal, L. (2015). Užívání glotalizace jako faktor umožňující identifikaci mluvčího. Master thesis. Charles University, Faculty of Phonetics, Institute of Phonetics. Supervisor: Radek Skarnitzl. [online]. [cit. 2023-01-01]. https://dspace.cuni.cz/bitstream/handle/20.500.11956/73628/DPTX_2013_2_11210_0_407997_0_150673.pdf?sequence=1&isAllowed=y.

Taiyun, W., Simko, V., Levy, M., Xie, Jin, Y., Zemla, J., Freidank, M., Cai, J., and Protivinsky, T. (2021). [online]. [cit. 2023-01-09]. https://CRAN.R-project.org/package=corrplot.

Titze, I., Horii, Y., and Scherer, R. (1987). Some Technical Considerations in Voice Perturbation Measurements. *In: Journal of speech and hearing research*, 30:pp.252–60. https://doi.org/10.1044/jshr.3002.252.

Titze, I. and Liang, H. (1993). Comparison of Fo Extraction Methods for High-Precision Voice Perturbation Measurements. *In: Journal of speech and hearing research*, 36:pp.1120–33. https://doi.org/10.1044/jshr.3606.1120.

Volín, J. (2012). Jak se v Čechách "rázuje". *In: Naše řeč*, 95(1):pp.51–54. [online]. [cit. 2023-01-05]. http://nase-rec.ujc.cas.cz/archiv.php?lang=en&art=8235.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019). Welcome to the Tidyverse. *In: Journal of Open Source Software*, 4(43):1686. [online]. [cit. 2023-01-05]. https://dx.doi.org/10.21105/joss.01686.

WikimeadiaCommons (2005). *File:Glottis positions.png*. [online]. [cit. 2023-01-08]. https://commons.wikimedia.org/wiki/File:Glottis_positions.png.

Yumoto, E., Gould, W. J., and Baer, T. (1982). Harmonics-to-noise ratio as an index of the degree of hoarseness. *In: The Journal of the Acoustical Society of America*, 71(6):pp.1544–1550. https://doi.org/10.1121/1.387808.

Šedivá, M. (2022). Kognitivní dopad použití rázu ve spontánní politické debatě. Praha, 2022. Master thesis. Charles University. Faculty of Arts. Institute of Phonetics. Supervisor: Pavel Šturm. [online]. [cit. 2023-01-01]. https://dspace.cuni.cz/bitstream/handle/20.500.11956/174541/120421685.pdf?sequence=1&isAllowed=y.

Štěpánová, V. (2019). *Vývoj kodifikace české výslovnosti*. Lingvistika. Academia, Praha, 1. edition. ISBN 978-80-200-2900-3.

List of Figures

List of Tables