Stability of Phonetic Features of Czech Plosives in Spontaneous Speech

Pavel Machač — Radek Skarnitzl

ABSTRACT:
This study examines speech reductions in the spontaneous Czech of six young adult speakers. Specifically, intervocalic plosives are analysed from the perspective of phonetic features, with the aim to discover these features' relative stability. Auditory analysis was used to determine the realisation types of plosives, and these types were then verified by acoustic analyses of duration, intensity range, harmonicity, and voicing profile. The results show that phonologically voiced plosives undergo reduction processes more (40%), with semi-vocalised realisation being the most frequent, while voiceless plosives are reduced less often (20%), with fricative-like realisation being the most frequent reduction. The least stable phonetic feature of Czech plosives is thus closure, as confirmed by all the analysed acoustic parameters.

ABSTRAKT:
Studie se zabývá řečovými redukcemi ve spontánních projevech šesti mladých mluvčích češtiny. Konkrétně jsme analyzovali realizace intervokalických exploziv z hlediska fonetických rysů s cílem popsat relativní stabilitu těchto rysů. Na základě auditivní analýzy jsme určili typy realizací a takto vzniklou kategorizaci jsme ověřovali prostřednictvím akustické analýzy trvání, rozdílu nejvyšší a nejnižší intenzity uvnitř explozivy, harmonicity a znělostního profilu. Výsledky ukazují, že fonologicky znělé explozivy podléhají redukcím častěji (40 %) než explozivy neznělé (20 %). U znělých exploziv je nejfrekventovanější redukcí polovokalická realizace, u neznělých realizace frikativizovaná. Všechny akustické parametry ukazují, že nejméně stabilním fonetickým rysem českých exploziv je jejich závěrovost.

KEY WORDS:
speech production, speech reductions, phonetic features, plosives, Czech

KLÍČOVÁ SLOVA:
produkce řeči, řečové redukce, fonetické rysy, explozivy, čeština

1 INTRODUCTION

One of the most significant sound characteristics of spoken speech is the rich variability of word and sentential forms in the acoustic signal (Blache & Meunier, 2004). A number of reduced forms belong here which differ from the full pronunciation in a greater or smaller degree (e.g. the Czech word *úplně* "completely" [ʔuːplɲɛ] > [uplɲɛ] > [upəɲɛ] > [upəɛ̃] > [upɛ̃ː] > [pɛ̃ː] > [pɛ̃]). To understand the principles of reductions at the word level, it is necessary to know the principles at the level of individual segments and their classes; that is the topic of the present study. Even though accumulating speech reductions may be typical of spontaneous speech in particular, it is also the phonetic and sociolinguistic studies that show that speech reductions are to be found in every speech style, across all social groups, with each speaker (e.g. Greenberg, 1999; Keune, Ernestus, Van Hout & Baayen, 2005; Schuppler, Ernestus, Scharenborg & Boves, 2011; Cangemi & Niebuhr, 2018).

It is the spontaneous speech and speech reduction research that is essential for discovering and modelling the principles of speech communication, the relationship between speech production and perception, mental storage of word forms (e.g. Lahiri & Marslen-Wilson, 1991; Mitterer & Blomert, 2003; Ranbom & Connine, 2007), or sound change (e.g. Ohala, 1981, 1989; Hamann, 2009).

Even though there may be only a slight difference between the reduced and the full (canonical) forms, it is not unusual for the degree of reduction to be so high that it prevents native speakers from adequately identifying individual words isolated from the naturally produced and commonly comprehensible sentences (e.g. Johnson, 2004; Ernestus & Warner, 2011; Niebuhr & Kohler, 2011), e.g. *pan kolega* "colleague" [pan kolɛga] > [β̞ãɰla]; *samozřejmě* "of course" [samozr̝ɛjmɲɛ] > [sãˑr̝ɛ̃ˑ]. Research into the dynamics of sentence reductions in the Czech language (in progress) presents an example that features an easily comprehensible compound sentence consisting of 19 words. More than 90% of listeners identified only 5 isolated words correctly, over a half of the listeners recognised only 9 words and there were 3 words that none of the listeners were able to identify. Within an experiment, identification of individual words in a continuous sentence is made possible by means of *linguistic context*. To be able to understand strongly reduced sentence forms, it is also the *extralinguistic context* which is necessary for the listener's understanding, e.g. the reduced sentential form *už to jede* "it's already going" [ʃtɛˑrɛ], canonically [ˈʔuʃtoˈjɛdɛ], is impossible to decipher without the situational context (waiting at a bus stop).

During his or her speech, the speaker, usually unknowingly and automatically, chooses from a number of learned potential articulation possibilities (at both the segmental and suprasegmental level in the form of phonetic detail) and determines the frequency and the degree of their use in time and space (see e.g. Browman & Goldstein, 1992; Fowler & Saltzman, 1993). In the case that the speaker does not pursue particular pragmatic or aesthetic aims, he or she naturally inclines to exerting only the speech effort necessary for adequate understanding under the given circumstances (Lindblom, 1990), i.e. understanding 'on the first attempt'. The resulting acoustic form of words and sentences and the degree of reduction in a particular speech act depends on the "coordination" between the communicating partners: the speaker offers a certain phonetic form while being directed in real time by the listener, who guides the speaker in the case of incomprehension or misunderstanding and thus determines the maximum acceptable degree of reduction in the particular situation (what we refer to as *minimal phonetic information*; see the Discussion for more information). As the communicating participants strive to find a balance between "speaker laziness" and the need to be immediately understood, as well as the need to understand one's communicating partner easily on the basis of acquired experience of speech (see the stochastic models of speech perception, such as Pierrehumbert, 2003), they develop a dynamic system of speech reductions whose functionality is verified in everyday conversations and is further modified by the practice of speech. Its fundamental principles are then shared by the respective linguistic community. A system of reductions comprises massive reductions at the level of words and phrases (see the examples above), as well as reductions at the level of segments (specifically plosives for this study). It is logical that the latter will participate in reductions at higher levels in fairly predictable ways.

2 PHONETIC FEATURES AND THEIR STABILITY

In order to better understand the principles of speech reduction at the level of semantic units, it is useful to consider the inherent tendencies towards reductions in different types of segments (for more information, see the Discussion). The aim of this study is to document the types of realisations of one class of consonants (plosives) in the Czech language and, at the same time, to determine how frequently these types appear in spontaneous speech and to determine their selected acoustic properties. Based on the quantification of canonical forms and the types of reduced forms, we deduce the relative inherent stability of phonetic features of a phone or a group of phones. According to Bybee (2003, p. 3) "all category members need not have all the features characterising the category, but a member is more central or more marginal depending on the number and nature of shared features". The reduced realisations of plosives particularly concern the absence or weakening of some of the inherent phonetic features (henceforth PFs; see below) and show characteristics of other groups of phones (e.g. semivowels, nasals).

For practical reasons, the description of the reductions is based on the "ideal phone" which we approach as a *complex of potential phonetic features* (i.e. a system of articulatory gestures, acoustic tracks and auditory perceptions), typical for a full, canonical pronunciation of a phone. Following Machač and Skarnitzl (2009) we call them the *inherent phonetic features*. The "ideal phone" is a theoretical construct used as a static listing of the inherent PFs of a phone for their comparison with the "real phone", i.e. the actually pronounced one. The *potentiality* of phonetic features is based on the fact that there is no need to implement all of the inherent PFs in a specific realisation of a phone. Quite the opposite — mostly under the influence of the phonetic environment — it will come to PFs not belonging to the inherent PFs inventory (see e.g. articulatory prosody, Kohler, 1999), which can have a significant impact on the final form of the phone (Machač & Zíková, 2015). PFs borrowed from their (not necessarily immediate) neighbours are called *extrinsic*; an example of an extrinsic acoustic PF in plosives can be *continuous noise* or *full formant structure*. Not using all of the inherent PFs (e.g., closure in plosives) and using the extrinsic PFs (e.g., full formant structure in plosives) usually, though not always (compare with the devoicing of Czech intervocalic fricatives, Machač, 2008), results in a weakened acoustic contrast between the neighbouring phones. Although the identification of the phone itself becomes more difficult, and sometimes even impossible, the adequate perception of semantic units in continuous speech does not tend to be affected.

It is also the historical development of the Czech language which documents the loss of some of the inherent PFs and the establishment of the extrinsic phonetic features in various types of phones. In the area of plosives, it was, for example, palatalisation leading to the establishment of palatal plosives /c ɟ/ or fricativisation /g/ > /ɣ/ (in contemporary Czech > /ɦ/, in common speech the "non-fricative" /ɦ/ or 0, e.g. *toho* "that" [togo] > [toɣo] > [toɦo] > [too] > [toˑ]).

Various PFs of phones can vary significantly in their stability (for examples, see Table 1), whereby the more stable and thus more frequent ones can be regarded as more important for appropriate perception than those that are often absent or weakened. The stability of features particularly designates the ability to resist the influence of the neighbouring segments (cf. Bladon & Al-Bamerni's, 1976, concept of *coarticulatory resistance*). At the same time, the most stable features tend to affect the neighbouring segments the most and the degree of stability thus positively correlates with the so-called *coarticulatory aggression* (Fowler & Saltzman, 1993).

speech sounds	**more stable features**	**less stable features**
vowels	open vocal tract, presence of f_0, formant structure	quantity, quality, oral character
nasals	nasality	occlusion, place of articulation
voiced sibilants	place and manner of articulation	presence of f_0
voiceless sibilants	place and manner of articulation	absence of f_0

TABLE 1: Examples of more stable and less stable PFs in Czech (based on Machač & Skarnitzl, 2009; Machač & Zíková, 2015).

The aim of this study is to determine the relative stability of PFs of Czech plosives in the intervocalic position on the basis of auditory categorisation of their realisations, verified by acoustic analyses. The type and frequency of deviations from canonical production of plosives is related to two factors which may play a role in the realisations: lexical stress and word frequency.

The Czech language has eight plosives: bilabial /p b/, alveolar /t d/, palatal /c ɟ/ and velar /k g/. Phonological voicing in Czech is implemented by means of phonetic voicing, i.e. the presence or absence of vocal fold vibration throughout the phone.

The inherent PFs of plosives, considered from three interconnected phonetic perspectives (production, acoustics and perception), are

a) articulatory features: closure, stop release, place of articulation, presence or absence of vocal fold vibration, raised soft palate;
b) acoustic features: lowered intensity, noise phase, presence or absence of fundamental frequency (f_0), absence of full formant structure;
c) perceptual features: lowered loudness, release burst, voicing or voicelessness, orality, place of articulation (bilabiality, alveolarity…).

This division of PFs corresponds to the ternary phonetic perspective of the same sound substance and to the objectives of this study. Here, we first use *perceptual* analysis to examine the stability of *articulatory* features (instrumental methods of articulatory phonetics are not suitable for studying spontaneous speech), and second, we identify *acoustic* characteristics of different ways of plosive articulation. For instance, *voicing/voicelessness* may be analysed perceptually; based on that, we may infer articulatory characteristics (*presence/absence of vocal fold vibration*) which, in turn, are manifested acoustically as *presence/absence of fundamental frequency*. When assessing the stability of PFs, we use traditionally used and unambiguous categorical terms like *voicing/voicelessness* and *closure* etc., rather than the more clumsy *presence/absence of vocal fold vibration* and *lowered loudness*, respectively.

3 METHOD

In this study, we first conducted auditory analysis of the recorded data to determine the realisation types of plosives. The realisation type corresponds to the group of plosives characterised by an identical manner of realization of PFs: canonical, semi-vocalised, fricativised, nasalised (see section 4.1 for more details). Subsequently, we studied their selected acoustic characteristics (duration, intensity range, harmonicity, voicing profile) and verified the relevance of the perceptually identified categories. Then, we investigated the dependence of the occurrence of the types of realisation of plosives on voicing, place of articulation (as the essential inherent PFs), word stress and word frequency (as external factors). Based on all these findings, we stated the relative stability of PFs for plosives in Czech.

3.1 MATERIAL

Six highly spontaneous dialogues were used for the analyses. Each dialogue pair consisted of a Charles University student who invited a good friend of theirs for an interview. It was the speech of the friends (i.e., $n = 6$) that was analysed. The speakers (four females, two males) were in the age range of 19 to 25. Each interview lasted approx. 40 minutes and was recorded in the sound-treated studio of the Institute of Phonetics in Prague.

In the analyses, we focused on intervocalic plosives, initially or medially with respect to word boundaries; it is precisely the intervocalic position where the highest level of using inherent PFs may be anticipated. Table 2 shows the number of the individual plosives based on lexical stress of the given syllable. The smaller number of tokens of palatal plosives is due to their overall relatively low frequency in Czech; in the case of /g/, the low number is due to their occurrence (in the intervocalic position) only within loan words in Czech. In total, 2,187 plosives were analysed in this study.

3.2 AUDITORY ANALYSIS

The aim of the auditory analysis was to determine the realisation types of the target plosives. Both authors evaluated the plosives by means of careful listening (with the use of waveform and spectrogram displays; see Figure 1 and the relevant comments in section 4.1 for more detail). Rare discrepancies in the two authors' evaluations were

	stressed	unstressed	total
p	179	178	357
b	134	183	317
t	184	184	368
d	183	184	367
c	21	147	168
ɟ	52	198	250
k	103	187	290
g	23	47	70

TABLE 2: Number of intervocalic plosives at the beginning of stressed and unstressed syllables.

inherent PFs	extrinsic PFs	
closure	semi-vocality	← opening of the vocal tract — larger
	fricativity	← opening of the vocal tract — smaller
	flap (only for /d/)	← extreme shortening of closure
voicing / voicelessness	voicelessness / voicing	
orality	nasality	

TABLE 3: Phonetic features monitored during the auditory analysis.

addressed by joint analysis. In the case of reduced forms, the absence of inherent and presence of extrinsic PFs was recorded; in the case of gradual reduction, the degree of their presence (the degree of opening of the vocal tract, expressed in phonetically meaningful categories) was recorded. The monitored PFs are listed in Table 3.

Neither the place of articulation nor its potential change are listed in Table 3, because the corresponding inherent auditory PFs — *bilabiality*, *alveolarity*, *palatality* and *velarity* — were preserved even in the reduced realisations. The *place of articulation* thus, in line with previous informal observations, proved to be a *highly stable* PF of intervocalic plosives (contrary to nasals, for instance).

3.3 ACOUSTIC ANALYSIS

The aim of the acoustic analysis was to investigate the relevance of the categories established on the basis of the auditory evaluation. Firstly, the segment boundaries were automatically aligned using HMM (Pollák, Volín & Skarnitzl, 2007) and then adjusted manually according to the rules stated in Machač & Skarnitzl (2009). The acoustic analyses were performed in Praat (Boersma & Weenink, 2011), focusing on the following acoustic parameters (see section 5 for more details):

a) plosive duration;
b) difference between the highest and the lowest intensity within the plosive (the difference would be large in a canonical plosive, especially a voiceless one);

c) mean harmonicity (HNR; Boersma, 1993) within the plosive (a canonical plosive would manifest low HNR values, corresponding to the prevalence of noise-like, rather than tone-like components in the spectrum);
d) voicing profile (Möbius, 2004), which enables the study of dynamic characteristics of voicing throughout a phone.

Specifics, as well as the purpose of measuring parameters b) to d), are provided in sections 5.2 to 5.4, along with the results and their interpretation.

Fundamental frequency (f_0), harmonicity and intensity were extracted in Praat using the default settings. For the voicing profile, f_0 was extracted in five equidistant steps throughout each plosive.

4 RESULTS I – AUDITORY EVALUATION

4.1 VARIABILITY IN THE MANNER OF ARTICULATION

The most important types of non-canonical plosive realisations are, from the perspective of the manner of articulation, those which were semi-vocalised and fricativised: the articulation organs do not reach the target position, a smaller or larger gap remains between them and the exhaled airflow is not interrupted. As a result, the least stable inherent PF is *closure*, and the most common extrinsic PF is *opening of the vocal tract*. Examples of waveform and spectrographic displays are shown in Figure 1.

A smaller opening of the vocal tract results in a continuous friction noise and a fricativised plosive (Fig. 1a, b). A larger opening of the vocal tract and the presence of tone (formant structure) in phonetically voiced plosives is typical of semi-vocalised realisations (Fig. 1c, d). It is clear that the gradient nature of this type of gestural reduction may prevent reliable identification of the stated categories.

Another extrinsic PF recorded in intervocalic plosives is *nasality*. In spite of a closure formed in the oral cavity, the soft palate lowers, for instance due to decreased speech effort and/or for reasons of distant coarticulation (e.g., [nɛbudɛmɛ] > [nɛmuɾ̃ɛmɛ] *nebudeme* "we will not"). With air flowing through the nasal cavity, the given phone loses the inherent PF *orality* and gains the extrinsic PF *nasality*. The voiced plosives thus transform into nasals with the same place of articulation: [b] > [m], [d] > [n], [ɟ] > [ɲ], [ɡ] > [ŋ]. For nasalised voiceless plosives, it is possible to record only a very faint friction noise originating during the passage of air through the nasal cavity: [p] > [m̥], [t] > [n̥], [c] > [ɲ̊], [k] > [ŋ̊]. Figure 1d shows an example of a voiced nasalised "plosive"; its nasal character is clearly audible and is also manifested through the nasal formant around 2 kHz.

The alveolar [d] is a plosive with the shortest duration in Czech, with the average duration of the closure phase being 47 ms (Machač, 2006). Its further shortening results in a flapped realisation, that is, a quick, ballistic tongue movement without a stable closure phase.

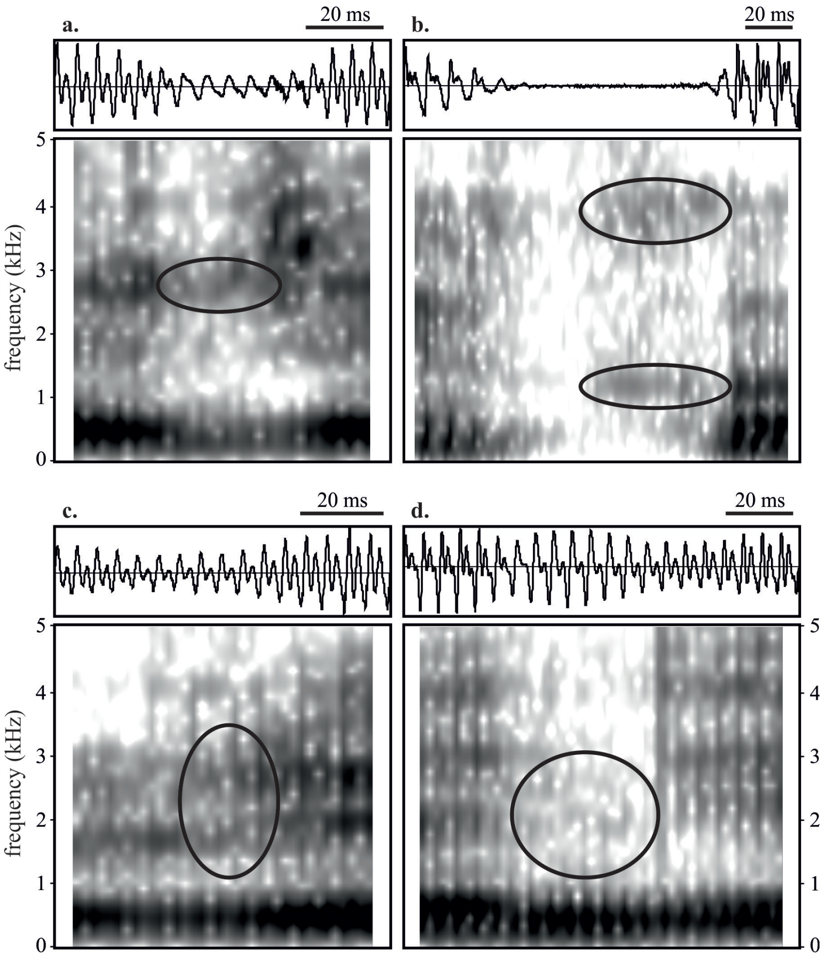

FIGURE 1: Examples of changes in the manner of articulation: **a.** fricativised /d/; **b.** fricativised /k/; **c.** semi-vocalised /d/; **d.** semi-vocalised and nasalised /ɟ/. The target extrinsic phonetic features are indicated by the ellipses.

4.1.1 FREQUENCY OF REALISATIONS TYPES BASED ON VOICING

The aerodynamic differences of voiced and voiceless plosives result in a diverse frequency of the types of realisations. Figure 2 shows some interesting aspects.

Voiced plosives are reduced significantly more often (38.1%) than the voiceless plosives (18.7%), as confirmed by the Yates-corrected chi-squared test: χ^2 (1; n = 2174) = 100.54; $p < 0.0001$. This result corresponds to the more difficult production of phonetically voiced plosives: during their production, there is a conflict created between the need to keep a sufficient transglottal pressure differential and, at the same time,

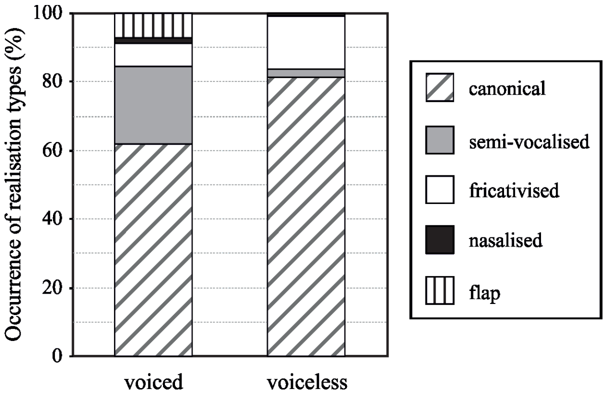

FIGURE 2: Frequency of reduction types based on voicing.

the need for a complete closure of the oral cavity. This conflict may be mitigated or eliminated applying various strategies. First of all, speakers may increase the volume of the vocal tract to accommodate voicing for a longer time (Westbury, 1983; Svirsky et al., 1997). In terms of reduction strategies, it seems that Czech speakers prefer the elimination of the *closure* as inherent PF — as a result, the difference in subglottal and supraglottal pressure can be maintained more easily, and the vocal folds vibrate throughout the plosive. This is important especially in languages such as Czech in which phonological voicing correlates mainly with phonetic voicing, that is with the presence/absence of f_0 for the whole duration of the phone. On the other hand, tenseness appears to be merely a secondary factor for perception, with the exception of whisper (Skarnitzl, Šturm & Machač, 2013).

Modification of the manner of articulation has been shown to be the prevalent type of reduction of plosives. Almost 40% of all voiced plosives lack the inherent PF *closure*. The most common type of reduction is semi-vocalisation (59.1% of all non-canonical realisations), which is well ahead of fricativisation (17.4%) and of realisation of /d/ as a flap (18.7%); nasalisation occurs marginally (4.7%).

In comparison, in the voiceless plosives, the *closure* as inherent PF disappears "only" in less than one fifth of all cases, and the most common type of reduction is fricativisation (82.3%). In the case that the extrinsic feature *opening of the vocal tract (large)* is joined by the extrinsic PF *tone-like character*, the result is semi-vocalisation (13.6%). There was also a marginal occurrence of "voiceless" nasalisation (4.1%).

4.1.2 FREQUENCY OF REALISATION TYPES BASED ON PLACE OF ARTICULATION

As seen in Figure 3, the results show a clear tendency towards a higher probability of occurrence of reductions at more posterior places of articulation and, at the same time, in plosives which require a larger mass of the active articulator; this does not seem to hold strictly only for /g/. The explanation for the overall tendency is obvious:

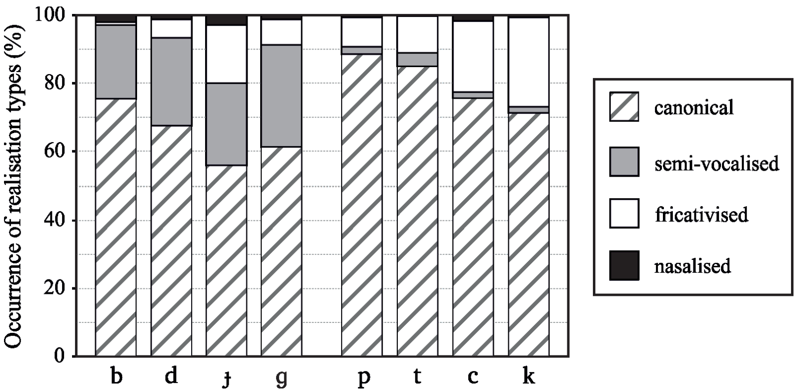

FIGURE 3: Frequency of reduction types based on place of articulation.

the smaller the space between the vocal folds and the place of articulation, the starker is the aerodynamic conflict mentioned above.[1]

4.1.3 ROLE OF LEXICAL STRESS

Lexical stress is not realised by means of clear acoustic prominence in Czech (Skarnitzl, 2018). Nevertheless, our material quite clearly implies that the inherent PFs of the voiced and voiceless plosives before stressed vowels are significantly more stable than before unstressed vowels ($p < 0.001$ for voiced plosives and $p < 0.005$ for voiceless plosives, according to the Yates-corrected chi-squared test). Naturally, it is unclear whether this reflects a role of word stress or whether it is the question of stability relating to the first syllable of the word; due to the fixed word stress on the first syllable of the prosodic word in the Czech language, these influences cannot be separated in our material.

4.1.4 ROLE OF WORD FREQUENCY

A frequency dictionary of Czech (Čermák et al., 2004) provided us with the Average Reduced Frequency (ARF) of all examined words, which were then separated into four groups. The results (Figure 5) confirm the assumption that plosives occurring in more frequent words (ARF > 100,000) are reduced more often. This supports the finding that word frequency belongs to the factors influencing the degree of reduction of words (e.g. Pluymaekers, Ernestus & Baayen, 2005).

1 Given its articulatory and aerodynamic specificities, the flap is counted as a canonical realization here. The main reason is that, in this realization of /d/, there is (unlike in the other plosives) a complete (albeit very short) closure. A secondary reason is the fact that this type of realization is only relevant for /d/, and cannot be compared across the plosive class.

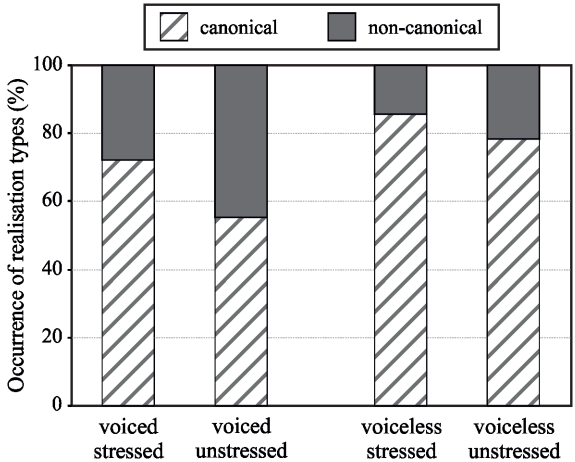

FIGURE 4: Frequency of canonical and non-canonical plosive realisations based on voicing and lexical stress.

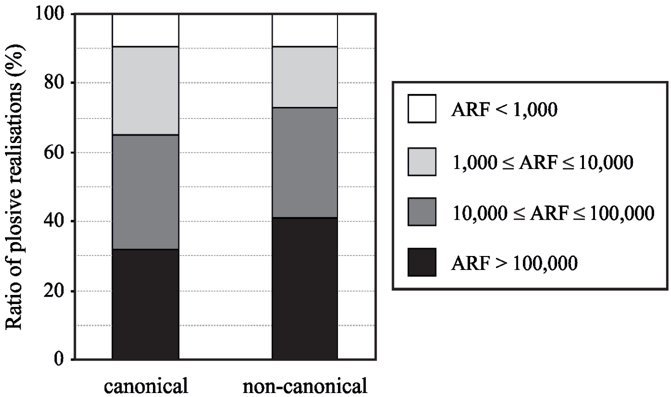

FIGURE 5: Proportion of canonical and non-canonical plosive realisations based on word frequency.

4.2 VARIABILITY OF VOICING

The results of auditory analyses confirmed our expectations: for the Czech phonologically voiced plosives, the inherent PF *voicing* is highly stable, with only 0.7% of them devoiced (this corresponds to less than 2 percent of all reduced realisations). As shown by Machač (2008) or Skarnitzl (2011, pp. 188–240), Czech fricatives manifest quite different behaviour; see also section 6.2 for more details on devoicing in fricatives.

More common is the voicing of phonologically voiceless plosives realised with a closure — this concerns 4.6% of all occurrences. If the semi-vocalised realisations are also included in the calculations (see Table 2), the share of the phonologically voiceless plosives realised as voiced grows to 7.1%. The difference in the frequency of voicing changes in the phonologically voiced and voiceless plosives is statistically significant: χ^2 (1; n = 2158) = 15.31; $p < 0.001$.

4.3 SUMMARY OF RESULTS — AUDITORY EVALUATION

On the basis of auditory analysis of over two thousand plosives, we stated their realisation types (canonical, semi-vocalised, fricativised, nasalised; devoiced, phonetically voiced) and we examined the frequency of their occurrence in relation to phonological voicing, place of articulation, word stress and word frequency. The results of auditory analysis may be summarized as follows:

- Voiced plosives are significantly more inclined to reduced pronunciation (almost 40% of all realisations) than the voiceless ones (almost 20%). This concerns particularly the resolution of the aerodynamic conflict between the need to keep a sufficient transglottal pressure differential through the non-realisation of the PF *closure*. For voiced plosives, the most common feature is thus *semi-vocalisation* (almost 60% of reduced realisations); for voiceless plosives, it is *fricativisation* (over 80%).
- The probability of reduced pronunciation (without a closure) rises with the decreasing distance between the obstruction and the vocal folds, as well as with the increasing mass of the active articulator: velars > palatals > alveolars > bilabials.
- Plosives in the onset of an unstressed syllable tend to be realised without a closure significantly more often than those in the onset of a stressed syllable.
- Plosives in highly frequent words are reduced more often than those in less frequent words.
- Phonetic voicing of phonologically voiceless plosives (realised both with and without a closure) is significantly more frequent (over 7% of all realisations) than devoicing of phonologically voiced plosives (only about 0.7%).

5 RESULTS II — ACOUSTIC ANALYSIS

The objective of the acoustic analyses was to identify (a) selected acoustic features of various types of plosive realisations and (b) the correlation between these features and types of realisations determined by means of listening.

5.1 DURATION

One of the phonetic universals is also the physiologically and aerodynamically conditioned and perceptually important fact that voiceless (tense) obstruents last relatively longer than their voiced (lax) counterparts (Ladefoged & Maddieson, 1996). It can therefore be expected that the voicing of phonologically voiceless plosives will lead to shorter durations, while the devoicing of phonologically voiced plosives will result in longer durations (note, however, that devoicing is quite rare in our material).

The tendency for phonologically voiceless plosives to become shorter when phonetically voiced is confirmed. (Please note that the term “phonetically voiced” here and also below refers to the process of an underlying voiceless plosive gaining voicing.) As shown in Table 4, their duration is shorter by 35–40%, as compared to the

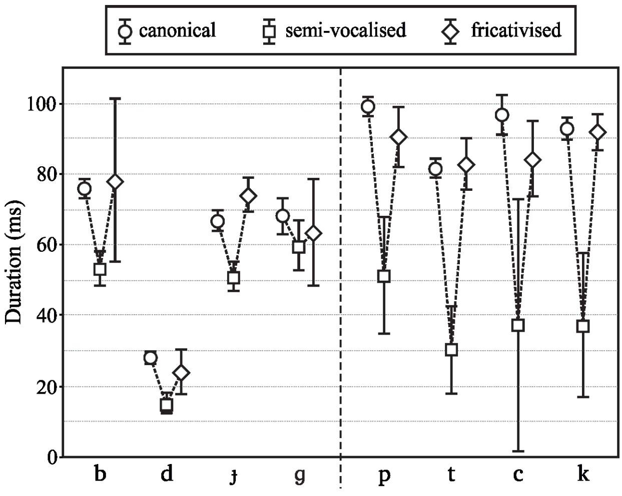

FIGURE 6: Duration of plosives realised canonically, as semi-vocalised and fricativised. The whiskers correspond to 95% confidence intervals.

duration of the corresponding canonical realisations, and the differences are highly significant (Tukey's post hoc test: $p < 0.001$).

	p	t	c	k
canonical	99.2 (24.3)	81.5 (23.1)	96.8 (31.8)	92.8 (21.7)
voiced	64.2 (16.7)	50.1 (14.0)	54.4 (16.4)	56.6 (15.0)

TABLE 4: Mean duration and standard deviation (ms) of phonologically voiceless plosives realised canonically and as voiced.

Turning to the effect of the individual types of plosive realisations on their duration, it is clear from Figure 6 that the difference in the duration of plosives realised canonically and those which were fricativised is minimal and statistically insignificant. Indeed, based on hitherto unpublished studies, the duration of the Czech canonical plosives and fricatives seems to be comparable. On the other hand, semivowels are considerably shorter in comparison to obstruents in general. That is also the case of the semi-vocalised realisations of the phonologically voiced and voiceless plosives in our material: the duration differences are, in most cases (except for /g/), statistically highly significant (Tukey: $p < 0.001$ for the voiced plosives and for /p t k/, $p < 0.01$ for /c/). It can be stated that these results support the perceptually introduced categories of plosives which become *semi-vocalised* or *voiced*.

5.2 INTENSITY DIFFERENCES

Another factor that can contribute to the acoustic characterisation of some of the realisation types of plosives is the difference between the intensity maximum and minimum within the plosive duration. To be able to capture relatively small dynamic changes, intensity values were measured in five equidistant steps.

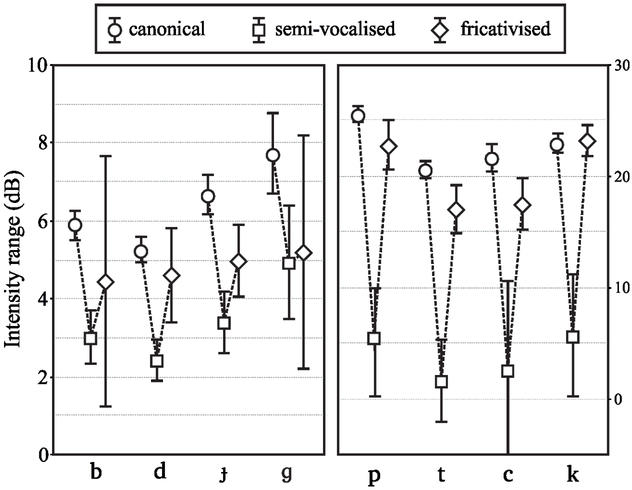

FIGURE 7: Intensity range in plosives realised canonically, as semi-vocalised and fricativised. The whiskers correspond to 95% confidence intervals.

An intensity maximum is often found in a canonically realised plosive either at its beginning (during the creation of the closure) or at its end (release of closure, release burst, onset of f_o). An intensity minimum is to be expected, in voiceless plosives, between the offset of f_o and the end of the closure phase. Both the release and f_o onset are included in plosive duration (Machač & Skarnitzl, 2009, p. 30). In the case of voiced plosives, the minimum is expected right before the release burst, due to the gradual equilibration of the transglottal pressure difference and decreasing amplitude of the vocal fold vibration.

The semi-vocalised plosives are characterised by two extrinsic PFs: (a) *opening of the vocal tract (large)*, which eliminates the potential aerodynamic conflict (see section 4.1.1); (b) continuous *full formant structure* resulting in a small intensity difference during the plosive (ca. 2–5 dB). As can be seen in Figure 7, the difference between the intensity maximum and the intensity minimum with the canonical and semi-vocalised plosives is highly significant (Tukey: $p < 0.001$; for /g/ $p < 0.05$). The intensity range of fricativised plosives does not differ from the canonical ones in an unequivocal way, although a slight tendency for lower intensity ranges can be observed in some of the plosives.

The intensity range also varies with voicing changes. Phonologically voiced plosives were found to have a bigger intensity range when devoiced (Tukey: $p < 0.001$; except for /b/ where the difference was insignificant). On the other hand, the intensity range of phonologically voiceless plosives, when voiced, was decreased ($p < 0.05$ for /k/; $p < 0.001$ for others). The average intensity range of voiceless and devoiced plosives is 7–20 dB higher than that of the phonetically voiced plosives (regardless of underlying phonological voicing).

It may be concluded that the above-stated findings support the meaningfulness of the perceptually determined categories of *semi-vocalised*, *voiced* (i.e., phonologically voiceless plosives which were phonetically voiced) and *devoiced* (phonologically voiced) plosives.

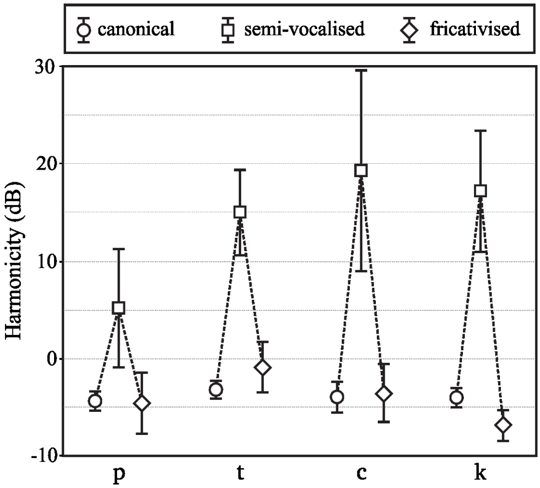

FIGURE 8: Harmonicity in phonologically voiceless plosives realised canonically, as semi-vocalised and fricativised. The whiskers correspond to 95% confidence intervals.

5.3 HARMONICITY

Harmonicity, which compares the energy of tone and noise components in the signal, is another measure which may acoustically differentiate between the perceptually identified categories of plosives. This expectation was fulfilled in that phonologically voiceless plosives realised in a semi-vocalised way manifest stronger tone components (higher HNR) than canonical ones due to the presence of a full formant structure (see Figure 8; $p < 0.05$ for /p/, $p < 0.001$ for others). A similar trend was recorded with phonologically voiced semi-vocalised plosives; nevertheless, the difference is only significant for /g/ ($p < 0.01$).

Voicing changes also induced significant differences in HNR values. In the case of /t/ and /c/ produced as voiced, harmonicity is significantly higher ($p < 0.001$) than in fully voiceless (canonical) realisations. The harmonicity of the devoiced plosives is significantly lower than that of the fully voiced plosives (except for /g/).

Harmonicity thus appears to be yet another factor which contributes to the differentiation between *canonical* and *semi-vocalised*, as well as *phonetically devoiced* and *voiced* realisations.

5.4 VOICING PROFILE

Voicing profile expresses the probability of phonetic voicing (presence of f_0) throughout a given sound. Figure 9 depicts values for phonologically voiceless (on the left, in the 0–1 range) and voiced plosives (on the right, in the 0.9–1 range) based on the realisation type. Note that the points marking the individual voicing profiles are placed so as not to overlap; in other words, they do not always correspond to one of the five measurement points.

A very similar voicing profile, with voicing probability not dropping below 90%, was found in all realisation types of phonologically voiced plosives (except for those

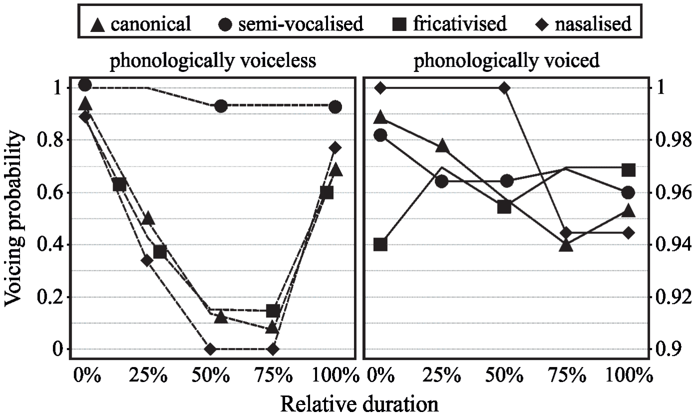

FIGURE 9: Voicing profile of plosives realised canonically, as semi-vocalised, fricativised, and nasalised.

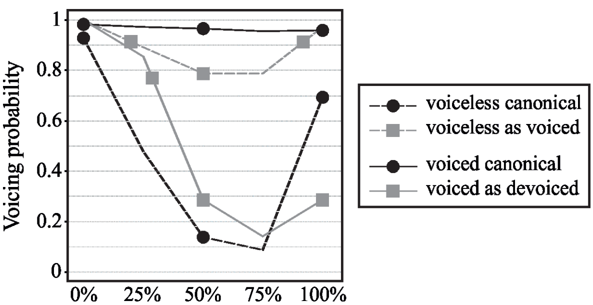

FIGURE 10: Voicing profiles in different changes of plosive voicing.

which were devoiced; see below) and also in semi-vocalised phonologically voiceless plosives. The remaining realisation types of voiceless plosives are characterised by a seemingly complicated and, at the same time, largely similar course of voicing probability: it decreases rapidly during approximately the first half of the segment due to the offset of f_0, voicing is least probable between 50 and 75% of the duration, and then voicing probability rises rapidly again as a result of f_0 onset associated with the following vowel. These differences confirm the relevance of the category *semi-vocalised* in phonologically voiceless plosives.

Voicing profiles of plosives with respect to their phonetic voicing are depicted in Figure 10. There are clear similarities in terms of phonetic voicing: the course of voicing probability is similar, on the one hand, across canonically realised phonologically voiced plosives and phonologically voiceless but phonetically voiced plosives, and on the other hand, across canonically realised voiceless plosives and phonologically voiceless but devoiced plosives. An exception among the phonologically voiced plosives is represented by the canonically realised /g/ which partially loses its voicing; this is in accordance with the smallest space between the vocal folds and the closure and the resulting intense aerodynamic conflict being at its most intense (see also, e.g., Ohala, 1983; Keating, 1984).

The above-stated findings support the importance of the perceptually stated categories of *phonetic voicing* and *devoicing*, although it is necessary to repeat that devoicing features only sporadically in our material.

5.5 SUMMARY OF RESULTS — ACOUSTIC ANALYSIS

The subsequent acoustic analysis was used (1) to identify acoustic characteristics corresponding to the individual types of realisations, (2) to confirm the meaningfulness of the categories identified perceptually and with the aid of acoustic displays and, at the same time, (3) to test ways of identifying the reduced forms in comparison to the canonical forms (duration, intensity range, harmonicity, voicing profile).

All the four examined acoustic parameters proved to be important for distinguishing between canonical and semi-vocalised realisations, and also for the identification of voicing changes (i.e., for distinguishing between canonically voiceless and phonetically voiced plosives, and between canonically voiced and devoiced ones). The results of the acoustic analyses are summarised in Table 5.

acoustic parameter	**manner of realisation**	**phonetic voicing**	**devoicing**
duration	can_vd > svoc	can_vl > vd+	rare, not verified
intensity difference	can_vd > svoc	can_vl > vd+	can_vd < vl–
harmonicity	can_vl < svoc (can_vd < svoc)	can_vl < vd+	can_vd > vl–
voicing profile	can_vd < svoc	can_vl < vd+	can_vd > vl–

TABLE 5: Comparison of acoustic parameter values for canonical (can) and semi-vocalised (svoc) realisations and for voicing changes (vl = phonologically voiceless, vd = phonologically voiced; vd+ phonetically voiced, vl– phonetically devoiced).

It is worth pointing out that fricativised variants did not show any statistically significant differences in any of the examined acoustic parameters. This may seem surprising; we hypothesised the intensity range within fricativised plosives to be lower than within canonical ones. As can be seen in Figure 7, however, such a tendency for lower values of intensity ranges can in fact be observed in /ɟ/, /t/ and /c/; overall, the difference as compared to the canonical plosives does not reach significance.

6 CONCLUSIONS

The aim of this study was to contribute to the description of the *relative stability of phonetic features* of phones, in particular through the study of the stability of phonetic features of Czech plosives in spontaneous speech. While describing the actual realisations of phones, the comparisons in the study are based on the "ideal phone", which serves as a theoretical construct including all phonetic features present in its full (canonical) realisation, that is the *inherent phonetic features*. In the actual realisation of the phone, these features tend to be weakened or completely absent and simultaneously replaced by *extrinsic phonetic features* which tend to be the inherent features of surrounding phones.

OPEN ACCESS

6.1 STABILITY OF PHONETIC FEATURES OF INTERVOCALIC PLOSIVES

The stability of an inherent phonetic feature of a given type of plosives may be quantified by determining the ratio of the number of plosives in which this feature was realised and the total number of plosives analysed. For instance, if 300 out of 500 voiced plosives are pronounced with closure, the stability of that feature is 0.6. Table 6 presents values of the relative stability of the phonetic features of plosives. It is interesting to compare these values with those concerning the stability of voicing in Czech phonologically voiced fricatives in the intervocalic position: [z] 0.67; [ʒ] 0.73; [r̝] 0.42; [v] 0.99 (based on Machač, 2008: Table 2). Unlike in plosives, phonetic voicing in the phonologically voiced fricatives is rather an unstable feature, especially with the fricative trill /r̝/ (0.42).

inherent PFs	**voiced plosives**	**voiceless plosives**
closure	0.62	0.81
orality	0.95	0.99
voicing/ voicelessness	0.99	0.93
place of articulation	1.00	1.00

TABLE 6: Relative stability of phonetic features expressing the ratio of Czech intervocalic plosives with the feature actually realised (see text).

The values obtained for intervocalic plosives imply the following hierarchy of the stability of phonetic features: *place of articulation* **>** *voicing and voicelessness* **=** *orality* **>** *closure*. The same hierarchy can be found for affricates and voiceless fricatives in Czech.

6.2 GENERAL DISCUSSION

The issue of speech reductions is approached in this paper from the perspective of phonetic features. We believe that if we know their relative stability in specific (classes of) phones in a given segmental context, we may be able, in modelling speech reductions, to distinguish more easily between reductions originating from the instability of a phonetic feature as an inherent trait of a given phone (such reductions are thus more easily predictable), and those originating from a joint influence of broader conditions. On the basis of such a distinction, it might be possible to be more successful in discovering the principles and limitations of reduction processes in relation to the context of speaking and the situation. The presence of such factors in everyday communication enables the listener to identify the meaning of the message, in spite of a possibly significant accumulation of very strongly reduced forms (see Introduction). The relationships within the system of speech reductions and their limitations can be assessed, for example, according to the specific form of *minimal phonetic information*, which represents for the listener the lowest acceptable level of the use of inherent phonetic features and the maximum level of the use of extrinsic features.

The following sentence may serve as an example of a probable minimal phonetic information: *já nevím, jestli jsem to už nepsala* “I don’t know whether I haven’t written it already” [jaːˈnɛviːmˈjɛstlɪsɛmtouʃˈnɛpsala] > [aˈə̃ĩˈɛ̰ʂ̩͡ʂəntəuʃˈə̃psaa]. Interpreting the reductions by means of phonetic features also provides for a rather apt description of features that are sometimes interpreted as elision. In full pronunciation, the speaker of the stated sentence would realise 27 segments, while only 17 segments are identified in the reduced form. That nevertheless does not mean that 10 segments have been elided. The number of segments which left no trace whatsoever is only 4: [v] in [nɛviːm] > [ə̃ĩ]; [t l] in [jɛstlɪ] > [ɛ̰ʂ̩]; [l] in [nɛpsala] > [ə̃psaa]. In the other 6 cases, the seemingly missing segment is represented by some of its phonetic features transferred to the neighbouring phone (parallel articulation), or it is indicated by a phonetic detail: [j] in [jaː] > [a] (residual soft voice onset at the beginning of the vowel); [n m] in [nɛviːm] > [ə̃ĩ] (nasalisation of vowels); [j] in [jɛstlɪ] > [ɛ̰ʂ̩] (palatalisation of the vowel); [ɪ] in [jɛstlɪ] > [ɛ̰ʂ̩] (duration and syllabicity of the voiceless obstruent); [n] in [nɛpsala] > [ə̃psaa] (nasalisation of the vowel). Machač & Zíková (2015) show that inherent phonetic features of “pseudo-elided” segments, transferred onto neighbouring phones where they act as extrinsic phonetic features, are in many cases able to distinguish the meaning of the message and are thus essential for adequate understanding.

Based on the analyses of strongly reduced speech units that represent a probable *minimal phonetic information* or are close to it, and with the knowledge of the *inherent stability of phonetic features*, it is possible to infer the stability of phonetic features from the perspective of wider semantic units; in these units, feature stability would be dependent on the structure, meaning and frequency of the lexical and sentential units, as well as on other factors. The knowledge of inherent and “contextual” stability should make it easier to model the minimal phonetic information in a given language, and can contribute to the understanding of the relationship between production and perception in ordinary speech.

Based on the existing observations of phone reductions (including elisions) in Czech, it can be for instance concluded that the most stable phones within words are those that are distinguished most by sonority. Specifically, should a word or a group of words contain vowels, sonorants, voiced obstruents and voiceless obstruents, then it is probable that vowels and voiceless obstruents (i.e., segments that are reciprocally most contrasting) will approximate full pronunciation most, while segments which support the acoustic contrast to a smaller degree (sonorants, voiced obstruents) will more likely be susceptible to lenition, to a loss of some of the phonetic features, or to elision.

It must be stressed that the method of phonetic features is regarded as a phonetic approach and not necessarily a phonological one. The primary objective is not to study the distinctive features within a given language system, but the study of phonetic or phonological rules in real speech. In this approach, we thus do not strive to create the smallest possible set of features; quite the opposite, all phonetic detail is taken into consideration.

REFERENCES

Blache, P., & Meunier, C. (2004). Language as a complex system: The case of phonetic variability. *Proc. of VI Congreso de Lingüística General*. Saint-Jacques de Compostelle.

Bladon, R. A. W., & Al-Bamerni, A. (1976). Coarticulation resistance in English /l/. *Journal of Phonetics*, *4*, 137–150.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *IFA Proceedings*, *17*, 97–110.

Boersma, P., & Weenink, D. J. M. (2011). *Praat: doing phonetics by computer (Version 5.2.30)*. Availablefrom http://www.praat.org.

Browman, C. P., & Goldstein, L. M. (1992). Articulatory phonology: An overview. *Phonetica*, *49*, 155–180.

Bybee, J. (2003). *Phonology and Language Use*. Cambridge: Cambridge University Press.

Cangemi, F., & Niebuhr, O. (2018). Rethinking reduction and canonical forms. In F. Cangemi, M. Clayards, O. Niebuhr, B. Schuppler & M. Zeller (Eds.), *Rethinking reduction* (pp. 277–302). Berlin: Walter de Gruyter.

Čermák, F., Křen, M. et al. (2004). *Frekvenční slovník češtiny*. Praha: Nakladatelství Lidové noviny.

Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics*, *39*, 253–260.

Fowler, C. A., & Saltzman, E. (1993). Coordination and coarticulation in speech production. *Language and Speech*, *36*, 171–195.

Greenberg, S. (1999). Speaking in short hand — A syllable-centric perspective for understanding pronunciation variation. *Speech Communication*, *29*, 159–176.

Hamann, S. (2009). The learner of a perception grammar as a source of sound change. In P. Boersma & S. Hamann (Eds.), *Phonology in perception* (pp. 111–149). Berlin: Mouton de Gruyter.

Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (Eds.), *Spontaneous speech: Data and analysis. Proceedings of the 1st Session of the 10th International Symposium* (pp. 29–54). Tokyo: The National International Institute for Japanese Language.

Keating, P. A. (1984). Physiological effects on stop consonant voicing. *UCLA Working Papers in Phonetics*, *59*, 29–34.

Keune, K., Ernestus, M., Van Hout, R., & Baayen, R. H. (2005). Social, geographical, and register variation in Dutch: From written MOGELIJK to spoken MOK. *Corpus Linguistics and Linguistic Theory*, *1*, 183–223.

Kohler, K. J. (1999). Articulatory prosodies in German reduced speech. *Proc. of the 14th ICPhS*, *1*, San Francisco, 89–92.

Ladefoged, P., & Maddieson, I. (1996). *The Sounds of the World's Languages*. Oxford: Blackwell Publishers.

Lahiri, A., & Marslen-Wilson, W. D. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, *38*, 245–294.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H Theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 403–439). Dordrecht: Kluwer.

Machač, P. (2006). *Temporální a spektrální struktura českých explozív*. Ph.D. dissertation, Institute of Phonetics, Faculty of Arts, Charles University, Prague.

Machač, P. (2008). Desonorizace českých intervokalických frikativ. In J. Volín & J. Janoušková (Eds.), *AUC Philologica 2/2007, Phonetica Pragensia XI* (pp. 105–116). Praha: Karolinum.

Machač, P., & Skarnitzl, R. (2009). *Principles of Phonetic Segmentation*. Praha: Epocha.

Machač, P., & Zíková, M. (2015). Parallel articulation: The phonetic base and the phonological potentiality. *Slovo a slovesnost*, *76*(1), 3–21.

Mitterer, H., & Blomert, R. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics*, *65*, 956–969.

OPEN ACCESS

Möbius, B. (2004). Corpus-based investigations on the phonetics of consonant voicing. *Folia Linguistica*, *38*, 5–26.
Niebuhr, O., & Kohler, K. J. (2011). Perception of phonetic detail in the identification of highly reduced words. *Journal of Phonetics*, *39*, 319–329.
Ohala, J. J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick & M. F. Miller (Eds.), *Papers from the Parasession on Language and Behavior* (pp. 178–203). Chicago, IL: Chicago Linguistic Society.
Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage (Ed.), *The Production of Speech* (pp. 189–216). New York, NY: Springer.
Ohala, J. J. (1989). Sound change is drawn from a pool of synchronic variation. In L. E. Breivik & E. H. Jahr (Eds.), *Language Change: Contributions to the Study of Its Causes* (pp. 173–198). Berlin: Mouton de Gruyter.
Pierrehumbert, J. B. (2003). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay & S. Jannedy (Eds.), *Probabilistic Linguistics* (pp. 177–228). Cambridge, MA: MIT Press.
Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America*, *118*, 2561–2569.
Pollák, P., Volín, J., & Skarnitzl, R. (2007). HMM-based phonetic segmentation in Praat environment. *Proc. of XIIth SPECOM 2007* (pp. 537–541). Moscow.
Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, *57*, 273–298.
Schuppler, B., Ernestus, M., Scharenborg, O., & Boves, L. (2011). Acoustic reduction in conversational Dutch: A quantitative analysis based on automatically generated segmental transcriptions. *Journal of Phonetics*, *39*, 96–109.
Skarnitzl, R. (2011). *Znělostní kontrast nejen v češtině*. Praha: Epocha.
Skarnitzl, R. (2018). Fonetická realizace slovního přízvuku u delších slov v češtině. *Slovo a slovesnost*, *79*, 199–216.
Skarnitzl, R., Šturm, P., & Machač, P. (2013). The phonological voicing contrast in Czech: An EPG study of phonated and whispered fricatives. *Proc. of Interspeech 2013* (pp. 3191–3195). Lyon.
Svirsky, M. A., Stevens, K. N., Matthies, M. L., Manzella, J. P., Perkell, J. S., & Wilhelms-Tricarico, R. (1997). Tongue surface displacement during bilabial stops. *Journal of the Acoustical Society of America*, *102*, 562–571.
Westbury, J. R. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America*, *73*, 1322–1336.

Pavel Machač | Ústav obecné lingvistiky FF UK
<pavel.machac@ff.cuni.cz>

Radek Skarnitzl | Fonetický ústav FF UK
<radek.skarnitzl@ff.cuni.cz>

This work was supported by the European Regional Development Fund project “Creativity and Adaptability as Conditions of the Success of Europe in an Interrelated World” (reg. no.: CZ.02.1.01/0.0/0.0/16_019/0000734) and by Charles University project Progres Q10, “Language in the shiftings of time, space, and culture”.