Neutralisation and contrast preservation: Voicing assimilation in Hungarian three-consonant clusters

Abstract

This paper studies the contextual variation in the voicing properties of three-consonant clusters (CC#C) in Hungarian. We investigate the velar–alveolar stop clusters /kt/ and /gd/, and the alveolar fricative–stop clusters /st/ and /zd/ in potentially voicing-neutralising and assimilating contexts. We show that in these contexts, regressive voicing assimilation in Hungarian is categorical, but partially contrast preserving, and that stops and fricatives are not affected in the same way. Fricatives resist voicing before a voiced obstruent and are devoiced utterance-finally. This is a phonetically unfavourable position, therefore other duration-related cues step up to prevent complete laryngeal neutralisation.

Keywords: voicing assimilation, devoicing, Hungarian, contrast, neutralisation

1 Introduction

Understood in its most well-known sense, phonological neutralisation refers to the case when two or more contrastive segments suspend their contrast under specific conditions, whereby only a limited set of the contrastive segments can occur in a particular position. Neutralisation processes are considered to be *complete* when there is no difference between the underlyingly contrasting members either in production or in perception for any of the possible phonetic correlates of a given contrast in a given context. That is, forms (e.g., voiced and voiceless obstruents) that are distinguishable in certain contexts (e.g., in intervocalic position) are indistinguishable in the neutralising contexts (e.g., word-finally or in pre-obstruent position). In such positions a devoiced segment cannot be distinguished from an underlyingly voiceless segment either in its phonetic properties or in its phonological behaviour in any way.

Voicing contrast may be neutralised in several ways. The most obvious example comes from regressive voicing assimilation during which the last member of an obstruent cluster (the "trigger" of the assimilation) expands its voice feature to the preceding obstruents (the "target"). Another possibility is positional neutralisation, typically syllable-, word- and utterance-final devoicing belong here. In both cases the issue of categoricality vs. gradience arises, that is, whether the segment in question coincides with the underlyingly voiced or voiceless segments of the language or can rather be placed on a scale somewhere in-between, namely, whether other phonetic cues (to be discussed below) can maintain the contrast at least to some degree. These phonetic cues of a phonological distinction may not necessarily be realised in the site of the distinction, but might be cued in the neighbouring segments, too, just as in the case of voicing contrast: the opposition between obstruents is often cued not during the consonant constriction phase but in the neighbouring vowels, in their duration or other acoustic properties (cf., for instance, Steriade 2008). It is also possible that a phonological process is (acoustically) neutralising for some speakers of the speech community, while for others it is not, or that under certain circumstances (e.g., fast, colloquial speech) it is neutralising, while in others (e.g., formal register, writing task) it is not.

One of the best studied topics regarding partial or incomplete neutralisation in the past decades has been word-final devoicing. The word-final position is a very common locus for non-assimilatory laryngeal neutralisation, a place where laryngeal distinctions (including aspiration and glottalisation) collapse to a single value in a number of unrelated languages. It has been reported for German (Port et al. 1981; O'Dell & Port 1983; Charles-Luce 1985), Catalan (Dinnsen & Charles-Luce 1984; Charles-Luce 1993), and Polish (Slowiaczek & Dinnsen 1985; Slowiaczek & Szymanska 1989) that word-final laryngeal neutralisation, contrary to traditional analyses, leaves some residual cues to the phonological voicing of obstruents. However, Fourakis & Iverson (1984) and Kahlen-Halstenbach (1990) found that word-final devoicing is indeed phonetically complete in German. Jassem & Richter (1989) report the same for Polish. Experimental evidence concerning voicing assimilation is less abundant, but also quite varied. There is experimental work demonstrating that voicing assimilation is non-neutralising, and it is a low-level phonetic process (e.g., Charles-Luce 1993 on Catalan and Burton & Robblee 1997 on Russian). In contrast, Hallé & Adda-Decker (2011) found that whenever it occurs, voicing assimilation is categorical in French. Strycharczuk & Simon (2013) claim the same about West-Flemish. The issue of complete vs. incomplete laryngeal neutralisation is far from being settled either empirically or theoretically.

Regressive voicing assimilation ("RVA") in Hungarian is generally considered to be a completely neutralising phonological process (see for instance Siptár & Törkenczy 2000, and the references therein). It is viewed as a feature changing process that changes the "+" value of the [voice] feature of a given consonant into "–", or the other way round, and thus voiceless and devoiced or contextually voiced and underlyingly voiced segments cannot be distinguished on the basis of their phonetic or phonological behaviour. A notably different approach to RVA is presented in Jansen (2004) who concludes that RVA in Hungarian leads to incomplete neutralisation of laryngeal distinctions in target sounds. He found residual traces of the underlying contrasts between /k/ and /g/, and /ʃ/ and /ʒ/ in terms of the voicing of the target consonants before the voiced obstruents /d/ and /z/. The difference between /ʃ/ and /ʒ/ in terms of the duration of the preceding vowel was also preserved in the presence of a following obstruent.

The present paper studies word-final obstruent clusters in Hungarian with the help of an acoustic (production) experiment. The word-final clusters /kt/−/gd/ and /st/−/zd/ are examined in the following positions: utterance-finally, word-finally when followed by a voiced obstruent, a voiceless obstruent, a sonorant consonant, and as a point of reference, in word-medial intervocalic position, too. This topic is of interest for two reasons: (i) there are very few studies dealing with the laryngeal characteristics of three-consonant clusters in general, and in Hungarian in particular, and acoustic phonetic analyses on obstruent voicing in Hungarian are very scarce too; (ii) the study of consonant clusters can shed further light on the issue whether regressive voicing assimilation (RVA) in Hungarian is a completely neutralising process or not.

We will seek to answer the following research questions:

1. How do the various phonetic correlates of voicing behave in Hungarian in potentially voicing-neutralising contexts?
2. Is voicing assimilation in Hungarian neutralising or incomplete?
3. In what ways does voicing in simplified clusters differ from voicing in singleton consonants?
4. How is the laryngeal contrast in utterance-final clusters implemented in Hungarian?

2 Experiment: material, procedure, statistical analysis

Words ending in /kt/–/gd/ and /st/–/zd/ were tested in the following three positions:

(1) absolute word-final (utterance-final) position,

(2) word-final sentence-medial position, where the target obstruents were followed by one of the following triggers: (a) voiced obstruent /b/, (b) voiceless obstruent /p/ (c) sonorant consonant (/l/ or /m/), and

(3) sentence-medial intervocalic position.

The target obstruents were always preceded by the vowels *a* /ɔ/ or *e* /ɛ/; in intervocalic position, the vowel following the target consonants was also *a* /ɔ/ or *e* /ɛ/. The test words were *kontakt* /ˈkontɔkt/ 'contact', *smaragd* /ˈʃmɔrɔgd/ 'emerald', *kereszt* /ˈkɛrɛst/ 'cross' and *gerezd*

/ˈgɛrɛzd/ ‘slice’. The intervocalic test words were *Magda* /ˈmɔgdɔ/ (proper name), *akta* /ˈɔktɔ/ ‘file’, and *kezdet* /ˈkɛzdɛt/ ‘beginning’. The use of minimal pairs was avoided on purpose because in our experience, despite the use of a fair number of distractors, subjects tend to overemphasise the differences in their pronunciation. Stimuli were embedded in carrier sentences: e.g., *Egy* ***gerezd mangó*** *díszíti a tálat.* ‘The dish is decorated with a slice of mango’. (The complete list of test sentences can be found in the Appendix.) The carrier sentences were 10–13-syllable-long, neutral declarative sentences, the target and the trigger occurred in the same intonational phrase; word stress in Hungarian falls on the first syllable. Note that the “fricative clusters” we tested are not composed of purely fricatives, they are the combination of an alveolar fricative followed by an alveolar stop. We chose fricative+stop clusters because due to phonotactic constraints, we could not have provided a balanced test set of voiced and voiceless fricative+fricative clusters. Also, our aim was to elicit a fairly natural speech tempo. These factors partly contributed to the second member of the cluster (“C_2”) being deleted in about half of the cases (we will refer to these as “simplified” clusters). In those clusters where all three consonants are obstruents, C_2 was deleted in 65% of the cases, this meant that for certain comparisons, statistical inference could not be reliably provided due to the low number of observations (for the affected cases, see the next section).

Six native speakers of Hungarian participated in the experiment, aged 22–26, none of them reported any speaking, hearing or reading disorder. They were all naive as to the aims of the experiment and participated as a courtesy to the authors. Subjects read the test sentences and fillers from a monitor screen in a randomised order, which was generated by SpeechRecorder.[1] Each test sentence was read five times, but the first reading of each subject was considered as the familiarisation phase, and was not taken into consideration. We investigated five contexts for four words by six subjects with four repetitions, which resulted

[1]http://www.bas.uni-muenchen.de/Bas/software/speechrecorder/

in altogether 96 × 6 items, one of which had to be removed as one of the subjects pronounced a trigger /v/ instead of /b/ on one occasion; altogether we collected 575 observations. Recordings were made in a sound-attenuated room, using a Sony ECM-MS907 microphone connected to a laptop through an M-Audio MobilePre USB preamplifier external sound card. The material was recorded at a 44,100 Hz sampling rate, and was resampled at 22,050 Hz for the various acoustic measurements.

The acoustic analysis was carried out in Praat (version 5.3.12; Boersma & Weenink 2012). The spectrograms were segmented manually by the authors and the following measurements were carried out on the basis of the inserted boundaries:

Correlates of laryngeal contrast measured in the experiment

1. *Phonation*-related correlates of laryngeal contrast
   a. the absolute length of the voiced interval ("voicing duration")
   b. ratio of the unvoiced part compared to the total length of the consonant ("% of devoicing")

2. *Duration*-related correlates of laryngeal contrast
   a. duration of the preceding vowel
   b. duration of the target consonant
   c. vowel-to-consonant duration ratio

Voicing was measured manually, based on the visual inspection of the spectrograms and oscillograms. In the case of stops, voicing was measured during the closure phase, i.e., up to the burst, but the release phase was not included (similarly to the methodology applied by Strycharczuk 2012, for instance). In the case of fricatives, voicing was measured during the whole duration of the frication noise and a low-pass filter with a cut-off frequency of 500 Hz was used to securely determine the exact portion of the voicing oscillation. We measured two parameters: the absolute length of the voiced interval within the target consonants in seconds (referred to as "voicing duration" in the figures below) and the ratio of the unvoiced part compared to the total length of the consonant (referred to as "devoicing" in the figures below).

As for the duration-related parameters, in the case of fricatives, the interval of frication noise was measured. In the case of stops, closure duration and release burst duration were measured. Since absolute segment durations are highly variable due to different speaking rates (cf., e.g., Port & Dalby 1982), the vowel-to-consonant duration ratio was also measured, that is, the duration ratio of the cluster under scrutiny and the preceding vowel. It has been long observed in the literature that the ratio between vowel duration to stop closure or fricative constriction remains relatively constant in words with the same underlying voicing feature: the vowel-to-consonant duration ratio is generally larger for voiced obstruents than for voiceless obstruents. Many perception-driven accounts derive the inverse patterning of voiced–voiceless obstruent length and preceding vowel duration as a form of mutual auditory enhancement for the voicing contrast. The idea is that increased vowel duration makes the duration of a following obstruent appear shorter, and conversely that a decrease in vowel duration increases the perceived duration of a following obstruent, and that vowel duration and obstruent duration are therefore integrated into a single percept (Port & Dalby 1982; Port & Leary 2005; Massaro & Cohen 1983; Kluender et al. 1988). This hypothesis has been largely supported by experimental evidence: listeners pay attention especially to the relative duration of a vowel and the constriction duration of a following obstruent (Javkin 1976; Parker et al. 1986; Kingston & Diehl 1994), which may serve to preserve the voicing contrast in phonetically unfavourable positions. The durational relationship between consonants and preceding vowels is often cited as "pre-fortis clipping" in the English phonetics/phonological literature (e.g., Harris 1994;

Wells 2000). For an extensive overview on the choice of these (and other) acoustic correlates of voicing contrast and voicing assimilation, see Jansen (2004), Bárkányi & Kiss (2007) and Strycharczuk (2012), and the references therein. In the present paper we will not discuss other cues that are also cited in the literature as correlates of laryngeal contrast, such as the intensity of the frication noise and the release burst, or the "low-frequency" spectral features – the fundamental frequency and the first formant – of the surrounding vowels.

The statistical analyses were carried out using R (R Development Core Team 2008), version 3.3.1. The acoustic correlates of voicing were analysed with linear mixed-effects models, using the *lme4* package (Bates et al. 2015) in the following way. For each phonetic outcome variable a mixed-effects model was fitted with a random intercept for subjects. The fixed predictors of the models were the various "target" undergoer sounds (the singleton consonants, /t/, /d/, /s/, /z/, and the clusters, /kt/, /gd/, /st/ and /zd/). The base or "null" model always consisted of these fixed predictors plus the random intercept for subject. This base model was then compared to another one which also contained the target predictor as a random slope for subjects. The random-slope model was only retained when the log-likelihood test showed a statistically significant improvement over the base model (at $\alpha = 0.05$). The results below come from the best-fitting model. We did not include "item" as a random effect because we only used one particular word for each of the phonetic targets, hence items did not vary across the various predictor variables. To make sure that only those comparisons were included in the model that our research focused on, we employed planned orthogonal contrast coding (for the method, see, among others, Field et al. 2012). The contrast coding distinguished between 1. target class: stop+stop (/kt, gd/) vs. fricative+stop (/st, zd/); 2. /kt/ vs. /gd/, and 3. /st/ vs. /zd/. We also contrasted the voicing of singletons to clusters in various ways (for details, see below). We report the results of the best-fitting model by including the beta-coefficient for the given contrast (indicated with *b* here), the *t*-statistic, together with its degrees of freedom, and a calculated *p*-probability of the *t*-statistic. The significance level was set at $\alpha = 0.05$. All the *t* tests were two-tailed (alternative hypothesis: the mean between the two contrasted groups is not zero). The *lme4* package does not provide a *p*-value for the *t* tests, and therefore we used the *lmerTest* package (Kuznetsova et al. 2016) to acquire probabilities; the *t* test of this package uses Satterthwaite approximations to degrees of freedom. The effect size measure used in the paper is Pearson's correlation coefficient *r* (see Field et al. 2012:457).

3 Results and discussion

3.1 Utterance-final position

In utterance-final position, Hungarian obstruents are usually claimed to fully preserve their laryngeal contrast as there is no word-final voicing neutralisation in this language. In an earlier experiment with singleton consonants (Bárkányi & G. Kiss 2015), we, however, found that Hungarian alveolar fricatives were realised voiceless in utterance-final position. This does not necessarily mean that the laryngeal contrast is neutralised in this context since other acoustic cues might help maintain the phonological contrast, which appeared to be the case in the mentioned study as well. Consonant duration (voiceless consonants were longer than their voiced counterparts), duration of the preceding vowel (vowels were longer before voiced consonants than before voiceless consonants), as well as their ratio were significantly different for the members of the voiced–voiceless fricative pairs. Consequently, the voicing contrast in absolute final position was found to be robust for alveolar stops and partial – though clearly present – for alveolar fricatives in Hungarian.

In the present study, in utterance-final position, neither the stop nor the fricative clusters were simplified, both word-final consonants (C_1 and C_2) were preserved. (This was not the case in word-final utterance-medial context as shown in detail below.) The duration of the clusters was on average 121–162% of that of singleton consonants. Contrary to earlier claims but in accordance with our findings on singletons, fricatives were very much devoiced in this context (Figure 1).

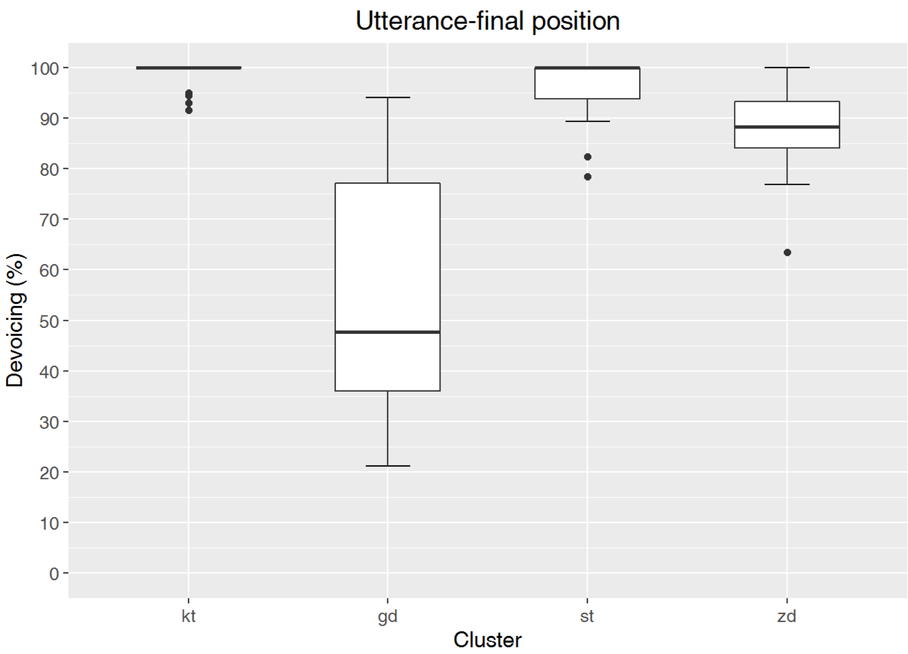

Figure 1

Boxplots showing the ratio of the unvoiced part to total consonant length in Hungarian utterance-final /kt gd st zd/ clusters.

As expected, /kt/ was realised voiceless (98.92% on average, SE = 0.52), while /gd/ was voiced (mean: 55.44%, SE = 4.43), the difference was statistically significant: $b = -21.74$, $t(87) = -12.98$, $p < 0.001$. The same ratio for /st/ and /zd/ was smaller, /st/ being devoiced in 96.46% (SE = 1.23) of the duration of the cluster on average, while /zd/ was devoiced in 87.93% (SE = 1.61). Unlike in the case of singletons, this difference was statistically significant, but the effect size was relatively small ($b = 4.26$, $t(87) = 2.55$, $p = 0.013$, effect size: $r = 0.42$). Figure 2 illustrates the lack of phonation in the utterance-final fricative + obstruent clusters.

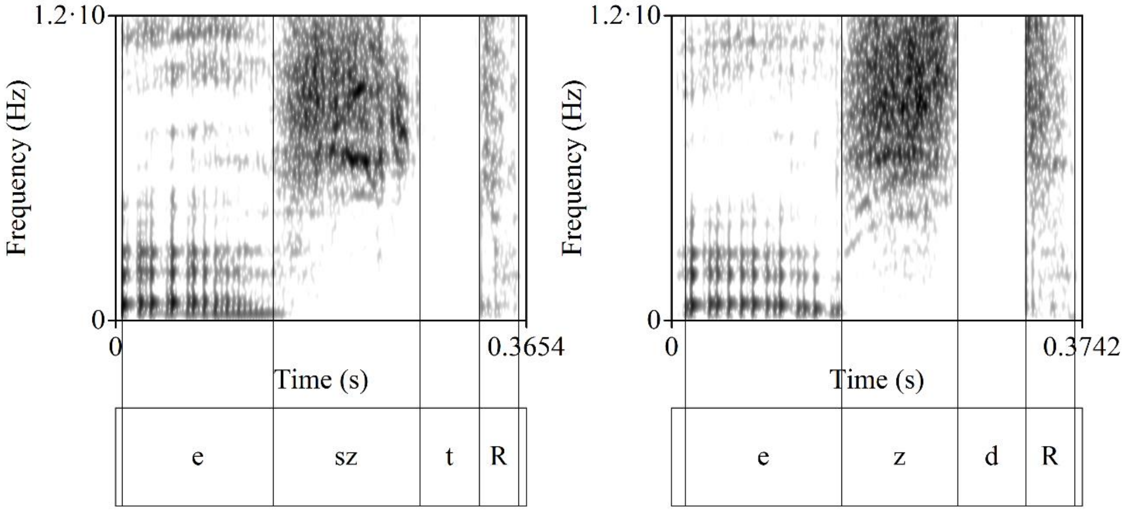

Figure 2

Spectrograms showing utterance-final *ker**eszt*** ‘cross’ and *ger**ezd*** ‘slice’ pronounced by Speaker 2 (“R” in the annotation stands for release burst). There is no phonation during the fricative+stop cluster intervals.

As for the actual amount of voicing, it also turned out to be highly significant for stops, just like in the case of singleton stops in our previous study (the voiced part in /kt/ clusters was 1.47 ms on average, SE = 0.7; in /gd/ it was 56.06 ms on average, SE = 5.77; $b = 27.296$, $t(5.17) = 6.67$, $p < 0.001$), and was also significant in the case of fricatives (/st/ clusters on average contained a voiced phase of 6.02 ms, SE = 2.02, while the voicing duration in /zd/ clusters was 17.00 ms, SE = 2.31; $b = -5.68$, $t(5.91) = -2.48$, $p = 0.048$).

The duration of the vowel preceding the obstruent cluster turned out to be a significant parameter differentiating voiced and voiceless obstruents and obstruent clusters in the utterance-final context. The vowel was always longer when preceding voiced obstruents (before /kt/ 83.75 ms, SE = 2.48 vs. before /gd/ 105.67 ms, SE = 2.66; $b = 10.95$, $t(5.67) = 5.15$, $p = 0.0018$; before /st/ 109.67 ms, SE = 4.64 vs. before /zd/ 138.54 ms, SE = 4.39; $b = -14.43$, $t(6.08) = -8.27$, $p < 0.001$), which makes the vowel-to-consonant duration ratio also significant

(/kt/ vs. /gd/ $b = 0.1$, $t(5) = 4.66$, $p = 0.0055$; /st/ vs. /zd/ $b = -0.13$, $t(5.79) = -10.58$, $p < 0.001$), despite the fact that the stop+stop clusters did not significantly differ in their length (/kt/ was 174.71 ms long on average, SE = 4.53 vs. /gd/ 157.79 ms, SE = 6.28; $b = -8.45$, $t(5) = 4.72$, $p = 0.13$). On the other hand, /st/ was significantly longer than /zd/ (mean duration of /st/: 208.17 ms, SE = 3.99 vs. /zd/ 174 ms, SE = 4.55; $b = 17.08$, $t(5) = 5.04$, $p = 0.004$). Gráczi (2010), testing nonsense words, did not find a statistically significant result (only a tendency) in vowel duration before fricatives in absolute final position, but consonant duration and vowel–consonant duration ratio did show vowel shortening effects in her study as well.

These results are consistent with the aerodynamic phonetic facts of obstruent phonation. In order to initiate voicing, the vocal folds must be lightly adducted and there must be sufficient air flowing through the glottis. When producing stops, a complete closure is made in the oral cavity, that is, all "exit valves" are closed by definition (Ohala 1983). This means that the air flowing through the glottis accumulates in the oral cavity, thus oral pressure will approach and exceed subglottal pressure and therefore, voicing is extinguished. The longer the closure, the greater its likelihood to devoice, and conversely, the shorter a stop closure is, the more likely it is to remain voiced. This explains the well-known tendency among languages to have shorter voiced stops (and fricatives as well) than their voiceless counterparts. Some languages shorten voiced stops so much that they cease to be stops and are realised as approximants. It also explains why languages in general prefer voiceless stops over voiced stops. It was also observed long ago (e.g., Greenberg 1970) that voiced stops at some places of articulation are "better" than at others: namely, velar stops and voicing show the greatest incompatibility, while labial stops and voicing the greatest compatibility. Certain articulatory gestures that enlarge the oral cavity (like lowering the larynx and/or the mandible) can help to maintain voicing for a longer time. Note that there is a certain degree of oral cavity enlargement that occurs naturally due to the natural compliance of the walls of the mouth.

As for fricative voicing, based on the description above, we could assume that voicing is easier in fricatives than in stops since they have a continuous venting of oral air pressure. This, however, is not the case (see, e.g., Stevens 1998 and Jackson & Shadle 2000). Aperiodic turbulent noise requires a large volume velocity as well as a narrow constriction (plus a sharp obstacle downstream) in the supraglottal vocal tract. As a result, the vocal folds are to be widely abducted, and supraglottal air pressure must exceed subglottal pressure. Voicing, on the other hand, as we have just discussed, requires the folds to be closely adducted, subglottal air pressure to be greater than supraglottal pressure, and the supraglottal vocal tract to be relatively open. The contradictory articulatory targets of voiced fricatives thus imply that aperiodic turbulent noise and passive voicing cannot be maintained simultaneously: an abducted glottis and a decrease in the transglottal pressure differential both remove the basic conditions for vocal fold oscillation. As Ohala (1983: 201) writes, "for the sake of continued voicing the oral pressure should be low, but for the sake of frication the oral pressure should be high. Meeting both of these requirements may be difficult. To the extent that if the segment retains voicing it may be less of a fricative, and if it is a good fricative it runs the risk of being devoiced". Due to the inherently contradictory conditions of turbulence and voicing, their simultaneous maintenance can only be achieved if the vocal tract is "actively reconfigured", thereby inhibiting the build-up of intraoral pressure as the supraglottal constriction area becomes narrow (and so voicing can be maintained). Several articulatory gestures can be used to produce active voicing in fricatives (and stops, for that matter, as mentioned above). The time interval over which the vocal folds can continue to oscillate can be extended (and so the build-up of intraoral pressure at the constriction can be delayed) if the vocal tract is actively expanded by, for example, raising the soft palate, by advancing the tongue root so that there is an outward movement of the neck surfaces, by lowering the larynx, by expanding the pharyngeal volume, by decreasing (laxing) the stiffness of the vocal tract walls, or a combination of these gestures. The implementation of these articulatory gestures is possible only within certain limits, especially in the aerodynamically unfavourable utterance-final position.

The question arises whether and to what extent our acoustic findings are relevant perceptually. Since /zd/ was realised 87.93% voiceless in this position, it is really questionable whether the remaining small amount of voicing is perceptible at all. There are very few perception studies on voicing contrast in Hungarian. Bárkányi & Mády (2012) examined the perception of utterance-final /s/ vs. /z/ using synthesised speech. Subjects heard the test words in isolation and had to respond in a forced-choice test. The test words were *méz* [meːz] 'honey' ending in underlying /z/ and *mész* [meːs] 'whitewash' with underlying /s/, /m/ being 50 ms long, /eː/ 250 ms and the fricative 210 ms. These durations of the segments were determined on the basis of several earlier acoustic studies. Synthesis was carried out in HLSyn (High-Level Speech Synthesis software) which is based on the combination of articulatory parameters of vocal tract aerodynamics and a Klatt-type formant synthesiser (Hanson et al. 1999). Voicing was added in 10% steps to the fricative, i.e., there were 11 different stimuli from completely voiceless to completely voiced items. The inflection point turned out to be at 30% voicing (SD = 8%), that is, if less than 70% of the fricative interval is voiceless the segment is more likely to be perceived as voiced, or to put it the other way, if a final fricative is over 30% voiced, it is more likely to be categorized as voiced. In both our studies (singletons and clusters), utterance-final fricatives contained considerably less than 30% voicing. This suggests either that fricatives in this context are fully neutralised for voicing, or that duration related parameters step up as secondary perceptual cues that help encode at least partial contrast preservation.

In order to tease apart the role of segment duration, Bárkányi & Mády (2012) carried out a second experiment. Since speakers typically talk at different rates, the absolute durations of the segments are highly variable. As mentioned in Section 2, it has been found for English and

German that the ratio of vowel duration to stop closure or fricative constriction remains rather constant in words with the same voicing feature. In this experiment too, synthesised tokens of the words *mész* [meːs] 'whitewash' and *méz* [meːz] 'honey' were used. Synthesis was carried out in HLSyn again. As the mean inflection point was at 30% in the first experiment, with a standard deviation of 8%, the ratio of voicing was kept constant on five levels: 14, 22, 30, 38 and 46% voicing of the fricative interval. (30% is the amount of voicing where there is exactly a 50-percent chance that a speaker perceives either /z/ or /s/ with a ±1 and ±2 SD.) The duration of V+C was set at 360 ms. The mean V/C duration ratio for unvoiced clusters in our present study was 0.48 and 0.8 for voiced ones, while it was 0.68 for utterance-final /s/ in our earlier acoustic experiment on singleton consonants and 1.2 for /z/. In Bárkányi & Mády's paper, the V/C ratio ranged from 0.57 to 1.77. The minimal segment duration for both vowels and consonants was 130 ms, the maximum 230 ms. At each voicing level, vowel and fricative lengths were changed in 10-ms steps starting with a 130-ms-long vowel and a 230-ms-long consonant, and ending up with a 230-ms vowel and a 130-ms consonant. In this way 55 different sound files were created. In the case of the most ambiguous stimulus, i.e., when 30% of the fricative interval is voiced, listeners were as likely to perceive the stimulus as voiced as they were to perceive it as voiceless if the vowel was at least 160 ms long (V/C duration ratio is 0.8). If the vowel was longer, subjects were more likely to perceive a final /z/; if it was shorter, they were more likely to perceive the final segment as /s/. Note that the value of vowel-to-consonant ratio was exactly the same as in our present study for /zd/, which, however, contained much less voicing. These results indicate that the V/C ratio does play a role in categorisation, at least if the amount of voicing is set to a borderline value. Bárkányi & Mády (2012) report that ratios higher than 1.0 resulted in the perception of a voiced fricative, that is, very long vowels can elicit a voiced response even with little voicing. In the case of 14% voicing this would be a very unlikely ratio, namely 2.1; with 22% voicing, the V/C duration ratio has to be at least 1.7 in order to induce voiced responses, while 46% voicing induces a voiced response independently of vowel length. In the case of 38% voicing this value is 0.53. This is in line with our earlier research and our claim that in the perception of laryngeal specifications, a whole cue-complex plays a role and not a single phonetic feature. These cues might mutually enhance or hinder each other. Based on our acoustic studies and the perception experiments presented here briefly, we can confirm that Hungarian word-final fricatives might have taken the very first step on their way to neutralising the voicing contrast in utterance-final position: the majority of actual utterance-final voiced fricative realizations fall into the perceptually ambiguous region, that is, they are partially neutralised, although secondary perceptual cues prevent complete neutralisation. These findings are in line of what Myers (2012) suggests: historically, the perceptual basis of word-final devoicing is limited to fricatives in utterance-final position (the context discussed here). The phonological pattern of devoicing is then generalised from utterance-final words to all words and from fricatives to all obstruents in a given language.

The difference in voicing duration between the voiced and voiceless clusters is very robust in the case of stops (as shown above); it is more similar – although still significantly different – in the case of fricative clusters. (Utterance-final /s/ and /z/ did not differ in the amount of voicing in our earlier experiment, Bárkányi & G. Kiss 2015.) It is noteworthy that the actual amount of voicing in the case of singleton /z/ was 17.90 ms on average and in the case of /zd/ in the present experiment it is 17.39 ms. This value might be a ceiling for fricatives to how much phonation they can be produced with in this phonetically unfavourable position. Schmidt and Willis (2011) examining /s/ voicing in Spanish claim that around 14–15-ms voicing is simply coarticulatory and does not indicate a planned gesture.

Evidence from the present experiment suggests that contrary to earlier claims (as in Vago 1980; Zsigri 1994; Siptár & Törkenczy 2000 for instance), alveolar fricatives in Hungarian are realised voiceless in utterance-final position. This, as we have argued, does not necessarily mean that the laryngeal contrast is neutralised in this context since other cues might maintain the contrast, which appears to be the case here as well. Consonant duration (voiceless consonants are longer), duration of the preceding vowel (vowels are longer before voiced consonants), as well as their ratio are significantly different for the members of the voiced–voiceless clusters. These results are novel in the sense that shortening before voiceless obstruents (known as "pre-fortis clipping" and described above in Section 2) is well-known to be present in aspirating languages like English or German, but not in voicing languages like Hungarian or Spanish. In English, it is the vowel length that prevents the complete laryngeal neutralisation of obstruents in final position. In Hungarian the vowel-to-consonant duration ratio also seems to be an important cue to maintain partial contrast preservation in utterance-final position, but it is not systematically present in other environments, such as intervocalically.

3.2 Intervocalic and pre-sonorant position

The intervocalic word-medial position and the word-final pre-sonorant position were included in the experiment as points of reference. We assumed that these contexts were phonetically "favourable" positions, where contrast-preservation should be robust. According to Steriade's p-map theory (Steriade 2008), which is based on the absolute and relative perceptibility and confusability of different contrasts across the different contexts where they might occur, the salience of a segmental contrast is cue-based and fundamentally depends on two factors: (i) the quality and quantity of the inherent acoustic cues of the given sound and (ii) the quality and quantity of the acoustic cues the sounds in its immediate context provide. (On the phonetically favourable nature of the prevocalic, intervocalic context for contrast preservation, see also Steriade 1997; Hayes 1999; Hayes & Steriade 2004; Wright 2001, 2004 and the references therein.) The laryngeal and place of articulation features of obstruents, especially stops, are not cued in the stop itself but rather in the transition phase to the neighbouring vowel. The vowel/sonorant on the right hand-side is more important than the transition from the preceding vowel/sonorant into the stop. This means that the most favourable phonetic context to maintain phonological contrast is the intervocalic or inter-sonorant position. It is in this position that the relevant cues for the contrast are available in number and quality: closure voicing, closure duration, the duration of the vowel, F1 values in the vowel, burst duration and amplitude, VOT value, f_0 and F1 values at the onset of voicing in the second sonorant.

Our results backed up the above expectations. Figure 3 shows the differences in the voicing ratio of the clusters (/kt/ had a devoiced portion of 96.98% on average, SE = 0.86 vs. /gd/ which is devoiced only in 8.78%, SE = 3.05; $b = -43.96$, $t(5.5) = -15.75$, $p < 0.001$; mean devoicing for /st/: 91.07%, SE = 1.55 vs. /zd/: 21.71%, SE = 6.28; $b = 34.68$, $t(5.02) = 6.4$, $p = 0.0013$).

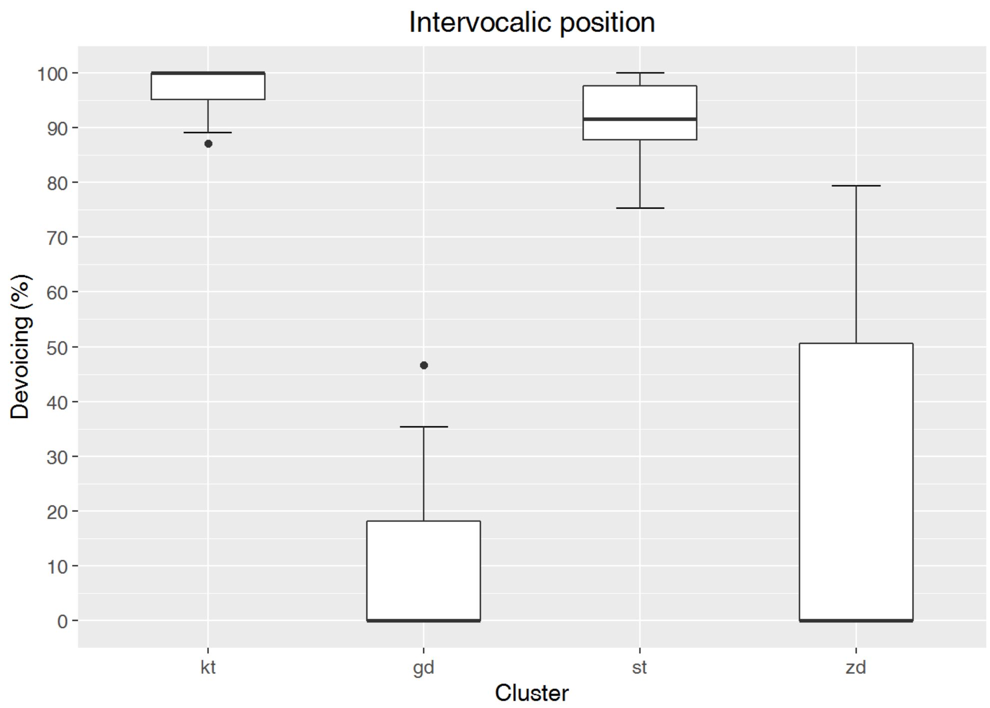

Figure 3

Boxplots showing the ratio of the unvoiced part to total consonant length in Hungarian intervocalic /kt gd st zd/ clusters.

One of our subjects seemed to have a difficulty in producing fully voiced fricatives even in this "ideal" phonetic environment, the proportion of voicelessness of this subject's /zd/ clusters was 74% on average, see Figure 4.

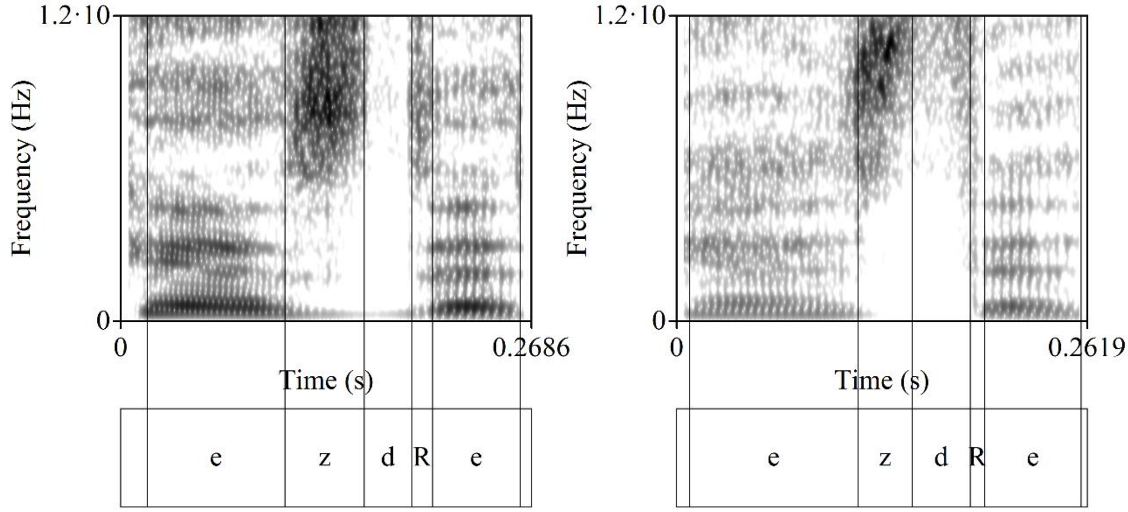

Figure 4

Spectrograms illustrating the pronunciation of *kez**d**et* 'beginning' with a fully voiced /z/ pronounced by Subject 7 (left) and a voiceless /z/ pronounced by Subject 4 (right) ("R" in the annotation stands for release burst)

In intervocalic position, similarly to the absolute final context, both members of the cluster were pronounced; we observed only one deletion: the second consonant of the cluster /kt/ was omitted on one occasion, and the velar stop was pronounced as a fricative rather than a stop. We observed that the velar stop had a fricative or approximant-like realisation with no closure phase and noticeable release burst in 23% of the cases in our data. Vowel length did not differ significantly either for stop clusters or for fricative clusters in intervocalic position (the vowel preceding /kt/ was 81.78 ms long on average, SE = 2.76 vs. /gd/ with a mean 81.12 ms-long vowel, SE = 2.3; $b = -1.16$, $t(5.74) = -0.54$, $p = 0.6$; /st/: 96.04 ms, SE = 3.09 vs. /zd/: 93.5 ms, SE = 1.79; $b = 1.27$, $t(5.05) = 0.35$, $p = 0.73$). The voiceless fricative cluster /st/ turned out to be significantly longer (139.75 ms, SE = 4.42) than its voiced counterpart /zd/ (89.92 ms, SE = 2.04; $b = 24.91$, $t(5.04) = 6.61$, $p = 0.0011$), this was not the case for the stop clusters (/kt/: 112.96 ms, SE = 3.05; /gd/: 108.62 ms, SE = 3.84; $b = -2.2$, $t(5.05) = -0.49$, $p = 0.63$). Due to the length difference in consonants, in /st/−/zd/ clusters, vowel-to-consonant duration ratio is also significantly different and shows the same vowel shortening effects as in the case of utterance-final clusters.

We now turn to the word-final pre-sonorant position. As we mentioned above, sonorant consonants do not trigger RVA in Hungarian, and therefore, in this context we expect full contrast preservation between the voiced−voiceless members of a cluster. Despite the fact that this is a phonetically favourable environment, in 33.8% of all $CC\#C_{sonorant}$ realisations, C_2 was deleted, that is, both members of the word-final cluster were preserved in 66.2% of the cases before a sonorant consonant. It was the stop+stop cluster /gd/ that was reduced most, and it was realised almost fully voiced. Considerable inter-speaker variation can be observed with regard to deletion: for example, Subject 5 had an equal number of simplified and unreduced clusters, while Subject 2 produced hardly any deletions. Our results are as expected: all the phonetic parameters examined including the duration-related parameters show a statistically significant difference between the voiceless−voiced members of a cluster pair. The situation is very similar in the case of reduced clusters, but as we had only three such instances of /st/, no statistically reliable comparison is possible in this context.

We conclude that intervocalic obstruent clusters and word-final pre-sonorant clusters are fully contrastive in Hungarian. The contrast is encoded by phonation itself. Due to aerodynamic reasons, fricatives are more likely to devoice in this phonetically favourable position than stops; therefore, laryngeal contrast is also enforced by segment duration. Let us now turn to the main focus cases of this paper, the voicing-assimilatory contexts.

3.3 Voicing assimilation in three-consonant clusters

According to the traditional generative literature, in pre-obstruent position, Hungarian obstruents neutralise their laryngeal opposition: voicing contrast is suspended in this environment. The inter-obstruent position (our focus of study) is highly unfavourable for contrast preservation (of laryngeal or place features) in the case of obstruent consonants since the all-important transition cues are unavailable. Therefore, it is not surprising that in this environment, we observed a large number of cluster simplifications, namely C_2 deletions. (On cluster simplification in Hungarian see, e.g., Côté 2000 and the references therein.)

3.3.1 Assimilation of voicelessness before /p/

First, we will present the results for those clusters that were followed by /p/ as C_3, and C_2 was preserved so that we can observe the voicing properties of consonant clusters in pre-consonantal position. Then, we will proceed to compare these clusters with those realizations where C_2 was deleted.

The cluster most often simplified before /p/ was /kt/; we were left with only four instances of proper cluster realisations. This allows us to see the tendencies of the devoicing process, but more reliable statistical results would require more data scores for the unreduced cluster groups. There was considerable inter-speaker variation in cluster simplification. Subject 6 deleted C_2 in all the clusters in pre-voiceless obstruent context, Subject 5 deleted C_2 in 87% of the cases while Subjects 2 and 7 hardly simplified any clusters (see Figure 5). Note that it is only in 39.6% of the cases that both members of the word-final cluster were preserved when followed by a voiceless stop in the next word.

Our expectations regarding devoicing assimilation were borne out: all the clusters were realized with around 86–90% devoicing (Figure 6), and we found no statistically significant differences in voicing ratio between the underlyingly voiced vs. voiceless clusters (stop+stop clusters: $b = -1.53$, $t(35) = -0.72$, $p = 0.47$; fricative+stop clusters: $b = 0.32$, $t(35) = 0.2$, $p = 0.83$). We did not find significant differences for any of the duration-related correlates either.

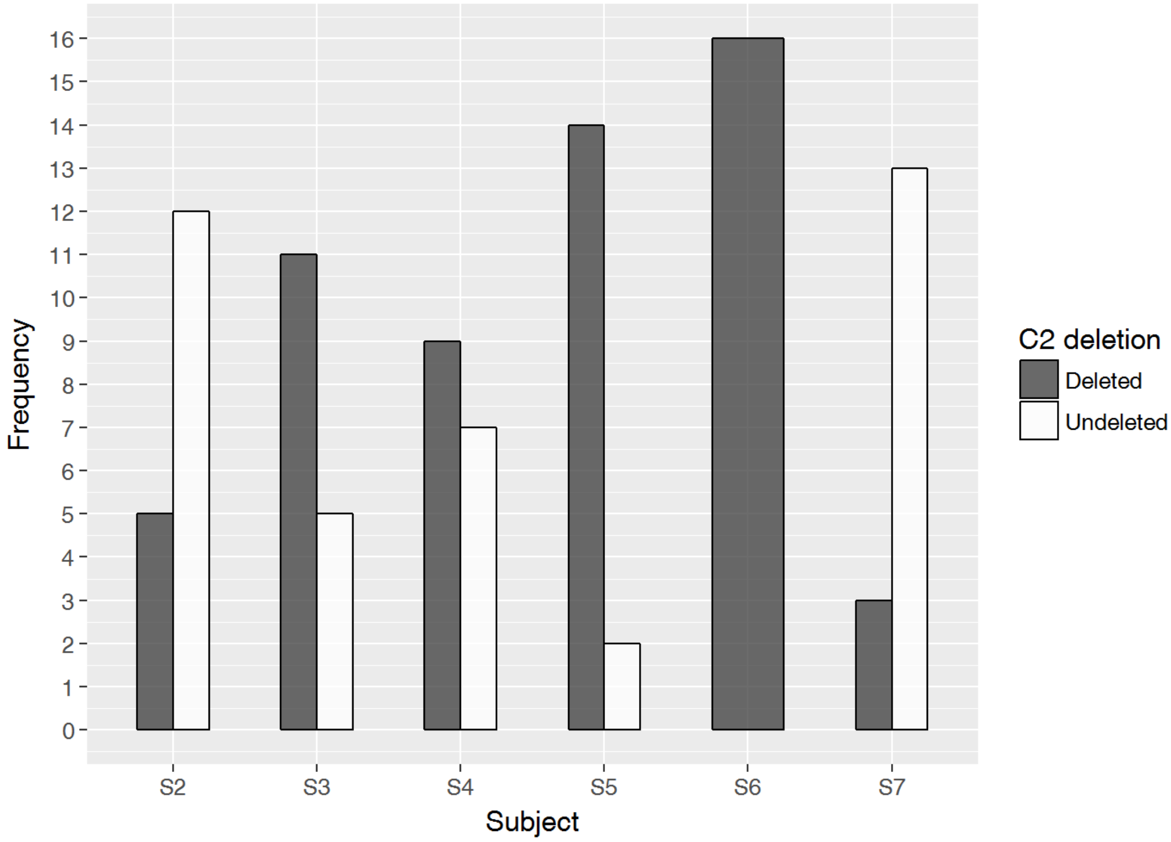

Figure 5

The number of C_2 deletions by subject in word-final utterance-medial position before /p/

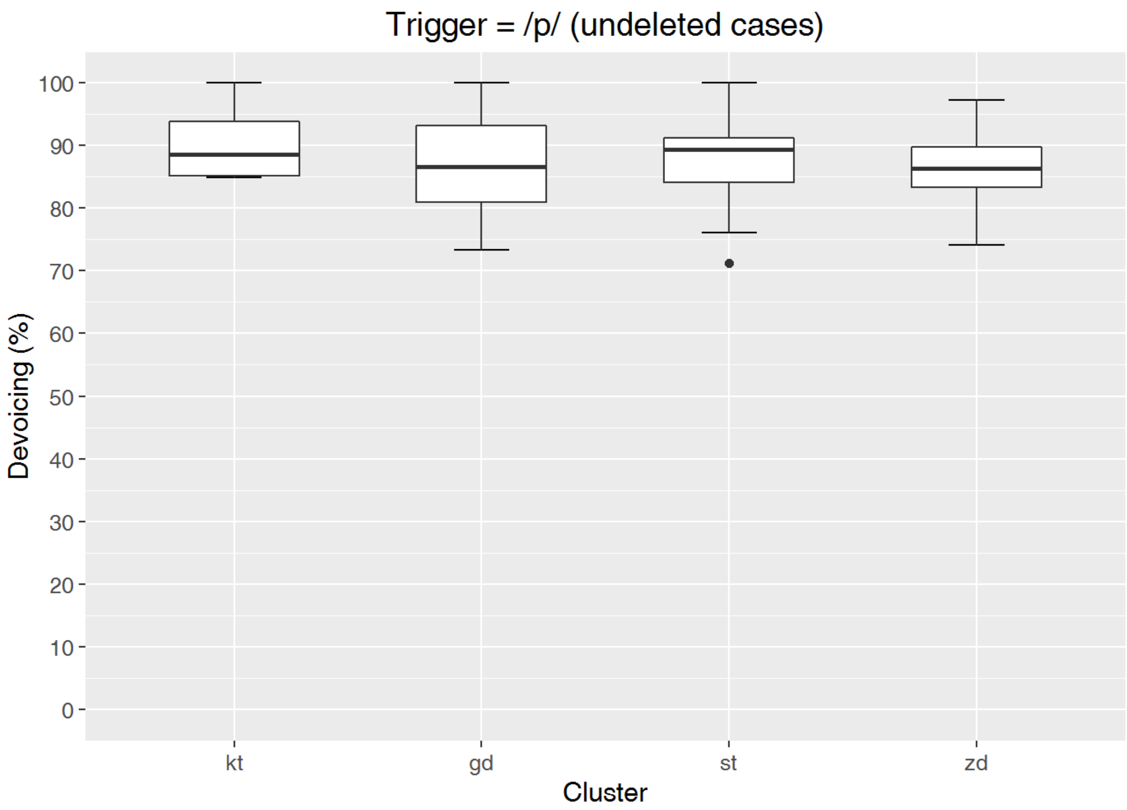

Figure 6

Boxplots showing the ratio of the unvoiced part to total consonant length in Hungarian word-final /kt gd st zd/ clusters followed by /p/ in the next word.

The situation was similar though not identical in the simplified clusters where C_2 was deleted. The simplified stop clusters /kt/ and /gd/ were on the verge of being significantly different in voicing ratio produced (mean devoicing: /kt/: 91.41%, SE = 1.85; /gd/: 79.44%, SE = 9.2; $b = -5.79$, $t(52.91) = -1.96$, $p = 0.054$). The near-significance was due to Subject 6 who realised /gd/ with 66.5% of devoicing on average, that is, with well over 30% of voicing as shown in Figure 7. Note that this is precisely the speaker who systematically, without exceptions, deleted C_2 in this position. One might suspect that due to the testing situation the speaker was hyper-articulating and not implementing otherwise natural assimilation processes. But if this were the case, the speaker would not systematically simplify clusters. We assume instead that this was a consistent strategy by the speaker to implement laryngeal contrast preservation in this highly assimilating context. If we exclude Subject 6 from the statistical analysis, there is no significant difference in voicing ratio between underlyingly voiced and voiceless clusters ($b = -0.75$, $t(36.21) = -0.41$, $p = 0.682$).

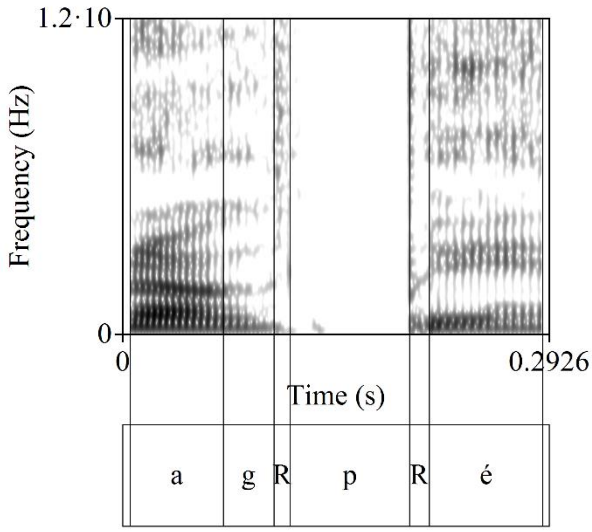

Figure 7

Spectrogram illustrating the deletion of word-final utterance-medial /d/ in *-gd#p-* (*smara**gd p**énzértéke* 'value of emerald') pronounced by Subject 6. Note that /g/ is fully voiced ("R" in the annotation stands for release burst).

The only phonetic parameter that turned out to show a difference between the underlyingly voiced and underlyingly voiceless clusters was the length of the preceding vowel in the case of simplified stop clusters. The vowel before /gd/ was significantly longer (mean: 83.25 ms, SE = 4.27) than before /kt/ (mean: 67.25 ms, SE = 3.28; $b = 7.99$, $t(50.48) = 4.07$, $p < 0.001$), the vowel-to-consonant ratio, however, was not significantly different.

These results imply that obstruent clusters are devoiced and seem to be neutralised for most speakers when followed by a voiceless stop in the next word as claimed by earlier generative analyses, but not for all native speakers. The role of sporadic traces of incomplete neutralisation cannot be disregarded though – especially, that deletion is more frequent in this context than cluster preservation − and should be clarified with perception experiments in the future.

3.3.2 Assimilation of voicing before /b/

We now turn to the voicing environment where all the aerodynamic difficulties mentioned above come into play, which is actually mirrored by the number of deletions we observed. Clusters before /b/ were simplified in 70.5% of the cases by deleting C_2, that is, in only 29.5% of the observations did we get a proper CC#C sequence. The cluster most prone to undergo reduction was /kt/ again: we obtained only two observations with /kt/, both were fully voiced and the velar segment was realised more like an approximant (with a formant structure) rather than a stop with proper closure throughout the segment (see Figure 8).

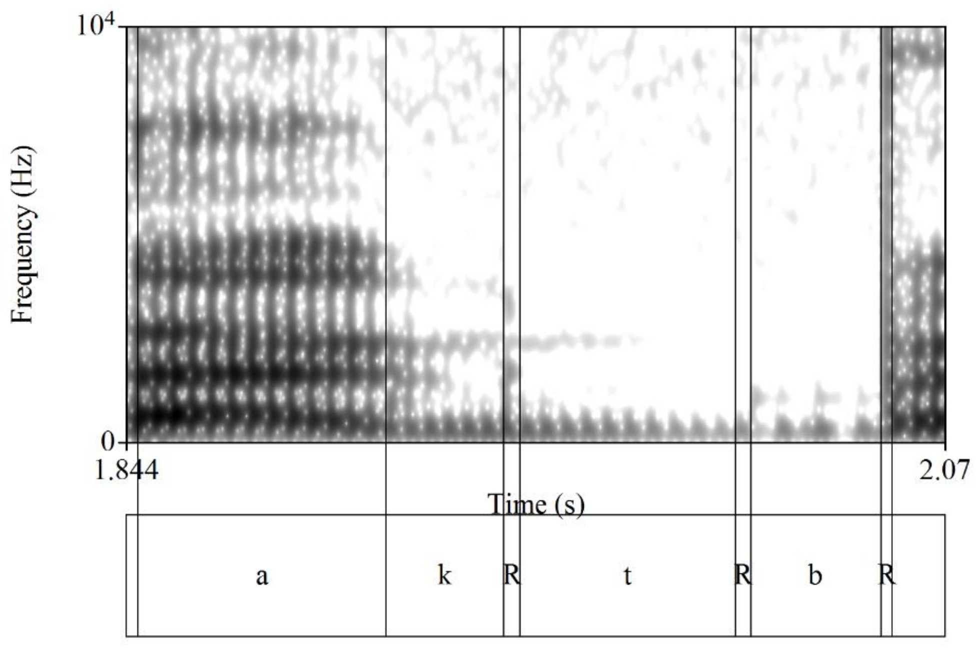

Figure 8

Spectrogram illustrating a voiced approximant-like realisation of /k/ in *-kt#b- (konta**kt b**őrgyulladás* ‘inflammation of the skin due to contact’) pronounced by Subject 3

When stop closure and voicing cannot be maintained for long enough at the same time, several repair strategies can be implemented. Most speakers opted for reducing the cluster (as mentioned above). Another possibility is to maintain the obstruent-like articulation but lose (at least part of) phonation, as in the case of fricatives and fricative clusters (presented below), or to maintain voicing, but "sacrifice" the manner of articulation, as in the present example.

As for the fricatives, the /st/ clusters contained considerably less voicing (mean devoicing: 70.23%, SE = 4.13) compared to /zd/ clusters (mean devoicing: 57.74%, SE = 8.53), results are close to statistically significant (b = 7.62, $t(22.02) = 1.97$, $p = 0.06$). Despite the fact that duration related parameters do not differentiate the underlyingly voiced and voiceless clusters, the duration of the voiced part of the cluster is significantly different: 23.88 ms on average (SE = 4.37) for /st/ and 35.57 ms on average (SE = 7.24) for /zd/ ($b = -7.86$, $t(21.27) = -2.39$, $p = 0.02$). This amount of voicing is more than pure coarticulation, it indicates a planned voicing gesture in both cases, i.e., for underlyingly voiced and for underlyingly voiceless clusters. We might assume that the difference in the amount of voicing is also due to a planned articulatory gesture that aims to prevent a complete neutralisation. Note that voicing generally comes from the left from the preceding vowel/sonorant and dies out during the closure phase of obstruent. We also observed several instances of actual devoicing of the trigger /b/ rather than voicing of the whole cluster which we also saw in the case of simplified clusters as illustrated in Figure 9. This suggests that the articulatory gesture of implementing phonation aims at the whole of the cluster, that is, target and trigger consonants together, and it is often the actual realization of the consonants involved that is "imperfect". This is in line with Markó et al. (2010) who found "progressive voicing assimilation" in CC clusters in 1.3% of the cases they investigated.

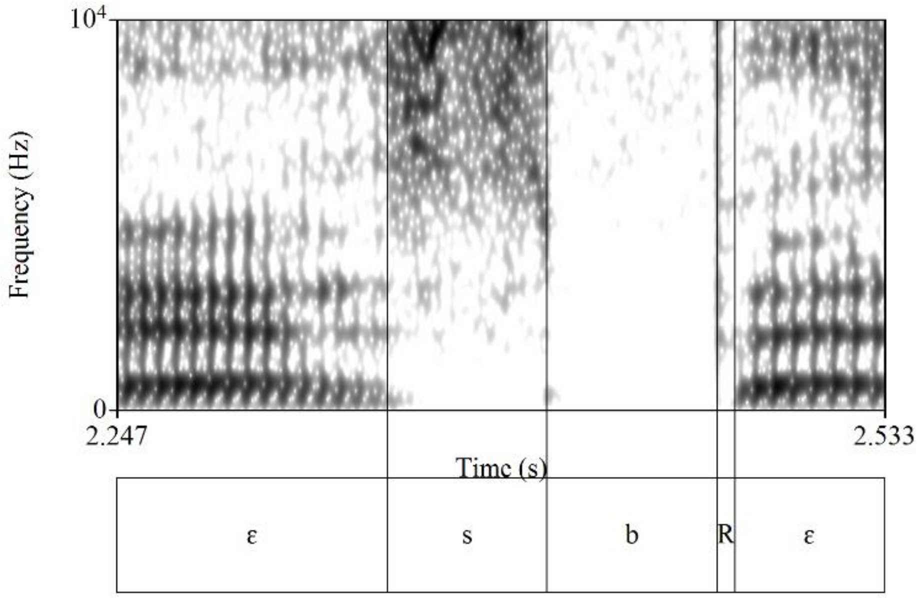

Figure 9

Spectrogram illustrating a simplified /st/ cluster followed by a devoiced /b/ in *-szt#b-* (*kere**szt** **b**erakás* 'decorated with a cross' pronounced by Subject 4 ("R" in the annotation stands for release burst)

Turning to simplified clusters before /b/ now, we found further evidence that fricatives partially preserve laryngeal contrast in this context. In Figure 10 we can see that while stops are generally realised fully voiced (except for a few outliers), fricatives are far from being completely phonated, and although there is considerable variation between the individual realisations, the underlyingly voiceless fricatives resist voicing more than the underlyingly voiced ones, the difference is statistically significant between the fricative pair (/st/: 53.98% mean unvoiced ratio, SE = 6.98; /zd/ 26.56%, SE = 6.83; b = 13.41, $t(59.1)$ = 3.51, p < 0.001) while it is negligible in the case of stops.

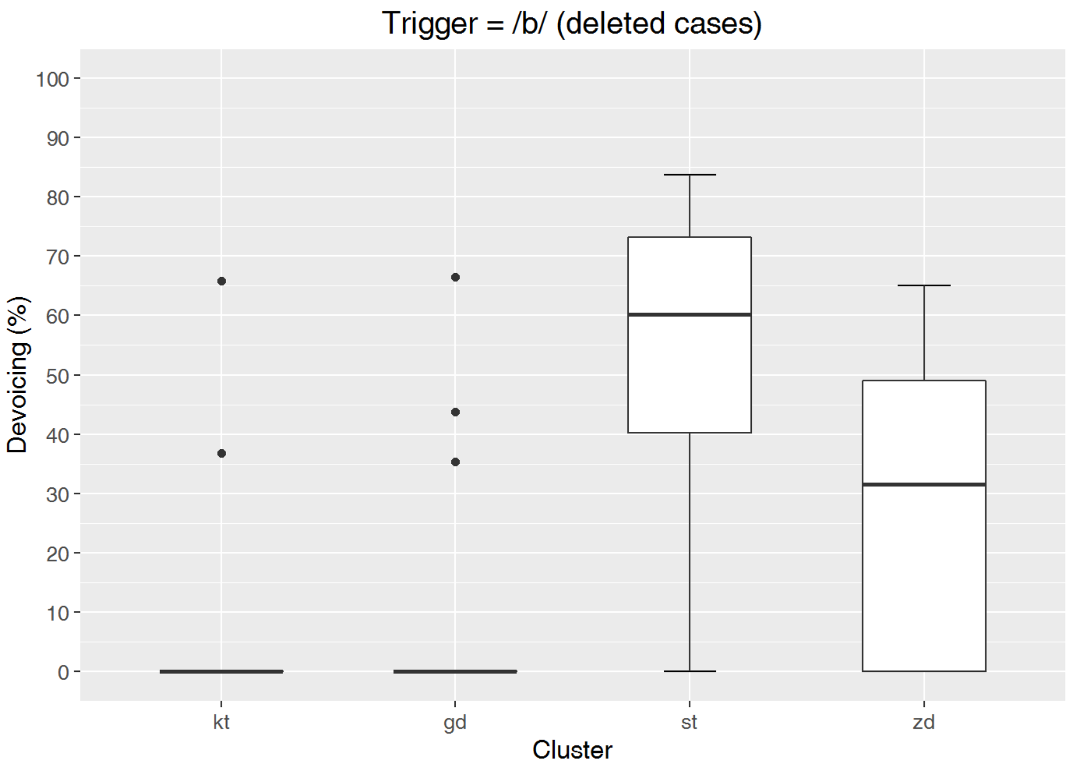

Figure 10

Boxplots showing the ratio of the unvoiced part to total consonant length in Hungarian word-final simplified /kt gd st zd/ clusters followed by /b/ in the next word

Voicing duration was also very different between the members of the fricatives: reduced /st/ was voiced for 24.04 ms on average (SE = 3.56), whereas the mean voicing duration of reduced /zd/ was 40.21 ms (SE = 4.91; $b = -7.72$, $t(60.32) = -2.38$, $p = 0.02$). Note how similar these values are to those in the case of non-reduced clusters above: in the case of reduced /st/ clusters vs. preserved /st/ clusters $b = -0.97$, $t(42.85) = -0.3$, $p = 0.765$; reduced /zd/ vs. preserved /zd/ $b = 6.01$, $t(44.55) = 1.83$, $p = 0.073$. Thus, simplified and fully preserved fricative clusters contain a voiced part of very similar length. Table 1 shows that none of the phonation-related parameters differentiate full clusters from reduced clusters. In our earlier study on singletons, fricatives had a longer voiced part in this context compared to fricative clusters (both reduced and non-reduced) in the present study, and this difference turned out to be statistically significant ($b = 4.87$, $t(90.32) = 2.53$, $p = 0.013$). Table 1 also shows that phonation-related parameters differentiate clusters (both reduced and non-reduced) from singleton fricatives before /b/.

Table 1: Phonation-related correlates of voicing in word-final fricative clusters reduced versus non-reduced clusters and clusters (both reduced and non-reduced) versus singletons before /b/. * stands for a statistically significant difference ($p < 0.05$).

Acoustic correlate	**reduced /st/– non-reduced /st/**	**reduced /zd/– non-reduced /zd/**	**cluster /st/–/s/**	**cluster /zd/– /z/**
devoicing ratio			*	*
voicing duration			*	*

We think that these results on fricative voicing before /b/ are an indication of (i) a planned but unrealised cluster the voicing of which is implemented in the preserved C_1. The fact that it is different from planned singleton fricatives also re-enforces cluster perception despite the fact that only one consonant is realised. (ii) Voicing induced by a following voiced stop is not coarticulatory, it is categorical, but it is not neutralising either. Fricatives do partially preserve their underlying laryngeal characteristics in this highly assimilating context as well. The duration-related parameters do not differentiate the voiced−voiceless members of reduced fricative clusters in the present environment. Stop clusters, again, show a "pre-fortis clipping" effect: the vowel is considerably longer before a reduced /gd/ cluster (mean: 82 ms, SE = 2.69) than before a reduced /kt/ cluster (mean: 69.14 ms, SE = 3.42; $b = 5.24$, $t(59.93) = 2.57$, $p = 0.012$).

A clear asymmetry is observed in the voicing and devoicing of word-final obstruent clusters. RVA induced by a voiceless stop is almost completely neutralising, we found only sporadic traces of the underlying laryngeal specification of the undergoer consonants cued by duration-related features. RVA induced by a voiced stop, on the other hand, is very different: neutralisation in this contexts is systematically incomplete. In the case of stops the contrast in phonetic voicing seems to be completely lost, but again durational effects crop up to prevent complete neutralisation. Fricatives are only partially neutralised with regard to phonation, as the underlyingly voiced clusters (independently of being reduced or not) are more voiced than the underlyingly voiceless ones. Examining C#C clusters in different prosodic contexts Mády & Bárkányi (2015) also observed that voiced fricatives before an accentual phrase boundary with no intervening pause were more voiced when followed by a voiced obstruent than voiceless fricatives, while this was not the case for stops. It is not surprising that stops and fricatives show an asymmetry in their behaviour with respect to voicing and voicing assimilation in many languages. The two groups of obstruents encode voicing in a very different way: while voicing is an inherent property of voiced fricatives, and so is cued during the fricative interval itself, voicing in stops is cued in the transition phases into and out of the closure. As for the articulatory side of voicing and devoicing, voicing is more difficult to initiate and maintain in fricatives than in stops (as mentioned above).

We conclude that voicing assimilation in Hungarian CC#C obstruent clusters is clearly phonological and categorical but incomplete, preserving traces of the underlying laryngeal specifications of the clusters. Partial contrast preservation is more robust in the voicing context (i.e., before a voiced obstruent) and in clusters containing fricatives. Since perceptual cut-off values with regard to voicing in assimilatory contexts have not been specified for Hungarian obstruents so far, we cannot say without further investigation to what extent these acoustic differences are relevant perceptually, that is, to what extent the attested incomplete neutralisation is identified by speakers.

4 Conclusions

In the present paper we examined the various phonetic correlates of voicing in Hungarian three-consonant clusters (CC#C) in various voicing-neutralising contexts as worded in our first research question. Our results show that neither the triggers (voiced obstruent vs. voiceless obstruent) nor the targets (stop clusters vs. fricative clusters) behave in the same way with regard to regressive voicing assimilation. We have discussed phonation-related and duration-related parameters in detail. It has been shown that fricatives are more likely to lose phonation than stops. Contrary to earlier claims, we have demonstrated that duration-related effects are present in Hungarian, too.

The second research question was whether regressive voicing assimilation in Hungarian is neutralising or incomplete. We have shown that the picture is far from being homogeneous. While devoicing is almost completely neutralising with only sporadic traces of contrast preservation, voicing assimilation is incomplete especially in the case of fricatives: underlyingly voiceless fricatives are systematically less voiced than underlyingly voiced fricatives when followed by the voiced stop /b/.

Our third research question aimed to focus on the difference between fully preserved clusters and simplified clusters. Interestingly, reduced and full-fledged clusters do not differ either in voicing ratio or in voicing duration, while the voicing properties of clusters and singleton consonants are different. This is a strong piece of evidence that regressive voicing assimilation in Hungarian is not a low-level coarticulatory mechanism but a planned phonological process despite the fact that it is often non-neutralising.

The fourth research question concerned the implementation of voicing contrast in utterance-final clusters. Again, the picture is asymmetrical: utterance-final stops have been found to be fully contrast preserving, while fricatives to be realised as voiceless; however, the phonological contrast has been shown to be preserved by duration-related acoustic cues.

Acknowledgements

This research was supported by grant 104897 by the Hungarian National Research Fund. We would like to thank Péter Siptár for his comments on earlier versions of this paper, as well as two anonymous reviewers for their valuable comments.

References

Bárkányi, Zsuzsanna & Katalin Mády. 2012. The perception of voicing in fricatives. Paper presented at the 9th Old World Conference in Phonology, Berlin. January.

Bárkányi, Zsuzsanna & Zoltán G. Kiss. 2015. Why do sonorants not voice in Hungarian? And why do they voice in Slovak? In Katalin É. Kiss, Balázs Surányi & Éva Dékány (eds.) *Approaches to Hungarian 14*. Amsterdam: John Benjamins. 65–94.

Bárkányi, Zsuzsanna & Zoltán Kiss 2007. A phonetically-based approach to the phonology of [v]: A case study from Hungarian and Slovak. Paper presented at the 4th Old World Conference in Phonology. Rhodes, Greece. 18–21 January.

Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1), 1–48. doi:10.18637/jss.v067.i01.

Boersma, Paul & David Weenink. 2012. Praat: Doing phonetics by computer. (Version 5.3.12) [Computer program]. Retrieved from http://www.praat.org/

Burton, Martha & Karen Robblee. 1997. A phonetic analysis of voicing assimilation in Russian. *Journal of Phonetics* 25. 97–114.

Charles-Luce, Jan. 1985. Word-final devoicing in German: Effects of phonetic and sentential contexts. *Journal of Phonetics* 13. 309–324.

Charles-Luce, Jan. 1993. The effects of semantic context on voicing neutralization. *Phonetica* 50. 28–43.

Côté, Marie-Hélène. 2000. *Consonant cluster phonotactics: A perceptual approach*. Doctoral dissertation. MIT.

Dinnsen, Daniel A. & Jan Charles-Luce. 1984. Phonological neutralization, phonetic implementation, and individual differences. *Journal of Phonetics* 12. 49–60.

Field, Andy, Jeremy Miles & Zoe Field. 2012. *Discovering statistics using R.* London: Sage.

Fourakis, Marios & Gregory Iverson. 1984. On the ‘incomplete neutralization’ of German final obstruents. *Phonetica* 41. 140–149.

Gráczi, Tekla Etelka. 2010. A spiránsok zöngésségi oppozíciójának néhány jellemzője [Some characteristics of the voicing contrast of fricatives]. In Mária Gósy (ed.) *Beszédkutatás 2010* [*Speech research 2010*]. Budapest: MTA Nyelvtudományi Intézet.42–56.

Greenberg, Joseph H. 1970. Some generalizations concerning glottalic consonants, especially implosives. *International Journal of American Linguistics* 36(2). 123–145.

Hallé, Pierre A. & Martine Adda-Decker. 2011. Voice assimilation in French obstruents: Categorical or gradient? In John Goldsmith, Elizabeth Hume & Leo Wetzels (eds.) *Tone and features: A festschrift for Nick Clements.* Berlin & New York: Mouten De Gruyter. 149–175.

Hanson, Helen M., Richard S. McGowan, Kenneth N. Stevens and Robert E. Beaudoin. 1999. Development of rules for controlling the HLsyn speech synthesizer. In *Acoustics, Speech, and Signal Processing. Proceedings*, vol. 1. IEEE International Conference on Acoustics, Speech, and Signal Processing. 85–88.

Harris, John. 1994. *English sound structure*. Cambridge, MA & Oxford: Blackwell.

Hayes, Bruce 1999. Phonetically driven phonology: The role of Optimality Theory and inductive grounding. In Michael Darnell, Edith Moravcsik, Michael Noonan, Frederick Newmeyer & Kathleen Wheatly (eds.) *Functionalism and formalism in linguistics.* Amsterdam & Philadelphia: John Benjamins. vol. 1. 243–285.

Hayes, Bruce & Donca Steriade. 2004. The phonetic bases of phonological markedness. In Bruce Hayes, Robert M. Kirchner & Donca Steriade (eds.) *Phonetically based phonology.* Cambridge: Cambridge University Press.1–33.

HLSyn (High Level Speech Synthesizer) Reference Manual.1999. Sensimetrics Corporation.

Jackson, Philip J.B.& Christine H. Shadle. 2000. Frication noise modulated by voicing, as revealed by pitch-scaled decomposition. *Journal of the Acoustical Society of America* 108(4). 1421–1434.

Jansen, Wouter. 2004. *Laryngeal contrast and phonetic voicing: A laboratory phonology approach to English, Hungarian, and Dutch*. Doctoral dissertation, Rijksuniversiteit Groningen.

Jassem, Wiktor & Lutoslawa Richter. 1989. Neutralisation of voicing in Polish obstruents. *Journal of Phonetics* 17. 317–325.

Javkin, Hector R. 1976. The perceptual basis of vowel duration differences associated with the voiced/voiceless distinction. *Report of the Phonology Laboratory, UC Berkeley* 1. 78–92.

Kahlen-Halstenbach, Birthe. 1990. Zur psychologischen Realität der Auslautverhartung im Deutschen. *Zeitschrift für phonetische Sprachwissenschaft und Kommunikationsforschung* 43. 645-655.

Kingston, John & Randy L. Diehl. 1994. Phonetic knowledge. *Language* 70. 419–454.

Kluender, Keith R., Randy L. Diehl & Beverly A. Wright. 1988. Vowel length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics* 16. 153–169.

Kuznetsova, Alexandra, Per Bruun Brockhoff & Rune Haubo Bojesen Christensen (2016). lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-32. https://CRAN.R-project.org/package=lmerTest

Mády, Katalin & Zsuzsanna Bárkányi. 2015. Voicing assimilation at accentual phrase boundaries in Hungarian. In Maria Wolters, Judy Livingstone, Bernie Beattie, Rachel Smith, Mike MacMahon, Jane Stuart-Smith & Jim Scobbie (eds.). *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow. 1–4.

Markó, Alexandra, Tekla Etelka Gráczi & Judit Bóna. 2010. The realisation of voicing assimilation rules in Hungarian spontaneous and read speech: Case studies. *Acta Linguistica Hungarica* 57. 210–238.

Massaro, Dominic W. & Michael M. Cohen. 1983. Consonant/vowel ratio: An improbable cue in speech perception. *Perception and Psychophysics* 33. 502–505.

Myers, Scott. 2012. Final devoicing: Production and perception studies. In Tony Borowsky, Shigeto Kawahara & Mariko Sugahara (eds.) *Prosody matters: Essays in honor of Elisabeth Selkirk*. London: Equinox Press. 148–180.

O'Dell, Michael & Port, Robert F. 1983. Discrimination of word final voicing in German. *Journal of the Acoustical Society of America* 73. Supplement 1.

Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In Peter MacNeilage (ed.) *The production of speech.* New York: Springer. 189–216.

Parker, Ellen M., Randy L. Diehl & Keith R. Kluender. 1986. Trading relations in speech and non-speech. *Perception and Psychophysics* 39. 129–142.

Port, Robert F. & Jonathan Dalby. 1982. C/V ratio as a cue for voicing in English. *Perception and Psychophysics* 2. 141–152.

Port, Robert F. & Adam P. Leary. 2005. Against formal phonology. *Language* 81. 927–964.

Port, Robert F., Fares Mitleb & Michael O'Dell. 1981. Neutralization of obstruent voicing in German is incomplete. *The Journal of the Acoustical Society of America* 70. Supplement 1.

R Development Core Team (2008). R: A language and environment for statistical computing. Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org.

Schmidt, Lauren B. & Erik W. Willis. 2011. Systematic investigation of voicing assimilation of Spanish /s/ in Mexico City. In Scott M. Alvord (ed.) *Selected proceedings of the 5th conference on Laboratory Approaches to Romance Phonology*. Somerville, MA: Cascadilla Proceedings Project. 1–20.

Siptár, Péter & Miklós Törkenczy. 2000. *The phonology of Hungarian*. Oxford: Oxford University Press.

Slowiaczek, Louisa M. & Daniel A. Dinnsen. 1985. On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics* 13. 325–341.

Slowiaczek, Louisa M. & Helena J. Szymanska. 1989. Perception of word-final devoicing in Polish. *Journal of Phonetics* 17. 205–212.

Steriade, Donca. 1997. Phonetics in phonology: The case of laryngeal neutralization. Manuscript. University of California Los Angeles.

Steriade, Donca. 2008. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In Sharon Inkelas & Kristin Hanson (eds.) *The nature of the word*. Cambridge, MA: MIT Press. 151–180.

Stevens, Kenneth N. 1998. *Acoustic phonetics*. Cambridge MA: MIT Press.

Strycharczuk, Patrycja. 2012. *Phonetics–phonology interactions in pre-sonorant voicing*. Doctoral dissertation, University of Manchester.

Strycharczuk, Patrycja & Ellen Simon. 2013. Obstruent voice before sonorants. The case of West-Flemish. *Natural Language and Linguistic Theory.* Available online, DOI 10.1007/s11049-013-9189-5.

Vago, Robert M. 1980. *The Sound Pattern of Hungarian*. Washington DC: Georgetown University Press.

Wells, John Christopher. 2000. *Longman pronunciation dictionary*. Harlow: Longman/Pearson Education.

Wright, Richard 2001. Perceptual cues in contrast maintenance. In Elizabeth Hume & Keith Johnson (eds.) *The role of speech perception in phonology*. San Diego: Academic Press. 251–277.

Wright, Richard 2004. A review of perceptual cues and cue robustness. In Bruce Hayes, Robert M. Kirchner & Donca Steriade (eds.) *Phonetically based phonology.* Cambridge: Cambridge University Press. 34–57.

Zsigri, Gyula. 1994. *Magyar mássalhangzószabályok.* [Rules for Hungarian consonants]. Doctoral dissertation. József Attila University (JATE), Szeged. (Published by JATEPress, Szeged, 2006.).

Appendix: Test sentences

The words tested are indicated in boldface.

1. Tegnap a **smaragd már** nem volt a kirakatban.
2. A **smaragd látványa** lenyűgözte a vásárlókat.
3. A **smaragd elrablásával** foglalkozott az egész sajtó.
4. Ezután a **smaragd pénzértéke** már erősen csökken.
5. A fűben **smaragd béka** ült csöndesen.
6. Lángolva szikrázott a türkiz és a **smaragd**.
7. **Magda** gyerekkori barátnőm.
8. Érintésre hatnak a **kontakt mérgek**.
9. A szemészetet a **kontakt lencse** forradalmasította.
10. A fizikatanár **kontakt elektromosságot** idézett elő.
11. A műtét **kontakt próbával** kezdődik.
12. A **kontakt bőrgyulladást** vegyi anyagok okozzák.
13. Sok erőfeszítés árán megvan a **kontakt**.
14. Az **akta** készítője minden részletre kitért.
15. Egy **gerezd mangó** díszíti a tálat.
16. A pici **gerezd látható** szélét penész borítja.
17. Teszünk bele egy **gerezd elkevert** fokhagymát.
18. A négy **gerezd puha** narancs a recept titka.
19. Mindegyik **gerezd beleillik** a mélyedésbe.
20. Lógott a fáról, mint egy nagy **gerezd**.
21. A **kezdet** a nehéz ilyen esetekben.
22. A régi **kereszt mása** került oda.
23. Ő a **kereszt legnagyobb** ellensége.
24. Az ott lévő **kereszt eredete** bizonytalan.
25. Minden **kereszt pirosra** volt festve a templomnál.
26. A csipkét **kereszt berakás** díszíti.
27. Az érmen látható a császári **kereszt**.
28. Az agronómus **keresztez** és génsebészkedik.