2 Is /v/ different?

Zsuzsanna Bárkányi and Zoltán Kiss
Eötvös Loránd University

Abstract

This article presents two acoustic experiments in which we demonstrate that voiced fricatives in utterance-final position in Hungarian are realized unphonated and the preceding vowel plays a crucial role in partial contrast preservation between the phonologically voiced and phonologically voiceless fricative. We also show that despite their very similar behavior in final position, /v/'s articulatory targets considerably differ from those of voiced sibilants in that /v/ is a passively voiced narrow approximant, while /z/ and /ʒ/ are true voiced obstruents. This phonetic difference can explain the differences in their phonological behavior, especially regarding their role in voicing assimilation.

1 Introduction

In recent studies (Kiss and Bárkányi 2006; Bárkányi and Kiss 2010) it has been demonstrated that the double-faced phonological behavior of /v/ in Hungarian can be explained in a model based on the phonetic properties of this segment and its linear context. The analysis is based on the claim that the phonetic targets of /v/ are contradictory on aerodynamic grounds (Ohala 1983) and can only be maintained in phonetically favorable positions. In this paper we explore the phonetic properties of the other two voiced fricatives of Hungarian, /z/ and /ʒ/, and discuss to what extent these properties can explain the differences in their phonology, especially, the different behavior they display compared to /v/ in voicing assimilation.

It is well known that for the articulatory system to target voicing and friction (turbulent noise) at the same time, an uneasy balance needs to be maintained. High-amplitude turbulent noise requires a relatively high volume velocity of the airflow as it blows out from a constriction. In order to achieve this condition, the glottis is widely abducted so that the intraoral pressure equals or approaches the subglottal pressure, and the oral cavity is relatively constricted, creating a pressure drop across the supraglottal constriction (see, among others, Shadle 1985; Stevens et al. 1992; Stevens 1998; Jesus 2001; Johnson 2003: 120–133; Krane 2005).

In contrast, for vocal fold vibration to be initiated the vocal folds must be set into modal phonation mode: they must be adducted; subglottal air pressure must build up below the adducted vocal folds, forcing the lower part of the folds to blow apart (with the consequence that subglottal pressure drops close to zero relative to atmospheric pressure); and the negative pressure that occurs as air passes between the folds must suck the elastic folds together again (Bernoulli effect). If the pressure above the folds builds up so that the pressure difference drops across the glottis, phonation ceases. To overcome devoicing a number of articulatory gestures, which aim at preserving a transglottal difference of pressure, need to be implemented to enlarge the oral cavity volume, e.g., raising the soft palate, advancing the tongue root so that there is an outward movement of the neck surfaces, lowering the larynx, expanding the pharyngeal volume, decreasing the stiffness of the vocal tract walls (reducing vocal tract compliance), or a combination of these gestures (see Stevens 1998: 465–486). These gestures can only be executed within certain limitations, which might have phonological consequences. Bárkányi and Kiss (2010) argue that in certain positions (which they call 'phonetically favorable contexts'), /v/ emerges as a passively voiced narrow approximant (following Padgett 2002) — in this context both phonetic targets of /v/ are maintained (voicing and friction). In other, less favorable positions, /v/ loses one of its articulatory targets, hence either voicing or turbulent noise is preserved, but not both at the same time. As a result of this, two realizations are possible: when /v/ devoices, it becomes a strongly fricated, noisy sound (narrow constriction and wide abduction of the vocal folds) — it is this realization that is implemented in Hungarian, for instance. When /v/'s voicing target is kept, it loses much of its friction (wider constriction), implemented in Slovak, for instance.

The above mentioned conflicting aerodynamic requirements and complex articulatory gestures are expected to hold for the other voiced fricatives as well, not only /v/. Therefore, we hypothesize that /z/ and /ʒ/ in Hungarian in phonetically unfavorable positions are also likely to devoice. One such environment is the utterance-final position. Note that Hungarian is not a final-devoicing language, i.e. there are minimal pairs which contrast due to the the voicing of the word-final obstruents: *mé*[z] 'honey' – *mé*[s] 'lime', *lá*[b] 'foot' – *lá*[p] 'marshland'. The phonetic and phonological literature on obstruents in Hungarian does not mention that this contrast is implemented by other parameters than phonation itself.

In Section 3 we present two acoustic experiments, which investigate whether /z/ and /ʒ/ indeed devoice in utterance-final position; we will pay special attention to other phonetic parameters which make the (partial) preservation of the contrast possible in this position. Section 4 provides a detailed discussion of our results and in Section 5 we discuss in what respects /v/ differs from the voiced sibilants and whether these phonetic differences can lay the basis of the differences in their phonological behavior.

2 The phonology of voiced sibilants in Hungarian

It is well-known in the phonological literature of Hungarian (see Siptár—Törkenczy 2000, for instance) that /v/ shows a two-fold patterning in voicing assimilation as well as in its phonotactic patterning. Namely, it undergoes voicing assimilation but does not trigger it. As for its static distribution, word-initially it patterns with sonorants, i.e., it can be the second member of a cluster (*tviszt* 'twist'), while word-finally it patterns with obstruents as it can stand after a sonorant (*terv* 'plan'). Voiced sibilants, on the other hand, are true obstruents in the sense that they both trigger and target regressive voicing assimilation just like any other obstruent (1). As for their distribution, they can stand after a sonorant word-finally: e.g., *torz* 'distorted', but they cannot form the second member of an initial cluster like /v/.

(1)	a.	/zt/ → [st]:	e.g., *torz-tól*	'distorted-abl.'	(vs. *torzul* 'to become distorted' [z])	
		/ʒh/ → [ʃh]:	e.g., *bézs-hez*	'beige-all.'	(vs. *bézsen* 'beige-super.' [ʒ])	
		/vt/ → [ft]:	e.g., *sav-tól*	'acid-abl.'	(vs. *savak* 'acid-pl.' [ʋ])	
	b.	/pz/ → [bz]:		e.g., *gép zaj*	'machine noise'	
		/tʒ/ → [dʒ]:		e.g., *két zsák*	'two sacks'	
		/kv/ → [kʋ][1]	(*[gʋ]):	e.g., *kék vár*	'blue castle'	

3 The acoustic properties of voiced sibilants

In this section, we turn our attention to the acoustic properties of utterance-final /z/ and /ʒ/. Experiment 1 focuses on the contrast between /s/ and /z/ in utterance-final position,[2] and in Experiment 2 the acoustic properties of /ʃ/ and /ʒ/ are examined in the same position.

[1] This is the case in Standard Hungarian, the focus of the present paper; in Western dialects, /v/ triggers voicing assimilation (see Kiss–Bárkányi 2006, 182).

[2] Experiment 1 was carried out in collaboration with Katalin Mády, and was presented at the Beszédkutatás 2009 conference, Budapest.

3.1 Experiment 1: Methods

Six women and one man aged 23–38 years participated in the experiment, all were speakers of Standard Hungarian living in Budapest and reported no speaking or hearing impairment. The material was recorded in a sound-proof cabin with a SonyECM-MS907 microphone through an M-AudioMobilePreUSB preamplifier connected to a laptop computer using the software SpeechRecorder. The acoustic analysis was carried out in Praat 5.1.07 (Boersma and Weenink 2009). The data recorded consisted of 11–13-syllable-long sentences and a text. The target sequences were -*ész* [eːs] and -*éz* [eːz] in one- and two-syllable words in utterance-final position, and as a base line, some of the test words were suffixed with a vowel-initial suffix, so they appeared in a sentence-medial intervocalic position. The experiment also contained a text in which test words appeared in word-final but sentence-medial position followed by a vowel-initial word. Some of the test words formed minimal pairs like *méz* 'honey' – *mész* 'lime', others did not, like *vész* 'peril' – *géz* 'gauze' (2). There were 38 test sentences and 6 repetitions, which gave 228 tokens per subject.

(2) a. *A fehér asztalon áll egy bögre* ***méz****.*
'There is a mug of honey on the white table.'
b. *A hátsó udvaron van egy talicska* ***mész****.*
'There is a barrow of lime in the backyard.'
c. *A beteg karján félrecsúszott a* ***géz****.*
'The gause slid aside on the patient's arm.'
d. *A védők feje fölül elmúlt a* ***vész****.*
'Peril is gone from above the defenders' heads.'

The experiment aimed to measure the following phonetic parameters:

(3) a. voicing in the fricative
b. duration of the fricative
c. duration of the preceding vowel

Voicing was measured on the basis of periodicity in the waveform, f_0 in the spectrogram, the presence/absence of voice striations in the spectrogram and Praat's voice report ("unvoiced frames percentage") manually checked on the basis of the spectrogram and oscillogram. We used Praat's default settings (pitch range: 75 Hz–500 Hz, and with the following advanced pulses settings, maximum period factor: 1.3, maximum amplitude factor: 1.6, pitch setting was optimized for voice analysis — see the Praat manual). Segmentation was carried out manually employing the following method. The fricative interval started where the preceding vowel's formants ceased, the spectrum became noisy, and the intensity level dropped. The end of the fricative was marked at the point where the noisy spectrum ceased, the intensity sharply dropped, and silence ensued.

3.2 Experiment 1 : Results

Figure 1 exhibits the boxplots of unvoiced frames (%) across subjects for /s/ and /z/ in utterance-final and word-medial position, Figure 2 illustrates the realization of *méz* and *mész* for Subject 5.

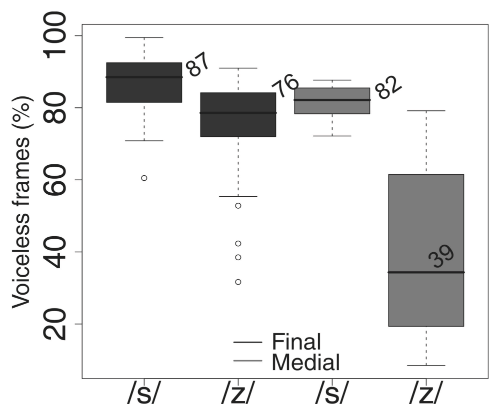

Figure 1: *Boxplots (with outliers) of unvoiced frames of /s/ and /z/ in utterance-final and word-medial position*

According to Figure 1, Hungarian /z/ is realized with over 70% of unvoiced frames in utterance-final position. The difference between utterance-final /s/ and /z/ is significant ($p < 0.001$). We can see that /z/ can devoice even in word-medial intervocalic position. This is in accordance with Gráczi (2010) who examined the voicing properties of Hungarian fricatives with the help of nonsense words of the form *la*C*al* and found that /z/ in this position was unvoiced in 42.1%. There is no statistically significant difference between word-medial /s/ and word-final /z/.

We must mention that considerable individual differences were observed in the voicing of the alveolar fricatives in both word-final and word-medial position. It is also noteworthy that one of our subjects produced /z/ in utterance-final position with more unvoiced frames than /s/ in the same position.

We also observed that the consonant that preceded *é* did not influence the voicing of the word-final fricative. Similarly, stress (whether the target syllable appears in a stressed or unstressed position), or the existence of minimal pairs (whether the

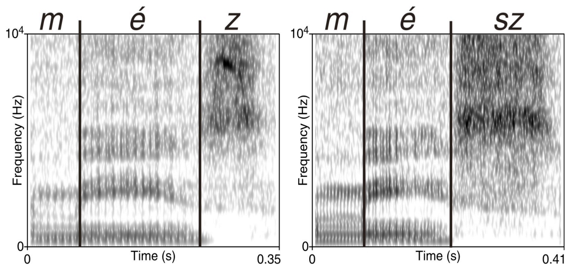

Figure 2: *The realization of* méz *and* mész *for Subject 5*

fricative under scrutiny appeared in a *méz–mész* or *géz–vész* type word pair) did not influence the voicing of the final fricative.

Figure 3 shows the boxplots of the duration of the fricative in utterance-final position, while Figure 4 demonstrates the duration of the preceding vowel.

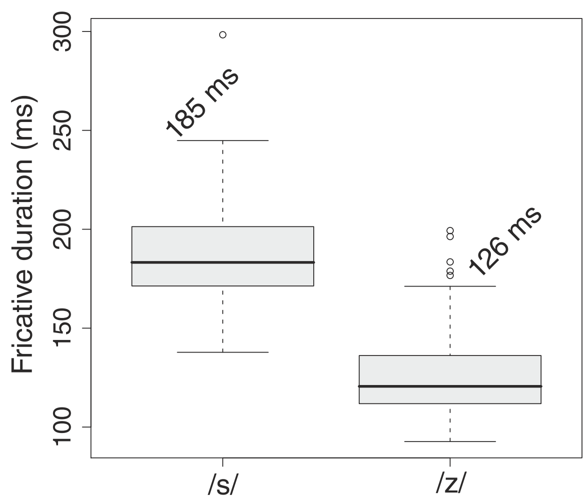

Figure 3: *Boxplots (with outliers) of duration of /z/ and /s/ in utterance final position*

As expected, /s/ is always realized considerably longer than /z/, the differences are statistically significant ($p < 0.001$), the differences are consistently significant for each subject individually as well. As for the vowel, it is always realized longer before the phonologically voiced fricative, the difference is statistically significant across subjects and for each subject as well ($p < 0.001$).

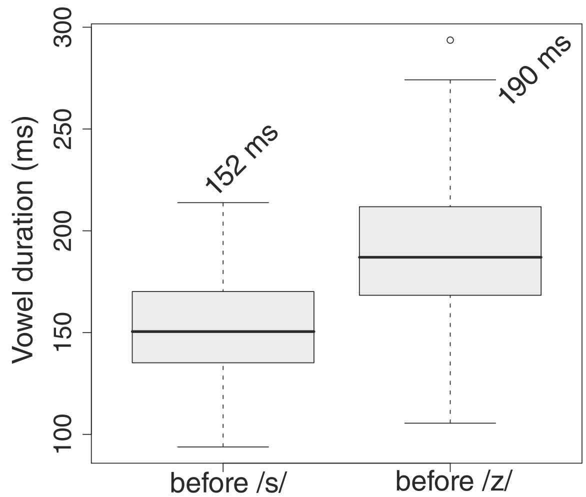

Figure 4: *Boxplots (with outliers) of the duration of the vowel preceding the utterance-final fricative*

In sentence-medial word-final position (in the text) the difference in voicing is statistically significant ($p = 0.047$), standard deviation is considerably larger, especially for some speakers (see Subjects 3, 6 and 7) as displayed in Figure 5.

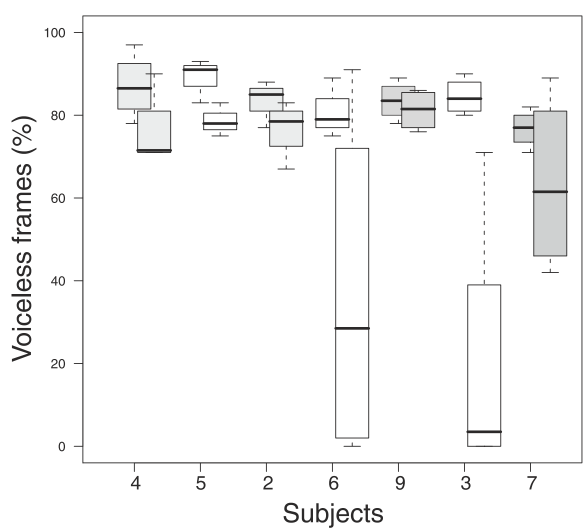

Figure 5: *Boxplots of unvoiced frames for /s/ and /z/ in text, in word-final sentence-medial position, by subjects*

As expected, /z/ in this position is also significantly shorter ($p = 0.0012$); however, vowel length is not significantly different in the two contexts ($p = 0.29$). Figure 6 shows the boxplots for vowel length by subject in word-final sentence-medial position. This is an important difference between sentence-medial and utterance-final position discussed in more detail in Section 4.

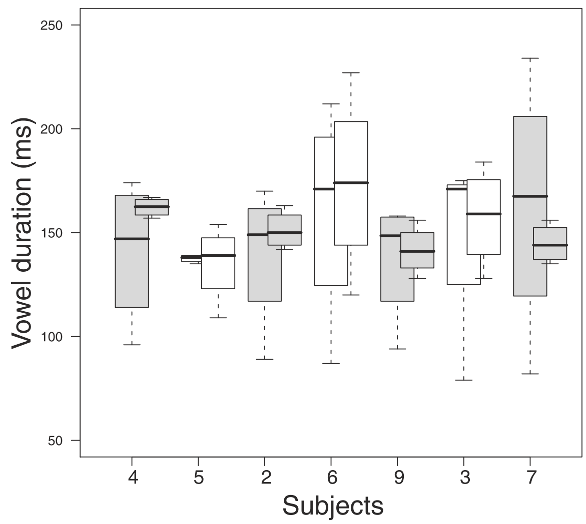

Figure 6: *Boxplots of the duration of the vowel before /s/ and /z/ in text, in word-final sentence- medial position, by subjects*

Let us now turn to the post-alveolar fricatives, /ʃ/ and /ʒ/.

3.3 Experiment 2: Methods

Five women and one man aged 19–30 years participated in the experiment, all were speakers of Standard Hungarian living in Budapest and reported no speaking or hearing impairment.

The material was recorded in a sound-proof cabin with a SonyECM-MS907 microphone through an M-AudioMobilePreUSB preamplifier to a laptop using the software SpeechRecorder. The acoustic analysis was carried out in Praat 5.1.07 (Boersma and Weenink 2009). The data recorded consisted of 11–13-syllable-long sentences and a coherent text. The target sequences were *-ás* [aːʃ] and *-ázs* [aːʒ] in two- and three-syllable words in utterance-final position, and as a base line some of the test words were suffixed with a vowel-initial suffix, so they appeared in a sentence-medial intervocalic position. The experiment also contained a text in which test words appeared in word-final but sentence-medial position followed by a vowel-initial word. Test words were chosen to contain pairs of identical final syllables which only differed in the voicing of the final fricative (4). There were 34 test sentences and 6 repetitions which gave 204 tokens per subject, two items were discarded due to a technical error.

(4) a. *A telek jobb oldalán áll a ga**rázs**.*
‘There is a garage on the right side of the lot.’
b. *A sebesült könyökén van egy ma**rás**.*
‘There is a bite on the injured person’s elbow.’

The experiment aimed to measure the same parameters as Experiment 1 and, in addition, the harmonics-to-noise ratio in the fricative:

(5) a. voicing in the fricative
b. duration of the fricative
c. duration of the preceding vowel
d. harmonics-to-noise ratio

The harmonics-to-noise ratio (HNR) measures the extent to which friction noise replaces the harmonic structure in the spectrogram. To compare the relation of voicing to friction in /ʒ/, we adopted Hamann and Sennema’s (2005) method of what they call the *harmonicity median*, i. e., the degree of acoustic periodicity. The harmonicity median was determined by calculating the average of the harmonics-to-noise ratio with time steps of 0.01 s, a minimum pitch of 75 Hz, a silence threshold of 0.1 and 1 period per window. The interpretation of the median values is the following (see Boersma 1993). A harmonicity median of 0 dB means that there is equal energy in the harmonics and noise signal, whereas a median near 20 dB indicates that almost 100% of the energy of the signal is in the periodic part. Based on this, a sound with a harmonicity median below 3 dB can be regarded as a noisy/fricatival sound.

3.4 Experiment 2: Results

Figure 7 illustrates the realization of *(da)rázs* ‘wasp’ and *(da)rás* ‘semolina-adj.’ for Subject 3, Figure 8 exhibits the boxplots of unvoiced frames (%) across subjects for /ʃ/ and /ʒ/ in utterance-final and word-medial position.

Although numerically the voicing of /ʃ/ and /ʒ/ differs from that of /s/ and /z/, as /ʃ/ and /ʒ/ are realized with more phonation, qualitatively the results are very

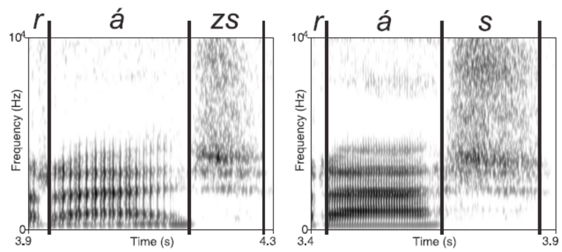

Figure 7: *The realization of* (da)rázs *and* (da)rás *for Subject 3*

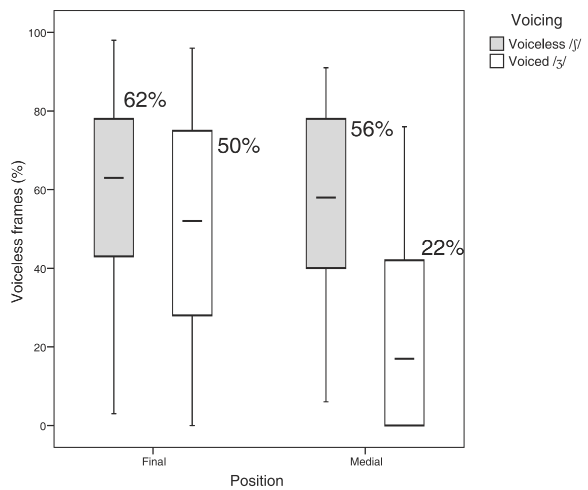

Figure 8: *Boxplots of unvoiced frames of /ʃ/ and /ʒ/ in utterance-final and word-medial position*

similar. Gráczi (2010) also found that /ʒ/ in intervocalic position was realized with more voicing (69.9%) than /z/ (57.9%). The difference between utterance-final /ʃ/ and /ʒ/ is significant, while the difference between final /ʒ/ and medial /ʃ/ are not significant — just like in Experiment 1. As expected, HNR significantly correlates with unvoiced frames (Pearson's correlation coefficient $r = -.779$ $p < 0.001$). This result indicates that the lower the harmonicity median, the higher the percentage of unvoiced frames, and the other way round, the higher the harmonicity median, the lower the percentage of unvoiced frames, i.e., periodicity negatively correlates with noise. This means that the more voiced a sound, the less turbulent or fricative-like it is. Figure 9 exhibits the HNR values for /ʃ/ and /ʒ/ in word-medial and utterance-final position.

In word-final, sentence-medial position the results are not statistically significant across subjects, and for any subject either, except for Subject 6. If, however, we exclude Subject 3 from the analysis, who produced an inverted pattern (she produced the phonologically voiced sibilant with less phonation than the phonologically voiceless one), we get a statistically significant result across subjects ($p = 0.002$).

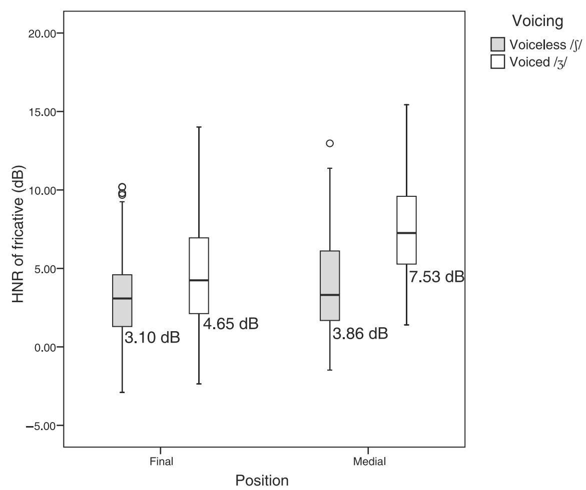

Figure 9: *Boxplots (with outliers) of harmonicity median values for /ʃ/ and /ʒ/ in utterance-final and word-medial position*

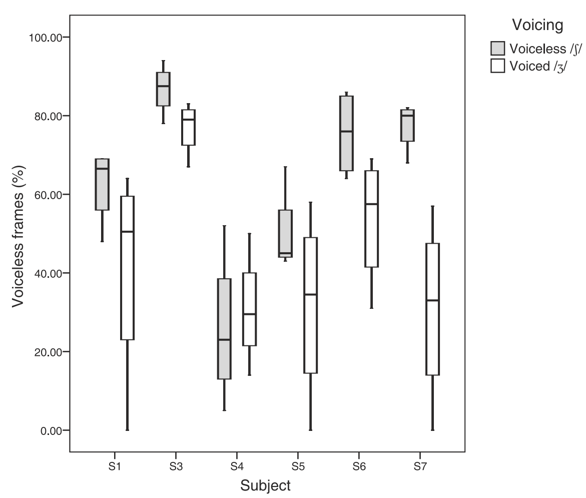

Figure 10: *Boxplots of unvoiced frames for /ʃ/ and /ʒ/ in text, in word-final sentence-medial position, by subjects*

Figure 11 shows the boxplots of the duration of the post-alveolar fricatives in utterance-final position and Figure 12 demonstrates the duration of the preceding vowel.

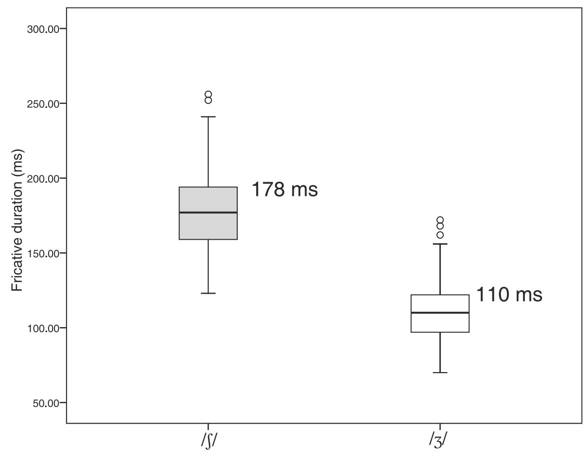

Figure 11: *Boxplots (with outliers) of duration of /ʃ/ and /ʒ/ in utterance final position*

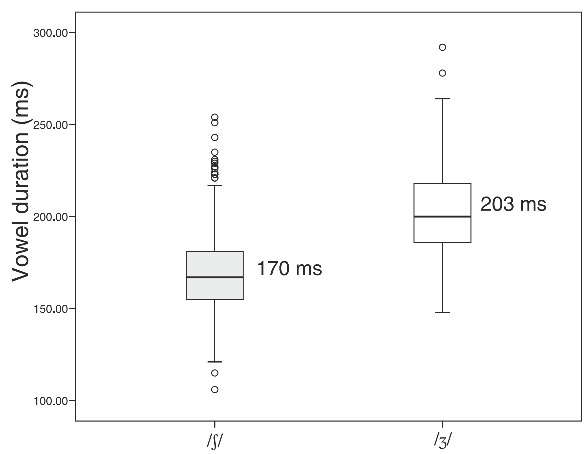

Figure 12: *Boxplots (with outliers) of the duration of the vowel preceding utterance-final /ʃ/ and /ʒ/*

The results are very similar to those in Experiment 1: as expected, /ʃ/ is always realized considerably longer than /ʒ/, the differences are statistically significant ($p < 0.001$), the differences are consistently significant for each subject individually as well. The preceding vowel is also always realized longer before the phonologically voiced fricative, the differences are statistically significant across subjects and for each subject as well ($p < 0.001$).

In sentence-medial position the results are again very similar. /ʒ/ is realized significantly shorter than its voiceless counterpart. However, the differences in the duration of the preceding vowel are not significant, unlike in utterance-final position (Figure 13), results are in accordance with Experiment 1.

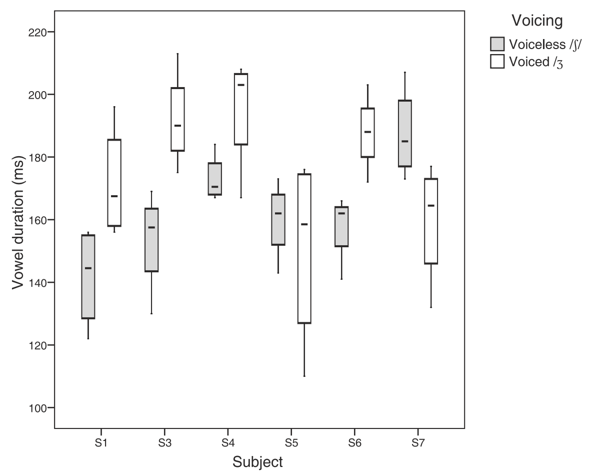

Figure 13: *Boxplots of the duration of the vowel before /ʃ/ and /ʒ/ in text, in word-final sentence-medial position, by subject*

4 General discussion

In accordance with our aerodynamically-grounded hypotheses, /z/ and /ʒ/ in Hungarian also devoice in utterance-final position, but there is no complete neutralization between the phonologically voiced and phonologically voiceless sibilants. (We cannot explain the substantial quantitative difference between the voicing of alveolar and post-alveolar fricatives.) One of the parameters that (partially) preserves the contrast is vowel length, or more precisely the ratio between vowel and fricative length.

It has been long observed that there is a correlation between the voicing properties of obstruents and the duration of preceding stressed vowels (or vowel + sonorant sequences), and the duration of closure or constriction of the obstruent (see, among others, House and Fairbanks 1953, Chen 1970, Lehiste 1970, Kluender et al. 1988). Voiceless obstruents as opposed to voiced obstruents are relatively long, and vowels (or vowel + sonorant sequences) before them are relatively short. This has been referred to in the English literature as Pre-Fortis Clipping (Wells 1982, Harris 1994). Since speakers typically talk at different rates, the absolute durations of the segments are highly variable. It has been found, however, for English and German for instance (Port and Dalby 1982, Port and Leary 2005) that the ratio of vowel duration to stop closure or fricative constriction remains rather constant in words with the same voicing feature.

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 and Dalby 1982, Port and Leary 2005, Massaro and Cohen 1983, Kluender et al. 1988). This hypothesis has been largely supported by experimental evidence. Thus, 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 and Diehl 1994).

Our results support that Hungarian also shows 'Pre-Fortis Clipping' since the length of the vowel is always longer in voiced tokens — a phonetic effect that seems to be redundant in most contexts, but turns out to be crucial in maintaining the contrast between voiced and voiceless sibilants in utterance-final position, just like in the case of maintaining /f/–/v/ contrast in the same position, as described in Bárkányi and Kiss (2009). The ratio between the duration of the consonant and that of the vowel for voiced fricatives is around twice as much as for their voiceless counterparts: for /s/ it is 0.82, for /z/ 1.51; /ʃ/: 0.97 and /ʒ/: 1.89. In Figures 14 and 15 the duration of the vowel+fricative sequence is shown. The length of the sequence in our study ranges between 322–347ms, which suggests that vowel and obstruent durations are indeed integrated into a single percept. Additional perception experiments are needed to corroborate this idea.

We must add that there are of course other phonetic properties of the preceding vowel that help maintain the contrast between voiced and voiceless fricatives, not only its duration. They include pre-aspiration or various low frequency spectral features. In this paper we do not discuss these properties, but see, among others, Gordeeva and Scobbie (2010).

Based on the above mentioned results we can claim that in the parameters we have examined voiced sibilants do not differ considerably from /v/. This questions the validity of the analyses given in Kiss and Bárkányi (2006) and Bárkányi and Kiss (2010) as they claim that it is the phonetic properties of /v/ (and its linear context) that explain its double-faced phonological behavior. Voiced sibilants, how-

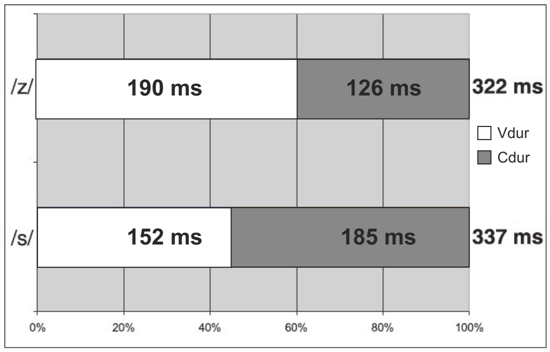

Figure 14: *Duration of the vowel plus /s/–/z/ in sentence-final postion*

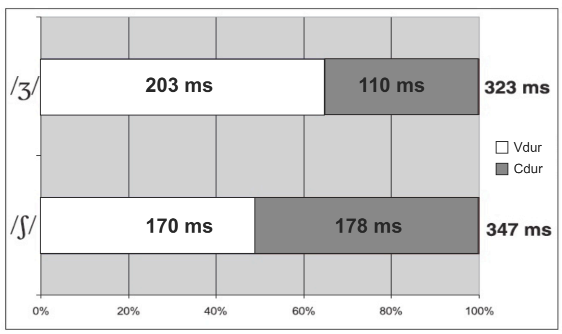

Figure 15: *Duration of the vowel plus /ʃ/–/ʒ/ in sentence-final position*

ever, do not show a peculiar behavior in voicing assimilation, they behave as proper obstruents by both triggering and undergoing voicing assimilation (see Section 2). This might be due to differences in other parameters that have not been measured in these experiments, such as spectral moments (frequency centroids, like the center of gravity and spectral standard deviation) or the intensity of the noise components of the fricative. The spectral center of gravity is a measure of how high the frequencies in a spectrum are on average. It is generally used to distinguish between the various places of articulation of obstruents as in Gordon et al. (2002), but Hamann and Sennema (2005) and Kiss and Bárkányi (2006) argue that it can be used to differentiate between the fricative vs. approximant realizations of labio-dentals in German and Dutch, and Hungarian, respectively. The spectrum of diffuse fricatives like /f/ and /v/ is spread out all through the frequency domain with low intensity, it shows no characteristic peaks, whereas the spectrum of comapct fricatives like

/z/ and /ʒ/ displays cahracteristic intensitiy peaks at various frequency postions, which contributes to their perceptual salience.

Alternatively, we might think that all voiced fricatives behave in a very similar manner in utterance-final position, i.e., they devoice in Hungarian, but the articulatory targets of /v/ still differ from those of /z/ and /ʒ/ in a meaningful way. In Section 5 this line of reasoning will be developed. Before further elaborating on this issue of why /v/ is still different, let us discuss the results obtained from the text.

We must admit that our results concerning the word-final but sentence-medial position are somewhat inconclusive. This position stands between word-medial and utterance-final position as for the voicing of the fricative, which is also explainable on aerodynamic grounds. The difference in the voicing of the voiced vs. voiceless fricative in this context is not as marked as in utterance-final position, but the voiced fricative on average is more phonated for all speakers except for Speaker 3 in the /ʃ/-/ʒ/ experiment, as shown in Figure 5 and Figure 9. We can see that standard deviation is larger and there are considerable individual differences. This means that phonetic voicing is possible but not necessary in intervocalic position. For some speakers probably the domain-final aspect of the context is dominant, while others treat it as an intervocalic position. We speculate that phonetic voicing is better perceived in this context. This phonetic environment exhibited no Pre-Fortis Clipping effect, that is to say, there were no consistent differences in the length of the preceding vowel between the phonologically voiced and phonologically voiceless context. We assume that the reason for this is that there is no need for this secondary cue to implement voicing contrast in this position, as it can adequately be perceived despite the lack of statistically significant differences for some speakers. Note that fricative length is a reliable indicator of phonological voicing for the sibilant fricatives, as mentioned above. Intensitiy—not measured here—is probably also a significant indicator of voicing: i.e., voiced fricatives have less energy at higher frequencies than their voiceless peers.

5 Is /v/ different?

The core of the analyses in Kiss and Bárkányi (2006) and Bárkányi and Kiss (2010) is based on the claim that the articulatory target of /v/ is a labiodental narrow approximant. The authors show that /v/ in utterance-final position is realized as a devoiced fricative, i.e., with little or no phonation and substantial turbulence. In this paper we have shown that /z/ and /ʒ/ are also realized with little or no phonation and substantial turbulence in utterance-final position. We claim, however, that while the articulatory target of /v/ is indeed a labiodental narrow approximant, /z/ and /ʒ/ are true voiced fricatives with active voicing and turbulence at the same time.

Gráczi (2010) shows that while /v/ is realized with 100% phonation (disregarding three extreme cases) in intervocalic position, /z/ and /ʒ/ contain 42.1% and 33.1% of unvoiced frames, respectively. This result is indicative of turbulence during the realization of /z/ and /ʒ/ and no friction during the realization of /v/.

In an additional experiment with two native speakers of Hungarian we examined the voicing and noise (HNR) properties of /z/ and /ʒ/ in contrast to /v/ in utterance-initial position. /v/ is voiced in 99% of the time, while /z/ only in 51% and /ʒ/ 79% in our study. Figure 16 shows the voicing of these sounds in word-initial position.

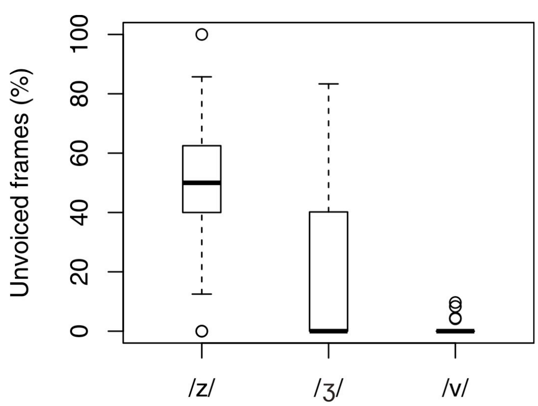

Figure 16: *Boxplots (with outliers) of unvoiced frames for /z/, /ʒ/ and /v/ in utterance-inital postion*

As expected, the HNR values indicate turbulence in the case of the voiced sibilants, but not in the case of /v/. The average HNR for utterance-initial /z/ is 5.37 dB in our experiment and 6.11 dB for /ʒ/. These results suggest a fair amount of noise during the production of the sibilants. Compare these mean HNR values with that of /v/, which is 10.82 dB — this suggests that /v/ is considerably less“noisy” and more approximant-like. Figure 17 shows the realization of word-initial /v/ in contrast to /ʒ/ for Speaker 1. In the case of the post-alveolar fricative substantial prevoicing is also observed (it is highlighted with a circle).

We can therefore conclude that the articulatory target for /v/ is an approximant, and as such, it is passively voiced, therefore its voicing gesture cannot propagate to the neighboring sounds, which means that it cannot trigger voicing assimilation, as described in detail in Kiss and Bárkányi (2006) and Bárkányi and Kiss (2010). On the other hand, /z/ and /ʒ/ are actively voiced obstruents. In order to achieve active voicing several articulatory gestures have to be realized as described in Section 1. This means that although the voiced sibilants devoice in utterance-final position,

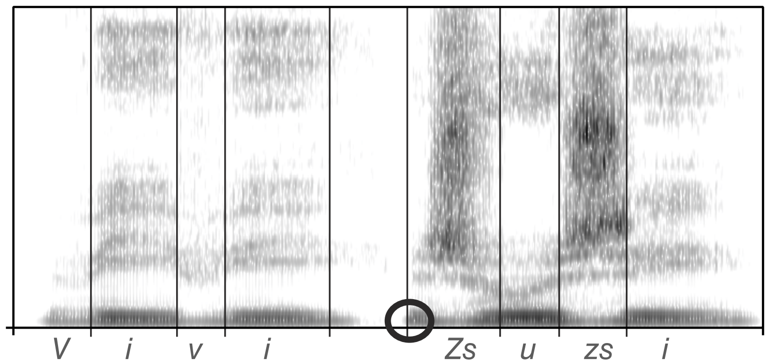

Figure 17: *The realization of* Vivi *and* Zsuzsi *for Speaker 1*

the active voicing quality of /z/ and /ʒ/ in other contexts can spill over to the surrounding sounds and, therefore, these sounds behave as proper obstruents, i.e., they both undergo and trigger voicing assimilation in Hungarian. It is open to further research whether voicing assimilation in Hungarian should be modelled on phonetic grounds, that is to say, it is a process governed by the temporal coordination of articulatory gestures, or it is rather a polarity-switching phonological process, or the combination of the two (as discussed in Jansen 2004, Kiss 2007, 301–307, for instance).

References

Bárkányi, Zs. and Z. Kiss (2010). A phonetic approach to the phonology of v: A case study from Hungarian and Slovak. In: Fuchs et al. (2010, 103–142).

Bárkányi, Zs., Z. Kiss and K. Mády (2009). Az /s/–/z/ oppozíció fenntartása megnyilatkozás végén [On the contrast of /s/ and /z/ in utterance-final position in Hungarian]. Paper presented at the Beszédkutatás 2009 [Speech Research 2009] conference. Budapest, 16–17 October 2009. http://www.budling.hu/~cash/papers/beszkut2009-slide.pdf

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In: *Proceedings of the Institute of Phonetic Sciences (University of Amsterdam)* 17, 97–110.

Boersma, P. and D. Weenink (2005). Praat: Doing phonetics by computer. (Version 5.1.38) [Computer program].

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. In: *Phonetica* 22, 129–159.

Fuchs, S., M. Toda and M. Żygis (eds) (2010). *Turbulent sounds. An interdisciplinary guide*. Berlin & New York: Mouton/de Gruyter.

Gordeeva, O. B. and J. M. Scobbie (2010). Preaspiration as a correlate of word-final voice in Scottish English fricatives. In: Fuchs et al. (2010, 167–207).

Gordon, M., P. Barthmaier and K. Sands (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association 32*, 141–174.

Gráczi, T. E. (2009). A spiránsok zöngésségi oppozíciójának néhány jellemzője [Properties of the voicing contrast of fricatives]. In: M. Gósy (ed.): Beszédkutatás 2010 [Speech research 2010], 42–54. Budapest: MTA Nyelvtudományi Intézet, Kempelen Farkas Beszédkutató Laboratórium.

Hamann, S. and A. Sennema (2005). Acoustic differences between German and Dutch labiodentals. In: *ZAS Papers in Linguistics 42*, 33–41.

Harris, J. (1994). English sound structure. Oxford : Blackwell.

House, A. and G. Fairbanks (1953). The influence of consonantal environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America 25*, 105–113.

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

Jesus, L. M. T. de (2001). *Acoustic phonetics of European Portuguese fricative consonants*. Doctoral dissertation, University of Southampton.

Johnson, K. (2003). Acoustic and auditory phonetics (Second edition). Oxford: Blackwell.

Kingston, J. and R. L. Diehl (1994). Phonetic knowledge. *Language 70*, 419–454.

Kiss, Z. (2007). *The phonetics–phonology interface: allophony, assimilation and phonotactics*. Doctoral dissertation, Eötvös Loránd University (ELTE), Budapest. http://www.budling.hu/∼cash/papers/papers.html

Kiss, Z. and Zs. Bárkányi (2006). A phonetically-based approach to the phonology of [v] in Hungarian. *Acta Linguistica Hungarica 53*, 175–226.

Kluender, K. R., Randy L. D. and B. A. Wright (1988). Vowel length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics 16*, 153–169.

Krane, M. H. (2005). Aeroacoustic production of low-frequency unvoiced speech sounds. *Journal of the Acoustical Society of America 118*, 410–427.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge MA: MIT Press.

Massaro, D. and M. Cohen (1983). Consonant/vowel ratio: an improbable cue in speech perception. *Perception and Psychophysics 33*, 502–505.

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

Padgett, J. (2002). Russian voicing assimilation, final devoicing, and the problem of [v]. Ms., University of California, Santa Cruz. http://people.uscs.edu/∼padgett/locker/newvoice.pdf

Parker, E., Randy L. Diehl and K. Kluender (1986). Trading relations in speech and nonspeech. *Perception and Psychophysics 39*, 129–142.

Port, R. F. and J. Dalby (1982). C/V ratio as a cue for voicing in English. *Perception and Psychophysics 2*, 141–152.

Port, R. F. and A. P. Leary (2005). Against formal phonology. *Language 81*, 927–964.

Shadle, C. H. (1985). The acoustics of frictaive consonants (RLE Technical Report 506). Cambridge MA: MIT Press.

Siptár, P. and M. Törkenczy (2000). *The phonology of Hungarian*. Oxford: Oxford University Press.

Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge MA: MIT Press.

Stevens, K. N., S. Blumstein, L. Glicksman, M. Burton and K. Kurowski (1992). Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *Journal of the Acoustical Society of America 91*, 2979–3000.

Wells, J. C. (1982). *Accents of English 1–3*. Cambridge: Cambridge University Press.