


# The perception of voicing contrast in assimilation contexts in minimal pairs: evidence from Hungarian

ZSUZSANNA BÁRKÁNYI<sup>1</sup> and ZOLTÁN G. KISS<sup>2\*</sup> 

<sup>1</sup> The Open University and Research Institute for Linguistics, HAS, Budapest, Hungary

<sup>2</sup> ELTE Eötvös Loránd University, Budapest, Hungary

Received: January 1, 2021 • Accepted: April 7, 2021

Published online: June 3, 2021



© 2021 The Author(s)

## ABSTRACT

It has been long acknowledged that the perception and production of speech is affected by the presence or absence of higher levels of linguistic information, too. The recoverability of meaning heavily relies on semantic context, similarly, the precision of articulation is inversely proportional to the presence of semantic information. The present study explores the recoverability of the voice feature of word-final alveolar fricatives in minimal pairs in Hungarian in phonetic contexts that trigger regressive voicing assimilation. Specifically, it aims to clarify whether the acoustic differences found in earlier studies are perceptually salient enough to distinguish underlying voicing in minimal pairs in semantically ambiguous contexts. For this reason, a perception study with the synthesised minimal pair *mész-méz* 'whitewash-honey' was carried out where the amount of voicing in the fricative, and the duration of the fricative and vowel were manipulated. The target words appeared in the following three phonetic contexts: before /p/, before /b/ and before the vowel /a/. Our results suggest that the observed acoustic differences in most of the cases remain below the perceptual threshold which means that phonological contrast is indeed neutralised before obstruents in Hungarian, and this may cause semantic ambiguity.

## KEYWORDS

regressive voicing assimilation, perception, neutralisation, compensation, Hungarian, minimal pairs

\* Corresponding author. E-mail: gkiss.zoltan@btk.elte.hu

## 1. INTRODUCTION

The speech signal is by nature highly variable, partly because of the individual physiological differences between speakers, partly because of intended differences between individual utterances due to speech rate or any other prosodic manipulations, and partly because of the phonetic context a speech sound is in. It is well accepted that adjacent speech sounds are influenced by each other. The change triggered in this way can be purely coarticulatory or provoked by language specific phonological rules. If a segment becomes more similar to the speech sound preceding or following it, we speak about assimilation. Assimilation can be so strong that a contrastive segment may lose its distinctive power fully or partially which might hinder its recoverability during perception. The perception of speech segments involves, as [Martin & Peperkamp \(2011, 10\)](#) put it, “[...] segmenting raw acoustic input and assigning each segment the appropriate category label. The probability that a given segment will be correctly categorised depends on what other categories it might be confused with, and where precisely the boundary between categories lies”. It has been long acknowledged that the perception and production of speech are affected by the presence or absence of higher levels of linguistic information, too. The recoverability of meaning heavily relies on semantic context, similarly, the precision of articulation is inversely proportional to the presence of semantic information ([Lieberman & Mattingly 1985](#)). Research on sound change (e.g., [Martinet 1952](#); [Silverman 2012](#)) also shows that emergent homophony and semantic misinterpretation militates against complete phonological neutralisation. There is ample evidence from speech production studies that lexical neighbours affect the fine phonetic realisation of words (see [Goldrick et al. 2013](#) and the references therein). For example, voiceless stops in words with a minimal pair neighbour (*cod* vs. *god*) are produced with longer VOTs than stops in matched words without minimal pairs (*cop* with no corresponding \**gop*) ([Bease-Berk & Goldrick 2009](#)).

Lexical neighbours seem to affect speech perception as well. It has been shown that category boundary shifts to the lexical end of an acoustic continuum. [Ganong \(1980\)](#) in an identification experiment demonstrated that voiced–voiceless stop pairs showed strong lexical effects, namely, listeners preferred words to nonwords in their categorisations. Test words were constructed along acoustic continua with variable VOT values where only one end of the continuum corresponded to an actual word (e.g., *dash–tash*; *dask–task*). The phenomenon has been known as the “Ganong effect” since this initial study. It has also been attested that voicing contrast in a neutralising context is more likely to be partially preserved in minimal pairs. [Charles-Luce \(1993\)](#) in an acoustic study examining the role of semantic information in regressive voicing assimilation (RVA) in Catalan, observed that if assimilation would lead to semantic ambiguity, it was more likely to be only partial. The length of the vowel preceding the obstruent systematically distinguished phonologically voiced and voiceless segments in her study in the assimilating environment significantly more frequently in minimal pairs than in non-minimal pairs where additional information was also present to recover meaning.

[Kitahara et al. \(2019\)](#) arrived at a similar conclusion. The authors investigated whether the voicing contrast in word-initial /k/ and /g/ in Japanese in spontaneous speech was affected by lexical factors, namely the presence of a minimal-pair competitor. The authors found that neither VOT nor closure duration were affected by lexical factors. However, an unexpected finding of the study was that the duration of the following vowel was significantly longer when a voicing competitor existed than when it did not. The authors conclude that this effect might



have two sources. On the one hand, pronunciation is more careful if a lexical competitor exists. On the other hand, it might be explained by a recent trend in Japanese whereby voicing contrast signalled by VOT is getting lost and being transferred to the pitch (and length) features of the following vowel.

Although a few studies on Hungarian have shown (e.g., Jansen 2004; Gráczki 2010; Bárkányi & G. Kiss 2015) that some phonetic correlates of the voicing contrast are systematically preserved in neutralising environments, there are no studies that investigate the influence of lexical factors on voicing neutralisation such as minimal pairs in this language. In a full-fledged study of the lexical effects in regressive voicing assimilation the behaviour of minimal pairs and non-minimal pairs should be compared in biasing and non-biasing contexts. In the present study, as a first step, we examine minimal pairs in non-biasing contexts. We seek to answer whether voicing contrast is recoverable in minimal pairs that only differ in the voice feature of the final segment. The specific research questions we aim to answer are:

1. To what extent does the perception of the fricatives /s/ and /z/ differ in minimal pairs in different phonetic contexts: before /p/, before /b/, and before a vowel across a word boundary?
2. To what extent are the acoustic differences found in production relevant in the perception of the contrast between the fricatives /s/ and /z/ in minimal pairs?

## 2. PREVIOUS STUDIES ON REGRESSIVE VOICING ASSIMILATION IN HUNGARIAN

It is a well-established view that adjacent obstruent clusters in Hungarian must agree in voicing and it is the last obstruent in the cluster that determines whether the cluster is voiced or voiceless. Obstruents in Hungarian contrast in terms of voicing word-initially (*pár* /pa:r/ 'pair' – *bár* /ba:r/ 'bar'), in intervocalic position (*ékig* /e:kiɡ/ 'wedge.TERM' – *égig* /e:giɡ/ 'sky.TERM'), and word-finally (*mész* /me:s/ 'whitewash' – *méz* /mez/ 'honey'). However, according to the traditional descriptive literature, regressive voicing assimilation in Hungarian is a completely neutralising process (see e.g., Siptár & Törkenczy 2000), and thus, voiceless and devoiced, or contextually voiced and underlyingly voiced segments cannot be distinguished on the basis of their phonetic or phonological behaviour, that is, e.g., *méztől* 'honey.ABL' and *mészről* 'whitewash.ABL' are identical in pronunciation: [me:stø:l].

In the new millennium, however, a number of different approaches have appeared. Jansen (2004) found that the underlying contrasts between /k/ and /ɡ/, and /ʃ/ and /ʒ/ are partially preserved before voiced obstruents: the underlyingly voiced segments showed more phonation than the voiceless ones; and in the case of /ʃ/ and /ʒ/ the duration of the preceding vowel was also systematically different depending on the voicing properties of the fricative. Similarly, Gow & Im (2004) also argue that RVA in Hungarian might be graded. The authors found that voiced segments showed shorter VOTs than assimilated segments and unvoiced segments; while assimilated segments showed shorter VOTs than unvoiced segments. Thus, they conclude that Hungarian voicing assimilation produces segments whose voicing is acoustically intermediate between those of voiced and unvoiced obstruents. Markó et al. (2010) in a study on spontaneous and read speech examining two and three-consonant clusters also conclude that RVA in



Hungarian is phonetically incomplete. In their production experiment voiced obstruents in around 80% of the cases preserved some degree of voicing before a voiceless obstruent while voiceless obstruents showed only partial voicing before a voiced obstruent in approximately 40% of the cases. The authors, however, did not examine whether voiced or voiceless obstruents were more likely to preserve their underlying properties, nor did they focus on the recoverability of the assimilated consonants.

Bárkányi & G. Kiss (2015) in an acoustic experiment on the /t/-/d/ and /s/-/z/ contrast before /p/ and /b/ also found traces of incomplete neutralisation. The authors examined parameters related to phonation and segment duration: the absolute length of the voiced interval, the ratio of the unvoiced part compared to the total length of the consonant, duration of the preceding vowel, duration of the target consonant, and vowel to consonant duration ratio. The devoicing context turned out to be highly neutralising, with only traces of vowel length difference in the case of the alveolar fricative pair, while the voicing contrast only seemed to be neutralising for stops, but not for fricatives: /s/ was significantly more voiceless than /z/ before /b/. Several questions arise in light of these acoustic studies. For example, are the acoustic differences observed in these experiments salient enough to be perceived by native speakers? Also, how are the segments exhibiting gradient and partial voicing mapped onto the phonological categories of voiceless vs. voiced?

### 3. PERCEPTION AND ASSIMILATION

A number of studies have investigated how listeners cope with assimilations, most of which focus on changes in the place of articulation (see e.g., Mitterer et al. 2013 for an overview). Most of these studies agree that listeners make use of the contextual information and compensate for coarticulatory/assimilatory changes. Viable assimilations, but not unviable assimilations, are often confused perceptually with canonical word forms in word identification tasks. This means that a changed word form is recognised as if it had not been changed only in the context that licenses such a change (i.e., in viable assimilatory context). Ohala (1981) in a study on vowel perception states that informants tolerated well the difference between the intended shape and the realisation of a vowel if it could be considered as a result of the phonetic context, which means that unintentional coarticulation is compensated for.

In a study on the perception of assimilated segments in RVA in French, Snoeren et al. (2008) investigated whether information from the actual word form was sufficient (or more important) to recover the underlying word form, or rather, information from the triggering context was more relevant. The authors used simple noun phrases (ending in /t/ and /d/) in an auditory-visual priming experiment in which the noun was never predictable in order to exclude bias from sentence meaning. Voiced final stops in the experiment were partially devoiced, while the voiceless stops were almost fully voiced in accordance with Snoeren et al.'s (2006) previous acoustic studies. When the triggering context was not present, reaction times were shortest for canonical forms, longer in the assimilated condition and longest in the unrelated condition, but there was no difference between the words with underlyingly voiced and voiceless segments. In the triggering context, however, word forms with voiceless stops were recognised more quickly than those with voiced ones. It seems that the assimilating context helps recover completely assimilated speech segments but not partially assimilated ones. The authors conclude that the



two sources of information (context and inherent cues) are complementary and both are taken into account by listeners when processing assimilated forms. In the perception of completely assimilated segments listeners rely on the following context, while in partially assimilated forms context has lesser importance.

In line with research on place assimilations in different languages, Mitterer et al. (2006) examining perceptual compensation for manner assimilation in Hungarian liquids with word and non-word stimuli concluded that viably assimilated words and canonical word forms were difficult to distinguish while this was not the case for unviably modified forms. Interestingly, this study did not find any effects of wordedness.

On the contrary, Kuzla et al. (2010) found that lexical factors do play a role in the recoverability of assimilated segments. The authors explicitly examined the role of minimal pairs in progressive voice assimilation in German. The focus of the study was the degree of assimilation of the lenis fricatives /v/ and /z/ after word- and phrase-boundaries preceded by the voiceless stop /t/. The test word for /v/ – *Wälder* /vældɐ/ ‘forests’ – has a minimal pair neighbour *Felder* /fældɐ/ ‘fields’, while the test word for /z/ – *Senken* /zɛŋkən/ ‘hollows’ – has no such close competitor in the lexicon as /s/ is not allowed word-initially in German. As for the acoustic side of the study, fricatives in the assimilation context were devoiced compared to fricatives in the non-assimilation context, and /z/ was more devoiced than /v/, but more importantly, assimilation did not affect the duration of the lenis fricatives, even though duration is an important cue in German for the fortis–lenis distinction. The perception experiment contained test words in which the initial fricatives had been manipulated: the two endpoints contained a completely voiceless token of /f/ and a completely voiced token of /v/ respectively, and 18 intermediate steps replacing the glottal cycles of the /v/-endpoint one by one by a part of the /f/-endpoint, starting from the left. The results showed that there were more /v/ responses in assimilation than in non-assimilation contexts, which means that listeners compensated for the loss of phonation when they perceived it as a consequence of the phonetic context. The authors conclude that the prosodic structure also played an important role as listeners accepted (almost) completely devoiced fricatives more readily as realisations of /v/ after word boundaries than after phrase boundaries but no prosodic conditioning of compensation for the devoicing of /z/ was found, which the authors explain with the lack of lexical ambiguity in the latter case.

#### 4. PERCEPTION OF VOICING IN HUNGARIAN

There are few studies examining the perception of voicing and especially the recoverability of the underlying voicing of assimilated obstruents in Hungarian. Bárkányi & Mány (2012) examined the perception of utterance-final /s/ vs. /z/ using synthesised speech. Subjects heard the test words *méz* /me:z/ ‘honey’ and *mész* /me:s/ ‘whitewash’ in isolation and had to respond in a forced-choice test. The length of the segments in the test words were determined in accordance with previous acoustic studies (see Section 2): /m/ being 50 ms long, /e:/ 250 ms and the fricative 210 ms. Voicing was added in 10% steps to the fricative, i.e., there were 11 different stimuli with end points as completely voiceless and completely voiced items. The mean inflection point turned out to be at 30% voicing (SD = 8%), that is, with only 30% of voicing during the fricative interval, the segment was more likely to be perceived as voiced (*méz*) than voiceless (*mész*). (Note that in Bárkányi & G. Kiss 2015, utterance-final fricatives contained less than 30% voicing).



In order to determine the perceptual role of secondary phonetic correlates of voicing contrast when the primary correlate – phonation – is partially lost, the authors carried out a second experiment. In that experiment too, synthesised tokens of the words *mész* and *méz* were used. 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 at  $30\% \pm 1 \times 8$  and  $\pm 2 \times 8$ , i.e., at the following five levels: 14, 22, 30, 38 and 46% voicing of the fricative interval. The duration of vowel plus consonant was set at 360 ms. 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. The authors found that in the case of the most ambiguous stimulus, i.e., when 30% of the fricative interval was voiced, listeners were as likely to perceive a /z/ as an /s/ if the vowel was 160 ms long; with longer vowels participants were more likely to identify that test word as *méz*, while with shorter ones they tended to hear *mész* in line with research according to which in the perception of laryngeal properties a whole cue-complex plays a role and not a single phonetic feature (cf. Javkin 1976; Port & Dalby 1982; Massaro & Cohen 1983; Parker et al. 1986; Kluender et al. 1988; Kingston & Diehl 1994; Port & Leary 2005).

The only study focussing on perception in RVA in Hungarian is Gow & Im (2004). In this paper the authors investigated the recognition of consonants following voiced, voiceless, and assimilated segments. They studied the effects of anticipation produced by the assimilated segment, that is, the recoverability of consonants that trigger RVA. Stimuli were extracted from meaningful speech and created by cross-splicing. The authors argue that while language-specific phonological processes systematically affect speech production, they do not appear to interfere with spoken word recognition as these rely on universal perceptual mechanisms. The role of lexical factors was not part of the study.

## 5. EXPERIMENT

### 5.1. Method, subjects and procedure

In this section we now turn to the discussion of an experiment that aimed to investigate the perception of the contrast between /s/ and /z/ in minimal pairs in voicing assimilatory environments. We used the same synthesised *mész* /me:s/ ‘whitewash’ – *méz* /me:z/ ‘honey’ minimal pair tokens that were used in the experiment of Bárkányi & Mány (2012), with the same durations and with the same five voicing ratios within the fricative interval: 14, 22, 30, 38, and 46% (see Section 4). Each of these tokens were embedded in the following three sentences:<sup>1</sup>

- (1) A \_\_\_\_ pakolás nem jelent nagyobb erőfeszítést.  
 A \_\_\_\_ berakás nem jelent nagyobb erőfeszítést.  
 A \_\_\_\_ átrakás nem jelent nagyobb erőfeszítést.

The carrier sentences were read out by a native speaker of Hungarian (male, in his 40s) at a natural speech rate, leaving as much space at the given position so that the synthesised forms

<sup>1</sup>Glosses: ‘The packing/placing/transfer of \_\_\_\_ doesn’t take much effort.’



could be inserted. Some of the acoustic parameters of the carrier sentences (amplitude, frequency range) were modified to minimise the difference between them and the synthesised tokens, although maximal fidelity was not possible to achieve, the embedded forms sounded somewhat less natural than the carrier sentences. The experiment investigated the perception of /s/ and /z/ in the minimal pair *mész*–*méz* across a word boundary before the plosives /p/ and /b/, and the vowel /a/.<sup>2</sup> The participants of the experiment heard the following sentences:

- (2) A *mész/z* pakolás nem jelent nagyobb erőfeszítést.  
 A *mész/z* berakás nem jelent nagyobb erőfeszítést.  
 A *mész/z* átrakás nem jelent nagyobb erőfeszítést.

The duration of the vowel and the fricative interval was modified in 20-ms steps, altogether in six steps: (step 1: 130 + 230 ms; step 6: 230 + 130 ms). The durational values are summarised in Table 1. For example in step 1, when the duration of the vowel was 130 ms, and that of the following consonant 230 ms, and when only 14% of the consonant had voicing, the duration of that voicing was 32 ms long, when the voicing duration was 22% it was 51 ms, etc.

The total number of tokens embedded in each of the three sentences was 30 (5 voicing ratios × 6 duration ratios). The experiment used a multiple forced choice test format in which the participants had to decide whether the word they heard was *mész* ‘whitewash’ (with final /s/) or *méz* ‘honey’ (with final /z/) by clicking on a computer screen showing these two choices. The experiment was created and carried out in the ExperimentMFC module of Praat (Boersma & Weenink 2015). Ten university students participated in the experiment, all were native speakers of Hungarian. Each of them heard all 30 stimuli (which were randomised) three times, this means that altogether 2,700 items could be analysed (10 participants × 3 rounds × 3 sentences × 30 tokens).

The statistical analysis (including the generation of the various plots) was carried out in R (R Core Team 2020) using various *tidyverse* packages (Wickham et al. 2019), as well as the

**Table 1.** Duration of segments (in ms) used in the experiment (V = vowel, C = consonant, V/C = vowel to consonant duration ratio, V/VC = ratio of the vowel’s duration to that of the whole vowel+consonant interval); the percentages indicate the ratio of voicing in the consonant

Step	V	C	V/C	V/VC	14%	22%	30%	38%	46%
1	130	230	0.57	0.36	32	51	69	87	106
2	150	210	0.71	0.42	29	46	63	80	97
3	170	190	0.89	0.47	27	42	57	72	87
4	190	170	1.12	0.53	24	37	51	65	78
5	210	150	1.40	0.58	21	33	45	57	69
6	230	130	1.77	0.64	18	29	39	49	60

<sup>2</sup>Since its length does not play a role in our discussion, we will simply transcribe the vowel as /a/, without the length mark.



*broom.mixed* package (Bolker & Robinson 2020) for the extracting of model components, the *MuMIn* package (Bartón 2020) for calculating  $R^2$  values for the final model, and the *patchwork* package (Pedersen 2020) during the composition of the plots. Generalized logistic mixed effects models (estimated using ML and Nelder-Mead optimizer) were used to model the data, using the package *lme4* (Bates et al. 2020). Voicing response was the dependent variable, giving predicted log odds of producing a voiced response as the model outcome, where a voiced response meant a choice of the word *mész* with the voiced final fricative (as opposed to *mész* with a final voiceless fricative). Random effects were used to model the experiment structure the following way. We fitted random intercepts and random slopes for the proportion of voicing and the vowel to consonant duration ratio varying across participants and items (the stimuli the participants heard). If the slopes for subject and/or items did not improve model fit relative to intercepts only, they were removed from the final model, and we retained only random intercepts. If a model did not converge with the default Nelder-Mead optimizer, we used “BOBYQA” optimizing (Bound Optimization by Quadratic Approximation), in which case models always converged. If a random variable had “singularity” issues (variances of the effect was (close to) zero), that random effect was removed from the model. The effect of the fixed and random variables was tested via model comparisons using the log-likelihood ratio test.

We will report the results of the model building and the model comparisons, and the properties of the final models using the guidelines in Meteyarda & Davies (2020). In the final model tables, the confidence intervals (95% CIs) and  $p$ -values were computed using the Wald approximation. During model building the numerical variables were scaled (standardised), i.e., they represent standard deviations from the mean of the given variable. Since the estimated random-slope coefficients measure how many times bigger the log-odds of one outcome is for one value of a predictor, compared to another value, i.e., they tell the direction and the strength of the relationship between the fixed effect and the odds that the response is voiced, these random-slope coefficients can also be interpreted as effect sizes.

The plots representing the fitted logistic regression models were generated using the non-standardised, “raw” data points of the given predictor. The inflection points (where the predicted probability of a voiced response is 0.5) reported in these plots were calculated using

the following formula:  $\frac{\ln\left(\frac{0.5}{(1-0.5)}\right) - \beta_0}{\beta_1} = \frac{-\beta_0}{\beta_1}$ , in which the betas were extracted from the given logistic regression model. The data points were drawn using some jitter so that they can be discerned better.

## 5.2. Results

**5.2.1. Before /p/.** First we begin with the results of the experiment when the test words occurred before voiceless /p/. Table 2 displays the properties of the model building and the model comparisons.

Based on Table 2, we retained in the final model only random intercepts for subject and item, and we did not include the interaction term between the proportion of voicing and the vowel/consonant duration ratio (this model is “mod.p.propv.vcrat” in the table). The properties of the final model are shown in Table 3.





**Table 2.** Model building and model comparison for the pre-/p/ environment

Model name	npar	AIC	BIC	logLik	Deviance	Chi-square	df	p
mod.p.ic.rdsb	2	1,208.08	1,217.68	-602.04	1,204.08	—	—	—
mod.p.ic.rdsb.rdit	3	883.39	897.80	-438.70	877.39	326.69	1	<0.0001
mod.p.propv	4	840.01	859.22	-416.01	832.01	45.38	1	<0.0001
mod.p.propv.rdsb	6	838.91	867.72	-413.45	826.91	5.1	2	0.0779
mod.p.propv.rdsb.rdit	8	841.92	880.34	-412.96	825.92	0.99	2	0.6097
mod.p.propv	4	840.01	859.22	-416.01	832.01	—	—	—
<b>mod.p.propv.vcrat</b>	5	806.18	830.19	-398.09	796.18	35.83	1	<0.0001
mod.p.propv.vcrat.rdsb	7	804.25	837.86	-395.12	790.25	5.93	2	0.0515
mod.p.propv.vcrat.rdsb.rdit	9	808.09	851.32	-395.05	790.09	0.15	2	0.9264
mod.p.propv.vcrat	5	806.18	830.19	-398.09	796.18	—	—	—
mod.p.propv.vcrat.iact	6	807.77	836.58	-397.88	795.77	0.41	1	0.5203

**Table 3.** Generalized linear mixed effects model results for proportion of voicing and vowel to consonant duration ratio in the pre-/p/ environment

Effect	Group	Term	Estimate	SE	95% CI	z	p
fixed		(Intercept)	0.66	0.23	0.22, 1.1	2.92	0.0035
fixed		prop.voice	1.75	0.12	1.5, 1.99	14.05	<0.0001
fixed		vc.ratio	0.78	0.1	0.58, 0.99	7.5	<0.0001
ran_pars	stimulus	sd__(Intercept)	0.16				
ran_pars	subject	sd__(Intercept)	0.64				

The total explanatory power of the final model shown in Table 3 is substantial (conditional  $R^2 = 0.55$ ) and the part related to the fixed effects alone (marginal  $R^2$ ) is of 0.50. The effect of both the proportion of voicing and the vowel to consonant duration ratio is significantly positive. The results indicate that both predictors greatly influence the voicing responses. If we convert the log-odds coefficient values to odds, we can say that the odds of a voiced response is increased by 5.76 times (log odds = 1.75) at each one-standard deviation increase of the proportion of voicing (while the vowel to consonant duration ratio is held at its average value). The same one-standard deviation increase in the vowel to consonant ratio results in a smaller predicted increase in the odds of voice responses: the increase will be 2.18 times as big (log odds = 0.78) as without this effect (while the proportion of voicing has its average value). This indicates that all else kept constant, the voicing ratio increase has a greater impact on voicing responses than the vowel's relative duration.



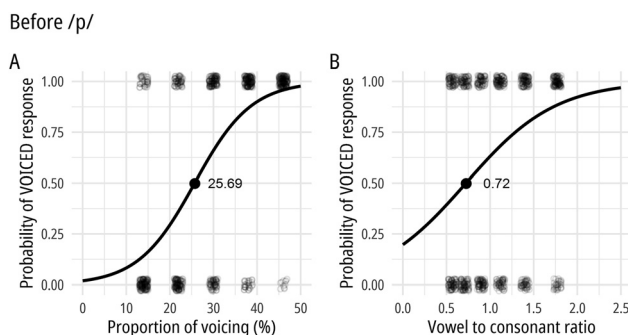
Figure 1 shows the voicing response as a function of the proportion of voicing in the fricative and the vowel to consonant duration ratio with a superimposed logistic regression fit curve; the black dot in the middle indicates the inflection point where a voiced response (i.e., *mész*) becomes more likely than an unvoiced response (i.e., *mész*).

As we can see in Figure 1, the model predicts that in order for the word-final fricative to be categorised as voiced before /p/, it needs to contain at least 25.69% of voicing (i.e., the inflection point of the regression model sigmoid curve is at 25.69%). On the other hand, the vowel to consonant duration ratio needs to be 0.72 or more for the fricative to be categorised as voiced /z/ rather than voiceless /s/, i.e., the vowel needs to be around three-quarter as long as the fricative. We note that at this point the models based on which the plots were generated still contain *all* data points: both relatively voiceless and voiced tokens, as well as those with different vowel/consonant ratios. We will tease these two predictors apart in Section 5.2.4.

**5.2.2. Before /b/. Table 4** displays the properties of the model building and the model comparisons for the data before /b/.

Including the fixed-effect predictor of the vowel duration ratio in the model created “singularity” issues (there was no variance in the predicted random intercepts for items), and for this reason, the item random effect was removed from the model. The remaining model (“mod.b.propv.vcrat.noit” in Table 4) was then compared to other models. Allowing slopes to vary for subjects for both the proportion of voicing and the vowel to consonant duration ratio improved model fit, but not their interaction. The final model (“mod.b.propv.vcrat.noit.rdsb” in Table 4) then contained both fixed predictors, plus varying intercepts and slopes across subjects only, and no interaction terms. Table 5 provides the summary of this final model.

Just like in the case of the pre-/p/ context, before /b/ too, the total explanatory power of the final model and the part related to the fixed effects are substantial (conditional  $R^2 = 0.63$ , marginal  $R^2 = 0.55$ ). The effect of both the proportion of voicing and the vowel/consonant duration ratio is significantly positive. The results indicate that both predictors greatly influence the voicing responses. Converting the log-odds coefficients to odds, we can say that the odds of a voiced response is increased by 7.69 times (log odds = 2.04) at each



**Fig. 1.** Perception of final /s/ vs. /z/ before /p/ as a function of (A) proportion of voicing in the fricative and (B) vowel to consonant duration ratio



**Table 4.** Model building and model comparison for the pre-/b/ environment

Model name	npar	AIC	BIC	logLik	Deviance	Chi-square	df	p
mod.b.ic.rbsub	2	1,248.88	1,258.49	-622.44	1,244.88	–	–	–
mod.b.ic.rbsub.rdit	3	874.69	889.10	-434.35	868.69	376.19	1	<0.0001
mod.b.propv	4	829.00	848.21	-410.50	821.00	47.69	1	<0.0001
mod.b.propv.rbsub	6	826.23	855.04	-407.11	814.23	6.77	2	0.0338
mod.b.propv.rbsub.rdit	8	829.72	868.14	-406.86	813.72	0.51	2	0.7763
mod.b.propv.vcrat.noit	6	789.36	818.18	-388.68	777.36	–	–	–
<b>mod.b.propv.vcrat.noit.rbsub</b>	9	776.22	819.44	-379.11	758.22	19.15	3	0.0003
mod.b.propv.vcrat.noit.rbsub	9	776.22	819.44	-379.11	758.22	–	–	–
mod.b.propv.vcrat.iact	14	784.88	852.12	-378.44	756.88	1.33	5	0.9313

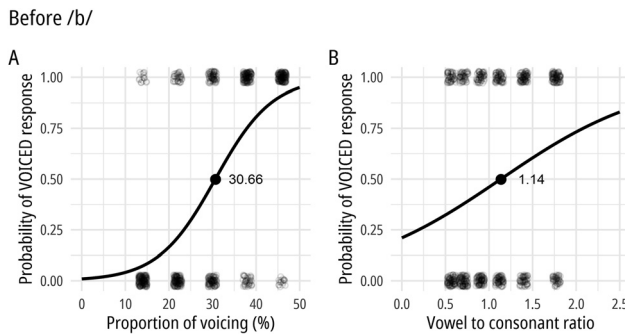
**Table 5.** Generalized linear mixed effects model results for proportion of voicing and vowel to consonant duration ratio in the pre-/b/ environment

Effect	Group	Term	Estimate	SE	95% CI	z	p
fixed		(Intercept)	-0.09	0.17	-0.42, 0.24	-0.56	0.5766
fixed		prop.voice	2.04	0.2	1.65, 2.43	10.3	<0.0001
fixed		vc.ratio	0.88	0.23	0.44, 1.33	3.9	0.0001
ran_pars	subject	sd__(Intercept)	0.44				
ran_pars	subject	cor__(Intercept).prop.voice	0.52				
ran_pars	subject	cor__(Intercept).vc.ratio	0.38				
ran_pars	subject	sd__prop.voice	0.44				
ran_pars	subject	cor__prop.voice.vc.ratio	0.30				
ran_pars	subject	sd__vc.ratio	0.62				

one-standard deviation increase of the proportion of voicing (while the vowel/consonant ratio has its average value). The same one-standard deviation increase in the vowel to consonant duration ratio results in a smaller predicted increase in the odds of voice responses: the increase will be 2.41 times as big (log odds = 0.88) as without this effect (while the proportion of voicing has its average value). This indicates again that all else kept constant, the voicing ratio increase has a greater impact on voicing responses than the vowel's relative duration before /b/, too.

Figure 2 shows the voicing response as a function of the proportion of voicing in the fricative and the vowel to consonant duration ratio with a superimposed logistic regression fit curve for the pre-/b/ position.





**Fig. 2.** Perception of final /s/ vs. /z/ before /b/ as a function of (A) proportion of voicing in the fricative and (B) vowel to consonant duration ratio

Figure 2 shows that in order for the word-final fricative to be categorised as voiced before /b/, it needs to contain 30.66% of voicing or more. The vowel to consonant duration ratio needs to be at least 1.14 for the fricative to be categorised as voiced /z/ rather than voiceless /s/, i.e., the vowel needs to be somewhat longer than the fricative. These values are higher than in the case of the pre-/p/ environment.

**5.2.3. Before /a/.** Finally, we turn to the prevocalic environment, i.e., where the test items occurred before /a/. Table 6 provides the details of the model building and comparisons.

Adding the fixed-effect predictor of vowel duration ratio caused no variance in the predicted random intercepts for items (singularity), and therefore the item random effect was removed from the model. The remaining model (“mod.a.propv.vcrat.noit” in Table 6) was then compared to other models. Since the other, more complex models did not improve model fit, this model

**Table 6.** Model building and model comparison for the pre-/a/ environment

Model name	npar	AIC	BIC	logLik	Deviance	Chi-square	df	p
mod.a.ic.rdsb	2	1,245.20	1,254.80	-620.60	1,241.20	–	–	–
mod.a.ic.rdsb.rdit	3	826.61	841.02	-410.31	820.61	420.58	1	<0.0001
mod.a.propv	4	779.44	798.65	-385.72	771.44	49.18	1	<0.0001
mod.a.propv.rdsb	6	770.25	799.06	-379.12	758.25	13.19	2	0.0014
mod.a.propv.rdsb.rdit	8	771.44	809.86	-377.72	755.44	2.81	2	0.2453
<b>mod.a.propv.vcrat.noit</b>	6	735.00	763.81	-361.50	723.00	–	–	–
mod.a.propv.vcrat.noit.rdsb	9	733.89	777.11	-357.95	715.89	7.1	3	0.0687
mod.a.propv.vcrat.noit	6	735.00	763.81	-361.50	723.00	–	–	–
mod.a.propv.vcrat.iact	14	742.52	809.76	-357.26	714.52	8.47	8	0.3888



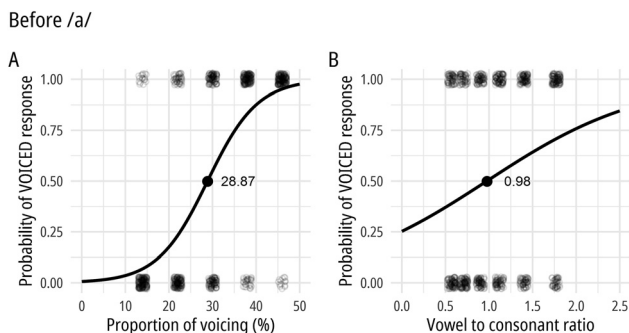
**Table 7.** Generalized linear mixed effects model results for proportion of voicing and vowel to consonant duration ratio in the pre-/a/ environment

Effect	Group	Term	Estimate	SE	95% CI	z	p
fixed		(Intercept)	0.22	0.21	-0.19, 0.64	1.07	0.283
fixed		prop.voice	2.24	0.26	1.73, 2.75	8.61	<0.0001
fixed		vc.ratio	0.86	0.1	0.65, 1.06	8.19	<0.0001
ran_pars	subject	sd__(Intercept)	0.58				
ran_pars	subject	cor__(Intercept).prop.voice	0.05				
ran_pars	subject	sd__prop.voice	0.67				

turned out to be the final one used for analysis. This model contained both fixed predictors, plus varying intercepts and slopes across subjects only, and no interaction terms. The properties of this final model are presented in Table 7.

According to the  $R^2$  values, the final model's total explanatory power (conditional  $R^2 = 0.66$ ) and the part related to the fixed effects alone are substantial (marginal  $R^2 = 0.59$ ). Just like before /p/ and /b/, the effect of both the proportion of voicing and the vowel to consonant duration ratio is significantly positive, both predictors greatly influence the voicing responses. Specifically, in the prevocalic position, the odds of a voiced response is increased by 9.3 times at each one-standard deviation increase of the proportion of voicing while the vowel duration ratio has its average value (log odds = 2.24). As far as the vowel duration ratio is concerned, a one-standard deviation increase results in the increase of the odds for voiced responses by 2.36 (log odds = 0.86) while the proportion of voicing has its average value. Similarly to the pre-/p/ and pre-/b/ environments, the voicing ratio increase has a greater impact on voicing responses than the vowel's relative duration.

The prevocalic voicing responses as a function of the proportion of voicing in the fricative and the vowel to consonant duration ratio are shown in Fig. 3 with a superimposed logistic regression fit curve indicating the predicted probability of voiced responses.

**Fig. 3.** Perception of final /s/ vs. /z/ before /a/ as a function of (A) proportion of voicing in the fricative and (B) vowel to consonant duration ratio

Based on Fig. 3, we can say that the model predicts that the fricative needs to contain at least around 29% voicing in order to be categorised as voiced before /a/, whereas the vowel to consonant duration ratio needs to be at least 0.98, i.e., the vowel needs to be around as long as the fricative. Just like before /b/, these values are higher than in the case of the pre-/p/ environment.

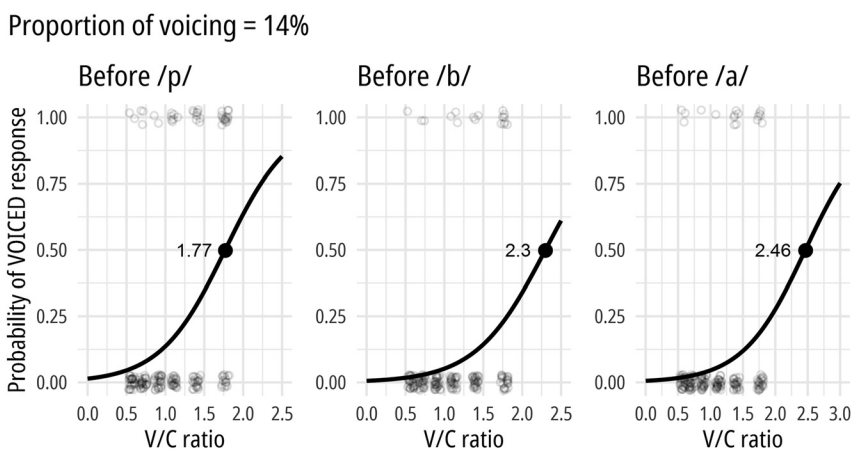
**5.2.4. Vowel duration effects by voicing proportion.** In what follows, we will look at the effect of the vowel to fricative duration ratio on the perception of voicing depending on the proportion of voicing in the fricative.

*Proportion of voicing = 14%.* The voicing responses as a function of the vowel to consonant duration ratio are shown in Figure 4. As before, the superimposed logistic regression fit curve indicates the predicted probability of voiced responses, while the black dot in the middle signals the inflection point where a voiced response becomes more likely than an unvoiced response.

As Figure 4 shows, the number of voiced responses when the fricative interval contained very little voicing was low, the vast majority of responses were voiceless. As we can see, if there is only very little fricative voicing, the model predicts that the preceding vowel needs to be around at least twice as long as the following consonant in order to be perceived as voiced. The inflection point is the smallest when the following sound is voiceless /p/ (1.77), while before /b/ and the vowel the values are similar (2.3 and 2.46 respectively).

*Proportion of voicing = 22%.* Figure 5 shows that when the fricative contains 22% voicing, the model predicts that the preceding vowel has to be at least around 1.5 times as long as the fricative for it to be perceived as voiced. Just as in the case of the fricative containing 14% voicing, here too, it is before /p/ that the inflection point is the smallest (1.53); before /b/, the model predicts that the vowel should be at least around twice as long as the fricative so that it can be categorised by listeners as voiced.

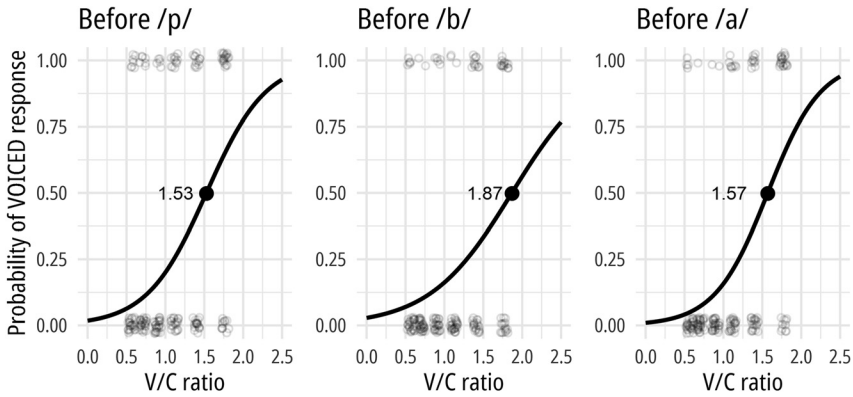
*Proportion of voicing = 30%.* As we can see in Figure 6, compared to the values when there is only 14 and 22% voicing in the fricative, at 30% voicing, the perceptual inflection points are



**Fig. 4.** Perception of final /s/ vs. /z/ before /p/, /b/, and /a/ as a function of vowel to consonant duration ratio when the proportion of voicing in the fricative is 14%

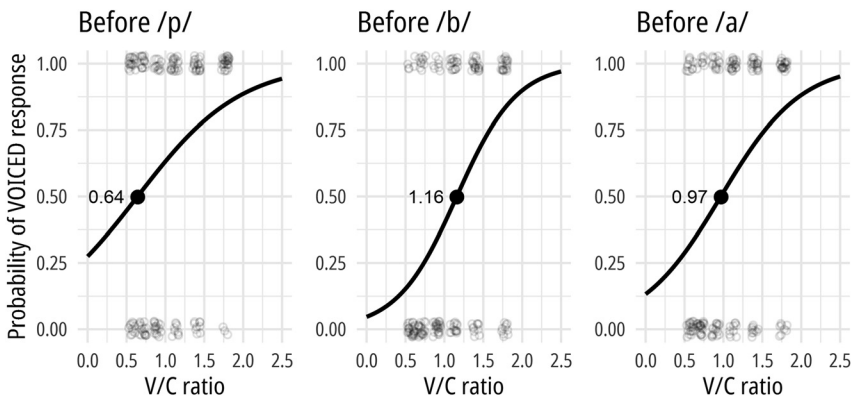


Proportion of voicing = 22%



**Fig. 5.** Perception of final /s/ vs. /z/ before /p/, /b/, and /a/ as a function of vowel to consonant duration ratio when the proportion of voicing in the fricative is 22%

Proportion of voicing = 30%

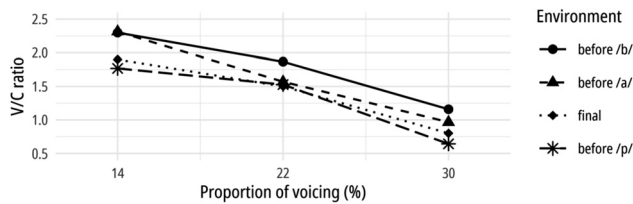


**Fig. 6.** Perception of final /s/ vs. /z/ before /p/, /b/, and /a/ as a function of vowel to consonant duration ratio when the proportion of voicing in the fricative is 30%

lower in all three environments. According to the prediction of the model, at 30% fricative voicing the preceding vowel should be about as long as the fricative so that the probability of voiced perceptions become more frequent if the next sound is /b/ or /a/. Again, the perceptual inflection point is lower when the following sound is /p/: in this case, the duration of the vowel is predicted to be a little more than half of that of the consonant.

In the remaining two voicing proportion classes (38%, 46%), the vowel to consonant duration did not play a role: voicing alone was a sufficient cue to categorise the fricative as voiced (the model predicted a vowel length close to or below zero).





**Fig. 7.** Perceptual inflection points as a function of vowel to consonant duration ratio at three voicing proportions in four environments

These results indicate that the length of the vowel plays a gradually lesser role as the amount of voicing increases, cf. Figure 7.<sup>3</sup>

Figure 7 shows that it is before /p/ that the perceptual inflection point is consistently the lowest across the three voicing proportions, i.e., it is in this position that the smallest vowel to consonant duration ratio is sufficient to perceive the final fricative as voiced, regardless how much voicing there is in the fricative. The pre-/b/ environment is the one in which the inflection points are consistently the highest (closely followed by the prevocalic position). Put simply, before /p/, listeners categorised the fricative as voiced more readily than before /b/, /a/, or in absolute word-final position. For example, when the fricative contained only 14% voicing, then the vowel had to be more than twice as long as the fricative for the fricative to be perceived as voiced by the participants of the experiment when the following sound was /b/. The vowel to consonant duration ratio at 14% voicing had to be at least 1.9 in word-final position, and only 1.77 before /p/.

## 6. DISCUSSION

The purpose of the present research was to study the perceptual consequences of regressive voicing assimilation in Hungarian in minimal pairs. The first research question asked to what extent the perception of word-final /s/ and /z/ in minimal pairs differs in different phonetic contexts. It has been mentioned in the Introduction that wordedness might influence the perception of lexical items and thus the perception of speech sounds in them. The identification of contrastive segments is generally biased towards words in contrast to non-words. The test words of the present study were chosen so that no such bias was present, they formed a minimal pair, i.e., they were both existing words, and the semantic context provided by the carrier sentences did not produce such bias either. In this way the potential impact of the lexical status of the test words was controlled for.

Based on the perceptual inflection point values shown in Figure 7, we can set up the following hierarchy of environments, in which the values of the proportion of voicing necessary to induce a voiced response gradually increase from left to right:

<sup>3</sup>Figure 7 also contains the values in the absolute word final (prepausal) environment, which were taken from B ark anyi & M ady (2012).



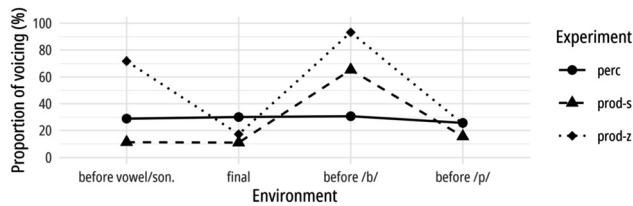


(3) before /p/ < absolute final position < before vowel < before /b/

It must be noted that fairly little voicing – 30% of the fricative interval or less – seems to be sufficient to favour a voiced response in all the examined phonetic contexts. The situation before /p/ is interesting. While the perceptual inflection points before /b/, /a/ and in absolute word-final position were rather similar, it was lower before /p/, i.e. a smaller amount of voicing was sufficient for the fricative to be categorised by listeners as voiced before /p/ than before /b/ and before the vowel /a/. We assume that this is due to perceptual compensation: in the voiceless environment (before /p/), listeners expect less voicing since they are used to hearing relatively less voicing in the phonologically voiced forms in this position, i.e., they perceptually compensate for the smaller amount of voicing. Thus, the overall probability that they hear a voiced form even with little voicing available will increase before /p/. This is in line with [Kuzla et al. \(2010\)](#) as it is an indication that speakers more readily identify a slightly voiced sibilant as voiced in the devoicing context than in the voicing context, which means that they compensate for the loss of phonation when they perceive it as a consequence of the phonetic context, but unlike in [Snoeren et al. \(2008\)](#), the perceptual compensation applies for partially assimilated segments as well. The perceptual compensation is noticeable in the temporal properties of these sequences, too. This is the context with the smallest V/C ratio, which means that the vowel does not have to be as long as in the other contexts to provoke a voiced response. In Hungarian, vowels before voiceless obstruents are typically shorter than before voiced obstruents (e.g., [Bárkányi & G. Kiss 2015, 2020](#)). The fricative before /p/ is likely to be voiceless as a result of RVA, our results show that in this position a fairly short vowel is sufficient to bring about a voiced percept. This suggests that listeners are more sensitive to voicing cues in a devoicing context than in contexts where they do not expect the voicing cues to be compromised.

It has been demonstrated that the vowel to consonant length ratio plays a role in the identification of the voice feature of the fricative, but its role gradually diminishes as the amount of voicing increases. Generally speaking, when the intrinsic acoustic properties of speech segments are robust, the disambiguating role of the contextual cues are lessened ([Stilp 2019](#)). However, it is not always easy to distinguish intrinsic and extrinsic acoustic cues. While phonation, i.e., the vibration of vocal folds, can be viewed as an intrinsic acoustic property and as such an intrinsic perceptual cue of a voiced fricative, the temporal properties of the preceding vowel are more likely to be interpretable in relation to the temporal properties of the fricative itself. It is well-attested in the literature, especially for English, that shorter vowels make the obstruent sound longer, and thus induce more voiceless responses (e.g., [Port & Dalby 1982](#); [Massaro & Cohen 1983](#); [Kluender et al. 1988](#); [Port & Leary 2005](#)). Our results indicate that both cues under scrutiny play an important role in identifying the fricative as voiced, but phonetic voicing in Hungarian has a superior role over durational cues: all else kept constant, the voicing ratio increase had a greater effect on voicing responses than the vowel's relative duration in all three contexts we investigated. A relatively long vowel and a short fricative, however, could induce a voiced response even if the fricative was only slightly voiced. It is not rare that when more than one acoustic cue is present in a contrast, listeners weigh one more heavily ([Goudbeek 2006](#); [Clayards 2008](#); [Clayards et al. 2008](#); [Goudbeek et al. 2008](#)). [Francis & Kaganovich \(2008\)](#) report that although both fundamental frequency at the onset of voicing and voice onset time are present, as well as other relevant cues, listeners weigh voice onset time more heavily in the recognition of voiceless–voiced syllable-initial stops in English. Similarly, Hungarian listeners





**Fig. 8.** Perceptual inflection points for the proportion of voicing and the mean proportions of voicing of the production experiment in four environments. Abbreviations: “perc” = inflection points from the perception experiment, “prod-s” = results for /s/ from the production experiment, “prod-z” = results for /z/ from the production experiment.

seem to weigh phonation more heavily than V/C ratio, which is in accordance with this language being a voicing language rather than an aspirating one.

If a fricative is fully voiced or completely voiceless, that is, stands at the endpoints of the voiced–voiceless continuum, its identification is straightforward; however, the acoustic characteristics of fluent coarticulated speech are rarely so clear (Lindblom 1963). In the present research, the segments that had to be identified were mid-continuum members of the voiced–voiceless and vowel/consonant ratio continua. According to Stilp (2019), such mid-continuum stimuli are more representative of the speech produced in everyday conversations.

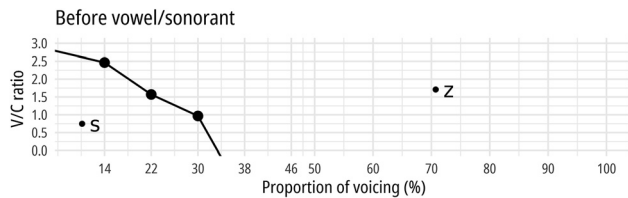
The second research question aimed to determine whether the subtle phonetic differences observed in earlier acoustic studies on RVA were relevant for the perception of laryngeal contrasts in Hungarian. For this reason, results from the current study are compared with the production data from Bárkányi & G. Kiss (2015) and (2020). The plot in Figure 8 displays the inflection points for the proportion of voicing before /p/, /b/ and /a/ measured in the present perception experiment (the “A” plots in Figures 1–3) and the mean proportions of voicing of the production experiments.<sup>4</sup> Since there are no relevant experimental results concerning the voicing properties of /s/ and /z/ before vowels across a word boundary, the plot in Figure 8 actually shows the pre-sonorant values from the production study of Bárkányi & G. Kiss (2015). Neither sonorant consonants nor vowels are known to trigger voicing assimilation in preceding obstruents in standard Hungarian, and therefore, the two environments can be merged into one set.

Figure 8 indicates that before vowels/sonorants the voicing contrast of /s/ and /z/ is significantly different: the mean voicing of /s/ (11.25%) is well below the perceptual inflection point, and therefore, it is assumed to be mostly perceived to be voiceless, while /z/ is well above it (71.84%), and so it is assumed to be perceived mostly voiced. The proportion of voicing thus seems to be a salient intrinsic perceptual cue before vowels/sonorants, and so the phonological contrast between /s/ and /z/ is maintained. This result corroborates the fact that vowels/sonorants do not trigger regressive voicing assimilation in standard Hungarian.

In absolute word-final position (data from Bárkányi & Mády 2012) however, the contrast between /s/ and /z/ seems to be neutralised: even though the mean values for the proportion of

<sup>4</sup>The perception results for the voicing of word-final (prepausal) /s/ and /z/ in Hungarian are from Bárkányi & Mády (2012).





**Fig. 9.** Perceptual inflection points before /a/ as a function of vowel to consonant duration ratio at five levels of fricative voicing (points connected with a line), and the results of the production experiment for /s/ and /z/. The categorisation below the line is voiceless, above it voiced.

voicing in the production experiment are different (/s/: 10.95%, /z/: 17.23%), they are below the perceptual inflection point, indicating that both /s/ and /z/ are likely to be perceived as voiceless. This suggests that the alveolar sibilant fricatives in Hungarian have taken the first step towards utterance-final voicing neutralisation, at least with regard to the phonetic parameters measured in these studies.

Before /b/, both values of the mean voicing proportions in the production experiment are above the perceptual inflection point (/s/: 65.39%, /z/: 93.22%), strongly suggesting that /s/ and /z/ are likely to be perceived as voiced in this environment, again, in spite of the fact that the mean voicing values are different. The contrast of the two fricatives thus seems to be neutralised before /b/.

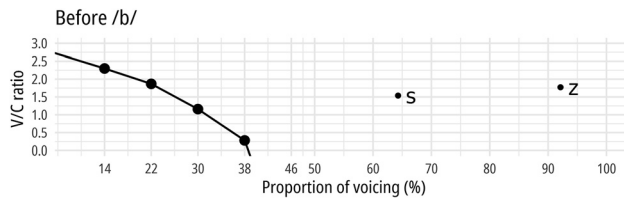
The situation before /p/ is similar. The mean voicing proportion of /s/ in the production experiment was 15.13%, which is below the perceptual inflection point (25.69%), indicative of it being categorised as voiceless. The mean voicing proportion in /z/ was 25.07%, which is very close to the inflection point but still below it. These results suggest that /s/ and /z/ are both likely to be perceived as voiceless before /p/.

In the following, we will disentangle the interplay between the two acoustic cues – voicing and V/C duration ratio. As shown in Figure 9, /z/ cannot induce a voiceless response, irrespective of the vowel to consonant length ratio (it is always in the voiced-response region, i.e., above the line in Figure 9), while /s/ could only induce a voiced response with an unrealistically long vowel. As we reported in Section 5.2.4, if the fricative only contained 14% voicing, our model predicted that the preceding vowel had to be at least 2.5 times as long as the fricative (assuming a 100-ms-long fricative, the vowel would have to be at least around 250 ms long); in the production experiment, the mean voicing proportion was even less, 11.25%, and so we predict that an even longer/more unrealistic vowel would be required for voiced responses.

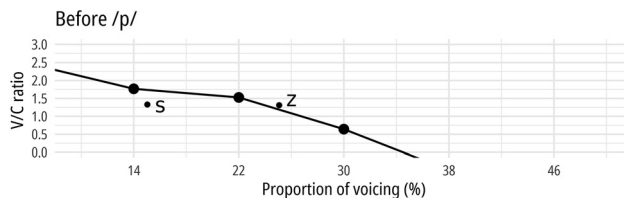
Figure 10 clearly shows that the acoustic differences between /s/ and /z/ are not translated into perceptual differences since if the proportion of the voiced interval in the fricative surpasses 38%, it is identified as /z/, that is, the duration of the preceding vowel is outweighed. The substantial differences in the voicing ratio are due to a longer voiceless fricative, not the actual amount of voicing (/s/: 38 ms, /z/: 47 ms), which, according to B ark anyi & G. Kiss (2015), is indicative of a phonologised voicing assimilation rather than coarticulatory voicing.

Our results before /p/ are somewhat less conclusive (Figure 11). In the production study the vowel before /s/ was 1.33 times longer than the fricative which puts /s/ below the perceptual voicing threshold. /z/, on the other hand, (with a 1.31 V/C ratio) is on the voiced–voiceless





**Fig. 10.** Perceptual inflection points before /b/ as a function of vowel to consonant duration ratio at five levels of fricative voicing (points connected with a line), and the results of the production experiment for /s/ and /z/. The categorisation below the line is voiceless, above it voiced.



**Fig. 11.** Perceptual inflection points before /p/ as a function of vowel to consonant duration ratio at five levels of fricative voicing (points connected with a line), and the results of the production experiment for /s/ and /z/. The categorisation below the line is voiceless, above it voiced.

category boundary. Note, though, that the perception experiment consisted of five voicing categories only (14%, 22%, 30%, 38%, 46%). This means that the jump between the points was relatively large. Had we applied smaller steps – especially between 22% and 30% where the voicing values of /z/ fall – it is likely that /z/ would be more below the perceptual voicing threshold, although still close to it.

The present research confirms that the phonological contrast between /s/ and /z/ in minimal pairs in regressive voicing assimilation contexts is neutralised in Hungarian. The acoustic differences observed in production studies are not mapped onto categorical perceptual differences. It requires further research whether other acoustic properties can still contribute to partial contrast preservation in these phonetic contexts. The acoustic differences that are systematically present before vowels are perceived by Hungarian listeners and thus voicing contrast is preserved in this context.

## 7. CONCLUSION

This research parted from the assumption that some acoustic correlates could potentially sustain the voicing opposition of obstruents in general – and alveolar fricatives in particular – in minimal pairs in regressive voicing assimilation contexts in Hungarian. After examining the proportion of voicing and the vowel to consonant duration ratio we can conclude that the acoustic differences observed in earlier studies do not surpass the perception threshold, that is, the phonological contrast is completely neutralised despite partial acoustic differences. This



confirms the findings of traditional descriptive and generative accounts according to which regressive voicing assimilation in Hungarian is a categorical neutralising process. It has also been demonstrated that listeners compensate for the loss of voicing if they perceive it as a result of the phonetic context. Further studies will clarify whether other phonetic correlates such as intensity and the spectral properties of vowels could still contribute to partial contrast preservation.

## ACKNOWLEDGMENTS

We wish to thank the editors and two anonymous reviewers for helpful comments that improved both the form and the content of this contribution. The usual disclaimers apply.

## REFERENCES

- Baese-Berk, Melissa and Mathew Goldrick. 2009. Mechanisms of interaction in speech production. *Language and Cognitive Processes* 24. 527–554. <https://doi.org/10.1080/01690960802299378>.
- Bárkányi, Zsuzsanna and Zoltán G. Kiss. 2015. Why do sonorants not voice in Hungarian? And why do they voice in Slovak? In Katalin É. Kiss, Balázs Surányi and Éva Dékány (eds.) *Approaches to Hungarian 14: Papers from the 2013 Piliscsaba Conference*. Amsterdam and Philadelphia: John Benjamins. 65–94. <https://doi.org/10.1075/atoh.14.03bar>
- Bárkányi, Zsuzsanna and Zoltán G. Kiss. 2020. Neutralisation and contrast preservation: Voicing assimilation in Hungarian three-consonant clusters. *Linguistic Variation* 20. 56–83. <https://doi.org/10.1075/lv.16010.bar>
- Bárkányi, Zsuzsanna and Katalin Mády. 2012. The perception of voicing in fricatives. Paper presented at the 9th Old World Conference in Phonology, Berlin, 18–21 January 2012.
- Bartoń, Kamil. 2020. MuMIn: Multi-model inference. <https://CRAN.R-project.org/package=MuMIn>. R package version 1.43.17.
- Bates, Douglas, Martin Maechler, Ben Bolker and Steven Walker. 2020. lme4: Linear mixed-effects models using ‘eigen’ and S4. <https://CRAN.R-project.org/package=lme4>. R package version 1.1-25.
- Boersma, Paul and David Weenink. 2015. Praat: Doing phonetics by computer. [Computer program]. Version 5.4.18, retrieved 2015-9-7 from <http://www.praat.org/>.
- Bolker, Ben and David Robinson. 2020. broom.mixed: Tidying methods for mixed models. <https://CRAN.R-project.org/package=broom.mixed>. R package version 0.2.6.
- Charles-Luce, Jan. 1993. The effects of semantic context on voicing neutralization. *Phonetica* 50. 28–43. <https://doi.org/10.1159/000261924>.
- Clayards, Meghan. 2008. The ideal listener: Making optimal use of acoustic-phonetic cues for word recognition. Doctoral dissertation. University of Rochester, Rochester, NY.
- Clayards, Meghan, Michael K. Tanenhaus, Richard N. Aslin and Robert A. Jacobs. 2008. Perception of speech reflects optimal use of probabilistic speech cues. *Cognition* 108. 804–809.
- Francis, Alexander L. and Natalya Kaganovich. 2008. Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. *Journal of the Acoustical Society of America* 124. 1234. <https://doi.org/10.1121/1.2945161>.



- Ganong, William F. 1980. Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 6. 110–125. <https://doi.org/10.1037/0096-1523.6.1.110>.
- Goldrick, Matthew, Charlotte Vaughn and Amanda Murphy. 2013. The effects of lexical neighbors on stop consonant articulation. *Journal of the Acoustical Society of America* 134. 172–177. <https://doi.org/10.1121/1.4812821>.
- Goudbeek, Martijn. 2006. The acquisition of auditory categories. Doctoral dissertation. Radboud University, Nijmegen, Netherlands.
- Goudbeek, Martijn, Anne Cutler and Roel Smits. 2008. Supervised and unsupervised learning of multidimensionally varying non-native speech categories. *Speech Communication* 50. 109–125.
- Gow, David and Aaron Im. 2004. A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language* 51. 279–296. <https://doi.org/10.1016/j.jml.2004.05.004>.
- Gráczy, Tekla Etelka. 2010. A spiránsok zöngésségi oppozíciójának néhány jellemzője [Some characteristics of the voicing contrast of fricatives]. *Beszédkutatás* 18. 42–56.
- Jansen, Wouter. 2004. Laryngeal contrast and phonetic voicing: A laboratory phonology approach to English, Hungarian, and Dutch. Doctoral dissertation. University of Groningen, Groningen, Netherlands.
- Javkin, Hector. 1976. The perceptual basis of vowel duration differences associated with the voiced/voiceless distinction. Report of the Phonology Laboratory, UC Berkeley 1. 78–92.
- Kingston, John and Randy L. Diehl. 1994. Phonetic knowledge. *Language* 70. 419–454.
- Kitahara, Mafuyu, Keiichi Tajima and Kiyoko Yoneyama. 2019. The effect of lexical competition on realization of phonetic contrasts: A corpus study of the voicing contrast in Japanese. In Sasha Calhoun, Paola Escudero, Marija Tabain and Paul Warren (eds.) *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia, 2019. 2749–2752.
- Kluender, Keith R., Randy L. Diehl and Beverly A. Wright. 1988. Vowel-length differences before voiced and voiceless consonants: an auditory explanation. *Journal of Phonetics* 16. 153–169. [https://doi.org/10.1016/S0095-4470\(19\)30480-2](https://doi.org/10.1016/S0095-4470(19)30480-2).
- Kuzla, Claudia, Mirjam Ernestus and Holger Mitterer. 2010. Compensation for assimilatory devoicing and prosodic structure in German fricative perception. In Cécile Fougeron, Barbara Kühnert, Mariapaola D'Imperio and Nathalie Vallée (eds.) *Laboratory phonology 10*. Berlin & New York: de Gruyter Mouton. 731–757.
- Lieberman, Alvin M. and Ignatius G. Mattingly. 1985. The motor theory of speech perception revised. *Cognition* 21. 1–36. [https://doi.org/10.1016/0010-0277\(85\)90021-6](https://doi.org/10.1016/0010-0277(85)90021-6).
- Lindblom, Björn E. 1963. Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35. 1773–1781.
- Markó, Alexandra, Tekla E. Gráczy and Judit Bóna. 2010. The realisation of voicing assimilation rules in Hungarian spontaneous and read speech: Case studies. *Acta Linguistica Hungarica* 57. 210–238.
- Martin, Andrew and Sharon Peperkamp. 2011. Speech perception and phonology. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.) *The Blackwell companion to phonology*. Malden, MA & Oxford: Wiley-Blackwell. 2334–2356.
- Martinet, André. 1952. Function, structure, and sound change. *Word* 8. 1–32. <https://doi.org/10.1080/00437956.1952.11659416>.
- Massaro, Dominic W. and Michael M. Cohen. 1983. Phonological context in speech perception. *Perception and Psychophysics* 34. 338–348.
- Meteyarda, Lotte and Robert A. I. Davies. 2020. Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language* 112. 1–22.



- Mitterer, Holger, Valéria Csépe and Leo Blomert. 2006. The role of perceptual integration in the recognition of assimilated word forms. *The Quarterly Journal of Experimental Psychology* 59. 1395–1424. <https://journals.sagepub.com/doi/abs/10.1080/17470210500198726>.
- Mitterer, Holger, Sahyang Kim and Taehong Cho. 2013. Compensation for complete assimilation in speech perception: The case of Korean labial-to-velar assimilation. *Journal of Memory and Language* 69. 59–83. <https://doi.org/10.1016/j.jml.2013.02.001>.
- Ohala, John J. 1981. The listener as a source of sound change. *Chicago Linguistic Society* 17. 178–203.
- Parker, Ellen M., Randy L. Diehl and Keith R. Kluender. 1986. Trading relations in speech and nonspeech. *Perception and Psychophysics* 39. 129–142.
- Pedersen, Thomas Lin. 2020. Patchwork: The composer of plots. <https://CRAN.R-project.org/package=patchwork>. R package version 1.1.0.
- Port, Robert F. and Jonathan, Dalby. 1982. Consonant/vowel ratio as a cue for voicing in English. *Perception and Psychophysics* 32. 141–152. <https://doi.org/10.3758/BF03204273>.
- Port, Robert F. and Adam P. Leary. 2005. Against formal phonology. *Language* 81. 927–964. <https://doi.org/10.1353/LAN.2005.0195>.
- R Core Team. 2020. R: A language and environment for statistical computing; 4.0.2. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Silverman, Daniel. 2012. Neutralization. Key topics in phonology. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139013895>.
- Siptár, Péter and Miklós Törkenczy. 2000. The phonology of Hungarian. Oxford: Oxford University Press.
- Snoeren, Natalie D., Pierre A. Hallé and Juan Segui. 2006. A voice for the voiceless: Production and perception of assimilated stops in French. *Journal of Phonetics* 34. 241–268. <https://doi.org/10.1016/j.wocn.2005.06.001>.
- Snoeren, Natalie D., Juan Segui and Pierre A. Hallé. 2008. On the role of regular phonological variation in lexical access: Evidence from voice assimilation in French. *Cognition* 108. 512–521. <https://doi.org/10.1016/j.cognition.2008.02.008>.
- Stilp, Christian. 2019. Acoustic context effects in speech perception. *WIREs Cognitive Science* e1517. 1–18. <https://doi.org/10.1002/wcs.1517>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemond, et al. 2019. Welcome to the tidyverse. *Journal of Open Source Software* 4(43). 1686. <https://doi.org/10.21105/joss.01686>.

---

**Open Access.** This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited, a link to the CC License is provided, and changes - if any - are indicated. (SID\_1)

