

1. Introduction

It has been long acknowledged that the production and perception 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 (Ganong 1980); similarly, the precision of articulation is inversely proportional to the presence of semantic information (Goldrick et al. 2013; Kitahara et al. 2019). Diachronic phonological processes, for example, are often reported to seek homophony avoidance (see, e.g., Silverman 2012). The question arises whether homophony avoidance is actively present in synchronic language use, too, and how (if at all) it interacts with phonological contrast maintenance or neutralisation.

A number of studies demonstrate that laryngeal processes previously considered to be neutralising (e.g., word-final devoicing, voicing assimilation) are not completely neutralizing phonetically. An underlyingly voiced obstruent often contains more phonation in devoicing contexts than an underlyingly voiceless obstruent, or if this is not the case, other phonetic features like the length of the preceding vowel, or the vowel/consonant duration ratio are systematically different, thereby maintaining the underlying laryngeal contrast (Bárkányi and G. Kiss 2019 provide an overview in Hungarian of this). The present study seeks to explore to what extent a particular *lexical* factor, homophony avoidance, i.e., whether or not a word forms a minimal pair with another word in the lexicon (“minimal pairhood”), affects the realisation of the primary laryngeal feature, the amount of phonation, in the word-final alveolar stops /t/ and /d/ and the fricatives /s/ and /z/ in potentially neutralising and non-neutralising contexts in the speech of Hungarian native speakers. To this end, acoustic experiments were carried out with test words ending in these obstruents in minimal pairs and non-minimal pairs that were placed in various phonetic environments.

* I would like to thank the two anonymous reviewers for their valuable comments and suggestions. This research was supported by grant nr. NKFIH K 142498 of the Hungarian National Research Fund (principal investigator: Péter Szigetvári).

The paper is structured as follows. First, I provide a brief overview of the laryngeal opposition and voicing assimilation of Hungarian obstruents in particular (section 2) and of homophony avoidance in general (section 3). In the second half of the paper, I present the details of the acoustic-production experiments (section 4), and their results (section 5), while I discuss the relevant conclusions of the experiments in section 6.

2. Voicing contrast and voicing assimilation in Hungarian

Based on vocal fold vibration and the timing of supraglottal articulatory gestures, distinct phonetic properties arise that languages use differently in their laryngeal oppositions. Most languages display a binary opposition. In Hungarian, for instance, the two values of the feature $[\pm\text{voice}]$ are in contrast. Because the timing of the abduction and adduction of the vocal folds may vary, several types of voiced and voiceless articulation are possible, and therefore, various types of laryngeal articulation may be associated with the $[\pm\text{voice}]$ feature. In Lisker and Abramson's (1964; 1967) classical work, these types of laryngeal differences in the production of initial stops are based on three phonetic categories according to the onset of periodic vocal fold vibration following the closure phase (referred to as Voice Onset Time, VOT): (i) negative VOT, where phonemically voiced stops are realised with voicing lead like in Hungarian; (ii) short-lag VOT, where phonation starts short after closure release, e.g., lenis stops in English; (iii) long-lag VOT where voicing starts at least 35–60 ms after closure release, e.g., voiceless aspirated stops in English. Languages where the contrast is based on negative VOT vs. zero/short-lag VOT are called *voicing* or *true voice* languages, while languages where the contrast is based on short-lag VOT vs. long-lag VOT are called *aspirating* languages. Hungarian belongs to the former group. According to traditional descriptions (see the summary in Jansen 2004; 2007 for example), the voicing properties of true voice languages are manifested not only in the VOT of stops, but also in that the $[\pm\text{voice}]$ feature of all obstruents (including affricates and fricatives) participates actively in phonological processes such as voicing assimilation, and that voicing contrast is typically preserved in absolute word final (prepausal) position as well.

Phonation, however, cannot be fully maintained in all phonetic contexts. Due to aerodynamic reasons, voicing contrast is fragile and may (partially or completely) disappear in absolute word-final position and before another obstruent. This is especially true for fricatives as they require high supraglottal pressure to maintain turbulent noise and high subglottal pressure to ensure voicing (Ohala 1983). According to the Hungarian descriptive tradition, adjacent obstruents cannot differ in their voice feature, thus within words, as well as across a morpheme or word boundary obstruents must agree in voicing (*akta*

‘file’, *labda* ‘ball’), unless a pause intervenes. Consequently, the laryngeal contrast of Hungarian obstruents is completely neutralised before another obstruent. Thus, while for example, /s/ and /z/ are in contrast word-initially (*szár* /sa:r/ ‘stem’ – *zár* /za:r/ ‘lock’), in intervocalic position (*mész* /me:ʃ/ ‘lime.terminative’ – *mézig* /me:ziɡ/ ‘honey.terminative’) and word-finally (*mész* /me:s/ ‘lime’ – *méz* /me:z/ ‘honey’), the contrast is thought to be completely lost before another obstruent: the /z/ in *méztől* ‘honey.ablative’ is claimed to be phonetically identical to the /s/ in *mésztől* ‘lime.ablative’. The same is true for regressive voicing: the /s/ in *mészből* ‘lime.elative’ is claimed to be phonetically identical to the /z/ in *mézből* ‘honey.elative’, thus, voicing contrast is neutralised (this traditional view is described in, e.g., Siptár & Törkenczy 2000).

- (1) Regressive voicing assimilation: spread of voicing
 /t/+/b/ → [db]: e.g., *hát-ba* ‘back-ill’; *két#barát* ‘two friends’
 /ʃ/+/b/ → [ʒb]: e.g., *hús-ba* ‘meat-ill’; *hús#beszerzése* ‘supply of meat’
- (2) Regressive voicing assimilation: spread of voicelessness
 /b/+/t/ → [pt]: e.g., *láb-tól* ‘foot-abl’, *láb#tisztítása* ‘cleaning of foot’
 /z/+/t/ → [st]: e.g., *víz-től* ‘water-abl’, *víz#tárolása* ‘storage of water’

According to the traditional descriptions, regressive voicing assimilation also affects consonant clusters (i.e., the rule is “iterative”, works from right to left, from one segment to the preceding one):

- (3) Regressive voicing assimilation in consonant clusters
 /st/+/b/ → [zdb]: e.g., *kereszt-ben* ‘cross-iness’
 /gd/+/p/ → [ktp]: e.g., *smaragd#pénzértéke* ‘value of emerald’

Sonorants and vowels do not trigger regressive voicing assimilation in standard Hungarian:

- (4) /p/+/n/ → [pn] (*[bn]): e.g., *kép-nél* ‘picture-ades’
 /s/+/n/ → [sn] (*[zn]): e.g., *rész-nél* ‘part-ades’

Over the past two decades an increasing number of studies have shown that phonological processes believed to be neutralising are often not (completely) neutralising in speech production. In acoustic studies, several authors have pointed out that regressive voicing assimilation in Hungarian is partially contrast preserving. Jansen (2004) found that /k/–/g/ and /ʃ/–/ʒ/ systematically differ in voicing before voiced obstruents, and vowel length before /ʃ/–/ʒ/ is also different. Grácz (2010), examining nonsense words, found that in word-final

position the vowel/consonant duration ratio differs according to the underlying voicing of the consonant. Markó et al. (2010) argue that although voicing assimilation seems to be obligatory, it is a gradient rather than a categorical process (unlike the present paper, the authors also examined environments where there was a pause between the target and the triggering consonant). Bárkányi and G. Kiss (2015) found a significant difference in vowel length before voiced and voiceless fricatives in regressive voicing assimilation contexts. Bárkányi and G. Kiss (2020) also found partial contrast preservation in the voicing of three-member consonant clusters. The authors also showed that stops and fricatives display a different behaviour in assimilation contexts. However, none of these studies address potential lexical effects – such as the existence of close lexical competitors, minimal pairhood, homophony avoidance, or wordedness – in the production of obstruents in Hungarian. The present study aims to fill this gap by examining the voicing of alveolar obstruents in minimal pairs vs. non-minimal pairs in various phonetic environments.

3. Homophony avoidance

It has been long observed that speakers maintain a comfortable buffer zone between a certain value of a segment and its immediate systemic neighbours (Martinet 1952), while listeners tolerate deviations from the intended value and the actual realisation as long as they perceive it as unintended coarticulation due to the phonetic context (e.g., Ohala 1981), a kind of compensatory effect. It is difficult to initiate and maintain voicing in obstruents in word-final and pre-obstruent position, thus keeping such a buffer zone is not easy, which may lead to homophony (like in the above-mentioned case of *méztől–mésztől*). It has also been long acknowledged that languages try to avoid homophony resulting from sound change mostly by morphosyntactic or lexical means (see the classic study of Gilliéron 1910 for French, and Silverman 2021 for Korean, for instance). Wedel et al. (2013) statistically analysed neutralising sound changes in nine languages and concluded that the number of minimal pairs distinguished by a pair of phonemes made significant predictions as to whether the contrast between them would be neutralised. Although the role of homophony avoidance in sound change remains debated (e.g., Sampson 2013), the question arises whether speakers seek to avoid homophony in synchronic language use as well.

In a study of the masculine noun paradigms of Russian, Munteanu (2021) concluded that the language displays a synchronic restriction against homophonous forms within the same paradigm. In nouns where the singular genitive would coincide with the plural nominative and where singular dative and prepositional cases would be identical, stress shift is much more frequent. So, the potentially homophonic forms are distinguished by their prosodic

properties. Yin and White (2018) in an artificial language learning experiment with native English speakers show that learners are less likely to learn neutralising phonological rules than non-neutralising ones, but only if these create homophony between lexical items that came up during learning. In the artificial language in their experiment plural was marked by /i/ which palatalised the final alveolar fricatives and stops of the singular forms. The process was either neutralising or created allophones.

Charles-Luce (1993), investigating regressive voicing assimilation in Catalan, observed that there is more likely to be incomplete neutralisation – as opposed to complete neutralisation – in contexts that would otherwise be semantically ambiguous. The author found that the length of the preceding vowel distinguished voiced and voiceless obstruents significantly more often in minimal pairs than in non-minimal pairs. Kharlamov (2014), examining word-final devoicing in Russian, claims that lexical competition and lexical density play an important role in partial contrast preservation. Thus, in shorter (monosyllabic) words and minimal pairs the author found greater acoustic differences between the voiced and voiceless final obstruents than in longer words and non-minimal pairs. Baese-Berk and Goldrick (2009) point out that word-initial voiceless stops in English are realised with longer VOT in words belonging to minimal pairs (e.g. *cod–god*) than in words that do not have such close competitors in the lexicon. Goldrick et al. (2013) reached a similar conclusion regarding word-final stops: the vowel is much longer before voiced stops in words like *bud* (forming a minimal pair with *but*) than in words with no such lexical neighbour. All these studies suggest that lexical and phonetic-phonological properties closely interact.

No similar studies have been made for Hungarian to the best of my knowledge, and so in this paper I will aim to compare the voicing of final alveolar obstruents in words forming a minimal pair with those which are in words that do not belong to a minimal pair. The hypothesis I will test is that based on the literature briefly overviewed above, in devoicing environments (in absolute word-final position/utterance-finally, and before another voiceless obstruent), the underlyingly voiced obstruents will contain significantly more voicing in a word that forms a minimal pair with another word than in non-minimal pairs (and consequently, the difference in voicing production will not or only partially neutralise). Similarly, I hypothesise that in voicing contexts (before another voiced obstruent), the underlyingly voiceless obstruents will contain significantly less voicing in minimal pairs than in non-minimal pairs – and thus the voicing contrast will be more readily maintained in the minimal pair group.

4. Subjects, material, method

The target consonants of the production experiments were word-final /s/–/z/ and /t/–/d/. The segments of both pairs were analysed in two lexical groups. In the first group, the final consonants in the words did not have existing counterparts with a voiceless or voiced final alveolar stop or fricative, i.e., these words did not form minimal pairs: *szesz* /sɛs/ ‘alcohol’, *mez* /mɛz/ ‘kit’, *net* ‘net’, *led* ‘led lamp’ (i.e., words such as “sze~~z~~” /sɛz/, “me~~z~~” /mɛs/, “ne~~d~~” /nɛd/ and “le~~t~~” /lɛt/ do not exist in Hungarian). I will refer to these words as the “non-minimal pair group” below. The other group consisted of the minimal pair *mész*–*méz* /mɛ:s/–/mɛ:z/ ‘lime’–‘honey’ and *vét*–*véd* /vɛ:t/–/vɛ:d/ ‘make an error’–‘protect’. I will refer to this group as the “minimal pair group”. The two groups involved different participants, and therefore, the group contrasts discussed in the following sections should be interpreted as comparisons between two different sets of participants. The target words were investigated in the following environments:

- (5) a. absolute word-final position (before a pause)
- b. across a word boundary before /p/
- c. across a word boundary before /b/
- d. across a word boundary before the sonorant consonants /m/ and /l/
- e. across a word boundary before the vowel /ɛ/
- f. intervocalically: /ɛ/___/ɛ/ and /e:/___/e:/ (here there was no word boundary after the target consonant)

No significant differences were found between the measured acoustic parameters in the presonorant and prevocalic environments, and so they were placed in the same group, which I will refer to simply as the “presonorant” environment, and will present the results of the statistical analysis for this unified group. In the minimal pair experiment the word pairs for the intervocalic position were *veszek el*–*vezek el* /vɛsɛkɛl/–/vɛzɛkɛl/ ‘I take away’–‘atone’; *vétek*–*védek* /vɛ:tɛk/–/vɛ:dɛk/ ‘I make an error’–‘I protect’. The sentences with the target words that the experiment participants read out can be found in the Appendix.

The non-minimal group was analysed in a previous experiment, whose results were partially published (Bárkányi & G. Kiss 2015; Bárkányi & G. Kiss 2019). However, here I will compare that group with the minimal pair group, and will focus not only on certain environments but all of the above in (5), aiming to provide a more comprehensive picture. The statistical analysis will also be presented in a unified framework for both lexical groups.

In the first experiment involving the non-minimal pairs, six participants took part, while in the minimal pair experiment there were ten subjects. In both production experiments the participants were university students whose age

ranged between 19 and 30 years (means: 21 ± 3.2 years). They read out every sentence (including ten fillers) five times. The sound files of the first round were excluded from the final analysis as these first-round readings are usually less natural due to the unusual experimental circumstances, and there is a greater chance of reading errors. Altogether thus four rounds were used for each participant. Overall then, 4 rounds of 28 sentences of 6 subjects were analysed in the minimal pair group, amounting to 672 data points (for each target sound there were 24 observations in the non-presonorant environments, and 72 in the presonorant one). In the minimal pair group 4 rounds of 28 sentences of 10 subjects were analysed, which amounted to 1120 data points (in this group then, for each target sound there were 40 observations in the non-presonorant environments, and 120 in the presonorant one). The complete data set that was analysed consisted of $672 + 1120 = 1792$ observations.

The sentences were recorded using SpeechRecorder (Draxler & Jänsch 2004). The sentences were randomised by the program, and it was these randomised sentences that the participants read out from a monitor screen. The amount of time available for each sentence was four seconds, which secured a relatively unified speech rate, which was neither too rapid nor too slow. An Audix f50 microphone and an Art USB Dual Pre preamplifier were used to make the recordings in a noise-free room at the Department of English Linguistics of Eötvös Loránd University.

The sound recordings were processed and analysed in Praat (Boersma & Weenink 2021). The segment boundaries and the voicing intervals were marked manually, using the methods discussed in for example G. Kiss (2013). In the case of /t/ and /d/, separate intervals were marked for the closure and (if there was one) the release. For the presence of voicing in the stops, only the closure interval was used, not the release. The boundaries of the fricatives were placed between the start and ending of the constriction phase (visible as aperiodic noise in the spectrograms and waveforms). It is in this interval that the proportion of voicing was measured. So that the presence of vocal fold vibration could be specified more securely, the frequencies above 300–500 Hz (depending on the given participant) were filtered out, the duration of voicing was measured on these filtered waveforms based on the presence of periodic vibrations. The end of the voicing interval was marked when the periodicity was no longer visible. The duration of the intervals was measured automatically with the help of a Praat script, which created the data tables that the statistical analyses used. In this paper, I will only focus on the voicing durations because the length of the pre-target vowels significantly differed in all environments (the underlying vowel was short /ɛ/ in the non-minimal pair group, while it was long /e:/ in the minimal pair group) and thus it was not possible to systematically compare the vowel durations, and based on that, the vowel/consonant duration ratios.

The statistical analysis (including the generation of the various plots) was carried out in *R* (version 4.0.2, R Core Development Team 2020) using various *tidyverse* packages (Wickham et al. 2019), as well as the *patchwork* package (Pedersen 2020) during the composition of the plots. Linear mixed effects models were used to model the data, using the package *lme4* (v. 1.1.27.1, Bates et al. 2015). To specify the *p*-values, the degrees of freedom were calculated using the Satterthwaite approximation available in the *lmerTest* package (v. 3.1.3, Kuznetsova et al. 2017). The fixed effects of the models were the underlying voicing of the obstruents (voiceless vs. voiced) as well as the minimal pairhood (non-minimal pair vs. minimal pair). The random effect structure contained the subjects. Random intercepts and random slopes were fitted for the proportion of voicing varying across participants. If the slopes for subjects did not improve model fit relative to intercepts only, they were removed from the final model, and only random intercepts were retained. If a model did not converge with the default Nelder-Mead optimizer, “BOBYQA” optimizing (Bound Optimization by Quadratic Approximation) was employed. If a model failed to converge even with this setting or if there was no significant variability between subjects for the given acoustic parameter in a given group (“singularity” issue), then the random slope for subjects was taken out of the model, in which case the models always converged. Due to the relatively low number of participants, to avoid further convergence issues, simpler models were fitted, i.e., instead of including three factors (underlying voicing, minimal pairhood, environment), and their interaction, plus including them as by-subject random effects, the models investigated the effect of voicing of the target sound (/t/ vs. /d/; /s/ vs. /z/) and their minimal pairhood in the four environments separately. I will refer to the marginal and conditional R-squared effect sizes below as “R2m” and “R2c” respectively. These values were calculated using the *MuMIn* package (Bartoń 2020). So that the components of the final models can be presented in a tabular format, the *broom.mixed* package (Bolker & Robinson 2021) was used which extracted the model parameters. The tables summarising the linear mixed-effects models in the following sections contain the terminology for the intercept as given by the *lme4* output (“(Intercept)”), the names of the slope coefficients (the *bs*) are “sound” (the sounds compared) and “minpair” (the lexical/minimal pair groups compared).

5. Results

Figure 1 shows the proportions of voicing in the two obstruent pairs in five environments, separately for non-minimal pairs and minimal pairs. The same data have been rearranged in Figure 2 in a way that the underlyingly voiceless sounds and their voiced counterparts appear in separate rows, while the proportions of voicing in a given sound in non-minimal pairs and in minimal pairs are shown together so that the sound pairs can be compared visually more easily with respect to their lexical group membership. The corresponding descriptive statistics can be found in Table 1. The detailed results for the different environments will be presented in the following sections: absolute word-final position (section 5.1), before /p/ (5.2), before /b/ (5.3), before the sonorants (5.4), and between two vowels (5.5).

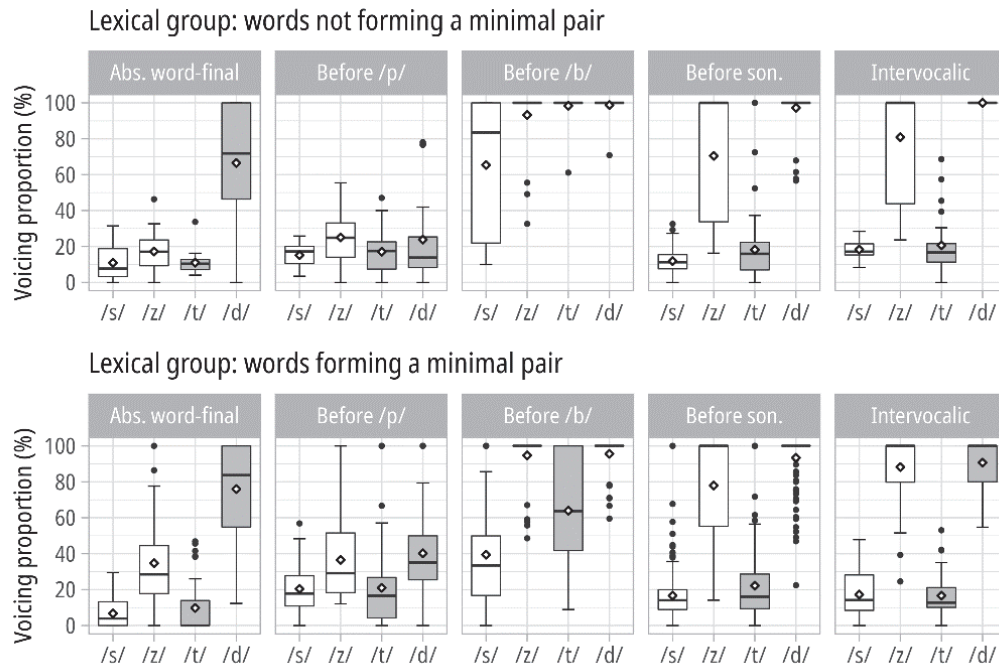


Figure 1: Proportion of voicing in /s/, /z/, /t/ and /d/ (the rectangles in the boxplots represent the means)

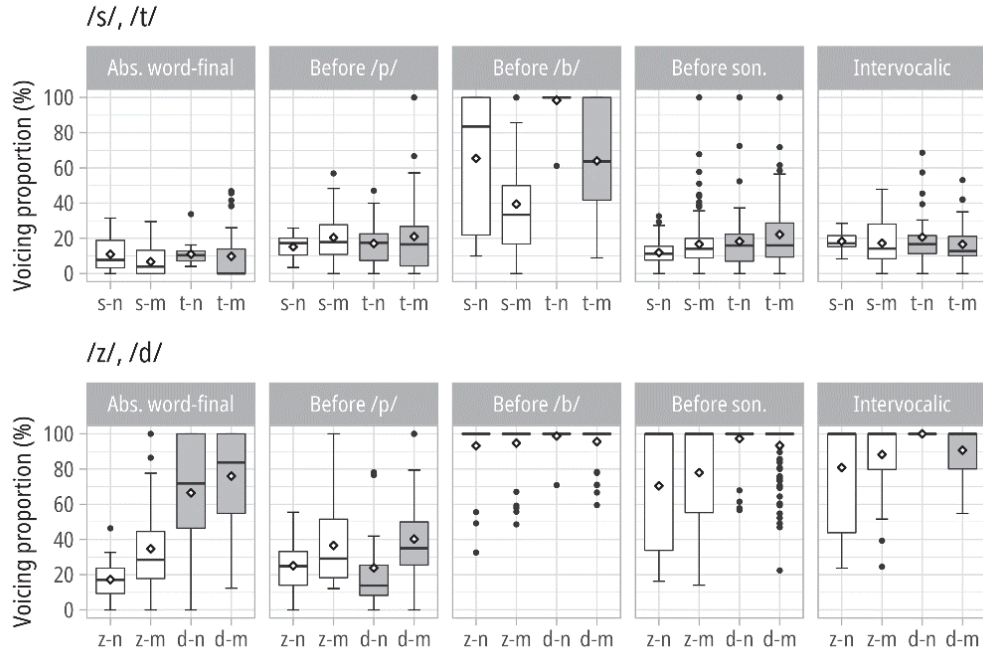


Figure 2: Proportion of voicing in /s/, /z/, /t/ and /d/ in non-minimal pairs and in minimal pairs. Abbreviations: “s-n, z-n, t-n, d-n” = word-final and intervocalic /s z t d/ in words that do **not** form a minimal pair with another word (*szesz, mez, net, led*); “s-m, z-m, t-m, d-m” = word-final and intervocalic /s z t d/ in words that form a **minimal pair** with another word (*mész, méz, vét, véd, veszek el, vezek el*); the rectangles in the boxplots represent the means.

Table 1: Proportion of voicing in /s/, /z/, /t/, /d/ – descriptive statistics (“non-mp” = word not belonging to a minimal pair; “mp” = word belonging to a minimal pair)

Lexical group	Environment	Sound	Mean	SD	Median	Min	Max	SE
non-mp	abs. word-final	/s/	10.95	9.83	7.69	0.00	31.50	2.01
		/z/	17.23	10.65	17.09	0.00	46.34	2.17
		/t/	10.99	6.10	10.36	0.08	33.72	1.25
		/d/	66.49	32.67	71.84	0.00	100.00	6.67
	before /p/	/s/	15.13	6.65	17.23	3.45	25.82	1.36
		/z/	25.07	13.72	24.87	0.00	55.51	2.80
		/t/	17.09	12.78	17.50	0.00	47.06	2.61
		/d/	23.81	26.11	13.86	0.00	78.12	5.33
	before /b/	/s/	65.39	38.06	83.50	9.98	100.00	7.77
		/z/	93.22	18.65	100.00	32.59	100.00	3.81
		/t/	98.38	7.94	100.00	61.11	100.00	1.62
		/d/	98.78	5.95	100.00	70.83	100.00	1.21

(continued on next page)

Lexical group	Environment	Sound	Mean	SD	Median	Min	Max	SE
non-mp	presonorant	/s/	11.89	6.57	11.26	0.00	32.57	0.77
		/z/	70.47	33.09	100.00	16.27	100.00	3.90
		/t/	18.23	18.87	15.91	0.00	100.00	2.22
		/d/	97.24	10.24	100.00	56.67	100.00	1.21
	intervocalic	/s/	18.28	5.35	17.12	8.32	28.46	1.09
		/z/	80.87	30.73	100.00	23.71	100.00	6.27
		/t/	20.64	16.87	16.66	0.00	68.63	3.44
		/d/	100.00	0.00	100.00	100.00	100.00	0.00
mp	abs. word-final	/s/	6.71	8.06	3.92	0.00	29.50	1.27
		/z/	34.71	24.42	28.51	0.00	100.00	3.86
		/t/	9.76	14.36	0.00	0.00	46.84	2.27
		/d/	76.00	27.53	83.75	12.31	100.00	4.35
	before /p/	/s/	20.42	13.33	17.81	0.00	56.86	2.11
		/z/	36.57	22.14	29.16	12.12	100.00	3.50
		/t/	21.00	23.83	16.55	0.00	100.00	3.77
		/d/	40.24	21.20	35.00	0.00	100.00	3.35
	before /b/	/s/	39.43	26.28	33.41	0.00	100.00	4.16
		/z/	94.71	14.33	100.00	48.57	100.00	2.27
		/t/	63.96	32.05	63.68	8.93	100.00	5.07
		/d/	95.61	10.89	100.00	59.52	100.00	1.72
	presonorant	/s/	16.70	14.08	14.07	0.00	100.00	1.29
		/z/	77.97	25.01	100.00	14.06	100.00	2.28
		/t/	22.18	21.21	15.99	0.00	100.00	1.94
		/d/	93.42	14.43	100.00	22.41	100.00	1.32
	intervocalic	/s/	17.27	12.96	14.16	0.00	47.83	2.05
		/z/	88.25	21.02	100.00	24.53	100.00	3.32
		/t/	16.62	12.29	12.71	0.00	53.06	1.94
		/d/	90.70	13.81	100.00	54.76	100.00	2.18

5.1. Absolute word-final position

As Figure 1 and Table 1 show, the mean proportion of voicing in /z/ and /d/ in absolute word-final position is higher than that of their voiceless counterpart. The difference is significantly greater between /t/ and /d/ in both lexical groups, and it is greater between /s/ and /z/ in the case of minimal pairs, i.e., while in the case of the non-minimal pair (*szesz–mez*) the difference between the fricatives is small (which is indicative of neutralisation, at least as far as this acoustic parameter is concerned), in the case of the minimal pair (*mész–mész*), the difference is larger (only 6.71% voicing on average with 8.06% standard deviation in /s/, while 34.71% with 24.42% standard deviation in /z/). This larger difference was confirmed by the linear mixed effects models, too (see Table 2). In the case of the non-minimal pair the difference between the voicing proportions of /s/–/z/ was non-significant, whereas it was significant in the case of the minimal pair. The marginal effect size R²_m for the minimal pair was 0.38, the conditional effect size R²_m was 0.81, that is, the fixed effect (the underlying voicing

of the fricative) largely explains the data, while the random effects (random intercept and slope by subject) significantly improve the model's total explanatory power. The difference between the mean voicing proportions of /t/–/d/ was significant in both lexical groups and the effect sizes were also relatively high (Table 2), i.e., both the fixed and the random effects significantly explain the variance.

The /d/ tokens in absolute word-final position had similar voicing proportions in both groups (*led*, *véd*) (Figure 2), their difference was not significant according to the model fitted (Table 2). This, however, was not the case for /z/ in this environment: the mean voicing proportion in the minimal pair /z/ was significantly larger than in the non-minimal pair /z/ (only $17.23 \pm 10.65\%$ voicing in *mez*, but $34.71 \pm 24.42\%$ in *méz*). That is, minimal pair /z/ is significantly different not only from its voiceless counterpart but also from its non-minimal pair counterpart (it contains more voicing than either). R2c was high (0.72), while R2m was lower (0.15), which suggests that the random intercept and slope by subjects fitted to the data largely contribute to the explanatory power of the model; however, in addition to the fixed effect (belonging to a minimal pair or not) other factors also contribute to /z/ having more voicing in absolute final position, but minimal pairhood also contributes to this effect.

Overall then we can say that in absolute word-final position, the difference in the voicing proportions of the obstruent pairs examined is maintained, except between /s/ and /z/ if they are parts of words that do not form a minimal pair. We can thus observe a strong minimal pairhood/allophony avoidance effect in this environment.

Table 2: Summary of the mixed effects models, outcome variable: voicing proportion, environment: **absolute word-final** (“non-mp” = word not belonging to a minimal pair; “mp” = word belonging to a minimal pair)

Contrast	Coefficient	<i>b</i>	SE	<i>t</i>	df	<i>p</i>	R2m	R2c
/s/–/z/, non-mp	(Intercept)	10.95	3.04	3.60	6.00	0.0113	0.09	0.46
	sound-z	6.28	4.30	1.46	6.00	0.1949		
/t/–/d/, non-mp	(Intercept)	10.99	4.03	2.73	27.94	0.011	0.6	0.72
	sound-d	55.51	8.43	6.59	7.03	0.0003		
/s/–/z/, mp	(Intercept)	6.71	2.05	3.27	10.00	0.0084	0.38	0.81
	sound-z	28.00	6.32	4.43	10.00	0.0013		
/t/–/d/, mp	(Intercept)	9.76	3.56	2.74	10.00	0.0209	0.7	0.88
	sound-d	66.24	5.84	11.34	10.00	<.0001		
non-mp–mp, /z/	(Intercept)	17.23	3.05	5.65	6.00	0.0013	0.15	0.72
	minpair-mp	17.49	7.38	2.37	13.58	0.0332		
non-mp–mp, /d/	(Intercept)	66.49	8.16	8.15	6.00	0.0002	0.02	0.45
	minpair-mp	9.51	10.98	0.87	14.10	0.4011		

5.2. Before /p/

Before /p/ (where voicing neutralisation is expected in the direction of devoicing) the mean voicing proportion of the underlyingly voiced obstruents was always larger than that of their voiceless counterparts, similarly to the absolute word-final position (see Figure 1 and Table 1). This difference between the means turned out to be significant for the fricatives in both lexical groups (see Table 3); however, the size of the effect was relatively small: marginal R-squared was low in both groups (R2m values: 0.18 and 0.17), conditional R-squared was smaller in the non-minimal group (R2c = 0.41; minimal pair: R2c = 0.53). This suggests that the size of the difference is smaller in the case of the non-minimal pair (*szesz–mez*).

The mean voicing proportion measured in /z/ in the minimal pair group was greater than in the non-minimal group. In non-minimal pair *mez*, voicing proportion ranged between 0% and 55.51% (mean: 25.07%, SD: 13.72%), while in minimal pair *mész* there were no values below 12.12%, and there were data points well over 55%, for example, fully voiced /z/ before /p/ also occurred in the data (mean: 36.57±22.14%). However, the difference between the two groups did not turn out to be significant based on the fitted mixed effects model (Table 3). These results suggest that the difference between /s/ and /z/ is significant within both groups but in the minimal pair group (*mész–méz*) the magnitude of the difference is even larger, i.e., we can observe the effect of homophony avoidance before /p/ just like in the absolute word-final position.

As far as the stops are concerned before /p/, the voicing proportion difference of /t/ and /d/ in the non-minimal pair group (*net–led*) did not turn out to be significant (Table 3). However, in the minimal pair group (*vét–véd*), the difference was significant. Thus, in the environment before /p/ – a potentially voicing neutralising environment – the mean proportion of voicing of /d/ in *véd* was significantly greater than in *vét* (/t/: 21.00±23.83%, /d/: 40.24±21.20%). The magnitude of the fixed effect was relatively low (R2m = 0.16) but this factor definitely contributes to the observed differences, as well as the random effects (R2c = 0.46).

The voicing proportions measured in /d/ before /p/ in the two lexical groups were different, the mean was higher in the minimal pair group, just like in the case of /z/ (non-minimal pairs: 23.81±26.11%, minimal pairs: 40.24±21.20%); however, this difference was not significant (Table 3) just as it was not significant for pre-/p/ /z/. Similarly to /z/ then, in the minimal pair group, /d/ contained more voicing before /p/ than in the non-minimal pair group, and this additional amount of voicing was enough to bring about a significant difference within the minimal pair group. Thus, we can definitely observe the effect of minimal pairhood in the case of the stops, too.

Table 3: Summary of the mixed effects models, outcome variable: voicing proportion, environment: **before /p/** (“non-mp” = word not belonging to a minimal pair; “mp” = word belonging to a minimal pair)

Contrast	Coefficient	<i>b</i>	SE	<i>t</i>	df	<i>p</i>	R2m	R2c
/s/–/z/, non-mp	(Intercept)	15.13	2.92	5.18	9.21	0.0005	0.18	0.41
	sound-z	9.94	2.59	3.84	42.00	0.0004		
/t/–/d/, non-mp	(Intercept)	17.09	3.25	5.26	6.00	0.0019	0.03	0.75
	sound-d	6.72	10.14	0.66	6.00	0.5319		
/s/–/z/, mp	(Intercept)	20.42	3.08	6.63	10.00	0.0001	0.17	0.53
	sound-z	16.15	4.40	3.67	10.00	0.0043		
/t/–/d/, mp	(Intercept)	21.00	5.47	3.84	10.00	0.0033	0.16	0.46
	sound-d	19.24	5.16	3.73	10.00	0.0039		
non-mp–mp, /z/	(Intercept)	25.07	6.03	4.16	16.00	0.0007	0.08	0.5
	minpair-mp	11.50	7.63	1.51	16.00	0.1513		
non-mp–mp, /d/	(Intercept)	23.81	7.69	3.10	16.00	0.0069	0.11	0.63
	minpair-mp	16.43	9.73	1.69	16.00	0.1106		

5.3. Before /b/

Before /b/, a potentially voicing environment, underlyingly voiceless /s/ contained less voicing on average than its voiced counterpart in both lexical groups. Figure 1 shows how wide a range values populated in the case of the non-minimal pair (*szesz–mez*) (we can observe values between 9.98% and 100%, with a mean of $65.39 \pm 38.06\%$). /s/ contained even less voicing before voiced /b/ in the minimal pair group (*mész–mész*), here in the lower quarter we can even find completely voiceless values (mean: $39.43 \pm 26.28\%$). Underlyingly voiced /z/, as expected, was largely voiced in this position (non-minimal pairs: $93.22 \pm 18.65\%$, minimal pairs: $94.71 \pm 14.33\%$). Based on this, it is not surprising that the mean voicing proportions of /s/ vs. /z/ were significantly different in both lexical groups (see Table 4). The effect size was larger in the minimal pair group based on the R-squared values. This indicates that contrast preservation in the minimal pair group is more likely than in the non-minimal pair group.

If we compare /s/ in the non-minimal group (*szesz*) with /s/ in the minimal group (*mész*), we find that the latter was produced by the participants with less voicing on average before /b/, that is, minimal pairhood seems to decrease the proportion of voicing (non-minimal pairs: $65.39 \pm 38.06\%$, minimal pairs: $39.43 \pm 26.28\%$). Based on the fitted mixed effects models, the difference was close to being significant ($p = 0.0575$). We note that if instead of Satterthwaite approximation Wald approximation was used to calculate the degrees of freedom (using the *parameters* function of the *parameters* R-package, see Makowski et al. 2021), the value of p was 0.045. Based on the R-squared values, we can say that in addition to the fixed effect (minimal pairhood) other factors also affect the variability of /s/’s voicing but this variable also contributes to it;

on the other hand, the model's total explanatory power is relatively large. Overall then, the voicing proportion in /s/ did not only show significant difference before /b/ within the lexical groups (it contained much less voicing compared to /z/) but it was also significant between the lexical groups (minimal pair /s/ contained much less voicing). That is, the minimal pair effect can be observed doubly.

The results were interesting for the pre-/b/ stops. While in the non-minimal pair group (*net*–*led*) both sounds were produced almost always 100% voiced, in the minimal pair group (*vét*–*véd*) /t/ contained much less voicing, as the boxplot in Figure 1 shows. In this group the values ranged between 9.93% and 100%, with a mean of $63.96 \pm 32.05\%$. The fitted models confirmed these observations (Table 4). The difference between /t/ and /d/ was not significant in the non-minimal pair group, but the proportion of voicing was significant in these sounds in the minimal pair group.

The amount of voicing in /t/ before /b/ clearly differed in the two lexical groups (see Figure 2): the /t/ in *vét* (which forms a minimal pair with *véd*) contained much less voicing than the /t/ in *net*, which does not form a minimal pair with another word (means: *net*: $98.38 \pm 7.94\%$, *vét*: $63.96 \pm 32.05\%$). This difference turned out to be significant, too, with relatively large effect sizes.

Overall then, we can say that there is a strong homophony-avoidance effect for both the fricatives and the stops before /b/; in essence, the effect is responsible for maintaining the underlying laryngeal contrast.

Table 4: Summary of the mixed effects models, outcome variable: voicing proportion, environment: **before /b/** (“non-mp” = word not belonging to a minimal pair; “mp” = word belonging to a minimal pair)

Contrast	Coefficient	<i>b</i>	SE	<i>t</i>	df	<i>p</i>	R2m	R2c
/s/–/z/, non-mp	(Intercept)	65.39	7.23	9.04	11.63	<.0001	0.19	0.31
	sound-z	27.83	7.79	3.57	42.00	0.0009		
/t/–/d/, non-mp	(Intercept)	98.38	1.40	70.16	48.00	<.0001	0	0
	sound-d	0.40	1.98	0.20	48.00	0.839		
/s/–/z/, mp	(Intercept)	39.43	7.13	5.53	10.00	0.0003	0.64	0.83
	sound-z	55.28	7.13	7.75	10.00	<.0001		
/t/–/d/, mp	(Intercept)	63.96	8.44	7.58	10.00	<.0001	0.31	0.74
	sound-d	31.65	7.69	4.12	10.00	0.0021		
non-mp–mp, /s/	(Intercept)	65.39	10.03	6.52	16.00	<.0001	0.15	0.59
	minpair-mp	–25.96	12.68	–2.05	16.00	0.0575		
non-mp–mp, /t/	(Intercept)	98.38	8.66	11.36	16.00	<.0001	0.3	0.72
	minpair-mp	–34.42	10.96	–3.14	16.00	0.0063		

5.4. Before sonorants

In this environment (which, as we said above, also contained the prevocalic position) voicing neutralisation was not expected, and this expectation was confirmed by the results (see Figure 1 and Table 1). /s/ and /z/ significantly differed

with respect to the voicing ratio within both lexical groups, and the effect size was also substantial (Table 5). It is interesting to note that the voicing of underlyingly voiced /z/ displayed a relatively large variation in this phonetically optimal environment for voicing contrast maintenance: there were subjects that almost always produced /z/ here almost voiceless. On the whole, however, the voicing proportions of /s/ and /z/ were saliently different (non-minimal pair: $11.89 \pm 6.57\%$ vs. $70.47 \pm 33.09\%$; minimal pair: $16.70 \pm 14.08\%$ vs. $77.97 \pm 25.01\%$).

No minimal pairhood effect was observed in this environment, i.e., the minimal pair membership did not significantly affect the proportion of voicing: /s/ and /t/ were similarly voiceless in both groups, and /z/ and /d/ similarly voiced (see Figure 2 and Table 5).

Table 5: Summary of the mixed effects models, outcome variable: voicing proportion, environment: **presonorant** (“non-mp” = word not belonging to a minimal pair; “mp” = word belonging to a minimal pair)

Contrast	Coefficient	<i>b</i>	SE	<i>t</i>	df	<i>p</i>	R2m	R2c
/s/–/z/, non-mp	(Intercept)	11.89	3.63	3.28	11.42	0.007	0.61	0.63
	sound-z	58.58	3.82	15.32	138.00	<.0001		
/t/–/d/, non-mp	(Intercept)	18.23	2.49	7.32	10.13	<.0001	0.87	0.88
	sound-d	79.01	2.40	32.92	138.00	<.0001		
/s/–/z/, mp	(Intercept)	16.70	2.12	7.86	10.00	<.0001	0.7	0.76
	sound-z	61.27	4.13	14.85	10.00	<.0001		
/t/–/d/, mp	(Intercept)	22.18	4.44	4.99	10.00	0.0005	0.8	0.87
	sound-d	71.24	4.04	17.64	10.00	<.0001		
non-mp–mp, /s/	(Intercept)	11.89	2.21	5.39	16.00	0.0001	0.04	0.17
	minpair-mp	4.80	2.79	1.72	16.00	0.1047		
non-mp–mp, /t/	(Intercept)	18.23	5.07	3.59	16.00	0.0024	0.01	0.33
	minpair-mp	3.95	6.42	0.62	16.00	0.5472		
non-mp–mp, /z/	(Intercept)	70.47	6.43	10.95	6.00	<.0001	0.02	0.21
	minpair-mp	7.49	7.74	0.97	11.22	0.3533		
non-mp–mp, /d/	(Intercept)	97.24	2.93	33.19	16.00	<.0001	0.02	0.26
	minpair-mp	–3.82	3.71	–1.03	16.00	0.3176		

5.5. Intervocalic position

Similarly to the presonorant position, no voicing neutralisation was expected in the word-internal intervocalic position, and this was confirmed by the results (see Figure 1). The voicing difference between /s/–/z/ and /t/–/d/ was significant with a large effect size in this environment (Table 6). Just like before sonorants, the voicing proportion in /z/ displayed relatively large variation, especially in the non-minimal pair group, but despite this, the average was around 80% and so the values showed a clear separation from those of /s/ (non-minimal pairs: $18.28 \pm 5.35\%$ vs. $80.87 \pm 30.73\%$; minimal pairs: $17.27 \pm 12.96\%$ vs. $88.25 \pm 21.02\%$).

Only /d/ displayed a minimal pairhood effect: it was significantly less voiced in the minimal pair group than in the non-minimal pair group. However, this result is not surprising considering that only 100% voiced /d/ tokens were found in the non-minimal pair group and so only a slight deviation from this proportion can result in a significant difference. And indeed, despite the statistically significant difference, the effect sizes were relatively small (see the R-square values in Table 6). The mean proportion of voicing of /d/ in the minimal pair group was also rather large ($90.70 \pm 13.81\%$); therefore, the difference between the /d/'s in the two groups is in fact small.

Table 6: Summary of the mixed effects models, outcome variable: voicing proportion, environment: **intervocalic** (“non-mp” = word not belonging to a minimal pair; “mp” = word belonging to a minimal pair)

Contrast	Coefficient	<i>b</i>	SE	<i>t</i>	df	<i>p</i>	R2m	R2c
/s/–/z/, non-mp	(Intercept)	18.28	5.67	3.23	10.15	0.0089	0.68	0.75
	sound-z	62.59	5.51	11.36	42.00	<.0001		
/t/–/d/, non-mp	(Intercept)	20.64	2.58	8.01	15.86	<.0001	0.92	0.93
	sound-d	79.36	3.28	24.23	42.00	<.0001		
/s/–/z/, mp	(Intercept)	17.27	2.58	6.70	10.00	0.0001	0.81	0.88
	sound-z	70.97	4.75	14.95	10.00	<.0001		
/t/–/d/, mp	(Intercept)	16.62	3.03	5.49	10.00	0.0003	0.89	0.93
	sound-d	74.08	3.34	22.21	10.00	<.0001		
non-mp–mp, /s/	(Intercept)	18.28	2.76	6.63	16.00	<.0001	0	0.21
	minpair-mp	–1.01	3.49	–0.29	16.00	0.7756		
non-mp–mp, /t/	(Intercept)	20.64	3.93	5.25	16.00	0.0001	0.02	0.32
	minpair-mp	–4.02	4.98	–0.81	16.00	0.4306		
non-mp–mp, /z/	(Intercept)	80.87	9.11	8.88	6.00	0.0001	0.02	0.44
	minpair-mp	7.38	10.42	0.71	9.72	0.4956		
non-mp–mp, /d/	(Intercept)	100.00	2.94	34.01	16.00	<.0001	0.15	0.37
	minpair-mp	–9.30	3.72	–2.50	16.00	0.0236		

6. Discussion

This paper hypothesised that in potentially devoicing environments (in absolute word-final position and before voiceless obstruents), the underlyingly voiced obstruents will contain more voicing in the case of minimal pairs than in the case of non-minimal pairs, and consequently, minimal pairs are less likely to completely neutralise in speech production. Similarly, the underlyingly voiceless obstruents in minimal pairs are assumed to be less voiced before voiced obstruents than in non-minimal pairs, thus the former group is more likely to preserve the voicing contrast. This hypothesis was indeed supported by the production experiments: the amount of voicing in /s/–/z/ and /t/–/d/ was systematically different between the minimal pair and non-minimal pair group word-finally and in regressive voicing assimilatory contexts. Minimal pairhood clearly acted against voicing neutralisation in the following ways.

In utterance-final position, the fricatives in the non-minimal pairs did not differ in voicing, while in the minimal pairs they did. To initiate and maintain voicing in fricatives requires active articulatory effort since simultaneous turbulent noise and vocal fold vibration are aerodynamically difficult (see, e.g., Stevens 1998). According to Myers (2012), if word-final devoicing appears in a language, it generally starts with fricatives in word-/utterance-final position and propagates over time to other obstruents and other domain-final environments (utterance-final > word-final > syllable-final position). Hungarian fricatives in non-minimal pairs might have taken the first step towards word-final obstruent devoicing as both /s/ and /z/ were produced with little voicing. Minimal pairs, however, seem to defy this as the voiced–voiceless categories were clearly kept apart in the production experiments. This suggests that lexical factors, such as homophony avoidance, can override the aerodynamically based phonetic effect of devoicing in final position.

While the utterance-final position triggers devoicing in phonetic terms, the position before voiceless obstruents – in this study across a word boundary before /p/ – triggers devoicing also in phonological terms and is expected to create homophony. The acoustic analysis showed that the voicing contrast between /s/–/z/ was attested in both lexical groups, but it was more pronounced in minimal pairs than in non-minimal pairs. As far as /t/ and /d/ are concerned, the difference between them was actually clearly maintained in words forming a minimal pair. This indicates that homophony avoidance counteracted phonetic/ aerodynamic effects (i.e., the fact that maintaining voicing is relatively difficult before another obstruent) and phonological rule application in this environment, too.

In the third potentially neutralising environment – across a word boundary before /b/ – both lexical and phonetic effects could be observed, and both counteracted complete neutralisation. Underlyingly voiceless /s/ in the non-minimal group did not become fully voiced even in this environment that favours phonetic voicing. This effect, which can be explained with aerodynamic reasons again, was further enhanced by the lexical effect since in minimal pairs /s/ was even less voiced. The most salient homophony-avoidance effect was observed in the case of /t/–/d/: while in non-minimal pairs the difference was neutralised (both were voiced to a similar degree), the difference was upheld in the minimal pairs in spite of the fact that maintaining devoicing before a voiced obstruent is relatively difficult phonetically.

The voicing-contrast maintenance in the minimal pair group in the three potentially neutralising environments is in accordance with the H&H theory (Lindblom 1990), according to which speakers take into account the purpose of speech production and the specifics of the communication situation (in this case the risk of ambiguity caused by the presence of minimal pairs), and adjust their

speech accordingly, which may lead to hyper-articulation. The speaker's aim to maintain contrast of course does not necessarily mean that listeners will always perceive the intended differences (see, e.g., Costa and Mattingly's 1981 description of an Eastern New England dialect of English whose speakers make a systematic distinction in vowel duration for the words *cod* and *card* despite the fact that they are unable to discriminate between the two tokens in perception experiments). The fact that the experiments presented in this paper included one-syllable words may also have contributed to the observed differences. Kharlamov (2014) demonstrates that word reading, the presence of minimal pairs, and short, monosyllabic words are more likely to induce the partial contrast preservation of laryngeal features.

Naturally, the question arises whether or not the measured acoustic differences are mirrored in perception, and if they are, to what degree. The role of perception in phonological contrast and its neutralisation is well known. This paper has brought up evidence that voicing differences in speech production/acoustics can remain in neutralising environments due to lexical reasons such as homophony avoidance; however, this does not necessarily mean that these acoustic differences will translate into perceptual – and consequently phonological – differences. There has been some perceptual research involving minimal pairs (e.g., Bárkányi & G. Kiss 2019; 2021) but future research must look into their systematic comparison with non-minimal pairs.

The findings in this paper provide further evidence for phonetically-based, functional phonological models according to which there is a direct link between phonetics (speech production/perception), phonology, and grammar, unlike in representational models which exclude such an active interface, and which therefore cannot adequately explain the influence of extra-grammatical factors such as homophony avoidance or the aerodynamics of voicing production on phonological processes (such as partial voicing neutralisation). Finally, these results highlight the importance of lexical factors in experimental design, too: choosing the type of lexical item can greatly influence the phonetic implementation of the sounds it contains. Ignoring such lexical factors can lead to misleading results.

Appendix

The test sentences used in the production experiments were the following (the test words are marked with bold):

Minimal pair group

1. Addigra a **mész** már régen elfogyott.
2. Gyógyításra a **mész** lehet a legjobb.
3. A **mész** elrablásával foglalkozott az egész sajtó.
4. A **mész** pénzértéke ezután csökkenni kezdett.
5. Sajnos a **mész** belefolyt a szemébe.
6. Sokféle felhasználásra alkalmas a **mész**.
7. Addigra a **méz** már régen elfogyott.
8. Gyógyításra a **méz** lehet a legjobb.
9. A **méz** elrablásával foglalkozott az egész sajtó.
10. A **méz** pénzértéke ezután csökkenni kezdett.
11. Sajnos a **méz** belefolyt a szemébe.
12. Sokféle felhasználásra alkalmas a **méz**.
13. Mindig **vezekel** az összes regényhős.
14. Egy inget **veszek el** a közös szekrényből.
15. Azóta **vét** minden idegen szabály ellen.
16. Gyakran **vét** legalább két intézkedés ellen.
17. Biztos nem **vét** ellenük ilyen módon.
18. Ez a műszer mindig **vét** párás időben.
19. A berendezés **vét** borús időben is.
20. A jól felszerelt készülék is **vét**.
21. Azóta **véd** minden idegen szabály ellen.
22. Gyakran **véd** legalább két intézkedés ellen.
23. Biztos nem **véd** ellenük ilyen módon.
24. Ez a műszer mindig **véd** párás időben.
25. A berendezés **véd** borús időben is.
26. A jól felszerelt készülék is **véd**.
27. Ezt ma **vétek** lenne elszalasztanunk.
28. Ma én **védek** a barátságos meccsen.

Non-minimal pair group

1. Azóta **net** már van a lakásban.
2. A **net** lehetett az oka a megakadásnak.
3. A **net** este mindig sokkal lassabb.

4. Egy **net** probléma lépett fel.
5. A **net** beállításokon múlik az egész.
6. A végére meg elromlott a **net**.
7. A **netet** tartják az évezred találmányának.
8. A vörös **led** már megint nem ég.
9. Egy kis **led** lámpát szereltek az oldalára.
10. A világító **led** erre nagyon jó.
11. Vészhelyzetben a **led** pirosan ég.
12. A **led** biztosítja a sötétben a világítást.
13. Semmi más nem világított csak a **led**.
14. A **ledet** kiszorítják a fénycsöves lámpák.
15. A bulin a **szesz** már éjfélre elfogyott.
16. Fertőtlenítésre a **szesz** lehet a legjobb.
17. Ilyenkor a **szesz** eredete a kérdés.
18. A **szesz** pirosra színezte a főzetet.
19. Sajnos a **szesz** belefolyt a szemébe.
20. Ipari felhasználásra is alkalmas a **szesz**.
21. A **szesz**es italok körében jól ismert.
22. Az új **mez** már ott várta a játékosokat.
23. Futás közben a **mez** lecsúszott a válláról.
24. A **mez** előtt hevert a stoplis cipő.
25. A futball **mez** párosával van csomagolva.
26. A foci **mez** belseje mikroszálas.
27. Szerencsét hozott a válogatott **mez**.
28. A játékos **meze** kétszer annyiért kelt el.

References

- Baese-Berk, Melissa and Matthew Goldrick. 2009. Mechanisms of interaction in speech production. *Language and Cognitive Processes* 24 (4): 527–54.
<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 & Philadelphia: John Benjamins. 65–94. <https://doi.org/10.1075/atoh.14.03bar>
- Bárkányi, Zsuzsanna and G. Kiss Zoltán. 2019. A fonetikai korrelátumok szerepe a zöngékонт-
raszt fenntartásában. *Beszédprodukciós és észleléses eredmények* [The role of phonetic
correlates in voicing contrast. Results from speech production and perception]. *Általános
Nyelvészeti Tanulmányok* 31: 57–102.

- 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 Zoltán G. Kiss. 2021. The perception of voicing contrast in assimilation contexts in minimal pairs: Evidence from Hungarian. *Acta Linguistica Academica* 68: 207–229. <https://doi.org/10.1556/2062.2021.00473>
- 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 Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67: 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Boersma, Paul and David Weenink. 2021. Praat: Doing phonetics by computer. Computer programme. Version 6.2.01.
- Bolker, Ben and David Robinson. 2021. broom.mixed: Tidying methods for mixed models. <https://CRAN.R-project.org/package=broom.mixed>. R package version 0.2.7.
- Charles-Luce, Jan. 1993. The effects of semantic context on voicing neutralisation. *Phonetica* 50: 28–43. <https://doi.org/10.1159/000261924>
- Costa, Paul and Ignatius Mattingly. 1981. Production and perception of phonetic contrast during phonetic change. Haskins Laboratories Status Report on Speech Research SR 67/68. 191–196. <https://doi.org/10.1121/1.386167>
- Draxler, Christoph and Klaus Jänsch. 2004. SpeechRecorder – A universal platform independent multi-channel audio recording software. Proceedings of the 4th International Conference On Language Resources And Evaluation, Lisbon. 559–562.
- G. Kiss, Zoltán. 2013. Measuring acoustic correlates of voicing in stops and fricatives. In: Péter Szigetvári (ed.). VLLXX: Papers presented to László Varga on his 70th birthday. Budapest: Department of English Linguistics, Eötvös Loránd University & Tinta Könyvkiadó/Tinta Publishing House. 289–311.
- Ganong, William F. 1980. Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 6(1): 110–125. <http://doi.org/10.1037/0096-1523.6.1.110>
- Gilliéron, Jules. 1910. Étude de géographie linguistique XII – mots en collision. Le *coq* et le *chat*. *Revue de philologie française* 4: 278–288.
- Goldrick, Matthew, Charlotte Vaughn and Amanda Murphy. 2013. The effects of lexical neighbors on stop consonant articulation. *The Journal of the Acoustical Society of America* 134(2): 172–177. <https://doi.org/10.1121/1.4812821>
- Gráci, Tekla Etelka. 2010. A spiránsok zöngésségi oppozíciójának néhány jellemzője [Some properties of the voicing opposition of spirants]. *Beszédkutatás* 2010: 42–56.
- Jansen, Wouter. 2004. Laryngeal contrast and phonetic voicing: A laboratory phonology approach to English, Hungarian, and Dutch. Doctoral dissertation. Rijksuniversiteit Groningen.
- Jansen, Wouter. 2007. Phonological ‘voicing,’ phonetic voicing and assimilation in English. *Language Sciences* 29: 270–293. <https://doi.org/10.1016/j.langsci.2006.12.021>

- Kharlamov, Viktor. 2014. Incomplete neutralisation of the voicing contrast in word-final obstruents in Russian: Phonological, lexical, and methodological influences. *Journal of Phonetics* 43: 47–56. <https://doi.org/10.1016/j.wocn.2014.02.002>
- Kiss, Zoltán. 2007. The phonetics–phonology interface: Allophony, assimilation and phonotactics. Doctoral dissertation. Eötvös Loránd University (ELTE).
- Kiss, Zoltán and Zsuzsanna Bárkányi. 2006. A phonetically-based approach to the phonology of [v] in Hungarian. *Acta Linguistica Hungarica* 53: 175–226. <https://doi.org/10.1556/ALing.53.2006.2-3.4>
- 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. Canberra: Australasian Speech Science and Technology Association. 2749–2752.
- Kuznetsova, Alexandra, Per B. Brockhoff and Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82: 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H&H theory. In: William J. Hardcastle and Alain Marchal (ed.). *Speech production and speech modelling*. Dordrecht: Kluwer. 403–440. https://doi.org/10.1007/978-94-009-2037-8_16
- Lisker, Leigh and Arthur Abramson 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20: 384–422. <https://doi.org/10.1080/00437956.1964.11659830>
- Lisker, Leigh and Arthur Abramson 1967. Some effects of context on voice onset time in English stops. *Language and Speech* 10: 1–28. <https://doi.org/10.1177/002383096701000101>
- Makowski, Dominique, Mattan S. Ben-Shachar, Indrajeet Patil and Daniel Lüdtke. 2021. Automated results reporting as a practical tool to improve reproducibility and methodological best practices adoption, v. 0.5.0. CRAN. <https://github.com/easystats/report>
- Markó, Alexandra, Tekla Etelka Grácsi and Judit Bóna. 2010. The realization of voicing assimilation rules in Hungarian spontaneous and read speech: Case studies. *Acta Linguistica Hungarica* 57: 210–238. <https://doi.org/10.1556/aling.57.2010.2-3.3>
- Martinet, André. 1952. Function, structure, and sound change. *Word* 8: 1–32. <https://doi.org/10.1080/00437956.1952.11659416>
- Munteanu, Andrei. 2021. Homophony avoidance in the grammar: Russian nominal allomorphy. *Phonology* 38: 401–435. <https://doi.org/10.1017/S0952675721000257>
- Myers, Scott. 2012. Final devoicing: Production and perception studies. In: Tony Borowsky, Shigeto Kawahara and Mariko Sugahara (eds.). *Prosody matters: Essays in honor of Elisabeth Selkirk*. London: Equinox Press. 148–180.
- Ohala, John J. 1981. The listener as a source of sound change. In: Carrie S. Masek, Roberta A. Hendrik and Mary Frances Miller (eds.). *Papers from the Parasession on Language and Behaviour conference (CLS 17)*. Chicago: Chicago Linguistics Society. 178–203.

- Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In: Peter F. MacNeilage (ed.). *The production of speech*. New York: Springer-Verlag. 189–216. https://doi.org/10.1007/978-1-4613-8202-7_9
- Pedersen, Thomas Lin. 2020. Patchwork: The composer of plots. <https://CRAN.R-project.org/package=patchwork>. R package version 1.1.0.
- R Development Core Team. 2020. R: A language and environment for statistical computing; 4.0.2. Vienna, Austria: R Foundation for Statistical Computing.
- Sampson, Geoffrey. 2013. A Counterexample to homophony avoidance. *Diachronica* 30: 579–91. <https://doi.org/10.1075/dia.30.4.05sam>.
- Silverman, Daniel. 2012. *Neutralisation*. Cambridge, MA: Cambridge University Press.
- Siptár, Péter and Miklós Törkenczy. 2000. *The phonology of Hungarian*. Oxford: Oxford University Press.
- Wedel, Andrew, Abby Kaplan and Scott Jackson. 2013. High functional load inhibits phonological contrast loss: A corpus study. *Cognition* 128: 179–86. <https://doi.org/10.1016/j.cognition.2013.03.002>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang et al. 2019. Welcome to the tidyverse. *Journal of Open Source Software* 4(43): 1686. <https://doi.org/10.21105/joss.01686>
- Yin, Sora Heng and James White. 2018. Neutralisation and homophony avoidance in phonological learning. *Cognition* 179: 89–101. <https://doi.org/10.1016/j.cognition.2018.05.023>

Zoltán G. Kiss

Eötvös Loránd University, School of English and American Studies,

English Linguistics Department

ORCID 0000-0001-5224-3563

gkiss.zoltan@btk.elte.hu