Zoltán G. Kiss
Péter Szigetvári

Telling fortis and lenis apart in English obstruent clusters[*]

We are going to cast doubt on the standard analysis of "voiceless" obstruent clusters in English. There is an almost unanimous agreement among those concerned that the language has clusters like [sp], [ft], [kt] (in, for example, *sport, after, packed, act*). We contend that this is not the case, these words contain [sb], [fd], [kd], and [gt], respectively.[1]

We begin with a historical/theoretical introduction to the issue, partly based on Szigetvári (2020) in §1 and then describe an experiment in §2, the results of which corroborate that the plosive clusters in *acting* and *packed in,* both traditionally transcribed as [kt], are not identical, there is reason to believe that the first one is [gt] and the second one is [kd].

1. The laryngeal identity of obstruent clusters in English

Many languages exhibit a two-way laryngeal contrast in obstruents. The two terms of this contrast are commonly called *voiced* and *voiceless.* Languages in which the *phonetic* difference between members of the two sets is indeed typically voicing are often referred to as voicing languages. English, however, is claimed to be an aspirating language, in which the difference between the two types of obstruent is realized by other phonetic cues, among them (post-) aspiration (Harris 1994; Iverson & Salmons 1995; Honeybone 2002; Jansen 2004; Beckman & al. 2013) and the shortening of the preceding sequence of sonorants, that is, a vowel and the following sonorant consonants (Zimmerman & Sapon 1958; Chen 1970; Davis & Van Summers 1989; Laeufer 1992).

The terms "voiced" and "voiceless" are phonetic categories that are not adequate to characterize the phonological contrast in an aspirating language. The use of these terms as phonological categories is misleading, since, for example, in English an obstruent that is phonologically "voiced" is not necessarily voiced phonetically. Alternative terms that have been proposed are lenis and fortis (Sievers 1876) or lax and tense (Jakobson & al. 1952). We will use Sievers's terms, *lenis* and *fortis,* in this paper. So the previous sentence now looks like this: in English an obstruent that is lenis is not necessarily voiced.

[*] We are grateful to Péter Siptár and Zsuzsanna Bárkányi for their helpful comments.

[1] What we give between square brackets are symbols representing phonological categories, not phonetic realizations.

The laryngeal identity of singleton obstruents can always be determined from phonetic cues in English. One of these cues is aspiration: prevocalic fortis plosives are followed by a significantly longer voice onset time (VOT) than their lenis counterparts. Thus, the VOT following the alveolar plosive is longer in *ten, try, contemn* than it is in *den, dry, condemn.* Another cue is the length of the sonorant interlude before the obstruent: this is shorter before a fortis than before a lenis obstruent. Thus the vowel is significantly shorter in *debt, less* than in *dead, Les,* and somewhat shorter in *metal, deafen* than in *medal, Devon* (Davis & Van Summers 1989). Likewise the vowel+nasal sequence is significantly shorter in *tense, tent* than in *tens, tend,* and somewhat shorter in *blunter* than in *blunder.* It is between sonorants that voicing is an important (but not indispensable) cue for telling fortis and lenis obstruents apart: the plosive in *simple* or the fricative in *alpha* is voiceless, while those in *symbol* or *Alva* are voiced.

The laryngeal identity of obstruents in a cluster of obstruents cannot always be determined. Referring to Swadesh's dilemma (1934), Twaddell argues that the plosive spelled *p* in *spill* could be categorized either as fortis [p] or as lenis [b]. Whichever we select of the two decisions, it will be arbitrary (1935: 30ff). The system is going to be defective either way: [sp] or [sb] is going to be missing from it. Zellig Harris claims that the clusters in question are either uniformly fortis or uniformly lenis, clusters with a fortis and a lenis member hardly exist (1944: 183). This follows from his theory of long components (autosegments): the component responsible for fortisness and/or lenisness is shared between two adjacent obstruents. Pike (1947) also argues that the cluster in *spill* had better be analysed as [sp], since the second member is "phonetically closer" to [p] than to [b] (furthermore this is what the standard orthography suggests). Trager and Smith (1957) subscribe to the same analysis, though they admit that there is room for doubt.

Lotz & al. (1960), Reeds & Wang (1961), and Davidsen-Nielsen (1969), on the other hand, report of experiments they carried out to determine whether the generally accepted analysis of clusters spelled *sp* as [sp] is indeed phonetically justified. Their results suggest a rather obvious no: in Lotz & al.'s perception experiments more than 90% of speakers of American English identified the second part of *sp, st,* and *sc/sch/sk*[2] clusters as [b], [d], and [g], respectively, after the [s] was removed from the recording presented to them. In Reeds & Wang's case the number was even higher, 98% (353 of 360 instances; 1961: 80). Davidsen-Nielsen corroborates the results of the two

[2] While *skill* provided rather uniform results, there was some uncertainty in the case of *score*: about 70% of the respondents identified the plosive as [g], 30% as [k]. Lotz & al. conclude that "vowel quality has some effect on the judgement" (1960: 76).

ISSN 2061–490X, http://seas3.elte.hu/even.

earlier experiments and adds the results of his auditory, acoustic, and physiological experiments to support the claim that these plosives are indeed not fortis, but lenis. He claims that since voicing does not necessarily characterize lenis plosives even intervocalically, the voicelessness of a plosive after [s] cannot be taken to be a symptom of its fortisness. In fact, the absence of aspiration in this environment suggests that they are lenis.

If we accept the results of these phonetic experiments, namely, that *[sp], *[st], and *[sk] are ruled out by some phonotactic constraint in English, we're back to Twaddell's dilemma: why? Why is the distribution of fricative+plosive clusters defective?

Note that such defectivity seems not to occur in plosive+plosive clusters: we find both [kt], in *lactose,* and [kd], in *anecdote.* The lenis+lenis cluster [gd] also occurs, in *amygdaloid,* but lenis+fortis [gt] appears to be impossible within a morpheme in the widely accepted accounts. (Any laryngeal combination of obstruents, so even [gt] occurs across a (strong) morpheme boundary though: *ragtime.*) What is interesting in the pair [kt]–[kd] is that the second plosive is not voiced in either cluster, just like in [st] or [sd], whichever of the two we think occurs in English. We identify the prevocalic plosive in *lactose* as fortis [t] because it is aspirated, and the one in *anecdote* as lenis [d] because it is not aspirated. This corroborates the view that *star* is [sdaː]: the plosive after a fricative is not aspirated, so it is not fortis. But again, why should aspiration be impossible after a voiceless fricative?

In fact, it is not. Fricative+aspirated plosive clusters do not occur word initially, but they do between vowels. In words like *Aztec, gazpatcho, Azkaban, lieutenant,*[3] the plosive after the fricative is aspirated. In most cases this is indicated by analysing/transcribing (often also spelling) the fricative as lenis. This fricative is voiceless, but if it were transcribed as such,[4] the following

[3] We find an interesting variation in the transcription of the recently coined word *cosplay* from *costume+play.* In dictionaries we find this word transcribed with [zp] (e.g., https://dictionary.cambridge.org/us/dictionary/english/cosplay) or both [zp] and [sp] (e.g., https://www.oxfordlearnersdictionaries.com/definition/english/cosplay_1). Impressionistically, most speakers featured on YouGlish for this word aspirate the [p], but very few voice the fricative before it. Our assumption is that the voiceless fricative is transcribed as [z] in this word to indicate that the following [p] is aspirated.

[4] The standard transcription of the British pronunciation of *lieutenant* is with [ft], not with [vt], as we would here expect (note the spelling though). This is because there is a tradition of formulating the occurrence of aspiration by reference to syllable boundaries, which, in turn, are defined by reference to word boundaries. Now [s]+C clusters do occur word initially in English, but [ft] does not, as a consequence, the latter is not considered to be syllable initial. Accordingly, in this view the [t] of this cluster is not expected to be

plosive would be predicted to be unaspirated (compare *Aztec* with aspiration vs. *hostile* without aspiration). The absence of word-initial fricative+aspirated plosive clusters follows from the general absence of lenis fricative+consonant clusters in this position (with the exception of [vj] and, in some varieties of English, [zj]).

That is, English has fortis+lenis (FL), (1a), and lenis+fortis (LF) fricative+plosive clusters, (1b), as well as lenis+lenis (LL), (1c), but not fortis+fortis, except, of course, across a # boundary, (1d).

(1) Fricative+plosive clusters in English
   a. FL: aspen [sb], mistake [sd], mascot [sg], after [fd], Afghan [fg]
   b. LF: gazpatcho [zp], Aztec [zt], Azkaban [zk], lieutenant [vt]
   c. LL: husband [zb], wisdom [zd], Glasgow [zg]
   d. F#F: less time [s#t], misplace [s#p], offcut [f#k]

The absence of a contrast between [sp], [st], [sk] and [sb], [sd], [sg] is parallelled by a similar lack of contrast between [ps], [ts], [ks] and [pz], [tz], [kz]. This fact is noted by Jones, who claims that the ending of words like *puts, drinks, box* could be transcribed [tz] and [kz] (1967: 47f). Such a lack of contrast does not only characterize word-final position, but all other positions too, we find no contrast between FF and FL plosive+fricative clusters prevocalically either: *Leipzig* may be transcribed either [ps] or [pz].

As already mentioned, one phonetic cue for identifying fortis obstruents is the relative shortness of the preceding sonorant interlude. This may be a vowel or a vowel+sonorant sequence, as in *hat, art, out, height, shalt, ant, pint, count* or *aunt* (as opposed to the vowel(+sonorant sequence)s in *had, hard, loud, hide, palled, hand, hind, hound* or *demand*). That is, the vowel may be followed by the fortis obstruent immediately, or separated by an intervening glide, liquid, or nasal.

But what if there were another obstruent between the vowel and the fortis obstruent? We have seen that there are lenis fricative+fortis plosive clusters, as in (1b), in fact, we contend that FF clusters of this type do not even exist within a morpheme. Looking at plosive+plosive clusters, LF is found in standard transcriptions of English, albeit also rather rarely. Examples include mostly names: *Abkhaz, jodhpurs, Nagpur, Netrebko, subtend, subtilize, Varadkar.* FL is not very common either: *anecdote, synecdoche.* Most plosive clusters are analysed (= transcribed) as FF.

---

unaspirated. In most other clusters transcribed [ft], the plosive is unaspirated (*after, fifty, kaftan*), in some cases it is rather variable (*fifteen*).

ISSN 2061–490X, http://seas3.elte.hu/even.

However, the hypothesis that fricative+plosive clusters may not be FF, while plosive+plosive clusters may, is spurious in itself and leads to a further oddity if we consider past tense suffixation. While according to our hypothesis the past tense suffix must remain lenis in *kissed* [-sd] or *sniffed* [-fd], it would be rather gratuitous to claim that it "turn into" fortis in *kicked* [-kt] or *slipped* [-pt].

Let us suppose that plosive+plosive clusters are not different from fricative+plosive clusters in their laryngeal phonotactics. That is, we have the same categories for these clusters as in (1): FL, LF and LL, but not FF. This means that all those plosive+plosive clusters that are currently analysed as FF have to be reinterpreted as either FL or LF. One reasonable consideration is that if the past tense suffix remains [d] after fortis fricatives, then it also remains so after fortis plosives, *kicked* and *slipped* end in [-kd] and [-pd]. Recall, the voicelessness of an obstruent does not entail that it is fortis. This consideration significantly simplifies past tense allomorphy: the number of phonologically conditioned allomorphs decreases from the traditionally assumed three ([-d], [-t], [-id]) to two ([-d], [-id]). The [-t] allomorph only occurs in lexically conditioned past tense forms (*spill~spilt*). In *spilt* we indeed have a fortis plosive, the vowel+[l] sequence preceding [t] is shortened, and accordingly *spilt* contrasts with *spilled.* The voicelessness of the last plosive in *kicked* and *slipped,* however, is not a cue for fortisness, the preceding vowels here are shorter because of the stem final fortis plosives, [k] and [p]. Furthermore, [kikd] and [slipd] do not contrast with the hypothetical forms [kikt] and [slipt].

The idea that the regular past tense suffix does not alter its laryngeal quality from lenis to fortis suggests that morphemes are relatively stable in this respect. Indeed, plosives in English do not regularly change from lenis to fortis or vice versa. To avoid any change in the laryngeal qualities of the plosives involved, we assume that while *packed* ends in [-kd], the apparently noncontrastive cluster in *act* ends in [-gt]. The second member of this cluster is aspirated in *active* or *activity,* therefore it is fortis in the latter two words and, by conjecture, also in *act.*

In the following section we discuss the results of an acoustic investigation which intended to find out whether the difference between these two clusters we propose – [-kd] in *packed* vs. [-gt] in *act* – is also manifested in subtle but statistically detectable acoustic differences.

ISSN 2061–490X, http://seas3.elte.hu/even.

2. The experiment

2.1. Methods

Material. The test items of the production experiment were *acting* and *packed in*. The data were collected from YouGlish (https://youglish.com) automatically making use of its extended captioning-based search query functionality. A list of 200 videos was generated for each test item containing their YouTube ID for each occurrence, together with the start time where they can be found in the video. YouGlish generated the video list randomly, the query setting for language included the three English accents that YouGlish makes available: "US", "UK", and "AUS", i.e., the search was not limited to one type of accent, although as we will see below, the majority of the videos featured American English speakers. With the help of this list, nine-second clips were downloaded for each item-token. The audio files were extracted from the videos and were converted to uncompressed wav files. These wav files were then resampled at 22050 Hz, converted to mono, low-pass filtered with a cut-off at 11025 Hz, and saved as wav files. These were the files that were used for the acoustic analysis in Praat (v. 6.1.16, Boersma & Weenink 2020).

Files were discarded for the following reasons: there was music or loud background noise during the audio, the test item was mislabelled (e.g., *acting* was the label according to YouGlish but it was actually another word, such as *interacting*); there was a pause between *packed* and *in*; the speaker clearly had a non-native English accent; the audio file contained the same speaker (only one token by a given speaker was kept in the data set); the speech rate was so quick that segments were deleted; the audio contained a technical error (skip of sound) or a slip of tongue, or it was considered to be of low general quality (volume too low or too loud, absence of higher frequencies, etc.). After discarding files for any of these reasons, 101 tokens could be used for the acoustic analysis for *acting* and 91 for *packed in*. The distribution of the data for the two test items according to accent is shown below:[5]

[5] In the "Accent" column, "else" refers to those tokens in which the speaker's accent was not American, British, or Australian English (but for example, South African English).

Table 1: Distribution of data according to accent

Item	Accent	n
acting	else	6
acting	UK	9
acting	US	86
packed in	AU	1
packed in	UK	12
packed in	US	78

The distribution of the data by gender was unbalanced (number of male speakers: 131, female speakers: 61). However, this should not be an issue as we do not expect differences for the measured acoustic parameters between male vs. female speakers. As much as it could be determined, the test items were usually in declarative sentences, typically in non-absolute word final/non-prepausal position, although intonation, the prosodic position was not controlled for in this experiment.

Procedure. We measured the following acoustic parameters in *acting* and *packed in*:

- segment duration
  - vowel duration: the interval of the vowel before the two consonants
  - consonant duration: the interval containing the two consonants between the vowels, we will simply refer to this parameter as "consonant duration" even though two consonants were involved
  - the vowel's duration ratio to the total duration (i.e., to the duration of the vowel plus the two consonants) and to the consonant duration
- voicing
  - duration in the two consonants, we will simply call this parameter "voicing duration"
  - voicing ratio in the two consonants, referred to as "voicing ratio" below
- voice onset time, VOT
  - absolute duration of the VOT
  - ratio of the VOT duration to the duration of the two consonants, referred to as "VOT ratio" below

ISSN 2061–490X, http://seas3.elte.hu/even.

Segmentation was carried out manually, the segment boundaries were always placed on the zero crossing in the waveform. The measurements themselves were carried out on the segment intervals automatically based on a Praat script written by the authors.

The duration of the segments (the vowel and the two consonants following it) were measured on a segment-duration tier in Praat the following way. The starting boundary of the preconsonantal vowel was marked when periodic vibrations *and* the presence of the first three formants were visible in both the waveform and the wideband spectrogram. The interval boundary between the first vowel and the first consonant was placed at the end of the vowel's first three formants. Periodicity – the low-level energy indicative of vocal fold vibration – is not a good enough indicator of the boundary as it may continue into the first consonant. The end of the consonant cluster was marked when the short abrupt release noise was detectable in the waveform and spectrogram, or if it was not visible, it was marked at the beginning of the aspiration noise. We present the duration results of the *whole* consonant cluster in this study because in most tokens the boundary between the members of the cluster could not be reliably detected (e.g., there was no observable release noise between the consonants).

The duration of voicing during the consonant cluster's interval was measured on a separate tier in Praat as follows. The start and end of the voicing interval coincided with the start and end of the consonant cluster (i.e., the release of the second consonant was not included in the voicing interval). Periodicity was detected primarily using the waveform of the speech signal. If there was a periodic waveform within the voicing interval, it was labelled "voiced". Note that the intensity of this periodic waveform was often very low, due to the dying-out of the vibrations of the vocal folds (coming from the vowel on the left) during stop-closure. Nonetheless, as long as these periodic waveforms were visible, they were considered as acoustic correlates of vocal fold vibration. If the waveform was flat, it did not include periodic waves, that portion was marked as "voiceless" within the voicing interval.

The VOT duration was also measured in Praat on a separate tier. The start of the VOT interval was the exact same position where the end of the consonant cluster was marked, i.e., it included the release noise, too, when it was visible or only the aspiration noise if the release noise was not detectable (cf. Abramson & Whalen 2017). The end of the VOT was marked when the following vowel's periodic vibrations and first three formants were detectable in both the waveform and the spectrogram.

The figure below shows an example for the segmentation of *acting*, in it, the duration of VOT is 39 ms.

ISSN 2061–490X, http://seas3.elte.hu/even.

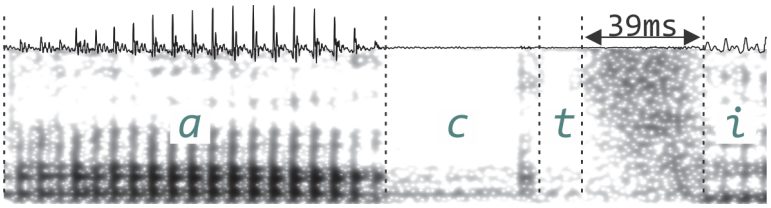

Figure 1

Statistical analysis. We aimed to measure the differences between the means of the acoustic parameters listed above for the two test items *acting* and *packed in*. To test whether the data are compatible or not with the null hypothesis that the means are not different from each other, the Welch independent-samples *t* test was used (two-tailed, $\alpha = 0.05$), the data scores belonged to two groups, they were randomly sampled, and came from different subjects, hence the independence condition could be assumed. As we will see below, in some cases the distribution of the data points could not be assumed to follow the normal distribution as they were right-skewed. However, the Welch independent-samples *t* test is known to be rather robust as far as the normality assumption is concerned (Krzywinski & Altman 2014), and it is extremely robust with respect to the "homogeneity of variance" requirement (Field et al. 2012: 373), which might be an issue if the sample sizes are different (as in the case of the present experiment). The effect size reported in this paper is Cohen's *d*, calculated with the *cohensD* R function (*lsr* package, Navarro 2015; this *d* is a value adjusted for the assumption that the corresponding populations may not have equal variances). The statistical analysis (including the generation of the plots) was carried out in R (R Core Team 2020; the non-base R packages used were: *tidyverse*, Wickham et al. 2019; *ggbeeswarm*, Clarke & Sherrill-Mix 2017; *patchwork*, Pedersen 2020).

2.2. Results

2.2.1. Segment duration

Total duration of vowel+consonants. Table 2 displays the descriptive statistics of the vowel plus consonant cluster interval, while Figure 2 shows the distribution of the data for this variable.

Table 2: Total duration of the vowel and the two consonants (in ms)

Item	N	Mean	SD	Median	Min	Max	SE
acting	101	228.68	35.81	227	144	338	3.56
packed in	91	223.16	42.44	216	143	368	4.45

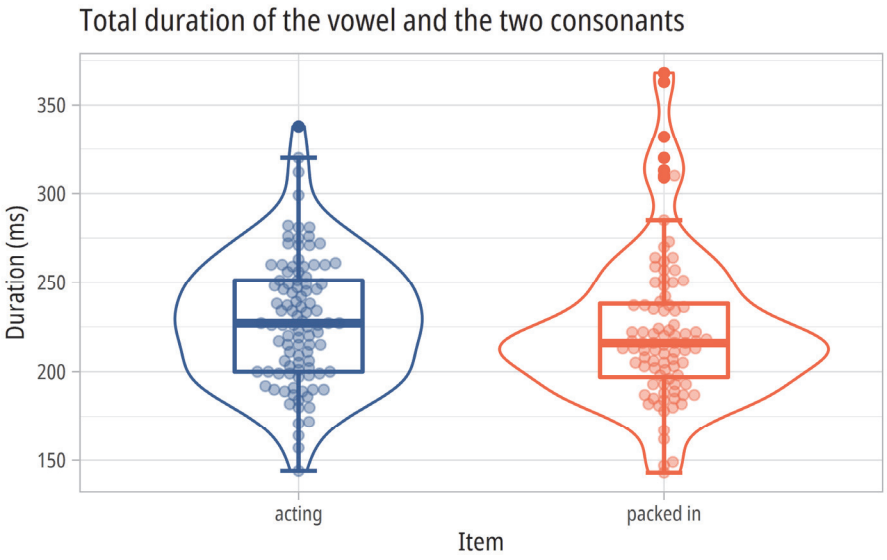

Figure 2

The Welch independent-samples *t* test showed that there was no statistically significant difference between the two test items with respect to the total duration of the vowel and the two consonants: $t(176.96) = 0.96818$, $p = 0.3343$. The 95% confidence interval (henceforth "CI95") for the mean difference is [−5.73, 16.77]. Cohen's $d = 0.14$, a very small effect size. This indicates that the two test items were produced at a similar speech rate.

Duration of the vowel and the two consonants. Tables 3 and 4 show the descriptive statistics for the duration of the vowel and the consonant cluster. The data distribution for these durational variables is shown in Figure 3.

Table 3: Duration of the vowel (in ms)

Item	N	Mean	SD	Median	Min	Max	SE
acting	101	126.78	25.42	123	65	222	2.53
packed in	91	110.62	30.75	109	46	203	3.22

Table 4: Duration of the two consonants (in ms)

Item	N	Mean	SD	Median	Min	Max	SE
acting	101	101.90	26.22	100	54	176	2.61
packed in	91	112.55	29.67	105	66	215	3.11

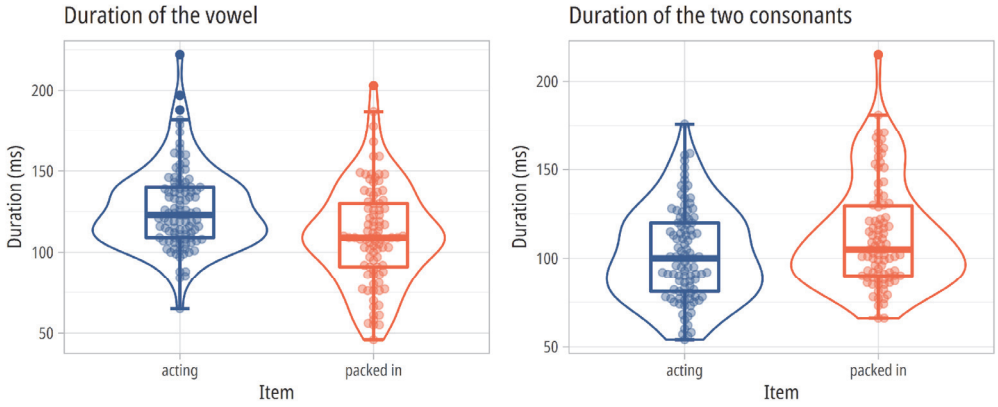

Figure 3

As we can see, the mean duration of the vowel in *acting* was longer than in *packed in*. This difference between the mean durations was statistically significant: $t(175.19) = 3.9459$, $p < 0.001$; CI95: [8.08, 24.25]; Cohen's $d = 0.57$, a medium effect. The two consonants were shorter on average in *acting* than in *packed in*. The difference between the mean durations was also statistically significant: $t(180.7) = -2.6227$, $p < 0.001$; CI95: [−18.66, −2.64]; Cohen's $d = 0.38$, a small effect.

Ratio of the vowel duration to the total and consonant duration. As we saw in Tables 3 and 4, and Figure 3, the mean duration of the vowel was longer in *acting* than in *packed in*, the consonant cluster was shorter on average in *acting* than in *packed in*. Since the overall duration of the vowel plus the consonants was similar (Table 2, Figure 2), this means that the vowel's mean duration ratio to the total duration (vowel+consonants) and to the consonant duration was also greater in *acting* than in *packed in*:

Table 5: Ratio of the vowel's duration to the total duration

Item	N	Mean	SD	Median	Min	Max	SE
acting	101	0.56	0.08	0.56	0.36	0.73	0.01
packed in	91	0.49	0.10	0.50	0.21	0.69	0.01

Table 6: Ratio of the vowel's duration to the consonant duration

Item	N	Mean	SD	Median	Min	Max	SE
acting	101	1.33	0.44	1.26	0.57	2.72	0.04
packed in	91	1.05	0.40	1.00	0.27	2.21	0.04

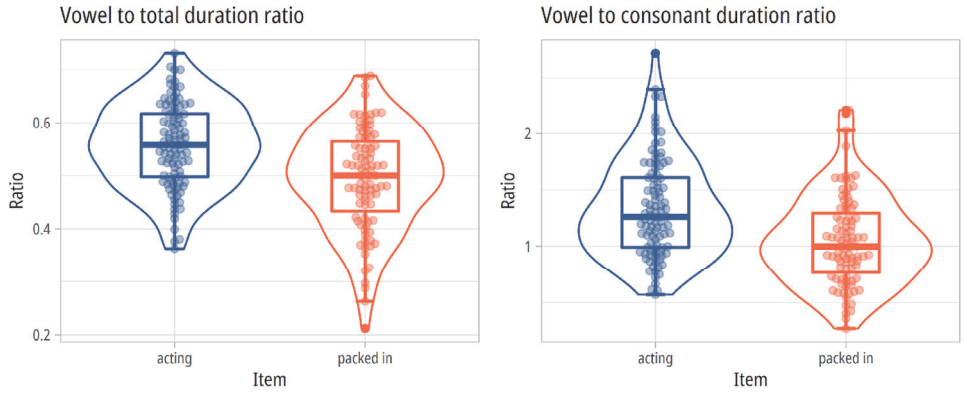

Figure 4

On average, 56% of the total duration was taken up by the vowel in *acting*, while the vowel in *packed in* was little less than half of the total duration in the case of *packed in*. This means that the vowel was longer than the two consonants by 1.33 on average. This difference between the mean vowel duration ratios turned out to be statistically significant; vowel-to-total ratio: $t(175.01) =$

4.8203, $p < 0.001$; CI95: [0.56, 0.49]; Cohen's $d = 0.70$, a medium effect; vowel-to-consonant ratio: $t(189.99) = 4.645$, $p < 0.001$; CI95: [1.33, 1.05]; Cohen's $d = 0.67$, a medium effect.

2.2.2. Voicing

As Table 7 shows, in almost 80% of the *acting* tokens, the consonant cluster had some (nonzero) voicing, while this proportion was only 55% in the case of *packed in* (i.e., almost half was completely voiceless):

Table 7: Voicing in the two consonants

Item	Voicing	N	Proportion
acting	voiced	80	0.79
acting	voiceless	21	0.21
packed in	voiced	50	0.55
packed in	voiceless	41	0.45

In what follows, we will consider only those tokens in which the consonants had nonzero voicing. Descriptive statistics for the voicing duration and ratio can be found in Tables 8 and 9. Figure 5 shows the distribution of these voicing variables.

Table 8: Voicing duration (in ms)

Item	N	Mean	SD	Median	Min	Max	SE
acting	80	19.26	15.23	16.5	1	98	1.7
packed in	50	12.54	4.95	13.0	3	23	0.7

Table 9: Voicing ratio

Item	N	Mean	SD	Median	Min	Max	SE
acting	80	0.20	0.15	0.16	0.01	1.00	0.02
packed in	50	0.12	0.05	0.11	0.03	0.21	0.01

ISSN 2061–490X, http://seas3.elte.hu/even.

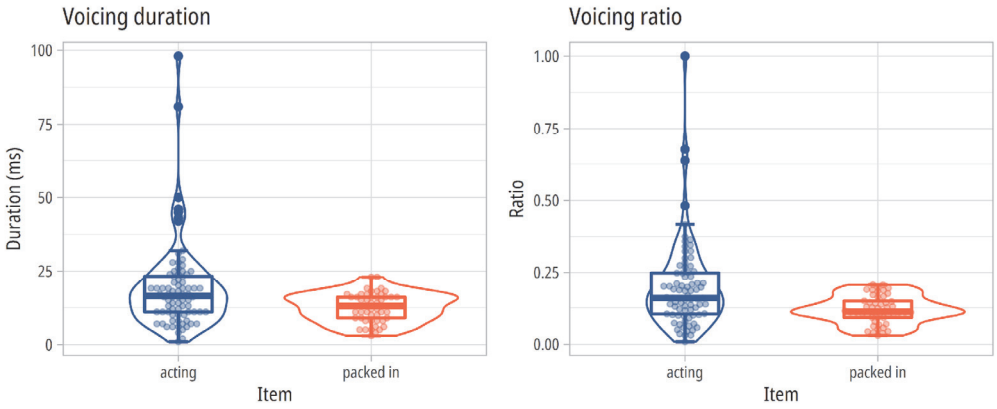

Figure 5

The mean duration of voicing in the two consonants was longer in *acting* than in *packed in*. This was mainly due to the following: the bulk of the data points were generally higher in *acting*; there were extreme values in the case of *acting*, too (i.e., in some tokens, three quarters of the consonant interval contained voicing, in one case the whole interval was voiced), and also, that *packed in* contained more scores close to zero voicing. The difference between the mean voicing durations was statistically significant: $t(103.22) = 3.6524$, $p < 0.001$; CI95: [3.07, 10.37]; Cohen's $d = 0.59$, a medium effect. The difference between the mean voicing ratios was also statistically significant: $t(101.65) = 4.3261$, $p < 0.001$; CI95: [0.1991857, 0.1193445]; Cohen's $d = 0.70$, a medium effect.

Figure 6 shows the correlation between vowel duration and the voicing ratio (here all tokens are included, the completely voiceless ones, too):

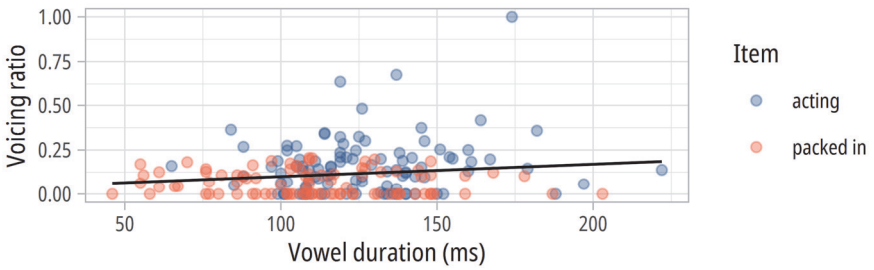

Figure 6

As we can see in Figure 6, there is a positive relationship between the duration of voicing and the ratio of voicing in the consonant interval: the longer the vowel, the more voiced the consonant tends to be, although the relationship is relatively weak (Pearson's correlation coefficient $r = 0.12$).

2.2.3. Voice Onset Time

The descriptive statistics of the VOT duration and the VOT ratio to the consonant interval can be found in Tables 10 and 11, while the distribution of the VOT data is displayed in Figure 7.

Table 10: Voice Onset Time (ms)

Item	N	Mean	SD	Median	Min	Max	SE
acting	101	31.42	14.82	27	11	88	1.47
packed in	91	21.65	9.52	20	8	45	1.00

Table 11: Ratio of VOT to the duration of the two consonants

Item	N	Mean	SD	Median	Min	Max	SE
acting	101	0.33	0.2	0.29	0.09	1.21	0.02
packed in	91	0.20	0.1	0.18	0.07	0.55	0.01

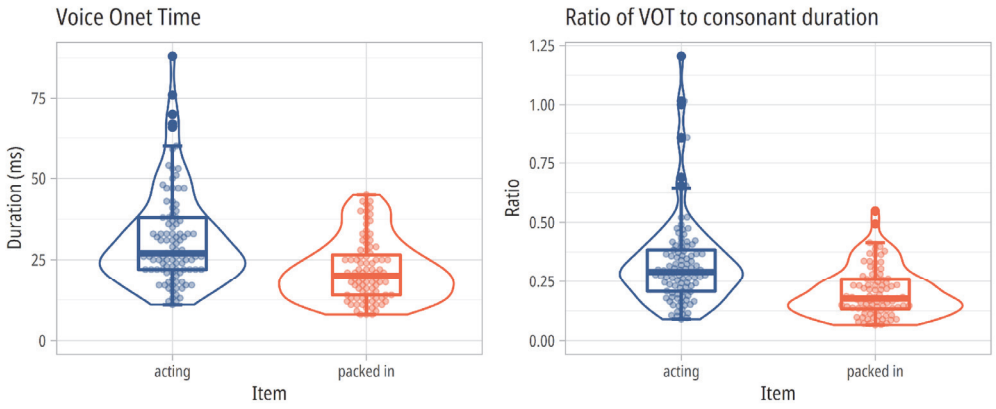

Figure 7

The mean Voice Onset Time turned out to be longer in *acting* than in *packed in*. The distribution of the data in *acting* was much more skewed towards larger values. The mean ratio of VOT compared to the duration of the two consonants was also larger in *acting*: in some cases, the VOT was as long as or even longer than the two consonants themselves. Both differences turned out to be statistically significant. Difference between the mean VOTs: $t(172.39) = 5.4865$, $p < 0.001$; CI95: [6.25, 13.28]; Cohen's $d = 0.79$, a medium-large effect. Difference between the mean VOT ratios: $t(154.97) = 5.7258$, $p < 0.001$; CI95: [0.08, 0.17]; Cohen's $d = 0.82$, a large effect.

As Figure 8 shows, there was no correlation between the VOT duration and the length of the two consonants for either item (Pearson's correlation coefficient $r = 0.055$ (*acting*); $r = 0.051$ (*packed in*)):

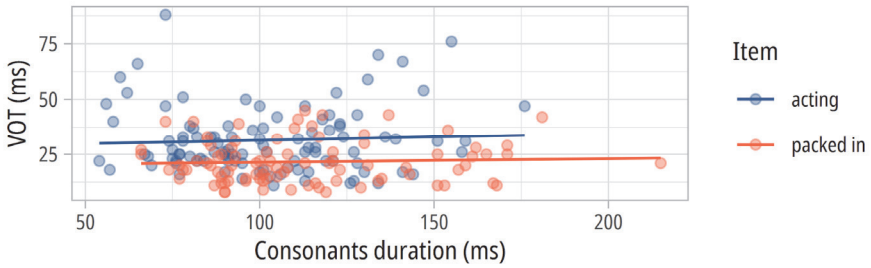

Figure 8

3. Discussion

The results for the segment durations indicate that while the vowel+consonant cluster had very similar lengths in both test items, the vowel was consistently longer in *acting* than in *packed in*, whereas it was just the reverse for the consonant duration: the two consonants were consistently shorter in *acting* than in *packed in*. As a result of this, the ratio of the vowel's duration to the vowel+consonant interval was also consistently greater in *acting* than in *packed in*. This finding is compatible with the well-known phonetic-phonological fact according to which, vowels tend to be phonetically shorter or "clipped" before phonologically "voiceless" obstruents than before phonologically "voiced" obstruents, especially in those environments where the postvocalic obstruent tends to lose its phonation, and there is a danger for the laryngeal contrast to be neutralized (see, among others, Bárkányi & G. Kiss 2019, to appear; Baroni & Vanelli 2000; Baum & Blumstein 1987; Docherty 1992; Gráczi 2012; Javkin 1976; Kingston & Diehl 1994; Kluender et al. 1988; Mair & Shadle

1996; Massaro & Cohen 1983; Parker et al. 1986; Port & Dalby 1982; Port & Leary 2005; Smith 1997). As far as English obstruents are concerned, such a phonetically "unfavourable" environment is next to another obstruent, just like the plosives in our test items. We can then assume that *act* can be analyzed as [agt], i.e., with a lenis+fortis cluster, while *packed* is [pakd], i.e., with a fortis+lenis cluster. This underlying contrast is reflected by the fact that the vowel is clipped before fortis [k] in *packed* but unclipped before [g] in *act.*

The voicing-related results also seem to support the [agt] vs. [pakd] analysis. Although both postvocalic stops had completely voiceless articulations, *act* had a lot more tokens in our data that contained at least some voicing. Only 55% of the *packed in* tokens contained voicing, and even those had only a mean duration of 12.54 ms, and the maximum value was only 23 ms. The distribution of these voiced tokens was rather compact, there were no extreme outlier values outside of this range. This is usually indicative of coarticulatory voicing, i.e., the voicing is just a by-product residue of the preceding vowel's final few vocal fold pulsations (cf., e.g., Schmidt & Willis 2011). In contrast to this, in 79% of the *acting* tokens, the postvocalic plosive contained some voicing, and the duration of voicing was longer on average (19.26 ms) than in the case of *packed in*, some tokens were well over the ceiling of coarticulatory voicing: values over 25 ms comprised the top 25% of the data scores, with a maximum value of 100 ms, which happened to be the full length of the consonant cluster. This suggests some planned voicing on the part of speakers, and not mere coarticulatory voicing: voicing may continue into the closure phase of the plosive longer in *act* than in *packed*. This planned extended phonetic voicing is also compatible with the segment duration results: the vowel in *acting* was longer, unclipped, and its voicing could therefore continue into the consonant closure longer. This was also corroborated by the correlation between vowel length and the voicing ratio: the longer the vowel, the greater the voicing ratio in the consonant tended to be, although the relationship was weak because of the relatively large number of completely voiceless tokens. All this is in line with the phonetic properties of English obstruents: while fortis obstruents are only voiced phonetically due to coarticulation, lenis obstruents may be phonetically voiced through allowing voicing to continue longer from the preceding vowel.

An interesting question is whether this small voicing difference that we found is sufficient to maintain the laryngeal contrast between obstruents in environments such as the one discussed here. After all, if phonetic voicing occurred, only 20% of the consonant closure was voiced on average in *acting*. The laryngeal contrast of obstruents can have various acoustic correlates, and segmental duration (of the vowel and of the obstruent) and the voicing dura-

tion within the obstruent are two of them (cf. Lisker 1986; Jansen 2004; Bárkányi & G. Kiss 2019, to appear). These correlates, and the interplay between them, have an important role in the perception of the laryngeal contrast. It generally seems to be the case that if there is sufficient voicing present in the obstruent, that fact in itself is a salient perceptual cue for the recognition of the contrast, and the durational parameters are redundant. If, however, the voicing ratio is small, the durational cues step up to play an important role in the perception and hence the maintaining of the contrast. For instance, preliminary results from Hungarian indicate that this cutoff point is at around 20 to 30% voicing (see Bárkányi & Mády 2012; Bárkányi & G. Kiss 2019, to appear). If the amount of voicing is below this value, it is more likely that the obstruent will be perceived as "voiceless". However, this perception can be counterbalanced by the durational cues: even if voicing is below 20–30%, the obstruent may be categorized by listeners as "voiced" if the vowel is long enough (compared to the consonant). Thus, just because an obstruent is (mostly) voiceless phonetically does not mean its laryngeal contrast is completely neutralized. For example, Bárkányi & Mády (2012) and Bárkányi & G. Kiss (2019, to appear) investigating the contrast between Hungarian [s] and [z] in various environments found that if only 14% of the alveolar fricative contained voicing, the vowel had to be at least around 2.1 times as long as the consonant for the fricative to be categorized by listeners as "voiced". At 22% voicing, the required minimum vowel-to-consonant duration ratio had to be 1.5 for the "voiced" responses. At 30%, the value went down to around 0.8, and below this value, the durational cues did not play a role in the perception of the fricative. In our experiment we found that the mean voicing ratio in *acting* was 20%, while the mean vowel-to-consonant duration ratio was 1.33. In the case of *packed in*, the mean voicing ratio was only 12% (among those tokens that had some voicing at least, a lot of tokens were completely voiceless), and the mean vowel-to-consonant duration ratio was 1.05. If what Bárkányi & Mády (2012) and Bárkányi & G. Kiss (2019, to appear) found for Hungarian (a non-aspirating, voicing language) can be applied to English plosives too, then our results are compatible with the assumption that the postvocalic plosives in *act* and *packed* are different and not neutralized despite being mostly phonetically voiceless, namely, that *act* is [agt], but *packed* is [pakd]. This is because in *act* the vowel is long enough and the plosive is voiced enough for it to be categorized as "lenis/voiced", whereas in *packed* the durational cues and the lack of voicing signal a "fortis/voiceless" plosive. But further perception research is necessary for English particularly before reliable conclusions can be made in this respect.

Finally, the VOT results indicate that the second consonant in *act* is fortis [t], while it is lenis [d] in *packed*. This is because the VOT was consistently longer in the *acting* tokens than in the *packed in* tokens. The mean VOT in *acting* was 31.42 ms, 25% of the data was over 35 ms, the maximum value was 88 ms (which was longer than the consonant cluster itself in that particular token). These long-lag VOT facts strongly indicate the presence of aspiration (Lisker & Abramson 1964; Keating 1984). In contrast to this, the VOT was much more compactly distributed in *packed in*, with a mean of 21.6 ms. This amount of VOT is usually not considered to be aspiration (Keating 1984). The lack of correlation between the consonant duration and VOT (Figure 8 above) indicates that VOT is independent of segment length (e.g., longer articulation did not correlate with longer VOT), that it is a relatively stable, "fixed" value (or range of values), and all that matters is whether the final plosive is fortis or lenis. If fortis, VOT will be relatively long (i.e., there will be aspiration), if lenis, the VOT will be relatively short (i.e., no aspiration). Therefore, based on these findings, we propose that the final plosive in *act* is fortis [t], while it is lenis [d] in *packed*.

References

Abramson, Arthur S. and D. H. Whalen. 2017. Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. Journal of Phonetics 63: 75–86. https://doi.org/10.1016/j.wocn.2017.05.002

Bárkányi, Zsuzsanna and Katalin Mády. 2012. The perception of voicing in fricatives. Paper presented at the 9th Old World Conference in Phonology (OCP9), Berlin, 18–21 January, 2012.

Bárkányi, Zsuzsanna and Zoltán G. Kiss. 2019. A fonetikai korrelátumok szerepe a zöngekontraszt fenntartásában. Általános Nyelvészeti Tanulmányok 31: 57–102.

Bárkányi, Zsuzsanna and Zoltán G. Kiss. to appear. The perception of voicing contrast in assimilation contexts in minimal pairs. Acta Linguistica Academica.

Baroni, Marco and Laura Vanelli. 2000. The relations between vowel length and consonantal voicing in Friulian. In: Lori Repetti (ed.): Phonological theory and the dialects of Italy. Amsterdam & Philadelphia: John Benjamins. 13–44. https://doi.org/10.1075/cilt.212.04bar

Baum, Shari R. and Sheila E. Blumstein. 1987. Preliminary observations on the use of duration as a cue to syllable-initial fricative consonant voicing in English. Journal of the Acoustical Society of America 82: 1073–1077. https://doi.org/10.1121/1.395382

Beckman, Jill, Michael Jessen and Catherine Ringen. 2013. Empirical evidence for laryngeal features: Aspirating vs. true voicing languages. Journal of Linguistics 49: 259–284. https://doi.org/10.1017/s0022226712000424

Boersma, Paul and David Weenink. 2020. Praat: doing phonetics by computer [Computer program]. Version 6.1.16. http://www.praat.org/

Chen, Matthew. 1970. Vowel length variation as a function of the voicing of the consonant environment. Phonetica 22: 129–159. https://doi.org/10.1159/000259312

Clarke, Erik and Scott Sherrill-Mix. 2017. ggbeeswarm: Categorical Scatter (Violin Point) Plots. R package version 0.6.0.

Davidsen-Nielsen, Niels. 1969. English stops after initial /s/. English Studies 50: 321–339. https://doi.org/10.1080/00138386908597340

Davis, Stuart and W. Van Summers. 1989. Vowel length and closure relations in word-medial VC sequences. Journal of Phonetics 17: 339–353. https://doi.org/10.1016/s0095-4470(19)30449-8

Docherty, Gerard J. 1992. The timing of voicing in British English obstruents. Berlin & New York: Foris.

Field, Andy, Jeremy Miles and Zoë Field. 2012. Discovering statistics using R. London: Sage.

Grászi, Tekla Etelka. 2012. Zörejhangok akusztikai fonetikai vizsgálata a zöngésségi oppozíció függvényében. Doctoral dissertation. Eötvös Loránd University, Budapest.

Harris, John. 1994. English Sound Structure. Oxford: Blackwell.

Harris, Zellig S. 1944. Simultaneous components in phonology. Language 20: 181–205. https://doi.org/10.2307/410118

Honeybone, Patrick. 2002. Germanic obstruent lenition: Some mutual implications of theoretical and historical phonology. Doctoral dissertation. University of Newcastle-upon-Tyne.

Iverson, Gregory K. and Joseph C. Salmons. 1995. Aspiration and laryngeal representation in Germanic. Phonology 12: 369–396. https://doi.org/10.1017/s0952675700002566

Jakobson, Roman C., Gunnar M. Fant and Morris Halle. 1952. Preliminaries to speech analysis: The distinctive features and their correlates. Cambridge, MA: The MIT Press.

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

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.

Keating, Patricia A. 1984. Phonetic and phonological representation of stop consonant voicing. Language 60: 286–319. https://doi.org/10.2307/413642

Kingston, John and Randy L. Diehl. 1994. Phonetic knowledge. Language 70: 419–454. https://doi.org/10.1353/lan.1994.0023

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

Krzywinski, Martin and Naomi Altman. 2014. Comparing samples – part I. Nature Methods 11: 215–216. https://doi.org/10.1038/nmeth.2858

Laeufer, Christiane. 1992. Patterns of voicing-conditioned vowel duration in French and English. Journal of Phonetics 20: 411–440. https://doi.org/10.1016/s0095-4470(19)30648-5

Lisker, Leigh. 1986. Voicing in English: A catalogue of acoustic features signaling /b/ vs. /p/ in trochees. Language & Speech 29: 3–11. https://doi.org/10.1177/002383098602900102

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

Lotz, John, Arthur S. Abramson, Louis J. Gerstman, Frances Ingemann, and William J. Nemser. 1960. The perception of English stops by speakers of English, Spanish, Hungarian, and Thai: A tape-cutting experiment. Language and Speech 3: 71–77. https://doi.org/10.1177%2F002383096000300202

Mair, Sheila J. and Christine H. Shadle. 1996. The voiced/voiceless distinction in fricatives: EPG, acoustic and aerodynamic data. Proceedings of Institute of Acoustics 18: 163–169.

Massaro, Dominic W. and Michael M. Cohen. 1983. Consonant/vowel ratio: An improbable cue in speech perception. Perception and Psychophysics 33: 501–505. https://doi.org/10.3758/bf03202904

Navarro, Danielle. 2015. lsr. R package version 0.5. University of Adelaide.

Parker, Ellen M., Randy L. Diehl and Keith R. Kluender. 1986. Trading relations in speech and nonspeech. Perception and Psychophysics 39: 129–142. https://doi.org/10.3758/bf03211495

Pedersen, Thomas Lin. 2020. patchwork: The Composer of Plots. R package version 1.1.0.

Pike, Kenneth L. 1947. Phonemics. Ann Arbor.

Port, Robert F. and Jonathan Dalby. 1982. C/V 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. R Foundation for Statistical Computing, Vienna, version 4.0.2. http://www.R-project.org/R.version

Reeds, J. A. and W. S.-Y. Wang. 1961. The perception of stops after s. Phonetica 6: 78–81. https://doi.org/10.1159/000258075

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

Sievers, Eduard. 1876. Grundzüge der Lautphysiologie zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen. Leipzig: Beitkopf und Härtel.

Smith, Caroline L. 1997. The devoicing of /z/ in American English: Effects of local and prosodic context. Journal of Phonetics 25: 471–500. https://doi.org/10.1006/jpho.1997.0053

Swadesh, Morris. 1934. The phonemic principle. Language 10: 117–129. https://doi.org/10.2307/409603

Szigetvári, Péter. 2020. Emancipating lenes: A reanalysis of English obstruent clusters. Acta Linguistica Academica 67: 39–52. https://doi.org/10.1556/2062.2020.00004

Trager, George L. and H. L. Smith. 1957. An Outline of English Structure. Oklahoma: Norman.

Twaddel, W. Freeman. 1935. On defining the phoneme. Language 11: 5–62. https://doi.org/10.2307/522070

Wickham, Hadley et al. 2019. Welcome to the tidyverse. Journal of Open Source Software 4: 1686. https://doi.org/10.21105/joss.01686

Zimmerman, Samuel A. and Stanley M. Sapon. 1958. Note on vowel duration seen cross-linguistically. Journal of the Acoustical Society of America 30(2): 152–153. https://doi.org/10.1121/1.1909521

Zoltán G. Kiss
Eötvös Loránd University
gkiss.zoltan@btk.elte.hu

Péter Szigetvári
Eötvös Loránd University
szigetvari@elte.hu

ISSN 2061–490X, http://seas3.elte.hu/even.