

SEGMENTAL DURATION AS A CUE TO SYLLABLE BOUNDARIES IN CZECH

PAVEL ŠTURM

ABSTRACT

The aim of the study is to establish whether the acoustic signal contains cues to the syllabification of words that are perceptually relevant, as suggested by previous research. Syllabification preferences of 27 speakers of Czech were examined in a behavioural experiment using disyllabic nonsense words with 10 CC clusters as stimuli. The C1/C2 duration ratio of the intervocalic cluster was manipulated by shortening and lengthening of both consonants. Participants repeated auditorily presented stimuli by syllables, with clear pauses between them (a pause-insertion task). Logistic regression analyses revealed significant effects of sonority type of the cluster, word-edge phonotactics and syllabification strategy reported by the participants in a post-test interview (only half of the participants reported not to have followed any strategy). However, the manipulation condition did not turn out to be a significant predictor, although the C1/C2 ratio correlated negatively with the rate of cluster division. The correlation was in compliance with the hypothesis stating that when C1 is longer than C2, the cluster has a higher probability of being maintained as the onset of the following syllable.

Key words: syllable, syllable boundaries, syllabification, onset maximization, perception, Czech

1. Introduction

Although the phonetic segment, or speech sound, is the smallest recurring linear segment in speech, the processes of human production and perception normally operate at a higher level, namely, the level of words or syllables (Sendlmeier, 1995; Coleman, 2002; Goldinger & Azuma, 2003; Port, 2007). There is mounting evidence that whole stretches of speech are stored as sound-sense associations in memory, which are then recovered in the processing of speech (Coleman, 2002; Goldinger & Azuma, 2003; Hawkins, 2003). This complements the well-known fact that acoustic cues to individual segments are distributed over adjoining segments as well. Phonetic research shows that the syllable plays an important role in speech acquisition, and its language-specific characteristics

also contribute, among others, to the rhythm type of a language. However, determining the location of syllable boundaries presents a considerable challenge to researchers. In phonological description, syllable boundaries are often derived from segments and from theoretical assumptions of the analyst. For instance, the *maximum onset principle* (MOP) predicts intervocalic consonants to belong to the following vowel, thus forming the onset of the syllable, with the provision that phonotactics of the language should not be violated (e.g., Pulgram, 1970; Kahn, 1976). Syllable boundaries can then be effectively viewed as predictable from underlying representations of segments (e.g., Ewen & van der Hulst, 2001: 141ff.).

Experimental evidence is therefore needed regarding the syllabification of words. A growing body of studies suggest that the boundaries of syllables do not appear to follow onset maximization (even in its weaker form) strictly. First of all, phonotactics seems to be gradual rather than categorical, i.e., there are *degrees* of phonotactic legality related to language use and the frequency of occurrence, which is systematically reflected in well-formedness judgments (Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997; Treiman, Kessler, Knewasser, Tincoff, & Bowman, 2000; Munson, 2001; Hay, Pierrehumbert, & Beckman, 2004). Very frequent sequences may be preferred over sequences that occur with relatively low frequency. Secondly, a wide range of both phonetic and phonological factors has been identified in behavioural tasks¹ that affect the syllabification of intervocalic consonants or clusters. We will now review some of these factors briefly.

On the one hand, several phonological effects consistently appear in the experiments. First, intervocalic consonants tend to be attracted to the vowel which is stressed (e.g. Fallows, 1981 and Treiman & Danis, 1988 for English). Second, results from different languages show that phonological length also affects the performance of participants in behavioural tasks: compared to long vowels, short vowels are associated with a higher probability of attracting a coda consonant (e.g. Treiman & Danis, 1988 for English; Schiller, Meyer, & Levelt, 1997 for Dutch; Ní Chiosáin, Welby, & Espesser, 2012 for Irish). Third, assignment to onsets or codas is to some degree influenced by the nature of the intervocalic consonant, specifically by its sonority. In the experiment of Treiman and Danis, sonorant singleton consonants were associated with the previous vowel more closely than obstruent consonants. Similar results were obtained by Goslin and Frauenfelder (2001) for CC clusters in French (obstruent-liquid clusters were treated as complex onsets, whereas other CC clusters were divided) and by Ní Chiosáin et al. (2012) for clusters in Irish (obstruent-liquid vs. sonorant-obstruent clusters).

On the other hand, many of these results might be motivated phonetically. For example, the effect of sonority can be related to its acoustic correlates (Parker, 2008; Clements, 2009), or the effect of vowel length to vowel duration. The experiment of Ní Chiosáin et al. (2012) is important because it took into account durational values as well. With increasing duration of the (stressed) vowel in the first syllable, the probability of the vowel

¹ For a detailed discussion of methods, see a review in Côté and Kharlamov (2011). The basic principle of behavioural experiments is that participants perform a simple task that is not directly related to the investigated issue (syllable boundaries). Subjects may double parts of the word, change the order of syllables or repeat one part of the word. For instance, in first syllable reduplication the response /pɪkpɪknɪk/ to the stimulus *picnic* would suggest the syllabification /pɪk.nɪk/, while /pɪpɪknɪk/ would suggest /pɪ.knɪk/. Usually, the term “syllable” is not used, favouring the term “part of word”.

attracting a coda consonant decreased. Another study that reports durational effects is that of Redford and Randall (2005). The authors mention another researcher (Christie, 1977) who found that intervocalic clusters were divided when the duration of both consonants in a CC cluster was similar, but kept as an onset when the first C was longer (Redford & Randall 2005: 30). However, the clusters were not word-medial but around word-boundaries (e.g. 'help us nail' [s#n] vs. 'help a snail' [#sn]) since Christie focused on juncture cues. In their own experiment using disyllabic nonsense words, Redford and Randall (2005) found that when C1 was shorter than C2, the medial clusters tended to be divided, whereas they were kept together as an onset when C1 was longer. However, this was true only for clusters that were not violating phonotactics – illegal sequences were syllabified invariably according to phonotactics.

The quoted study by Christie is unavailable to us, but Christie (1974) also brings interesting findings. He synthesized 100 tokens of one nonce word ([asta]), varying the formant transients of [a] (flat × more movement), aspiration of [t] (unaspirated × aspirated) and using 25 steps in the duration of the closure interval. The listeners were forced to choose between V.CCV and VC.CV syllabifications. Formant transitions did not seem to have any effect on the listeners. However, the results showed that aspiration of [t] was associated with more C.C syllabifications, which is in accord with the allophonic variation of English plosives where aspiration after [s] in the same syllable is disallowed. Moreover, there was a gradual increase in the proportion of C.C syllabifications in response to increasing duration of the closure interval of [t], i.e., along with lengthening the second consonant of the cluster. This suggests that the C1/C2 ratio is indeed important in syllabification judgments (at least in synthetic speech).

The aim of the current experiment is to replicate such findings with Czech listeners using acoustic manipulations of the signal. Redford and Randall (2005) did the durational analysis *ex post*, taking the variability in duration into account. Our experiment is instead designed to investigate this effect explicitly. In addition, the authors used a written task, in which syllable boundaries were marked by the subjects on paper (the listeners were asked to write down nonce words that they heard and divide them into syllables). This is burdened with metalinguistic awareness to a higher degree than the behavioural experiments reported above (see Goslin & Frauenfelder, 2001). In contrast to Christie (1974), we would like to investigate a wider range of clusters, which also increases ecological validity of the experiment.

It is reasonable to believe that a significant array of phonetic features in the acoustic signal may contribute to the perception of boundaries between syllables. However, we focus only on the C1/C2 durational ratio. H0 would state that syllabification of medial clusters is not influenced by the temporal structure of the cluster. In contrast, our alternative hypothesis (H1) predicts that there will be more V.CCV syllabifications for tokens where the ratio has been raised (or less V.CCV syllabifications where it has been lowered). In other words, the longer the C1 (or the shorter the C2), the more CC onsets are predicted compared to unaltered tokens. This assumption is substantiated by the literature reported above and in general by domain strengthening effects, where for instance segments close to the initial boundary are longer or differently articulated (Fougeron & Keating, 1997).

In addition to this main line of interest, some of the other factors shown to be affecting syllabification judgments can be examined. Given the practical limitations of the experiment in terms of time and scope, we do not investigate the effects of stress or the phonological length of the preceding vowel. However, the target clusters vary in their frequency of occurrence word-initially from very frequent clusters to clusters that do not occur at all, which allows us to evaluate the contribution of word-edge phonotactics. It is predicted that sequences with a higher frequency of occurrence will be more likely to be preserved as CC onsets than less frequent sequences (Hypothesis 2). Moreover, the clusters are of different sonority and manner of articulation types (combinations of stops, fricatives and sonorants). Therefore, the sonority sequencing principle can be taken into account as well. Hypothesis 3 thus states that clusters with a rising sonority (obstruent-sonorant sequences) will be more likely to be preserved as CC onsets than clusters with a plateau in sonority (sequences of two obstruents or two sonorants).

2. Method

The material comprised a set of nonsense words that were modelled on the Czech language. Most of the pseudo-words differed from genuine words by a single phoneme substitution (e.g., /zolja/ ~ *rolba*), but several items included more substitutions (e.g., /knetrem/ ~ *svetrem*). The test session included 10 target nonsense words that represented 10 unique clusters differing in frequency and manner of articulation:

- two items with a plosive-plosive sequence (/tʃakt̪em/, /natka/)
- one item with an affricate-plosive sequence (/smat̪ski/)
- three items with a fricative-plosive sequence (/lesta/, /viskem/, /zaxti/)
- two items with a plosive-sonorant sequence (/knetrem/, /xɛbra/)
- one item with a fricative-sonorant sequence (/kɛslo/);
- one item with two sonorants (/xarmu/).

A set of distractor items varying in the number of syllables and cluster length was added to the test session as a source of variability (/louʃc/, /fnus/, /mostlma/, /kotxarma/, /toʒnit̪se/, /xɛt̪ɛ/, /sa:nɛf/); these items assisted to conceal the medial CC sequences in target items. In addition, a set of training items was constructed that was used for the initial pre-test session in which the participants familiarized themselves with the range of material to be presented later (/vras/, /vufe/, /t̪jeka/, /zolja/, /mɛspekt/, /kejsɛk/, /dorsko/, /loskame/, /pa:nɛstouʃ/, /nɛspozm/).

A female native speaker of Czech (22 years old) read a list of nonsense words including distractor items (hereafter we will refer to them simply as “words”). The experimenter was present during the recording, which took place at the Institute of Phonetics in Prague (sound-treated recording booth, condenser microphone, 16-bit 32-kHz audio). When a correction was necessary, the experimenter asked the speaker for a new version of the item. The aim was that the final speech production should sound as natural as possible, without emphasis, without syllable lengthening.

The acoustic signal of all target items and of selected other items was manipulated in *Praat* (Boersma & Weenink, 2014). First, the boundaries of the two intervocalic consonants

(C1 and C2) and the preceding vowel (V) were determined, and the duration of these segments was extracted (see Appendix). The rules for segmentation of the acoustic signal can be summarized as follows (see also the recommendations in Machač & Skarnitzl, 2009):

- the decisive factor was the formant structure, which defines vowels and sonorants; the boundary between an obstruent and sonorant segments was therefore placed at the point where full formant structure began/ceased to appear (in ambiguous cases, the boundary was placed in the middle of the transitory region);
- nasals have full formant structure, but are associated among others with a drop of energy in the higher frequencies and the presence of a nasal formant;
- lateral approximants are very similar to vowels, but are associated with lower formant values and a drop of energy in the higher frequencies; when the lateral was visually indistinguishable from the vowel, the decision was based primarily on audition alone; however, the sequence [sl] is problematic because of the synchronization of articulatory gestures – we can distinguish friction of [s], friction of devoiced [l̥] and full formant structure of [l]; the lateral in our analysis included both the [l̥] and [l] parts.

Figure 1. A spectrogram of the stimulus [kɛslo] with marked boundaries of target speech sounds. The devoiced ([l̥]) and sonorant ([l]) parts of the liquid are considered to be one speech sound.

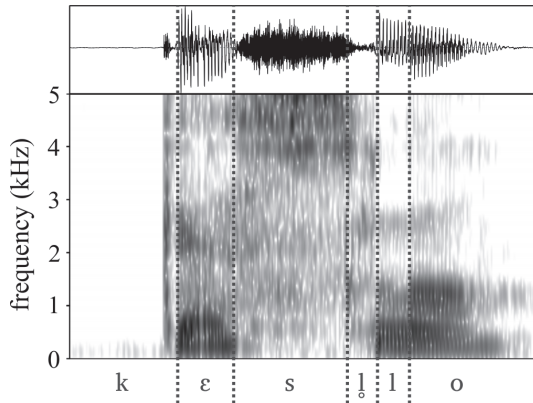


Figure 1 shows the segmentation of [kɛslo] as an example of boundary placement in plosive-vowel, vowel-fricative, [sl] and lateral-vowel contexts.

In the next step, five Manipulation objects were created from the Sound object (using default parameters): with *original* C1 and C2 durations (relative duration = 1.0), with C1 or C2 *lengthened* by half of its duration (relative duration = 1.5), and with C1 or C2 *shortened* by half of its duration (relative duration = 0.5). The duration of the preceding vowel was not altered. PSOLA resynthesis was used to create new audio files from these Manipulation objects. The perceptual test thus included 18 training items, 26 distractor items and 50 target items (5 manipulations × 10 words), totalling 94 items.

The perceptual test was administered in DMDX (Forster & Forster, 2003) without any information displayed on the screen. Each new item was introduced by a short warning signal (a combination of noise and tones), which also functioned as a simple means of perceptual desensitization. After 800 milliseconds the participants heard the stimulus itself,

and their response was recorded (see below for the instructions). The time to respond was limited to 3–4 seconds. Individual items were played automatically so that no further activity was necessary on the side of the participants. The experiment was divided into a training session followed by four test blocks. The participants were prompted to take a short break after each block. The order of the 19 items in a block (and the order of the blocks themselves) was randomized for each participant. The total duration of the experiment did not exceed 17 minutes. In a post-test interview, the participants were asked to report whether they had followed some strategy in the task.

The task was to *repeat the presented word by syllables*, with clear pauses between them. The participants were asked to follow only their first impression of the sound, how they perceived it. They were urged to listen very carefully to the stimuli. The participants were told that they were going to hear different variants of individual words, which were not supposed to have the same characteristics (or outcome of the division into syllables). They should consider and divide each word separately, individually, without reference to previous cases. The speech production of the participants was recorded with a microphone, and the location of syllable boundaries was identified with the aid of this recording (with silence between sound intervals implying a syllable boundary percept on part of the listener).

27 subjects participated in the experiment (19 females, 6 males, median age = 21 years), all students of English at a pedagogical faculty. However, two subjects were discarded prior to performing any analyses, one for being bilingual and one for showing signs of miscomprehension of the task during the training session. Thus, only 25 speakers were analysed (yielding 1250 tokens). Furthermore, 63 tokens (5% of the data set) with missing or ambiguous syllabification were removed. This comprised cases where, for instance, the speaker produced a given word without a break between the syllables, or hesitated, or produced a different consonant in the target cluster. The final number of analysed tokens was therefore 1187.

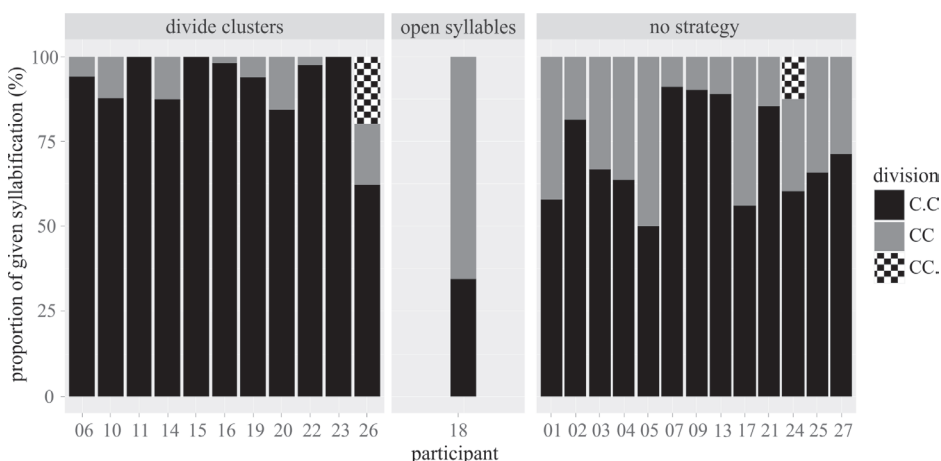
Statistical analyses were performed in the *R* software (R Core Team, 2016) using the package *lme4* (Bates, Mächler, Bolker, & Walker, 2015). Figures were drawn using the package *ggplot2* (Wickham, 2009). The data were analysed with a logistic mixed-effects regression that allows investigating the effect of predictors on a binary dependent variable (the type of syllabification outcome: V.CCV × VC.CV). Individual predictors are introduced in the results section. The statistical significance of a predictor was evaluated in a goodness-of-fit test using maximum likelihood ratio by comparing the full model (a given set of predictors and their interactions) with a reduced model (lacking one predictor or interaction). Maximum random-effect structure was used that still allowed the model to converge. In addition, other basic statistical functions were used in the analyses (t-tests, correlations, binomial tests).

3. Results

Overall, the preference was to divide the intervocalic clusters between two syllables (79%), followed by V.CC syllabifications (20%) and CC.V syllabifications (1%, $n = 15$). Given that three speakers divided the intervocalic clusters in all stimuli (thus yielding

100% of C.C syllabifications), it is likely that individual participants may have different strategies. Figure 2 therefore shows the response patterns of all participants depending on the strategies reported in the post-test questionnaire. Approximately half of the participants said they did not follow any specific strategy in the task. A similar number of participants admitted to divide any intervocalic clusters. Only one participant reported that he endeavoured to pronounce open syllables, keeping the cluster as an onset. Accordingly, this speaker is associated with the lowest proportion of C.C responses, and the “no strategy” group generally seems to yield a lower proportion of C.C responses than the “divide clusters” group. Interestingly, the CC.V responses were produced by only two speakers. Given its speaker specificity and low occurrence, this category was therefore also excluded from the results, leaving 1172 tokens for analysis using logistic regression with a binary response variable. (However, we will return to the VCC.V syllabifications in the Discussion.)

Figure 2. Proportion of syllabification response types for individual speakers in relation to their reported strategies in the experimental task.



The logistic regression analysis therefore includes STRATEGY as a predictor, which proved to be statistically highly significant ($\chi^2(2) = 21.9, p < 0.001$). Adding this predictor reduced the variance of the random effect of PARTICIPANT (from 2.7 to 0.9). The goodness-of-fit of the model was further improved when SONORITY was included as a three-level predictor ($\chi^2(2) = 16.2, p < 0.001$), differentiating between clusters of two obstruents \times two sonorants \times an obstruent followed by a sonorant. The residual variance of the random effect of WORD decreased from 2.0 to 0.3. Specifically, S-S clusters were associated with the highest odds of division, whereas O-S clusters with lowest. The SONORITY effect was also added as a slope to PARTICIPANT, allowing individual participants to differ in sensitivity to the sonority classes ($\chi^2(5) = 15.1, p < 0.01$). Further, it is likely that the frequency of occurrence of the cluster may play a role in the syllabification behaviour of the participants. The predictor of FREQUENCY – log ipm frequency of occurrence of the sequence as a word-initial onset, adopted from Šturm and Lukeš (2017) – was therefore added to the model, which increased its goodness-of-fit signif-

icantly ($\chi^2(1) = 4.6, p < 0.05$). However, the direction of the influence was unexpected: more frequent clusters had somewhat higher odds of division than less frequent clusters. There was no significant interaction of FREQUENCY and SONORITY. Figure 3 shows the effect plots for STRATEGY, SONORITY and FREQUENCY in terms of the probabilities of cluster division (= C.C syllabification).

The main investigated factor was MANIPULATION – we predicted that syllabification would be affected by changes in the temporal relation between C1 and C2 in the intervocalic cluster. The overall results do not support this conclusion: adding this effect into the model did not increase its goodness-of-fit significantly. Moreover, the interaction term for MANIPULATION*SONORITY was not significant either. However, the manipulation of C1 and C2 duration cannot be treated as equal for all items, since lengthening or shortening may not always change the value for the C1/C2 ratio. Therefore, we substituted MANIPULATION with a binary parameter ASYMMETRY (C1 is longer \times C1 is not longer)

Figure 3. Probability of cluster division (C.C syllabification) as a function of strategy (left), sonority type (middle) and log frequency (right). (O = obstruent, S = sonorant)

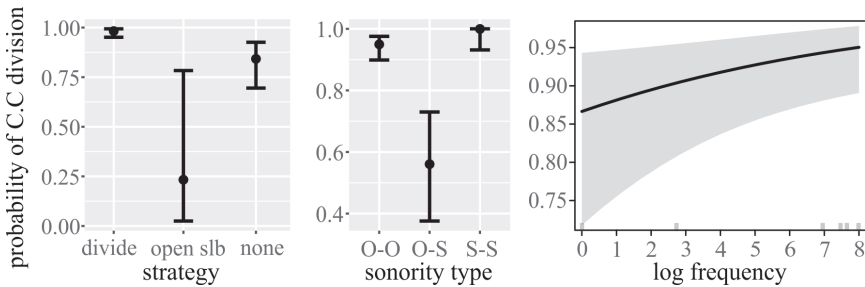


Figure 4. Correlation between C1/C2 ratio and proportion of cluster division (C.C syllabification) computed on word-by-manipulation cells. The regression line is indicated along with 95% confidence intervals of the regression model. Only speakers without a reported task strategy ($n = 13$).

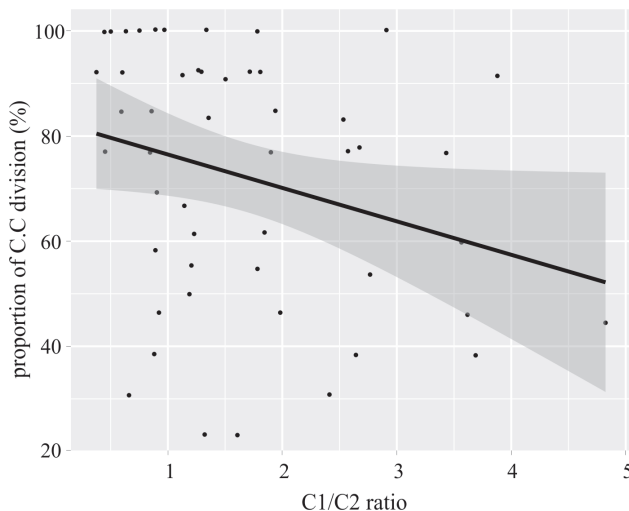
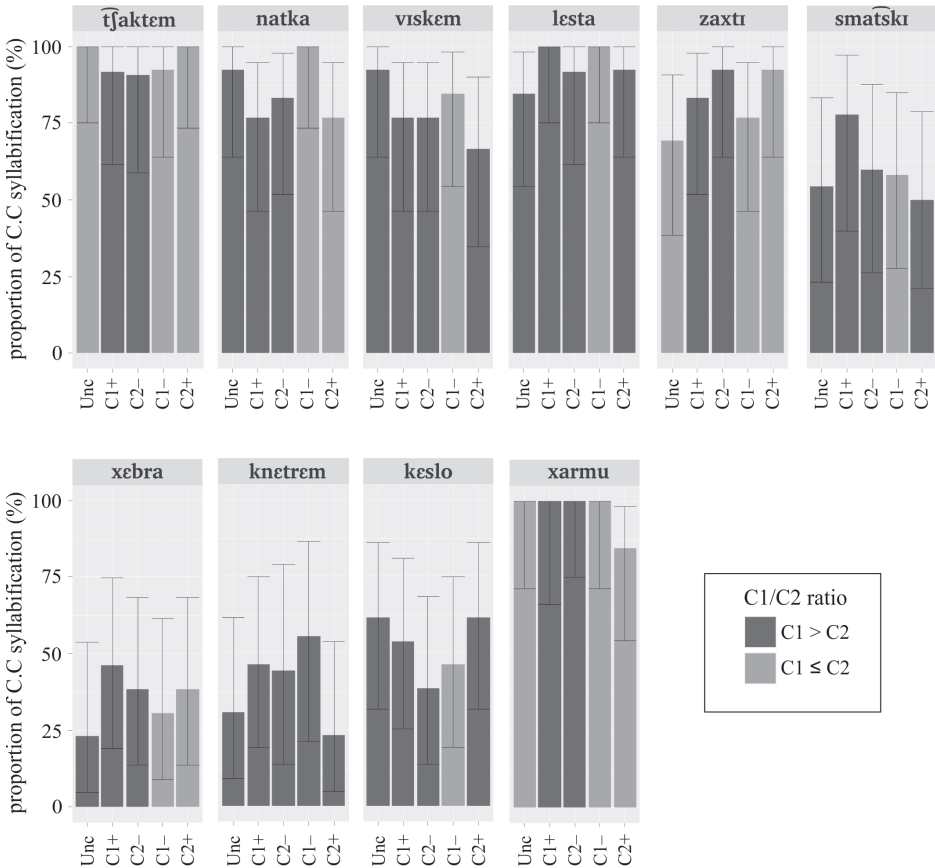


Figure 5. Proportion of cluster division (C.C syllabification) as a function of individual words and manipulation condition (Unc = unchanged, C1/C2 = 1st/2nd consonant, +/- = lengthened/shortened). The colour indicates whether or not the resulting C1/C2 ratio was greater than one (C1 > C2). The whiskers indicate 95% confidence intervals from a binomial test. Only speakers without a reported task strategy ($n = 13$).



in order to see whether the acoustic structure of the manipulated clusters is relevant. Although it reached smaller p -values than MANIPULATION, it was far from statistical significance ($\chi^2(1) = 0.2, p = 0.66$). Lastly, we filtered out all participants that reported to follow some strategy, narrowing the analysed sample to data from 13 participants. The effects observed in the subsample were by and large identical to the previous results, confirming the effects of SONORITY and FREQUENCY, and confirming the lack of effect of ASYMMETRY. Thus, not even the speakers without a task-related strategy seemed to be influenced by the acoustic manipulations in their syllabification behaviour.

A correlation analysis based on the subsample nevertheless showed a statistically significant relationship between the C1/C2 ratio and the proportion of C.C responses ($r = -0.28, p < 0.05$). This is visually represented in Figure 4: the higher the C1/C2 ratio, the lower the rate of cluster division (i.e., a greater preference for CC syllable onsets). However, the

weak correlation coefficient indicates that only 8% of the variance in the parameters could be explained. Additionally, the ratio was transformed into a binary variable *asymmetry* (like in the logistic model); listeners divided the stimuli with C1 longer than C2 68% of the time, while it was 80% of the time for stimuli with C1 shorter than or equal to C2 (the difference was not significant in a t-test, $p = 0.12$). The general preference, notwithstanding the C1/C2 ratio, is thus for cluster division. Durational manipulations of the items seem to exert only a small influence on the participants.

Finally, Figure 5 shows the proportion of cluster division for individual words and manipulations. The colour of the bars indicates whether or not the given item has C1 longer than C2 (for instance, the word [knetrem] had consistently longer C1s in all conditions, i.e., even shortening of C1 did not change the direction of the C1/C2 ratio). It is immediately apparent that the confidence intervals completely overlap for the manipulations within words, suggesting no change in syllabification across conditions. The only trend is that O-S clusters seem to behave differently from O-O or S-S clusters. With a possible exception of the word [smatski], individual words in a sonority group do not diverge substantially. Despite the lack of clear evidence suggesting an effect of manipulation in the hypothesized direction (higher C1/C2 ratio will lead to lower rate of cluster division), an important finding is that there was simultaneously no opposite effect, i.e. a lower ratio being associated with lower rates of division.

4. Discussion

The aim of the experiment was to establish whether the acoustic signal contains cues to the syllabification of words that are perceptually relevant. This has already been approached in previous research (Christie, 1974; Redford & Randall, 2005; also Christie, 1977 cited in Redford & Randall, 2005), but has not been investigated for Czech. The latter two studies concluded that when the first member of a two-consonant cluster is longer than the second member, listeners have a tendency to treat the cluster as the onset of the following syllable; in the opposite case, the probability of division of the cluster increases. Similarly, Christie (1974) discovered, on a synthesized speech token, that increasing the duration of the closure interval in intervocalic [st] was associated with a gradual increase in the proportion of C.C syllabifications.

The current experiment was designed to replicate these findings with Czech listeners. We repeat the hypotheses of the study here for the sake of convenience:

- H1: more V.CCV syllabifications are expected for tokens where the C1/C2 ratio is raised (or less V.CCV syllabifications where it is lowered);
- H2: sequences with a higher frequency of occurrence are more likely to be preserved as CC onsets than less frequent sequences;
- H3: clusters with a rising sonority are more likely to be preserved as CC onsets than clusters with a plateau in sonority.

The first hypothesis was not confirmed in the statistical analysis based on the whole data set. Manipulated items did not differ significantly from non-manipulated items. Depending on the duration ratios in the original stimulus, a critical boundary ($C1/C2 = 1$) could

be crossed by the manipulation, but it was not the case in all the words. Therefore, the lack of a manipulation effect could be explained in some items by the stability of the ratio. However, not even categorizing individual tokens into “C1 longer than C2” vs. “C1 not longer than C2” suggested any significant changes in the response patterns.

Furthermore, the extent of the acoustic manipulations was quite massive and above the just noticeable difference. We can expect that in natural speech the differences in duration would be of lower magnitude, which would obscure the potential syllable boundary effect even more. This suggests that the null hypothesis should not be rejected (H0: syllabification is not influenced by the duration of consonants in an intervocalic cluster). In a similar vein, the authors mentioned above admit that the results only concerned words for which the syllabification was ambiguous (Redford & Randall, 2005: 42–43), i.e., when there were several syllabifications options, all of them allowed by the phonotactics of the language. For the most part, our sample included precisely these cases (even the words /xarmu/ or /tʃaktem/ could be syllabified in other ways than C.C: compare the initial onsets in *rmoutit* /rmōucit/ or *který* /kteri:/). We can expect that syllabification of illegal sequences, ruled out by the phonotactics, would be even more resistant to acoustic manipulations, favouring invariably the C.C division. However, this prediction seemed to be compatible only with the illegal cluster in /zaxti/ but not with the illegal cluster in /smatski/, which was, quite unexpectedly, associated with a significant amount of V.CCV responses (*tsk/* will be discussed in detail below).

Importantly, it must be stressed that we only focused on the duration of consonants in the cluster. Thus the manipulations involved stretching or shortening of C1 (or C2) duration by 50%, while no manipulations were performed on the vowels. Yet it is clear that perception utilizes many other cues apart from the temporal structure of the intervocalic cluster, e.g. the C1/V1 ratio (see Kingston, Kawahara, Chambless, Mash, & Brenner-Alsop, 2009 for geminates; Maddieson, 1985). However, in a preliminary analysis this parameter did not seem to contribute to the syllabification responses in any way.

A possible reason behind the lack of a manipulation effect could be that some participants reported to have followed a certain strategy in the task, which might have reduced their sensitivity to the acoustic manipulations. Although the effects observed in a subsample of 13 participants – those who reported to have “no strategy” – were by and large identical to the previous results, a correlation analysis showed a statistically significant relationship between the C1/C2 ratio and the proportion of C.C responses (specifically, there was a greater preference for CC onsets with higher C1/C2 ratios). This finding at least is in accord with H1. Moreover, the results of our experiment – both of the subsample and of the whole data set – do not contradict the H1 in the sense of opposite direction. Although durational manipulations of the items seemed to exert none or only a small influence on the participants, there was simultaneously no change in syllabification across conditions that would suggest that a *lower* C1/C2 ratio is associated with *higher* rates of V.CCV syllabifications.

The prediction of H2 was not borne out. On the contrary, a positive – not negative – correlation was ascertained between cluster frequency of occurrence and the probability of cluster division. For instance, the relatively frequent clusters /st/ and /sk/ were most often split into two syllables, which counters the expectation. Moreover, the phonotactically illegal cluster *tsk/* was not predominantly split, as would be expected, but was

ambivalent between C.C and V.CC syllabifications. Thus, a substantial number of participants² produced an ill-formed onset cluster in response to the word /smatskɪ/, which seems to contradict the well-formedness principle whereby only syllables encountered at the edges of a word result from word-medial syllabification. If this is the case, then either the phonotactic principle should not be given such a prominent place in syllabification, or speakers might not perceive the /tsk/ cluster as illegal (e.g., they might treat its absence from Czech word onsets as an accidental gap). Figure 5 reveals that the patterns of /tsk/ were quite similar to the obstruent-sonorant clusters, especially /sl/, but the data offer no clear explanation for this behaviour, apart from the fact that /tsk/ is the only cluster with an affricate sound as its member. However, Šturm (2017, p. 89) found in a similar experiment using genuine Czech words that the proportion of C.C syllabification of the clusters /tsk/ and /tʃk/ was higher, approximately 80%, which is more in line with the hypothesis. Therefore, an alternative explanation concerns the material used: since the participants in the current experiment responded to nonsense words, they might have treated the sequences differently from real speech material (namely, with more benevolence towards certain sequences). This will be discussed in more detail below.

With regard specifically to the cluster /kt/, one explanation may relate to how the frequency was computed. The counts are based on written corpora, where the cluster /kt/ – associated with one of the highest rates of C.C division in the experiment – is more common than in spoken corpora due to the frequent use in written texts of relative clauses with the pronoun *kerý* (“which/who”). However, it must be admitted that a separate experiment is needed for investigating the effect of cluster frequency on syllable division. In the current state, only 10 clusters were taken into account, which is too small a number given that there were also differences in sonority and manner of articulation that might represent more decisive factors in syllabification.

H3, concerning the sonority type of the cluster, was substantiated by the results of the experiment and is completely in compliance with previous findings (Goslin & Frauenfelder, 2001; Ní Chiosáin et al., 2012). Obstruent-sonorant clusters were most frequently maintained as onsets, whereas clusters of two obstruents were more often divided, and the cluster of two sonorants was divided almost always. In fact, the difference between the latter two – i.e. clusters with sonority plateaus – was not substantial. The only exception was /tsk/, which has already been discussed. It is especially interesting to compare plosive-plosive clusters with fricative-plosive clusters. In several approaches to the sonority hierarchy, but not in some others, fricatives are placed higher on the scale than plosives (Zec, 2007: 178; Gordon, 2016: 99). Since the /sk st xt/ clusters would then violate the sonority profile in the syllable onset, we may expect a higher proportion of C.C syllabifications compared to the plosive-plosive clusters. This seems to be the case possibly with /st/, but not with the other two clusters. However, the very similar behaviour of both types of clusters does not necessarily present a case for treating fricatives and plosives together as obstruents because sonority plateaus, represented by the plosive-plosive sequences, are avoided as well. In other words, both approaches to sonority classes would lead to the same conclusion, namely, a strong preference of C.C. syllabifications.

² Eight out of thirteen participants in the subset. Interestingly, all but one were associated with the highest proportions of CC onsets generally according to Figure 2. Therefore, the increased number of CC syllabifications might reflect their general strategy rather than the specifics of the /tsk/ cluster.

However, the case of /sl/ is different. The concept of minimal sonority distance (see Zec, 2007) assumes that a certain difference in sonority is needed between the first and second member of a CC onset. If fricatives are higher on the sonority scale than plosives, then the difference from liquids is smaller for /s/ than for plosives, implying that e.g. a /pl/ cluster is more ideal than a /sl/ cluster. Although Figure 5 suggests that /sl/ might be more often divided by the participants than /br/ and /tr/, the difference is not significant. Moreover, it could be the case that the ambiguous syllabification of /keslo/ (bordering around 50% of C.C or CC responses) is related to the ambiguous phonetic segmentation of the cluster, as shown in Figure 1. We do not know whether the devoiced part of the lateral belongs, perceptually, to the fricative or to the approximant. An analysis of individual participants revealed that the V.CCV syllabification was again especially linked to those speakers who, compared to the other speakers in the subset, generally produced more CC onsets (Fig. 2, see also note 2). With regard to the extreme cases, speakers 5 and 17 syllabified all five tokens of /keslo/ as V.CCV, whereas speakers 7 and 13 syllabified them in all instances as VC.CV. Five other speakers showed a less strong preference for one of the options, and four speakers did not incline to either of the options.

With the exception of the /rm/ cluster, which was almost unanimously divided, the syllabification outputs were not clear-cut (dichotomous, either-or). Since eight out of the ten clusters were phonotactically legal (including /rm/, although it is not frequent word-initially), the prevailing strategy should have been that of *onset maximization*. This is a common assumption of many researchers and writers about syllabification (Pulgram 1970; Kahn, 1976; Fallows, 1981; Hall, 2006; see also Bičan, 2017), but it is clearly contradicted by the results of the current experiment and of other experiments in Šturm (2017). The clusters were syllabified – regardless of whether the two illegal clusters were included – as VC.CV in 78% of cases, whereas the V.CCV division, conforming to onset maximization, occurred in only 20% of the cases. Moreover, there were 15 cases (1%) of VCC.V syllabification. Although we might discard these outputs as marginal (and we indeed excluded them from the main analysis), the syllabification pattern nevertheless occurred and it represents further evidence against the maximal onset principle. The data come from two participants (S24 and S26) and 6 words/clusters (in descending order of frequency of occurrence: /xarmu/, /zaxtu/, /tʃaktem/, /viskem/, /lesta/, /natka/). Four of the clusters display a falling sonority pattern typical of syllable codas, and two have sonority plateaus. Thus, despite the speaker-specificity, the occurrence cannot be declared unexpected or unnatural.

Our decision to use nonsense words entails the acceptance of some assumptions along with it. As pointed out by a reviewer, one implicit assumption is that Czech participants pronounce non-Czech words like Czech words. Yet it is common to experience difficulties – and change the tempo or manner of speech, for instance – when we encounter an unfamiliar word in a text. This may have contributed to the significant number of invalid responses in the experiment, such as hesitations or slips of the tongue. However, the participants were expecting nonsense words to appear because they familiarized with the task and the range of stimuli in the training session. Moreover, there is no reason to believe that the participants would pronounce a nonsense word like /xarmu/ any differently from the Czech word /farmu/. The second assumption, saying that Czech participants *syllabify* non-Czech words like Czech words, is more difficult to substantiate. We can only highlight

again the close similarity between the nonsense words used and genuine Czech words. Also, the results of the experiment in terms of the sonority factor seem to suggest that the two types of stimuli are to a large degree treated similarly (but note the deviant cluster / $\overline{\text{tsk}}$ /). A crucial difference is the exclusion of lexical information; however, this is rather beneficial, since it prevents any morphological effects from influencing the results.

Finally, 12 out of 25 analysed participants reported that they followed a strategy in the behavioural task, such as onset maximization or cluster division, despite being explicitly instructed not to do so. There were no significant differences between male and female participants, except that males had a greater tendency towards CC onsets regardless of the stated strategy. Moreover, it is questionable whether the 13 remaining participants really performed the task “without a strategy”, as they stated. This might be a serious limitation to the study. Figure 2 suggests that the “no strategy” group forms two clusters of listeners, with five participants resembling the performance of subjects from the “divide clusters” strategy. However, it is very difficult to persuade participants to break free from all syllabification rules and other deep-rooted habits, if not outright impossible. In any case, future research would benefit from a larger sample of participants.

ACKNOWLEDGEMENT

This research was supported by the Charles University project Progres 4, *Language in the shiftings of time, space, and culture*.

REFERENCES

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bičan, A. (2017). Slabikování. In P. Karlík, M. Nekula, & J. Pleskalová (Eds.), *Nový encyklopedický slovník češtiny online*. Retrieved from <http://www.czechency.org>
- Boersma, P., & Weenink, D. (2014). Praat – doing phonetics by computer (Version 5.4). Retrieved from <http://www.praat.org>
- Christie, W. M. (1974). Some cues for syllable juncture perception in English. *Journal of the Acoustical Society of America*, 55(4), 819–821.
- Clements, G. N. (2009). Does sonority have a phonetic basis? Comments on the chapter by Bert Vaux. In E. Raimy & C. E. Cairns (Eds.), *Contemporary Views on Architecture and Representations in Phonological Theory* (pp. 165–175). Cambridge, Mass: MIT Press.
- Coleman, J. (2002). Phonetic representations in the mental lexicon. In J. Durand & B. Laks (Eds.), *Phonetics, Phonology, and Cognition* (pp. 96–130). Oxford: Oxford University Press.
- Côté, M.-H., & Kharlamov, V. (2011). The impact of experimental tasks on syllabification judgments: A case study of Russian. In C. Cairns & E. Raimy (Eds.), *Handbook of the Syllable* (pp. 273–294). Leiden: Brill.
- Ewen, C. J., & van der Hulst, H. (2001). *The Phonological Structure of Words: An Introduction* (Vol. 39). Cambridge: Cambridge University Press.
- Fallows, D. (1981). Experimental evidence for English syllabification and syllable structure. *Journal of Linguistics*, 17(2), 309–317. <https://doi.org/10.1017/S0022226700007027>
- Forster, K. I., & Forster, J. C. (2003). DMDX: a windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc*, 35(1), 116–124.

- Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America*, 101(6), 3728–3740. <https://doi.org/10.1121/1.418332>
- Goldinger, S. D., & Azuma, T. (2003). Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics*, 31(3–4), 305–320. [https://doi.org/10.1016/S0095-4470\(03\)00030-5](https://doi.org/10.1016/S0095-4470(03)00030-5)
- Gordon, M. K. (2016). *Phonological Typology*. Oxford: Oxford University Press.
- Goslin, J., & Frauenfelder, U. H. (2001). A comparison of theoretical and human syllabification. *Language and Speech*, 44(4), 409–436. <https://doi.org/10.1177/00238309010440040101>
- Hall, T. A. (2006). English syllabification as the interaction of markedness constraints. *Studia Linguistica*, 60(1), 1–33. <https://doi.org/10.1111/j.1467-9582.2006.00131.x>
- Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, 31(3–4), 373–405. <https://doi.org/10.1016/j.wocn.2003.09.006>
- Hay, J., Pierrehumbert, J., & Beckman, M. (2004). Speech perception, well-formedness and the statistics of the lexicon. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in Laboratory Phonology VI* (pp. 58–74). Cambridge: Cambridge University Press.
- Kahn, D. (1976). *Syllable-based generalizations in English phonology [dizertační práce]*. Cambridge, Mass.: MIT.
- Kingston, J., Kawahara, S., Chambless, D., Mash, D., & Brenner-Alsop, E. (2009). Contextual effects on the perception of duration. *Journal of Phonetics*, 37(3), 297–320. <https://doi.org/10.1016/j.wocn.2009.03.007>
- Machač, P., & Skarnitzl, R. (2009). *Principles of Phonetic Segmentation*. Praha: Epoque.
- Maddieson, I. (1985). Phonetic cues to syllabification. In V. Fromkin (Ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged* (pp. 203–221). New York: Academic Press.
- Munson, B. (2001). Phonological pattern frequency and speech production in adults and children. *Journal of Speech, Language, and Hearing Research*, 44(4), 778–792.
- Ní Chiosáin, M., Welby, P., & Espesser, R. (2012). Is the syllabification of Irish a typological exception? An experimental study. *Speech Communication*, 54(1), 68–91. <https://doi.org/10.1016/j.specom.2011.07.002>
- Parker, S. (2008). Sound level protrusions as physical correlates of sonority. *Journal of Phonetics*, 36(1), 55–90. <https://doi.org/10.1016/j.wocn.2007.09.003>
- Port, R. (2007). How are words stored in memory? Beyond phones and phonemes. *New Ideas in Psychology*, 25(2), 143–170. <https://doi.org/10.1016/j.newideapsych.2007.02.001>
- Pulgram, E. (1970). *Syllable, Word, Nexus, Cursus*. The Hague: Mouton.
- R Core Team. (2016). R: A language and environment for statistical computing (Version 3.2.4). Vienna: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Redford, M. A., & Randall, P. (2005). The role of juncture cues and phonological knowledge in English syllabification judgments. *Journal of Phonetics*, 33(1), 27–46. <https://doi.org/10.1016/j.wocn.2004.05.003>
- Schiller, N. O., Meyer, A. S., & Levelt, W. J. M. (1997). The syllabic structure of spoken words: evidence from the syllabification of intervocalic consonants. *Language and Speech*, 40, 103–140.
- Sendlmeier, W. F. (1995). Feature, phoneme, syllable or word: how is speech mentally represented? *Phonetica*, 52(3), 131–143.
- Šturm, P. (2017). *Určování slabičných hranic v češtině [dizertační práce]*. Praha: Filozofická fakulta Univerzity Karlovy.
- Šturm, P., & Lukeš, D. (2017). Fonotaktická analýza obsahu slabik na okrajích českých slov v mluvené a psané řeči. *Slovo a slovesnost*, 78(2), 99–118.
- Treiman, R., & Danis, C. (1988). Syllabification of intervocalic consonants. *Journal of Memory and Language*, 27(1), 87–104. [https://doi.org/10.1016/0749-596X\(88\)90050-2](https://doi.org/10.1016/0749-596X(88)90050-2)
- Treiman, R., Kessler, B., Knewasser, S., Tincoff, R., & Bowman, M. (2000). English speakers' sensitivity to phonotactic patterns. In N. B. Broe & J. Pierrehumbert (Eds.), *Laboratory Phonology V: Acquisition and the Lexicon* (pp. 269–282). Cambridge: Cambridge University Press.
- Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, (40), 47–62.
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.
- Zec, D. (2007). The syllable. In P. de Lacy (Ed.), *The Cambridge Handbook of Phonology* (pp. 161–194). Cambridge: Cambridge University Press.

APPENDIX

Temporal structure of the target stimuli (V = first vowel, C1 = first consonant in the intervocalic cluster, C2 = second consonant). Onset frequency was adopted from Šturm and Lukeš (2017), and refers to the ipm (items per million) frequency of occurrence of the cluster as a word-initial onset in written texts.

sonority	cluster	word	onset freq.	V dur (ms)	C1 dur (ms)	C2 dur (ms)	C1/C2 ratio				C1/V ratio		C2/V ratio	
							unchanged	C1 long	C2 short	C1 short	C2 long	unchanged	unchanged	unchanged
O-O	/sk/	viskem	2 100	67	163	95	1.72	2.57	3.43	0.86	1.14	2.43	1.42	
O-O	/st/	lesta	8 371	75	192	99	1.94	2.91	3.88	0.97	1.29	2.56	1.32	
O-O	/xt/	zachty	0	82	122	135	0.90	1.36	1.81	0.45	0.60	1.49	1.65	
O-O	/kt/	čaktem	6 380	71	100	133	0.75	1.13	1.50	0.38	0.50	1.41	1.87	
O-O	/tk/	natka	15	82	133	105	1.27	1.90	2.53	0.63	0.84	1.62	1.28	
O-O	/tsk/	smacky	0	70	205	115	1.78	2.67	3.57	0.89	1.19	2.93	1.64	
O-S	/br/	chebra	1 049	67	111	84	1.32	1.98	2.64	0.66	0.88	1.66	1.25	
O-S	/tr/	knetrem	1 748	80	123	51	2.41	3.62	4.82	1.21	1.61	1.54	0.64	
O-S	/sl/	keslo	2 950	88	166	90	1.84	2.77	3.69	0.92	1.23	1.89	1.02	
S-S	/rm/	charm	0	102	98	110	0.89	1.34	1.78	0.45	0.59	0.96	1.08	

RESUMÉ

Cílem studie bylo ověřit, zda akustický signál obsahuje percepčně relevantní vodítka ohledně dělení slov na slabiky. Sylabifikační odezvy 27 českých mluvčích byly zkoumány v behaviorálním experimentu za použití dvojslabičných pseudoslov s deseti shluky CC. Poměr trvání C1 a C2 intervokálního shluku byl manipulován čtyřmi způsoby: zkrácením/prodloužením prvního/druhého konsonantu. Úkolem mluvčích bylo opakovat slyšené stimuly po slabikách (metoda vkládání pauz). Logistická regrese potvrdila signifikantní efekt sonoritního typu shluku, fonotaktiky na okrajích slov a sylabifikační strategie (jen polovina účastníků uvedla, že se v experimentu neřídila žádnou strategií). Typ manipulace se na jednu stranu neprojevil jako signifikantní prediktor v logistické analýze; na druhou stranu poměr C1/C2 v souladu s hypotézou negativně koreloval s mírou rozdělení shluku. Tato korelace naznačuje, že čím je C1 delší než C2, tím vyšší má shluk pravděpodobnost, že bude zachován jako přetura následující slabiky.

Pavel Šturm
Institute of Phonetics
Faculty of Arts, Charles University
pavel.sturm@ff.cuni.cz