Jaw movement in vowels and liquids forming the syllable nucleus

Štefan Beňuš [1,2], Marianne Pouplier[3]

[1] Department of Eng. and Am. Studies, Constantine the Philosopher University, Nitra, Slovakia
[2] Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia
[3] Institute of Phonetics and Speech Processing, Ludwig-Maximilians-University, Munich, Germany
sbenus@ukf.sk, pouplier@phonetik.uni-muenchen.de

Abstract

This paper investigates jaw movements in the production of Slovak syllables with and without vowels. We test the hypothesis that /l, r/ in the syllable nucleus position show a degree of jaw opening comparable to vowels, therefore providing a rising-falling sonority profile even in syllables lacking vowels. We also investigate whether the phonemic length distinction occurring for both vowels and syllabic consonants is implemented in a similar fashion for the different nucleus types. Our articulatory data show that the jaw activity during syllabic liquids is indeed comparable to that of vowels, and that the jaw is recruited to help maintain the main lingual articulation. This became evident in particular in an interaction between nucleus type and phonemic length effects.

Index Terms: syllabic consonants, sonority, jaw cycle

1. Introduction

The syllable is unquestionably a fundamental unit of speech that plays a crucial role in both segmental and suprasegmental aspects of spoken language. Advancements in the identification and processing of features describing syllables have potential for improving the quality and reliability of models used in automatic processing of speech [e.g. 1, 2].

The syllable has often been defined in terms of acoustic sonority [3]: Syllables constitute rising and falling cycles of acoustic energy. It is generally agreed that the degree of jaw opening is one of the major determinants of sonority. Vowels, which generally have a lower jaw position compared to consonants and allow thus greater radiation of acoustic energy, occupy the syllable nucleus. Consonants, on the other hand, have generally a more closed jaw position associated with lower acoustic energy, and typically form syllable margins. Moreover, consonants within a syllable are preferably arranged so that those with a higher jaw position (e.g., /s/) are further away from the nucleus than those with a lower jaw position (e.g. /l/). Hence /sla/ is cross-linguistically a more common syllable than /lsa/. Segments are preferably arranged such that there is a regular opening and closing cycle of the jaw, giving rise to increasing and falling sonority [4]. Furthermore, the positive correlation between sonority and jaw opening has been shown for various prosodic features such as accent or emphasis [5, 6]. The alternating pattern of sonority and jaw-openings peaks and valleys is thus crucially important for speech segmentation in humans and its deeper understanding may prove useful in speech technology as well.

While syllabic consonants are considered as a marginal phenomenon in English since they can occur only in unstressed positions, some languages allow vowel-less syllables to occur in an almost unrestricted fashion. Slovak, for example, similarly to some other Slavic languages, freely allows alveolar liquids /l/ and /r/ to function as syllable nuclei in both stressed and unstressed syllables. This results in entire words containing no vowels (*vlk* 'wolf' vs. *vak* 'bag', *mrk* 'wink' vs. *mak* 'poppy') and allows for long consonant sequences (e.g. *štvrťvrstva* [ʃtvrc.vrs.tva] 'quarter layer' with 10 consonants in row arranged into three syllables with the first two /r/s forming syllable nuclei). The consonantal sequences run counter the general consonant-vowel alternation pattern underlying spoken language, and it is an open question whether and how these sequences participate in sonority and jaw-opening patterns. A detailed investigation of jaw position changes in syllables with consonantal nuclei thus contributes to our understanding of how the acoustic signal is modulated in the absence of vowels. The current paper is the first one to provide a comprehensive investigation of jaw position in syllables with syllabic consonants on the basis of Slovak /l, r/ syllable nuclei.

It is well known that consonants differ in their intrinsic jaw position. Although alveolar consonants have generally a higher jaw position than velars and labials [7], German non-syllabic liquid /l/ was reported as having a relatively low jaw position compared to alveolar consonants of different manner [8]. This opens the possibility that liquids can serve as consonantal nuclei because, due to their low jaw position, enough acoustic energy is radiated to provide a sonority contrast to the surrounding consonants. Whether this reasoning could be extended to trills is unclear: To our knowledge, the jaw position during (syllabic) trills has not been previously studied. If, similarly to German non-syllabic /l/, a relatively low jaw position can be shown for both Slovak liquids in the nucleus position, this may explain why it is cross-linguistically /l, r/ that are common syllabic consonants.

Finally, Slovak liquids are phonologically similar to vowels not only in that they are able to form a syllable nucleus, but also in that they both signal phonemic contrast in quantity. Both consonantal and vocalic nuclei can be contrastively short and long (*vykŕm* [vi.kr:m] 'to feed' vs. *výkrm* [vi:.krm] 'feeding', *sĺz* [sl:z] 'tear-Gen.Pl.' vs. *slz* [slz] 'to tear-Imper.'). Assuming that longer duration decreases target undershoot, jaw movement in the long nuclei may reveal a difference in jaw targets which may be minimal for the short nuclei. Furthermore, the comparison between the short and long nuclei allows testing if the syllabicity of liquids, and their functional similarity to vowels can also be observed in any potentially present systematic variations in jaw movement as a function of phonemic length.

The examination of the articulatory control of the phonological quantity contrasts is also important for speech processing since it presents a difficult challenge for achieving reliable recognition and natural-sounding synthesis of speech. This is due to a non-trivial relationship between the segmental and prosodic requirements for durational variation since the nucleus duration simultaneously signals phonological quantity, and participates in cuing the presence of pitch accents, prosodic boundaries, or changes in speech rate.

In sum, in this paper we examine jaw opening as one of the major determinants of signal modulation in syllables with and without vowels, and we investigate to what extent the phonological durational contrast in Slovak nuclei is similarly implemented for vowels and consonants in terms of jaw kinematics.

2. Methodology

Two female and three male native speakers of Slovak read identical sentences containing nonsense words in the form of C_iNC_ia, in which C represents one of three voiceless stop consonants {/p/, /t/, /k/}, and N represents all 14 possible nuclei in Slovak: 10 vowels {[i], [e], [a], [o], [u], [i:], [e:], [a:], [o:], [u:]} and 4 liquids {[r], [l], [r:], [l:]}. Tokens [tlta] and [tl:ta] represent sequences that do not occur in Slovak and were not collected. Subjects repeated each sentence 5 times at a normal rate and 5 times at a fast rate, which produced 2000 tokens in total (3 consonant and 14 nuclei minus [tlta] and [tl:ta] makes 40 tokens x 10 repetitions x 5 subjects).

Acoustic and articulatory data were collected; the former at a 16 kHz sampling rate with a Sennheiser MKH 40 microphone and the latter by means of electro-magnetic articulography at a sampling rate of 200Hz (AG 500, Carstens Medizinelektronik, IPS Munich). One of seven sensors tracking the movements of active articulators was placed on the lower incisors to record jaw movement.

A trained annotator manually aligned boundaries between the consonants and vowels to the acoustic signal following standard procedures and using primarily the visual information in the oscillogram and spectrogram. From this labeling, we took the temporal landmarks of the acoustic release of the first consonant and the onset of the closure of the second consonant, and extracted the time function of the vertical jaw movement between these two landmarks for further analyses.

Our dependent measures include the vertical jaw position at the temporal midpoint between the two acoustic landmarks (coinciding roughly in the acoustic middle of the nucleus). We also calculated an estimate of displacement as the difference between the jaw position at the midpoint and at the acoustic release of the first consonant. Additionally, we performed a discrete cosine transform on the extracted time functions of vertical jaw movement and examined the third coefficient of this transform, which describes the curvature of the function [9]. Finally, we extracted the acoustic duration of the nucleus.

A mixed-model approach as implemented in the R software package is used for statistical testing, since it allows us to specify multiple random factors [10]. We test the effect of three independent factors: nucleus type (7 levels: 5 vowels, /l/, /r/), consonant type (3 levels: /p, t, k/), and phonological length (2 levels: long, short). Subject is a random factor in all tests, which effectively normalizes differences between subjects. Due to the known difficulty of computing degrees of freedom in a linear mixed model, we report the output F-values as the main indicator of the robustness of effects and employ a conservative approach for estimating the p-values [11] in which $F > 8.49$ is considered significant at $p < 0.01$ for data with degrees of freedom well beyond 60.

3. Results

3.1. Jaw height at nucleus midpoint

Our first analysis concerns the question whether syllabic liquids both show a jaw height comparable to that of vowels at nucleus midpoint. Fig. 1 shows that the jaw height of vowels roughly corresponds to the phonological division of low /a/, mid /e/, /o/, and high /i/, /u/ vowels. This contrast is most robust for the p_p environment, slightly less for k_k, and the least for t_t in which /i/ is more similar to mid vowels than to /u/. The broad pattern immediately observable in the figure is that the jaw position of liquids is always within the range of the position of vowels. The factor of nucleus type significantly affects jaw height at the temporal midpoint of the nucleus ($F = 179.8$ with all other factors treated as random). Separate pair-wise tests for /l/ and five other vowels and treating all other factors as random (Subject, Consonant, Length) show that the jaw position for /l/ is significantly higher than for /a/, /e/, /o/, and significantly lower than for /i/, /u/. The smallest separation was observed for the /l/-/i/ pair ($F = 18.6$). This result holds even when /t/ is excluded; that is, the absence of *tlta* tokens does not bias the results. The same observation applies to /r/ in that it shows a significantly higher jaw position than /a/, /e/, /o/, and a significantly lower position compared to /i/, /u/ with the smallest difference from /o/ ($F = 20.1$). The two liquids do not differ significantly from each other ($F = 0$).

The effect of flanking consonants is highly significant in that the jaw position during the nucleus is lowest in the /p/ context, medium in the /k/ and highest in the /t/ context ($F = 768.8$ with all other factors treated as random). This effect is robustly present for each nucleus type.

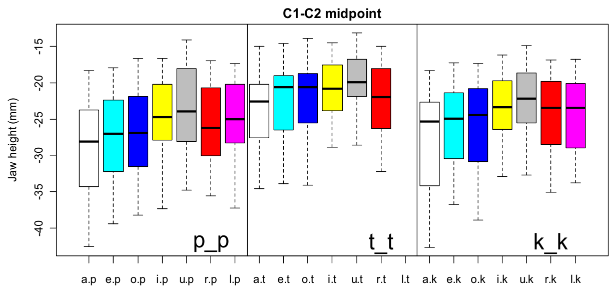

Figure 1: *Jaw height at the temporal midpoint of the nucleus split by consonant type and nucleus type; the 5 leftmost bars represent vowels ordered /a/, /e/, /o/, /i/, /e/, followed by consonants /r/, /l/.*

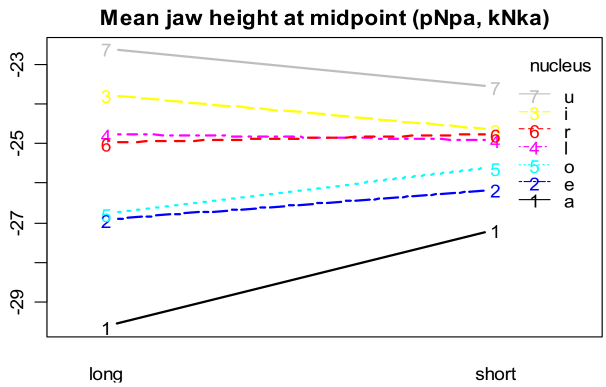

Figure 2: *Mean jaw height at the temporal midpoint of the nucleus divided by nucleus length and type.*

Fig. 2 shows the data collapsed across the labial and velar contexts, separately for the two phonemic lengths. Long nuclei /a/, /o/ and /e/ have lower jaw positions than the short ones (in that order of robustness). The positions for the short and long liquids are similar, but the relationship for high vowels is reversed: long high vowels show a higher jaw than their short counterparts. Consonant type t_t is omitted from this plot due to the absence of *tlta* tokens. The mean jaw height of liquids appears between those for /i/ and /o/ for both long and short nuclei, corroborating thus the observation in Fig. 1. Statistically, the effects of phonemic length on jaw height is significant, albeit the least robust of the three tested factors (F = 50.7 with all other factors treated as random), and interacts significantly with nucleus type (F = 25.6).

3.2. Nucleus duration

We now examine how the phonological length distinction affects the acoustic duration of nuclei. Fig. 3 shows that phonologically long nuclei were robustly longer than the short ones (F = 3429). Nucleus type had a significant effect (F = 17.4), which, however, was much weaker than the effect of nucleus type on the jaw height at midpoint (F = 179.8). The interaction between the two factors was not significant (F = 5.0). The nuclei /r/, /l/ and /a/ did not differ from each other, and all three were significantly longer than other four nuclei.

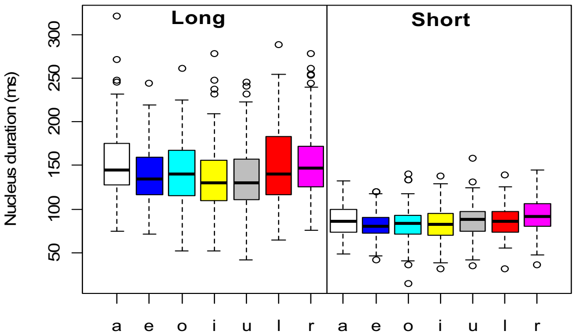

Figure 3: *Acoustically determined duration of nuclei.*

3.3. Displacement and curvatures

Displacement, shown in Fig. 4, was defined as the difference in jaw height between C1 offset and midpoint (negative values mean lowering, positive raising). On this measure, nucleus type has a significant effect (F = 167.6). Pair-wise tests show that /l/ does not differ from /i/ (F = 2.2), /r/ is similar to /o/ (F = 0.7), and /l/ is less displaced than /r/ (F = 36.3). The effect of phonological length on displacement is qualitatively similar to its effect on midpoint jaw position presented in Fig.2: /a/ and /o/ are affected the most, other nuclei minimally, and long /u/ show clear positive displacement (i.e., jaw raising).

Curvature of the time function describes the difference of the jaw position at the midpoint with respect to the position at the release of the preceding consonant and target achievement of the following one. Hence, assuming that nuclei have lower jaw positions than the surrounding onset consonants /p, t, k/, the curvatures should be positive, and flatter for high and short nuclei than for low and long ones. Nucleus type affects curvatures significantly (F = 140). Visual inspection of Fig. 5 shows that the effect of nucleus type on movement curvatures is similar to its effect on midpoint and displacement: curvatures are greatest for /a/, medium for /e/, /o/, the smallest for /i/ and /u/, and negative for long /u/. /l/ has similar curvatures to /i/ (not significantly different, F = 0.4) but significantly greater curvature values than /u/ (F = 52.6), and significantly smaller values than /a/, /e/, /o/. /r/ has significantly flatter curvatures than /a/ but rather similar ones to both mid vowels (F = 6.4 for /o/ and F = 9.6 for /e/), and significantly greater curvature values than high vowels and /l/ (F = 15.7). Hence, for this measure, the tendencies observed earlier for the jaw height receive further corroboration: /r/ patterns with mid vowels and /l/ patterns with /i/. In contrast, the pattern observed for curvatures and displacement are different from the ones for nucleus length for which /a/, /l/, /r/ pattern similarly and are different from the other 4 vowels.

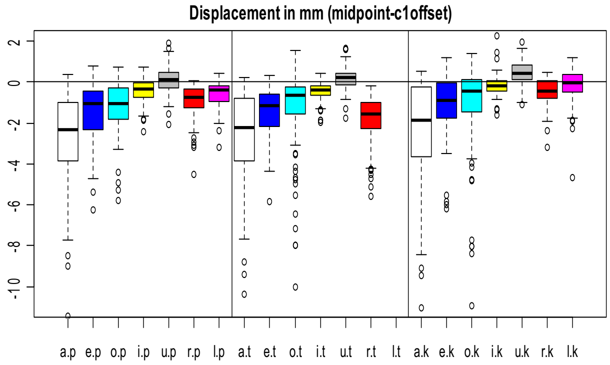

Figure 4: *Jaw displacement (midpoint – c1-offset) split by the flanking consonants and nucleus types.*

The effects of phonemic duration on displacement and curvatures are significant: short nuclei have flatter curvatures and smaller displacements than long ones (F = 144.7 and F = 275.6 respectively). There is also a significant interaction for both measures between duration and nucleus type (F = 45.0 and F = 71.8, respectively). The patterns for these derived measures are similar to the results obtained for jaw height: curvatures and displacements for /a/ are affected the most, /r/s are affected to an intermediate degree, /l/ and /i/ are affected minimally, etc. Long /u/ has more negative curvatures than short /u/, which, together with Fig. 2 and 4, suggests that for /u/, long vowels have a higher target jaw position than the short ones, especially in the k_k context. This might be due to the fact that /k/ and /u/ have very similar tongue positions, so they reinforce each other.

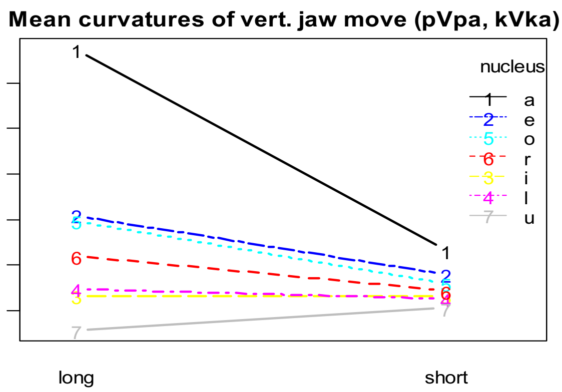

Figure 5: *The effect of phonemic length and nucleus type of the curvatures of vertical jaw movements.*

The effect of consonant type on displacement and curvature is also significant (F = 60.7): /k/ has the flattest trajectories and smallest displacements, /t/ has the most curved trajectories and greatest displacements, and /p/ shows intermediate values differing significantly from /t/ for curvature (F = 32.9) but not for displacement (F = 7.4). It is interesting that the relationship between consonants on the curvature measure differs from the jaw height results: the highest positions and greatest curvatures can be observed in t_t, but p_p shows the lowest jaw height and medium curvatures while k_k shows a medium jaw height and smallest curvature. Hence, nucleus jaw positions are in general more influenced by a flanking /k/ than by a flanking /p/.

Finally, Table 1 shows that greater jaw lowering correlates with greater curvature very robustly. Since displacement is calculated between the midpoint and c1-offset and curvatures between c1-offset and c2-onset, jaw height at c2-offset affects jaw lowering minimally. Longer nucleus duration only modestly correlated with lower jaw position at the midpoint, especially compared to the strength of the correlation with derived measures of displacement and curvature. This suggests an important effect of the jaw position at the offset of the consonant preceding the nucleus, i.e. the range of jaw movement is a more important factor than jaw position per se. Other relationships show a medium strength.

Table 1. *Correlations between dependent variables, all significant at p < 0.001*

Correlations	nuc.dur	midpoint	displ.	curv.
nuc.dur		-.08	-.32	.42
jaw height	-.08		.50	-.41
displacement	-.32	.50		-.94
curvature	.42	-.41	-.94	

4. Discussion

The goal of this paper was to examine degree of jaw opening for syllables with and without vowels. The hypothesis was that /l, r/ can serve as syllabic nuclei because, when in nucleus position, they show a degree of jaw opening similar to vowels. Another question we investigated was whether the phonemic length distinction occuring for both vowels and syllabic consonants in Slovak is implemented in a similar fashion in terms of jaw movement. Our data show two broad patterns. First, the temporal and spatial features of vertical jaw movement of consonantal and vocalic nuclei are comparable since the values for syllabic liquids were always within the range of vowels, determined always by the extremes of /a/ and /u/ vowels, in all examined variables. Second, jaw targets depend on the different tongue position requirements of the nucleus type. This is shown in the comparison of long and short vowels: Long high vowels have a higher jaw position compared to short high vowels. The opposite was the case for low vowels: the longer nucleus, the lower the jaw. For long nuclei, there will be less duration dependent undershoot [12]. For high vowels this means, that a higher tongue position will be achieved, hence, in a synergistic effect, the jaw will come higher to aid the tongue to achieve the high target position. Low vowels have a pharyngeal target, for which the lower jaw position will provide synergistic effects.

The first observation is in line with the proposal that the syllable should be defined on the basis of an independent jaw cycle that governs the jaw positions of segmental articulations [12]. In this view, segmental articulations accommodate their intrinsic jaw position to the basic opening-closing cycle of the jaw and a relatively low jaw position observed for syllabic liquids may be dictated by the jaw cycle. However, further corroborating evidence would need to show that liquids in nucleus position have a lower jaw position compared to their non-syllabic /l, r/ in onset or coda position. Future research will address this question. A relative independence of jaw within the functional synergy of the lips and jaw is also supported in our data by the observation that the p_p environment results in the robustly lowest jaw positions of the nuclei compared to t_t and k_k environments, despite the requirement for a high lower lip position in forming the bilabial closure.

Interestingly, our second observation concerning a higher jaw position for long high vowels suggests that the lingual constriction degree of the nucleus dominates the jaw height and that the jaw has mainly a synergistic role. This argues against the proposal that the jaw cycle is a segment-independent determinant of jaw height in the syllable [13]. Also, syllabic liquids rank among the longest nuclei, patterning with /a/, but at the same time they display jaw curvatures and displacements similar to mid-high vowels, i.e. the jaw movement for the liquids is less curved and shows a lesser displacement compared to /a/. This again supports the notion that the jaw serves predominantly a synergistic function for the lingual target (a tongue tip constriction for /l, r/), instead of the jaw dominating any intrinsic jaw position specifications of a given segment, as posited by the jaw cycle hypothesis. Hence, the potential of extracting reliable features identifying syllables based on an independent jaw cycle seems less promising than an effort based on syllable features dependent on lingual constriction degree.

Overall, our initial hypothesis that liquids might show jaw positions similar to vowels in nucleus position was confirmed. While the jaw showed synergistic effects for both vowels and consonants, it was overall still the case that the jaw position for /l, r/ is comparable to the one of mid vowels. This can be interpreted to the effect that liquids may cross-linguistically frequently serve as syllabic consonants because they allow for a relatively open jaw position. This may ensure overall a rising sonority profile from syllable margin to nucleus and a falling profile away from the nucleus, and thus sufficient signal modulation even in vowel-less words to allow for reliable speech segmentation in speech perception.

5. Acknowledgements

Work supported by Alexander von Humboldt fellowship, EPSF project ITMS 26240220060, and Deutsche Forschungsgemeinschaft (PO 1269/1-1).

6. References

[1] Zhang, L. and Edmondson, W.H. "*Speech Recognition Using Syllable Patterns", in* Proceedings of ICSLP, 1237-1240, 2002.
[2] Fujimura, O. "The C/D model and prosodic control of articulatory behavior", Phonetica 57: 128-138, 2000.
[3] Ohala, J.J., Kawasaki-Fukumori, H. "Alternatives to the sonority hierarchy for explaining the shape of morphemes", in S. Eliasson, E.H. Jahr [Eds.], Language and its ecology: Essays in memory of Einar Haugen, 343-365, 1997.
[4] Lindblom, B. "Economy of speech gestures", in P. F. MacNeilage [Ed.], The Production of Speech, 217-246, 1983.
[5] Harrington, J., Fletcher, J., Beckman, M. "Manner and place conflicts in the articulation of accent in Australian English", in M. Broe, J. Pierrehumbert [Eds.], Papers in Lab. Phonology V: Language acquisition and the lexicon, 40-51, 2000.
[6] Erickson, D. "Articulation of Extreme Formant Patterns for Emphasized Vowels", Phonetica 59:134–149, 2002.
[7] Keating, P., Lindblom, B., Lubker, J., Kreiman, J. "Variability in jaw height for segments in English and Swedish VCVs", J. of Phonetics, 22: 407-422, 1994.
[8] Mooshammer, C., Hoole, P., Geumann, A. "Jaw and Order", Language and Speech, 50(2): 145-176, 2007.
[9] Harrington, J. "Phonetic Analysis of Speech Corpora", 2010.
[10] Baayen, R.H. "Analyzing Linguistic Data. A Practical Introduction to Statistics Using R, pp. 241-302, 2008.
[11] Reubold, U., Harrington, J., Kleber, F. "Vocal aging effects on F0 and the first formant: A longitudinal analysis in adult speakers", Speech Communication 52, 638-651, 2010.
[12] Lindblom, B. "Spectrographic study of vowel reduction. J. Acoust. Soc. Am. 35, 1773-1781, 1963.
[13] MacNeilage, P. F., Davis, B. L. "On the origin of internal structure of words", Science, 2888(April): 527-531, 2000.