

Prosodic Phrasing in Lithuanian: Preparatory Study

Sigita DEREŠKEVIČIŪTĖ, Asta KAZLAUSKIENĖ

Vytautas Magnus University, V. Putvinskio str. 23-204, LT-44243 Kaunas, Lithuania

`sigita.dereskeviciute@vdu.lt, asta.kazlauskiene@vdu.lt`

Abstract. The aim of this study is to determine the most significant features of prosodic phrasing which should be analysed in detail in the future. The research material consists of 12 records of the 20th chapter from Antoine de Saint-Exupéry's *The Little Prince*. This text was read by four male actors 3 times. The results show that acoustic cues such as pitch lowering, pausing, segmental lengthening, decrease in intensity, and increase in amplitude perturbations may be manipulated by speakers to distinguish among distinct levels of prosodic boundaries.

Keywords: Lithuanian, prosodic phrasing, phrase, pause, intensity, duration, voice quality

1. Introduction

Prosodic phrasing is grouping of words within an utterance (Jun, 2003). This grouping is influenced by a syntactic and prosodic structure of a language, semantic and pragmatic factors, individual characteristics of a speaker (speech rate, emotions, experience of speaking aloud, etc.). On the other hand, prosodic phrasing can lead to the perception and production of language phenomena: for appropriate clause and sentence processing, prosodic and syntactic boundaries have to match (for research on this relationship, see e.g., Watson, Gibson, 2004; Himmelmann, 2022).

There is no single opinion on the amount of levels of prosodic phrasing; however, two levels are usually distinguished in almost all languages: an intonation phrase (IP) and an intermediate phrase (ip) which is intermediate between an IP and a word. These two levels are in the spotlight of the study. Proper phrasing is very important in speech synthesis and recognition; unfortunately, the Lithuanian language has not been studied on this issue. Therefore, it is very important to find out what basic features can signal an IP and ip boundaries.

It is obvious that pitch (and its acoustics correlates F0) is the most important component of intonation and setting the edges of IP and ip. In addition to pitch, other articulatory and acoustic features may signal boundaries of prosodic units, such as pitch lowering, pausing, segmental lengthening, amplitude (loudness) changes (Peters, 2003; Ulbrich, 2006; Yoon et al., 2007; Michelas and D'Imperio, 2010; Petrone et al., 2017; Brugos et al., 2018; Žygis et al. 2019), non-modal phonation, for example, phrase-final creak (Kohler, 2000; Redi and Shattuck-Hufnagel, 2001; Garellek, 2015; Crowhurst, 2018), etc. Crosslinguistic studies show that these phonetic and acoustic features are not

universal indicators of IP and ip boundaries (Mylberg and Riad, 2015; Kwan Ip and Cutler, 2022).

2. Aim, Material and Methods

This is the initial stage of the study of prosodic phrasing in Lithuanian. The aim of this stage is to determine the most significant features of prosodic phrasing. For this reason, a controlled-read speech was chosen for the analysis. In order to carry out the experiment, four native male actors read the 20th chapter from Antoine de Saint-Exupéry's *The Little Prince*. The text consists of 165 lexemes, about 45 phrases (declaratives, interrogatives, and exclamatives), and it was read 3 times. The actors were asked to read the text clearly, paying attention to the semantical and grammatical connections between words, but without emotional emphasis. They were free to choose the phrasing of the sentences and it did not necessarily have to be the same in all records.

The acoustics data of the investigated segments were measured using Praat (Boersma and Weenink, 2018).

3. Results and Discussion

3.1 Duration of pauses

Speech pauses can be made for a variety of reasons: they can be caused by psychological, physiological, rhythmic, and other reasons. One of the functions of pauses is to divide speech into certain intonation units. Therefore, it is likely that pauses (or prosodic breaks) may be indicators of phrase boundaries.

The results show that pauses are always made after intonation phrases. The actors identified intonation phrases with punctuation marks at the end of a sentence and always made a prosodic break (Kazlauskienė and Kalašinskaitė, 2017). 70% of the intermediate phrases are separated by pauses. It is important to mention that pausing in this case depends on the readers, and the absence of pauses does not indicate the absence of an intermediate phrase.

One of the tasks of this study is to find out whether the duration of the pauses is different after intonation and intermediate phrases. Duration was measured only for those pauses that are at the end of the phrases; pauses before narrow focus were not analysed.

The duration of pauses varies; it is well illustrated by the high standard deviation (see Table 1). However, one tendency is obvious: pauses after intonation phrases are longer (on average 1.8 times) than pauses which separate intermediate phrases. This difference is statistically significant.

Table 1. Duration of pauses

Researched pauses	Median (ms)	AVR (ms)	CI (ms)	<i>t-test</i> ($p_{0.95}=2.0$)	Ratio
Between ip	592	626 (± 283)	588÷665	13.2	1.8
After IP	994	1094 (± 429)	1037÷1151		

3.2 Intensity

In order to find out whether the intensity can signal the end of a phrase, intensity of the whole intonation or intermediate phrase and the last syllable of a phrase was measured.

The results show that the intensity of the last syllable of a phrase is lower than of the whole phrase (see Table 2). On average the whole phrase is produced with 1.1- or 1.2-times higher intensity than the last syllable. Although the difference is small, it is statistically significant. The change in intensity slightly depends on the nature of a phrase; the intensity is slightly lower at the end of intonation phrases than at the end of intermediate phrases.

Table 2. Intensity of phrases and the last syllables¹

Researched unit	Median (dB)	AVR (dB)	CI (dB)	<i>t-test</i> ($p_{0.95}=2.0$)	Ratio
ip	69	70 (± 6)	70÷71	13.6	1.1
σ_{ip}	62	63 (± 7)	62÷64		
IP	68	69 (± 6)	68÷70	14.2	1.2
σ_{IP}	59	59 (± 8)	58÷60		
ip	69	70 (± 6)	70÷71	2.2	1.0
IP	68	69 (± 6)	68÷70		
σ_{ip}	62	63 (± 7)	62÷64	5.1	1.1
σ_{IP}	59	59 (± 8)	58÷60		

We can confirm that intensity decreases at the end of a phrase, especially at the end of intonation phrases (see typical examples in Figure 1).

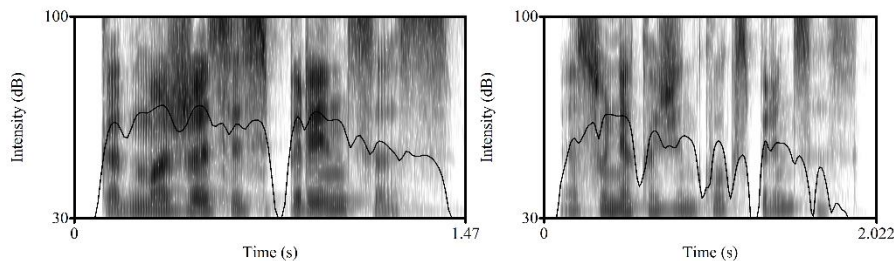


Figure 1. The samples of intensity curves of an intermediate phrase [kəd_mə'zės'is 1'p'ir'intsės] (on the left) and an intonation phrase [e'so n'ė_kəz'it' 1'ko:ks 1'p'ir'intsės] (on the right)

3.3 Duration of the last sound of a phrase

Féry (2017) mentions that a signal of the end of a phrase may be lengthening of the last sound. For this reason, it was decided to find out whether this could be characteristic of Lithuanian as well. The empirical material does not cover all sounds; therefore, only the duration of more frequent sounds [s] and [æ:] was chosen to be presented here. These sounds are very common at the end of Lithuanian words. The duration of the last sounds of a phrase and the last sounds of words in the middle of a phrase was compared.

¹ σ_{ip} marks the last syllable of an intermediate phrase, σ_{IP} marks the last syllable of an intonation phrase.

Consonant [s] at the end of a phrase is almost one and a half times longer than in the middle of a phrase (see Table 3). This difference is statistically significant. However, its duration does not depend on the phrase type, as the consonant is only 1.1 times longer at the end of an intonation pause. This difference is statistically insignificant. Thus in the future, the duration of consonants at the end of a phrase may be investigated irrespective of the type of a phrase.

The duration of long vowels at the end of a word is a very complicated issue due to their frequent shortening. As in case of [s], the duration of [æ:] does not depend on the phrase type. [æ:] is almost 1.4 times longer at the end of a phrase than in the middle of a phrase.

Table 3. Duration of sound at the end and the middle of a phrase²

Researched sounds	Median (ms)	AVR (ms)	CI (ms)	<i>t-test</i> ($p_{0.95}=2.0$)	Ratio
[s] / —	124	129 (±38)	120÷137	1.8	1.1
[s] / —	133	138 (±33)	132÷144		
[s] / —	124	129 (±38)	120÷137	7.5	1.4
[s] / — #	88	90 (±26)	84÷96		
[s] / —	133	138 (±33)	132÷144	10.7	1.5
[s] / — #	88	90 (±26)	84÷96		
[æ:] / —	90	93 (±24)	83÷103	0.5	1:1
[æ:] / —	84	89 (±27)	80÷98		
[æ:] / —	90	93 (±24)	83÷103	4.8	1.4
[æ:] / — #	67	67 (±11)	63÷71		
[æ:] / —	84	89 (±27)	80÷98	3.9	1.3
[æ:] / — #	67	67 (±11)	63÷71		

The comparison of other sounds also shows a similar tendency; consonants and vowels may be lengthened at the end of a phrase. Moreover, our preliminary comparison shows that the long vowels are more tense (on average 1.4 times) in the middle of a phrase than at the end. However, these assumptions still need to be verified with a larger database.

3.4 Voice quality features

For the voice quality analysis, Praat software parameters were selected: the mean of fundamental frequency (F0 mean), jitter (local), shimmer (local), harmonic-noise ratio (HNR), and noise-harmonic ratio (NHR). All these measures provide information about voice signal aperiodicity, stability, noise, and frequency levels. F0 changes at the end of utterances (more specifically, F0 falls) may indicate to listeners the end of an utterance (Ishimoto and Koiso, 2014). Jitter (voice frequency cycle-to-cycle perturbation) is a parameter which evaluates minor glottal pulse irregularities and reflects harsh voice or voice noise in general. Shimmer (voice amplitude cycle-to-cycle perturbation) is the increase of disturbances. It may be related with a decreased or inconsistent vocal fold contact and may reflect noise in voice in general as well. HNR (harmonic noise ratio) and

² / — | marks a sound at the end of an intermediate phrase, / — | marks a sound at the end of an intonation phrase, / — # marks a sound at the end of a word in the mid-phrase position.

NHR (noise harmonic ratio) measures assess the degree of acoustic periodicity and the presence of noise in a voice signal and establish a general perception of hoarseness in a voice signal (Finger at all., 2009; Teixeira at all., 2013; Boersma and Weenink, 2018).

Using Praat's Voice Report command, we manually extracted all five voice quality parameters for the same vowel: 1) at the end of an intonation phrase (||), 2) at the end of an intermediate phrase (|), 3) and at the end of a word in the middle of a phrase (#). The unstressed low, mid and high vowels [ɐ], [ɛ], [æ:], [e:], [i], [u:] were analyzed in the current stage of the research. Table 4 illustrates the results of the voice quality analysis.

Table 4. The means of the voice quality features at boundaries of different prosodic units³

Researched vowel	F0 mean (±Hz)	Jitter (±%)	Shimmer (±%)	HNR (±dB)	NHR ratio
[ɐ] / —	81 (±23)	5.8 (±2)	15.85 (±4)	8.26 (±4)	0.22 (±0.1)
[ɐ] / —	99 (±24)	4.3 (±3)	13.5 (±8)	10.60 (±9)	0.20 (±0.2)
[ɐ] / — #	125 (±23)	2.3 (±2)	7.36 (±4)	10.76 (±5)	0.15 (±0.1)
[ɛ] / —	85 (±12)	4.43 (±4)	18.91 (±7)	9.27 (±4)	0.22 (±0.2)
[ɛ] / —	96 (±30)	5.61 (±4)	15.05 (±7)	9.84 (±4)	0.29 (±0.1)
[ɛ] / — #	124 (±27)	3.62 (±4)	9.32 (±5)	9.95 (±3)	0.20 (±0.1)
[æ:] / —	83 (±14)	4.05 (±3)	14.58 (±6)	9.70 (±3)	0.17 (±0.1)
[æ:] / —	115 (±30)	3.22 (±1)	12.50 (±5)	10.19 (±3)	0.17 (±0.1)
[æ:] / — #	117 (±26)	1.76 (±1)	8.33 (±3)	11.19 (±3)	0.11 (±0.1)
[e:] / —	89 (±14)	3.62 (±3)	12.36 (±6)	9.15 (±4)	0.22 (±0.2)
[e:] / —	103 (±18)	3.28 (±1)	10.97 (±5)	10.58 (±3)	0.15 (±0.1)
[e:] / — #	99 (±17)	1.59 (±2)	6.58 (±3)	11.39 (±3)	0.14 (±0.1)
[u:] / —	84 (±17)	5.87 (±2)	12.12 (±5)	6.13 (±2)	0.37 (±0.4)
[u:] / —	102 (±19)	5.11 (±4)	12.54 (±8)	8.35 (±2)	0.26 (±0.1)
[u:] / — #	111 (±24)	6.48 (±4)	13.05 (±7)	7.20 (±2)	0.28 (±0.1)
[i] / —	87 (±17)	5.32 (±3)	21.00 (±9)	7.02 (±3)	0.32 (±0.1)
[i] / —	104 (±21)	5.04 (±3)	16.50 (±8)	7.63 (±2)	0.24 (±0.1)
[i] / — #	117 (±18)	4.23 (±3)	8.15 (±4)	7.65 (±3)	0.27 (±0.1)

F0 mean. F0 is lower in vowels at the phrase boundaries compared with their phrase-internal counterparts (except the vowel [e:]). F0 declines more if vowels occur at the end of a larger phrase, i.e., an intonation phrase. On average, vowels here are statistically significantly 1.2 times lower than at the end of an intermediate phrase, and statistically significantly 1.4 times lower compared with the vowels in the middle of a phrase. At the end of an intermediate phrase vowels tend to be lower than their phrase-internal counterparts as well; however, differences in mean values of F0 are statistically relevant only for short unstressed vowels [ɐ], [ɛ], [i].

Jitter. Higher jitter values indicate higher frequency perturbations, which in the case of the current analysis are higher at the end of an intonation phrase, smaller at the end of an intermediate phrase and smallest at the end of a word in the middle of a phrase (with some exceptions for the vowels [ɛ] and [u:], where the differences are not statistically significant).

³ Statistical significance was assessed using *t-test*, the significance level - 0.05.

Depending on the vowel and phrase level, the statistical significance varies. The significantly higher perturbation appears in the vowels [ɐ], [æ:], [e:] when uttered at the end of an IP. Compared with their phrase-internal counterparts, they are on average 2.4 times more irregular in frequency. The same vowels also differ substantially (on average 1.6 times) when produced at the end of the ip in comparison to the same vowels produced phrase internally at the end of a word. For the jitter values between boundaries (ip and IP) no statistical significance is observed (except the vowel [ɐ]).

Shimmer. Like jitter, shimmer measurements show the increased amplitude perturbations at the end of larger prosodic units, i.e., the highest shimmer values are in vowels uttered at the end of an IP, smaller at the end of an ip, and markedly decreased at the end of a phrase-internal word (except the vowel [u:]). These differences are statistically important between vowels at the phrase boundaries in comparison to the vowels produced phrase-internally, but not within the boundaries themselves. The values are 2.1 times statistically significantly higher at the end of an IP and 1.7 times statistically significantly higher at the end of an ip compared with their counterparts in the middle of a phrase.

HNR. The harmonicity results show that vowels at the end of utterances tend to be produced with less periodicity (lower HNR values show more F0 irregularity, noise). In the case of Ips, these values are smaller than those of ips, but statistically insignificant. The only relevant difference is between vowels [ɐ] and [e:] uttered at the end of a word in the middle of a phrase and at the end of an intonation phrase. In addition, the low vowels ([ɐ], [æ:], [e:]) show the highest periodicity index in general, whereas the high vowels [ɪ] and [u:] – the lowest.

NHR. Similarly to HNR, NHR measurements assess the amount of noise in a signal: lower values, less noise. The data of the current research shows that the highest values, signalling more noisy production, are in vowels uttered at the end of an intonation phrase. However, the difference is significant only in the cases of the front vowels [e:] and [æ:] (and in relation only to phrase-internal vowels). In other positions NHR is lower and varies. Again, we can see that the high vowels [ɪ] and [u:] are the ‘noisiest’ sounds, whereas the low and middle vowels [æ:] and [e:] are the least noisy.

Reviewing the results, we can see that voice quality features at the boundaries of larger prosodic units (IP) become more salient in comparison with smaller prosodic units (ip and phrase-internal word): F0 declines the most, perturbations in frequency and amplitude increase, periodicity decreases, and noise ratio gets higher.

The evaluation of the statistic differences among vowels at the end of different prosodic units shows that the most reliable features would be F0 mean and shimmer, where 78% and 56% of cases were statistically significant. Jitter, HNR and NHR are less relevant parameters for describing voice quality at the boundaries (33%, 11%, 17%, respectively).

We can assume that these voice quality parameters may indicate non-modal phonation at the boundaries, for example, a creaky phonation (low F0, F0 irregularity, see Figure 2); however, this assumption should be verified by a more thorough analysis in the future.

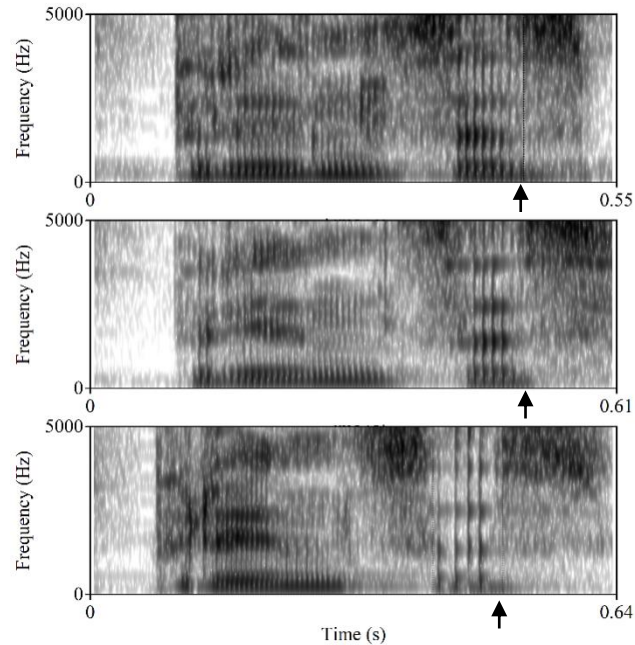


Figure 2. Spectrograms of the vowel [ɐ] (indicated by arrows) in the same word [ˈpʁintsɐs] (En. 'a prince') at the end of a phrase-internal word (top panel), at the end of an intermediate phrase (middle panel), at the end of an intonation phrase (bottom panel). The bottom panel reflects the non-modal phonation (a creaky phonation with irregular and larger spaced vocal pulses).

4. Conclusions

This preparatory study leads to the following conclusions:

- The decrease in intensity is a sign of the end of a phrase.
- A pause is possible but not mandatory indicator of the end of a phrase.
- Sounds can be extended at the end of a phrase.
- F0 lowering and increase in amplitude perturbations are the voice quality features that signal about the end of a larger or smaller prosodic unit most significantly.

It is necessary to mention that not all speakers use the same strategies for marking phrase boundaries. In addition, the junction of phrases can be shown not only by the features of a previous phrase end, but also by the beginning features of the next phrase. These phenomena should be studied in more detail.

Punctuation in Lithuanian is quite complex and is mainly based on syntax. In terms of phrasing and intonation, it has not been studied at all. Therefore, the relationship between phrase boundaries and punctuation should be analysed as well.

In order to integrate the results of prosodic phrasing into the development of speech technologies, they need to be verified on a larger and more varied dataset. Based on these results, the speech corpus which is used for speech technologies will be expanded.

Acknowledgment

This research was funded by a grant (No. S-LIP-21-5) from the Research Council of Lithuania.

References

- Jun, S.-A. (2003). Prosodic phrasing and attachment preferences, *Journal of Psycholinguistic Research*, **32**(2), 219-249.
- Watson, D., Gibson, E. (2004). The relationship between intonational phrasing and syntactic Structure in language production, *Language and Cognitive Processes*, **19**(6), 713-755. DOI: 10.1080/01690960444000070.
- Himmelmann, N. P. (2022). Prosodic phrasing and the emergence of phrase structure, *Linguistics*, **60**(3), 715-743. <https://doi.org/10.1515/ling-2020-0135>.
- Yoon, T.-J., Cole, J., Hasegawa-Johnson, M. (2007). One the edge: Acoustic cues to layered prosodic domains, *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, 1017-1020.
- Žygis, M., Tomlinson, J., Petrone, C., Pfütze, D. (2019). Acoustic cues of prosodic boundaries in German at different speech rate, *Proceedings of ICPhS 2019*, Melbourne, Australia. (hal-02097707)
- Peters, B. (2003). Multiple cues for phonetic phrase boundaries in German spontaneous speech, *Proceedings 15th ICPhS*, Barcelona, p. 1795-1798.
- Brugos, A., Breen, M., Veilleux, N., Barnes, J., Shattuck-Hufnagel, S. (2018). Cue-based annotation and analysis of prosodic boundary events, *Proceedings of the 9th international conference on Speech Prosody 2018*, p. 245-249.
- Michelas, A., D'Imperio, M. (2010). Durational cues and prosodic phrasing in French: Evidence for the intermediate phrase. *Speech Prosody*.
- Petrone, C., Truckenbrodt, H., Wellmann, C., Holzgrefe-Lang, J., Wartenburger, I., Höhle, B. (2017). Prosodic boundary cues in German: Evidence from the production and perception of bracketed lists, *Journal of Phonetics*, **61**, 71-92.
- Kohler, K. J. (2000). Linguistic and paralinguistic functions of non-modal voice in connected speech, *Proceedings of the 5th Seminar on Speech Production: Models and Data*, 89-92.
- Redi, L., Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers, *Journal of Phonetics*, **29**(4), 407-429.
- Garellek, M. (2015). Perception of glottalization and phrase-final creak, *The Journal of the Acoustical Society of America*, **137**(2), 822-831.
- Crowhurst, M. J., 2018. The joint influence of vowel duration and creak on the perception of internal phrase boundaries, *Journal of Acoustic Society of America*, **143**(3).
- Mylberg, S., Riad, T. (2015). The prosodic hierarchy of Swedish, *Nordic Journal of Linguistics*, **38**(2), 115-147.
- Kwan Ip, M. H., Cutler, A. (2022). Juncture prosody across languages: Similar production but dissimilar perception. *Laboratory Phonology*, **13**(1), 1-49.
- Boersma, P., Weenink, D. (2018). Praat: doing phonetics by computer, version 6.035. www.praat.org [computer program].
- Kazlauskienė, A., Kalašinskaitė, E. (2017). Pauzių trukmė skaitomoje lietuvių kalboje, *Kalby studijos = Studies about languages*, **31**, 71-82.
- Féry, C. (2017). *Intonation and Prosodic Structure*, Cambridge University Press, Cambridge.
- Ishimoto, Y., Koiso, H. (2014). Utterance-final F0 changes in Japanese monologs and dialogs, Conference paper of the *Oriental COCOSDA 2014*.

- Finger, L. S., Cielo, C. A., Schwarz, K. (2009). Acoustic vocal measures in women without voice complaints and with normal larynxes, *Brazilian Journal of Otorhinolaryngology*, **75**(3), 432-440.
- Teixeira, J. P., Oliveira, C., Lopes, C. (2013). Vocal acoustic analysis – jitter, shimmer and HNR parameters, *Procedia Technology*, **9**, 1112-1122.
- Ulbrich, C. (2006). Prosodic phrasing in three German standard varieties. *University of Pennsylvania Working Papers in Linguistics*, **12**(1), 361-374.

Received August 2, 2022, accepted August 6, 2022